Sex chromosome and sex determination evolution in African clawed frogs (*Xenopus* and *Silurana*)

Sex chromosome and sex determination evolution in African clawed frogs (*Xenopus* and *Silurana*)

Adam J Bewick, B.Sc. Hons., B.Ed.

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

McMaster University

DOCTOR OF PHILOSOPHY (2012) McMaster University
(Biology) Hamilton, Ontario

TITLE: Sex chromosome and sex determination evolution in African clawed frogs (*Xenopus* and *Silurana*)

AUTHOR: Adam J Bewick, B.Sc. Hons., (Laurentian University), B.Ed. (Laurentian University)

SUPERVISOR: Dr. Ben J Evans

NUMBER OF PAGES: [xi], 142

ABSTRACT

Sex chromosomes have evolved independently multiple times in plants and animals. According to sex chromosome evolution theory, the first step is taken when an autosomal mutation seizes a leading role in the sex determining pathway, such that heterozygotes develop into one sex, and homozygotes into the other. In the second step, sexually antagonistic mutations are expected to accumulate in the vicinity of this gene, benefiting from linkage disequilibrium. Recombination in the heterogametic sex is suppressed because of mutations that eliminate homology between the sex chromosomes, providing epistatic interactions between the sex determining and sexually antagonistic genes. However, suppressed recombination also lowers the efficacy of selection causing accumulation of deleterious mutations. Additionally large segments of non-functional DNA can be deleted in the sex chromosome and can reduce their physical size. Collectively, this leads to divergence between non-recombining portions of each sex chromosome, causing drastic differences at the sequence level and cytologically. However, sex chromosome degeneration is not always the case, and evolutionarily old, and young, but nondegenerate sex chromosomes have been observed. African clawed frogs (*Xenopus* and *Silurana*) have homomorphic sex chromosomes due to a recent turnover event. However, occasional recombination between the sex chromosomes may contribute to the maintenance of homomorphic sex chromosomes in African clawed frogs. Mechanisms that prevent divergence and heteromorphy of sex chromosomes may be related to polyploidization, which is frequently observed in African clawed frogs. The studies herein construct a phylogenetic framework to test alternative hypotheses for selection on sex-linked and autosomal genes involved in sex determination, map sex chromosomes and compare sex chromosomes across African clawed frogs. I have also explored the relationship between phenomena like recent turnover events, recombination and polyploidization to sex chromosome degeneration (or lack thereof). In this dissertation, I have discussed the potential for multiple mechanisms of sex determination and the unique pseudoautosomal nature of sex chromosomes within this group of frogs. This body of work provides a comprehensive study of sex chromosomes in a group lacking phylogenetic resolution, anura, and sheds light on the origin and evolution of sex chromosomes in other organisms.

ACKNOWLEDGEMENTS

My academic career would not exist without Ben Evans. I thank my supervisor for his commitment, relentless encouragement and guidance on successful and many unsuccessful research projects, unprecedented enthusiasm and humour, for inadvertently introducing me to my wife, and for allowing me to have a life outside of the lab.

My committee members Drs. Brian Golding, JP Xu and Jonathon Stone for pushing me to know and do more. Past undergraduate supervisors Drs. Albrecht Schulte-Hostedde and Thomas Merritt for their continual support and encouragement. I am also grateful to Dr. Wilfried Haerty and Wilson Sung for contributing their time to ease my suffering with bioinformatics. Many other students (past and present) have provided support, among them Drs. Melanie Lou, Carlo Artieri and Abha Ahuja, and many more who deserved to be mentioned

Members of the Evans lab that I had the privilege to interact with over many years. Dr. Freddy Chain and Mr. Iqbal Setiadi were among the first to entertain me and I am greatly honoured to have worked along side them. Mrs. Shireen Bliss, Ms. Tia Harrison, Ms. Simone Mendel and Ms. Laura Pin for their help with my frog duties and various research projects. I especially acknowledge Simone for her unwavering work ethic and dedication to research. Master Jane Shen for being brave and trying new things, Hermina Ghenu for teaching me something new every day, and Ben Furman for putting up with my antics and for being an outstanding field companion in South Africa.

My family and friends for their unwavering support and curiosity of my progression in my studies. My family for always reminding me that I still have a thesis to complete. My wife Emily for emotional and academic support, and for leaning on me as much as I have leaned on her.

DECLARATION OF ACADEMIC ACHIEVEMENT

Each chapter of this thesis has been written as a separate manuscript accepted for publication with the exception of Chapter 4, which is a manuscript in preparation. Data collection, programming, analysis and manuscript preparation for each chapter was primarily an individual effort, with contributions in data preparation, analysis and editing from Dave W Anderson, Frédéric JJ Chain, Ben J Evans, Joseph Heled, and from collaborators at the MRC National Institute for Medical Research in London, Charles University in Prague and the Veterinary Research Institute in the Czech Republic.

TABLE OF CONTENTS

List of Figures

List of Tables

Part I

INTRODUCTION

Sex chromosome and sex determination evolution

Sex chromosomes have independently evolved from autosomes multiple times in plants and animals (Ohno, 1967; Bull, 1983; Charlesworth, 1996). According to models of sex chromosome evolution, the first step is taken when an autosomal mutation seizes a leading role in the sex determining pathway, such that heterozygotes develop into one sex, and homozygotes into the other (Ohno, 1967; Charlesworth et al., 2005). In the second step, sexually antagonistic mutations are expected to accumulate in the vicinity of this gene, benefiting from linkage disequilibrium (Bull, 1983; Rice, 1996). Recombination in the heterogametic sex is suppressed because of mutations that eliminate homology between the sex chromosomes, providing epistatic interactions between the sex determining and sexually antagonistic genes (Rice, 1996; Charlesworth and Charlesworth, 2000). However, suppressed recombination also lowers the efficacy of selection, manifested as slow adaptive evolution of sex-linked genes compared with genes on the X or Z chromosomes (Orr and Kim, 1998), and as reduced effectiveness of purifying selection, causing accumulation of deleterious mutations (Agulnik et al., 1997; Fridolfsson and Ellegren, 2000; Filatov et al., 2001; Bachtrog and Charlesworth, 2002; Filatov and Charlesworth, 2002; Wyckoff et al., 2002; Tucker et al., 2003; Berlin and Ellegren, 2006; Kaiser, 2010; Kaiser and Charlesworth, 2010). Additionally large segments of non-functional DNA can be deleted in the sex chromosome and can reduce their physical size (Bachtrog, 2013). Collectively, this leads to divergence between non-recombining portions of each sex chromosome, causing drastic differences at the sequence level and cytologically. The degeneration of sex chromosomes can be exacerbated by time, and the effects of such processes are exemplified by the drastic size and molecular differences between sex chromosomes.

The vertebrate sex-determining pathway is extremely conserved at the molecular and the physiological level, but many different factors trigger it. In many vertebrate lineages, the trigger for the sex-determining pathway is a gene (or genes), and this type of sex determination is termed genetic sex determination (GSD). In other lineages the sex-determining pathway is triggered through an environmental stimulus, most commonly temperature, termed environmental (temperature) sex determination (TSD). GSD can be achieved by a male inducing factor, which defines the male-specific Y chromosome in a male heterogametic system (XX females, XY males), by a female inducing factor, which defines the female-specific W chromosome in a female heterogametic system (ZZ males, ZW females), or through dosage by differences in allelic copy number between the sexes. Few vertebrate sex determining genes have been identified: *SRY* in mammals, *Amhr2* in pufferfish, and *DMRT1* in birds and its homologs in fish and frogs. The reappearance of these genes in vertebrates can be explained by shared ancestry and reemergence in different lineages or by evolution rediscovering the few genes suitable for sex determination (Graves and Peichel, 2010). But these hypotheses are not mutually exclusive. For example, DM domain-containing genes, like *DMRT1*, play a remarkably conserved role as an activator of male differentiation in metazoans including worms, flies, coral, birds, and humans (Burtis and

Baker, 1989; Raymond et al., 1999; Yi and Zarkower, 1999; Raymond et al., 2000; Miller et al., 2003; Haag and Doty, 2005) and this may represent shared ancestry or suitableness of *DMRT1* as a sex determining gene.

Duplication

Duplications of individual genes, chromosomal segments, or entire genomes is suspected to provide the primary material for the origin of evolutionary novelties, including new gene functions and expression patterns (Ohno, 1970; Lynch and Conery, 2000). For example, in the Japanese medaka fish (*Oryzias latipes*), which has an XY system, a duplicated copy of *DMRT1* (*DMY*) defines a novel Y chromosome. This novel Y chromosome is genetically the same as the X chromosome, with the addition of ∼258 kb of sequence that includes the *DMY* gene. *DMY* encodes a fully functional *DMRT1*-like protein and has been shown to be necessary and sufficient to trigger male development (Matsuda et al., 2002; Nanda et al., 2002; Kondo et al., 2006; Matsuda et al., 2007). Also, the only known amphibian sex determining gene, *DM-W*, is a partial duplicate of *DMRT1* on the W chromosome in the frog *Xenopus laevis* (Yoshimoto et al., 2008). In this female heterogametic species, genetic male (ZZ) tadpoles transgenic for *DM-W* were feminized (i.e., phenotypically female), implying that *DM-W* acts as a dominant-negative, antagonizing *DMRT1* activation of male-specific genes by binding to and inhibiting regulatory regions recognized by both proteins (Yoshimoto et al., 2008, 2010).

Duplication also plays an integral role in genome evolution. *In lieu* of previous statements, sex chromosome degeneration is not always observed and many species with evolutionarily old sex chromosomes are morphologically similar, or homomorphic (Matsubara et al., 2006; Tsuda et al., 2007). Processes that deter degeneration are not understood, but polyploidization (Orr, 1990; Evans et al., 2012), frequent turnover of sex chromosomes (i.e., the "high-turnover" hypothesis), or recombination between sex-reversed individuals (i.e., the "fountain-of-youth" hypothesis) could explain the lack of degeneration in some cases (Perrin, 2009; Stöck et al., 2011; Guerrero et al., 2012).

Most species of amphibians have homomorphic sex chromosomes (reviewed in Schmid et al. (2010)). In one group of hylid frogs (genus *Hyla*), genomic regions that are tightly linked to the sex determining locus are not substantially diverged between males and females, indicating that the sex chromosomes of these frogs recombine, at least occasionally (Stöck et al., 2011; Guerrero et al., 2012). Sex chromosomes are estimated to have changed many (∼32) times during amphibian evolution and these changes are evidenced, for example, by considerable variation among and within species in male versus female heterogamy (Evans et al., 2012). These two examples support the fountain-of-youth and high-turnover hypotheses, respectively, helping to explain the abundance of homomorphic sex chromosomes in frogs and perhaps amphibians as a whole. Sex chromosome degeneration also cre-

ates imbalances in allelic copy number between the sexes, which can lead to the evolution of dosage compensation – a factor that is also potentially relevant to genome duplication (Orr, 1990). Dosage compensation is a frequent consequence of sex chromosome degeneration that equalizes expression levels of a gene that has a different number of alleles in each sex. Polyploidization might be less common in species with degenerate sex chromosomes because a degenerate ancestral sex chromosome could segregate as a new autosomal chromosome, and the resulting homozygous null genotypes could be detrimental (Evans et al., 2012). It has also been proposed that dosage compensation in species with a degenerate sex chromosome could act as a barrier to genome duplication, because dosage compensation is disrupted when a newly formed triploid individual backcrosses with a diploid parental individual during the first stages of polyploid speciation (Orr, 1990). Speciation through whole genome duplication is common in African clawed frogs, and species investigated to date all possess homomorphic chromosome pairs. Hence, whole genome duplication and sex chromosome nondegeneration may be interrelated.

Speciation of polyploid frogs

Extant pipoid frogs consist of two sister families: the fossorial Rhinophrynidae and the highly aquatic Pipidae. Together these two families form the superfamily Pipoidea. The family Rhinophrynidae contains one species, the Mexican burrowing toad (*Rhinophrynus dorsalis*), and is found from the southern part of North America (Texas, USA) to Central America (Costa Rica). This species remains predominantly underground and emerges after long periods of rain to reproduce in temporary bodies of water. In contrast, members of Pipidae live a principally aquatic life in still or slow-moving water. A derived morphology facilitates aquatic life, characterized by a lateral line system that persists through metamorphosis; laterally positioned limbs and dorsoventral compression of the body; pelvic modifications that facilitate swimming but impede terrestrial jumping; and modifications of the head and cranium including the loss of a tongue (Trueb, 1996). Due to this last characteristic these frogs are commonly referred to as tongueless frogs. Calls are produced underwater using a unique mechanism of sound production within the larynx consisting of trains of clicks (Yager, 1996). The larynx is exceptionally sexually dimorphic; it is larger in males and sexual dimorphism resonates in song production (Sassoon and Kelley, 1986). There are 5 genera within Pipidae: *Xenopus*, *Silurana*, *Hymenochirus*, *Pseudhymenochirus*, and *Pipa*, which include 19, 2, 4, 1, and 7 described species, respectively (Frost, 2011). *Pipa* is found in South America and the others are found in Sub-Saharan Africa. Although some nodes are unresolved, a consensus has been reached about other aspects of pipid phylogeny (reviewed in Evans (2008)). It is widely accepted that each genus is monophyletic, that (*Xenopus* + *Silurana*) is a clade, that (*Hymenochirus* + *Pseudhymenochirus*) is a clade, and most studies agree that Pipidae is the sister clade to the New World frog family Rhinophrynidae (Kluge and Farris, 1969; Lynch, 1973; Trueb and Cannatella, 1986; Cannatella and Trueb,

1988a,b; Cannatella and de Sá, 1993; Ford and Cannatella, 1993; Graf, 1996; Kobel et al., 1998; Evans et al., 2004; Roelants and Bossuyt, 2005; Frost et al., 2006; Roelants et al., 2007; Irisarri et al., 2011; Pyron and Wiens, 2011; Wiens, 2011).

African clawed frogs comprise >19 (described and un-described) species of frogs native to Sub-Saharan Africa. Phylogenetic relationships among African clawed frog species are characterized by a combination of regular bifurcating speciation, in which one ancestral species splits into two descendant species, and reticulating speciation via allopolyploidization, in which two ancestral species merge into one descendant species (Evans, 2008). As a result, the species tree of African clawed frogs is composed of branches that both split and merge. Speciation by allopolyploidization occurs when interspecies hybridization leads to fusion of their complete genomes into a new descendant species. The genome of a new allopolyploid species is therefore duplicated relative to that of each ancestral species. Therefore, within *Xenopus* there exist at least 12 tetraploid species (4n = 36), 6 octoploids (8n = 72), and 2 dodecaploids (12n = 108). The sister genus *Silurana* includes one diploid (2n = 20) and three tetraploids (4n = 40). Together these two genera form the superfamily Xenopodinae. Given the way allopolyploidization events have occurred autosomal genes are in 2, 4, and 6 paralogous copies in tetraploids, octoploids, and dodecaploids, respectively. However loci on the sex chromosome are not duplicated (reviewed in Evans, 2008).

Goals

The overarching aim of this dissertation is to provide insights into the series of events that gave rise to the origin and evolution of new sex chromosomes in a recent ancestor of African clawed frogs (*Xenopus* and *Silurana*). Pipid frogs are an evolutionarily old group and cladogenesis of genera may have been triggered by continental drift. Resolution of species relationships of pipid frogs is important to understanding the timing of sex determination and sex chromosome evolution. Frequent whole genome duplication via allopolyploidization within African clawed frogs offers a compelling opportunity to investigate the fate of duplicate genes involved in sex determination, and how barriers to sex chromosome divergence and heteromorphy are related to polyploid speciation. Nondegenerate sex chromosomes in African clawed frogs suggests that sex chromosomes are evolutionarily young or that other mechanisms may be operating to maintain the sex chromosomes in a state of homomorphy. Therefore, this system also allows for testing hypotheses related to the initial steps of sex chromosome evolution, prior to degeneration.

Chapter 1

The pipid root

Bewick AJ, Chain FJJ, Heled J, and Evans BJ (2012) *Systematic Biology*, 61(6):913-926

1.1 Preface

Many challenges are presented when molecularly estimating species relationships. To elucidate some of these challenges, and how to overcome them, we set out to resolve a debated, and unresolved, relationship among pipid frogs using high throughput sequencing technology and multilocus species tree estimation methods.

1.2 Abstract

The estimation of phylogenetic relationships is an essential component of understanding evolution. Accurate phylogenetic estimation is difficult, however, when internodes are short and old, when genealogical discordance is common due to large ancestral effective population sizes or ancestral population structure, and when homoplasy is prevalent. Inference of divergence times is also hampered by unknown and uneven rates of evolution, the incomplete fossil record, uncertainty in relationships between fossil and extant lineages, and uncertainty in the age of fossils. Ideally, these challenges can be overcome by developing large "phylogenomic" data sets and by analyzing them with methods that accommodate features of the evolutionary process, such as genealogical discordance, recurrent substitution, recombination, ancestral population structure, gene flow after speciation among sampled and unsampled taxa, and variation in evolutionary rates. In some phylogenetic problems, it is possible to use information that is independent of fossils, such as the geological record, to identify putative triggers for diversification whose associated estimated divergence times

can then be compared *a posteriori* with estimated relationships and ages of fossils. The history of diversification of pipid frog genera *Pipa*, *Hymenochirus*, *Silurana*, and *Xenopus*, for instance, is characterized by many of these evolutionary and analytical challenges. These frogs diversified dozens of millions of years ago, they have a relatively rich fossil record, their distributions span continental plates with a well characterized geological record of ancient connectivity, and there is considerable disagreement across studies in estimated evolutionary relationships. We used high throughput sequencing and public databases to generate a large phylogenomic data set with which we estimated evolutionary relationships using multilocus coalescence methods. We collected sequence data from *Pipa*, *Hymenochirus*, *Silurana*, and *Xenopus* and the outgroup taxon *Rhinophrynus dorsalis* from coding sequence of 113 autosomal regions, averaging ∼300 bp in length (range: 102-1695 bp) and also a portion of the mitochondrial genome. Analysis of these data using multiple approaches recovers strong support for the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology, and geologically calibrated divergence time estimates that are consistent with estimated ages and phylogenetic affinities of many fossils. These results provide new insights into the biogeography and chronology of pipid diversification during the breakup of Gondwanaland and illustrate how phylogenomic data may be necessary to tackle tough problems in molecular systematics.

1.3 Introduction

Estimation of phylogenetic relationships among species using molecular data must overcome challenges associated with estimating individual genealogies (e.g., phylogenetic error stemming from short branch lengths or homoplasy) and challenges associated with discordance between gene trees and species trees. Differences between gene trees and species trees can have a biological basis, including ancestral polymorphism, simultaneous divergence of multiple species (hard polytomies), and other phenomena such as balancing selection, gene conversion, horizontal gene transfer, interspecies hybridization, and allopolyploidization (Ioerger et al., 1990; Maddison, 1997; Brooks and McLennan, 2002; Evans, 2008; Degnan and Rosenberg, 2009). When ancestral population size is large, structured, or when the time between internodes is brief, genealogical discordance with a species tree can be common. Gene trees and species trees are similar when internal branch lengths are on the order of $\sim 5 \times 2N_e$ generations (Degnan and Salter, 2005) but tend to differ when internal branch lengths are brief and when population size is large. Strikingly, for some asymmetrical species trees with short internodes, the most likely gene tree may not be the same as the species tree (Degnan and Rosenberg, 2006) because the probability of a symmetrical (balanced) tree is greater than an asymmetrical (unbalanced) one (Rosenberg, 2002). When the most likely gene tree is different from the species tree, the gene tree is said to be "anomalous" (Degnan and Rosenberg, 2006). Mutational variance increases the parameter space that anomalous gene trees are observed, although this effect is offset by

many anomalous trees being unresolved because of short internal branch lengths (Huang and Knowles, 2009). Disparities between gene trees and species trees can also have an analytical basis, including phylogenetic error (Hillis et al., 1994) and inappropriate homology statements due to gene duplication or incorrect sequence alignment (Wong et al., 2008). Together, these factors can lead to incorrect inferences of topology, branch lengths, credible intervals, and parameter values of species trees.

Another challenge to understanding evolutionary history is the estimation of divergence times. Molecular clocks that use fossils for calibration generally rely on (i) an inference of phylogenetic affinities between a fossil and the group of extant species and (ii) an inference of the age of the fossil. Even if phylogenetic affinities and ages of fossils are well characterized, in most situations a fossil provides only a minimum (most recent) divergence time for a particular node. Geological calibrations suffer from similar challenges and often provide minimum divergence times based on an assumption that diversification was either triggered by or preceded some event, such as the drifting apart of continental plates. Geological calibrations can potentially provide a maximum (most ancient) limit to the timing of divergence. For example, ignoring ancestral polymorphism, divergence of a terrestrial island endemic from a sister lineage on a continent could be assumed to have occurred more recently than the age of the island. Ancient DNA provides a third tool for calibration of molecular clocks. This information is perhaps optimal if its source is a direct ancestor of an extant species – a condition rarely met by most studies. In general, resources available for calibration of many groups are sparse, and the resulting calibrations can be vague or misleading.

An interesting example of these challenges is presented by the diversification of genera of the frog family Pipidae. Phylogenetic relationships among these taxa are unresolved and their resolution faces many of these topological and temporal challenges despite a relatively rich fossil record and a putative role for well-timed geological events in their diversification. There are 5 pipid genera: *Xenopus*, *Silurana*, *Hymenochirus*, *Pseudhymenochirus*, and *Pipa*, which include 19, 2, 4, 1, and 7 described species, respectively (Frost, 2011). *Pipa* occurs in South America and the others are found in sub-Saharan Africa. Pipids are suspected to have drifted apart $\sim$100 Ma (Figure 1.3; (Cannatella and de Sá, 1993; Roelants et al., 2007)). Various studies have supported the (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) topology (Roelants and Bossuyt, 2005; Roelants et al., 2007; Irisarri et al., 2011; Pyron and Wiens, 2011; Wiens, 2011), the (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) topology (Frost et al., 2006), and the ((*Hymenochirus*, *Pipa*)(*Xenopus*, *Silurana*)) topology (Báez and Pugener, 2003; Evans et al., 2004, 2005; Trueb et al., 2005; Trueb and Báez, 2006). These studies are not independent because many use data from the same genes, notably mitochondrial DNA and the autosomal locus *RAG1*. Nonetheless, these contradictions raise the possibility that internodes at the base of the pipid phylogeny are short, that the ancestral population sizes of pipid lineages were large or some combination.

Although some nodes are unresolved, a consensus has been reached about other as-

pects of pipid phylogeny (reviewed in Evans, 2008). It is widely accepted that each genus is monophyletic, that (*Xenopus* + *Silurana*) is a clade, that (*Hymenochirus* + *Pseudhymenochirus*) is a clade, and most studies agree that Pipidae is the sister clade to the New World frog family Rhinophrynidae (Kluge and Farris, 1969; Lynch, 1973; Trueb and Cannatella, 1986; Cannatella and Trueb, 1988a,b; Cannatella and de Sá, 1993; Ford and Cannatella, 1993; Graf, 1996; Kobel et al., 1998; Evans et al., 2004; Roelants and Bossuyt, 2005; Frost et al., 2006; Roelants et al., 2007; Irisarri et al., 2011; Pyron and Wiens, 2011; Wiens, 2011). An analysis of morphological characters by Cannatella and Trueb (1988a) and Cannatella and Trueb (1988b) suggested that *Xenopus* is sister to a clade containing (*Silurana* + *Hymenochirus* + *Pipa*) but a reevaluation of these characters and molecular data by Cannatella and de Sá (1993) found (*Xenopus* + *Silurana*) to be monophyletic with respect to *Hymenochirus*. A phylogeny illustrating resolved and unresolved relationships among pipid frogs is presented in Figure 1.1, along with names for resolved nodes following Cannatella and de Sá (1993). As pointed out by Frost et al. (2006), resolution of phylogenetic relationships among pipids may boil down to ascertaining where the root is in only 1 of the 3 possible unrooted topologies for *Xenopus*, *Silurana*, *Hymenochirus*, and *Pipa* (topology 1 in Figure 1.1).

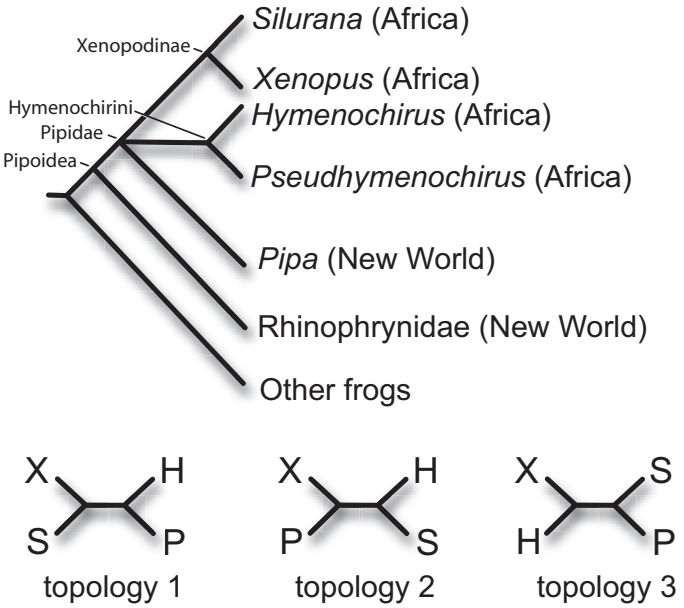

Figure 1.1: Monophyly of *Silurana* + *Xenopus*, of *Hymenochirus* + *Pseudhymenochirus*, and of Pipidae + Rhinophrynidae has been established based on morphological and molecular data, but relationships among 3 major lineages within the frog family Pipidae remain unresolved. Node-based names of clades are indicated on the topology. Topologies 1-3 are the 3 possible unrooted topologies for *Silurana* (S), *Xenopus* (X), *Pipa* (P), and *Hymenochirus* (H); the available data supports unrooted topology 1. The genus *Pseudhymenochirus* and the family Rhinophrynidae both include only one species (*Pseudhymenochirus merlini* and *Rhinophrynus dorsalis*, respectively).

With an aim of further exploring the controversial phylogenetic relationship of pipid frogs, we assembled a large multilocus data set using a combination of 454 pyrosequencing of cDNA and publicly available sequence data. We analyzed each locus individually and also jointly using coalescent-based approaches that accommodate differences between gene trees and species trees. We used a multilocus coalescent-based method (*BEAST) to develop a temporal framework for diversification based on the hypothesis that continental drift triggered cladogenesis of one or more lineages in this group. We then used the relatively rich fossil record of these frogs to evaluate a posteriori the plausibility of these proposed geological mechanisms for diversification. We uncovered a high degree of genealogical discordance among loci, underscoring the utility of phylogenomics and non-concatenated analyses for contextualizing the evolutionary history of this group.

1.4 Materials and methods

1.4.1 Taxon sampling and multilocus data

We collected multilocus sequence data from the pipid frogs *Xenopus laevis*, *Silurana tropicalis*, *Hymenochirus curtipes*, *Pipa carvalhoi*, and the outgroup taxon *Rhinophrynus dorsalis*. Data from *X. laevis* and *S. tropicalis* were obtained from GenBank; most data from the other species were obtained by 454 pyrosequencing of cDNA. An *H. curtipes* individual and a *P. carvalhoi* individual of unknown geographic origin were obtained from animal suppliers (PetSmart and Xenopus Express, respectively), and a sample of *R. dorsalis* was obtained from the tissue archive at the Museum of Vertebrate Zoology at the University of California at Berkeley. We also included data from 6 autosomal loci and one mitochondrial region that were available in GenBank.

For 454 pyrosequencing, total RNA was extracted using QIAGEN RNeasy Mini Kit from liver tissue. Normalized double-stranded cDNA was prepared using SMART cDNA Synthesis Kit (Clontech) and Advantage 2 PCR Kit, and TRIMMER cDNA Normalization Kit (Evrogen), according to the manufacturer s protocol and as described in Chain et al. (2008). The normalized cDNA was used for 454 pyrosequencing (ROCHE GS FLX). A titration run was performed for four 1/16th portions of a plate for each of 3 species (*H. curtipes*, *P. carvalhoi*, and *R. dorsalis*) and an additional 1/4th portion of a plate with the appropriate titration was then run for each species.

Contigs were assembled from the 454 data using GS De novo Assembler (Roche). We averaged 5, 20, and 29 reads per locus for *R. dorsalis*, *P. carvalhoi*, and *H. curtipes*, respectively, and single reads were used for portions of some loci for some species (Supplementary Table 1.S1, Dryad database: doi:10.5061/dryad.bt92r9f1). The lower coverage of *R. dorsalis* is probably related to the lower quality of cDNA synthesized from this old flash frozen sample. Single nucleotide polymorphisms in the resulting contigs were encoded us-

ing International Union of Pure and Applied Chemistry (IUPAC) nucleotide codes. Custom Perl scripts and MUSCLE version 3.6 (Edgar, 2004) were used to align the data to *X. laevis* and *S. tropicalis* sequences from GenBank, retaining as putative orthologs those sequences that all had the reciprocal best BLAST hit (Altschul et al., 1997). Our previous studies used this approach to identify putative singletons and duplicates in *X. laevis* (Chain and Evans, 2006; Chain et al., 2008, 2011), and the resulting data sets are highly concordant with other independent studies (Morin et al., 2006; Hellsten et al., 2007; Sémon and Wolfe, 2008) and with NCBIs UniGene database. For *X. laevis*, a tetraploid species, we identified 2 paralogs for most loci. In order to have the same number of terminals for all loci in our analysis, we randomly selected one *X. laevis* paralog to be included. Both *X. laevis* paralogs are monophyletic with respect to other pipid genera because they formed by genome duplication in *Xenopus* that occurred after divergence from *Silurana* (Evans et al., 2005; Evans, 2007; Bewick et al., 2011). Thus, discarding one paralog from *X. laevis* should have no impact on our inferences of evolutionary relationships among pipid genera.

When analyzing sequences from species as diverged as those in this study, homology statements (alignment) of sequence data can be challenging. For this reason, all noncoding sequences from the 5′ or 3′ untranslated region were discarded and alignments were adjusted manually with MacClade version 4.08 (Maddison and Maddison, 2000) using codon frame inferred from complete transcripts of *Xenopus* and *Silurana* to facilitate homology statements. We retained only alignments with complete information from all 5 focal taxa and a length of at least 99 bp that was devoid of stop codons in all taxa. The only gaps in the alignments in the nuclear data are due to amino acid insertion/deletion polymorphisms, which were infrequently encountered.

One concern is that some of the data partitions could inadvertently include paralogous instead of orthologous sequences. Because we have large databases from *X. laevis* and *S. tropicalis*, we can be fairly confident that the reciprocal best BLAST hits between these species are orthologous. However, as illustrated in Figure 1.2 A-C, it is conceivable that we included nonorthologous sequences from *Pipa*, *Hymenochirus*, or *Rhinophrynus* due to gene duplication or missing data. Even without missing data, it is also conceivable that incomplete lineage sorting could cause a *R. dorsalis* sequence to be an inappropriate outgroup, even though this species is sister to pipids (Figure 1.2D). One possible signal of non-orthology or incomplete lineage sorting in the outgroup would be an atypically low or high divergence between one or more ingroup sequence(s) and the outgroup sequence compared with the other ingroup sequences (Figure 1.2D). To identify alignments that potentially contain nonorthologous sequences, we therefore calculated a JukesCantor corrected pairwise distance between each ingroup sequence and the outgroup sequence using PAUP* (Swofford, 2022). We then divided each of these 4 distances by their maximum to generate a "standardized outgroup distance ratio" for each ingroup taxon. With equal rates of evolution among orthologous sequences, each standardized outgroup distance ratio should be near one, and departures from one would be due to stochastic variation in mutation. If nonorthologous sequences are present, one or more standardized outgroup distance ra-

tios should be low. For example in Figure 1.2A, the standardized outgroup distance ratio between *Hymenochirus* and the outgroup would be lower than those between *Xenopus*, *Silurana*, or *Pipa* and the outgroup. Likewise in Figure 1.2B, the standardized outgroup distance ratio between *Hymenochirus* and the outgroup would be the highest (and therefore equal to one) and the others would be much lower than one. Thus, as a conservative measure, we removed from the analysis all loci with a standardized outgroup distance ratio less than 0.4. This cutoff was arbitrary and is independent of the phylogeny supported by each locus. By imposing this cutoff, we excluded 14 data partitions with high similarity between at least one ingroup sequence and the outgroup sequence compared with the other ingroup sequences. The autosomal portion of the data set thus comprised portions of 114 loci, which are treated as 113 loci because the tightly linked genes *RAG1* and *RAG2* were treated as a single locus.

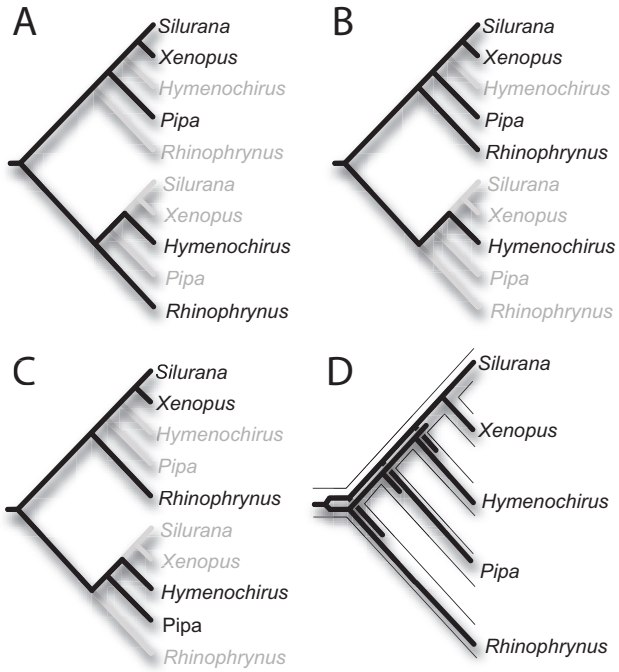

Figure 1.2: Gene family extinction or missing data could generate misleading phylogenetic affinities. In (A-D), the example "true" phylogeny is (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))). However, as a consequence of unsampled orthologs in (A) and (B), rooting the phylogeny with the *Rhinophrynus* sequence suggests that *Pipa* is sister to (*Xenopus* + *Silurana*). In (C), rooting the phylogeny with the *Rhinophrynus* sequence suggests that (*Hymenochirus* + *Pipa*) and (*Xenopus* + *Silurana*) are reciprocally monophyletic. In (D), ancestral polymorphism causes the topology rooted with *Rhinophrynus* sequences to be (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))). In (A-D), divergence between *Rhinophrynus* and some ingroup taxa is lower than that between *Rhinophrynus* and other ingroup taxa, an observation we exploit with the standardized outgroup distance ratio.

We also included in our analysis portions of a ∼2400 bp sequence from the 12S and 16S rDNA genes and the $tRNA_{val}$ of the mitochondrial genome (Evans et al., 2004). To facilitate unambiguous alignment, we discarded loop regions of these rDNA enzymes based on an analysis of secondary structure (Cannone et al., 2002). To avoid violation of the assumption of independent evolution of each site, we discarded half of the stem sequences that paired with the retained portion of the data set, leaving a total of 432 bp. These mitochondrial data are therefore composed exclusively of independently evolving stem region sites that do not form doublets with each other. Six sites in the mitochondrial DNA alignment had gaps as a result of length variation at the junction with a loop region.

1.4.2 Gene tree estimation

We estimated phylogenies from each locus independently with Bayesian analysis as implemented by MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) using a model of evolution selected by the Akaike Information Criterion (AIC) with MrModelTest version 2 (Nylander, 2004). Similar to the BEST analyses described below, these models were not partitioned by codon position. Two Markov Chain Monte Carlo (MCMC) runs were performed with 4 chains per run, each for 10 million generations, with the temperature parameter set to 0.2. Based on inspection of the posterior likelihood surface with Tracer version 1.5 (Drummond and Rambaut, 2007), a burn-in of 2 million generations was discarded for all individual locus analyses. The effective sample size (ESS) of the post burn-in parameter MCMC sampling was calculated using Tracer and an ESS value >200 was used as an indication that the MCMC sample had converged on the posterior distribution. These and other computationally intensive analyses were performed on the SHARCENT computer cluster (www.sharcnet.ca) and also the computer network of Brian Golding at McMaster University.

1.4.3 Concatenated analysis

Under situations where discordance between gene trees and a species tree is expected, concatenation of sequence data with different evolutionary histories can distort branch lengths, relationships, and credibility intervals recovered from phylogenetic analysis (Kubatko and Degnan, 2007). Nonetheless, for comparative purposes, we performed a partitioned analysis on the concatenated data using the preferred model of evolution for each locus as ascertained with the AIC. Two MCMC runs, each with 4 chains for 20 million generations, were performed. A burn-in of 2 million generations was discarded and convergence assessed based on the Tracer analysis discussed above.

1.4.4 Species tree estimation – BEST and *BEAST

We used Bayesian estimation of species trees (BEST) version 2.3 (Liu and Pearl, 2007) and *BEAST pre-release version 1.7.0 (Drummond and Rambaut, 2007; Heled and Drummond, 2010) to estimate a species tree from the multilocus data. These methods assume that discrepancies between gene trees and the species tree are due exclusively to lineage sorting, free recombination between genes, no recombination within genes, and no gene flow after speciation. In both of these analyses, the tree topologies are unlinked across partitions (i.e., estimated independently for each gene) and a species tree is estimated from these potentially discordant tree topologies.

For both analyses, we evaluated 3 models of evolution using Bayes factors as described by (Nylander et al., 2004). For the BEST analysis, the first model (hereafter JC + Γ + d) employed an equal transition rate among nucleotides, with a Γ – distributed rate heterogeneity and base frequencies estimated from the data with a Dirichlet prior. The second model (hereafter HKY + Γ + d) was the same as the first except that the substitution rates between transitions and transversions were estimated separately. The third model (hereafter GTR + Γ + d) is the same as the first except that the transition rates between each nucleotide were estimated separately. For all analyses, the parameter estimations were performed independently for each partition and the gene mutation prior was set at (0, 114) to conservatively allow for a different mutation rate for each partition. The prior for the effective population size parameter theta (θ) was inverse gamma (3, 0.018) based on silent site nucleotide diversity in *X. laevis* (Bewick et al., 2011) as suggested in the program documentation. Depending on the model, multiple independent MCMC runs were performed (37 for JC + Γ + d, 33 for HKY + Γ + d, and 41 for GTR + Γ + d),each with 2 chains that were initiated at different starting seeds for at least 40 million generations, sampling every 2000 generations, with the temperature parameter set to 0.20, *R. dorsalis* set as the outgroup, the mtDNA partition set as haploid, and the other partitions set as diploid. To assess convergence of the MCMC runs, we first inspected a plot of the posterior distribution of likelihoods from each independent MCMC chain and based on this inspection, we discarded 4-70 million generations as burn-in. We then calculated the ESS of the post burn-in parameter values from all runs using Tracer and convergence of the MCMC run on the posterior distribution was again assumed when ESS values exceeded 200.

We also used *BEAST pre-release version 1.7.0 (Drummond and Rambaut, 2007; Heled and Drummond, 2010) to estimate a species phylogeny and provide estimates of divergence times based on geological calibration points enumerated below. We again set out to consider 3 models of evolution using the approach of Nylander et al. (2004) but for this analysis the models were more complex. All the *BEAST models estimated parameters for each codon position of each partition separately. They differed in whether a single rate was used for nucleotide transitions (hereafter JC codon), 2 separate rates for transitions and transversions (hereafter HKY codon), or separate rates for all transitions and transversions (hereafter GTR codon). Because the mtDNA data were from stem regions only, we did

not use a codon model on this partition. For all models, base frequencies were assumed to follow the empirical proportions. A correction term was applied to the default pure birth (Yule) prior, ensuring that the marginal density of the combined prior was identical to the calibration density (Heled and Drummond, 2011). This is in contrast to the default BEAST construction which results in a prior that neither preserves the calibration densities nor the Yule prior (Heled and Drummond, 2011). The correction terms are described in further detail in Supplementary information for each of the 3 calibration regimes discussed below. The prior on birth rate x was set to the uninformative 1/x with a hard bound of (0.0007,1) and the prior on the evolutionary rate y was set to the uninformative 1/y. Each partition had one rate parameter and 3 "subrates" for each codon position, with either the JC, HKY, or GTR model for each codon position. *R. dorsalis* was set as the outgroup by enforcing monophyly of Pipidae. Ploidy of mitochondrial DNA was set to haploid and all other loci to diploid and a strict molecular clock was assumed. For each *BEAST analysis, dozens of independent runs were performed, each for at least 200 million generations. A burn-in of 5-80 million generations was discarded based on inspection of the posterior likelihoods using Tracer. Convergence was evaluated using ESS values after discarding burn-in generations that were identified by eye as described above.

We considered 2 geological calibration points in the *BEAST analysis (Figure 1.3). The first hypothesizes that divergence of *Pipa* from other pipids was triggered by the rifting apart of South America and Africa. From the Late Jurassic through the Middle Cretaceous, South America and Africa were a continuous landmass called West Gondwana (Smith et al., 1994; Gheerbrant and Rage, 2006). Based on dates provided by previous studies (Pitman et al., 1993; Maisey, 2000; McLoughlin, 2001; Sereno et al., 2004; Ali and Aitchison, 2008), we assigned a normal prior for this divergence time with a mean of 102 Ma and a standard deviation of 7 myr in order to accommodate uncertainty in this estimate. The second calibration point hypothesizes that pipoid divergence into the families Rhinophrynidae and Pipidae was triggered by the opening of the North Atlantic Ocean as North America rifted apart from West Africa (Duellman and Trueb, 1994). This occurred during the Late Jurassic (McHone and Butler, 1984) and we again used a normal prior for this divergence time with a mean of 190 Ma and the same standard deviation (7 myr) in order to accommodate uncertainty (Withjack et al., 1998). We performed 3 *BEAST analyses – one with each of these calibration points (*BEAST analysis 1 and 2, respectively) and one with both (*BEAST analysis 3).

1.5 Results

Using data from GenBank and 454 pyrosequencing, we generated 114 5-taxon alignments for *X. laevis*, *S. tropicalis*, *H. curtipes*, *P. carvalhoi*, and the outgroup taxon *R. dorsalis* for a combined total of 35,673 bp per taxon. This included sequences from the coding region of 114 nuclear loci (2 of which were linked and thus concatenated) and a portion of

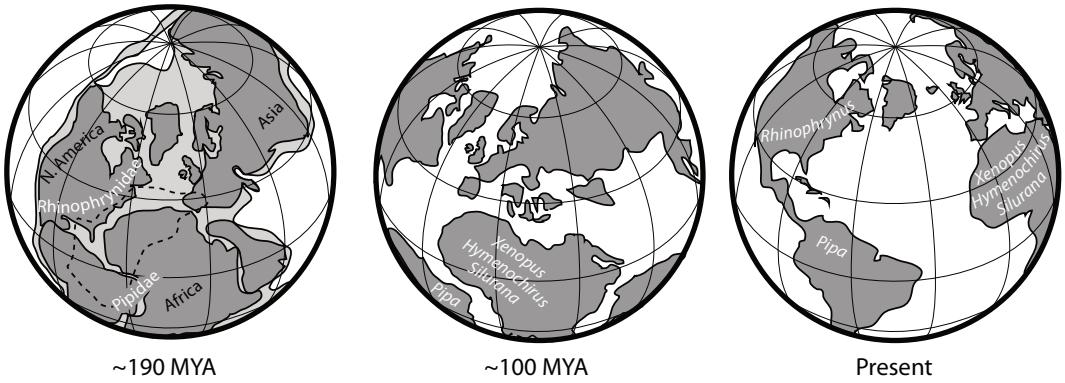

Figure 1.3: Evolution of the Atlantic ocean and of pipoid frogs: our geological calibration points hypothesize that cladogenesis of pipoids was influenced by continental drift. Roughly 190 Ma, divergence of Rhinophrynidae and Pipidae may have been triggered by the opening of the North Atlantic Ocean and consequent separation of North America and West Africa. The Central Atlantic Magnetic Province associated with this event is depicted with dotted lines and putative land positive regions in light grey following Korte et al. (2009) and McHone (2000). About 100 Ma, divergence of *Pipa* from African pipids may have been triggered by the rifting apart of South America and Africa. Pipoid genera now occur in North America, South America, and Africa.

rDNA sequence from the mitochondrial genome in which sites evolve independently. Each data partition has sequence from all 5 taxa and, therefore, there are no missing data in this study. A total of 9,106 positions (25.5%) were variable including 5,940 (16.6%) that were parsimony uninformative and 3,166 (8.9%) that were parsimony informative. Fifty-eight nucleotides in the data set were encoded with ambiguous IUPAC symbols. A summary of variation in partition sequence length is presented in Figure 1.4A and additional information on each partition and GenBank accession numbers for sequences >200 bp in length are available in Supplementary information. Input files for BEST and *BEAST that include all sequences are available in the Dryad database: doi:10.5061/dryad.bt92r9f1.

Figure 1.4B illustrates how the minimum standardized outgroup distance ratio varied among the different data partitions with respect to the length of each partition. Shorter alignments tended to have higher variance in standardized outgroup distance ratios. Higher variance is expected for shorter alignments, but this could also suggest that non-orthologous sequences are more prevalent in the shorter alignments.

1.5.1 Analysis of individual and concatenated loci

The posterior distribution of a phylogenetic tree can be summarized either with the posterior distribution of tree topologies or with the posterior distribution of clades (Sukumaran and Linkem, 2009). The former is accomplished by counting how many times each possible

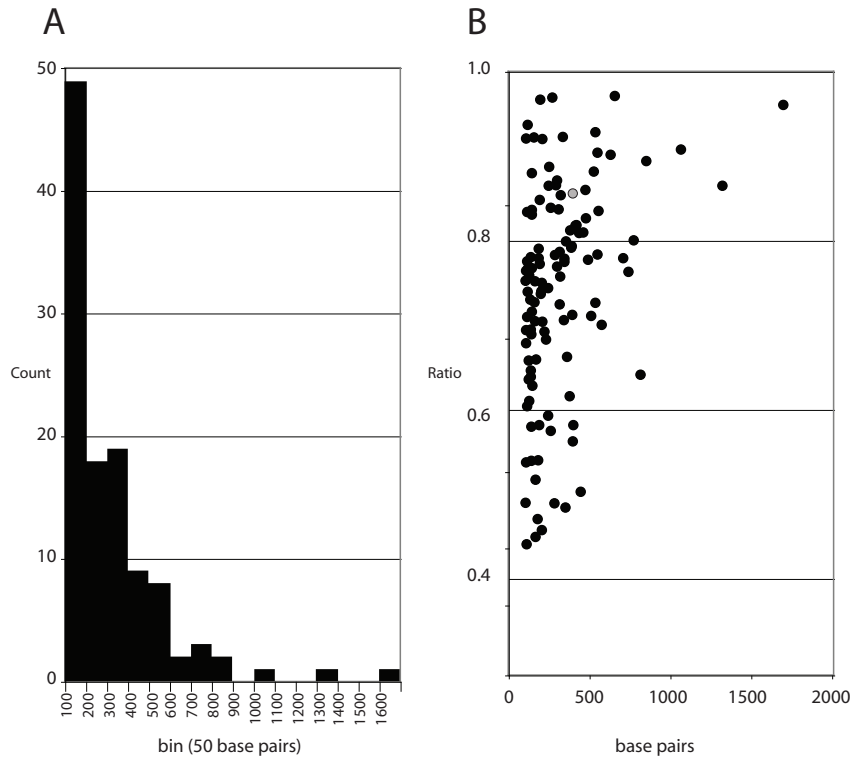

Figure 1.4: Descriptive information about data analyzed in this study. (A) Distribution of the sizes of the data partitions. (B) The standardized outgroup distance ratio, that is the ratio of the lowest Jukes-Cantor-corrected pairwise distances between each ingroup taxon and the outgroup divided by the highest pairwise distance, as a function of the size in base pairs of each partition. In (B), mitochondrial DNA is indicated with a gray circle.

topology appears in the posterior distribution with the most common topology being the maximum a posterior probability tree (hereafter the MAP tree). The latter is accomplished by counting how many times each possible clade appears in the posterior distribution and is commonly summarized with a majority rule consensus topology. Most studies focus on the posterior distribution of clades, in part because many distinct tree topologies are present in posterior distributions of analyses with many taxa. For this reason, the majority rule consensus topology may differ from the MAP tree. Here, we discuss both types of summaries because the posterior distribution includes only 15 possible topologies due to the small number of ingroup taxa.

Figure 1.5 summarizes the posterior distribution of tree topologies recovered from individual Bayesian analysis of each of the 114 data partitions. Analysis of individual loci illustrates that 3 of the 5 possible rootings of topology 1 in Figure 1.1 are far more probable than the other 2 possible rootings of topology 1 or than any of the rootings of topologies 2 and 3. Of the 5 possible rootings of topology 1 in Figure 1.1, only 3 – ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)), (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))), and (*Hymenochirus* (*Pipa*

(*Xenopus*, *Silurana*))) – account for 80.5% of the combined post-burn-in posterior probability distribution of topologies across all individual loci. The breakdown of this distribution is 28.8% for the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology, 26.7% for the (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) topology, and 25.0% for the (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) topology. The remaining 19.5% consisted of the other 12 possible rooted topologies.

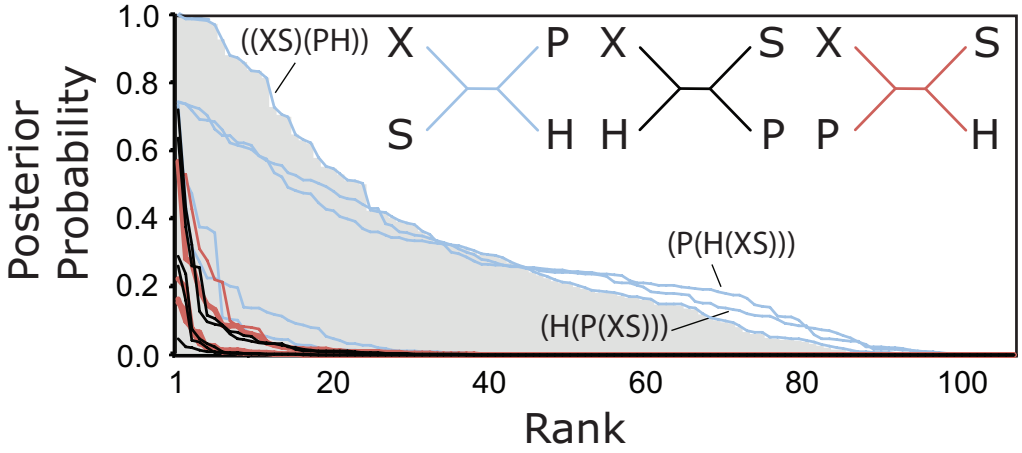

Figure 1.5: Posterior probability distribution of each of the 15 possible rooted topologies recovered from analysis of 114 individual alignments. For each rooted topology, the posterior probabilities (expressed as a decimal) from each partition were ranked from highest to lowest. Each set of 5 distributions are shaded according to their corresponding unrooted 4-taxon topology. The distribution for ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) is shaded to emphasize that this topology is the only one that has strong support (posterior probability >80) at multiple partitions. The next most strongly supported topologies are (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) and (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))). Letters refer to taxa as defined in Figure 1.1.

The MAP tree was ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) for 39 loci, (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) for 28 loci, and (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) for 27 loci. However, the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology comprises >80% of the posterior topology distribution for 13 partitions, but the (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) and (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) do not comprise >80% of the posterior topology distribution for any partition (Figure 1.5). For the mitochondrial DNA partition, the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) comprised 83.7% of the posterior topology distribution, which is consistent with phylogenetic analysis of a larger portion of mitochondrial DNA sequence that includes rDNA stem and loop regions (Evans et al., 2004). Overall then, analysis of individual partitions suggests that the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology has the highest posterior probability.

We contrasted the MAP tree with the majority rule consensus topology for each partition. The ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) consensus topology was recovered

for 29 loci (including mtDNA), the (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) consensus topology was recovered for 29 loci, and the (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) consensus topology was recovered for 22 loci. The consensus topologies of 34 loci were either unresolved with respect to the trichotomy depicted in Figure 1.1 or were one of the other 12 possible topologies. The difference in support for the top 3 rooted topologies was not due to differences in the length of the alignments. The mean length of the loci with the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) consensus topology was 337 bp, whereas the mean length of the loci with the (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))) or (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) consensus topologies was 381 and 334 bp, respectively. However, support for the other possible rooted topologies was probably related to differences in information content – the mean length of 34 loci with either an unresolved consensus tree or a consensus topology that was not ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)), (*Pipa* (*Hymenochirus* (*Xenopus*, *Silurana*))), or (*Hymenochirus* (*Pipa* (*Xenopus*, *Silurana*))) was only 221 bp. A concatenated MrBayes analysis of all partitions also produced a consensus topology with 100% posterior probability for the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology.

1.5.2 Multilocus coalescent analysis

Using Bayes factors, the GTR + Γ + d was strongly favored for BEST analysis (Table 1.1) based on the criterion of Kass and Raftery (1995). Using this model, BEST analysis of the full data set recovered the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) consensus topology with 100% posterior probability for the (*Xenopus*, *Silurana*) clade and 95% posterior probability for the (*Pipa*, *Hymenochirus*) clade.

Table 1.1: Statistics on coalescent-based analyses of species trees (BEST and *BEAST Analyses 1-3) including the model, the post burn-in log likelihood (-LnL), the number of states sampled (# States), 2log(Bayes Factor), the sample frequency (Sampling), the ESS of the likelihood (LnL ESS), and the number of parameters that had ESSs greater than 200 (ESS >200), between 100 and 200 (100 <ESS <200), and less than 100 (ESS <100)

BEST model	-LnL	# States	2log(Bayes Factor)	Sampling	LnL ESS	ESS >200	100 <ESS <200	ESS <100
JC + Γ + d	107,838	527,788,000	—	2000	400	56	134	45
HKY + Γ + d	105,564	885,066,000	4548	2000	961	147	89	0
GTR + Γ + d	105,203	1,336,892,000	722	2000	1389	241	0	0
*BEAST Analysis 1								
JC codon	101,887	1,935,300,000	—	25,000	23,962	931	0	0
HKY codon	97,481	2,009,100,000	8812	25,000	17,590	1271	0	0
*BEAST Analysis 2								
JC codon	101,165	1,929,025,000	—	25,000	25,116	931	0	0
HKY codon	97,482	2,147,475,000	7366	25,000	22,873	1271	0	0
*BEAST Analysis 3								
JC codon	101,867	1,987,000,000	—	25,000	24,236	931	0	2
HKY codon	97,481	2,147,475,000	8812	25,000	21,137	1271	0	0

Note: Values are not reported for the GTR codon model in *BEAST Analyses 1-3 because convergence was not achieved at most of the 2631 parameters.

Our *BEAST analyses using the GTR codon model failed to converge for multiple parameters, even after sampling a large number of states in multiple independent runs. Thus, we focus our discussion on the next most complex model (HKY codon) in which analysis the MCMC chain converged on the posterior distribution for all parameters for each of the 3 differently calibrated analyses (Table 1.1). This model was preferred over the JC codon model according to the criterion of Kass and Raftery (1995) (Table 1.1). The topology of the species tree recovered from the HKY codon model in *BEAST was identical to that recovered from the GTR codon model in *BEAST even though most parameters did not converge in the GTR codon analysis. Results using the HKY codon model in *BEAST were also consistent with BEST analyses. Specifically, the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology was the most probable one and monophyly of *Xenopus* and *Silurana* was supported by 100% posterior probability in each of the 3 differently calibrated *BEAST analyses (Figure 1.6). When only one calibration point was used (*BEAST analysis 1 or 2), posterior probability on the clade containing *Pipa* and *Hymenochirus* was 71% or 73%. When both calibration points were used the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology was even more strongly supported with 86% posterior probability on the clade containing *Pipa* and *Hymenochirus* (Figure 1.6). In *BEAST analysis 1, divergence of *Xenopus* and *Silurana* was estimated to have occurred 50.4 Ma with a 95% credible interval of 38.6-63.2 myr. In *BEAST analyses 2 and 3, divergence estimates of *Xenopus* and *Silurana* were almost identical ($\sim$65 Ma with a 95% credible interval of $\sim$57-74 Ma). In *BEAST analysis 1, estimated divergence time of the clade containing *Xenopus* and *Silurana*, hereafter "Xenopodinae" (Cannatella and de Sá, 1993) from the clade containing *Pipa* and *Hymenochirus*, hereafter "Pipinae" (Cannatella and de Sá, 1993), is 109.4 Ma with a 95% credible interval of 89.2-131.1 Ma. In *BEAST analyses 2 and 3, estimated divergence time of Xenopodinae from Pipinae are similar (143.4 and 134.7 Ma with 95% credible intervals of 122.8-163.4 and 111.9-156.1 myr, respectively). In *BEAST analysis 1, estimated divergence time of Pipidae and Rhinophrynidae is 144.7 Ma with a 95% credible interval of 115.7-176.4 Ma.

The geological calibration points provide estimates of evolutionary rates that appear reasonable (Supplementary Figure 1.S1). Rates of evolution varies across partitions but are $\sim 4.7 \times 10^{-10}$ substitutions per site per year, assuming a generation time of 1 year. Interestingly, the rate of evolution of the stem region of mitochondrial DNA ribosomal genes appears to be similar to the rates of evolution of autosomal coding regions, even though the overall rate of mitochondrial DNA is faster than autosomal DNA in some groups, such as primates (Brown et al., 1982). Although we do not attempt to directly compare the models used by BEST and *BEAST, use of the codon model in the *BEAST analysis is probably preferable given the substantially different rates of evolution of each codon position (Supplementary Figure 1.S2). Taken together, analysis of the individual loci and the collective coalescent-based analyses provide strong support for the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) relationship. We relate divergence time estimates from the *BEAST analyses to geological and fossil information below.

1.6 Discussion

The complexity of the phylogenetic problem explored here is relatively simple in the sense that we seek to estimate relationships among only 4 lineages (4 genera of frogs in the family Pipidae) and there are therefore only 3 possible unrooted topologies (Figure 1.1) and only 15 possible rooted topologies. However, these relation- ships are actually quite difficult to resolve because they are old, because the internodes are probably short, and because ancestral populations were probably very large and/or structured. Our analyses (Figure 1.5 and Figure 1.6) support the contention of Frost et al. (2006) that the primary challenge to resolving relationships among pipids is locating the position of the root of only one of the 3 possible unrooted topologies (topology 1 in Figure 1.1). Using a combination of publicly available sequences, new data from 454 pyrosequencing, and multilocus coalescent phylogenetic analysis, the relationship ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) is most strongly supported. Thus, the pipid root is on the lineage connecting Xenopodinae and Pipinae (Figure 1.1). This result is consistent with the morphology-based results of Báez and Pugener (2003); Trueb et al. (2005); Trueb and Báez (2006), a mitochondrial DNA analysis by Evans et al. (2004), an analysis of one autosomal gene by Evans et al. (2005), and a maximum parsimony analysis of concatenated autosomal loci by Roelants and Bossuyt (2005). However, it is not consistent with the results of Frost et al. (2006) or various maximum likelihood analyses (Roelants and Bossuyt, 2005; Roelants et al., 2007; Irisarri et al., 2011; Pyron and Wiens, 2011). What could account for these discrepancies?

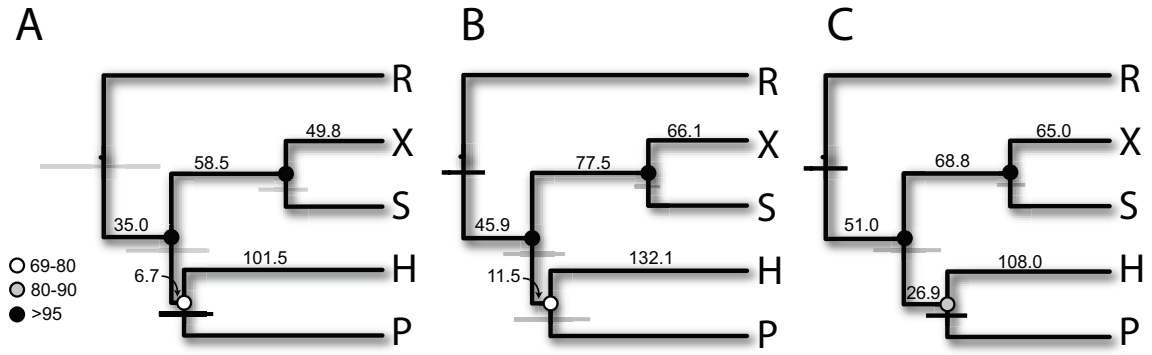

Figure 1.6: Consensus species trees from (A) *BEAST Analysis 1 which uses only the calibration for divergence of *Pipa* from *Hymenochirus*, (B) *BEAST Analysis 2 which uses only the calibration for divergence of Rhinophrynidae from Pipidae, and (C) *BEAST Analysis 3 which uses both of these calibrations. Numbers above branches represent time in millions of years and posterior probabilities (expressed as percentages) are indicated by shaded circles over each node. Grey bars represent the 95% highest posterior density intervals for time estimates in the posterior distribution and dark bars represent the nodes that were fixed in each analysis. Letters refer to taxa as defined in Figure 1 and R refers to Rhinophrynidae.

For a 3-taxon problem, even when internodes are short, each possible gene topology is equally likely so anomalous gene trees do not exist (Degnan and Rosenberg, 2006). If we accept the uncontroversial relationships depicted in Figure 1.1, the phylogenetic issue addressed in this study boils down to a 3-taxon problem with no anomalous trees. However, if we do not assume *Xenopus* + *Silurana* to be a clade for all portions of their genomes, this becomes a 4-taxon problem and an "anomaly zone" may exist depending on the length of the internodes (Degnan and Rosenberg, 2006). However, analysis of individual loci suggests that anomalous gene trees had very little impact on our phylogenetic inference. If anomalous gene trees were common, one would expect similar frequencies of all possible rootings of each of the 3 unrooted topologies depicted in Figure 1.1, but this is not the case (Figure 1.5). Instead, most (80.5%) of this distribution is comprised of only 3 alternative rootings of only one labeled unrooted topology (topology 1 in Figure 1.1), with the other 12 rootings being very rare in the posterior distribution.

Here, it was not possible to estimate effective population size for the extant species because we lack intraspecific data, but we are able to get rough estimates of the ancestral population size of the ancestral nodes. In all *BEAST analyses, the estimated ancestral population size of Xenopodinae is the lowest ($\sim$10 million individuals) and that of pipoids (the most recent common ancestor of *Rhinophrynus*, *Pipa*, *Hymenochirus*, *Pseudhymenochirus*, *Xenopus*, and *Silurana*) is second lowest ($\sim$26 million individuals). When only one node is fixed for calibration, the effective population size of Pipinae is smaller ($\sim$75 million individuals) than that of Pipidae (the most recent common ancestor of *Xenopus*, *Silurana*, *Pipa*, *Hymenochirus*) ($\sim$81 million individuals) but when both calibration points are enforced the opposite is true (163 and 107 million individuals for Pipinae and Pipidae, respectively). The relative magnitudes of these estimates in comparison to the ancestor of Xenopodinae are consistent with the observation that most genealogical discordance involves the location of the pipid root (i.e., whether or not "Pipinae" is a clade). These estimated ancestral population sizes are huge and this is unexpected in frogs and even high for some species of fruit flies. This may be a result of the sensitivity of *BEAST to the absence of intraspecific polymorphism data, a lack of phylogenetic signal in some partitions (which increases genealogical discordance), nonneutral evolution of portions of the sequences, or nonorthology of some sequences. Another important factor is that these ancestral populations were probably not panmictic and that ancestral population structure has the potential to inflate inferred effective population sizes (Wright, 1943; Nei and Takahata, 1993). This last possibility seems particularly plausible given the enormous distribution of ancestral pipoids which spanned much of North America, Europe, the Middle East, Africa, and South America (Trueb et al., 2005). Using different methods, a recent analysis of ancestral effective population size of 2 subdivided species of *Xenopus* from Ethiopia also recovered estimates that were in the millions for each species, probably also as a consequence of ancestral population structure (Evans et al., 2011a). Another possibility is that gene flow among lineages occurred after speciation via hybridization. The models assumed by BEST and *BEAST do not include gene flow after speciation and this biological phenomenon could lead to inflated estimates of ancestral effective population size by increasing genealogical discordance. In-

terspecies hybridization during pipoid diversification also seems plausible in light of the frequent allopolyploid evolution in *Xenopus* and *Silurana*.

There are analytical caveats to the multilocus coalescent methods we used. First, these methods generally assume neutral evolution of the data. Here, we included synonymous and nonsynonymous sites in order to maximize the phylogenetic signal of the data and because of the concern that synonymous substitutions would be saturated over the protracted divergence between the ingroup and outgroup (>300 Ma in total; Figure 1.6). However, rates of evolution of each codon position strongly suggest nonneutral evolution, minimally of the first and second positions which evolve slower than the third position (Supplementary Figure 1.S1). It is not yet clear to what degree the inclusion of nonsynonymous sites affected our inference of topology or parameter estimates (including divergence time and effective population size) using these methods and undoubtedly future research will explore this issue. A second important concern relates to the possibility that we included nonorthologous sequences in these alignments (Figure 1.2 and Figure 1.4B). This is difficult to quantify and could account for the inflated estimates of ancestral population size recovered from *BEAST and also for some of the genealogical discordance observed from the analysis of individual loci (Figure 1.5). We attempted to cope with this problem using the standardized outgroup distance ratio but it is not clear to what degree, if any, we succeeded. Another concern is that divergence times are partially confounded with ancestral population size due to ancestral polymorphism. Intraspecific sampling would permit the estimation of extant population sizes and potentially improve accuracy of the estimates of ancestral population sizes and divergence times. Simulation studies (J. Heled, unpublished results) suggest that the credible intervals for divergence times and effective population sizes of internal nodes tend to be larger with only one extant sequence per species as compared with analyses with more than one per extant species. Another caveat is that we were unable to achieve convergence with *BEAST analysis using the GTR codon model. This presumably is either because the data were insufficient to inform the posterior distributions of the many parameters in this model, or because we did not run the analysis for long enough or some combination of these possibilities. In a few of the GTR codon runs, we observed large improvements in likelihood that were not achieved in the other runs. We note that while convergence was not achieved for this model, the most strongly supported topology sampled in this high likelihood parameter space was the same as that recovered from the other *BEAST and BEST analyses – that is the ((*Xenopus*, *Silurana*)(*Hymenochirus*, *Pipa*)) topology.

1.6.1 Biogeographical implications

Dispersal of frog lineages can occur over marine barriers (e.g., Evans et al., 2003; Vences et al., 2003; Heinicke et al., 2007), although this is clearly much less frequent than dispersal over most terrestrial habitats. Here, we consider 2 marine barriers (the separation of South America and Africa ∼100 Ma and the separation of North America from West Africa ∼190 Ma) to calibrate our divergence estimates. In doing so, we generated hypothe-

ses and associated divergence time predictions that can be independently evaluated using other information.

At least 2 lines of evidence would reject the hypotheses associated with the calibration points in our *BEAST analyses. The first would be evidence of either *Hymenochirus* or *Pipa* on both sides of the Atlantic Ocean, evidence of Rhinophrynidae in Africa, or evidence of Pipidae in North America. To our knowledge, there is no such evidence. The second would be if date estimates recovered from these analyses were inconsistent with the fossil or geological record. To explore this further, we compared the fossil record and associated inferences about phylogenetic relationships to our geologically calibrated estimates of divergence times from *BEAST Analyses 1-3 (Table 1.2). For the most part, the timing and inferred phylogenetic affinities of fossil taxa agreed with the BEST and *BEAST analyses. However, there were some discrepancies between the age and inferred phylogenetic affinities from the fossil record and the 95% credible intervals for the divergence time between Rhinophrynidae and Pipidae that were recovered from *BEAST analysis 1 (Table 1.2). In particular, the extinct taxa *Thoraciliacus*, *Cordicephalis*, and *Shomronella* from the Early Cretaceous (145.5-99.6 Ma) of Israel, are postulated to be more closely related to Pipidae than to Rhinophrynidae (Nevo, 1969; Estes et al., 1978; Báez et al., 2000; Trueb and Báez, 2006). These fossils and their inferred relationships suggest that the most recent common ancestor of these fossils and Pipidae diverged from Rhinophrynidae prior to $\sim$145.5 Ma. This time overlaps with the 95% credible interval for the divergence of these clades from *BEAST analysis 1 (115.7-176.4 Ma). Similarly, a fossil of the extinct species *Rhadinosteus parvus* from the Kimmeridigian (155.7-150.8 Ma) is thought to be more closely related to extant Rhinophrynidae than to extant Pipidae, suggesting that these families diverged before this (Henrici, 1998). However, again, the 95% credible interval for the divergence of these clades from *BEAST analysis 1 overlaps with this time.

Table 1.2: Pipid fossils can be used to assess credibility of divergence times estimated using geological calibrations of *BEAST analysis 1 (*Pipa*), *BEAST analysis 2 (Root), and *BEAST analysis 3 (*Pipa* + Root); for each geological calibration, the credible intervals of estimated divergence times either agree (Y) or do not agree (N) with the estimated sister taxon (Sister) and age (Age) and not overlapping or by overlapping, respectively, with the age of the fossils.

Fossil	Location	Age	Sister	Citation	Pipa	Root	Pipa + Root
Avitabatrachus uliana	Argentina	Late Albian-Early Cenomanian (93.5-112.0 mya)	Pipidae, not Rhinophrynidae	Báez et al. (2000); Nevo (1969); Trueb and Báez (2006)	Y	Y	Y
Palaeobatrachus	Europe	Late Cretaceous (65.6-99.6 mya)	Pipidae, not Rhinophrynidae	Báez et al. (2000); Nevo (1969); Trueb and Báez (2006)	Y	Y	Y
Thoraciliacus	Israel	Early Cretaceous (145.5-99.6 mya)	Pipidae, not Rhinophrynidae	Báez et al. (2000); Nevo (1969); Trueb and Báez (2006)	N	Y	Y
Cordicephalis	Israel	Early Cretaceous (145.5-99.6 mya)	Pipidae, not Rhinophrynidae	Báez et al. (2000); Nevo (1969); Trueb and Báez (2006)	N	Y	Y
Shomronella jordanica	Israel	Early Cretaceous (145.5-99.6 mya)	Pipidae, not Rhinophrynidae	Estes et al. (1978)	N	Y	Y
Rhadinosteus parvus	USA (Utah)	Kimmeridigian (155.7-150.8 mya)	Rhinophrynidae, not Pipidae	Henrici (1998)	N	Y	Y
Saltenia ibanezi	Argentina	Santonian-Campanian (85.7-70.6 mya)	Xenopodiane, not Pipinae	Báez and Pugener (2003); Trueb and Báez (1997); Trueb and Báez (2006)	Y	Y	Y
Shelania	Chile + parts of South America	Palaeogene (65.5-23.0 mya)	Xenopodiane, not Pipinae	Báez and Pugener (2003); Trueb and Báez (1997); Trueb and Báez (2006)	Y	Y	Y
Llankibatrachus	Chile + parts of South America	Palaeogene (65.5-23.0 mya)	Xenopodiane, not Pipinae	Báez and Pugener (2003); Trueb and Báez (1997); Trueb and Báez (2006)	Y	Y	Y
"*Xenopus*" *romeri*	Brazil	Palaeocene (65.5-55.8 mya)	Xenopodiane, not Pipinae	Báez and Pugener (2003); Trueb and Báez (1997); Trueb and Báez (2006)	Y	Y	Y
Xenopus arabiensis	Yemen	Late Oligocene (28.4-23.0 mya)	*Xenopus*, not *Silurana*	Estes (1977); Henrici and Báez (2001)	Y	Y	Y
Eoxenopoides reuningi	South Africa	Maastrichtian/Selandian (70.6-58.7 mya)	Pipinae, not Xenopodinae	Báez and Harrison (2005); Báez and Pugener (2003); Trueb and Báez (2006); Trueb et al. (2005)	Y	Y	Y
Vulcanobatrachus mandelai	South Africa	Senonian (89.3-65.5 mya)	Pipinae, not Xenopodinae	Báez and Harrison (2005); Báez and Pugener (2003); Trueb and Báez (2006); Trueb et al. (2005)	Y	Y	Y
Singidella latecostata	Tanzania	∼45 mya	*Hymenochirus*, not *Pipa*	Báez and Harrison (2005); Báez and Rage (1998); Trueb and Báez (2006)	Y	Y	Y
Pachycentrata (Pachybatrachus) taqueti	Niger	Coniacian-Santonian (83.5-89.3 mya)	*Hymenochirus*, not *Pipa*	Báez and Harrison (2005); Báez and Rage (1998); Trueb and Báez (2006)	Y	Y	Y

If migration over marine barriers is difficult for these frogs, the estimated time of divergence of Pipidae from Rhinophrynidae from *BEAST Analysis 1 also is inconsistent with the geological record which indicates that by this time North America had already drifted apart from West Africa forming the North Atlantic Ocean, (Figure 1.3; McHone and Butler 1984), which eventually connected with the Tethys Sea to circumscribe the globe. This suggests that the calibrations used in *BEAST Analyses 2 and 3 are most appropriate if the North Atlantic Ocean was a formidable barrier to amphibian dispersal that triggered diversification of pipoids (Duellman and Trueb, 1994). While we do not know which of these 2 sets of divergence times are more accurate, the estimated divergence times of Xenopodinae are very similar in *BEAST analyses 2 and 3: $\sim$65 Ma with a 95% credible intervals of $\sim$57-76 Ma. These results underscore how ancient various pipid diversification events are, including the age of the subfamily Xenopodinae. Underestimates of divergence, for example by Bisbee et al. (1977), have contributed to nomenclatural confusion associated with the nonubiquitous recognition of "*Silurana*".

Although it does not speak to the validity of the geological calibration points, it is interesting that some fossils from South America are purportedly more closely related to Xenopodinae than to Pipinae, suggesting that both of these lineages were widely dispersed over West Gondwana prior to its breakup (Báez and Pugener, 2003; Trueb et al., 2005; Trueb and Báez, 2006), even though today Xenopodinae is found only in Africa. An interesting direction for future research would be to add data from divergent taxa within these clades, such as *Pseudhymenochirus merlini* (Cannatella and Trueb, 1988b), *Pipa parva* (Cannatella and Trueb, 1988a), or *X. borealis* (Evans, 2008) in order to provide more detailed information about the timing of diversification within pipids.

1.7 Conclusions

We generated a phylogenomic data set with an aim of providing further resolution to ancient relationships among a group of frogs (family Pipidae). In compiling these data, we encountered challenges with data quality in that next generation sequence reads from an old sample of *R. dorsalis* were generally shorter and sparser than those from fresh samples. We encountered challenges with alignment, which drove us to conservatively discard noncoding sequences and randomly discard one *X. laevis* paralog when 2 were identified. In the interest of being confident in our alignments and to avoid violation of model assumptions, we further discarded data from loop regions and one side of stem regions of ribosomal genes from mitochondrial DNA. Despite these conservative measures, the final data set was sufficiently large that it proved difficult to analyze using a relatively complex model, forcing us to resort to a model with less parameters. The ((*Xenopus*, *Silurana)*(*Pipa*, *Hymenochirus*)) topology was recovered using 2 multilocus coalescent methods and multiple models, and this topology was also suggested by analysis of individual loci and a concatenated analysis. Calibration regimes that included a role for the opening of the North Atlantic Ocean $\sim$190

Ma produced divergent time estimates that were consistent with the fossil record. Perhaps most striking was the high degree of genealogical discordance, which probably stems from a combination of phylogenetic error, inadvertent analysis of nonorthologous (paralogous) data, and large or structured ancestral population sizes. Together, these findings highlight the exciting prospects for the field of phylogenomics and also the daunting analytical challenges that lie ahead.

1.8 Acknowledgements

We thank Jim McGuire for suggesting that we try to extract RNA from a *R. dorsalis* sample at the Museum of Vertebrate Zoology and for providing this tissue loan, Brian Golding for use of computational resources, and Ron DeBry, David Blackburn, Rich Glor, Laura Kubatko, and 3 anonymous reviewers for helpful discussion and comments on an earlier version of this manuscript.

1.9 Supplementary information

Table 1.S1: Information on sequences in this study including gene name, gene acronym, total number of base pairs (bp), the number of parsimony uninformative positions (PUI), the number of parsimony informative positions (PI), model used in analysis of individual loci (Model), accession numbers of *S. tropicalis*, *X. laevis*, *H. curtipes*, *P. carvalhoi*, and *R. dorsalis* (ST accession, XL accession, H accession, P accession, and R accession, respectively), and number of reads per partition for *H. curtipes*, *P. carvalhoi*, and *R. dorsalis* (H, P, and R, respectively). Sequences with portions that have less than 200 bp of continuous sequence are no archived by GenBank and the accession number is therefore indicated with "XXX"; these sequences are available in the files in the Dryad database with three "NNN"s reflecting gaps.

#	Gene name	Acronym	bp	PUI	PI	Model	ST accession	XL accession	H accession	P accession	R accession	H	P	R
1	ATP Synthase, H+ Transporting, Mitochondrial FO Complex, Subunit B1	ATP5F1	318	60	26	HKY+I+d	NM_203787.1	NM_001086241.1	JU062532	JU062530	JU062531	26	9	3
2	Desmin, Gene 1	DES1	189	25	22	GTR+G+d	NM_001093564.1	NM_001093564.1	XXX	XXX	XXX	17	20	16
3	Scaffold Attachment Factor A	SAF-A	120	5	4	GTR+I+d	BC167650.1	BC084742.1	XXX	XXX	XXX	11	3	9
4	Ribosomal Protein L4	RPL4	846	140	67	GTR+G+d	BC043895.1	NM_001171532.1	JU062613	JU062611	JU062612	143	47	10
5	Elongation Factor-1 Delta	EEF1D	219	42	22	HKY+I+d	NM_001011450.1	BC072139.1	JU062661	JU062659	JU062660	47	53	1
6	Fragile X Mental Retardation, Autosomal Homolog 1	FXR1	111	17	11	GTR+G+d	XM_002936875.1	BC131884.1	XXX	XXX	XXX	6	1	3
7	Glyceraldehyde 3-Phosphate Dehydrogenase	GAPDH	546	108	52	SYM+G+e	NM_001004949.1	BC043972.1	JU062610	JU062608	JU062609	22	45	8
8	Hematopoietic Prostaglandin D Synthase	HPGDS	474	136	67	GTR+G+d	NM_001005090.1	NM_001090049.1	JU062496	JU062494	JU062495	39	68	7
9	Heat Shock 70kDa Protein 9	HSPA9	267	41	20	GTR+I+d	NM_001001229.2	NM_001086697.1	JU062550	JU062548	JU062549	14	3	9
10	Poly(A) Binding Protein,	PABPC1	627	110	48	GTR+G+d	NM_001005051.1	NM_001091600.1	JU062643	JU062641	JU062642	16	73	6
11	Ferritin Light Chain	FERRITIN	396	78	41	K80+I+e	NM_203881.1	NM_001086183.1	JU062595	JU062593	JU062594	11	1	9
12	Proteasome (Prosome, Macropain) 26S Subunit, Non-ATPase, 1	PSMD1	138	24	7	HKY+G+d	CR762098.2	BC084152.1	XXX	XXX	XXX	5	5	8
13	Ribosomal Protein L5	RPL5	258	54	20	K80+I+e	NM_203550.1	NM_001085908.1	JU062655	JU062653	JU062654	114	52	4
14	Ribosomal Protein S3	RPS3	486	20	23	SYM+G+e	NM_203788.1	BC041299.1	JU062622	JU062620	JU062621	26	21	8
15	Uncoupling Protein 2 (Mitochondrial, Proton Carrier)	UCP2	408	69	47	GTR+G+d	NM_203848.1	NM_001091378.1	JU062544	JU062542	JU062543	116	8	1
16	Adiponectin Receptor 2	ADIPOR2	108	15	7	HKY+I+d	XM_002938095.1	NM_001093867.1	XXX	XXX	XXX	6	1	2
17	Eukaryotic Translation Elongation Factor 1 Gamma	EEF1G	441	84	37	GTR+I+d	NM_203931.1	NM_001086297.1	JU062628	JU062626	JU062627	84	83	3
18	Heterogeneous Nuclear Ribonucleoprotein D-like	HNRPDL	105	16	12	HKY+I+d	NM_001011015.1	NM_001091139.1	XXX	XXX	XXX	7	9	1
19	Heat Shock 70kDa Protein 5a	N/A	123	24	15	SYM+G+e	XM_002941644.1	NM_001087993.1	XXX	XXX	XXX	78	22	6
20	Heat Shock 70kDa Protein 1-like	HSPA1L	552	82	46	GTR+G+d	XM_002937528.1	NM_001086599.1	JU062508	JU062506	JU062507	93	12	12
21	Ribosomal Protein S6	RPS6	705	101	47	GTR+G+d	NM_203821.1	NM_001086702.1	JU062568	JU062566	JU062567	56	42	4
22	ATP Synthase, H+ Transporting, Mitochondrial F1 Complex, Gamma Polypeptide 1	ATP5C1	282	44	27	GTR+G+d	NM_001016407.2	NM_001087012.1	JU062556	JU062554	JU062555	34	22	2
23	Acyl-CoA Dehydrogenase, C-4 to C-12 Straight Chain	ACADM	246	40	25	GTR+I+d	XM_002936083.1	NM_001094193.1	JU062541	JU062539	JU062540	1	5	2
24	Prolyl 4-Hydroxylase, Beta Polypeptide	P4HB	135	31	9	GTR+G+d	BC161026.1	NM_001086346.1	XXX	XXX	XXX	38	3	11
25	Phosphatidylethanolamine Binding Protein 1	PEB1	105	17	10	K80+I+e	NM_001016825.2	NM_001092157.1	XXX	XXX	XXX	5	20	4
26	Peptidylprolyl Isomerase B (Cyclophilin B)	PPIB	105	21	9	K80+I+e	NM_001017063.2	NM_001087036.1	XXX	XXX	XXX	37	3	3
27	Staphylococcal Nuclease and Tudor Domain Containing 1	SND1	294	53	17	GTR+G+d	NM_203852.1	NM_001087031.1	JU062538	JU062536	JU062537	18	4	4
28	Small Nuclear Ribonucleoprotein D3 Polypeptide 18kDa	SNRPD3	132	18	13	K80+I+e	NM_001017093.2	NM_001091467.1	XXX	XXX	XXX	1	4	1
29	Keratin 19	KRT19	135	26	20	HKY+I+d	NM_203587.1	NM_001096855.1	XXX	XXX	XXX	66	17	1
30	ATP Synthase, H+ Transporting, Mitochondrial F1 Complex, O Subunit	ATP5O	348	72	35	K80+G+e	NM_001044412.1	BC078592.1	JU062625	JU062623	JU062624	27	13	1
31	Basic Transcription Factor 3	BTF3	186	28	13	HKY+I+d	NM_001016697.2	NM_001094887.1	XXX	XXX	XXX	23	14	1
32	Tubulin, Alpha 3C	TUBA3C	459	78	43	SYM+I+e	BC135250.1	NM_001086054.1	JU062580	JU062578	JU062579	26	5	5
33	Pinin, Desmoaome Associate protein	PNN	294	44	27	HKY+G+d	NM_001005707.1	NM_001087556.1	JU062499	JU062497	JU062498	10	1	3
34	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 21	DDX21	312	71	27	HKY+G+d	NM_001045759.1	NM_001088564.1	JU062517	JU062515	JU062516	45	17	3
35	Ribosomal Protein S27a	RSPS27A	159	20	22	K80+G+e	NM_001016172.2	NM_001092596.1	XXX	XXX	XXX	16	22	2
36	LUC7-like	LUC7L	111	23	12	K80+I+e	NM_001005458.1	NM_001092889.1	XXX	XXX	XXX	5	1	2
37	Glutaminyl-tRNA Synthetase	QARS	129	25	15	K80+G+e	XR_097203.1	NM_001086510.1	XXX	XXX	XXX	14	1	3
38	Amyloid Beta (A4) Precursor-like Protein 2	APLP2	132	20	12	SYM+G+e	XM_002940006.1	AJ608933.1	XXX	XXX	XXX	14	5	1
39	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 3, X-linked	DDX3X3	141	23	10	GTR+G+d	NM_203865.1	NM_001086814.1	XXX	XXX	XXX	6	1	10
40	Glycoprotein, Synaptic 2	GPSN2	195	35	12	GTR+I+d	NM_001008198.1	NM_001086992.1	XXX	XXX	XXX	1	2	8

Table 1.S1: *Continued from previous page*

#	Gene name	Acronym	bp	PUI	PI	Model	ST accession	XL accession	H accession	P accession	R accession	H	P	R
41	NDP Dipeptidase 2 (Metallopeptidase M20 Family)	CNDP2	141	23	15	SYM+G+e	NM_203958.1	NM_001086840.1	XXX	XXX	XXX	8	3	1
42	Ribosomal Protein L18	RPL18	315	48	24	K80+G+e	NM_001011030.1	X06222.1	JU062571	JU062569	JU062570	38	35	6
43	Nonmuscle Myosin II Heavy Chain A	MYH9	120	19	10	SYM+I+e	XM_002934718.1	BC170424.1	XXX	XXX	XXX	19	11	5
44	Stress-Induced-Phosphoprotein 1	STIP1	240	41	26	GTR+G+d	NM_204029.1	NM_001086794.1	XXX	XXX	XXX	29	6	3
45	Nucleosome Assembly Protein 1-Like 1, Transcript Variant 2	NAP1L1	531	84	34	GTR+G+d	NM_001015984.3	NM_001088541.1	JU062502	JU062500	JU062501	44	49	6
46	Non-Metastatic Cells 2, Protein (NM23B) Expressed In (NME2)	NME2	111	14	10	SYM+G+e	NM_001005140.1	BC079795.1	XXX	XXX	XXX	26	35	4
47	RNA Binding Motif Protein 39	RBM39	183	32	16	GTR+G+d	NM_001016788.2	NM_001093481.1	XXX	XXX	XXX	1	1	3
48	Ribosomal Protein L13	RPL13	507	88	29	GTR+I+d	NM_203780.1	NM_001093003.1	JU062505	JU062503	JU062504	74	46	4
49	Ribosomal Protein S20	RPS20	330	52	22	SYM+I+e	NM_001005106.1	BC041524.1	JU062598	JU062596	JU062597	40	47	2
50	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein, Beta Polypeptide	YWHAB	198	39	21	K80+G+e	NM_001011116.1	NM_001172053.1	XXX	XXX	XXX	1	28	2
51	Chromosome 14 Open Reading Frame 166	C14ORF 166	153	31	11	GTR+G+d	XM_002936905.1	NM_001095368.1	XXX	XXX	XXX	3	8	2
52	Ribosomal Protein S7	RPS7	390	46	26	GTR+I+d	NM_001113856.1	NM_001091033.1	JU062589	JU062587	JU062588	58	49	2
53	Heat Shock Protein 90kDa Beta (Grp94), Member 1	HSP90B1	342	58	33	GTR+G+d	BC121250.1	NM_001090811.1	XXX	XXX	XXX	149	93	3
54	Ribosomal Protein SA	RPSA	738	127	62	SYM+I+e	NM_203737.1	NM_001095637.1	JU062649	JU062647	JU062648	39	66	3
55	Tubulin, Beta 3	TUBB3	201	44	20	K80+G+e	NM_204077.1	BC044125.1	JU062658	JU062656	JU062657	2	3	9
56	Ribosomal Protein S5	RPS5	105	13	9	K80+I+e	BC161332.1	NM_001086331.1	XXX	XXX	XXX	62	57	6
57	Ribosomal Protein L27	RPL27	204	22	23	K80+I+e	NM_001016180.2	NM_001086954.1	XXX	XXX	XXX	30	34	1
58	Ribosomal Protein L27a	RPL27A	372	57	25	GTR+G+d	NM_001130329.1	NM_001086720.1	JU062607	JU062605	JU062606	29	25	2
59	40S Ribosomal Small Subunit Protein S10	RPS10	141	18	13	SYM+G+e	NM_001016718.2	BC169445.1	XXX	XXX	XXX	30	16	1
60	NSA2 Ribosome Biogenesis Homolog	NSA2	183	30	15	HKY+G+d	NM_001017326.3	NM_001174021.1	XXX	XXX	XXX	27	24	1
61	Cytochrome C Oxidase Subunit Va	COX5A	258	43	14	GTR+I+d	NM_001017049.3	NM_001091923.1	XXX	XXX	XXX	7	10	2
62	Ribosomal Protein L26	RPL26	432	63	28	GTR+G+d	NM_001005104.1	BC075124.1	JU062514	JU062512	JU062513	53	37	5
63	Eukaryotic Translation Initiation Factor 3, Subunit E	EIF3E	342	47	24	HKY+G+d	BC080327.1	BC128927.1	JU062577	JU062575	JU062576	19	6	5
64	Ribosomal Protein L17	RPL17	351	51	20	GTR+G+d	NM_001005078.1	NM_001093152.1	JU062553	JU062551	JU062552	46	37	5
65	Pancreatic Progenitor Cell Differentiation and Proliferation Factor Homolog	PPDPF	180	26	13	K80+I+e	NM_001097343.1	NM_001096856.1	XXX	XXX	XXX	1	4	2
66	Aminoacyl tRNA Synthetase Complex-Interacting Multifunctional Protein 1	AIMP1	162	30	11	SYM+I+e	NM_001017050.2	NM_001086641.1	XXX	XXX	XXX	1	9	1
67	Ribosomal Protein, Large, P0	RPLP0	393	66	29	SYM+G+e	NM_203736.1	NM_001086665.1	JU062547	JU062545	JU062546	40	40	6
68	Eukaryotic Translation Elongation Factor 1 Alpha 1	EEF1A1	813	135	82	SYM+G+e	BC157768.1	BC130144.1	JU062574	JU062572	JU062573	82	107	25
69	Transketolase	TKT	417	89	44	K80+G+e	NM_001015866.1	NM_001089843.1	JU062586	JU062584	JU062585	7	55	3
70	H3 Histone, Family 3A	H3F3A	192	36	12	SYM+I+e	NM_001005101.1	NM_001098432.1	XXX	XXX	XXX	6	5	2
71	Tripartite Motif-Containing 7	TRIM 7	174	49	18	K80+I+e	XM_002943043.1	NM_001087987.1	XXX	XXX	XXX	1	1	2
72	Methionine Aminopeptidase 2	METAP2	132	22	6	GTR+G+d	XM_002944458.1	NM_001087003.1	XXX	XXX	XXX	16	1	3
73	Ribosomal Protein S4, X-Linked	RPS4X	387	51	22	SYM+I+e	NM_203581.2	NM_001097003.1	JU062640	JU062638	JU062639	29	38	6
74	Chaperonin Containing TCP1, Subunit 5	CCT5	201	28	15	GTR+G+d	NM_204009.1	NM_001086984.1	JU062526	JU062524	JU062525	11	1	1
75	Alcohol Dehydrogenase 1B (Class I), Beta Polypeptide	ADH1B	135	28	14	GTR+G+d	NM_001011431.1	BC083003.1	XXX	XXX	XXX	9	28	1
76	Ribosomal Protein S3A	RPS3A	546	82	45	GTR+G+d	NM_001008074.1	NM_001110740.1	JU062529	JU062527	JU062528	38	21	6
77	Actin, Alpha 2, Smooth Muscle, Aorta	ACTA2	162	16	10	SYM+G+e	NM_001011250.1	NM_001091337.1	XXX	XXX	XXX	1	1	23
78	Actin, Gamma 1	ACTG1	570	121	64	GTR+I+d	NM_204001.1	NM_001172163.1	JU062631	JU062629	JU062630	56	10	23
79	Glutamyl-Prolyl-tRNA Synthetase	EPRS	228	48	24	GTR+G+d	XM_002936018.1	NM_001127874.1	XXX	XXX	XXX	12	12	4
80	ATP Synthase, H+ Transporting, Mitochondrial F1 Complex, Beta Polypeptide	ATP5B	384	51	38	GTR+I+d	NM_001001256.1	NM_001086657.1	JU062637	JU062635	JU062636	33	26	1
81	ATP Synthase, H+ Transporting, Mitochondrial Fo Complex, Subunit C3 (Subunit 9)	ATP5G3	306	35	19	SYM+I+e	BC159382.1	NM_001086614.1	JU062559	JU062557	JU062558	32	32	3
82	Ribosomal Protein S25	RPS25	159	26	15	GTR+G+d	NM_001005084.1	NM_001093006.1	XXX	XXX	XXX	43	50	1
83	Solute Carrier family 25 (Mitochondrial Carrier; Phosphate Carrier), Member 3	SLC25A3	471	60	33	HKY+G+d	NM_203597.1	NM_001086726.1	JU062583	JU062581	JU062582	63	30	7
84	Ribosomal Protein L37a	RPL37A	114	7	6	K80+I+e	NM_001005137.1	NM_001086257.1	XXX	XXX	XXX	28	20	1
85	Heat Shock 70kDa Protein 1	HSPA1	378	74	33	K80+G+e	XM_002937895.1	NM_001174009.1	JU062634	JU062632	JU062633	1	8	17
86	GTP Binding Protein 4	GTPBP4	522	119	69	GTR+I+d	XM_002935177.1	NM_001110745.1	JU062664	JU062662	JU062663	13	37	10
87	DEAD (Asp-Glu-Ala-Asp) Box Polypeptide 46, Transcript Variant 2	DDX46	189	40	22	GTR+G+d	XM_002935899.1	BC123328.1	XXX	XXX	XXX	11	12	2
88	Radixin	RDX	111	20	11	HKY+G+d	XM_002937962.1	BC089125.1	XXX	XXX	XXX	1	1	2
89	Alcohol Dehydrogenase 5 (Class III), Chi Polypeptide	ADH5	204	35	20	GTR+I+d	NM_001011502.1	AY393843.1	JU062619	JU062617	JU062618	24	2	2
90	Tubulin, Beta 2B	TUBB2B	243	42	30	SYM+I+e	NM_203944.1	NM_00108606	JU062616	JU062614	JU062615	1	1	9

Table 1.S1: *Continued from previous page*

#	Gene name	Acronym	bp	PUI	PI	Model	ST accession	XL accession	H accession	P accession	R accession	H	P	R
91	Epiplakin 1	EPPK1	156	27	13	K80+G+e	BC125783.1	BC128679.1	XXX	XXX	XXX	3	6	2
92	Moesin	MSN	123	22	12	K80+I+e	XM_0029367	NM_00109656	XXX	XXX	XXX	18	4	2
93	Arginyl-tRNA Synthetase	RARS	105	16	9	K80+G+e	NM_28034.1072.1	BC0957.1633.1	XXX	XXX	XXX	10	1	1
94	Tubulin, Alpha 4b	TUBA4B	165	23	10	GTR+G+d	BC171015.1	NM_00109446	XXX	XXX	XXX	4	3	7
95	Nuclear Receptor Coactivator 2	NCOA2	132	21	9	HKY+G+d	NM_0011421	NM_0081.108767	XXX	XXX	XXX	3	3	2
96	Transaldolase 1	TALDO1	240	54	22	HKY+G+d	NM_5090.110171	NM_001.109473	JU062523	JU062521	JU062522	18	4	1
97	Coiled-Coil Domain Containing 109B	CCDC109B	102	9	13	K80+I+e	NM_001126576.1	NM_001094368.1	XXX	XXX	XXX	1	1	1
98	Enhancer of Polycomb Homolog 2	EPC2	102	18	9	HKY+I+d	NM_001005683.1	NM_001094140.1	XXX	XXX	XXX	1	4	2
99	Ribosomal Protein L23	RPL23	288	34	17	HKY+I+d	NM_0010112	NM_00109245	JU062562	JU062560	JU062561	16	25	1
100	Heat Shock Protein 90kDa Alpha (Cytosolic), Class A Member 1, Gene 1	HSP90AA1.1	339	50	40	SYM+G+e	NM_3010.11079297.1	NM_0021.1092129.1	JU062565	JU062563	JU062564	108	52	10
101	Glucose-6-Phosphate Isomerase	GPL	114	19	14	SYM+I+e	NM_0010110	NM_00109229	XXX	XXX	XXX	26	1	1
102	RAB10, Member RAS Oncogene Family	RAB10	141	22	19	HKY+G+d	NM_001126746.1	NM_001089799.1	XXX	XXX	XXX	1	1	1
103	Ribosomal Protein L3	RPL3	768	100	62	GTR+G+d	NM_0011266	NM_00108687	JU062601	JU062599	JU062600	53	25	9
104	Eukaryotic Translation Initiation Factor 3, Subunit M	EIF3M	144	22	19	GTR+G+d	NM_001004794.1	NM_001086592.1	XXX	XXX	XXX	17	5	7
105	Ribosomal Protein L10a	RPL10A	534	80	42	HKY+G+d	NM_0011265	NM_00108673	JU062520	JU062518	JU062519	48	50	8
106	Seryl-tRNA Synthetase	SARS	117	18	15	K80+G+e	BC067920.1	NM_00108701	XXX	XXX	XXX	13	1	1
107	Vesicle-Associated Membrane Protein-Associated Protein B/C	N/A	357	70	39	GTR+G+d	NM_203574.1	NM_0081.1095059.1	JU062667	JU062665	JU062666	45	28	1
108	Recombination Activating Genes 1 and 2	RAG1, RAG2	1695	336	207	GTR+I+d	XP_002937338.1, XP_002937337.1	NM_001172083.1, NM_001097900.1	AY874305, EF535956	AY874303, JU062590	EF535901.1, EF535955	–	–	–
109	Solute Carrier Family 8 (Sodium/Calcium Exchanger), Member 3	SLC8A3	1062	170	90	GTR+G+d	XM_002936892.1	NM_001093062.1	EF107344.1	EF107351.1	AY948894.1	–	–	–
110	Chemokine (C-X-C Motif) Receptor 4	CXCR4	651	121	78	GTR+G+d	NM_001097362.1	NM_001137581.1	AY523685.1	AY364174.1	AY523699.1	–	–	–
111	Histone H3.2-Like	H2A	279	35	44	GTR+G+d	XM_0029452	BC084311.1	JU062646	JU062644	DQ284156	–	–	–
112	Sodium/Calcium Exchanger 1 Gene, Exon 2	NCX-1	1317	199	96	GTR+I+d	AY5283.7121.1	X90839.1	AY523702.1	AY523711.1	AY5213722.1	–	–	–
113	Rhodopsin	RHO	312	46	22	HKY+G+d	NM_0010973	NM_00108704	DQ347348	DQ283922	DQ347405	–	–	–
114	Mitochondrial DNA	12S, 16S	432	62	44	GTR+G+d	AY581668	AY581639.1	AY581623	AY581621	AY581620	–	–	–

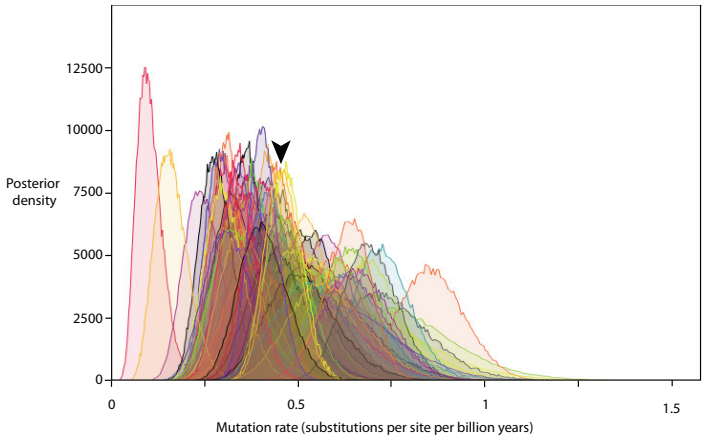

Figure 1.S1: Posterior densities of mutation rates from *BEAST analysis 3 with the HKY codon model. An arrowhead indicates the posterior density for the loop region of the mitochondrial DNA and rDNA.

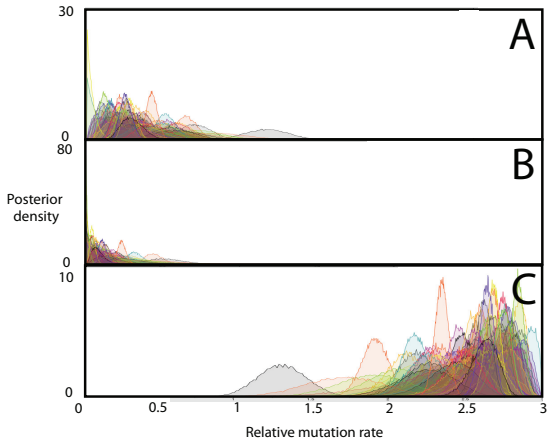

Figure 1.S2: Posterior densities of relative mutation rate from *BEAST analysis 3 with the HKY codon model for (A) the first codon position, (B) the second codon position, and (C) the third codon position of the 113 autosomal partitions.

Correction terms for *BEAST analyses provided by J. Heled

**BEAST Analysis 1 correction term*
The analysis has five taxa with an outgroup, or four taxa monophyly, and the calibration is on the parent of *Pipa*. There are 6×3 ranked trees: 9 of those have 3 internal nodes above the calibrated parent, 6 have 2 above and 1 below, and the remaining 3 have 2 below and one (the root) above.

As described in Appendix 3 of Heled and Drummond (2011). the total density for a above and b below is:

$$f_{a,b}(h) = \lambda e^{-\lambda h} \frac{e^{-\lambda(a+1)h}}{(a+1)!} \frac{(1-e^{-\lambda h})^b}{b!}$$

So the total correction is:

$$\begin{aligned} f(h) &= \frac{5!}{18}(9f_{9,3}(h) + 6f_{2,1}(h) + 3f_{1,2}(h)) \\ &= \frac{5\lambda e^{-3\lambda h}}{6}(e^{-2\lambda h} - 4e^{-\lambda h} + 6) \end{aligned}$$

**BEAST Analysis 2 correction term*
The correction term for this analysis is given by equation 11 in Heled and Drummond (2011).

**Beast Analysis 3 correction term*

$$f_{a,b}(h_0, h_1) = \lambda e^{-2\lambda h_0} \lambda e^{-2\lambda h_1} \frac{(e^{-\lambda h_1} - e^{-\lambda h_0})^{a-1}}{(a-1)!} \frac{(1-e^{-\lambda h_0})^b}{b!}$$

and the total correction is:

$$\begin{aligned} f(h_0, h_1) &= \frac{5!}{18}(9f_{3,0}(h_0, h_1) + 6f_{2,1}(h_0, h_1) + 3f_{1,2}(h_0, h_1)) \\ &= 10e^{-\lambda(h_1+2h_0)}(1 - e^{-\lambda h_0})(1 + 2e^{-\lambda h_1} - 3e^{-\lambda h_0}) \end{aligned}$$

Chapter 2

Evolution of the closely related, sex-related genes *DM-W* and *DMRT1* in African clawed frogs (*Xenopus*)

Bewick AJ, Anderson DW, and Evans BJ (2011) *Evolution*, 65(3):698-712

2.1 Preface

Gene duplication is suspected of being a fundamental source of genetic novelty. Here we investigated many aspects of the evolution of a newly identified, partially paralogous female sex-determining gene, *DM-W*, across the breadth of the African clawed frog (*Xenopus*) phylogeny, and at a finer scale between closely related species.

2.2 Abstract

DM-W is a dominant, female-specific, regulator of sex determination in the African clawed frog *Xenopus laevis*. This gene is derived from partial duplication of *DMRT1*, a male-related autosomal gene. We set out to better understand sex determination in *Xenopus* by studying this pair of genes. We found that *DM-W* evolved in *Xenopus* after divergence from the sister genus *Silurana* but before divergence of *X. laevis* and *X. clivii*, and that *DM-W* arose from partial duplication of *DMRT1*β, which is one of the two *DMRT1* paralogs in the tetraploid ancestor of *Xenopus*. Using the rate ratio of nonsynonymous to synonymous substitutions per site and multilocus polymorphism data, we show that *DM-W* evolved non-neutrally. By cloning paralogs and using a pyrosequencing assay, we also demonstrate that

DMRT1 underwent phylogenetically biased pseudogenization after polyploidization, and that expression of this gene is regulated by mechanisms that vary through development. One explanation for these observations is that the expression domain of *DMRT1*β was marginalized, which would explain why this paralog is dispensable in *Xenopus* polyploids and why *DM-W* has a narrow expression domain. These findings illustrate how evolution of the genetic control of stable phenotypes is facilitated by redundancy, degeneration, and compartmentalized regulation.

2.3 Introduction

Sex determination in frogs is genetically triggered (Hayes, 1998) but is achieved via variable mechanisms, making this group a compelling but understudied model for studying evolution of this crucial phenotype. Recently, a female-specific, W-linked gene called *DM-W* was discovered in the African clawed frog *Xenopus laevis* (Yoshimoto et al., 2008), supporting earlier conclusions that *X. laevis* females are heterogametic (Mikamo and Witschi, 1963). *DM-W* is the first master regulator of primary sex determination to be identified in amphibians. The 5′-coding region of *DM-W* is very similar to the 5′-coding region of the *doublesex* and *mab-3* related transcription factor (*DMRT1*) gene, suggesting an origin of *DM-W* by partial gene duplication of *DMRT1* (Yoshimoto et al., 2008). This is also suggested by the exon structure of these genes, which are similar in the 5′ portion but not the 3′ portion (Figure 2.1). The homologous region of *DMRT1* and *DM-W* contains a DNA-binding motif called a "DM domain" (Yoshimoto et al., 2008) and both genes bind to shared DNA sequences (Yoshimoto et al., 2010), suggesting that they are transcription factors that interact with the same regulatory elements. Analysis using full-length sequences suggested *DM-W* arose after the diversification of ancestors of frogs and fish but before divergence of the ancestors of frogs, birds, and mammals (Supplementary information of Yoshimoto et al. (2008)) but the genome sequence of the Nigerian clawed frog (*Silurana tropicalis*) does not contain *DM-W* (Yoshimoto et al., 2008), even though the individual sequenced was female (Hellsten et al., 2010).

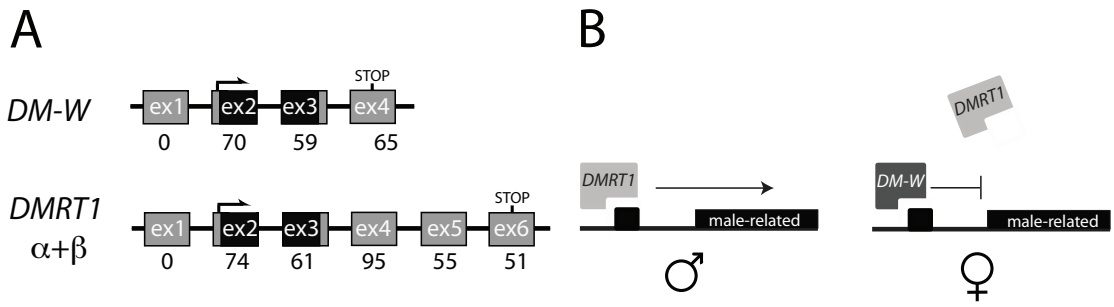

Figure 2.1: Regions of homology between *DM-W* and *DMRT1* and proposed competitive binding during sexual differentiation of females. (A) Exons (boxes) and introns (lines) of *X. laevis DM-W*and *DMRT1* with homologous region used for analysis of ω in black. *DM-W* exon structure follows Yoshimoto et al. (2008) and *DMRT1* exon structure follows Osawa et al. (2005) and is based on *S. tropicalis*. Numbers below each exon reflect how many amino acids are encoded; the locations of start (arrow) and stop codons are indicated. (B) *DM-W* (black) and *DMRT1* (gray) bind to the same *cis*- regulatory factors indicated by black squares (Yoshimoto et al., 2010). In males, *DMRT1* drives expression of male-related genes. In females, *DM-W* is expressed slightly earlier during primary gonadal differentiation and is thought to repress transcription of male-related genes (Yoshimoto et al., 2008).

DM domain-containing genes play a remarkably conserved role as an activator of male differentiation in metazoans including worms, flies, coral, birds, and humans (Burtis and Baker, 1989; Raymond et al., 1999; Yi and Zarkower, 1999; Raymond et al., 2000; Miller et al., 2003; Haag and Doty, 2005). *DMRT1* is broadly expressed during development of *X. laevis* and has been detected in unfertilized eggs, early tadpole development of both sexes, in the gonads of both sexes during primary gonadal differentiation, and in postmetamorphic testis and ovary, with expression becoming increasingly male-biased in testis compared to ovary by 1-5 months after metamorphosis (Osawa et al., 2005; Yoshimoto et al., 2006, 2008). No sex difference in *DMRT1* expression was detectable during gonadal differentiation (Yoshimoto et al., 2008). Although, paralogous sequences were not considered. In contrast, *DM-W* appears to be expressed only in female gonads during primary gonadal differentiation, with its peak expression level just prior to a surge of *DMRT1* expression in the gonads of both sexes (Yoshimoto et al., 2008). In females, *DM-W* expression may antagonize *DMRT1*-activation of male-specific genes by binding to and inhibiting regulatory regions recognized by both proteins (Figure 2.1B; Yoshimoto et al., 2008, 2010).

African clawed frogs include the genera *Silurana* and *Xenopus* and a diversity of species that evolved through allopolyploidization (Evans, 2008). Species of *Xenopus* have 36, 72, or 108 chromosomes and are tetraploid, octoploid, or dodecaploid, respectively. *Xenopus laevis* is tetraploid but considered "pseudotetraploid" because of disomic inheritance (Tymowska, 1991). Species of *Silurana* have 20 or 40 chromosomes and are diploid (such as *S. tropicalis*), or tetraploid. Because of the unique way that allopolyloidization occurs in *Xenopus* (Kobel, 1985, 1996), *DM-W* and linked female-specific genes on the nonre-

combining portion of the W chromosome (hereafter NRW) are not duplicated by allopolyploidization (Figure 2.2). In contrast, autosomal genes such as *DMRT1* (Uno et al., 2008; Yoshimoto et al., 2008) are duplicated by allopolyploidization. Thus females of tetraploid, octoploid, and dodecaploid species are expected to have only one allele of *DM-W* but up to four, eight, or 12 *DMRT1* alleles organized in two, four, or six paralogous loci, respectively.

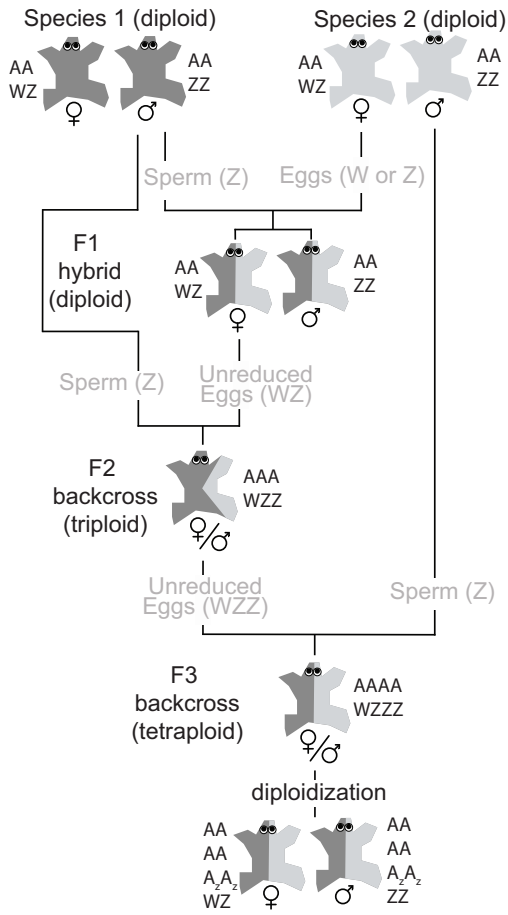

Figure 2.2: Autosomal genes (A) such as *DMRT1* are duplicated by allopolyploidization in clawed frogs but *DM-W* (W) is not. Whole-genome duplication (WGD) proceeds via F1 hybrid and backcrossed females that produce unreduced eggs. The mechanisms of sex determination in the triploid and tetraploid progeny is unclear, although both sexes do exist in laboratory-generated polyploids. Presumably diploidization causes two Z chromosomes (Z) to become autosomes (A_Z). This figure is adapted from Kobel and Pasquier (1986) and Kobel (1996).

How evolutionarily conserved is the "*DM-W/DMRT1* system" of sex determination in frogs and how did this system arise? To what extent and in what way are the components of this crucial system variable across species and ploidy levels, and how is this influenced by natural selection? To begin to address these questions, we examined the molecular evolutionary history of *DM-W* and *DMRT1*. Using a PCR assay, we screened frog species

closely related to *X. laevis* for evidence of *DM-W* and we analyzed phylogenetic relationships between homologous portions of *DM-W* and cloned paralogs of *DMRT1*. Using these data and polymorphism data from natural populations, we tested for evidence of nonneutral evolution of *DM-W* and linked regions of the NRW. We also examined patterns of pseudogenization in paralogs of *DMRT1* in multiple polyploid species of *Xenopus*. Additionally, with an aim of better understanding regulatory evolution of this system, we used a pyrosequening assay to characterize the mechanisms of expression divergence of one *DMRT1* paralog at two crucial stages of development: (1) during primary sexual differentiation in tadpoles and (2) in adult testis.

2.4 Methods

2.4.1 Genetic samples, phylogenetic estimation, codon analysis

We analyzed evolution and expression of *DM-W* and *DMRT1* using genetic samples obtained from a *Xenopus* colony that was in Geneva, field collections by Ben J. Evans, and animals from a commercial supplier (Xenopus Express, Brooksville, FL). Samples from a natural populations of *X. laevis* and *X. gilli* were obtained from Cape Province, South Africa in ponds near Bettys Bay as detailed elsewhere (Evans et al., 1997, 1998) and samples from the population in the Democratic Republic of the Congo (DRC) were collected by Ben J. Evans in the town of Lwiro.

Sequences from *DMRT1* paralogs and *DM-W* were obtained using primers detailed in Supporting information. For *DMRT1*, paralogs were co-amplified and cloned using the TOPO TA cloning kit (Invitrogen). Estimation of evolutionary relationships among homologous portions of *DMRT1* paralogs and *DM-W* was performed using Bayesian phylogenetics as implemented by MrBayes version 3.2 (Ronquist and Huelsenbeck, 2003) using a model selected by the Akaike information criterion (AIC) with MrModeltest (Nylander, 2004). Dirichlet priors were used for nucleotide frequencies, two independent Markov Chain Monte Carlo runs were performed, each for 1,000,000 generations. A *DMRT1* sequence from the toad *Bufo marinus* (Genbank accession number FJ697175) was used as an outgroup. Sationarity was assessed using plots of likelihood values from the posterior distribution using Tracer version 1.5 (Drummond and Rambaut, 2007) and based on this, a burn-in of 100,000 generations was discarded.

Analysis of the *dN/dS* ratio (ω) of the homologous region of *DM-W* and *DMRT1* was performed with the codeml module of PAML version 4.2 (Yang, 1997) using models described in further detail below. The coding sequence analyzed included 98 codons. It did not include the first 17 codons of exon 2 because our forward PCR primer annealed in this region and it did not include the last six codons of exon 3 because this region was deemed nonhomologous. We included *DMRT1* sequences amplified from cDNA so there was not

missing data surrounding the junction of exons 2 and 3 in this analysis.

To explore whether changes within the DM domain of *DM-W* are unique to this gene (and therefore potentially functionally significant), we compared this 68 amino acid long region to the DM domain of the *DMRT1* gene in other lineages including *DMRT1* paralogs of six fish (*Gasterosteus aculeatus* AY867870, *Paralichthys olivaceus* EU490514, *Takifugu rubripes* NM_001037949, *Xiphophorus maculates* AF29187, *Epinephelus merra* EU555179, *Oryzias latipes* AY442916), chicken (*Gallus gallus* AF123456), a toad (*B. marinus* FJ697175), a crocodile (*Crocodylus palustris* EU531727), and various mammals (*Homo sapiens* NC_000009, *Pan troglodytes* XM_528528, *Mus musculus* NM_015826, *Bos taurus* XM_002689628, *Oryctolagus cuniculus* XM_002708188, *Canis familiaris* XM_846402, and *Sus scrofa* NM_214111). We also compared the DM domain of *DM-W* to the DM domain of the *DMY* gene of *O. latipes* (NM_001104680), which is a male-specific sex determining gene, in the context of each genes' closest paralog.

2.4.2 Multilocus polymorphism data

By analyzing expressed sequence tag databases with the assistance of Frédéric Chain, we identified genes in *X. laevis* that we suspected to be singletons due to loss of one paralog after whole-genome duplication. To test for evidence of recent nonneutral evolution, we sequenced these putative single-copy genes and also portions of *DM-W* and linked flanking regions of the NRW. After accounting for differences in mutation rate and the number of silent sites sequenced, our null expectation is that the level of polymorphism at the NRW should be one-fourth that of the other loci as a consequence of differences in mode of inheritance and copy number. The likelihood of neutral evolution, where the effective population size of the NRW is one-fourth that of autosomal loci (i.e., the inheritance scalar equal to 0.25), and nonneutral evolution (the inheritance scalar <0.25), was estimated using a maximum likelihood version of Hudson, Kreitman, Aguadé test (Hudson et al., 1987) with mlHKA version 2 (Wright and Charlesworth, 2004). Divergence from an outgroup was used to accommodate differences in rates of evolution among loci. *Xenopus gilli* and *X. laevis* are sister species and sequences from each were used as an outgroup for the other. PCR primers are provided in Supporting information.

2.4.3 Phylogenetically biased pseudogenization

To better understand pseudogenization after genome duplication in *DMRT1*, we sequenced 367 clones amplified from *DMRT1* exon 2 and 155 clones amplified from *DMRT1* exon 3 from genomic DNA of a total of 18 species including tetraploids: *X. laevis*, *X. borealis*, *X. muelleri*, *X. pygmaeus*, *X. fraseri*, *X. largeni*, *X. clivii*, and an undescribed species *X. new tetraploid 1* (Evans et al., 2004), octoploids: *X. itombwensis*, *X. boumbaensis*, *X. vestitus*,

X. andrei, *X. wittei*, *X. ameiti*, and an undescribed octoploid *X. new octoploid 2* (Evans et al., 2011b), and dodecaploids: *X. longipes*, *X. ruwenzoriensis*, and an undescribed dodecaploid *X. cf. boumbaensis* (Evans, 2007). We used a model-based approach implemented by the program BayesTraits (Pagel and Meade, 2006) to test for phylogenetic bias in the pattern of pseudogenization in this gene. We compared the likelihood of a model with one rate of pseudogenization in both major paralogous lineages of *DMRT1* (α and β) to the likelihood of a model with a separate rate of pseudogenization in each lineage. The rate of change of pseudogenes to functional paralogs was fixed to zero under the assumption that once a paralog became a pseudogene, functional resuscitation is not possible. A phylogeny for this analysis was estimated using up to 3461 base pairs of sequence from cloned paralogs of the recombination activating genes 1 and 2, as detailed in Evans (2007). Following model selection and analytical methods of Evans (2007), phylogenetic relationships were estimated with the data partitioned by codon position using MrBayes version 3.2 (Ronquist and Huelsenbeck, 2003). Two independent runs were performed, each for 2,000,000 generations, and 100,000 generations discarded as burn-in. For the BayesTraits analysis, a chronogram was constructed from the consensus topology recovered from the Bayesian analysis using penalized likelihood as implemented by r8s version 1.71, with the smoothing parameter obtained from a cross-validation procedure (Sanderson, 1997, 2002).

2.4.4 Expression divergence

Gene regulation is orchestrated by *cis*-acting factors that independently affect transcription of each allele and by *trans*-acting factors that affect expression of both alleles. To quantify the degree to which *cis*- and *trans*-acting factors drive *DMRT1* expression divergence between species, we used an experimental approach that compares the expression ratios of species-specific alleles in each species and in their F1 hybrid (Wittkopp et al., 2004, 2008; Landry et al., 2005). In hybrids, *trans*-acting factors from both parental species interact with the *cis*-acting factors of species-specific alleles, so divergent expression of species-specific alleles can be attributed exclusively to *cis*- divergence. In the parental species, expression divergence is the culmination of *cis*- and *trans*-factors. Comparison of the expression ratio of species-specific alleles in hybrids to the parental species therefore allows one to dissect apart and quantify *cis*- and *trans*-contributions to expression divergence between species. For example, *trans*- only divergence is inferred if species-specific alleles are expressed at similar levels in hybrid individuals but at different levels in the parental species. Note that this approach compares expression ratios of species-specific alleles (which are identified by species-specific nucleotide polymorphisms), and does not compare the magnitude of expression. This is because the magnitude of expression between parental individuals and hybrids could vary as a consequence of upstream factors that impact entire genetic networks – even in the absence of expression divergence between species – whereas the expression ratio in parental individuals and hybrids should be affected by mechanisms that operate at the level of an individual locus (Wittkopp et al., 2004). We performed this assay using *X.*

laevis and *X. borealis* on tissue from adult testis and tissue from tadpole stage 53, which is after the gonads develop but before they differentiate in *X. laevis*.

The ratio of expression of species-specific alleles of *DMRT1*α in the parental species and their F1 hybrid was quantified using a Biotage PSQ96 pyrosequencer based on species-specific and paralog-specific single nucleotide polymorphisms in exon 3 (Supporting Information). Expression ratios were estimated in F1 hybrids generated from a cross between an *X. laevis* female and an *X. borealis* male (six tadpoles and two adults) or the reciprocal cross (four tadpoles and two adults), and compared to ratios the corresponding expression ratios in mixtures of tissues from each parental species (seven tadpole and four adult parental mixes). According to the manufacturers protocol, genomic DNA is efficiently removed from the RNA extraction but we nonetheless implemented an optional DNAse digestion step in this procedure. Expression ratios in parental mixes were normalized by the amount of genomic DNA in each mix (which was extracted from an aliquot taken before the DNAse digestion step) following Landry et al. (2005). For genomic DNA normalization we used data from *DMRT1*α, one paralog of the recombination activating gene 2 (*RAG2*), and both paralogs of the recombination activating gene 1 (*RAG1*). Pyrosequencing primers and normalized expression ratios are provided in Supplemental Information and elsewhere (Anderson and Evans, 2009).

Statistical analysis of mechanisms of expression divergence followed Landry et al. (2005) and used the "proc mixed" implementation of restricted maximum likelihood of SAS version 9.1.3 (SAS Institute) with modified scripts provided by Patricia Wittkopp. Students t-tests were computed within the mixed procedure and locus-level significance was interpreted after sequential Bonferroni correction for two tests (Rice, 1989). Application of the Bonferroni correction makes the detection of antagonistic *cis*- and *trans*- divergence more conservative (Anderson and Evans, 2009). Conclusions were identical to those drawn from analysis using a standard Students t-test.

2.5 Results

2.5.1 *DM-W* arose after divergence of *Xenopus* and *Silurana* but before divergence of *X. laevis* and *X. clivii*

The *DM-W* gene was first identified in *X. laevis* but a search of the genome of *S. tropicalis* failed to recover this gene (Yoshimoto et al., 2008). To investigate what other species have *DM-W*, we attempted to amplify and sequence portions of *DM-W* orthologs in species that are more closely related to *X. laevis* than to *S. tropicalis*. We found *DM-W* in *X. andrei*, *X. clivii*, *X. gilli*, *X. itombwensis*, *X. laevis*, *X. largeni*, *X. pygmaeus*, and *X. vestitus* (Genbank accession numbers HQ220848-54; HQ220877-8). These species are tetraploid except *X. itombwensis* and *X. vestitus*, which are octoploid.

Based on a previously published phylogeny of clawed frogs (Figure 2.3), the most distantly related species from *X. laevis* in which we detected a *DM-W* ortholog was *X. clivii*. The phylogenetic position of *X. clivii* is unclear. Mitochondrial DNA strongly supports a sister relationship between *X. clivii* and Clade 1 in Figure 2.3 (Evans et al., 2004) but two tightly linked nuclear loci (*RAG1* and *RAG2*) strongly support a sister relationship between *X. clivii* and Clade 2 in Figure 2.3 (Evans, 2007).

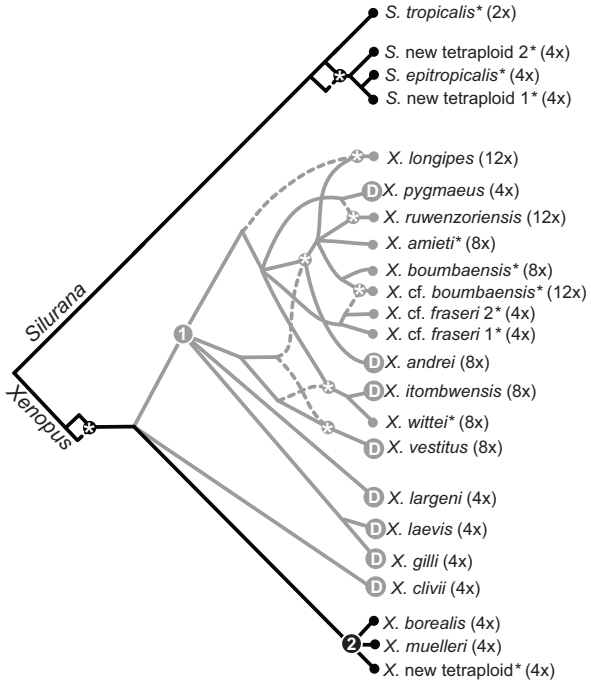

Figure 2.3: *DM-W* was detected in a subset of *Xenopus* species whose most recent common ancestor and descendants are indicated in grey. A "D" next to species names indicates species from which portions of *DM-W* was amplified. Nodes with 1 or 2 inside refer to Clades 1 and 2, respectively. An asterisk follows species names for which amplification of *DM-W* was not attempted. This phylogeny represents reticulating relationships among species that stem from allopolyploidization; allopolyploid speciation is indicated by asterisks on internal nodes and the ploidy level of each species is indicated after the species name. Maternal or biparental relationships are indicated with solid lineages and paternal contributions to allopolyploid speciation events are indicated with dashed lineages. Resolved nodes have >95% posterior probability and are based on the linked autosomal genes *RAG1* and *RAG2* and mitochondrial DNA; details of phylogenetic estimation are given in Evans et al. (2004, 2005) and Evans (2007). Strongly supported but conflicting relationships were recovered for *X. clivii* with respect to numbered Clades 1 and 2 (see text) so this relationship is represented as a polytomy.

We were not able to amplify *DM-W* from the dodecaploid species *X. longipes* and *X. ruwenzoriensis* or the tetraploid species *X. muelleri* and *X. borealis*. The dodecaploids are closely related to species that have *DM-W* (Figure 2.3) so our failure to amplify this gene in these species could be either due to sequence divergence or to loss of *DM-W*. We put extensive effort into attempts to amplify *DM-W* in *X. borealis* using a total of 20 primer combinations that independently targeted each coding exon of *DM-W* over a gradient of different annealing temperatures. As a positive control, amplification of both paralogs of *DMRT1* from *X. borealis* was successful. We therefore suspect that *DM-W* may not be present in this species or potentially in other closely related species in Clade 2 in Figure 2.3 (i.e., *X. muelleri* and an undescribed tetraploid species).

Because *DM-W* is a chimeric protein formed only partially from *DMRT1* (Yoshimoto et al., 2008), evolutionary relationships among homologous portions of these genes (Figure 2.1A) can provide insights into the origin of *DM-W*. To this end, we sequenced or downloaded homologous portions of *DMRT1* from paralogs of the tetraploid species *X. laevis*, *X. borealis*, and *X. muelleri* (accession numbers NM_001096500, NM_001085483, HQ225638-41), and compared these sequences to the homologous region of *X. laevis DM-W* (accession number AB259777). For phylogenetic analysis, the AIC selected the general time reversible model with a gamma distributed among site rate heterogeneity parameter, and base frequencies fixed at equal frequencies. Phylogenetic analysis points to a much more recent origin than was previously suggested (Figure 2.4), and indicates with strong statistical support that the 5′ portion of *DM-W* is closely related to *DMRT1*β of *Xenopus* tetraploids. (The designation of paralogs as α or β is arbitrary and we followed GenBank annotations here). This phylogeny establishes a recent origin of *DM-W* after divergence of *Silurana* and *Xenopus*. Because the relationships within the clade containing *DM-W* and *DMRT1*β are unresolved, we cannot determine from this analysis whether *DM-W* appeared before or after divergence of the ancestor of *X. laevis* from the ancestor of (*X. borealis* + *X. muelleri*). In an attempt to address this, we sequenced portions of the intron between exon 2 and 3 in *DMRT1* paralogs and *DM-W*. However, extensive divergence and insertion/deletion polymorphisms prevented us from confidently assessing homology (data not shown), so we restricted our analysis to the coding region as presented above.

We explored alternative scenarios for the origin of *DM-W* by filtering the post-burn-in posterior distribution of tree topologies recovered from the Bayesian analysis with constraint trees depicted in Figure 2.4. Constraint A retains topologies in which *DMRT1*β of *X. laevis*, *X. borealis*, and *X. muelleri* are monophyletic; these topologies are consistent with an origin of *DM-W* before the divergence of the most recent common ancestor (MRCA) of *X. laevis* and *X. borealis*. Constraint B retains topologies in which *X. laevis DMRT1*β and all of the *DM-W* sequences are monophyletic; these topologies are consistent with an origin of *DM-W* after the divergence of the MRCA of *X. borealis* and *X. laevis*. Constraint C retains topologies in which *DMRT1*β, of *X. borealis* and *X. muelleri* and all of the *DM-W* sequences are monophyletic. Topologies consistent with Constraint C require an explanation involving ancestral polymorphism or gene transfer via hybridization

because they would suggest an origin of *DM-W* after the divergence of the MRCA of *X. laevis* and *X. borealis*, but in the *X. borealis* lineage, even though the *X. laevis* lineage carries *DM-W*. The posterior probabilities of constraints A, B, and C were 0.38, 0.245, and 0.375, respectively. This indicates that there is insufficient statistical power to distinguish between these possibilities with the available data. We repeated this exercise using additional data from partial sequences from *DMRT1* paralogs from other species (discussed below) and recovered essentially identical results (Figure 2.S1).

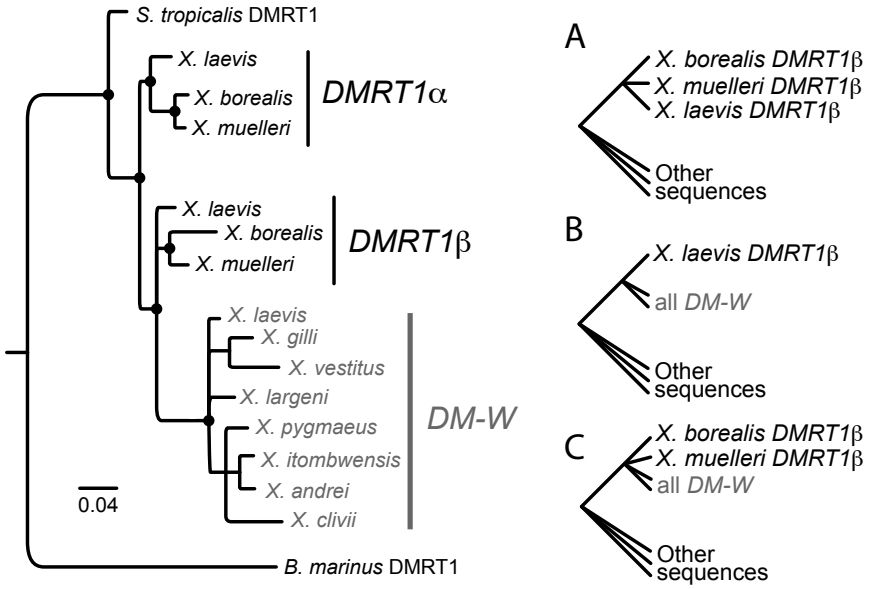

Figure 2.4: *DM-W* originated after divergence of *Silurana* and *Xenopus*. Evolutionary relationships of *X. laevis DM-W* with respect to homologous of *DMRT1* paralogs (α and β) of the tetraploid species *X. laevis*, *X. borealis*, and *X. muelleri*, and the diploid species *S. tropicalis* show that this gene is most closely related to *Xenopus DMRT1*β. *DMRT1* from *Bufo marinus* is used as an outgroup. Nodes with posterior probabilities $\geq$0.99 are indicated with black dots. This analysis fails to resolve whether *DM-W* originated before or after the divergence of ancestors of *X. laevis* and (*X. borealis* + *X. muelleri*). To evaluate alternative scenarios for the origin of *DM-W*, we filtered the post-burn-in posterior distribution of tree topologies with constraint trees A-C. The posterior probabilities of constraints A, B, and C were 0.38, 0.245, and 0.375, respectively.

Overall, these analyses pin down the origin of *DM-W* to a window of time after divergence of *Silurana* + *Xenopus* but before divergence of *X. clivii* + Clade 1 (Figure 2.3). Rough divergence time estimates based on mitochondrial DNA (Evans et al., 2004) suggest that *DM-W* therefore arose about 32-64 million years ago. Another divergence time estimate based on nuclear DNA suggests *DM-W* could be even younger, about 13-38 million years ago (Chain and Evans, 2006).

2.5.2 *DM-W* evolved under natural selection

How did *DM-W* seize control of the sex determination system? One possibility is that this was catalyzed by natural selection on *DM-W* or a linked region (e.g., Rice, 1986; van Doorn and Kirkpatrick, 2007; Vuilleumier et al., 2007). To explore this possibility, we used two approaches to test for evidence of natural selection on *DM-W* and/or the NRW.

We first analyzed the rate ratio of nonsynonymous to synonymous substitutions per site (ω) and recovered significant support for positive selection during the ascendancy and evolution of *DM-W*. Using the codeml module of PAML (Yang, 1998), we performed a branch-site test for positive selection (Yang et al., 2005) on a subset of the data analyzed in Figure 2.4 including homologous regions of both *DMRT1* paralogs from *X. laevis* and *X. muelleri*, *DMRT1* paralog α of *X. borealis*, *DMRT1* orthologs of *B. marinus* and *S. tropicalis*, and the *DM-W* gene of *X. laevis*. *DM-W* sequences from the other species (*X. gilli*, *X. vestitus*, *X. largeni*, *X. pygmaeus*, *X. itombwensis*, *X. andrei*, and *X. clivii*) were excluded because of missing data surrounding the junction of exons 2 and 3. *Xenopus borealis DMRT1* paralog β was excluded because it had premature stop codons at amino acid positions 36 and 59.

This test compares models in which four classes of sites are considered. Sites in one class evolve under purifying selection, sites in another class evolve neutrally, and sites in two other classes evolve under positive selection. The proportion of sites in each class is estimated separately for different portions of the tree (in this case for the *DM-W* branch and for the rest of the topology) and the likelihood of this model is compared to a null model in which no positive selection occurs and sites either evolve neutrally or under purifying selection (Yang et al., 2005). Comparison of the alternative model to the null model recovers significant evidence of positive selection (P <0.01, Table 2.1). In the alternative model, the maximum likelihood estimate of the proportion of sites under positive selection is 0.134 with ω of these sites equal to 10.058.

Table 2.1: Codon analysis of *DM-W* and *DMRT1* supports positive selection on sites in the 5′ region of *DM-W*. Results are shown for a site-branch test, a branch test, and a branch test for neutrality, including the likelihood of the null model (–lnL H_o), likelihood of the alternative model (–lnL H_a), degrees of freedom (df), and *P*-value (*P*). The *P*-value of the site-branch test is based on a 50:50 mixture of point mass 0 and χ^2_1. P-values for the other tests are based on a standard χ^2_1 approximation.

Model	–lnL H_o	–lnL H_a	df	*P*
Site-branch	877.363	873.68	1	<0.01
Branch	889.151	879.181	1	<0.00001
Test for neutrality	879.256	879.181	1	0.698

Analyzing the same data using a model with a different ω across all sites on the *DM-W* branch is significantly preferred over a model in which ω is the same in *DM-W* and *DMRT1* (P <0.00001, Table 2.1). The maximum likelihood estimate of the ω over all sites on the branch leading to *X. laevis DM-W* is 0.8152 as compared to a ratio of 0.0715 across all sites over the rest of the phylogeny. This is also more likely than the neutral expectation (where ω = 1 for the *DM-W* branch), but not significantly so (P = 0.698, Table 2.1). In this analysis, a total of 98 codons were analyzed, 70 of which occur in the DM domain. Of 14.67 inferred nonsynonymous substitutions, slightly fewer occurred in the DM domain than expected (nine substitutions), but not significantly so (P = 0.4279, G test). By contrast, over a similar period of time only four or five amino acid substitutions occurred in the 98 codon region of *X. muelleri DMRT1*β and *X. laevis DMRT1*β, respectively, and only two occurred in *X. muelleri DMRT1*α and *X. laevis DMRT1*α.

Interestingly, there is considerable variation among species in the amino acid sequence of *DM-W*. Seven of 111 codons sequenced from *X. laevis* and *X. gilli* encoded different amino acids in *DM-W* and these divergent codons were evenly distributed among the exons, including in the DM domain. *X. laevis* and *X. gilli* diverged ∼17 million years ago (Evans et al., 2004). In *X. clivii DM-W* from exon 2, 8 of 55 amino acids are diverged from the homologous portion of *X. laevis DM-W*. These species diverged ∼32 million years ago (Evans et al., 2004).

We then used the Hudson, Kreitman, Aguade (HKA) test (Hudson et al., 1987) to test for more recent evidence of natural selection based on a multilocus polymorphism dataset, and again recovered a significant departure from neutrality. We collected polymorphism data from 23 genes from a natural population of *X. laevis* and *X. gilli* from Cape Province, South Africa. Although both species are tetraploid, we sequenced loci that are either single copy due to post-polyploidization deletion, or with sufficiently diverged primer sites that made possible amplification of only one pair of alleles. We compared silent site polymorphism at these loci to silent polymorphism at *DM-W* and linked female-specific flanking regions of the NRW (amplified using primers designed from accession AB365520) in both species (Table 2.2).

Table 2.2: Polymorphism data from a natural population of *X. laevis* from South Africa including names of unlinked loci (locus), the number of base pairs sequenced (bp), number of silent sites (silent), number of segregating sites for *X. laevis* and *X. gilli* (SXL and SXG respectively), number of chromosomes sequenced for *X. laevis* and *X. gilli* (nXL and nXG respectively), divergence from the outgroup (diverge), and Genbank accession numbers (Accession). For *DM-W*, different amounts of polymorphism data were obtained for *X. laevis* and *X. gilli* and seperated by a forward slash respectively. Acronyms for locus names refer to names in Supplementary material.

Locus	Bp	Silent	SXL	nXL	SXG	nXG	Diverge	Accession
AR	402	99	0	30	0	32	5	HQ220654-84
BTB domain protein 6	525	121	0	26	2	24	6	HQ220685-709
LOC100487412	481	458	-	-	3	30	34	HQ225642-57
efcab5	483	473	30	16	4	14	17	HQ220931-42
fem1c	497	112	1	20	4	26	7	HQ221309-31
LOC100488701	581	581	19	30	0	18	14	HQ220994-1016
gja3	717	717	19	14	-	-	21	HQ220943-50
kiaa1919	660	144	0	10	2	26	7	HQ220951-64
LOC100145442	267	267	9	18	-	-	15	HQ220965-93
LOC100158283	690	140	15	14	0	26	15	HQ220974-93
mastl	539	119	2	24	5	26	8	HQ221046-70
mogs	705	170	4	20	0	24	13	HQ221071-92
nfil3	337	77	1	24	0	24	3	HQ221093-120
pcdh1	507	122	1	20	2	26	5	HQ221121-43
pigo	494	126	3	22	1	26	4	HQ221144-67
prmt6	666	157	5	22	6	22	13	HQ221168-90
RAG2	1062	237	11	22	5	18	18	HQ221191-210
rassf10	489	98	3	24	3	26	7	HQ221332-56
zbed4	472	110	2	24	3	22	4	HQ221262-84
LOC100496658	435	435	7	32	0	26	20	HQ221017-45
C7orf25	532	119	2	24	0	24	4	HQ220710-32
sugp2	463	113	3	24	0	26	7	HQ221211-35
znf238.2	526	118	0	24	2	24	11	HQ221285-308
DM-W	2781	2549/1677	0*	12	5	14	48/49	HQ220822-47; HQ220855-76; HQ220879-930

* Two segregating sites were observed in an insertion that occurred in an ancestor of the South African population of *X. laevis* after divergence from *X. gilli*.

In the South Africa population of *X. laevis*, very little polymorphism was detected at the NRW; over 2500 silent sites from the NRW contained only two polymorphisms (Table 2.2). Both polymorphisms were in an insertion/deletion that was present in the South Africa population of *X. laevis* but absent in *X. gilli*. Because gaps were deleted to estimate divergence with respect to an outgroup, this meant that there were no polymorphisms in the region of the *X. laevis* NRW that was homologous to the outgroup (*X. gilli*). These data departed significantly from the neutral expectation ($P = 0.0034$, degrees of freedom (df) = 1, mlHKA). The maximum likelihood estimate of the *DM-W* inheritance scalar was 0.003 as compared to the neutral expectation of 0.250. When the two polymorphisms with no outgroup homology are included in the analysis, departure from the neutral expectation is not significant ($P = 0.319$, df = 1) although the maximum likelihood estimate of the inheritance scalar (0.079) is still below the neutral expectation. Both *X. laevis* polymorphisms were completely linked and had intermediate frequency (5 of 12 individuals).

To further understand the evolution of polymorphism on the *X. laevis* NRW, we sequenced portions of the NRW from a divergent population of *X. laevis* from the Democratic Republic of the Congo (DRC). By sequencing the NRW from this more closely related population, we were able to determine that the insertion containing the polymorphisms in the South Africa *X. laevis* population occurred after divergence from the DRC population, as opposed to the alternative that a deletion occurred in the *X. gilli* lineage. This is because, like *X. gilli*, the *X. laevis* population from the DRC lacks the region carrying the two NRW polymorphisms in the *X. laevis* population from South Africa. Twenty-nine of 1549 sites were diverged between the two *X. laevis* populations in the 5′ upstream region of *DM-W*. No polymorphic sites were detected in the NRW of 11 individuals from the DRC population (data not shown). This could be due to small population size, natural selection, or both, but we did not investigate these possibilities here.

In contrast to the South Africa population of *X. laevis*, the NRW of the *X. gilli* population was highly polymorphic. The NRW of the *X. gilli* contained five polymorphic sites out of about 1700 silent sites, even though polymorphism at autosomal loci was similar in magnitude to the South Africa population of *X. laevis* (Table 2.2). In all of the *X. gilli* NRW polymorphisms, the rare allele was derived relative to *X. laevis*, and four of these rare derived substitutions were present in only one individual. These data from *X. gilli* do not depart significantly from the neutral expectation ($P = 0.260$, df = 1) and the maximum likelihood estimate of the inheritance scalar (0.691) was greater than the neutral expectation of 0.250.

Out of eight substitutions in the DM domain of *DM-W*, four were at positions that were invariant (conserved) in the DM domains of *DMRT1* in the other species we examined, and four were in positions that were variable (non-conserved) in the DM domains of *DMRT1* in these species. By comparison, the DM domain of the *DMY* gene of the fish *O. latipes*, a male-specific sex determining gene, has 13 amino acid substitutions in this region relative to its *O. latipes DMRT1* paralog. Two are in positions that are conserved in other *DMRT1*

loci, and the others are at non-conserved positions in *DMRT1*. Thus although *DMY* is more diverged from its closest *DMRT1* paralog than is *DM-W* in terms of the number of amino acid substitutions, *DM-W* has more divergence at highly conserved sites in *DMRT1* than does *DMY*.

Taken together, our tests for natural selection uncover a dynamic evolutionary history of *DM-W*. It appears that the 5′ portion of this gene was subject to positive selection during early evolution that increased ω of some sites beyond the neutral expectation. More recently, *DM-W* and linked portions of the NRW continued to evolve non-neutrally in *X. laevis* leading to a dearth of polymorphism in this region of the genome. In *X. gilli*, however, recent evolution of the NRW did not depart from the neutral expectation.

2.5.3 Biased pseudogenization of *DMRT1* paralogs after allopolyploidization

How did the male-related gene *DMRT1* respond to the ascendance of *DM-W* and genome duplication? One step to further understanding the evolution of this system is quantify how many paralogs of *DMRT1* are functional in polyploid species of *Xenopus*. To accomplish this, we amplified, cloned, and sequenced portions of *DMRT1* paralogs in 18 polyploid species (Table 2.3). We focused our efforts on exons 2 and 3 because these 5′ regions contain the DM domain, which is key to the DNA-binding function of the protein (Yoshimoto et al., 2010).

Table 2.3: *DMRT1* paralogs cloned from representative tetraploid, octoploid, and dodecaploid species of *Xenopus*. Cloned exons that appear functional are indicated with a "Y", missing data indicated with a dash, and degenerate exons (with premature stop codons, frameshift mutations, or large deletions) are indicated with a "D" with a forward slash dividing results from exon 2 and 3 because each exon was independently amplified for cloning. Two paralogs (α, β) are expected in tetraploids, four (α_1, α_2, β_1, β_2) in octoploids, and six (α_1, α_2, α_3, β_1, β_2, β_3) in dodecaploids. In some instances paralogous divergence could not be positively distinguished from allelic variation and is indicated with a "?". Also indicated are the proportion of paralogs with data that appear functional at the coding level (proportion) and Genbank accession numbers (Accession).

Tetraploid species	α			β			Proportion	Accession
X. laevis	Y/Y			Y/Y			2/2	NM_001096500; NM_001085483
X. largeni	Y/-			Y/- and D*/-			2/2	HQ220775-7
X. muelleri	Y/Y			Y/Y			2/2	HQ225640-1
X. borealis	Y/Y			D/Y			1/2	HQ225638-9
X. fraseri	Y/-			D/Y			1/2	HQ220733-4; HQ220792
X. pygmaeus	-/Y			D/Y			1/2	HQ220737; HQ220790-1
X. new tetraploid 1	Y**/-			Y/-			2/2	HQ220778-83
X. clivii	Y**/Y			-/Y			2/2	HQ220735-6; HQ220788-9
Octoploid species	α_1	α_2		β_1	β_2			
X. vestitus	Y/Y	Y/Y		Y/Y	D***/-		3/4	HQ220738-42; HQ220793-5
X. itombwensis	Y/Y	Y/Y		Y/Y	Y/Y		4/4	HQ220745-9; HQ220796-800
X. boumbaensis	Y/Y	?/Y		Y/Y	-/-		3/3	HQ220750-2; HQ220801-4
X. andrei	Y/-	-/-		Y/-	Y/-		3/4	HQ220772-4;
X. amieti	-/Y	Y/Y		Y/Y	-/-		3/3	HQ220753-4; HQ220805-7
X. wittei	Y/-	Y/-		D/-	Y/-		4/4	HQ220784-7
X. new octoploid 2	-/Y	-/Y		Y/Y	Y/?		3 or 4/3 or 4	HQ220743-4; HQ220818-21
Dodecaploid species	α_1	α_2	α_3	β_1	β_2	β_3		
X. longipes	Y/Y	Y/Y	Y/Y	Y/Y	Y/Y	D/Y	5/6	HQ220755-61; HQ220808-13
X. ruwenzoriensis	Y/-	Y/-	Y/-	Y/Y	-/-	D/-	4/5	HQ220762-6; HQ220814
X. cf. boumbaensis	Y/Y	Y/-	-/-	D/Y	-/?	Y/-	3/4 or 5	HQ220767-71; HQ220815-7

* Segmental duplicate of exon 2 paralog b is degenerate.

** Segmental duplicate of exon 2 paralog a not degenerate.

*** *X. vestitus* β2 is heterozygous for stop codon in exon 2.

Copy number and evolutionary relationships of *DMRT1* paralogs are generally consistent with duplication by whole-genome duplication (Table 2.3). As expected, two *DMRT1* paralogs were usually detected in tetraploid species, up to four were detected in octoploid species, and up to six were detected in dodecaploid species. In some species, fewer paralogs were detected than expected; these could have been lost by deletion or not cloned in our assay. In three tetraploid species, an extra divergent sequence was identified (Table 2.3); these could stem from segmental duplication of one paralog or possibly be examples of very high allelic variation. Two of the extra divergent sequences were *DMRT1*α paralogs and one was a *DMRT1*β paralog.

Translation of the cloned *DMRT1* sequences indicates multiple paralogs of multiple species are degenerate due to stop codons, frameshift mutations, or large deletions (Table 2.3); pseudogenization therefore has a marked impact on the stoichiometry of *DMRT1* and *DM-W* in *Xenopus* females. Eight instances of degeneration were observed in *DMRT1*β and a ninth in a putative singleton duplicate copy of *DMRT1*β in *X. largeni*. To our surprise, however, degeneration was never observed in *DMRT1*α or in either of the putative singleton duplicates of *DMRT1*α (Table 2.3). Under the assumption of allopolyploidization, the "β" paralogs are inherited from one of the two diploid ancestors whose genomes were fused to form an ancestral allotetraploid *Xenopus*. The eight to nine instances of degeneration were observed in exon 2 of *DMRT1*β (and never in exon 3), and each instance was species specific (e.g., a premature stop codon was never observed in the same position in multiple species).

Using BayesTraits (Pagel and Meade, 2006), we recovered support for significant phylogenetic bias in the pattern of pseudogenization in *DMRT1*. When missing data were coded as missing, a model with a different rate of pseudogenization in *DMRT1* paralog α and β (–lnL = –30.8618) was significantly preferred over a model with one rate of pseudogenization in both paralogs (–lnL = –34.3794, $P = 0.00776$, df = 1). In the more parameterized model the rate of pseudogenization was 10 times faster in *DMRT1* paralog β than in *DMRT1* paralog α. As a conservative measure, we also performed this test with missing data coded as functional (there were more missing data from *DMRT1*β than *DMRT1*α), and the result was the same ($P = 0.00781$). This result is consistent with the notion that natural selection preferentially targeted *DMRT1*β for pseudogenization in multiple polyploid species or that natural selection preferentially favoured *DMRT1*α to remain functional. A caveat is that we may have underestimated the number of *DMRT1* pseudogenes if degenerative mutations are present regions that we did not sequence, if seemingly functional paralogs are not expressed due to mutations in the regulatory region, or if we failed to clone some pseudogenized or deleted paralogs.

2.5.4 *DMRT1*α expression is regulated by distinct mechanisms over development

Is it possible that the pattern of biased pseudogenization in *DMRT1*β is somehow linked to the origin of *DM-W* from partial duplication of *DMRT1*β? To explore this question, we examined mechanisms of expression divergence (*cis*- and/or *trans*-) of a *DMRT1*α, in *X. laevis* and *X. borealis* during two developmental stages (tadpole and adult). As discussed earlier and in Table 2.3, *DMRT1*α paralogs appear to be fully functional at the sequence level in both of these species.

A first concern in this analysis is whether expression levels in hybrid individuals differ depending on the direction of cross (i.e., whether the mother of the hybrid individuals is *X. laevis* or *X. borealis*). However, tests for parent of origin effects were not significant (Supporting information) so data from both directions of hybrid cross were considered jointly. Another concern is whether expression levels differ in males and females. In hybrid tadpoles with a *X. borealis* mother, we were not able to test for sex effects because we do not know the sex-determining gene in this species. However, we did test for sex effects in hybrid tadpoles when the mother was *X. laevis* (Supporting information); these tests also were not significant after Bonferroni correction, which is consistent with (Yoshimoto et al., 2008), allowing us to combine expression data from male and female tadpoles. Because the sex effect was individually significant before Bonferroni correction, we repeated analysis presented below using only males and conclusions were essentially identical (Supporting information).

Comparison of expression levels of alleles from each species indicates that *DMRT1*α expression in *X. laevis* and *X. borealis* is significantly different during primary gonadal differentiation in tadpoles, but not in adult testis (indicated by significant "$P = 0$" tests in tadpoles but not adults; Table 2.4). Significant expression divergence of *DMRT1*α in tadpoles during primary gonadal differentiation is interesting from the standpoint that *DM-W* is expressed in *X. laevis* females but possibly not in *X. borealis* if this species lacks this gene. Expression of *DMRT1*α is 1.74 fold higher in *X. laevis* than in *X. borealis*; this is indicated by the $\log_2$-transformed *X. laevis*/*X. borealis* parental ratio of 0.8027 in Table 2.4. Because *DMRT1*β is a pseudogene in *X. borealis* (Table 2.3), the disparity in *DMRT1* expression between *X. laevis* and *X. borealis* during primary gonadal differentiation could be even higher if in *X. laevis* the *DMRT1*α and *DMRT1*β loci are co-expressed, translated, and functionally interchangeable.

How can we account for these expression differences? In the simplest scenario, a single mutation could cause different levels of expression in two species; such a mutation could occur in *cis*- (e.g., affecting an upstream promoter) or in *trans*- (e.g., affecting expression level or binding capacity of a transcription factor). If *DMRT1*α expression were governed by the same regulatory elements throughout development and across multiple tissue types, one would expect a similar direction of divergence (e.g., upregulation of *DMRT1*α in *X. lae-*

Table 2.4: Log_2-transformed parental and hybrid expression ratios (*X. laevis*/*X. borealis*) of *DMRT1*α and probabilities of the null hypothesis that this ratio is equal to zero (P = 0), that ratios in parental species and hybrids are equal (P = H), and that the hybrid ratio is equal to zero (H = 0). Interpretations include *cis*- divergence (C) and *trans*- divergence (T) with each mechanism causing upregulation of the *X. laevis* (XL) or *X. borealis* (XB) allele. Inferences are based on two tests, (P = H) and (H = 0), and significant departures of these test after Bonferroni correction for two tests are indicated with asterisks. Significant departure of (P = 0), is also indicated with asterisks.

	Parental ratio	*P*-value (P = 0)	Hybrid ratio	*P*-value (P = H)	*P*-value (H = 0)	Interpretation
Tadpole	0.8027	0.0068*	–0.3982	0.0070*	0.0053*	C: XB up, T: XL up
Adult	1.1491	0.0846	3.0141	0.0250*	0.0001*	C: XL up, T: XB up

vis compared to *X. borealis*) and a similar relative magnitude of divergence (e.g., twofold) in all developmental stages and all tissue types where the gene is expressed. Contrary to these expectations, we found that *DMRT1*α regulation occurs in a developmental-stage specific manner. In tadpoles, *DMRT1*α upregulation of *X. laevis* compared to *X. borealis* alleles is attributable to *trans*- acting factors (this is indicated by a significantly higher parental ratio than the hybrid ratio in tadpoles according to the "*P* = H" test; Table 2.4). But this is counteracted to some degree by antagonistic *cis*- upregulation of the *X. borealis DMRT1*α allele (because the log_2-transformed *DMRT1*α hybrid tadpole ratio is significantly higher than 0 according to the "H = 0" test; Table 2.4). In adult testis, however, the opposite mechanisms operate: *trans*- divergence upregulates the *X. borealis* allele (the adult hybrid ratio is significantly higher than the adult parental ratio according to the "*P* = H" test; Table 2.4) and *cis*- divergence upregulates the *X. laevis* allele (the adult hybrid ratios are both significantly higher than zero according to the "H = 0" test; Table 2.4).

In *X. laevis*, *DMRT1* is autosomal (Uno et al., 2008; Yoshimoto et al., 2008). But if *DMRT1* inheritance is not biparental in *X. borealis* (e.g., if a sex-linked allele of one paralog was deleted), this could affect conclusions relating to mechanisms of expression divergence in tadpoles because some *X. borealis* tadpoles would be hemizygous for one *DMRT1* paralog, and because some hybrids would lack one *DMRT1* paralog from *X. borealis*. To investigate this possibility, we co-amplified and sequenced exon 2 of *DMRT1*α and *DMRT1*β from genomic DNA of F1 hybrids generated by crossing an *X. borealis* female to an *X. laevis* male, and also from the reciprocal cross with an *X. laevis* mother. If both paralogs of *DMRT1* are biparentally inherited in both species, then hybrids from both crosses should both inherit an allele of both paralogs from both species. Alternatively, if both alleles of one *DMRT1* paralog were not biparentally inherited in *X. borealis*, 50% of the F1 hybrids from one of these crosses would inherit an allele from only one paralog from *X. borealis*. This is because they would either receive different sex chromosomes from their *X. borealis* mother (if females are heterogametic) or from their *X. borealis* father (if males

are heterogametic). We used adult female and male hybrids (12 females, six males from the cross with the *X. borealis* mother and seven of each sex from the cross with the *X. borealis* father) to test this. Species- and paralog-specific single nucleotide polymorphisms allowed us to confirm that F1 hybrids inherit both alleles from *DMRT1* paralogs from both species. We note that biparental inheritance of both alleles of both *DMRT1* paralogs in *X. borealis* does not necessarily mean that *DMRT1* is not sex-linked in this species.

Overall then, analysis of *DMRT1*α expression divergence illustrates two main points with respect to the evolution of variation in mechanisms of sex determination in clawed frogs. First, *DMRT1*α is expressed significantly higher during gonadal differentiation in a *DM-W*-containing species (*X. laevis*) compared to one that may lack *DM-W* (*X. borealis*) but not significantly different in adult testis. Second, genetic mechanisms that govern expression of *DMRT1*α during primary gonadal differentiation in tadpoles are at least partially independent from those that govern expression in adult testis. We relate these expression findings to our results concerning biased pseudogenization of *DMRT1* below.

2.6 Discussion

Almost every vertebrate species has separate sexes, but mechanisms by which sexual differentiation is orchestrated vary considerably. In frogs, variation in male versus female heterogamy is observed among frog families and genera (Hillis and Green, 1990) and within the species *Rana rugosa* (Miura, 2007). Presence/absence polymorphism of a W sex chromosome occurs in *Leiopelma hochstetteri* (Green et al., 1993), and mapping of sex-linked genes demonstrates that ranids and pipids have different sex chromosomes (Uno et al., 2008). Here we illustrate that mechanisms of sex determination also vary within the African clawed frogs (*Xenopus* + *Silurana*) – and possibly within the genus *Xenopus* – as a consequence of the appearance of a novel female- specific sex-determining gene called *DM-W*. In *X. laevis*, *DM-W* triggers primary (gonadal) female development, and this may occur via competitive inhibition in females of genes that are activated by *DMRT1* in males (Yoshimoto et al., 2008, 2010). In contrast to analysis of the full *DM-W* and *DMRT1* genes (Yoshimoto et al., 2008), analyses presented here demonstrate that this gene evolved in *Xenopus* after divergence from *Silurana* but before divergence of *X. laevis* and *X. clivii*. That we did not succeed in amplifying *DM-W* in *X. borealis* despite considerable effort opens the possibility that this and other closely related species in Clade 2 (Figure 2.3) do not have *DM-W*. Similar lines of evidence suggested that the mammalian sex-determination locus *SRY* was not present in platypuses (Grützer et al., 2004) before confirmation with the complete genome sequence (Warren (2008) and others), or in various vole species (Just et al., 1995), and that *DMY* (a paralog of *DMRT1*) is not present in species closely related to the medaka (Kondo et al., 2003, 2004).

At this point, it is not clear whether *DM-W* evolved in a diploid genome with 18 chro-

mosomes or a tetraploid genome with 36 chromosomes. The probability of fixation of *DM-W* in a diploid versus a tetraploid is presumably influenced by multiple factors such as dominance interactions between *DM-W* and the ancestral trigger for sex determination, whether a tetraploid genome is polysomic or disomic (Otto and Whitton, 2000), whether *DM-W* arose on the same chromosome or a different chromosome as the ancestral sex-determining locus, whether *DM-W* conferred advantages to one or both sexes, and whether the ancestral heterogametic sex was female or male. If *DM-W* did evolve after divergence of the most recent common ancestor of *X. laevis* and *X. borealis*, various scenarios depicted in Figure 2.S2 involving multiple episodes of allo-tetraploidization in *Xenopus* are possible. Identification of the sex-determining locus in *S. tropicalis*, and a complete genome sequence for *X. laevis* will assist in exploring these possibilities. Also relevant to the question of exactly when *DM-W* arose is the relationship among *X. clivii* and Clades 1 and 2 in Figure 2.3. Most relevant is the relationship among genomic regions linked to *DM-W* and *DMRT1*β in these lineages. We attempted to address this question by sequencing portions of the intron between exons 2 and 3 of *DM-W* and *DMRT1*β in *X. laevis*, *X. borealis*, *X. muelleri*, and *X. clivii* but divergence prevented us from making reliable homology statements for these sequences. Analysis of the available data with unambiguous alignment provided insufficient statistical support to distinguish between an origin before or after divergence of the MRCA of *X. laevis* and *X. borealis*.

Multiple mechanisms have been proposed to account for variation among species and within populations in the mechanisms of sex determination. For example, sexual antagonism and other types of genetic conflict could potentially drive sex chromosome turnover (Rice, 1986; Werren and Beukeboom, 1998; van Doorn and Kirkpatrick, 2007). Sex chromosomes have a smaller effective population size than the autosomes (barring extreme sex-specific demography) and genetic drift thus has a more profound impact on their evolution. Genetic drift, possibly combined with sex-ratio selection, could also contribute to sex chromosome turnover (Bull, 1983; Kozielska et al., 2996; Viulleumier et al., 2007). Another possibility is that sexual selection and addition of upstream elements drives evolution of sex determination systems (Pomiankowski et al., 2004; Wilkins, 1995). Consistent with a role for natural selection in sex chromosome turnover and evolution, we found that some sites in the 5′ portion of *DM-W* evolved under positive selection since divergence from *DMRT1*β and that *DM-W* and linked regions have a significantly lower level of polymorphism in the South Africa population of *X. laevis*. Other examples of rapidly evolving sex determining loci have been reported (Tucker and Lundrigan, 1993; Whitfield et al., 1993; Zhang, 2004; Graves, 2008), but it is still unclear whether there is a general role for natural selection during sex chromosome turnover and evolution, or whether purifying selection on these loci tends to be relaxed.

Arguing against relaxed purifying selection on *DM-W*, we found a dearth of molecular polymorphism in the NRW of a natural population of *X. laevis* from South Africa. The NRW of a DRC population of *X. laevis* is diverged from the South African population, so it appears that non-neutral evolution of *DM-W* or linked portions of the NRW occurred in the

South African population after divergence from the DRC population, as opposed to being a species-wide phenomenon. It is also worth pointing out that the DRC population may warrant separate species status (Measey and Channing, 2003), in which case one would not expect a selective sweep in the *X. laevis* from South Africa to involve the DRC population. An alternative demographic explanation for these data is a higher variance in reproductive success in females than males (Lande and Barrowclough, 1987; Engen et al., 2007). This could happen, for instance, if females skipped some mating cycles to develop a full clutch of eggs. However, this was not the case in a population of *X. gilli*, which had a higher level of polymorphism than the neutral expectation (although not significantly so).

2.6.1 Marginalized expression of *DMRT1*β

In general, the probability that one paralog of a duplicate pair is a pseudogene increases with the age of the duplicates (Lynch and Conery, 2000). Given that $\sim$60% of duplicate genes in the pseudotetraploid species *X. laevis* have degenerated to singletons (Sémon and Wolfe, 2008), it is no surprise that some *DMRT1* paralogs of the *Xenopus* species we surveyed are pseudogenes. What is surprising, however, is that pseudogenization of *DMRT1* paralogs occurred in a phylogenetically biased pattern in multiple polyploid species, affecting only paralogs closely related to *DM-W* (that is, the *DMRT1*β paralogs). A similar pattern of biased pseudogenization was observed at the *RAG1* locus in African clawed frogs (Evans, 2007). How can these findings be explained?

Analysis of *DMRT1* in *S. tropicalis* indicates that this gene is expressed during much of tadpole development, including well before gonadal differentiation and adult testis (Yoshimoto et al., 2006). A broad expression pattern of *DMRT1* is also observed in *X. laevis* based on real-time PCR assays (Osawa et al., 2005; Yoshimoto et al., 2006, 2008), suggesting that a broad expression domain is the ancestral condition of *DMRT1* in African clawed frogs. However, these *X. laevis* assays probably jointly quantified expression of *DMRT1*α and *DMRT1*β, and it is therefore possible that the expression domain of each of these paralogs is diverged.

One explanation for these observations is that the expression domain of *DMRT1*β was marginalized soon after it was formed by whole-genome duplication, making this paralog dispensable in *Xenopus* polyploids. Consistent with this speculation, our results demonstrate that expression of *DMRT1*α is orchestrated by a nonidentical suite of regulatory mechanisms during development that are therefore distinct and independently mutable (Table 2.4). If this developmentally compartmentalized regulation is the ancestral condition for both *DMRT1* paralogs in *Xenopus*, this opens the possibility that mutation could have degenerated regulatory machinery of *DMRT1*β that drives expression in adult testis and other tissues, while leaving intact expression during primary gonadal differentiation. Also consistent with the hypothesis that the expression domain of *DMRT1*β was marginalized soon after genome duplication is the observation that the expression domain of *DM-W* – a

gene formed from partial gene duplication of the 5′ region of *DMRT1*β and potentially also portions of its regulatory region – is restricted to a brief window during primary differentiation in the female gonad (Yoshimoto et al., 2008). Thus, we speculate that the narrow expression domain of *DM-W* could offer clues into why *DMRT1*β paralogs appear to be dispensable in *Xenopus* polyploids.

Future work aimed at independently characterizing the expression domain of *DMRT1*α and *DMRT1*β in multiple *Xenopus* species could further test the hypothesis that the ancestral expression domain of the progenitor locus of *DMRT1*β and *DM-W* was marginalized prior to the origin, functional diversification, and biased pseudogenization of *DMRT1*β. It would also be useful to evaluate whether the upstream portion of *DMRT1*β is homologous to that of *DM-W* when *X. laevis* genome sequences are available. We were not able to detect homology between available sequences from the NRW of *X. laevis* and sequences upstream of *S. tropicalis DMRT1* obtained from the genome sequence (Hellsten et al., 2010).

Why is *DMRT1*α expression in tadpoles upregulated in *X. laevis* compared to *X. borealis*? At this point, we do not know, but one possibility is that this is a consequence of sexual antagonism in *X. laevis* where upregulation of *DMRT1* is favoured in males. If this is the case, developmental compartmentalization of *DMRT1*α regulation could have facilitated regulatory response to the invasion of *DM-W* by decoupling regulatory control in tadpoles and adults (we did not detect significant difference in expression level of *DMRT1*α in adult *X. laevis* and *X. borealis*). Alternatively this expression divergence could be unrelated to *DM-W* and instead be due to neutral drift, or related to other physiological differences between *X. laevis* and *X. borealis*.

We did not consider mechanisms of expression divergence between *X. laevis* and *X. borealis DMRT1*β because of premature stop codons in the *X. borealis* paralog. However, in *X. muelleri*, a close relative of *X. borealis*, the reading frames of both *DMRT1* paralogs appear intact. If Clade 2 (Figure 2.3) really does lack *DM-W*, an interesting direction for further work would compare expression levels and mechanisms of expression divergence of both paralogs of this species to a *DM-W*-containing species such as *X. laevis* during gonadal differentiation. Another question that remains unanswered is how expression stoichiometry of *DMRT1* and *DM-W* expression varies in species with higher ploidy levels at the RNA and protein level.

2.7 Conclusions

Almost every vertebrate species has separate sexes, but mechanisms by which sexual differentiation is orchestrated vary considerably. Here we investigated many aspects of the evolution of a newly identified, partially paralogous female sex-determining gene, *DM-W*, across the breadth of the African clawed frog (*Xenopus*) phylogeny, and at a finer scale between closely related subspecies. Our findings suspect that mechanisms of sex deter-

mination may vary within the African clawed frogs (*Xenopus* + *Silurana*) – and possibly within the genus *Xenopus* – as a consequence of the appearance of *DM-W*. *DM-W* evolved some time after the split of *Silurana* from *Xenopus*, but before the split of *X. clivii* from the remainder of *Xenopus* species, and is more closely related to *DMRT1*β than *DMRT1*$alpha$. The β paralog underwent phylogenetically biased pseudogenization after polyploidization, and the expression of this gene is regulated by mechanisms that vary through development. We also demonstrated that *DM-W* evolved non-neutrally. Together, the infiltration of *DM-W* as a sex determining gene may have been influenced by marginalization of the expression domain of *DMRT1*β, and fixed through natural selection. These findings illustrate the influence and consequences of gene duplication in the evolution of a novel genetic mechanism of sex determination.

2.8 Acknowledgements

We thank F. Chain for identifying singleton genes in *X. laevis*, P. Wittkopp for providing SAS scripts for analysis of expression divergence, A. Smith for assistance with pyrosequencing, and Catherine Peichel and anonymous reviewers for constructive comments.

2.9 Supplementary information

Below we provide supplemental information on (1) PCR primers, (2) phylogenetic relationships between the homologous portion of *DM-W* and *DMRT1*, (3) expression ratio data, (4) tests for parent of origin and sex effects, and (5) possible evolutionary scenarios for *DM-W* under the assumption that it evolved after divergence of the most recent common ancestor of *X. laevis* and *X. borealis*.

(1) PCR primers

Primers used to amplify the 5′ UTR and exons of *DM-W*

DMW_5pr_for_71: GCAGCAGTAACCCGACAGCAGTCC
DMW_5pr_rev_810: AGCATTCATATGGGCTATGAGTTTG

DMW_5P_for_1300: CTCTGGCTGCTGGAGTGCTTGTG
DMW5Prev2229_in_repeat2: ACCCACAGATTGCCAGTCCAG

DMW_5pr_for_2118: CTGTGAATCAGATAGAGATG
DMW_5pr_rev_2870: GCCACCCCTTTAAGTKCCATCAG

DMW_5pr_for_2762: TATTCAGTGTTCTGGCATTGAGTGC
DMW_5pr_rev_3122: GCTTTCCACTGCATGCCCAGTC

DMW_exon2_for2b: GAACCATATAACACCGGGCAGT
DMW_exon2_rev2: CTCCAGATACAGAAGCGCTTA

DMW_exon3_for4: TGCATTGCAAAGACAGCAAGCT
DMW_exon3_rev4: GGTTGTCTGTGTGGAACTGC

DMW_exon4_for2: AGGAAGAAGAGGTGGCTAAAC
DMW_exon4_rev2: CATGAGCTGCTGGATCATCGC

Primers used to amplify *DMRT1*α and *DMRT1*β

DMRT1_exon1_for_16: ATACAGAGAATGCAAAACAATGAGG
DMRT1_exon1_for_43: AAMCATATAGCAAGACCCGTARCWCCG
DMRT1_exon1_rev: ACCTGWGCCGCCATAACYCG

DMRT1_exon2_for: TGCGAAGACAGCAAGCCC
DMRT1_exon2_rev_new: CTGAACTGGYTGTKGAAC

Primers used for analysis of mechanisms of *DMRT1* expression divergence

Forward Primer: GGAATMAGCCATCCWATCCMTTTGC
Starts at position 4693

Reverse Primer (biotinylated): TGTKGAACCTGAAGTGGGTGTGC
Ends at position 4805

Sequencing primer #3 (identifies *X. laevis DMRT1*α): TCTTGCTTGATGYTGGAAARC-AGT
Starts at position 4765
SNP target is at position 4791
T = *X. laevis DMRT1*α
C = Other paralogs

Sequencing primer #1 (identifies *X. borealis DMRT1*β): GGAATMAGCCATCCWATC-CMTTTGC
Starts at position 4693
SNP target is at position 4719
T = *X. borealis DMRT1*β
C = Other paralogs

Sequencing primer #4 (identifies *X. borealis DMRT1*α): TGATAAAAARGGARCATG-GTGGTAGCA
Starts at position 4736
SNP target is at position 4764
T = *X. borealis DMRT1*α
C = Other paralogs

Primers used for analysis of polymorphism

androgen receptor alpha isoform (AR)
XLAR_for_40: AGGGCTCGGCGGGGTATACAAACAGC
XLAR_rev_431: GGCGCTATCAGAGATGCCTTCG

BTB domain protein 6, mRNA (BTB)
Exon19_fora: AGGTTTGCCAATCACTCCAG
Exon19_reva: TCTGTCATTCCCTCCTGTCC

Xenopus (*Silurana*) *tropicalis* clone CH216-35D21 (CH216-35D21)
Scaf351_for_8b: GAAGTCTGAYTGTGAAGTG
Scaf351_rev_8a: CCGCACACCTTCTGAGCCA

Xenopus (*Silurana*) *tropicalis* clone CH216-60D12 (CH216-60D12)
Scaf482_for_11a: TGCAGTTCCAAGAATGGAC
Scaf482_rev_11a: GATTTTAAKGCACTTTGTACC

fem-1 homolog c (fem1c)
Exon21_fora: TTTGTTGTCGTTTGCAGAGC
Exon21_reva: TGTGCGAATTCGTAGAGTCG

Xenopus (*Silurana*) *tropicalis* clone ISB1-27B20 (ISB1-27B20)
Scaf1495_for_5a: GTGGGATCTGTGAKTGGATG
Scaf1495_rev_5b: CCTTCTTTGCCMMCTGTGATT

Xenopus (*Silurana*) *tropicalis* clone ISB-266M19 (ISB-266M19)
Scaf486_for_15a: GGAAGAGCTGCTTATCATG
Scaf486_rev_15a: CCTTAATGTACCCATTAGC
Scaf486_for_15b: CTTCAAATGGAAGAGCTGC
Scaf486_rev_15b: CGCTGCTWACACTCTCCCC

sodium-dependent glucose transporter 1 (kiaa1919)
Exon3_for1: GTCTATGTTATCCTATATAGTCATTG
Exon3_rev2: CTCTGRTATCACAGTAACTG

Xenopus (*Silurana*) *tropicalis* hypothetical protein LOC100145442 (LOC100145442)
Scaf1027_for_7c: GCTCCTCRACCTACACYCTGACC
Scaf1027_rev_7b: CACTGTCTGAYAAACTGC

Xenopus laevis hypothetical protein LOC100158283 (LOC100158283)
Exon2_for1: ACATCAGGGAGATACGCTATACGTGCAGGG
Exon2_rev2: CAGGTGGGAGATGACGCTGAAGC

microtubule associated serine/threonine kinase-like (mastl)
Exon13_fora: CAGGCAAGAAGCAAGAAACC
Exon13_reva: GGTACGAGGTGCGGATATGT

mannosyl-oligosaccharide glucosidase (mogs)

Exon5_for2: CTGAAGATGAGCGGCATGTGGATCTG
Exon5_rev2: CTTCAGCCATGATTAGTACCAC

nuclear factor, interleukin 3 regulated (nfil3)
Exon10_fora: AGCAACTGCAGAAGCACTGA
Exon10_reva: GATGCCACTGACCTGGTTTT

protocadherin 1 (pcdh1)
Exon9_fora: CCTCTCCAGCATCTCCTTTG
Exon9_reva: CGTTGTTTGCTTCGCTCATA

phosphatidylinositol glycan anchor biosynthesis, class O (pigo)
Exon11_fora: GTGAAGAGACTCCCGACTGC
Exon11_reva: CTTCCTCTTCATCCGTTCCA

protein arginine methyltransferase 6 (prmt6)
Exon4_for1: GACCRSGAGTATTTCCAGTGCTACTC
Exon4_rev2: TGCGAATCCGTGCAACAAG

recombination activation (RAG-2) gene (RAG-2)
Rag2_for_45: CTGGGAGTAATACATCATGATC
Rag2_rev_1149: CCTCGTCAAAATGTTCCCGTCTCTG

Ras association (RalGDS/AF-6) domain family (N-terminal) member 10 (rassf10)
(RalGDS/AF-6; N-terminal; rassf10)
Exon16_fora: CTCGGTGGAGAAAATGGAAA
Exon16_reva: GTTCAGCCTCAACCCAATGT

zinc finger, BED-type containing 4 (zinc finger; BED 4)
Exon14_fora: CAATTTGTTCTGCCGACTCA
Exon14_reva: TGTCCGACTGCTCATCCATA

Xenopus (*Silurana*) *tropicalis* unknown (unknown protein)
Scaf1434_for_14a: GCGACTCATGGAACTTACGG
Scaf1434_rev_14a: CTTCTTCTTCTCMATCAGC

UPF0415 protein C7orf25 homolog (UPF0415)
Exon7_fora: CTGGTGGTTGATGTTGTTGC
Exon7_reva: GTGGAAGCACCTTTTCTTG

splicing factor, arginine/serine-rich 14 (sfrs14)
Exon12_fora: CTGAACGCCGATTACAGGAT
Exon12_reva: GTTGAACTTGGCCCACCTTA

zinc finger protein 238, gene 2 (znf238.2)
Exon25_fora: CAAGCCGGTAGACTCTGAGG
Exon25_reva: TCCATTTCATCCTCGCTTTC

(2) Phylogenetic relationships estimated using additional partial data from other *DMRT1* paralogs.

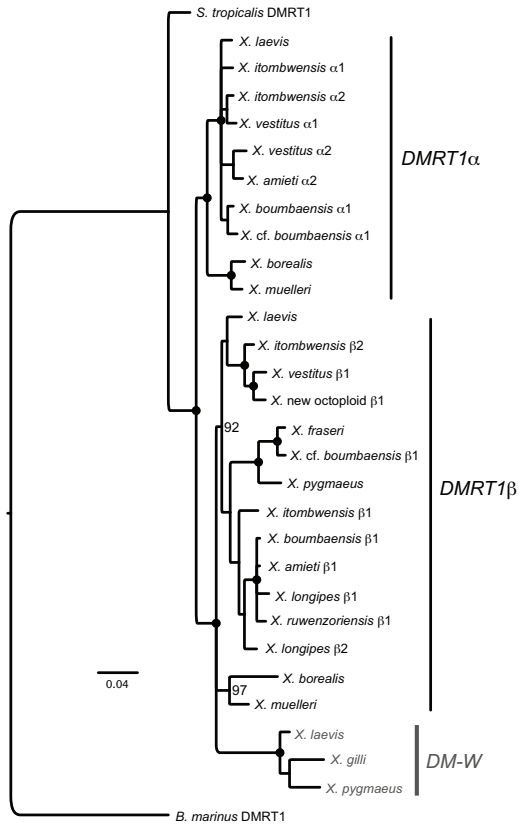

Figure 2.S1: For some species were able to clone and sequence both exons from a paralog or multiple paralogs, and unambiguously concatenated them based on their phylogenetic relationships (Table 2.3). These sequences have missing data from portions of the coding region because the primers annealed to this region and the analysis in Figure 2.4 is thus taxonomically more limited but with more complete data for each *DMRT1* paralog. In this analysis, we did not include sequences from paralogs for which we were only able to obtain sequence from one exon or for which we could not unambiguously concatenate based on phylogenetic relationships. Labelling as in Figure 2.4 with the additional indication of posterior probabilities above 90 expressed as percentages and results are similar to those presented in Figure 2.4. The posterior probabilities of constraints similar to A, B, and C in Figure 2.4 are 44.8%, 19.8%, and 33.4% respectively.

(3) Expression ratio data

Table 2.S1: Raw data used in analysis of mechanisms of expression divergence in *DMRT1*. Indicated are individual codes for each biological replicate (Replicate), the name of the alleles being compared (Alleles), whether the mother of the hybrid (Mother) was *X. laevis* or *X. borealis* (not applicable (NA) for parental mixes), the sex of the individual (not applicable for hybrid tadpoles with a *X. borealis* mother and refers only to the sex of the *X. laevis* individual in tadpole parental mixes), what type (Type) of ratio was assayed (H for hybrid or P for parental mix), the age of the sample (T for tadpole or A for adult), and the $\log_2$ transformed expression ratio (*X. laevis*/*X. borealis* for *DMRT1*α or *DMRT1*β and *DMRT1*α/β for within species ratios of paralogs).

Replicate	Alleles	Mother	Sex	Type	Age	Ratio
FPM12	*DMRT1*α	NA	F	P	T	1.08588
PTM_2	*DMRT1*α	NA	F	P	T	-0.03767
FPM_2	*DMRT1*α	NA	M	P	T	1.88342
FPM_7	*DMRT1*α	NA	M	P	T	0.96046
FPM_9	*DMRT1*α	NA	M	P	T	1.18894
PTM_5	*DMRT1*α	NA	M	P	T	0.11365
PTM_6	*DMRT1*α	NA	M	P	T	0.42398
MIX_1	*DMRT1*α	NA	M	P	A	0.65424
MIX_3	*DMRT1*α	NA	M	P	A	2.81420
MIX_6	*DMRT1*α	NA	M	P	A	0.46942
MIX_7	*DMRT1*α	NA	M	P	A	0.65835
HT_11	*DMRT1*α	XL	F	H	T	-0.48161
TM_16	*DMRT1*α	XL	F	H	T	-0.76975
TM_17	*DMRT1*α	XL	F	H	T	-1.06831
TM_22	*DMRT1*α	XL	F	H	T	-0.84362
HT_12	*DMRT1*α	XL	M	H	T	0.07486
TM_20	*DMRT1*α	XL	M	H	T	-0.27481
HT_1	*DMRT1*α	XB	NA	H	T	0.03807
HT_4	*DMRT1*α	XB	NA	H	T	-0.09645
HT_7	*DMRT1*α	XB	NA	H	T	-0.33118
HT_9	*DMRT1*α	XB	NA	H	T	-0.22904

Table 2.S1: *Continued from previous page*

Replicate	Alleles	Mother	Sex	Type	Age	Ratio
HYB_A	*DMRT1*α	XL	M	H	A	3.46003
HYB_B	*DMRT1*α	XB	M	H	A	3.51541
HYB_X	*DMRT1*α	XB	M	H	A	2.29764
HYB_Y	*DMRT1*α	XB	M	H	A	2.78351
MIX_1	XB *DMRT1*α/β	NA	M	P	A	1.37601
MIX_6	XB *DMRT1*α/β	NA	M	P	A	1.04006
MIX_7	XB *DMRT1*α/β	NA	M	P	A	1.50863
FPM12	XB *DMRT1*α/β	NA	F	P	T	1.23466
FPM13	XB *DMRT1*α/β	NA	F	P	T	0.49849
PTM_2	XB *DMRT1*α/β	NA	F	P	T	1.57153
FPM_2	XB *DMRT1*α/β	NA	M	P	T	1.04054
FPM_7	XB *DMRT1*α/β	NA	M	P	T	0.36053
FPM_9	XB *DMRT1*α/β	NA	M	P	T	1.47578
FPM10	XB *DMRT1*α/β	NA	M	P	T	2.90647
PTM_5	XB *DMRT1*α/β	NA	M	P	T	1.90597
PTM_6	XB *DMRT1*α/β	NA	M	P	T	1.10201
HYB_A	XB *DMRT1*α/β	XL	M	H	A	1.81863
HYB_B	XB *DMRT1*α/β	XB	M	H	A	0.85962
HYB_Y	XB *DMRT1*α/β	XB	M	H	A	1.68598
HT_11	XB *DMRT1*α/β	XL	F	H	T	0.58025
TM_16	XB *DMRT1*α/β	XL	F	H	T	1.09437
TM_17	XB *DMRT1*α/β	XL	F	H	T	2.07516
TM_22	XB *DMRT1*α/β	XL	F	H	T	1.83508
TM_23	XB *DMRT1*α/β	XL	F	H	T	1.80169
TM_24	XB *DMRT1*α/β	XL	F	H	T	0.79931
HT_12	XB *DMRT1*α/β	XL	M	H	T	0.67214
TM_18	XB *DMRT1*α/β	XL	M	H	T	0.58620
TM_20	XB *DMRT1*α/β	XL	M	H	T	1.86825

Table 2.S1: *Continued from previous page*

Replicate	Alleles	Mother	Sex	Type	Age	Ratio
HT_1	XB *DMRT1*α/β	XB	NA	H	T	1.22747
HT_4	XB *DMRT1*α/β	XB	NA	H	T	0.58637
HT_7	XB *DMRT1*α/β	XB	NA	H	T	0.91004
HT_9	XB *DMRT1*α/β	XB	NA	H	T	1.20500

(4) Tests for sex-specific and parent of origin effects in *DMRT1*

In hybrid tadpoles, we tested whether expression ratios of species-specific alleles varied by parent of origin (i.e., whether having an *X. laevis* (XL) or and *X. borealis* (XB) mother impacted expression). No significant effect was detected in hybrid tadpoles (P = 0.1056 and 0.4304 for tadpoles with an XL or an XB mother respectively). The parent of origin test was not possible for hybrid adults because of small sample size of the cross with an XL mother. Based on these results, we pooled tadpole data from both types of hybrid cross for subsequent analysis.

Because we have no way to identify sex in XB tadpoles with undifferentiated gonads, we were not able to test for sex effects and we assume no sex bias in expression in this species. We did, however, use a PCR assay (Yoshimoto et al., 2008) to determine the sex of hybrids tadpoles with an XL mother and also XL tadpoles that were used in the parental mixes for the tadpoles. We used this information to test whether there were sex differences in the following two expression ratios: XLDMRT1α/ XBDMRT1α and XBDMRT1α/ XBDMRT1α. After sequential Bonferroni correction, these ratios did not vary significantly by sex in the hybrid tadpoles (P = 0.0306 and 0.5057 respectively) and neither varied significantly by sex in the tadpoles used in the parental mixes (P = 0.5545 and 0.5375 respectively). These results are consistent with another study (Yoshimoto et al., 2008) that reported similar *DMRT1* expression levels in XL tadpoles of each sex and allowed us to pool tadpole data from both sexes.

Because the sex effect was near significance for the XLDMRT1α/XBDMRT1α ratio, we repeated the analysis using only males and hybrid tadpoles with unknown sex (because their mother was XB). Conclusions drawn from this analysis were similar to the analysis of both sexes combined and significantly support different regulatory mechanisms acting at different developmental stages (Table 2.S2).

Table 2.S2: Log_2 transformed parental and hybrid expression ratios (*X. laevs*/*X. borealis*) and probabilities of the null hypothesis of tests detailed in Results and in Table 2.2. Interpretations and abbreviations follow Table 2.2. Conclusions are identical to the analysis based on combined data from both sexes with the exception that textitcis-upregulation of the XB allele in tadpoles is no longer significant.

DMRT1α	Parental ratio	*P*-value (P = 0)	Hybrid ratio	*P*-value (P = H)	*P*-value (H = 0)	Interpretation
Tadpole	0.9141	0.0158*	–0.1364	0.0088*	0.0790	T: XL up
Adult	1.1491	0.0846	3.0141	0.0250*	0.0001*	C: XL up, T: XB up

(5) In whom did *DM-W* arise?

The 36 chromosomes of allotetraploid *Xenopus* species such as *X. laevis* and *X. borealis* are divided into two subgenomes (α and β) that probably correspond with separate diploid ancestors, each with 18 chromosomes, whose genomes were fused by allopolyploidization. Descendants of these ancestors are extinct or undiscovered, so genome duplication by allopolyploidization as opposed to autopolyploidization is an untested hypothesis (Evans, 2007, 2008). As reflected by the topology of relationships among *DMRT1* paralogs (Figure 2.4), phylogenetic relationships among many paralogs of *X. laevis* and *X. borealis* indicate that each tetraploid species contains pairs of paralogs, or "co-orthologs", that diverged from one another by speciation (Figure 2.4; Chain and Evans (2006); Chain et al. (2008)). However, these relationships could stem from multiple evolutionary scenarios (Supplementary Figure 2.S2), including one autopolyploidization event involving a diploid with 18 chromosomes, one allopolyploidization event between two diploids with 18 chromosomes, two allopolyploidization events between two diploids with 18 chromosomes but with different maternal and paternal contributions, or allopolyploidization between different maternal and paternal combinations of three or four diploids with 18 chromosomes. If we assume that absence of *DM-W* in *X. borealis* is the ancestral condition (rather than an instance of gene loss), these scenarios illustrate how *DM-W* could have originated in a tetraploid species (Figure 2.S2A) or in a diploid species (Figure 2.S2B-D).

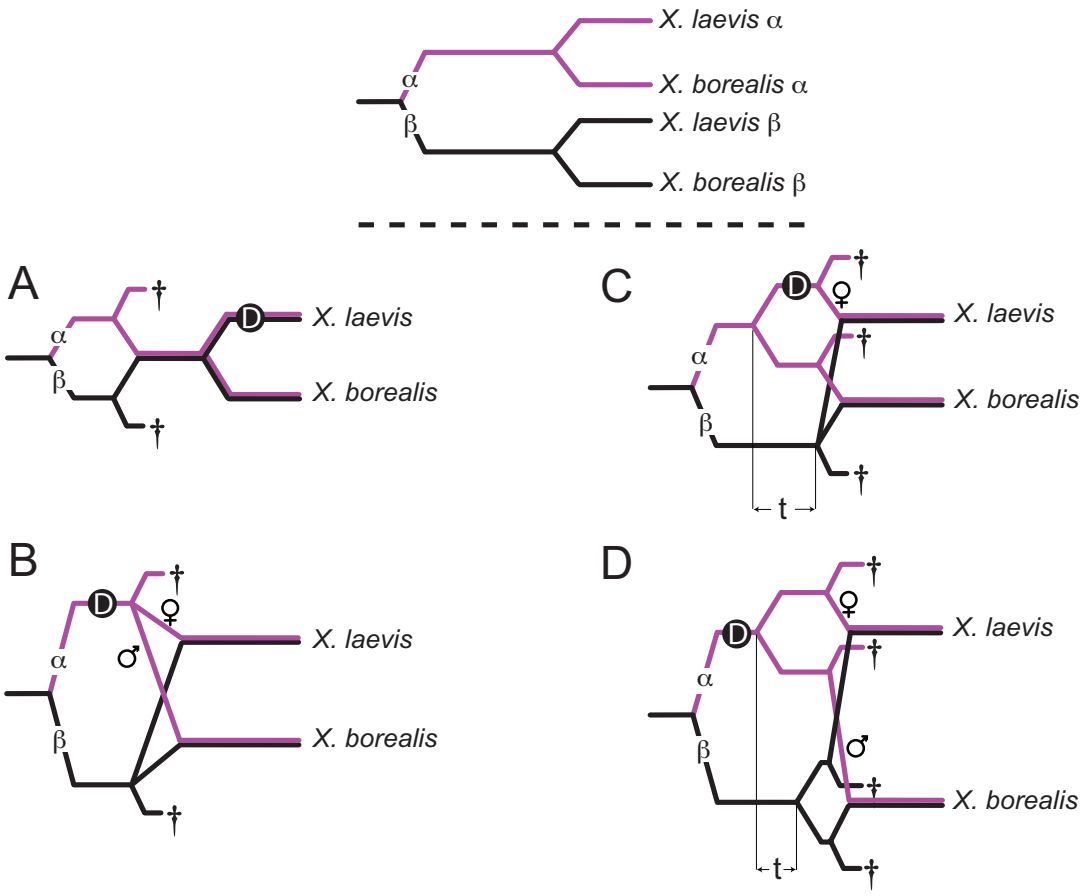

Figure 2.S2: Alternative scenarios for evolution of *DM-W*, indicated by a "D" on each phylogeny, that are compatible with estimated relationships among paralogs. Extinct or undiscovered diploid species are indicated with daggers. Many paralogous genes in *X. laevis* and *X. borealis* have the topology depicted above the dotted line (Chain et al., 2008). Under the assumption that *X. borealis* does not have *DM-W*, multiple scenarios involving a single origin and no subsequent loss could explain the origin of this gene: (A) *DM-W* could have originated in the allotetraploid ancestor of *X. laevis* and other members of Clade 1 (Figure 2.3). Alternatively, (B) *DM-W* could have originated in a diploid prior to two instances of allotetraploidization that involved the same species but with different maternal and paternal combinations, and that gave rise to species in Clade 1 and 2 respectively. Another possibility is that (C) two instances of allotetraploidization occurred, but involving three species, or that (D) two instances of allotetraploidization occurred, but involving four species. The longer time (t) is, the more statistical power there is to distinguish scenarios (C) and (D) from scenarios (A) and (B).

Chapter 3

A large pseudoautosomal region on the sex chromosomes of the frog *Silurana tropicalis*

Bewick AJ, Chain FJJ, Zimmerman LB, Sesay A, Gilchrist MJ, Owens ND, Seifertova E, Krylov V, Macha J, Tlapakova T, Kubickova S, Cernohorska H, Zarsky V, and Evans BJ (2013) *Genome Biology and Evolution*, *in press*

3.1 Preface

Sex chromosome evolution theory predicts that suppression of recombination around a sex determining gene should spread to much of the chromosome to due linkage and epistatic interactions with sexually antagonistic alleles, and the reduction of efficacy of selection. Why is it then that sex chromosomes in some species do not degenerate even after substantial amounts of evolutionary time? We set out to better understand mechanisms that are potentially related to sex chromosome homomorphy by characterizing the sex chromosomes of the female heterogametic frog *Silurana tropicalis* in terms of the size, divergence, and molecular evolution and expression of the sex-linked genes.

3.2 Abstract

Sex chromosome divergence has been documented across phylogenetically diverse species, with amphibians typically having cytologically non-diverged ("homomorphic") sex chromosomes. With an aim of further characterizing sex chromosome divergence of an amphib-

ian, we used "RAD-tags" and Sanger sequencing to examine sex-specificity and heterozygosity in the Western clawed frog *Silurana tropicalis* (also known as *Xenopus tropicalis*). Our findings based on ∼20 million genotype calls and ∼200 PCR amplified regions across multiple male and female genomes failed to identify a substantially sized genomic region with genotypic hallmarks of sex chromosome divergence, including in regions known to be tightly linked to the sex determining region. We also found that expression and molecular evolution of genes linked to the sex determining region did not differ substantially from genes in other parts of the genome. This suggests that the pseudoautosomal region, where recombination occurs, comprises a large portion of the sex chromosomes of *S. tropicalis*. These results may in part explain why African clawed frogs have such a high incidence of polyploidization, shed light on why amphibians have a high rate of sex chromosome turnover, and raise questions about why homomorphic sex chromosomes are so prevalent in amphibians.

3.3 Introduction

Sex can be advantageous because it decouples beneficial from deleterious mutations via recombination, which increases the variance in fitness effects of linked mutations, and thus the efficiency with which natural selection operates. In species with genetic sex determination, developmental differences between the sexes are initiated by genetic differences between the sex chromosomes. In some lineages, the genes responsible for triggering sex determination vary, and the sex chromosomes (which carry the sex determining region) are routinely reassigned from one to another ancestral pair of autosomal chromosomes (Evans et al., 2012; Fridolfsson et al., 1998; Pease and Hahn, 2012; Ross et al., 2009). Ironically, suppression of recombination within a sex-specific region is often favoured by natural selection, lest a sex-specific, sex determining allele loses its sex-specificity.

The origin of sex chromosomes could be initiated by sexual antagonism (van Doorn and Kirkpatrick, 2007) and in many species this is associated with cessation of recombination between a portion of the sex chromosomes that makes possible unisexual inheritance of a key genomic region that triggers sex determination. Cessation of recombination between the sex chromosomes can be achieved by reducing or eliminating homology (Charlesworth, 1991), for example through point mutations, inversion, deletion, or insertion of DNA. Strikingly, the extent of the non-recombining region may increase overtime, although this is not necessarily the case (Bergero and Charlesworth, 2009; Charlesworth et al., 2005). The expansion of non-recombining regions may be influenced by the nature of evolution in non-recombining genomic regions, which is influenced by Mullers ratchet, background selection, Hill-Robertson effects, and genetic hitchhiking of deleterious alleles with beneficial mutations (Charlesworth and Charlesworth, 2000). Suppressed recombination between sex chromosomes thus has important implications for genome evolution, speciation, and adaptation.

Sex chromosome "degeneration" can be associated with sex chromosome divergence resulting from suppressed recombination, and involves the loss of coding regions, the accumulation of repetitive regions, and structural changes such as insertions, deletions, and inversions on the sex-specific chromosome (the Y or W). Thus, degenerate sex chromosomes have differences that extend beyond the fundamental difference in the presence or absence of a sex determining allele. However, evolutionarily young sex chromosomes start out being similar to each other because they originated from an essentially identical pair of autosomal chromosomes. In the medaka fish, for example, the male-specific region of the Y-chromosome contains a newly evolved sex determining locus (*dmrt1bY*) and is only ∼258,000 base pairs long (Kondo et al., 2006). Likewise, the sex chromosomes of the tiger pufferfish appear to be distinguished by only one nonsynonymous substitution (Kamiya et al., 2012). In theory, natural selection may drive the expansion of the sex-specific region of suppressed recombination (Charlesworth et al., 2005), and old sex chromosomes may evolve distinct suites of genes with unique, non-homologous, and/or extensively diverged functions. In humans, for example, almost all of the ancestral genes persist on the X-chromosome, but have been lost on the Y-chromosome (Skaletsky et al., 2003). The limit of divergence is achieved if the sex-specific sex chromosome is lost altogether, as occurred in the Ryukyu spiny rat (Kuroiwa et al., 2010). In contrast, however, the sex chromosomes of ratite birds and boid snakes are old but each pair is morphologically non-diverged (homomorphic), suggesting that the size of the region of suppressed recombination on the sex-specific sex chromosome does not necessarily expand over time (Matsubara et al., 2006; Tsuda et al., 2007).

3.3.1 Frog sex chromosomes

Most species of amphibians have homomorphic sex chromosomes (reviewed in Schmid et al., 2010). One possible explanation for the high incidence of homomorphic sex chromosomes in amphibians is that their sex chromosomes tend to be young because there has been frequent switching of the sex determining locus during evolution. Consistent with this explanation is the inference that sex chromosomes have changed ("turned over") ∼32 or more times during evolution based on variation in male (XY) versus female (ZW) heterogamy (Evans et al., 2012; Schmid et al., 2010). Another explanation for widespread homomorphic sex chromosomes is that there is periodic recombination between the sex chromosomes over most of their length without changes in the sex determining locus. In one group of hylid frogs for instance, genomic regions that are tightly linked to the sex determining locus are not substantially diverged between males and females, indicating that the sex chromosomes of these frogs recombine, at least occasionally, over most of their length (Stöck et al., 2011). This suggests that sex chromosomes of these frogs have large "pseudoautosomal" regions where inheritance of genetic information resembles autosomal genes, and where recombination prevents divergence between the sex chromosomes. Thus, frequent turnover and recombination clearly both play a role in homomorphy of amphib-

ian sex chromosomes, but which phenomenon plays the dominant role remains an open question.

The only known amphibian sex determining gene is called *DM-W* and was discovered in the African clawed frog *Xenopus laevis* (Yoshimoto et al., 2008). *DM-W* is female-specific, and originated after divergence from the sister genus *Silurana* but before diversification of most or all extant species of *Xenopus* (Bewick et al., 2011). Because *Silurana tropicalis* (also known as *Xenopus tropicalis*) lacks *DM-W*, sex determination in this species must be triggered by another as of yet unidentified genetic trigger. A high quality draft genome sequence is available for *S. tropicalis* that was generated from a female (Hellsten et al., 2010), but the sex specific region of this genome has not been characterized. Using amplified fragment length polymorphisms (AFLPs), Olmstead et al. (2010) identified 22 AFLPs linked to the sex determining locus in the "golden" strain of *S. tropicalis* (Table 3.1), and proposed that females are the heterogametic sex in this strain. Four of these 22 AFLPs placed to the distal tip of chromosome/linkage group 7 in a linkage map developed by Wells et al. (2011), also represented by scaffold 7 in version 7.1 of the *S. tropicalis* genome sequence. However, the linkage map contains a large (15 cM) gap between the distal most two markers where the sex-determining region is likely to reside (Wells et al., 2011). Many of the other sex-linked AFLPs identified by Olmstead et al. (2010) map to other major or small chromosome/linkage groups (Table 3.1); this is presumably because some scaffolds are chimerical (for example a portion of scaffold 2 in version 7.1 is probably actually derived from *S. tropicalis* chromosome 7) and because linkage relationships between some small contigs ("orphan scaffolds") and the larger scaffolds has not yet been established. Furthermore, additional experiments with other strains suggest that sex determination may occur through the action of multiple alleles at one locus, or multiple tightly linked genes (A. Olmstead, personal communication).

The goal of this study is to further characterize the sex chromosomes of *S. tropicalis* in terms of the size and level of divergence of the sex specific region, and to compare molecular evolution and expression of sex-linked and non-sex-linked genes. To this end, we used "RAD tags" (Baird et al., 2008), a reduced representation next generation sequencing approach, to genotype millions of homologous nucleotide positions in male and female individuals including positions that are monomorphic in both sexes, polymorphic in one or both sexes, and positions in which a genotype inference (i.e., homozygous or heterozygous) was only possible in one sex due either to sex-specificity of the genotyped position or differences in coverage of that position between the sexes. The RAD tag approach produces sequences of thousands of small regions that are adjacent to a rare cutting restriction enzyme site. Because the sequenced portions of the genome are associated with a restriction enzyme site, many homologous sequences are obtained from multiple individuals. Missing data among individuals can arise in unusual cases where mutation generates polymorphism in the presence or absence of the restriction enzyme sites or because of variation among individuals in the depth of sequencing coverage for a particular region. Our analysis incorporated information on sex-linked regions from Olmstead et al. (2010), information from a

laser-dissected chromosome arm 7p from a male individual (Seifertova et al., 2013) which is linked to the sex linked region identified by Olmstead et al. (2010), and the most recent genome assembly (version 7.1, reference accession PRJNA12348). This study thus provides, for the first time, a comprehensive perspective on the extent of sex chromosome divergence in this species by evaluating the distribution of homozygous and heterozygous genotypes, molecular evolution, and gene expression of sex chromosomes in the context of the rest of the genome.

3.4 Methods

Four female and four male *S. tropicalis* individuals were obtained from Xenopus Express (Brooksvile, FL, USA). Sex was confirmed by dissection, and species assignment achieved by comparing between 809 and 812 bp of mitochondrial DNA sequence from a portion of the 16S gene from each sample to homologous sequence data from all other known species of African clawed frog (Evans et al., 2011a). We performed a phylogenetic analysis on these eight sequences, 27 sequences from individuals used in the PCR screen detailed below, all *Silurana* sequences from Evans et al. (2004), six *S. tropicalis* samples from Ghana (obtained from tissue archive at the Burke Museum, University of Washington, accession numbers UWBM5957-8, UWBM5961-63, and UWBM5969), and sequences from six individuals from the "golden" strain used by Olmstead et al. (2010) that were provided by Richard Harland. We used an *X. laevis* sequence from South Africa as an outgroup in this analysis and the total alignment length was 817 bp. Model selection for phylogenetic analysis was accomplished using MrModeltest2 (Nylander, 2004). Phylogenetic analysis was performed with MrBayes version 3.1.2 (Huelsenbeck and Ronquist, 2001) using the best-fit model based on the Akaike Information Criterion, with two independent Markov Chain Monte Carlo (MCMC) runs, each for 2,000,000 generations. Convergence of the MCMC runs on the posterior distribution was assessed by inspecting parameter trends and effective sample sizes using Tracer version 1.5 (Rambaut and Drummond, 2007). Based on these analyses, a burn-in of 500,000 generations was discarded before constructing a consensus tree with MrBayes.

Genomic DNA was extracted from liver using QIAGEN DNeasy kit, purified using QIAGENs spin purification protocol, and RAD tag library preparation performed by Floragenex, Inc (Eugene, Oregon). For each individual, two libraries were generated – one used the restriction enzyme *SbfI* and another used *NotI*. The RAD tag libraries were multiplexed on three Illumina flow cells using individual barcodes, and Illumina sequencing was performed at the University of Oregon. These data have been deposited in Genbank (accession number XXX).

Illumina sequence reads were sorted by barcode with RADtools v1.2.4 using the "fuzzy_MID" option, which assigns reads with barcode errors to the nearest barcode (Baxter et al.,

2011). Data from each individual were independently aligned to the *S. tropicalis* v7.1 genome using bwa-0.6.2 (Li and Durbin, 2009) and samtools.0.1.18 (Li et al., 2009). The "MarkDuplicates" function in picard (http://picard.sourceforge.net) was used to mark putative PCR amplified duplicates, which were then excluded from the genotyping analysis with an aim of minimizing genotyping error. The Genome Analysis toolkit (GATK) version 2.2-15 was then used to realign indels using the"RealignerTargetCreator" and "IndelRealigner" functions (DePristo et al., 2011; McKenna et al., 2010). The "FixMateInformation" function of picard was then used to adjust mate pair alignments.

Following "Best Practices" guidelines on the GATK website and forum (http://gatkforums.broadinstitute.org/) for analysis of genomes that lack known SNPs, the "UnifiedGenotyper", "BaseRecalibrator", and "PrintReads" functions of GATK were used to iteratively genotype, recalibrate base quality scores, and generate new input (bam) files, using the genotype files generated from "UnifiedGenotyper" as known polymorphic positions to be ignored for base recalibration in each iteration. Convergence was reached by the 5^{th} iteration, in that variable positions recovered from this analysis were 99.8% identical to those from the 4^{th} iteration. The "VariantFiltration" and "SelectVariants" functions of GATK were then used to identify and exclude genotyped positions that (1) were within 10 bp of an insertion/deletion, (2) had a Phred genotype quality score (Ewing and Green, 1998) of less than 30, which means that we removed positions that had a probability of error of greater than 0.001, or (3) had more than one tenth of the reads mapping equally well to another position, and where there were at least 4 of these reads.

The *S. tropicalis* genome assembly 7.1 consists of 7,730 scaffolds aggregated from 55,234 contigs connected by "N"s within each scaffold. The total number of bases is 1,437,594,934, of which 5% (n = 71,599,926) are "N"s. This assembly includes fourteen large "super scaffolds" that were assembled using meiotic map, synteny, and cytological data, corresponding to the 10 haploid chromosomes, with some chromosomes being represented by multiple scaffolds (3a and 3b; 5a and 5b; 8a, 8b, and 8c). The rest of the scaffolds are "orphan scaffolds" whose chromosomal locations are not yet known. We divided the genomic regions into five mutually exclusive groups based on (a) the inferred level of recombination with the sex determining region by Olmstead et al. (2010), (b) the linkage groups in the genome assembly 7.1 (Table 3.1), and (c) the results of the Illumina sequencing of the dissected petite arm of chromosome 7 (Seifertova et al., 2013). The first of the five groups ("completely sex-linked") included contigs from assembly 7.1 that contain regions that had no recombination (0%) with the sex determining region in Olmstead et al. (2010). This means that recombination between an AFLP polymorphism and the sex determining region was not observed in any of 300 individuals assayed by Olmstead et al. (2010). The second group ("partially sex-linked") included contigs from assembly 7.1 that contain regions that had a recombination rate $>0\%$ and $<3.0\%$ in Olmstead et al. (2010). The third group ("chromosome 7p") contained sections of scaffolds in assembly 7.1 that are located on chromosome 7p according to (Seifertova et al., 2013), and not in the "completely sex-linked" or "partially linked" categories. The fourth group ("non-7p

chromosomes") contained the remaining sections on the chromosome-scale scaffolds in assembly 7.1, including the portion of scaffold 7 that did not map to chromosome 7p. The fifth group ("other orphans") contained orphan scaffolds in assembly 7.1 that (a) have not been linked to a chromosome, (b) have no evidence of sex linkage according to Olmstead et al. (2010), and (c) did not map to chromosome arm 7p according to Seifertova et al. (2013). More specific information on the scaffold or scaffold portions in each of these groups is provided in Table 3.1.

Table 3.1: Genomic regions of *Silurana tropicalis* that are putatively sex linked based on linkage study of Olmstead et al. (2010) and sequencing of chromosome arm 7p by Seifertova et al. (2013). AFLP refers to the name of the amplified fragment length polymorphism from Olmstead et al. (2010) if provided. Recombination refers to the recombination rate with the sex determining locus from that study. Scaffold and position of AFLPs are provided for genome assembly version 4.0 (v4) and 7.1 (v7.1). Sex linked portions that were included in categories in Table 2 and 3 based on the level of recombination (Portion sex-linked, with NA meaning not applicable) either refer to base pair positions of a contig within a larger scaffold that is not interrupted by unknown sequence, or the entire scaffold was assumed to be sex linked (ALL).

AFLP	Recombination	v4	v7.1	7p?	Portion sex-linked
E33.M72.143	0	605:241571-241691	7:4966286-4966166	yes	4435335-5175370
E33.M81.275	0	494:27646-27898	no hits	–	NA
E33.M90.327	0	no hits	211:76840-76535	yes	ALL
E38.M93.218	0	953:138210-138402	278:109297-109490	no	ALL
No name	0	494:31633-32115	no hits	–	NA
No name	0	494:27541-27902	no hits	–	NA
No name	0	379:817889-818000	2:149826489-149826600	yes	149787496-150105127
No name	0	736:292586-293067	78:248343-247864	yes	ALL
No name	0.3	605:245039-245621	7:4963243-4962661	yes	4435335-5175370
No name	0	605:116800-117215	22:1040592-1040177	yes	ALL
E40.M52.572	0.4	859:57522-58049	7:3195527-3196069	yes	76939-3370464
E33.M61.177	0.5	1778:6156-6312; 1778:9971-9815	144:136541-136385	yes	ALL
E33.M61.177	0.5	1778:6156-6312; 1778:9971-9815	662:22621-22465	no	NONE because 114 is on chr7
No name	0.5	810:261995-262744	94:264236-264985	yes	ALL
No name	0.5	810:261995-262744	94:263563-263460	yes	ALL
E32.M94.406	1.9	multiple hits	22:58856-59241	yes	ALL
E37.M52.423	1.9	810:325559-325959	144:95531-95931	yes	ALL
E37.M52.423	1.9	1151:130316-130719	144:95931-95531	yes	ALL
E41.M83.506	1.9	810:276803-277288	94:279867-279382	yes	ALL
E32.M35.552	2.6	multiple hits	no hits	–	NA
E37.M60.232	2.6	6092: 2392-2601	7931:611-828	yes	ALL
E32.M59.335	2.9	735:292141-292443	7:7903155-7902853	yes	6687308-9940823

3.4.1 Genome-wide distribution of genotypes in female and male *S. tropicalis*

For genomic regions in each of the six categories described above, we tabulated genotype patterns for three scenarios (Figure 3.1) in 500,000 base pair windows across the *S. tropicalis* genome; smaller windows were examined at the ends of scaffolds or when a scaffold was smaller than 500,000 base pairs. Genotype patterns in each sex (i.e., the distribution of homozygous or heterozygous positions) are relevant to sex chromosome evolution in the following ways. First, divergence between the sex chromosomes due to suppressed recombination generates positions that are either heterozygous in all females (for a ZW sex determining system) or heterozygous in all males (for an XY sex determining system). We call this pattern "Scenario 1A" and "Scenario 1B" respectively (Figure 3.1). We note that in "Scenario 1A" regions, some positions can also be heterozygous in males due to polymorphism on the Z-chromosome, and in "Scenario 1B" some positions can also be heterozygous in females due to polymorphism on the X-chromosome. In any case, in genomic regions consistent with Scenario 1, heterozygosity observed in all samples from one sex is expected to exceed heterozygosity observed in all samples from the other sex. We therefore searched for regions with heterozygosity present in all females or in all males. For both of these statistics, we ignored positions that are heterozygous in all genotyped individuals. To account for variation in coverage in males and females, for each window we divided these counts by the total number of positions in each window for which genotype calls were made in at least one female and at least one male.

Another genotypic scenario for sex chromosomes is that a genomic region may be present only on the Z-chromosome (with female heterogamy) or only on the X-chromosome (with male heterogamy) (Scenario 2A and 2B; Figure 3.1). No counterpart exists on the W-chromosome (or Y-chromosome) due to deletion, insertion, or divergence. To detect such a genomic region, we searched for regions with heterozygous positions present in one sex but not the other. For such positions, we required a genotype call in at least one individual of each sex but heterozygous calls to be present in only one sex. To account for variation in coverage in males and females, for each window we divided these counts by the total number of positions in each window for which genotype calls were made in at least one female and at least one male.

A third genotypic scenario for sex chromosomes is that a genomic region may be present only on the W-chromosome, or only on the Y-chromosome (Scenario 3A and 3B, Figure 3.1). Thus we searched for positions that had genotype calls only in females (or only in males), and that are all homozygous. To account for variation in coverage, we standardize the counts in each window by the sum of the number of positions in each window for which genotype data is available for (i) at least one female and at least one male, (ii) at least one female but no males, and (iii) at least one male but no females. Thus by evaluating these three genotype scenarios in genomic windows across the *S. tropicalis* genome assembly, we attempted to identify genomic windows that either had significantly more heterozygous po-

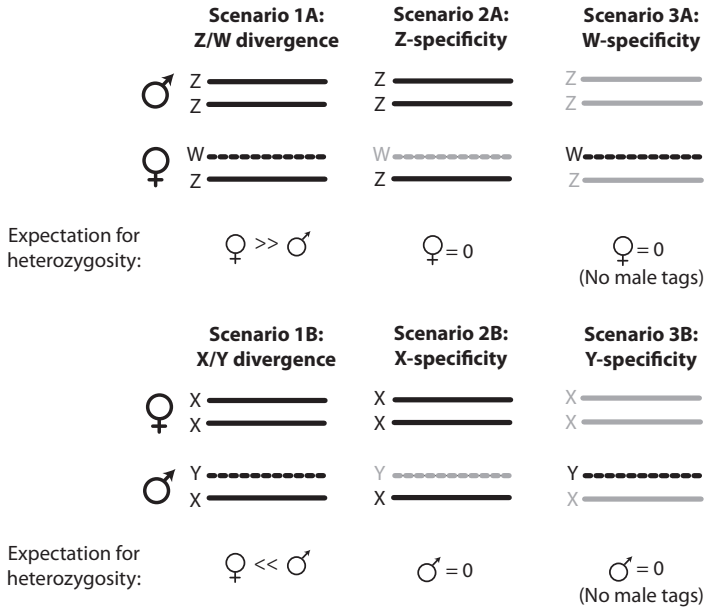

Figure 3.1: Genotypic scenarios for sex linked regions. Expectations for heterozygosity depend on which sex is heterogametic, and which portion of the sex chromosome the genotypes are. Scenarios 1-3 shown below assume female heterogamy and include female biased heterozygosity (Scenario 1A), male only heterozygosity (Scenario 2A), or female only homozygosity with no male genotypes (Scenario 3A). Corresponding scenarios (Scenarios 1B, 2B, and 3B) apply to the opposite sex for male heterogamy.

sitions in one sex (Scenario 1), had heterozygous positions only in one sex (Scenario 2), or that had homozygous positions in only one sex and no homologous genotypes in the other (Scenario 3). Higher values for each ratio are suggestive of genotype patterns characteristic of degenerate sex chromosomes.

3.4.2 Expression and molecular evolution

As described in Chain et al. (2011), we estimated gene expression levels based on sequences across 26 expressed sequence tag (EST) libraries from the following tissues or developmental stages: egg, gastrula, neurula, embryo, tailbud, tadpole, metamorphosis, adipose tissue, bone, brain, head, heart, intestine, kidney, limb, liver, lung, ovary, oviduct, skeletal muscle, skin, spleen, stomach, tail, testis, and thymus. We summarized patterns of gene expression across EST libraries using the non-independent "total", "intensity", and "evenness" statistics described in Chain et al. (2011). The "total" expression of a gene (T) is the proportion of times that a gene was sequenced in each EST library (L_i) summed across all libraries (T = ΣL_i). The "intensity" of expression (I) is the mean expression level from the perspective of a gene, and is calculated following this equation: I = $\Sigma L_i^2/\Sigma L_i$. "Evenness" of expression

(E) can be thought of as the "effective number" of tissues in which a gene is expressed, and is calculated following this equation: E = T/I.

For a subset of the sex-linked and non-sex-linked genes, we also calculated the rate ratio of nonsynonymous to synonymous substitutions per site (*dN/dS*) along the *S. tropicalis* lineage using PAML version 4.5 (Yang, 1997). This ratio was calculated using a maximum likelihood model that individually estimates *dN/dS* for each branch in a phylogeny, following Chain et al. (2011). Our phylogeny was estimated from sequences from *S. tropicalis*, *X. laevis*, and using sequences from another pipid frog (*Pipa carvalhoi* or *Hymenochirus curtipes*) as an outgroup. To avoid undefined values we added 0.02 to all *dS* values before calculating *dN/dS*, following Chain et al. (2011). We made this adjustment a priori by looking only at *dS* values, in order to make better use of the data. Because extreme values for *dN* and *dS* were occasionally estimated, we excluded from the analysis genes with an estimated *dN* or *dS* value above 2, and any genes whose available data comprised less than 100 synonymous positions.

We used one-sided permutations to test whether the expression and molecular evolutionary statistics differed between genes that either (a) were or (b) were not on the same chromosome as the sex determining locus. The permutations randomly divided the set of (a + b) values into two groups of size a and b, and then calculated the difference between the averages of each group. We repeated this 1000 times to generate a distribution for the null hypothesis that the values were drawn from the same underlying distribution, and then compared this to the observed differences, which is the test statistic of each test. A significant difference was inferred if the observed difference was greater than 95% of the differences from the permutations. Because these tests are one sided, the operands of the test statistic (that is, the minuend and subtrahend of each difference) were defined according to specific expectations for sex chromosome degeneration discussed below.

3.5 Results

3.5.1 Mitochondrial DNA variation within *S. tropicalis*, including the "golden" strain

We analyzed phylogenetic relationships among ∼810 bp region of mitochondrial DNA from the commercially obtained *S. tropicalis* individuals we used for RAD tags and PCR screens, six individuals from the golden strain used by Olmstead et al. (2010), and several other wild caught *S. tropicalis* individuals and individuals from other *Silurana* species. An identical mitochondrial DNA sequenced was obtained from the six golden strain individuals, one of the samples we used for RAD tag sequencing (a female) and 20 of the samples we used for PCR screens (9 females, 11 males), and one individual sampled from Nigeria. Mitochondrial sequences from five samples used in the RAD tag sequencing (2 females, 3

males) were identical and differed from the golden strain sequence by one nucleotide substitution. Mitochondrial sequences from two other samples used in RAD tag sequencing (1 female, 1 male) and 7 samples used in PCR screens (4 females, 3 males) differed from the golden strain mitochondrial sequence by a different single nucleotide substitution than the previously mentioned sequence present in five of the RAD tag samples. Mitochondrial sequences from another sample from Nigeria differed from the golden strain mitochondrial sequence by two nucleotide substitutions. Phylogenetic analysis of these and other sequences indicates that the commercially obtained *S. tropicalis* samples used in this study form a well-supported clade that includes two sequences from Nigeria and the six sequences from the golden strain of *S. tropicalis* (Figure 3.2). This clade is possibly common in individuals east of the Dahomey Gap, a savannah corridor that interrupts the West African rain forest (Salzmann and Hoelzmann, 2005).

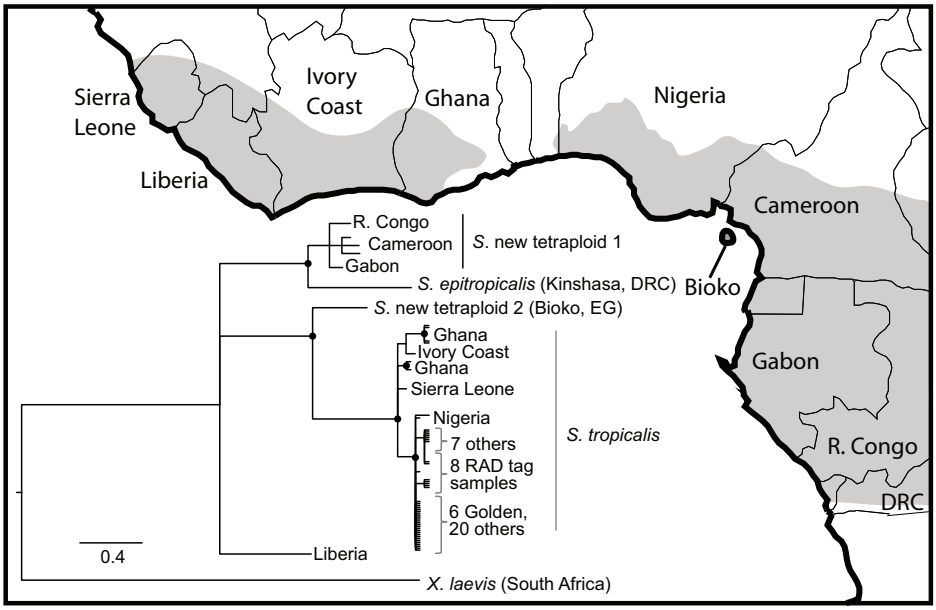

Figure 3.2: Phylogenetic analysis of mitochondrial DNA sequences suggest that the "golden" strain used by Olmstead et al. (2010) and samples used in this study (8 for RAD tag analysis and 27 others for PCR assays) originate from Nigeria. Nodes with $\geq$95% posterior probability are indicated with a black circle. Species names, including those undescribed, follow Evans et al. (2004). Abbreviated country names include the Republic of the Congo (R. Congo), the Democratic Republic of the Congo (DRC), and Equatorial Guinea (EG). The scale bar refers to the number of substitutions per site, and grey areas on the map indicate the distribution of tropical forest in West Africa.

3.5.2 Reduced representation genome-wide genotyping from RAD tags

We used a reduced representation genome sequencing approach called "RAD tags" to sequence many small but homologous portions of the *S. tropicalis* genome in four female and four male individuals. An average of 9,696,525 Illumina reads were mapped in each individual, with the average number of reads mapped per female or per male being 9,445,507 and 9,947,543 reads respectively. After excluding positions in the reference sequence with no data, within an individual the average depth of coverage was 18.4 reads per position. Genotypes were called for a total of 19,624,843 positions, and 193,199 single nucleotide polymorphisms (0.98%) were detected. For each position at which at least one genotype was called, an average of 7.14 individuals were genotyped.

If *S. tropicalis* has a large female-specific genomic region on the W-chromosome, we expected a higher proportion of the Illumina reads from females to map to the genome assembly because this assembly was generated from a female individual. Contrary to this expectation, a slightly higher proportion of reads from males (average per male individual 89.2%, range 87.8%-90.8) than from females (average per female individual 87.9%; range 85.0%-90.8%) mapped to this genome assembly, arguing against there being a large female-specific region in the *S. tropicalis* genome. Another indication of a large female specific genomic region on the W-chromosome would be a substantially higher number of positions genotyped in females than in males. Out of a total of 19,624,824 positions that were genotyped with high confidence in at least one individual, slightly more genotypes were recovered in females than in males: 912,738 (4.7%) positions were genotyped only in one or more females, and 462,792 (2.4%) positions genotyped only in one or more males. However, in the ten largest scaffolds, the number of genotype calls in at least one female was consistently 1.1-3.3% higher than the number of genotype calls in at least one male, with scaffold 7 having 2.5% more genotype calls in females than males. This suggests that the higher number of unique genotype calls in females is primarily a technical artifact related to differences in coverage among individuals in the RAD tag libraries. The RAD tag data did not provide high quality genotypes from any positions on 5721 scaffolds, which together comprise 36,133,437 bp ($\sim$2.1% of the genome).

3.5.3 Genome-wide genotype patterns and nucleotide diversity similar in males and females

We searched 500,000 bp windows for various genotypic patterns consistent with sex chromosome divergence expected under female and male heterogamy (Figure 3.1). In general this effort failed to identify any regions with a pronounced genotypic signature of sex chromosome divergence expected by female heterogamy (Table 3.2). One exception was a significant excess of windows with female-only homozygous genotypes (Scenario 3A) in orphan scaffolds, but we suspect this was an artifact related to the broader coverage in fe-

males. Most notably, portions of linkage groups 2 and 7 that were identified as “partially sex-linked” and “completely sex-linked” to the sex determining region in Olmstead et al. (2010) did not exhibit a genotypic pattern consistent with degenerate sex chromosomes based on the RAD tag genotypes.

Table 3.2: Putatively sex-linked regions do not exhibit genotypic hallmarks of diverged sex chromosomes. Analysis of 500,000 bp genomic windows indicates that genotype patterns of genomic regions linked to sex do not resemble scenarios expected for sex chromosomes substantially more than other parts of the genome. Numbers indicate average values for three genotypic scenarios depicted in Figure 3.1 and described in Methods. For the "Non-7p chromosomes", 95% confidence intervals are in parentheses. Asterisks indicate values that are higher than the 95% confidence intervals from the "Non-7p chromosomes".

Region	Number of genotype calls	Scenario 1A: All females heterozygous	Scenario 1B: All males heterozygous	Scenario 2A: All females homozygous	Scenario 2B: All males homozygous	Scenario 3A: All females homozygous; males no genotype	Scenario 3B: All males homozygous; females no genotype
"Non-7p chromosomes"	17026460	0.00013 (0.00000-0.00087)	0.00018 (0.00000-0.00123)	0.00124 (0.00000-0.00399)	0.00179 (0.00000-0.00504)	0.04980 (0.00390-0.15189)	0.02670 (0.00050-0.09960)
"Chromosome 7p"	830811	0.00019	0.00019	0.00072	0.00456	0.10066	0.03760
"Other orphans"	1618864	0.00034	0.00040	0.00305	0.00524*	0.33587*	0.18522*
"Partially sex-linked"	132177	0.00024	0.00014	0.00182	0.00263	0.06098	0.03920
"Completely sex linked"	16512	0.00026	0.00003	0.00228	0.00308	0.02848	0.01176

Considerable caution is needed in the interpretation of the average genotype frequencies in genomic windows for the "other orphans" category because in many cases the scaffold is smaller than the window size (500,000 bp), and the resulting truncated genomic windows are therefore expected to have an increased variance in the frequency of various genotypic patterns. Additionally, because unusual scaffolds that have a genotypic signature of sex chromosome divergence might not substantially affect the averages across all orphan scaffolds, average genotype frequencies in these genomic windows could fail to detect small scaffolds that have genotypic patterns consistent with sex chromosome divergence. In any case, "other orphans" had higher than expected values for Scenarios 2B and 3B, which are consistent with male heterogamy, but this is probably related to the small size of these scaffolds and consequent increase in the sex-specific genotypes in truncated windows for essentially all of the scaffolds.

Additional insights are gained by examining nucleotide diversity in each sex within 500,000 bp windows. If a portion of the sex chromosomes is substantially diverged, we expected much higher average nucleotide diversity per site in one sex (females for female heterogamy) in genomic windows spanning this diverged region. However, average nucleotide diversity per site is essentially identical in males and females throughout these scaffolds, including chromosome arm 7p, which is linked to the sex determining region (Figure 3.3). To explore the possibility that there could be variation within the RAD tag samples in sex chromosome divergence that corresponds with the three mitochondrial DNA haplotype groups detailed above, we explored nucleotide diversity in male and female individuals from each group. This analysis also did not identify a pronounced signature of sex chromosome divergence (Supplementary Figure 3.S1 and S2).

3.5.4 Genotype patterns based on Sanger sequencing

We amplified 65 genomic regions identified by Olmstead et al. (2010) to be linked to the sex determining region, including 18 and 46 amplicons from "completely sex-linked" and "partially sex-linked" regions respectively (Table 3.1, Supplementary Table 3.S1, Supplemental Figure 3.S3). None had female-specific amplifications, allowing us to dismiss Scenario 3 for all of these regions (Figure 3.1). We sequenced 45 of these amplifications in multiple male and female individuals. Single nucleotide polymorphisms or insertion/deletion polymorphisms were shared between males and females in at least one amplicon for essentially all scaffolds (no polymorphism was observed in amplicons from scaffold 144 and some amplicons from scaffold 662 were not sequenced). This suggests that Scenario 1 is unlikely for these regions, with the caveat being that a heterozygous position could arise in both sexes in a region consistent with Scenario 1 through convergent evolution on the W and Z.

We also used PCR to examine an additional 173 regions that exhibited signs of sex linkage based on our analyses of the RAD tag data, including regions of chromosome 7p and elsewhere as detailed in Supplementary Figure 3.S3 and Supplemental Table 3.S1.

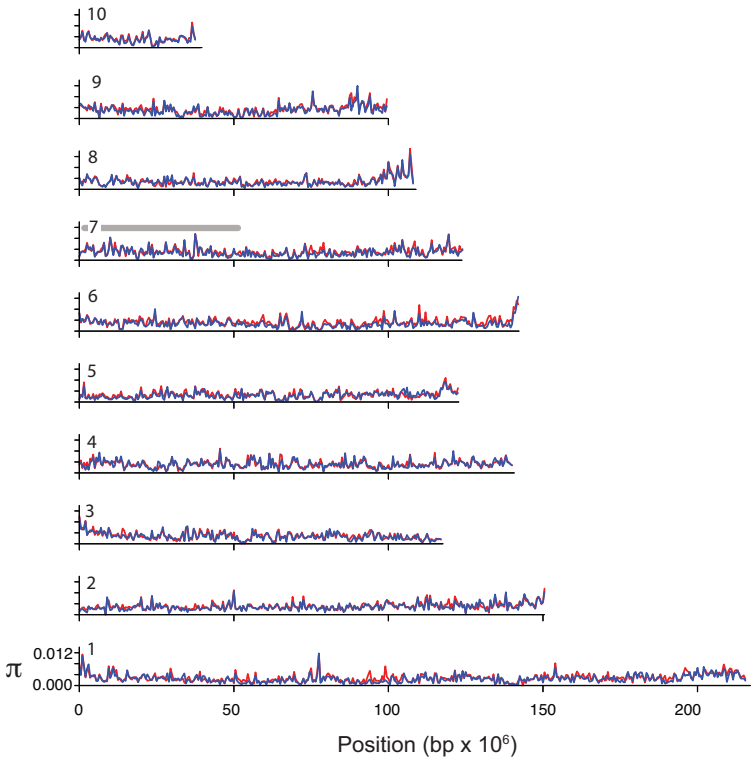

Figure 3.3: Nucleotide diversity (π) in 500,000 bp windows is similar in males (blue) and females (red) throughout much of the *S. tropicalis* genome. Plots are labeled with numbers which refer to scaffolds 1-10, which collectively comprise ~75% of the genome. A grey bar on scaffold/chromosome 7 indicates the petite arm based on the linkage map of Wells et al. (2011), which carries the sex determining region in at least one *S. tropicalis* strain Olmstead et al. (2010).

None had sex-specific amplifications, allowing us to dismiss Scenario 3 for all of these regions. We sequenced 94 of these amplifications from male and female individuals. Thirty of these were not polymorphic in any of the individuals we sequenced. Fourty-eight had polymorphisms shared between males and females, allowing us to conclude that Scenario 1 is unlikely for these regions. Five had polymorphisms in both sexes with none being shared across sexes. Ten had polymorphisms only in females and two had polymorphisms only in males.

Two amplifications were of particular interest. An amplification on scaffold 7 that spanned positions 10,128,301-10,129,920 was highly polymorphic in females but not males, although no polymorphism was fixed in females (out of 7 females and 3 males sequenced; Supplementary Table 3.S1). This region failed to amplify in four females and three males. Another amplification, which targeted a region on scaffold 163 had 31 polymorphisms in three females but only one polymorphism in 3 males that was not heterozygous.

3.5.5 Gene expression and molecular evolution

Expression was detected in a total of 37,790 transcripts in at least one of the 26 EST libraries we surveyed (Table 3.3). Based on studies of recently diverged neo-sex chromosomes in fruit flies (*Drosophila*) (reviewed in Bachtrog, 2013), we expected expression of genes situated near the sex determining locus to be expressed (i) at a lower total level, (ii) higher intensity, (iii) lower evenness, and to have (iv) higher *dN/dS* compared to genes in other parts of the genome (i.e., the "non-7p chromosomes"). For the most part, these expectations were not met for genes that were demonstrably very close to the sex determining region, with the one exception that the intensity of "completely sex-linked" genes were individually significantly higher than the "non-7p chromosomes" ($P<0.05$, Table 3.3). Evenness of "chromosome 7p" and "other orphans" was also significantly lower than "non-7p chromosomes" as was total expression of "other orphans". *dN/dS* was significantly higher only in "chromosome 7p" compared to "non-7p chromosomes" but the magnitude of this difference was small.

A conspicuous gap in Unigene blast hits was identified on scaffold 22, which is tightly linked to the sex determining region (Olmstead et al., 2010), that included no top hit Unigene matches even though it was 1,156,260 bp long. We examined this scaffold using Xenbase (Bowes et al., 2009) and found that it contained a cluster of olfactory receptors, none of which had BLAST hits to any of the EST libraries. The presumed expression profile of these genes (mostly nasal epithelia) is potentially consistent with the observation that completely sex linked genes tend to be more intense than the genome-wide average.

Table 3.3: Average expression and molecular evolution statistics for *S. tropicalis* genes in five genomic categories. See Methods for description of statistics. Asterisks indicate values that are individually significantly different from the the "Non-7p chromosomes" ($P<0.05$, one-sided permutation tests).

Region	Number of genes (Expression)	Total	Intensity	Evenness	Number of genes (*dN/dS*)	*dN/dS*
"Non-7p chromosomes"	35,134	0.00065	0.00015	3.19655	9183	0.2702
"Chromosome 7p"	2,114	0.00079	0.00019	3.17054*	546	0.2850*
"Other orphans"	260	0.00040*	0.00012	2.59443*	55	0.2670
"Partially sex-linked"	246	0.00090	0.00018	3.37658	71	0.2751
"Completely sex-linked"	35	0.00101	0.00049*	2.56020	8	0.2615

3.6 Discussion

To explore sex chromosome divergence in an amphibian, we used genotype calls from ∼20 million positions, information about sex linkage, EST databases, and molecular evolution-

ary analyses to further characterize the sex chromosomes of the Western tropical frog *S. tropicalis*. Phylogenetic analysis of mitochondrial DNA sequences suggests our samples originated in Nigeria, which is also the source of the female individual from which the genome sequence was generated (Hellsten et al., 2010). Additionally, our analysis also suggests that the golden strain analyzed by Olmstead et al. (2010) is from Nigeria.

Known sequences in the *S. tropicalis* genome sequence assembly version 7.1 comprise ∼80.4% of the ∼1.7 Gbp genome, and scaffolds, including "N"s, comprise ∼84.5% of the genome. Thus the RAD tag data could not be compared to 15-20% of the genome because of gaps in the genome sequence. Due to variation in coverage, high confidence genotype calls were not made on scaffolds that together comprise an additional 2.1% of the genome. Thus, in this study we lack information from a non-trivial portion of this genome.

Mindful of these substantial gaps in genome sequence and the uncertainty in linkage relationships among many unassembled (orphan) scaffolds, we leveraged information from a targeted sequencing effort of chromosome arm 7p and also the linkage analysis by Olmstead et al. (2010) to guide our analysis. The dearth of genotypic patterns consistent with divergent sex chromosomes, and particularly patterns that are consistent with female heterogamy (Table 3.2), and the similar level of pairwise nucleotide diversity in males and females throughout the petite arm of chromosome 7 (Figure 3.3) argues strongly against there being a large sex-specific region of the *S. tropicalis* chromosomes. This inference is consistent with the findings of Uno et al. (2008) who detected no sex differences in C-banded heterochromatin in *S. tropicalis*.

Based on studies of fruit flies (reviewed in Bachtrog, 2013), we expected genes linked to the sex determining locus to potentially exhibit lower total expression and higher specificity (that is, higher intensity and lower evenness as defined in Methods). We also expected molecular evolution of these genes to be consistent with relaxed purifying selection. However, based on a small sample size, we only observed a significant increased expression intensity of "completely linked" genes compared to the rest of the genome, with none of these expectations met in "partially sex-linked" genes (Table 3.3). Some of these expectations were also met in orphan scaffolds, which have undetermined linkage relationships with respect to the sex determining locus, and regions of chromosome arm 7p. It is not clear that these latter observations are related in any way to linkage to the sex determining region.

Caveats exist in our interpretation of these data. First, non-recombining portions of the genome tend to accumulate repetitive sequences that can be difficult to sequence and map. For this reason, the sex-specific portion of the *S. tropicalis* genome may be underrepresented in the current genome assembly and/or our mapped Illumina reads. Second, it is conceivable that there is polymorphism in the sex determining mechanism (Olmstead, personal communication). Polymorphism in genetic sex determination could occur at a single locus wherein multiple, differently functioned sex determining alleles are segregating at a single locus that have distinct and not necessarily transitive dominance relationships,

or at multiple loci distributed on the same or different chromosomes. Sex determination in zebrafish, for example, appears to be orchestrated by genes on different chromosomes (Anderson et al., 2012). Genotypic patterns expected with these types of polymorphisms are unclear, and could include a dearth or absence of pronounced sex chromosome divergence. Third, polymorphism among females could also potentially exist in the extent of divergence between the W and Z chromosomes. Under this scenario, it is conceivable that there could be variation among populations in the extent of sex chromosome divergence. Further exploration of these possibilities will be assisted by the identification of the sex determining locus in *S. tropicalis*, the completion of high quality sequencing and assembly of sex-linked regions, and the exploration of variation within and among populations in sex determination and sex chromosome evolution.

3.6.1 Polyploidization, dosage compensation and sex chromosome turnover

Within a species, the propensity to undergo genome duplication and sex chromosome evolution are potentially interrelated. For example, polyploidization might be less common in species with divergent sex chromosomes where one has degenerated because, after duplication, the degenerate ancestral sex chromosome would segregate as a new autosomal chromosome, and the resulting homozygous null genotypes could be detrimental (Evans et al., 2012). Sex chromosome degeneration also creates imbalances in allelic copy number between the sexes, which can lead to the evolution of dosage compensation – a factor that is also potentially relevant to polyploid speciation (Orr, 1990). Dosage compensation is a process that equalizes expression levels in each sex of a gene that has a different number of alleles in each sex. This could evolve in a species with female heterogamy, for example, through inactivation of one of the Z alleles in males, or through upregulation of the Z allele in females. Orr (1990) proposed that dosage compensation in species with a degenerate sex chromosome could act as a barrier to polyploid speciation, because dosage compensation is disrupted when a newly formed triploid individual backcrosses with a diploid parental individual. Our analyses suggest that the sex-specific region of *S. tropicalis* is small, that sex chromosome divergence is minimal, and therefore that dosage compensation associated with degeneration of the sex specific sex chromosome would have evolved in very few genes or none at all. These features may have facilitated (or at least not impeded) polyploidization in *Silurana*, which occurred at least once (reviewed in Evans, 2008). Interestingly, the sister genus *Xenopus* has a newly evolved sex determining gene called *DM-W* (Bewick et al., 2011; Yoshimoto et al., 2008). Species in this group also probably have minimally diverged sex chromosomes, and have undergone polyploid speciation multiple times (Evans, 2008). Clearly, however, this is not the only consideration in the propensity of species to tolerate polyploidization because many amphibian groups that have homomorphic sex chromosomes lack polyploid species.

The extent of sex chromosome degeneration is also relevant to the propensity of species to experience future sex chromosome turnover – a change in which pair of chromosomes carries the trigger for sex determination (Charlesworth and Mank, 2010). If sex chromosome turnover occurs in a species with a diverged and degenerate sex chromosome, the ancestral degenerate chromosome could segregate autosomally, and some individuals could inherit two copies and be homozygous for degenerate alleles (Charlesworth and Mank, 2010). Thus, sex chromosome turnover may be more likely in species that have sex chromosomes that are not substantially degenerated.

If sex chromosome turnover were common, this could maintain homomorphy of sex chromosomes (the "frequent turnover" hypothesis). Recent work on sex chromosomes in African clawed frogs has established non-homology between the sex chromosomes of *X. laevis* and *S. tropicalis* (Uno et al., 2013). Thus it appears that the origin of a new sex determining gene in *Xenopus* (Bewick et al., 2011; Yoshimoto et al., 2008) was associated with a reassignment of sex chromosomes without a change in heterogamy. Another possibility is that the sex determining mechanism of non-diverged sex chromosomes could be old, but that divergence is prevented by periodic recombination, possibly facilitated by breeding individuals that are phenotypically sex reversed (the "fountain-of-youth" hypothesis; Perrin, 2009). Because we do not yet know the sex determining gene(s) of *S. tropicalis* or other frogs that could share this system (e.g., genera *Hymenochirus*, *Pseudhymenochirus*, or *Pipa*), we cannot determine at this time which hypothesis best accounts for the lack of extensive divergence of the sex chromosomes of *S. tropicalis*. Additional identification of sex determining genes in amphibians, and analysis of their evolutionary histories and genomic context, is a promising direction for future research that would further our understanding of how sex chromosomes evolve in general.

3.7 Conclusions

We used RAD tags, Sanger sequencing, sequences from a laser-dissected chromosome arm, EST databases, and molecular evolutionary analyzes to further characterize the sex chromosomes of the Western tropical frog *S. tropicalis*. We used genotype calls from $>$20 million sites to search for genotypic patterns consistent with sex chromosome divergence. Our findings strongly suggest that the sex-specific region of *S. tropicalis* is small based on (i) a dearth of regions with a pronounced genotypic signature of sex-specific portion of sex chromosomes, (ii) a lack of this genotypic signature in regions "completely linked" to the sex determining region inferred by Olmstead et al. (2010), and (iii) the observation of only minor differences in patterns of expression, and similar levels of purifying selection on "completely linked" genes.

3.8 Acknowledgements

We thank Richard Harland, Mustafa Khokha, James Evans and Maura Lane for assistance with samples and information on the "golden" strain of *S. tropicalis*. We thank 3 anonymous reviewers for helpful comments on an earlier draft of this manuscript. We additionally thank Brian Golding and Sharcnet (www.sharcnet.ca) for access to computational facilities. This research was supported by grants from the National Science and Engineering Research Council of Canada, an Ontario Internal Prestige Scholarship, an Early Researcher Award from the Ontario Ministry of Economic Development and Innovation, and McMaster University. Support for sequencing of the laser dissected of chromosome arm 7p was provided by UK Medical Research Council U117560482 and U117597137, the Grant Agency of Charles University in Prague (407311), the Grant Agency of Czech Republic (P502/11/P522), the Ministry of Education, Youth and Sports of Czech Republic (MSM0021620858), UNCE (204013), SVV (265211) and the Ministry of Agriculture of the Czech Republic (MZE 0002716202).

3.9 Supplementary information

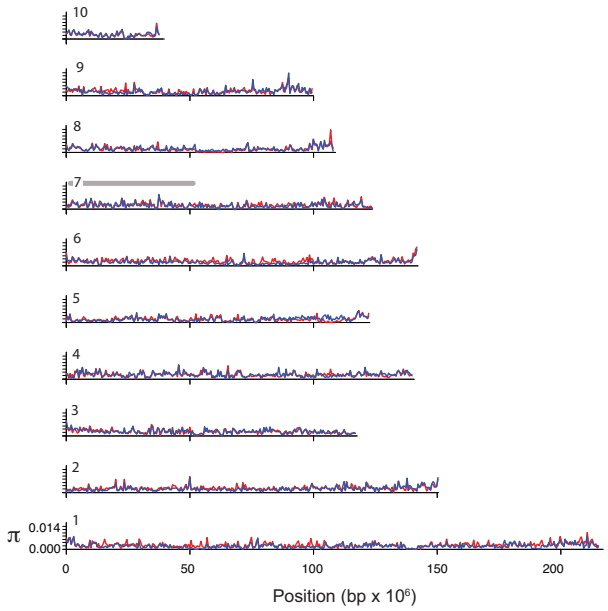

Figure 3.S1: Nucleotide diversity (π) in 500,000 bp windows is similar in three males (blue) and two females (red) that had an identical mtDNA genotype that differed from the golden strain genotype by one nucleotide substitution. Labelling follows Figure 3.3.

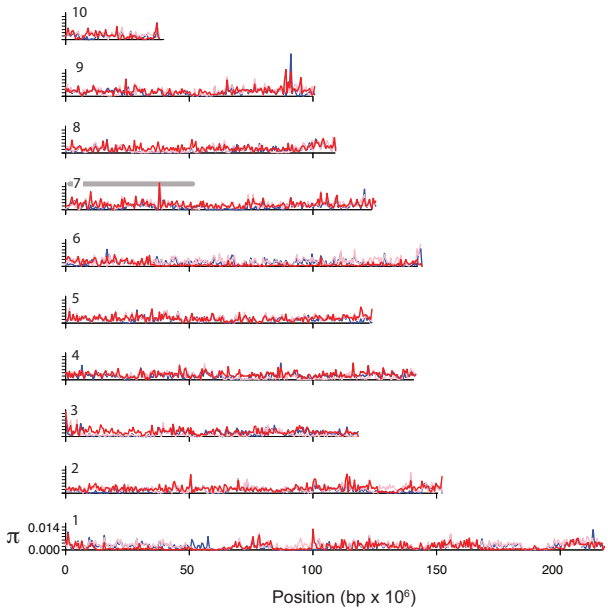

Figure 3.S2: Nucleotide diversity (π) in 500,000 bp windows is similar in one male (blue) and one female (pink) that had an identical mtDNA genotype that differed from the golden strain genotype by one nucleotide substitution, and a female (red) that had an identical mtDNA genotype as the golden strain. Labelling follows Figure 3.3. Breaks in the line for each individual correspond to windows that lack genotype data for that individual.

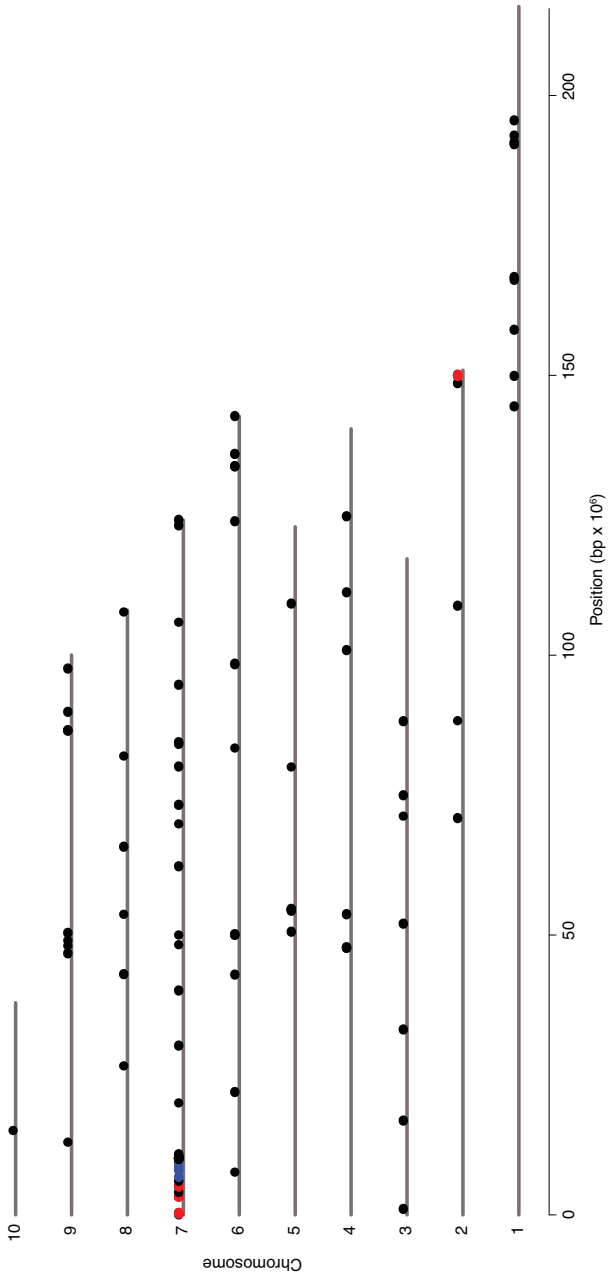

Figure 3.S3: Regions amplified on the ten large scaffolds. Black dots are regions not linked to the sex determining region, blue and red indicate regions that are linked and tightly linked respectively to the sex determining region based on the study of Olmstead et al. (2010).

Table 3.S1: PCR primers used in this study.

Scaffold	Forward	Primer sequence (5' - 3')	First base position	Reverse	Primer sequence (5' - 3')	Last base position	Number of females sequenced	Number of males sequenced	Female-specific polymorphic sites	Male-specific polymorphic sites	Shared polymorphic sites	Comments	Interpretation[a,b]
Primers targeting scaffolds designated as "completely linked" to the sex determining gene based on Olmstead et al. (2010).													
2	scaf2_149790591_f	GATCAGAACTCGTTGAATCTCTTT	149790591	scaf2_149791341_r	CGGGGGATTCACTGGTTATT	149791341	3	2	4	0	0	four indels present; sex specific	Not Scenario 3
2	scaf2_149810241_f	GGAGACAAGCGCTCAGAGG	149810241	scaf2_149810658_r	TCATACCTTCAAGCCCTTTACC	149810658	8	8	0	1	0		Not Scenario 3
2	scaf2_149954039_f	TGAAATTGGGCATCACAAAA	149954039	scaf2_149954800_r	GCTGTACTCTTGAATTGTGTTACCAT	149954800	8	8	1	0	0		Not Scenario 3
2	scaf2_149963385_f	ATCCCTAGTTTGGCCTGCTT	149963385	scaf2_149964156_r	TCTTTTGCTGATTCAGGGTACA	149964156	7	8	0	0	18		Not Scenario 3
2	scaf2_149987128_f	CATTAGCGTTGGCAATGATTT	149987128	scaf2_149987914_r	ATTGGCTGAAGGCCTAGTGA	149987914	7	7	0	7	8	one indel present; not sex specific	Not Scenario 3
2	scaf2_150031923_f	AAAATCTTCACACTATTTTTGATTTGA	150031923	scaf2_150032715_r	TGCAGGTAGAAGACGGTTAAGG	150032715	7	7	2	2	0		Not Scenario 3
2	scaf2_150071042_f	GATTGGTCCCCTTCCGTAAT	150071042	scaf2_150071793_r	GGATTTCAGGCCAGCAGTTA	150071793	5	3	1	0	0		Not Scenario 3
78	sc78v7_342950_f	TACCGTTTCTGGTGTTGCAG	342950	sc78v7_342218_r	CACGGCTGACATTAGGTTGA	342218	7	8	0	0	17	two indels present; one sex specific, one shared	Not Scenario 3
211	sc211v7_24962_f	TCTGCTTTTCTATCATTGCCTTT	24962	sc211v7_25832_r	GGAGCAGTCTCGGCATAAAA	25832	7	8	0	0	1		Not Scenario 3
211	sc211v7_29765_f	ACCGTGAGATTTCCCTGAGA	29765	sc211v7_30316_r	CAATATGAAACATGCATGATAGAGG	30316	7	1	0	0	0		Not Scenario 3
211	sc211v7_32722_f	GGGCTTAATTATGTTGCTTCG	32722	sc211v7_33330_r	AGCGTTTGGCATGTTCTACTT	33330	4	5	0	0	8	two indels present; not sex specific	Not Scenario 3
278	scaf278_338_f	GGCTCAGTTTAGTCCGGAAG	338	scaf278_865_r	CGGGGTATGCTGACCTTTT	865							Not Scenario 3
278	sc178v7-XPMC2-F	AGAAGAAGTGCTGGGAAGC	22449	sc178v7-XPMC2-R	GGAATGGGCTTAAGTCACG	22002	3	3	0	0	6		Not Scenarios 2 or 3; unlikely to be Scenario 1
278	scaf278_25301_f	TGAAAGTCCTTAGCCATTTGC	25301	scaf278_25900_r	GTCGGATTGGACGGAGAG	25900							Not Scenario 3
278	scaf278_26321_f	GGGCGGAGGATATGTTACC	26321	scaf278_26932_r	CAAATTCCCAGAAACAGACGA	26932							Not Scenario 3
278	scaf278_61762_f	TGATTCAGTGGAGGACTAAGGT	61762	scaf278_62556_r	CCCATGAAACTGGGTCTGTT	62556							Not Scenario 3
278	scaf278_76962_f	TGTCCCTTTACATGATGTTTATCG	76962	scaf278_77600_r	GGACAGAGTTTGTTGGTACGG	77600							Not Scenario 3
278	scaf278_127041_f	CGGAAGTAGACCACTTTTTCG	127041	scaf278_127830_r	GTCCCCATGGGAAAAATCTC	127830							Not Scenario 3
Primers targeting scaffolds designated as "partially linked" to the sex determining gene based on Olmstead et al. (2010).													
7	uphos-ex9-sc7v7-for1	AACCACTGCCCTTTTGTTTG	213466	uphos-ex9-sc7v7-rev1	TTTGGCCTTGCTTTAGTGCT	213873	2	4	0	0	2	two indels present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
7	uphos-ex9-sc7v7-for2	GGCTGATTCGGCCTGATTAT	213974	uphos-ex9-sc7v7-rev2	AACCACCAAAGCTGCATCA	213341	2	4	0	0	5	three indels present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
7	uphos-ex8-sc7v7-for2	CAGCTAGCACACATTATACACTGC	215033	uphos-ex8-sc7v7-rev2	CACAAAATGCAATGAGGTGTG	214409	4	4	0	1	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-u-phos-F	GTTTGTGGGTGCACTTCCAT	215406	sc7v7-u-phos-R	TTCCTGTTGGTGATGGCATA	214928							Not Scenario 3
7	236112-nonrep-sc7v7-for	TTGGCTGCCTAAGTTGAACA	236112	236764-nonrep-sc7v7-rev	CTTTGGGAGCTTTTCTGCTG	236764	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	285016-nonrep-sc7v7-for	TTCTGCTGTGTGTTGGCTGT	285016	285575-nonrep-sc7v7-rev	TCATGTATTGTGATAGAGAGGCAGA	285575	4	4	1	0	0		Not Scenario 3
7	323979-nonrep-sc7v7-for	TTGTGCTATTGTGTCCTTATTGC	323979	324489-nonrep-sc7v7-rev	GATTTCCATTGGCCTCTGTG	324489	4	4	0	0	0		Not Scenario 3
7	scaf7_3180963_f	TCATTGGCACTGCCCGTA	3180963	scaf7_3181432_r	GGCACAGCAGGGTCAAAT	3181432	4	4	0	0	0		Not Scenario 3
7	scaf7_3181766_f	GATTCTCTGAAATAACTGGGAAGG	3181766	scaf7_3182319_r	TGCTTTATACCCTGTAAGAATTCCA	3182319	4	4	0	0	0		Not Scenario 3
7	scaf7_3193681_f	ACAGGTGTTGCCTTTCTCC	3193681	scaf7_3194079_r	TGATGCTTATTCCCACCACA	3194079							Not Scenario 3
7	scaf7_3194602_f	TGGGGATTTACCTTGATGTG	3194602	scaf7_3195403_r	GTCTGTGTACCAGGGGCATT	3195403							Not Scenario 3
7	pdz.for	ATCAAGCAAGCCGGWGACAG	3212443	pdz.rev	GGGYTTKGTAKAAAGTGTTGC	3212141	4	4	0	0	0		Not Scenario 3
7	g.signaling.for	TGCTGAAGAAGGCKCAYGAT	4983824	g.signaling.rev	GGCTCYTCCARCATCTTCTC	4983932							Not Scenario 3
7	sc7v7-5000961-f	TTTCCCCCAACAAAGAAGAG	5000961	sc7v7-5001510-r	GGTAGGCTCCTGACCAATCA	5001510	6	8	1	0	7	two indesl present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
7	emx2-for1	AGTYTGGTTTCAGAACAGAAGGA	6698262	emx2-rev1	TGAATMCCTGKGGRTGACCTT	6697881							Not Scenario 3
7	rsubcomp.for	GGKCARTTTCCATGGMTGAT	7943624	rsubcomp.rev	CTCTRGAGMGCAACACTCAC	7944380							Not Scenario 3
7	sc7v7-8022490-necap1-f	TTCCGAAACATCACTTTCTACA	8022490	sc7v7-8023018-necap1-r	CCCCAAAGGTCTGATGAGGT	8023018	8	8	0	0	0		Not Scenario 3
7	sc7v7-8502172-f	GGCATGCCATCTGGAATAA	8502172	sc7v7-8502792-r	TTCTTTTTCCTAGCTCTTGCAC	8502792	8	8	2	0	0		Not Scenario 3
7	sc7v7-9027768-f	TGCTCCTCTGCTCTCTGACA	9027768	sc7v7-9028290-r	CTGCGCAGACAATCTTCTTG	9028290	8	8	1	0	0	one indel present; not sex specific	Not Scenario 3
7	sc7v7-9504085-f	AGACGCTGAGTCCGGATG	9504085	sc7v7-9504634-r	AGCTTCCAAACTGCGTTAGC	9504634	4	7	0	0	9	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-9935192-f	CCCGCAGCGTAATTGTATTT	9935192	sc7v7-9935749-r	CTTGCCGTTTTCTCTTTTGC	9935749	4	4	2	0	6		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-9935616-for	TCAATTAGTTTTAATGAGACAGATCCA	9935616	sc7v7-9936066-rev	GTTGGACCCTGCCCACAC	9936066	4	4	1	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-9937888-for	TCATTATCAGGCAAGAGATCTTCA	9937888	sc7v7-9938143-rev	TGCCAGCATAATCTAATATCCTCA	9938143	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-9937919-f	TCCCATTTTTGAAGGCAATAG	9937919	sc7v7-9938436-r	GATCCCCTAATCGTCTGAAG	9938436	8	8	1	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-9937919-f	TCCCATTTTTGAAGGCAATAG	9937919	sc7v7-9939600-rev	CTGCCAAAGAAACAACAGAAA	9939600	4	4	4	0	0		Not Scenario 3
7	sc7v7-9939052-f	ATAATGTGTTGAACAACGG	9939052	sc7v7-9939585-r	CAGAAAGGGAATATGGGTA	9939585							Not Scenario 3
7	sc7v7-9939361-for	TCCTGTGTAGATGTCGGCCTAA	9939361	sc7v7-9939600-rev	CTGCCAAAGAAACAACAGAAA	9939600	4	4	4	0	0		Not Scenario 3
22	scaf22_205044_f	AGAAGATATTCAAAGTTCATTTAGCA	205044	scaf22_205678_r	CCCATTTTGAGTTCAAACCATT	205678	4	4	0	0	6	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
22	scaf22_214162_f	TTTCTCCAGTGGCCCAATTT	214162	scaf22_214951_r	GGATATCCCTTTTCCCTAACA	214951	4	3	0	0	9	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
22	scaf22_265764_f	CAGAAGGTCTTCAATTTCTTTGAGA	265764	scaf22_266593_r	AGGGCCTCTGATCTCCCTAA	266593	4	4	0	0	0		Not Scenario 3
22	scaf22_437441_f	AACTTGATAGTCATCGAAAGTGATG	437441	scaf22_438155_r	GATCACCAAGTTATCATGAACAGG	438155							Not Scenario 3

Table 3.S1: *Continued from previous page*

Scaffold	Forward	Primer sequence (5' - 3')	First base position	Reverse	Primer sequence (5' - 3')	Last base position	Number of females sequenced	Number of males sequenced	Female-specific polymorphic sites	Male-specific polymorphic sites	Shared polymorphic sites	Comments	Interpretation[a,b]
Primers targeting scaffolds designated as "partially linked" to the sex determining gene based on Olmstead et al. (2010).													
22	scaf22_511489_f	CTGCCACTTTCTCCCTTGAT	511489	scaf22_512470_r	TCCAGCTCTTCTAAAGTTGGACT	512470	1	3	0	1	6		Not Scenarios 2 or 3; unlikely to be Scenario 1
22	scaf22_1073710_f	GGATCTATTTCTACGGAATCTGG	1073710	scaf22_1074239_r	TTGAGAAATTTCACCGAATCTG	1074239	3	2	1	0	7		Not Scenarios 2 or 3; unlikely to be Scenario 1
22	scaf22_1107366_f	ACAGTTGCAGTATTGGGCACAG	1107366	scaf22_1107680_r	TCCTGCCACGTGCAATTT	1107680	2	4	0	0	0		Not Scenario 3
94	scaf94_3414_f	AGAATTCCAGGCAGGATCAA	3414	scaf94_4165_r	CCTGTGTCTCTTCCCTCCAG	4165	4	4	0	5	0		Possibly Scenario 2
94	scaf94_37132_f	AGGGCTTAGTGAAGGGAGGA	37132	scaf94_37957_r	AGCACTAAGGCCCCTGGTAT	37957							Not Scenario 3
94	scaf94_40122_f	CAGGCAGAACTAGCCCAATC	40122	scaf94_40878_r	TACAACTGCCGTCAGCTCAT	40878	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
94	scaf94_47783_f	CCTGCAACCCTCTCACAAG	47783	scaf94_48532_r	CTCCCAAAGCAAAACAGGAA	48532	1	3	0	0	0	one indel present; not sex specific	Not Scenario 3
94	scaf94_113125_f	CAGAAGAATCTTTTACTACTGGTATCC	113125	scaf94_114079_r	TGGTGGTACATCTATTCACTGAACTGT	114079							Not Scenario 3
94	scaf94_165640_f	GGAAGAAGATCCCCATGGTT	165640	scaf94_166399_r	AGCCACCAGTAATTTGCTGTC	166399							Not Scenario 3
94	scaf94_315361_f	CCACTGTGATCCACTCATCTC	315361	scaf94_315830_r	ATTCTGTGGTACCTGGGCATA	315830	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
144	scaf144_5058_f	TGTTGGCCAATAATCCCTCA	5058	scaf144_5357_r	GATACAGGGTAAAAGTTCTCATCCA	5357	4	4	0	0	0		Not Scenario 3
144	scaf144_5211_f	CGTTCCCTACCTCCCCACTA	5211	scaf144_6240_r	TTGTGTATTACGCCTTTAACATTCA	6240							Not Scenario 3
662	scaf662_82_f	GAGGGGGAATACAAGTAGTATACAGG	82	scaf662_1515_r	GCGCTTGCCTACAGTTTGAT	1515							Not Scenario 3
7931	scaf7931_544_f	TCATGCGACCAACTGTGTCT	544	scaf7931_1308_r	GGCCAGCTGATTTCAAAGAT	1308							Not Scenario 3
7931	scaf7931_1217_f	GATTAAAGGCGTGGTCTGGA	1217	scaf7931_2097_r	AACGAACGACAAAGGGAAAA	2097	1	2	5	2	0	two indels present; not sex specific	Scenario 2 unlikely; Not Scenario 3
7931	scaf7931_1778_f	AACAGGGACCCCCTAAACAG	1778	scaf7931_2306_r	TCAACCTCTCTCTGCCCAAG	2306	2	3	5	0	0	one indel present; not sex specific	Not Scenario 3
Primers targeting other regions of the *S. tropicalis* genome.													
1	scaf1_ythdc1_f	GTGGTGAGCGGAAGAAGAAG	144373226	scaf1_ythdc1_r	TGCATTCTCCACAAGTCTGC	144374007	2	1	1	0	0		Not Scenario 3
1	scaf1_cisd2_f	CCTGTTTGTGACGGATCTCA	149812197	scaf1_cisd2_r	TGGTCCACTGCAGTCACTATG	149811529	4	4	0	0	0		Not Scenario 3
1	scaf1_phf17_f	GTACAGTGGCTCGCTCATCA	158057919	scaf1_phf17_r	GCAAGTCACAGTGCAAGGAA	158058687	4	4	2	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
1	sc1v7-166981981-for	CTCGGGGAGCACTGTAAAAA	166981981	sc1v7-166982371-rev	TGTGGCATTTTCTTCCTTCTTT	166982392	3	3	0	0	0		Not Scenario 3
1	sc1v7-167482335-for	AACCCACACTGCAGATTTCC	167482335	sc1v7-167482911-rev	AGTTCAGCCCTGTACCATGC	167482930	4	4	0	1	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
1	tapt1-for	CCATCCYCCAGAYGAAACTG	191199282	tapt1-rev	GAAATCAGTCWATYCTGTTKCC	191199511	4	4	1	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
1	lap3-for1	YGCCTTYCTGAAAGAGTTTG	191564827	lap3-rev1	TTAAAYAGTYTGWTTGTCTWCACTRAG	191564998	4	4	0	0	6		Not Scenarios 2 or 3; unlikely to be Scenario 1
1	scaf1_LOC100488800_f	TCCCTAAAAGAAGCCCATCA	192752336	scaf1_LOC100488800_r	TGCTTTCTTTAATCTGCTGCAA	192751628	4	4	2	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
1	c4orf23-for	GGTCAATCTCMCTYTWYACGTT	195472634	c4orf23-rev	ACCTTCACCATRTCTCTGTATTTTTC	195472856	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
2	sc2v7-70802215-for	GAGCAATGAGTGCAGCAGAG	70802215	sc2v7-70802758-rev	CACACAGGACTGTGGGAATG	70802777	4	4	4	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
2	sc2v7-88294523-for	TGAAGGAACGTGGGATCTTC	88294523	sc2v7-88295129-rev	AATTTGTGGCCCTCTTTGTG	88295148							Not Scenario 3
2	sc2v7-108750145-for	CTGCCAACCCAGATACACCT	108750145	sc2v7-108750751-rev	TTCATCTGGCCCATTTCTTC	108750770	4	4	3	1	9		Not Scenarios 2 or 3; unlikely to be Scenario 1
2	sc2v7-148515264-for	TGATCCCTTTTGGCTCTGTC	148515264	sc2v7-148515831-rev	TCTCTCGGGGCTCTAACAAA	148515850	3	2	1	2	0		Scenario 2 unlikely; Not Scenario 3
2	sc2v7_150720770_f	CGAGGCTCTCAAGCTCCTGT	150720770	sc2v7_150721049_r	CTCGAACACCCCCTCGAA	150721049	8	8	0	0	0		Not Scenario 3
2	sc2v7_150805291_f	ACGGAACCTGTAAGGGGAAC	150805291	sc2v7_150805678_r	GGCATTAGGCTGAATGACTG	150805678	8	7	0	0	16		Not Scenario 3
2	sc2v7_150811044_f	CATTCCCACGTGTCATAGCA	150811044	sc2v7_150811593_r	AGATTGCCCCTAAGTCTTGC	150811593	3	2	3	0	2	two indels present; sex specific	Not Scenario 3
3	scaff809-lg3-jhdm1d-for1	GGAACAGGCCGGTCAAAT	977299	scaff809-lg3-jhdm1d-rev1	CTCTTCCCACGGGGACAT	977100							Not Scenario 3
3	scaff40-lg3-fah-for	GCCAGAGACATCCAAAAGTG	16761161	scaff40-lg3-fah-rev	CCTGCACTGKATTWGGCAGAA	16761029							Not Scenario 3
3	scaff363-lg3-unknown-for	CTGACCTGGTTAGTGACCAACTT	33035760	scaff363-lg3-unknown-rev	ATGCCTTAAAAATTGGGTGAAA	33035191							Not Scenario 3
3	scaff177-lg3-phf15-for	CACAACCTGGAAATGCACAC	51946074	scaff177-lg3-phf15-rev	CYTCCAGCTCCATGAGTTTC	51946281							Not Scenario 3
3	acs5.for	ACCCTGAAAATMGTKGACAG	67879906	acs5.rev	TGTCGCCRTGYACAAACACC	67880033							Not Scenario 3
3	scaff230-lg3-unknown-for	TCTAGGTGAGCACTAGGGGAAT	71251069	scaff230-lg3-unknown-rev	TTCAGTATATACAGCAGGAAATCAGC	71250420							Not Scenario 3
3	scaff251-lg3-fgfr4-for	ACATCTACTTCTAGACCCYGGRAAT	74883409	scaff251-lg3-fgfr4-rev	CAACAACRGATATRGAGAACCTTCK	74883638							Not Scenario 3
3	sc3v7-88117924-for	CCCAGACCCAAGTCCAGTAA	88117924	sc3v7-88118500-rev	AAGCAGCAGGTAGGGAAACA	88118519	3	4	0	0	0		Not Scenario 3
4	sc4v7-47619381-for	TGAGGAAAATCCCATTGAGG	47619381	sc4v7-47619911-rev	AGTCCCACTGCAAAACATCC	47619911							Not Scenario 3
4	sc4v7-47643348-for	ATTGTGCCAGGCAATAAACC	47643348	sc4v7-47643877-rev	TTATTGGGGTGCCTTCTGAC	47643877							Not Scenario 3
4	sc4v7-47733072-for	ATCCCTCAATTGTGCGTGAC	47733072	sc4v7-47733591-rev	GGTCACACACAGCCTTGCTA	47733591							Not Scenario 3
4	scaf4_trpt1_f	AAGTGGCCGCTGGATAGAT	53615338	scaf4_trpt1_r	CAAACCCTTCACACCCTCAT	53616151	3	3	0	1	3		Not Scenarios 2 or 3; unlikely to be Scenario 1
4	scaf4_LOC100486965_f	CTTCCCCGTGGTCTATTTGA	53823997	scaf4_LOC100486965_r	AGCAAATGGATGCCAAAATC	53823205	4	4	0	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1

Table 3.S1: *Continued from previous page*

Scaffold	Forward	Primer sequence (5' - 3')	First base position	Reverse	Primer sequence (5' - 3')	Last base position	Number of females sequenced	Number of males sequenced	Female-specific polymorphic sites	Male-specific polymorphic sites	Shared polymorphic sites	Comments	Interpretation[a,b]
Primers targeting other regions of the *S. tropicalis* genome.													
4	scaff1-lg4-dmap1-for	GACTTGAACCCAATGCCAAC	100823319	scaff1-lg4-dmap1-rev	GGCCCCRACAACATCRAT	100823050							Not Scenario 3
4	scaff6-lg4-e2f4-for2	CTYTCYCCTCCTCCYGGTGA	111146693	scaff6-lg4-e2f4-rev2	GSCAAATAYGGRTTTATTTTG	111147417							Not Scenario 3
4	sc4v7-124733060-for	GGTGAGAGCCAATTCCGATA	124733060	sc4v7-124733630-rev	ACAAAGACGCTGAGCAAGGT	124733649	4	4	0	0	4		Not Scenarios 2 or 3; unlikely to be Scenario 1
5	scaff250-lg5-cdc40-for1	AACATGGCGGCTCTTACATC	50506344	scaff250-lg5-cdc40-rev1	CTTSACCRCTACCTCYGGTG	50506499							Not Scenario 3
5	agmat.for	TTGGGCCRCGYCAYATYAG	54248778	agmat.rev	GCCATGATYTTCTGGTAGG	54248877							Not Scenario 3
5	sc5v7-54429387-f	TTGACATTGCCTCCCTTTTT	54429387	sc5v7-54429915-r	GCAGTAGGAGACGTGAGATGC	54429915							Not Scenario 3
5	sc5v7-54551736-f	GCTTGGTTGAAGGCAAACAT	54551736	sc5v7-54552237-r	GGGAGGCATTACGACACATC	54552237							Not Scenario 3
5	sc5v7-54585339-f	CAATCATTTCCATCCCAACA	54585339	sc5v7-54585851-r	GGAGCAGAGTGCAATGGTTT	54585851							Not Scenario 3
5	sc5v7-54663872-f	CCGGGATTAACATCCTTTCA	54663872	sc5v7-54664392-r	TTTCGGCGACTTTTCCATAC	54664392							Not Scenario 3
5	sc5v7-80030129-for	GGCTTGCTCCTGTAACACCT	80030129	sc5v7-80030756-rev	GGAATTCCAAAGCATCAACTG	80030776	3	2	2	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
5	scaff25-lg5-prepl-for1	GGTGAAGAYAGYCTTGAGGTTT	109135596	scaff25-lg5-prepl-rev1	AACATTTTCTATACAATGYGTTAYTTG	109135790							Not Scenario 3
6	sc6v7-7634749-for	ACATCAATGCATGGGTCTCA	7634749	sc6v7-7635340-rev	AGTCCTGGGGTACTGCTTCA	7635359	2	3	3	0	0		Not Scenario 3
6	scaff96-chr6-esrp1-for	TCCAAYGAAGTGGCACAGT	21862771	scaff96-chr6-esrp1-rev	ACCTCTGCGGCAGTGCTC	21862482							Not Scenario 3
6	scaff420-chr6-lpin2-for	GTATAGCCGMTTAAGTGAGTTGGTG	42842481	scaff420-chr6-lpin2-rev	ATGMCATTTKACATGTTTCCCTGA	42842137							Not Scenario 3
6	sc6v7-49941142-for	CATGGGGGTATGTTCACTCC	49941142	sc6v7-49941684-rev	GCCAAGACCCTAAAATGTGG	49941684	6	5	0	0	9	multiple indels present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
6	sc6v7-50002329-for	TGCCCACTGAGCTCCTAAAT	50002329	sc6v7-50002837-rev	TGCACTAGCACAGCATCTCC	50002837	6	6	0	0	0		Not Scenario 3
6	sc6v7-50063642-for	CTGCAAAGGGGAAAACAACT	50063642	sc6v7-50064159-rev	TCACTGCTGGTTTACTCTCCCTA	50064159	6	6	0	0	5		Not Scenarios 2 or 3; unlikely to be Scenario 1
6	scaff57-lg6-fignl1-for	TYTGTATGGTGGTGSCAAAA	83415629	scaff57-lg6-fignl1-rev	AAACCATCAGCCTGCAAAAC	83416471							Not Scenario 3
6	sc6v7-98287997-for	AGGGGGTTCACCATAACACA	98287997	sc6v7-98288618-rev	TTGTCAGGCTGAAGTGATGC	98288637	4	4	1	2	0		Scenario 2 unlikely; Not Scenario 3
6	sc6v7-98404974-for	ATGTGCTGTGAGCTGAATGG	98404974	sc6v7-98405537-rev	TGCAAACACTGCTTTCCTTG	98405556	2	3	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
6	scaff133-lg6-pard3-for1	CACAGAATATGTTCCGCCAAG	123843736	scaff133-lg6-pard3-rev1	TCARCTGAATGTAAAACTTCTTTC	123843362							Not Scenario 3
6	scaf6_LOC100492461_f	CTAGTGGTGACTCCGCAACA	133642396	scaf6_LOC100492461_r	TCTCTGGGTCATTGTGGTCA	133643184	4	4	0	0	0		Not Scenario 3
6	scaf6_c11orf51_f	GGCTGATATGGAAGGAGCTG	133693222	scaf6_c11orf51_r	TCCACATTCCCAATGAACCT	133693736	3	1	3	0	0		Not Scenario 3
6	scaf6_ngfr_f	TGGCAGCTCTGTGGTGTAAC	133774579	scaf6_ngfr_r	TGTAATGCAGCGATCAGGAG	133775337	4	4	0	0	0		Not Scenario 3
6	scaf6_LOC100485553_f	TGATCATGATGGTCGGAGAA	135876575	scaf6_LOC100485553_r	TTCGTTCCATCAGCACTTTG	135875847	4	3	2	1	0		Scenario 2 unlikely; Not Scenario 3
6	scaf6_LOC100486202_f	GGCTGAACAACAAACCGTTC	135939642	scaf6_LOC100486202_r	CTGAGCAGGAAAATGGCTTC	135940426	2	3	1	2	3	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
6	scaf6_LOC733912_f	GGATACACTCACTTATTCACACA	142405296	scaf6_LOC733912_r	TGGGATAGCTCAGGAACGT	142404576							Not Scenario 3
6	scaf6_LOC100486242_f	TTGGCCCTCAGTTTTAGTGG	142642112	scaf6_LOC100486242_r	CCCCTCGCCAGTCATAATAA	142642793	4	4	5	2	11		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	680-nonrep-sc7v7-for	GTAGCAGTGCCTTGCTTTCC	661	1111-nonrep-sc7v7-rev	CACATATACCAGAGGGGGACA	1131	3	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7_882_f	CACACATCTATGCAGTAATGGATTT	882	sc7v7_1537_r	ACACTGATGGGCTTTTGTGA	1537	4	4	0	0	0		Not Scenario 3
7	sc7v7-3999931-f	AAAAAGGTTTTTGTTCCCGTA	3999931	sc7v7-4000480-r	TGCGAAACCAGAGATTATCC	4000480							Not Scenario 3
7	sc7v7-6002882-f	TCCCCTTTTAATGTTTAGAATGC	6002882	sc7v7-6003323-r	TGGTTACACCCAGTAGTGTAATTTCT	6003323							Not Scenario 3
7	sc7v7-9943085-for	CCTTGCAGACGAGCACATAA	9943085	sc7v7-9943599-rev	TTCTCACGCCAGAACTGATG	9943599	2	4	0	1	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-10005350-f	GAAAGTTCATTCTCATTGGGCG	10005350	sc7v7-10006756-r	AGTGGGGCTGATGTTGCTTT	10006756							Not Scenario 3
7	sc7v7-10005402-f	GATTAAATAACGTTACTTCC	10005402	sc7v7-10006756-r	AGTGGGGCTGATGTTGCTTT	10006756							Not Scenario 3
7	sc7v7-10005661-f	TCACGTAACACGTTGTGTGT	10005661	sc7v7-10006756-r	AGTGGGGCTGATGTTGCTTT	10006756							Not Scenario 3
7	sc7v7-10006382-for	CAGCAATGCCATTTTCCATA	10006382	sc7v7-10007200-rev	AGTTCTGTTTCTCCATGTGC	10007200	6	2	1	0	0		Not Scenario 3
7	sc7v7-10006382-for	CAGCAATGCCATTTTCCATA	10006382	sc7v7-10008060-rev	TGCACGAGCATGTGCGTGTG	10008060							Not Scenario 3
7	sc7v7-w-10006458-for	TTTTACTGGGGCCTCAGTTG	10006458	sc7v7-w-10006780-rev	GGCTTGTAAGGGTCTACAGTGGGG	10006780							Not Scenario 3
7	sc7v7-w-10006458-for	TTTTACTGGGGCCTCAGTTG	10006458	sc7v7-w-10006907-rev	TTATGCAACAGCGGGGGTGTTACG	10006907	4	2	0	0	0		Not Scenario 3
7	sc7v7-10006458-f	TTTTACTGGGGCCTCAGTTG	10006458	sc7v7-10007200-rev	AGTTCTGTTTCTCCATGTGC	10007200							Not Scenario 3
7	sc7v7-10006458-f	TTTTACTGGGGCCTCAGTTG	10006458	sc7v7-insert-r	AAAGTTGCATTTGTACTTTGTGCAT	10007046							Not Scenario 3
7	sc7v7-10006467-f	GGCCTCAGTTTCAAAAGTCC	10006467	sc7v7-10006969-r	GATGCAACTGGGTGTCTGTG	10006969							Not Scenario 3
7	sc7v7-10006467-f	GGCCTCAGTTTCAAAAGTCC	10006467	sc7v7-10008060-rev	TGCACGAGCATGTGCGTGTG	10008060							Not Scenario 3
7	sc7v7-10006546-f	GTAGGTTCAAATCATTGTGC	10006546	sc7v7-insert-r	AAAGTTGCATTTGTACTTTGTGCAT	10007046							Not Scenario 3
7	sc7v7-10006546-f	GTAGGTTCAAATCATTGTGC	10006546	sc7v7-w-10006907-rev	TTATGCAACAGCGGGGGTGTTACG	10006907							Not Scenario 3
7	sc7v7-10059835-f	TGTGTCTGGGACAGTTACACA	10059835	sc7v7-10060881-r	ATCCAGGGACCCTCCTGCTC	10060881	6	3	0	0	0		Not Scenario 3
7	sc7v7-10060881-f	GAGCAGGAGGGTCCCTGGAT	10060881	sc7v7-10062301-r	CTGCAGCACGCTGATATTAA	10062301							Not Scenario 3
7	sc7v7-10123137-f	ACATCCTATGGGAGGGGTG	10123137	sc7v7-10123299-r	GATGTAGGACTTGGGTGTCAGA	10123299							Not Scenario 3
7	sc7v7-10125922-f	CAAGGTGTTGTTTTGTAAAG	10125922	sc7v7-10128904-r	AACTCTACTTAAACCAGC	10128904							Not Scenario 3

Table 3.S1: *Continued from previous page*

Scaffold	Forward	Primer sequence (5' - 3')	First base position	Reverse	Primer sequence (5' - 3')	Last base position	Number of females sequenced	Number of males sequenced	Female-specific polymorphic sites	Male-specific polymorphic sites	Shared polymorphic sites	Comments	Interpretation[a,b]
Primers targeting other regions of the *S. tropicalis* genome.													
7	sc7v7-10128301-f	GCAACTATGAGTTGGAACTT	10128301	sc7v7-10128904-r	AACTCTACTTAAACCAGC	10128904							Not Scenario 3
7	sc7v7-10128301-f	GCAACTATGAGTTGGAACTT	10128301	sc7v7-10129920-r	CTCATTGCTCATTGTTTTGGTG	10129920	7	3	26	0	0	Highly polymorphic in females	Not Scenario 3
7	sc7v7-10128997-f	GCTTTAATCAGAGAAGAAATGAAA	10128997	sc7v7-10129269-r	CTGCTAAAAACATTATTCAAATA	10129269	6	5	3	0	0	polymorphisms are present in 5 out of 6 females	Not Scenario 3
7	sc7v7-10129571-f	GTCCTCCTATGACAGAGAAAAAGG	10129571	sc7v7-10129920-r	CTCATTGCTCATTGTTTTGGTG	10129920							Not Scenario 3
7	sc7v7-w-10129683-for	GACCCATTCCTGTATACA	10129683	sc7v7-10129920-r	CTCATTGCTCATTGTTTTGGTG	10129920							Not Scenario 3
7	sc7v7-w-10129683-for	GACCCATTCCTGTATACA	10129683	sc7v7-10130625-rev	TTCCGGTCGCAGCAGTAA	10130625							Not Scenario 3
7	sc7v7-w-10129683-for	GACCCATTCCTGTATACA	10129683	sc7v7-10129920-r	CTCATTGCTCATTGTTTTGGTG	10129920	5	3	0	0	0		Not Scenario 3
7	sc7v7-10133921-f	CCTGTCTATGATTCCTTCTTCAA	10133921	sc7v7-10134218-r	CACCCGTAGGCCACCTTTAC	10134218	2	1	0	0	0		Not Scenario 3
7	sc7v7-10179281-f	GGTGTTGCTTGTCACAATGG	10179281	sc7v7-10179640-r	CACCAGGTTGTAAATTTCTGTATTTC	10179640	4	4	0	0	0		Not Scenario 3
7	sc7v7-10187376-f	GCAGTGAGAAGCAGTTGTGC	10187376	sc7v7-10187923-r	GAGCTTCGCAATTGACCTGT	10187923	4	4	0	0	7		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-10242496-f	TTTGTGATCAAATGGGATGG	10242496	sc7v7-10243024-r	TGGTGTTTGCCTGTTACCAA	10243024	1	2	1	1	2	two indels present; not sex specific	Not Scenario 2; Not Scenario 3
7	sc7v7-10480345-f1	CTGGGGTACAGAGGGAGAG	10480345	sc7v7-10481485-r1	GGGGGCTAAATGTGCTGTAA	10481485							Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-10480402-f2	ATTTTGCTTTTATACACAGGAAGGA	10480402	sc7v7-10481421-r2	AAGACTGTAGGGGCCACACC	10481421							Not Scenario 3
7	sc7v7-10749020-stambpl1-f	CGCTTCTCTCCCTTTCAGAC	10749020	sc7v7-10748725-stambpl1-r	CCCCTCCATCTAAATGCACA	10748725	8	8	0	0	0		Not Scenario 3
7	sc7v7-20002402-f	TTCTGCATTTCTGTTTGTGTCC	20002402	sc7v7-20003101-r	TCATTTTTCACATGCACTGCT	20003101	3	3	0	1	0	one indel present; not sex specific	Not Scenario 3
7	sc7v7-entpd1-F	CCAGTCTTCAGAAGCCGTTC	30157656	sc7v7-entpd1-R	AGCAAATGGCAAAGTCTGCT	30157176	1	3	0	0	0		Not Scenario 3
7	sc7v7-40000563-f	CTTGCACTATGTAACCAGTCCCTA	40000563	sc7v7-40001022-r	ATTGCAATTCAACCCCCACT	40001022	8	8	0	0	0		Not Scenario 3
7	scaf7_nrg3_f	TAGAAGGCAGCCACTGTGTG	48274758	scaf7_nrg3_r	GGTGCCCGAGTTGTTGTAGT	48275557	4	4	0	0	3		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-50009764-f	GGGTTTCCTTAACCCCCTTT	50009764	sc7v7-50010233-r	GTATTATCACATTCACTCCCCATGA	50010233	7	6	0	0	1		Not Scenario 3
7	sc7v7-tardbp-F	TCATGGAATTCGTACCAGCA	62192643	sc7v7-tardbp-R	CAACTGCAGCAAACAGCATT	62193140	3	3	1	0	0		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	cbl.for	GGAGTCTTCYCAGGTTCCTG	69862353	cbl.rev	TTCTTGCTCTCCTCCCCTTT	69862670	1	1	0	0	0		Not Scenario 3
7	scaf7_ncam1_f	CACGGGTCTTAAGCCAGAAA	73183463	scaf7_ncam1_r	ACGCAGGCTCTTCACCTAAA	73184057	4	4	0	0	0		Not Scenario 3
7	scaf7_LOC100491957_f	TAACGGCAGCCTTTGATTTT	73230423	scaf7_LOC100491957_r	AATGGAGGAACGTGATCTGC	73231163							Not Scenario 3
7	sc7v7_80025325_f	GTTTTCCCGTCCARYAATAA	80025325	sc7v7_80025474_r	TAGCTCTCCTGSGTGTASGC	80025474	1	1	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	scaf7_TTbA018g09_f	CAAGCTCATGGCCTCTCTTC	84065585	scaf7_TTbA018g09_r	CAATGTCACACAGCCTTGCT	84064793							Not Scenario 3
7	scaf7_gnb1_f	GATGGCCAAATGCTAGGAAA	84240264	scaf7_gnb1_r	GAGGGTAGCGCTTGTGAAAG	84239561	4	4	1	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	scaf7_tmem52_f	AGTGAGCAAGAGGGGACAGA	84366752	scaf7_tmem52_r	CCAACCATCTCAAGGCAACT	84366027							Not Scenario 3
7	sc7v7-dnajc11-F	GCACTACGTGGGACTTGGAT	94624781	sc7v7-dnajc11-R	CTGTGGATGCTGGAACCTTT	94624303	3	3	1	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	fut1.for	AGGATGATGCTTTTGMTKACTG	105893743	fut1.rev	AGAGGCGAAAGATCTGCAGG	105892707	1	1	0	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-kirrel2-F	CAGGGGTCATCGTCTCATTT	123108605	sc7v7-kirrel2-R	CTTATGGCCACCGTCAAAGT	123109086	3	3	0	0	0		Not Scenario 3
7	sc7v7_124100849_f	TTGGGCCRCGYCAYATYAG	124100849	sc7v7_124101024_r	GCCATGATYTTCTGGTAGG	124101024	1	1	0	1	7		Not Scenarios 2 or 3; unlikely to be Scenario 1
7	sc7v7-insert-f1	CTCAGCTATAACTGACTGCC	NA	sc7v7-insert-r	AAAGTTGCATTTGTACTTTGTGCAT	10007046							Not Scenario 3
7	sc7v7-insert-f2	AGTGGGTCACAAGCAACATG	NA	sc7v7-insert-r	AAAGTTGCATTTGTACTTTGTGCAT	10007046							Not Scenario 3
7	sc7v7-w-f	GAGTAAGATCTGCCAGCAAGA	NA	sc7v7-insert-r	AAAGTTGCATTTGTACTTTGTGCAT	10007046							Not Scenario 3
7	sc7v7-w-f	GAGTAAGATCTGCCAGCAAGA	NA	sc7v7-w-10006907-rev	TTATGCAACAGCGGGGGTGTTACG	10006907							Not Scenario 3
8	sc8v7-atg4a-F	TGTCACCTTACGCAAGCAAG	26623040	sc8v7-atg4a-R	TCAGCCATTGATGGCACTTA	26623539	3	3	0	0	0		Not Scenario 3
8	sc8v7-42921671-for	GCTGGGAAATTCCTGTTTAGG	42921671	sc8v7-42922207-rev	TGCTAGAAGTGATGGCAGTTTT	42922228	4	4	0	0	3		Not Scenarios 2 or 3; unlikely to be Scenario 1
8	sc8v7-LOC548739-F	GGCACCTGAAAATTCACACA	53704500	sc8v7-LOC548739-R	TTGGAACACTGGACACAATGA	53704026	3	3	0	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
8	sc8v7-65697357-for	GGCTTAACCAGGTGTGGAAA	65697357	sc8v7-65697931-rev	AAGCAGCACCAGTAGGGAGA	65697950	4	4	1	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
8	sc8v7-daam1-F	TTTGCAGAAAGAGGGCAACT	81985187	sc8v7-daam1-R	ACATGGCATTCCCTATTCCA	81984710	3	3	0	0	0		Not Scenario 3
8	sc8v7-crabp2-F	TGATTGCCGATGATGTTGTC	107718976	sc8v7-crabp2-R	TATGGTGGAGCAGGCCAAT	107719425	3	3	8	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
9	scaf9_itgb6_f	GTGTAAGCATTACGTCTGCGTCT	12993896	scaf9_itgb6_r	GGTGCCTCAAGCAAGAGAAA	12993522	4	4	0	0	0		Not Scenario 3
9	scaf9_ube2f_f	TCAAGCGTTATGCCAGATGA	46638662	scaf9_ube2f_r	TAGCAGGACGGGCAGAGTTA	46639161	3	4	0	0	0		Not Scenario 3
9	scaf9_LOC100485964_f	GGCTTCAAAGGCATACAGG	48086512	scaf9_LOC100485964_r	GATGGAATGAAGGTGGCACT	48087300	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
9	scaf9_col18a1_f	AGAGCTCATGGGTGCTCATT	49031258	scaf9_col18a1_r	TTGGTTTTCCAACTGTGCAA	49030551	2	4	0	0	0		Not Scenario 3
9	sc9v7-50298166-for	TTGGGGATGCGTAAAGTTTC	50298166	sc9v7-50298760-rev	ATCTTGGTAACCCCCTGGTC	50298760	4	4	0	1	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
9	sc9v7-86414721-f	GGCCTTGACTGGGGTGAG	86414721	sc9v7-86415120-r	CCCCCAAAATTACTAGCAGCA	86415120							Not Scenario 3

Table 3.S1: *Continued from previous page*

Scaffold	Forward	Primer sequence (5' - 3')	First base position	Reverse	Primer sequence (5' - 3')	Last base position	Number of females sequenced	Number of males sequenced	Female-specific polymorphic sites	Male-specific polymorphic sites	Shared polymorphic sites	Comments	Interpretation[a,b]
Primers targeting other regions of the *S. tropicalis* genome.													
9	sc9v7-86517567-f	TGGCTACATTTCAGCACCAA	86517567	sc9v7-86518100-r	TAATGGGCACAACCAACTCA	86518100							Not Scenario 3
9	sc9v7-86556168-f	AGGTCCCACCAGTCCTAACA	86556168	sc9v7-86556419-r	GGAAGGCCGAACCATAATTT	86556419							Not Scenario 3
9	scaf9_LOC100488116_f	GTGGATTTAGCGGGAAGTGA	89790577	scaf9_LOC100488116_r	AAGTCGGTGCATTTTCCTTG	89791336	4	4	0	6	0		Not Scenario 3
9	scaf9_prss29_f	CTTAGATGGGGCCTTTAGGG	97526553	scaf9_prss29_r	ATCGCTAGGTTTTGCCACAG	97525785	4	4	0	1	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
10	sc10v7-15075106-for	AAACATGCTGCACAGTGCTC	15075106	sc10v7-15075764-rev	TAATGCAGTTCCCCCTTGAC	15075783	3	2	3	0	6	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
18	scaf18-1434882f	GGGGAACAAACACAAAACCA	1434882	scaf18-1435585r	ATGGCCAGTCAAACCAAGTT	1435585							Not Scenario 3
18	scaf18-1459781f	CAAATCAGACTTTCAGTTCTTCAA	1459781	scaf18-1460554r	GGGACAATGGTACGACAACA	1460554							Not Scenario 3
18	scaf18-1470993f	CGCCGCAACCTGTTTCTAT	1470993	scaf18-1471599r	ACCTTTATTTCCCCCATTCC	1471599							Not Scenario 3
163	scaf163_nr1_f	CCCTATATAGTGAACAACAGATGG	155463	scaf163_nr1_r	GCAATGATATCAGGCATAA	156032	3	3	31	1	0		Not Scenario 3
264	scaf264_nr1_f	TGCCACTTGCAAACAGAGAC	14101	scaf264_nr1_r	CACAACACAGCATAACGTACATATT	14800	4	4	1	0	0		Not Scenario 3
267	scaf267_LOC_f	AAGATGTCCACGGAGAGCAC	134707	scaf267_LOC_r	CAGAGAAAATAACTGCAATGAAACC	134443	4	4	0	0	1		Not Scenarios 2 or 3; unlikely to be Scenario 1
344	scaf344_nr1_f	CTGTCCAACGGGCCTTACT	81	scaf344_nr1_r	TTGCTGCTCCCAAACTCTTC	860	2	2	5	0	0	multiple indels present; not sex specific	Not Scenario 3
355	scaf355_nr1_f	GCAGGATTACCCAAAGAAGG	69284	scaf355_nr1_r	CTTCTCCGTCTTCCGAACCT	70061	3	3	1	0	2	multiple indels present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
386	scaf386_LOC_f	CCGTCAAAACCCACAGAAAT	20750	scaf386_LOC_r	ATGTGAGCGCACACTGAAAC	21480	4	4	0	0	3		Not Scenarios 2 or 3; unlikely to be Scenario 1
400	scaf400_31448_f	TGGTCTGTTTTGCAGCACCT	31448	scaf400_32158_r	GGAGCAAAATTAGAGCAATAAATCA	32158	4	4	0	0	0		Not Scenario 3
400	scaf400_32326_f	ATAAATTCACCAATATCTGTTTTGAGG	32326	scaf400_32875_r	GACTAATTTCAGGATATGACTAATTGC	32875	4	4	1	1	0		Not Scenarios 2 or 3; unlikely to be Scenario 1
423	scaf423_nr1_f	CCAGCAATGTGCTTCAGTTG	6734	scaf423_nr1_r	TTTTTCCAGAATAATAAAATCCCTAAA	7357	4	3	0	0	0		Not Scenario 3
423	scaf423_nr2_f	TTTGCCCATAAACATAAAATGC	17043	scaf423_nr2_r	CAGCTGCTGCTTCTCAAATG	17832							Not Scenario 3
486	scaf486_nr1_f	CAAGTAGGCAAGGTAGAATAAAGGA	6254	scaf486_nr1_r	AGCTGAGACTGTTGGTGCTG	6960							Not Scenario 3
554	scaf554_nr1_f	TGTTGGTACGTGCCTGTTTG	31797	scaf554_nr1_r	GGCTCGCCATTGTGTAAGAT	32683							Not Scenario 3
578	scaf578-10894f	CACCTATGGCTGCCTCTCAT	10894	scaf578-11688r	AGCAGGTCCTTGCGACTTTA	11688							Not Scenario 3
678	scaf678-11724f	CGGGGTTGTACATTTCAAGG	11724	scaf678-11947r	TCAAGGGGACAAAATTCAGC	11947							Not Scenario 3
691	scaf691-10219f	ATTGCCAATCCACCACAAAT	10219	scaf691-10819r	CAACATTCGATTGCCATGAT	10819							Not Scenario 3
698	scaf698-1175f	TCTGTTCTAAACGCTGCTGCT	1175	scaf698-1685r	TTCAGCAAAGGCAAGTGAGA	1685							Not Scenario 3
703	scaf703-17654f	TTTCAGAATTGTGATGGACTGTG	17654	scaf703-18373r	AACTAGGCGGAGACAACCTG	18373							Not Scenario 3
749	scaf749-894f	ATTATGGGCACTCTGCTTCG	894	scaf749-1569r	GAGCCTACAGCTGGAAATGC	1569							Not Scenario 3
781	scaf781-11842f	GACACCAACAAGTTTAAAGT	11842	scaf781-12692r	CCTTAGCATACCAGCTTATG	12692							Not Scenario 3
793	scaf793-20901f	GACCCAATCAAGTTGCATTA	20901	scaf793-21566r	AATAGCAACCGGAGGGTTAC	21566							Not Scenario 3
800	scaf800-77f	GCTATCGCTCCAGATGGTGT	77	scaf800-429r	AGGGTAGCAGACGGGGTAGT	429							Not Scenario 3
835	scaf835_16643_f	TATTTATTAACGCCATCATTAAAGG	16643	scaf835_17422_r	CTTGTGGCGCAGCAGAAA	17422	3	3	2	1	13	multiple indels present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
835	scaf835_20658_f	ATATGAGTCTGAGCGGATCG	20658	scaf835_20960_r	TTTAATCACGTTAAATGGAGACG	20960	4	4	0	0	4		Not Scenarios 2 or 3; unlikely to be Scenario 1
838	scaf838-4542f	GGGGAGACTAACATTTTTGC	4542	scaf838-5013r	ATGGCACTGAGGTACAGTAG	5013							Not Scenario 3
842	scaf842-17833f	GAAGCAGTTGCAACAATGCC	17833	scaf842-18435r	AATCAGCCACTATGACCTCT	18435							Not Scenario 3
851	scaf851_19627_f	CTTAGCTTTTTACTGTCAGTCTAACCA	19627	scaf851_20226_r	TTGCATACTTGTCACATTTCAGG	20226	2	4	0	0	0		Not Scenario 3
856	scaf856-4333f	TTCCTGTATTTTATTTGACC	4333	scaf856-4864r	TAACTCTGCAGCTTCTTGTT	4864							Not Scenario 3
880	scaf880-3922f	AAGGGCAACATGACCTTGTAA	3922	scaf880-4480r	GGCAGACACTAACTGGCACT	4480							Not Scenario 3
906	scaf906-16722f	CTGAGGCTAACTCAGGCTTGT	16722	scaf906-17278r	TGCATGATTTTACCTACTTTGCTT	17278							Not Scenario 3
919	scaf919-33527f	CCTACCAATAAGCCCCCTGT	33527	scaf919-33759r	GATTGTTTGGCCCCAGTTT	33759							Not Scenario 3
953	scaf953_7503_f	ATGGTTACGCCGCTGATTAC	7503	scaf953_8155_r	TCGGCACAATAAGGTTTTCC	8155	3	2	2	1	1	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
953	scaf953_15516_f	ATCTGCCAGTTCCCATGTTG	15516	scaf953_16295_r	TGTTACCCCAAATCCTGCAT	16295	3	3	8	0	1	one indel present; not sex specific	Not Scenarios 2 or 3; unlikely to be Scenario 1
1898	iqcb1-for1	RCGCCAGTTAGATGAGGATT	7369	iqcb1-rev1	CCTGTTGGTAAGAACTAGAGTACAGG	6914							Not Scenario 3
2472	scaf2472_7180_f	TACCTCCAGTCCATGCAACA	7180	scaf2472_7660_r	CAGCTAGAAATTTTTGCTACAATCC	7660	2	3	0	0	0		Not Scenario 3
5248	scaf5248_3331_f	GAGTAAGCAAGCTGGGCAAA	3331	scaf5248_3990_r	TGGTAACTTAACCAACAGAAAACAA	3990	2	4	0	3	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
5485	sc5485_29_f	GCACAATAAACAACAAGATACCG	29	sc5485_420_r	CTGATCCCCATTCACCTGTT	420	12	12	1	0	2		Not Scenarios 2 or 3; unlikely to be Scenario 1
5485	sc5485_1356_f	GAGCACCCTATCTAACAAACCTC	1356	sc5485_2030_r	AGCACAATTTGCCAATGG	2030	12	12	0	0	0		Not Scenario 3
6052	scaf6052_1241_f	TTAATTGAACTATGACTGTGAGTATGG	1241	scaf6052_1840_r	CTTAGGGTCCAACTATTCCTGCAT	1840	1	1	0	0	0	multiple indels present; not sex specific	Not Scenario 3
6052	scaf6052_3213_f	TGTGCTATTATCTTCCCCAAGC	3213	scaf6052_3927_r	GGCATCTGCAGGTTCCTA	3927	4	4	1	0	0		Not Scenario 3
8518	scaf8518_802_f	CCAGCAATGATATAGGGAGTCAG	802	scaf8518_1565_r	GGAAGGGGCTTTTGACTTTC	1565	2	2	1	1	3		Not Scenarios 2 or 3; unlikely to be Scenario 1

[a] In amplifications with shared polymorphisms, Scenario 1 is unlikely because it would require convergent mutation on the sex chromosomes.
[b] None of the amplifications were sex specific, allowing us to exclude Scenario 3.

Chapter 4

The sex chromosomes of frogs

Bewick AJ and Evans BJ

4.1 Preface

Molecular evolutionary studies of sex chromosomes have focused on understanding the consequences of suppressed recombination between each sex chromosome (the X and Y or Z and W). However, little is understood about why and how some organisms maintain recombination throughout most of the sex chromosomes and thereby circumvent some of the consequences of nonrecombination. Here we review recent studies of sex chromosome evolution in frogs and provide insights into phenomena that cause and have maintained homomorphic sex chromosomes.

4.2 Abstract

Cold-blooded organisms, like frogs (order Anura), commonly possess cytologically indistinguishable (homomorphic) sex chromosomes. Explanations for the presence of homomorphy in frogs include recent turnover events and recombination between sex chromosomes. The absence of dosage compensation in frogs may also act as a barrier to sex chromosome divergence and heteromorphy, and together these phenomena may be related to polyploid speciation. Minimally divergent sex chromosomes provide a window into the past and can elucidate early evolutionary processes that acted during sex chromosome evolution in species with heteromorphic sex chromosomes, including humans.

4.3 Introduction

Sex chromosomes independently evolved from autosomes multiple times in many groups of plants and animals (Ohno, 1967; Bull, 1983; Charlesworth, 1996) including frogs (Hillis and Green, 1990; Evans et al., 2012). According to models of sex chromosome evolution, the first step is taken when an autosomal mutation seizes a leading role in the sex determining pathway, such that heterozygotes develop into one sex, and homozygotes into the other (Ohno, 1967; Charlesworth et al., 2005). In the second step, sexually antagonistic mutations are expected to accumulate in the vicinity of this gene, benefiting from linkage disequilibrium (Bull, 1983; Rice, 1996). Recombination in the heterogametic sex is suppressed through mutations that eliminate homology, providing epistatic interactions between the sex determining and sexually antagonistic genes (Rice, 1996; Charlesworth and Charlesworth, 2000). However, suppressed recombination also lowers the efficacy of selection, manifested as slow adaptive evolution of sex-linked genes compared with genes on the shared sex chromosomes (Orr and Kim, 1998) and as reduced effectiveness of purifying selection, causing accumulation of deleterious mutations (Agulnik et al., 1997; Fridolfsson and Ellegren, 2000; Filatov et al., 2001; Bachtrog and Charlesworth, 2002; Filatov and Charlesworth, 2002; Wyckoff et al., 2002; Tucker et al., 2003; Berlin and Ellegren, 2006; Kaiser, 2010; Kaiser and Charlesworth, 2010). This leads to divergence between non-recombining portions of each sex chromosome. Sex chromosome divergence involves the loss of coding regions, the accumulation of repetitive regions, and structural changes such as insertions, deletions, and inversions on the sex-specific chromosome (the Y or W). Thus, diverged (also called "degenerate") sex chromosomes have differences that extend beyond the fundamental difference in the presence or absence of a sex determining allele (Figure 4.1). Not surprisingly, degeneration of sex chromosomes can increase over evolutionary time, and the effects of such processes are exemplified by the drastic size and molecular differences between sex chromosomes. In humans, for example, almost all of the ancestral genes persist on the X-chromosome, but have been lost on the Y-chromosome (Skaletsky et al., 2003). In contrast, recently evolved sex chromosomes are expected to be nondegenerate since suppressed recombination has not taken affect; therefore recombination is still occurring across the length of the chromosome, which contains most of the ancestral sequence. However, sex chromosome degeneration is not an inevitable evolutionary outcome and many species with evolutionarily old sex chromosomes are morphologically similar, or homomorphic (Ogawa et al., 1998; Matsubara et al., 2006; Tsuda et al., 2007). The persistence of nondegenerate sex chromosomes over vast amounts of evolutionary time is evidence against sex chromosome evolution theory in terms of degeneration of the sex-specific sex chromosome.

Processes that maintain homomorphic sex chromosomes are not well understood, but frequent turnover of sex chromosomes (i.e., the "high-turnover" hypothesis), or recombination between sex-reversed individuals (i.e., the "fountain-of-youth" hypothesis) could explain the lack of degeneration in some cases (Perrin, 2009; Stöck et al., 2011). Also,

the rate of polyploidization may potentially be related to sex chromosome evolution (Orr, 1990; Evans et al., 2012). Interestingly, homomorphic sex chromosomes seem to be more prevalent in cold-blooded vertebrates like amphibians (frogs: Table 4.1, and salamanders) and reptiles. Also, unlike other vertebrate groups, like therian mammals, the sex chromosome system is not fixed in frogs (Table 4.1). Here we review current literature on sex chromosome evolution in frogs and provide insights into phenomena that cause and have maintained homomorphic sex chromosomes.

Figure 4.1: A generalized model of sex chromosome degeneration (adopted from Bachtrog (2013)). (A) Sex chromosomes formed from ordinary autosomes, which contain identical sets of genes. (B) A potential first step in the evolution of heteromorphic sex chromosomes is the acquisition of a sex-determining locus on a proto-sex chromosome, such as a male-determining gene. (C) Accumulation of sexually antagonistic mutations close to the sex-determining region select for suppression of recombination on the proto-sex chromosomes, which can be achieved by chromosomal inversion(s). (D) The non-recombining region can increase if other mutations with sex-specific fitness effects accumulate on proto-sex chromosomes. Moreover, a lack of recombination results in the accumulation of loss-of-function mutations at Y-linked genes (pseudogenization). (E) Lack of recombination also results in an accumulation of repetitive DNA, which can lead to an increase of the size of the diverging Y chromosome. (F) Large segments of non-functional DNA can be deleted in Y chromosomes and can reduce their physical size. The evolutionary outcome of this process is heteromorphic sex chromosomes, in which the X chromosome largely resembles the ancestral autosome, and the Y chromosome has lost most of its ancestral genes and may instead have accumulated repetitive DNA, and beneficial sex-specific alleles and genes. Additionally, at least some portion of the Y chromosome will resemble the X chromosome, and will recombine with the X chromosome during meiosis – this region is known as the pseudoautosomal region (PAR) and is important for the proper segregation of the sex chromosomes.

Table 4.1: A list of known sex chromosomes in frogs compiled from primary literature searches. Heteromorphic/Homomorphic sex chromosomes were scored based on gross morphological differences, banding patterns using various stains for heterochromatin, and/or genomic rearrangements (i.e., inversions and autosomal fusions (A_Y)).

Family and species	Sex chromosomes	Heteromorphic/Homomorphic	Description	Reference
Alsodidae				
Eupsophus insularis	XY	Heteromorphic	Pericentric heterochromatin on Y; absent on X	Cuevas and Formas (1996)
Eupsophus migueli	XY	Heteromorphic	Metacentric Y; telocentric X	Iturra and Veloso (1981)
Eupsophus roseus	XY	Homomorphic	–	Iturra and Veloso (1989)
Bufonidae				
Bufo bufo	ZW	Homomorphic	–	Engel and Schmid (1981)
Bufo marinus	ZW	Homomorphic	–	Abramyan et al. (2009)
Bufo viridis subgroup	ZW	Homomorphic	–	Schmid (1978); Stöck et al. (2005)
Bufo viridis (Moldovian)	ZW	Heteromorphic	Inversion on W	Odierna et al. (2007)
Centrolenidae				
Centrolenella antisthenesi	XY	Heteromorphic	Centromeric and pericentromeric heterochromatin on X; centromeric heterochromatin only, and on the long arm telomeres on Y	(Schmid et al., 1989)
Discoglossidae				
Discoglossus pictus	ZW	Heteromorphic	No confidence	Morescalchi (1964)
Hemiphractidae				
Gastrotheca ovifera	XY	Heteromorphic	X >Y; large amounts of heterochromatin on X; extremely little on Y	Schmid et al. (1988, 2002a)
Gastrotheca walkeri	XY	Heteromorphic	X >Y; large amounts of heterochromatin on X; extremely little on Y	Schmid et al. (1988)
Gastrotheca pseustes	XY^A	Homomorphic	–	Schmid et al. (1990)
Gastrotheca pseustes	XY^B	Heteromorphic	X <Y; large heterochromatin region on long arm telomere on Y; absent on X	Schmid et al. (1990)
Gastrotheca riobambae	XY	Heteromorphic	X <Y and Y almost entirely heterochromatic	Schmid et al. (1983); Schmid and Klett (1994)
Hylidae				
Pseudis tocantins	ZW	Heteromorphic	Metacentric Z <subtelocentric W; inversion; increased heterochromatin on W	Busin et al. (2008)
Hyla femoralis	XY	Heteromorphic	X >Y; deletion of NOR on Y	Schmid and Steinlein (2003); Wiley (2003)
Hyla arborea japonica	XY	Heteromorphic	X >Y	Yoshida (1957)
Hyperoliidae				
Hyperolius viridiflavus ommatostictus	XY	Homomorphic	–	de Almeida et al. (1990)
Leiopelmatidae				
Leiopelma archeyi	ZW	Heteromorphic	Z <W; mostly heterochromatic short arm on W	Green (2002)
Leiopeima hamiltoni	ZW	Heteromorphic	Z <W	Green (1988, 2002)
Leptodactylidae				
Eleutherodactylus albipes	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus casparii	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus cuneatus	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus emiliae	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus glamyrus	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus turquinensis	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus sp. n. G	ZW	Heteromorphic	Z <W	Schmid et al. (2010)
Eleutherodactylus johnstonei	XY	Homomorphic	–	Schmid et al. (2010)
Eleutherodactylus oxyrhyncus	XY	Heteromorphic	X <Y	Schmid et al. (2010)
Eleutherodactylus cavernicola	XY	Heteromorphic	X <Y; A_Y fusion	Schmid et al. (2010)

Table 4.1: *Continued from previous page*

Family and species	Sex chromosomes	Heteromorphic/Homomorphic	Description	Reference
Eleutherodactylus euphronides	ZW	Heteromorphic	Z <W; heterochromatic W	Schmid et al. (2002a, 2010)
Eleutherodactylus shrevei	ZW	Heteromorphic	Z <W; heterochromatic W	Schmid et al. (2002a, 2010)
Eleutherodactylus riveroi	XY	Homomorphic	–	Schmid et al. (2003)
Eleutherodactylus maussi	XY	Heteromorphic	Multiple sex chromosomes	Schmid et al. (1992, 2002a)
Physalaemus ephippifer	ZW	Heteromorphic	Z <W; additional segment in the short arm composed of a distal NOR and an adjacent terminal DAPI-positive C-band on W	Nascimento et al. (2010)
Physalaemus petersi	XY	Heteromorphic	X >Y	Lourenço et al. (1999)
Pseudopaludicola saltica	XY	Heteromorphic	Duarte et al. (2010)	
Myobatrachidae				
Crinia bilingua	ZW	Heteromorphic	Acrocentric Z <submetacentric W; heterochromatin at centromere and proximal region of the short arm of the W	Mahony (1991)
Odontophrynidae				
Proceratophrys boiei	ZW	Heteromorphic	Heterochromatin restricted to centromeric region on Z; mostly heterochromatic W	Ananias et al. (2007)
Pipidae				
Silurana tropicalis	ZW	Homomorphic	–	Olmstead et al. (2010); Bewick et al. (2013)
Xenopus laevis	ZW	Homomorphic	–	Mikamo and Witschi (1963, 1964); Graf (1989); Watson and Kelley (1992); Yoshimoto et al. (2008)
Pyxicephalidae				
Pyxicephalus adspersus	ZW	Heteromorphic	Z >W; heterochromatic W	Schmid (1980); Schmid and Bachmann (1981); Engel and Schmid (1981)
Pyxicephalus delalandii	ZW	Homomorphic	–	Schmid (1980)
Ranidae				
Rana esculenta	XY	Homomorphic	–	Schempp and Schmid (1981)
Rana rugosa (Niigata)	ZW	Heteromorphic	Pericentric inversion	Uno et al. (2008)
Rana rugosa (Hamakita)	XY	Heteromorphic	Pericentric inversion	Uno et al. (2008)
Rana rugosa (Hiroshima and Isehara)	XY	Homomorphic	–	Miura et al. (1998)
Rana japonica	XY	Heteromorphic	Heterochromatin at basal portion of the long arm on X; no such banding on Y	Miura (1994)
Rana tagoi	XY	Heteromorphic	Heterochromatin on the long arm of X; absent on Y	Ryuzaki et al. (1999)
Rana sakuraii	XY	Heteromorphic	Metacentric X >submetacentric Y	Ryuzaki et al. (1999)
Rana ridibunda	XY	Heteromorphic	X >Y	Al-Shehri and Al-Saleh (2005)
Rana temporaria	XY	Homomorphic	–	Popov and Dimitrov (1999)
Rana tigrina	ZW	Homomorphic	–	Chakrabarti et al. (1983)
Rhacophoridae				
Buergeria buergeri	ZW	Heteromorphic	Subtelocentric Z >submetacentric W; satellite near the end of the long arm on Z	Schmid et al. (1993), Hanada (2002)
Strabomantidae				
Pristimantis euphronides	ZW	Heteromorphic	Z <W	Schmid et al. (2002b, 2010)
Pristimantis shrevei	ZW	Heteromorphic	Z <W	Schmid et al. (2002b, 2010)
Pristimantis pulvinatus	XY	Heteromorphic	X <Y; A_Y fusion	Schmid et al. (2010)
Pristimantis riveroi	XY	Heteromorphic	X <Y; A_Y fusion	Schmid et al. (2003, 2010)
Pristimantis sp. n. H	XY	Heteromorphic	X <Y; A_Y fusions	Schmid et al. (2010)
Strabomantis bioporcatus	XY	Heteromorphic	X <Y; A_Y fusion	Schmid et al. (1992, 2002a, 2010)

4.4 Evidence for recent evolution of sex chromosomes in frogs

The only known amphibian sex determining gene, *DM-W*, was first discovered in the pseudotetraploid African clawed frog *Xenopus laevis* and is a partial duplicate of the functional DM domain of the autosomal, paralogous gene *DMRT1* on the W-chromosome (Yoshimoto et al., 2008). DM domain-containing genes play a remarkably conserved role as an activator of male differentiation in metazoans including worms, flies, coral, birds, and humans (Burtis and Baker, 1989; Raymond et al., 1999; Yi and Zarkower, 1999; Raymond et al., 2000; Miller et al., 2003; Haag and Doty, 2005). *DMRT1* is broadly expressed during development of *X. laevis* and has been detected in unfertilized eggs, early tadpole development of both sexes, in the gonads of both sexes during primary gonadal differentiation, and in postmetamorphic testes and ovaries, with expression becoming increasingly male-biased in testes compared to ovaries by 1-5 months after metamorphosis (Osawa et al., 2005; Yoshimoto et al., 2006, 2008). No sex difference in *DMRT1* expression was detected during gonadal differentiation (Yoshimoto et al., 2008). In contrast, *DM-W* appears to be expressed only in female gonads during primary gonadal differentiation, with its peak expression level prior to a surge of *DMRT1* expression in the gonads of both sexes at Nieuwkoop and Faber stage 50 (Yoshimoto et al., 2008). Yoshimoto et al. (2008) demonstrated that genetic male (ZZ) tadpoles transgenic for *DM-W* were feminized (i.e., phenotypically female), implying that *DM-W* acts as a dominant-negative, antagonizing *DMRT1* activation of male-specific genes by binding to and inhibiting regulatory regions recognized by both proteins. Binding in females has been hypothesized to occur as a *DM-W* homodimer or as a *DM-W-DMRT1* heterodimer and in males it has been hypothesized that primary (gonadal) development is achieved by homodimerization of *DMRT1* (Yoshimoto et al., 2010).

Since Yoshimoto et al. (2008) discovered *DM-W* in *X. laevis*, Bewick et al. (2011) were able to identify *DM-W* in in several other *Xenopus spp.*, including *X. andrei*, *X. clivii*, *X. gilli*, *X. itombwensis*, *X. largeni*, *X. pygmaeus* and *X. vestitus* (Figure 4.2A: Clade 1). These species are tetraploid except *X. itombwensis* and *X. vestitus*, which are octoploid; however, no *DM-W* sequence has been identified in a dodecaploid species (Bewick et al., 2011). Though *DM-W* is female-specific in these other *Xenopus spp.*, its role in sex determination has not been conclusively tested – *DM-W* could be downstream of a more dominant, upstream, sex determining gene. Phylogenetically *DM-W* is restricted to one clade of *Xenopus spp* and is more closely related to *DMRT1*β than *DMRT1*α (Figure 4.2B). Therefore, the duplication event of *DMRT1*β gave rise to *DM-W* and occurred in *Xenopus* after the divergence from the sister genus *Silurana* but before divergence of most or all extant species of *Xenopus* (Bewick et al., 2011). Attempts to amplify *DM-W* in another African clawed frog, *Xenopus borealis* (Figure 4.2A: Clade 2), were performed using similar polymerase chain reaction (PCR), cloning, and sequencing techniques as described in Bewick et al. (2011). However, amplification was unsuccessful and resulted in no or unspecific amplification. 454 sequencing (ROCHE GS FLX) was also performed on normalized cDNA extracted

from *X. borealis* tadpoles at Nieuwkoop and Faber stages 48-52 and BLAST searches using full and partial *DM-W* sequences from this library did not recover a *DM-W* homolog (unpublished results).

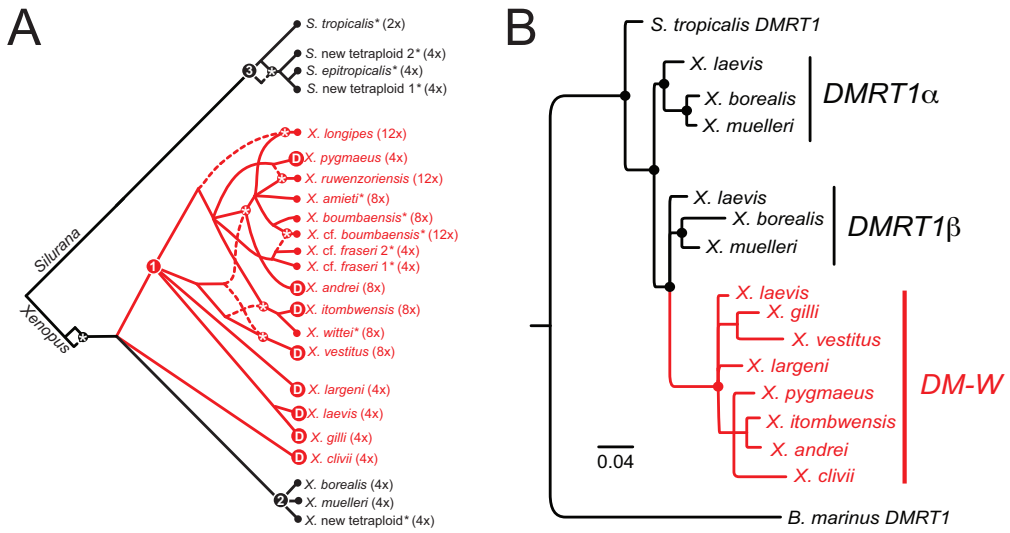

Figure 4.2: (A) *DM-W* was detected in a subset of *Xenopus* species whose most recent common ancestor and descendants are indicated in red. A "D" next to species names indicates species from which portions of *DM-W* was amplified. Nodes with 1, 2 or 3 inside refer to Clades 1, 2 and 3, respectively. An asterisk follows species names for which amplification of *DM-W* were not attempted. Allopolyploid speciation is indicated by asterisks on internal nodes and the ploidy level of each species is indicated after the species name. Maternal or biparental relationships are indicated with solid lineages and paternal contributions to allopolyploid speciation events are indicated with dashed lineages. Resolved nodes have >95% posterior probability and are based on the linked autosomal genes RAG1 and RAG2 and mitochondrial DNA; details of phylogenetic estimation are given in Evans et al. (2004, 2005) and Evans (2007). Strongly supported but conflicting relationships were recovered for *X. clivii* with respect to numbered Clades 1 and 2; this relationship is represented as a polytomy. (B) *DM-W* originated after divergence of *Silurana* and *Xenopus*. Evolutionary relationships of *X. laevis DM-W* with respect to homologous portions of *DMRT1* paralogs (α and β) of the tetraploid species *X. laevis*, *X. borealis*, and *X. muelleri*, and the diploid species *S. tropicalis* show that this gene is most closely related to *Xenopus DMRT1*β. *DMRT1* from *Bufo marinus* is used as an outgroup. Nodes with posterior probabilities $\geq$99% are indicated with dots. This analysis fails to resolve whether *DM-W* originated before or after the divergence of ancestors of *X. laevis* and (*X. borealis* + *X. muelleri*).

A complete genome sequence is available for a female *Silurana tropicalis* (Hellsten et al., 2010), which is ~65 million years (MY) diverged from *Xenopus* (Bewick et al.,

2012). But the sex specific region or sex determining gene has not been characterized. Using full and partial *DM-W* sequences, BLAST searches did not recover homologous proteins in the *S. tropicalis* genome (Figure 4.2A: Clade 3). However, using amplified fragment length polymorphisms (AFLPs), Olmstead et al. (2010) identified 22 AFLPs linked to the female-specific region of *S. tropicalis*. In a linkage map developed by Wells et al. (2011), four of these regions were mapped to linkage group 7. Using high throughput sequencing, and PCR and Sanger sequencing Bewick et al. (2013) attempted to identify the boundaries of the sex-specific region and loci. Results from their study suggest that the sex chromosomes of *S. tropicalis* are mostly pseudoautosomal, and thus recombining along most of the chromosome. The genome of *S. tropicalis* does not contain *DM-W*, nor is linkage group 7 syntenic with scaffolds 173120 or 443201 that contain *DM-W* of the *X. laevis* draft genome. This suggests that sex chromosomes have evolved independently since the ancestor of *Xenopus* and *Silurana*. Also, a comparative study by Uno et al. (2008) could not identify *DM-W* in *S. tropicalis* using fluorescent in situ hybridization (FISH).

Turnover of the sex chromosome resets the clock of degeneration and may be accompanied by a change in the heterogamy (XY to ZW or ZW to XY) and/or the dominant sex determining trigger. It has been shown that the load of deleterious mutations that accumulate on the nonrecombining sex chromosomes may be enough to cause a turnover event (Blaser et al., 2012). Given the rate of sex chromosome turnover in amphibians estimated by Evans et al. (2012), it is not unfathomable to suspect that the sex determining mechanism has changed more than once within African clawed frogs and in other frogs. The phylogenetic restriction of *DM-W* to one clade of *Xenopus* suggests that the mechanism of sex determination within *Xenopus* is recently evolved and has changed since the last common ancestor (LCA) of *Xenopus* + *Silurana* ($\sim$65 MYA, (Evans et al., 2004; Roelants and Bossuyt, 2005; Bewick et al., 2012)) and even more recently since the LCA of *X. laevis* + *X. borealis* ($\sim$27 million years ago (MYA), (Evans et al., 2004)). The former implies a loss of *DM-W* and a turnover of sex chromosomes in *X. borealis*. *DM-W* was not identified in dodecaploid species nor in every species of *Xenopus*, so there could have been a more recent turnover event within the last several million years.

In frogs, sex chromosome turnover has also been recorded in Palearctic green toads (*Bufo viridis* subgroup) and Japanese wrinkled frog (*Rana rugosa*). All *Bufo spp.* are female heterogametic (Hillis and Green, 1990); however, Stöck et al. (2011) recorded a heterogametic transition in *B. balearicus* and *B. siculus* that may have occurred in the Oligocene/Early Miocene (>23.8-5.3 MYA) or as early as the Pliocene (5.3-1.8 MYA) – divergence of the *B. viridis* subgroup and *B. balearicus*, respectively (Stöck et al., 2006). *R. rugosa* underwent a heterogametic transition; however both XX/XY and ZZ/ZW sex determining systems share the same ancestral proto sex chromosome (chromosome 7) (Miura et al., 2011). There are examples of closely related species with different sex chromosomes and mechanisms of sex determination, with relationships that are much younger (Tanaka et al., 2007; Ross et al., 2009), which speak for the plausibility of sex chromosome turnover events in frogs.

Recent evolution of sex chromosomes resets the clock of sex chromosome degeneration. Turnover events can be accompanied by a switch in the heterogametic sex (Evans et al., 2012) and/or a switch in the sex determining mechanism (Tanaka et al., 2007). If a turnover event happens after degeneration, there is a potential for a double-null genotype, and if deleterious or lethal it will be lost from the population (Evans et al., 2012) (see below). If these events occur frequently (i.e., before degeneration has occurred), the sex chromosomes will be in a constant, dynamic, homomorphic state. Rates of degeneration differ between species (Graves, 2004; Bachtrog et al., 2008; Evans et al., 2012; Zhou and Bachtrog, 2012), so the rate of turnover to maintain homomorphic sex chromosomes is unknown. Also, sex chromosome turnover may be used as a mechanism to deter the complete loss of the sex-specific sex chromosome (Kuroiwa et al., 2010). In frogs, given the data presented above, it seems plausible that recent turnover events have maintained homomorphic sex chromosomes of many different frog genera.

4.5 Low rates of recombination may be sufficient to maintain homomorphic sex chromosomes in frogs

Frequent turnover of sex chromosomes implies evolutionarily young sex chromosomes that are recombining across almost the entire chromosome, and the effects of suppressed recombination in the form of degeneration have yet to occur. In many disparate animals, recombination is determined by phenotypic, rather than genotypic sex (Inoue et al., 1983; Wallace et al., 1997; Lynn et al., 2005; Campos-Ramos et al., 2009; Matsuba et al., 2010); therefore species with a low incidence of sex-reversed individuals will see a rare but steady rate of crossing over between sex chromosomes. This acts to rejuvenate the sex chromosome (Perrin, 2009) and can result in anciently yet largely homomorphic sex chromosomes (Stöck et al., 2011). However, we cannot rule out the possibility that the sex determining region is small and most of the sex chromosomes are pseudoautosomal (Bhalla and Craig, 1970). The former phenomenon is referred to as the "fountain-of-youth" hypothesis and has been given support in several recent studies of frog species belonging to the genera *Hyla* (Stöck et al., 2011; Guerrero et al., 2012) and *Bufo* (Stöck et al., 2013) and may be occurring in *S. tropicalis* (Bewick et al., 2013).

European tree frogs (*Hyla arborea*, *H. intermedia* and *H. molleri*) are estimated to be ∼5.4-7.1 MY diverged and share the same pair of sex chromosomes with complete absence of X-Y recombination in males estimated by sibship analyses of microsatellite polymorphisms (Stöck et al., 2011). However, sequences of sex-linked loci show no divergence between the X- and Y-chromosomes; gametologs form a clade by species in a phylogeny (Stöck et al., 2011). Using Approximate Bayesian Computation (ABC) and the data generated by Stöck et al. (2011), Guerrero et al. (2012) estimated the rates of recombination between the X- and Y-chromosome to be 105 times smaller than that between

X-chromosomes. This suggests very low rates of recombination (i.e., 1 in every 100,000 individuals) may be sufficient to maintain homomorphism in sex chromosomes. Similarly, in Palearctic green toads (*Bufo viridis* subgroup) with similar divergence time (∼3.3 MY), Stöck et al. (2013) demonstrated that gametologs form a clade by species rather than X- and Y-chromosomes. Interestingly, populations of the *B. virdis* subgroup have demonstrably heteromorphic sex chromosomes, yet it is suggested that recombination is still occurring between the sex chromosomes (Odierna et al., 2007; Stöck et al., 2013). Male heteromorphy in the *B. viridis* subgroup is also surprising, since recombination has been proposed to deter degeneration. The average rate of recombination for sex-linked loci in males (0.02) was slower than the average for females (0.38), but estimated X-Y recombination in *Hyla spp.* is much lower than that of *Bufo spp.* (Guerrero et al., 2012; Stöck et al., 2013). Bewick et al. (2013) showed that the African clawed frog *S. tropicalis* has indistinguishable Z and W sex chromosomes at the molecular level, even in regions linked to sex (Olmstead et al., 2010). Although recombination rates between Z and W chromosomes were not estimated, the homomorphic state of the sex chromosomes suggests recombination may be occurring throughout most of the chromosomes.

The effects of temperature could be related to the "fountain-of-youth" hypothesis and can explain sex reversed individuals within a population (Matsuba et al., 2010). Temperature can alter sex ratios of larvae in frogs, and can alter the phenotypic sex of a developing larvae. For example, when larvae of *Rana sylvatica* are reared at 32°C for up to 33 days, half of the individuals are phenotypic males and the remainder show more or less masculinized ovaries at the end of the treatment (Witschi, 1929). Although the temperature of 32°C for rearing larvae is extremely high, this experiment suggests that the sex of some species of frogs can be determined by temperatures. Similar results have been documented in other frogs (Makoto, 1963; Hsü et al., 1971), salamanders (Uchida, 1937a,b; Chardard et al., 1995; Dournon et al., 1984), and in other animals that are external fertilizers (Baroiller et al., 1999; Crews, 2003; Sato et al., 2005; Barske and Capel, 2008; Abozaid et al., 2011). In the wild, if extreme temperatures are experienced for prolonged periods of time or briefly at thermosensitive stages (Dournon et al., 1990), it may be enough to perturb the genetic sex determination (GSD) and cause sex-reversed individuals. This opens the possibility for mating and recombination of the sex chromosomes in sex-reversed individuals.

Recombination between sex chromosomes does not explicitly test the "fountain-of-youth" hypothesis. The sex determining region (SDR), which includes the sex determining gene, is unknown in the species mentioned above. Therefore, the SDR could be minuscule, making the sex chromosomes mostly pseudoautosomal and thus expected to recombine across most of the chromosome. Species other than frogs possess homomorphic sex chromosomes (e.g., ratites: Ogawa et al. (1998) and pythons: Matsubara et al. (2006)), but the "fountain-of-youth" hypothesis has not been suspected to operate in these species. A more global mechanism to deter degeneration may be operating. It is known that real world populations and environments that organisms inhabit are dynamic, and that selection varies over time and space. Also, restricting recombination lowers the effective population size of

a genomic region, making this region susceptible to deleterious mutations and the efficacy with which natural selection operates, and thus the population is limited in the direction of selection with which it can take. Therefore, stabilizing selection might be acting to maintain alleles at intermediate frequencies within the population and certain combinations of alleles could be brought together through recombination during periods that require rapid evolutionary change (Otto and Barton, 1997). Another possibility is that species with homomorphic sex chromosomes have failed to evolve a mechanism of dosage compensation and has somehow constrained sex chromosome divergence (Adolfsson and Ellegren, 2013). Related to this hypothesis are different mechanisms to resolve sexual conflict, and the consequences of these mechanisms on genome structure. Sexual conflict occurs because the divergent reproductive interests of the sexes generate different selection pressures on many traits (Pischedda and Chippindale, 2006). Sexual conflict resulting from the accumulation of sexually antagonistic mutations that accumulate near the sex determining gene, but within the PAR, can be resolved either by expanding the region of suppressed recombination between the sex chromosomes or by down-regulating the expression of genes in the sex that they harm (consequently creating sex-biased gene expression in the sex that benefits from the mutation) (Vicoso et al., 2013). The former mechanism eventually leads to the evolution of heteromorphic sex chromosomes (Vicoso et al., 2013). However, the latter mechanism eliminates the selective pressure to suppress recombination promoting the retention of homomorphic sex chromosomes (Otto et al., 2011).

4.6 Absence of a global mechanism of dosage compensation may effect sex chromosome evolution

The genomes of males and females are almost identical with the exception of genes on the Y- (or W-) chromosome or sex determining alleles. Dosage compensation equalizes expression levels of genes that have one allele in one sex and two alleles in the other sex. In this way the relative expression of X-linked (or Z-linked) and autosomal genes is constant, or "balanced", in males and females. Dosage compensation has evolved independently at least three times, relying on different mechanisms: upregulation of the single X in *Drosophila* males, inactivation of one X in female mammals, and downregulation of both Xs in *Caenorhabditis elegans* hermaphrodites. Dosage compensation has also been recorded in the plant *Silene latifolia* and the mechanism is similar to *Drosophila*: upregulation of the X-linked allele (Muyle et al., 2012).

In vertebrates, there is no evidence for global dosage compensation outside mammals. It seems absent from sticklebacks (Leder et al., 2010) and in birds compensation varies by tissue and ontogenetically (Mank and Ellegren, 2009), occurring on a gene-by-gene basis, when and where balanced transcription is needed (Mank et al., 2011). A recent study by Adolfsson and Ellegren (2013) showed no evidence for dosage compensation in the ostrich.

Ratites, like the ostrich, have maintained homomorphic sex chromosomes for more than 120 MY (Tsuda et al., 2007). Adolfsson and Ellegren (2013) attributed the lack of evolution of dosage compensation as a constraint on sex chromosome divergence and that dosage compensation should only evolve if the maintenance of ancestral expression levels is vital for interactions with other genes. Therefore, mutations (i.e., chromosomal rearrangements) that reduce recombination between sex chromosomes should be deleterious (Adolfsson and Ellegren, 2013). This means homomorphic sex chromosomes may not elicit dosage compensation.

In amphibians evidence of dosage compensation has not been found (Ohta, 1986; Hayes, 1998; Schmid et al., 1986; Schmid and Steinlein, 2001). Similar to ratites, a lack of a dosage compensation mechanism may have constrained the evolution of sex chromosomes in amphibians. Dosage compensation has evolved within $\sim$10 MY of sex chromosome evolution (Muyle et al., 2012) and pseudogenization of genes on a newly formed sex chromosome can occur with in an extremely short period of time ($\sim$1 MY) (Bachtrog et al., 2008), implying that dosage compensation could arise even earlier. Sex chromosomes in African clawed frogs have remained homomorphic for $>$10 MY, hence enough time has passed for a dosage compensation mechanism to evolve. However, assessment of dosage compensation in amphibians is biased since only one species of frogs (*Buergeria buegeri*) has been tested for one Z-linked gene (Ohta, 1986). Also, this species is female heterogametic and dosage compensation has been suggested to be a generality in this sex chromosome system (Graves and Disteche, 2007; Mank, 2009; Vicoso and Bachtrog, 2009; Zha et al., 2009; Vicoso and Bachtrog, 2011; Pease and Hahn, 2012). A global dosage compensation mechanism is most likely not present in amphibians, but a more generalized; gene-by-gene dosage compensation mechanism is possible. This has yet to be exhaustively tested and we are only now beginning to acquire whole genome information on amphibians. This does not rule out the possibility of other phenomena, which could be at work maintaining sex chromosomes in a homomorphic sate.

4.7 Homomorphic sex chromosomes and the rate of polyploidization

Polyploidization occurs more commonly in plants than in animals and reasons for this trend is a central question in biology (Mable, 2004; Muller, 1925; Orr, 1990). Multiple explanations have been put forward (reviewed in Gregory and Mable, 2005; Mable, 2004; Orr, 1990; Otto and Whitton, 2000). One possibility is that the propensity for a species to undergo polyploidization is related to the extent of sex chromosome degeneration.

Polyploidization might be less common in species with degenerate sex chromosomes because, after duplication, a degenerate ancestral sex chromosome could segregate as a new autosomal chromosome, and the resulting homozygous null genotypes could be detri-

mental (Evans et al., 2012). This is caused by "diploidization" of a polyploid genome. Diploidization refers to the change in inheritance of chromosomes from polysomic, where multivalents are formed during cell division, to disomic, where bivalents are formed (Wolfe, 2001). For example, with a degenerate Y-chromosome the nascent autosomal pair that was previously a pair of sex chromosomes would initially have three possible genotypes: $A_X A_X$, $A_X 0$, and 00 where A_X refers to an autosomal allele derived from an ancestral X-chromosome and 0 refers to a missing allele that was lost on the ancestral Y-chromosome (Evans et al., 2012) (Figure 4.3). If the 00 genotype is deleterious or lethal, there would be reproductive incompatibilities in the early stages of diploidization until the degenerate chromosome is lost (Evans et al., 2012). Also, species with a degenerate sex chromosome could act as a barrier to genome duplication because dosage compensation is disrupted when a newly formed polyploid individual backcrosses with a diploid parental individual during the first stages of polyploid speciation (Orr, 1990). It is widely accepted that increase in selective pressure for dosage compensation occurs as the gene content disparity increases between diverging sex chromosomes (Bergero and Charlesworth, 2009; Charlesworth et al., 2005). Consequently, these two hypotheses are not mutually exclusive. Nevertheless, if sex chromosome degeneration acts as a barrier to polyploidization, this would predict that polyploid species or species with polymorphism in ploidy levels would have minimally degenerate sex chromosomes as compared with other species.

Fifty polyploid frog species have been described, including seven triploids, 30 tetraploids, 11 octoploids, and two dodecaploids derived from 15 families and 20 genera (reviewed in Schmid et al., 2010; Mable et al., 2011; Evans et al., 2012). Three tetraploids and two dodecaploids have been reported from the genus *Xenopus* but not yet formally described as species (Evans, 2007, 2008; Evans et al., 2004, 2005; Tymowska, 1991). Stable triploids are known from three frog genera (*Bufo*, *Eusophus*, and *Rana*), tetraploids from 16 (*Aphantophryne*, *Astylosternus*, *Bufo*, *Chaismocleis*, *Dicroglossus*, *Eleuthrodactylus*, *Hyla*, *Neobatrachus*, *Odontophrynus*, *Phyllomedusa*, *Pleurodema*, *Pyxicephalus*, *Scaphiophryne*, *Silurana*, *Tomopterna*, and *Xenopus*), octoploids from three (*Ceratophrys*, *Pleurodema*, and *Xenopus*), and dodecaploids only from *Xenopus*. Spontaneous or experimentally induced polyploidy has been reported in at least five frog species (Evans et al., 2012).

About one third of the described polyploid frog species belong to the genus *Xenopus*. Why is polyploidization rampant in *Xenopus*? Potentially relevant to the high incidence of polyploidization in *Xenopus* is the recent evolution of *DM-W* (Bewick et al., 2011). Not surprisingly, the sex chromosomes of *Xenopus* are not cytologically distinct (Tymowska, 1991; Yoshimoto et al., 2008). Gene contents of the W- and Z-chromosomes of *Xenopus* are therefore probably very similar, and *Xenopus* species presumably lack mechanisms of dosage compensation operating over most sex-linked genes because both sexes have two alleles at most loci on the sex chromosomes. The preponderance of polyploids in *Xenopus* is therefore consistent with the proposal that polyploidization is more likely to occur in lineages with young, minimally degenerate sex chromosomes.

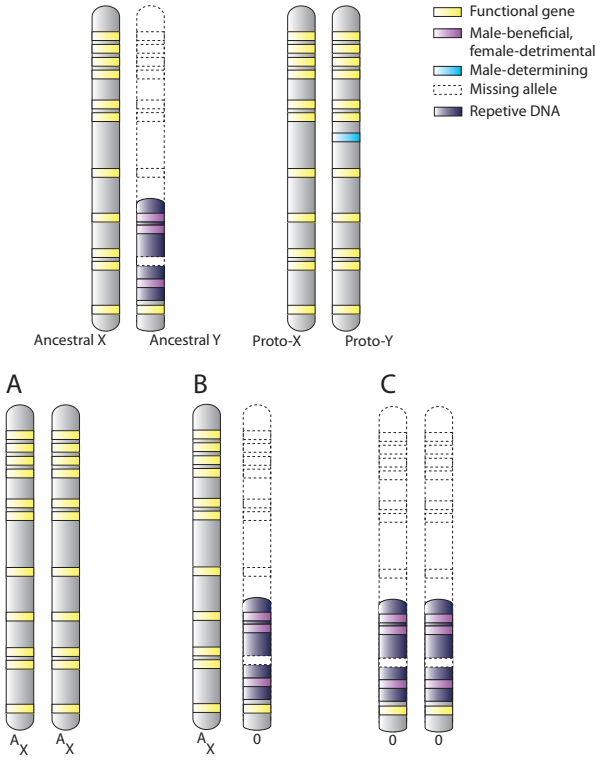

Figure 4.3: Segregation of the degenerate ancestral sex chromosome during diploidization of a polysomic genome. With a degenerate Y-chromosome the nascent autosomal pair that was previously a pair of sex chromosomes would initially have three possible genotypes: (A) $A_X A_X$, (B) $A_X 0$, and (C) 00 where A_X refers to an autosomal allele derived from an ancestral X-chromosome and 0 refers to a missing allele that was lost on the ancestral Y-chromosome. Similar inheritance would be observed if the sex determining gene moved to a different autosome, which now become the proto-X and -Y sex chromosomes in a male heterogametic system.

It is not clear whether novel mechanisms for sex determination are more likely to evolve and persist in species that have nondegenerate sex chromosomes, but this seems plausible under the same reasoning discussed above with respect to the propensity for lineages to experience polyploidization, and evidence put forward by Evans et al. (2012). Also, *Xenopus* provides anecdotal support for a negative negative correlation between the age of sex chromosomes and a species propensity to undergo polyploidization. Another possible link between sex chromosome evolution and polyploidization is provided by *Leiopelma hochstetteri*. This species has intraspecific variation in the presence of a recently evolved univalent W-chromosome that governs sex determination in females (Green, 1988). *L. hochstetteri* is a diploid but also has spontaneous triploidy (that is, polyploidy without speciation; Green et al. (1984)), which suggests tolerance of polyploidy (Evans et al., 2012).

4.8 Conclusions

Sex chromosome evolution theory predicts the gradual divergence ("degeneration") of sex chromosomes over evolutionary time due to suppressed recombination. However, sex chromosome degeneration has not always occurred; there exist evolutionarily old, but not diverged sex chromosomes. Additionally, there exist examples where sex chromosomes are nondegenerate either because of recent turnover of the sex chromosomes or recombination between the sex chromosomes. The question of why some sex chromosomes degenerate and others do not thus has important implications for understanding evolutionary processes at the initial stage of sex chromosome differentiation. Using frogs as a model we reviewed phenomena that cause and maintain homomorphic sex chromosomes.

Recent turnovers of sex chromosomes and recombination between sex chromosomes have been documented in frogs, helping to explain the prevalence of homomorphy in this group of animals. Turnover of the sex chromosomes has been estimated to have occurred $\sim$32 times in amphibians, and recent turnover events have been observed in African clawed frogs and *Bufo*. The documented cases of recombination between sex chromosomes in frogs have been defaulted to the "fountain-of-youth" hypothesis. However, direct testing of this hypothesis needs to be performed, and further research into other mechanisms that promote the retention of homomorphic sex chromosomes need to be explored. Additionally, size homoplasy of microsatellites may interfere with interpretations of recombination between sex chromosomes. Size homoplasy is problematic under high mutation rates and large population sizes together with strong allele size constraints (Estoup et al., 2002). Guerrero et al. (2012) estimated rates of recombination between the X and Y in *Hyla spp* to occur 1 in every 100,000 individuals; it is likely that the population size is large and homoplasy needs to be considered. An alternative to the "fountain-of-youth" hypothesis is recombination between sex chromosomes could be due to the presence of a small sex-specific region (and subsequently a large pseudoautosomal region). Under this alternative hypothesis the sex chromosomes are expected to recombine. Additionally, down-regulating the expression of sexually antagonistic genes in the sex they harm can eliminate the selective pressure of reduced recombination, thus maintaining homomorphic sex chromosomes.

Few vertebrate sex determining genes have been identified: *SRY* in mammals, *Amhr2* in fugu, and *DMRT1* in birds and its homologs in fish and frogs. What other sex determining genes might exist? Recently, it is has been suggested that some autosomes may have a predisposition to become sex chromosomes, possibly due to the presence of conserved sex determining genes or suitable genes capable to capture the dominant position of sex determination (Graves and Peichel, 2010; O'Meally et al., 2012). The transcription factor *DMRT1* is a prime example of a gene involved in sex determination in deeply divergent taxa (Brunner et al., 2001; Matson and Zarkower, 2012; Gamble and Zarkower, 2012). But the role of *DMRT1* in sex determination could be due to shared ancestry or suitableness. Also, little is known about how often heteromorphic sex chromosomes have either evolved convergently from different autosomes or in parallel from the same pair of autosomes, or

how universal patterns of molecular evolution on sex chromosomes really are. Empirical studies including the comparison of sex chromosomes across taxa will provide support for these hypotheses.

Polyploidization in some genera of frogs may be relevant to sex chromosome evolution. Whole genome duplication via allopolyploidization occurred numerous times during the evolution of African clawed frogs and provides support for polyplopdization occurring more frequently in species with homomorphic sex chromosomes. Partial duplication of *DMRT1* gave rise to the female sex determining gene *DM-W* and the interaction of these two transcription factors control gonad development. Paralogous copies of *DMRT1* generated from whole genome duplication may interact differently with *DM-W*. Phylogenetically biased pseudogenization of the *DMRT1*β paralog offers some insights into the functionality of *DMRT1* paralogs, but further work is needed to determine the relative importance of each copy. Duplication may also play a role in more recent events that produce homomorphic sex chromosomes, in addition to maintaining homomorphic sex chromosomes through evolutionary time.

Several hypotheses can explain the lack of sex chromosome degeneration in frogs, and some are more applicable to certain genera of frogs over others. Hypotheses that effect all of frogs (and possibly all vertebrates) include recent turnover of the sex chromosome, recombination between the sex chromosomes and expression regulation to resolve sexual conflict. Since we do not know the sex determining gene(s) in all but one species of frogs, we cannot determine at this time which hypothesis best accounts for the lack of sex chromosome degeneration in frogs. However, evidence exists that support both the recent turnover and fountain-of-youth hypothesis. With the complete genome of *S. tropicalis*, current sequencing of *X. laevis* and the identification of sex-linked genes in *R. rugosa* we will be able to test hypotheses of dosage compensation and resolution of sexual conflict through sex-specific gene regulation in terms of sex chromosome evolution. Precise testing of each of these hypotheses would be of direct benefit to understanding the prevalence of homomorphic sex chromosomes in frogs. Identification of sex determining genes in frogs, and analysis of their evolutionary history and genomic context, is a promising direction for future research that would further our understanding of how sex chromosomes evolve in general and the early processes that altered autosomes during their transition to sex chromosomes. The advent of less expensive and increasingly powerful next generation sequencing technology will only advance the field of sex chromosome evolution further.

Part II

CONCLUSION

The family Pipidae is an evolutionarily old group of frogs found in the Old and New Worlds with a derived morphology that facilitates aquatic life. Phylogenomic analysis recovered strong support for the ((*Xenopus*, *Silurana*)(*Pipa*, *Hymenochirus*)) topology, and geologically calibrated divergence time estimates that are consistent with estimated ages and phylogenetic affinities of many fossils. Therefore, diversification of pipid frog genera *Pipa*, *Hymenochirus*, *Silurana* and *Xenopus* occurred dozens millions of years ago during the breakup of Gondwanaland. The estimation of phylogenetic relationships is an essential component for understanding evolution.

Females and males are almost ubiquitous in vertebrates, however the genomic and genetic mechanisms that govern this evolutionary conserved phenotype vary. Using African clawed frogs (genera *Xenopus* and *Silurana*) I investigated many aspects of sex chromosome evolution including mechanisms that determine sex and the effects of whole genome duplication. Sex determination in *Xenopus* evolved at least once and the mechanism of sex determination is most likely different in other genera of Pipidae and possibly within *Xenopus*. Sex in *Xenopus* is determined by a dominant, female-specific gene called *DM-W*. This gene evolved from a partial duplication event from one paralog of *DMRT1*; an autosomal gene with conserved function in male sex determination. The exact pathway of sex determination is unknown, but *DM-W* and *DMRT1* may act competitively for binding sites. A series of events including biased pseudogenization of the *DMRT1* paralogs, differences in regulation of *DMRT1* through development, and non-neutral evolution of *DM-W* probably led to the infiltration and maintenance of *DM-W* as the dominant female sex determining gene. *DM-W* is the only known sex determining gene in Anura (and in amphibians), but homologous proteins involved in sex determination in other taxa have been identified. The reappearance of certain proteins is interesting because it suggests shared ancestry of sex determining genes or rediscovery by evolution of a few, suitable, genes for sex determination.

Sex chromosomes have evolved independently from different autosomes in plant and animal species. Through evolutionary time many of these sex chromosomes have evolved along a convergent path of restricted recombination and degeneration, and increased genetic specialization. However, some sex chromosomes remain in an ancestral state of autosome-like characteristics – notably, almost fully recombining and nondegenerative. African clawed frogs represent a group of Anura with nondegenerate, or homomorphic, sex chromosomes. In support of this, I demonstrated through next generation sequencing technology and mapping of over 20 million single nucleotide polymorphisms (SNPs) that male and female nucleotide diversity along the 10 major linkage groups of *Silurana tropicalis* is the same. This suggests that sex chromosomes of *S. tropicalis* constitute a large pseudoautosomal region (PAR) and are recombining across most of their length. Homomorphic sex chromosomes in *S. tropicalis* most likely represent a recent turnover event of the sex chromosomes (i.e., the "high-turnover" hypothesis). However, the persistence of homomorphic sex chromosomes could be due to recombination between sex-reversed individuals (i.e., the "fountain-of-youth" hypothesis), or a small sex determining region (large PAR).

Whole genome duplication (WGD) through allopolyploidization has occurred frequently during African clawed frog diversification, providing novel genetic material for natural selection and has drastically changed their genomic architecture. Within a species, the propensity to undergo WGD and sex chromosome evolution are potentially interrelated. For example, WGD might be less common in species with degenerate sex chromosomes because, after duplication, a degenerate ancestral sex chromosome could segregate as a new autosomal chromosome, and the resulting homozygous null genotypes could be detrimental. Sex chromosome degeneration also creates imbalances in allelic copy number between the sexes, which can lead to the evolution of dosage compensation – a factor that is also potentially relevant to genome duplication. Species with a degenerate sex chromosome could act as a barrier to genome duplication, because dosage compensation is disrupted when a newly formed triploid individual backcrosses with a diploid parental individual during the first stages of polyploid speciation. Our analyses suggest that the sex-specific region of *S. tropicalis* is small, that sex chromosome degeneration is minimal, and therefore that dosage compensation would have evolved in very few genes or not at all. These features may have facilitated (or at least not impeded) WGD in *Silurana* and potentially in other African clawed frogs.

Part III

REFERENCES

Bibliography

Abozaid, H., Wessels, S., and Horstgen-Schwark, G. (2011). Effect of reating temperatures during embryonic development on the phenotypic sex in zebrafish (*Danio rerio*). *Sexual Development*, 76:1879–1902.

Abramyan, J., Ezaz, T., Graves, J. A. M., and Koopman, P. (2009). Z and W sex chromosomes in the cane toad (*Bufo marinus*). *Chromosome Research*, 17:1015–1024.

Adolfsson, S. and Ellegren, H. (2013). Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. *Molecular Biology and Evolution*, Epub ahead of print.

Agulnik, A. I., Bishop, C. E., Lemer, J. L., Agulnik, S. I., and Solovyev, V. V. (1997). Analysis of mutation rates in the *SMCY/SMCX* genes shows that mammalian evolution is male driven. *Mammalian Genome*, 8(2):134–138.

Al-Shehri, A. H. and Al-Saleh, A. A. (2005). Karyotype of amphibians in Saudi Arabia 1: The karyotype of *Rana ridibunda*. *Journal of Biological Sciences*, 5(3):335–338.

Ali, J. R. and Aitchison, J. C. (2008). Gondwana to Asia: Plate tectonics, paleogeography and the biological conectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166-35 Ma). *Earth-Science Reviews*, 88(3-4):145–166.

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.

Ananias, F., Álvaro Dhimas S. Modesto, Mendes, S. C., and Napoli, M. F. (2007). Unusual primitive heteromorphic ZZ/ZW sex chromosomes in *Proceratophrys boiei* (Anura, Cycloramphidae, Alsodinae), with description of C-Band interpopulational polymorphism. *Hereditas*, 144:206–212.

Anderson, D. W. and Evans, B. J. (2009). Regulatory evolution of a duplicated heterodimer across species and tissues of allopolyploid clawed frogs (*Xenopus*). *Journal of Molecular Evolution*, 68:236–247.

Anderson, J. L., Marí, A. R., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., and Postlethwait, J. H. (2012). Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One*, 7(7):e40701.

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

Bachtrog, D. and Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature*, 416:323–326.

Bachtrog, D., Hom, E., Wong, K. M., Maside, X., and de Jong, P. (2008). Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biology*, doi: 10.1186/gb-2008-9-2-r30.

Báez, A. M. and Harrison, T. (2005). A new Pipinae frog from an Eocene crater lake in north-central Tanzania. *Palaeontology*, 48(4):723–737.

Báez, A. M. and Pugener, L. A. (2003). Ontogeny of a new Paleogene pipid frog from southern South America and xenopodinomorph evolution. *Zoological Journal of the Linnean Society*, 139:439–476.

Báez, A. M. and Rage, J.-C. (1998). Pipid frogs from the upper Cretaceous of In Beceten, Niger. *Palaeontology*, 41:669–691.

Báez, A. M., Trueb, L., and Calvo, J. O. (2000). The earliest known pipoid frog from South America: A new genus from the middle Cretaceous of Argentina. *Journal of Vertebrate Paleontology*, 20(3):490–500.

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

Baroiller, J. F., Guiguen, Y., and Fostier, A. (1999). Endocrine and environmental aspects of sex differentiation in fish. *Cellular and Molecular Life Sciences*, 55(7):910–931.

Barske, L. A. and Capel, B. (2008). Blurring the edges in vertebrate sex determination. *Current Opinion in Genetics and Development*, 18(6):499–505.

Baxter, S. W., Davey, J. W., Johnston, J. S., Shelton, A. M., Heckel, D. G., Jiggins, C. D., and Blaxter, M. L. (2011). Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One*, 6(4):e19315.

Bergero, R. and Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology and Evolution*, 24(2):94–102.

Berlin, S. and Ellegren, H. (2006). Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *Journal of Molecular Evolution*, 62(1):66–72.

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

Bewick, A. J., Chain, F. J. J., Heled, J., and Evans, B. J. (2012). The pipid root. *Systematic Biology*, 61(6):913–926.

Bewick, A. J., Chain, F. J. J., Zimmerman, L., Sesay, A., Gilchrist, M., Owens, N., Seifertova, E., Krylov, V., Macha, J., Tlapakova, T., Kubickova, S., Cernohorska, H., Zarsky, V., and Evans, B. J. (2013). A large pseudoautomsal region on the sex chromosomes of the frog *Silurana tropicalis*. *Genome Biology and Evolution*, submitted.

Bhalla, S. C. and Craig, G. B. (1970). Linkage analysis of chromosome I of *Aedes aegypti*. *Canadian Journal of Genetics and Cytology*, 12(3):425–435.

Bisbee, C. A., Baker, M. A., Wilson, A. C., Haji-Azimi, I., and Fischberg, M. (1977). Albumin phylogeny for clawed frogs (*Xenopus*). *Science*, 195(4280):785–787.

Blaser, O., Grossen, C., Neuenschwander, S., and Perrin, N. (2012). Sex-chromosome turnovers induced by deleterious mutation load. *Evolution*, 66(10):3155–3166.

Bowes, J. B., Snyder, K. A., Segerdell, E., Gibb, R., Jarabek, C., Noumen, E., Pollet, N., and Vize, P. D. (2009). Xenbase: A *Xenopus* biology and genomics resource. *Nucleic Acids Research*, 36:D761–72.

Brooks, D. R. and McLennan, D. A. (2002). *The nature of diversity*. Chicago (IL): Chicago University Press.

Brown, W. M., Prager, E. M., Wang, A., and Wilson, A. C. (1982). Mitochondrial DNA sequences of primates: Tempo and mode of evolution. *Journal of Molecular Evolution*, 18:225–239.

Brunner, B., Hornungb, U., Shana, Z., Nandac, I., Kondod, M., Zend-Ajuscha, E., Haafa, T., Ropersa, H.-H., Shimad, A., Schmidc, M., Kalscheuera, V. M., and Schartl, M. (2001). Genomic organization and expression of the *doublesex*-related gene cluster in vertebrates and detection of putative regulatory regions for *DMRT1*. *Genomics*, 77(1-2):8–17.

Bull, J. J. (1983). *Evolution of sex determining mechanisms*. Benjamin Cummings, Menlo Park.

Burtis, K. C. and Baker, B. S. (1989). *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mrnas encoding related sex-specific polypeptides. *Cell*, 56(6):997–1010.

Busin, C. S., Andrade, G. V., Bertoldo, J., Grande, M. L. D., Uetanabaro, M., and Recco-Pimentel, S. M. (2008). Cytogenetic analysis of four species of *Pseudis* (Anura, Hylidae), with the description of ZZ/ZW sex chromosomes in *P. tocantins*. *Genetica*, 133:119–127.

Campos-Ramos, R., Harvey, S. C., and Penman, D. J. (2009). Sex-specific differences in the synaptonemal complex in the genus *Oreochromis* (Cichlidae). *Genetica*, 135:325–332.

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

Cannatella, D. C. and Trueb, L. (1988a). Evolution of pipoid frogs: Intergeneric relationships of the aquatic frog family Pipidae (Anura). *Zoological Journal of the Linnean Society*, 94(1):1–38.

Cannatella, D. C. and Trueb, L. (1988b). Evolution of pipoid frogs: Morphology and phylogenetic relationships of *Pseudhymenochirus*. *Journal of Herpetology*, 22(4):439–456.

Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Müller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002). The Comparative RNA Web (CRW) Site: An online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, 3:2.

Chain, F. J. J., Dushoff, J., and Evans, B. J. (2011). The odds of duplicate gene persistence after polyploidization. *BMC Genomics*, 12:599.

Chain, F. J. J. and Evans, B. J. (2006). Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. *PLoS Genetics*, 2(4):e56.

Chain, F. J. J., Ilieva, D., and Evans, B. J. (2008). Duplicate gene evolution and expression in the wake of vertebrate allopolyploidization. *BMC Evolutionary Biology*, 8:43.

Chakrabarti, S., Banerjee, S. N., Neogi, L. N., and Roy-Choudhuri, S. (1983). C-band positive W chromosome in the female Indian frog. *Experientia*, 39(3):321–322.

Chardard, D., Desvages, G., Pieau, C., and Dournon, C. (1995). Aromatase activity in larval gonads of *Pleurodeles watl* (urodele amphibia) during normal sex differentiation and during sex reversal by thermal treatment effect. *General and Comparative Endocrinology*, 99(1):100–107.

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science*, 251:1030–1033.

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Current Biology*, 6(2):149–162.

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

Charlesworth, D., Charlesworth, B., and Marais, G. (2005). Steps in the evolution of heteromorphic sex chromosomes. *Heredity*, 95(2):118–28.

Charlesworth, D. and Mank, J. E. (2010). The birds and the bees and the flowers and the trees: Lessons from genetic mapping of sex determination in plants and animals. *Genetics*, 186(1):9–31.

Crews, D. (2003). Sex determination: Where environment and genetics meet. *Evolution and Development*, 5(1):50–55.

Cuevas, C. C. and Formas, J. R. (1996). Heteromorphic sex chromosomes in *Eupsophus insularis* (amphibia: Anura: Leptodactylidae). *Chromosome Research*, 4:467–470.

de Almeida, C. G., Grafe, T. U., Guttenbach, M., and Schmid, M. (1990). Karyotype and chromosome banding in the reed frog *Hyperolius viridiflavus ommatostictus* (Amphibia, Anura, Hyperoliidae). *Experientia*, 46(5):509–511.

Degnan, J. H. and Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genetics*, 2(5):e68.

Degnan, J. H. and Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology and Evolution*, 24(6):332–340.

Degnan, J. H. and Salter, L. A. (2005). Gene tree distributions under the coalescent process. *Evolution*, 59(1):24–37.

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., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43:491–498.

Dournon, C., Guillet, F., Boucher, D., and Lacroix, J. C. (1984). Cytogenetic and genetic evidence of male sexual inversion by heat treatment in the newt *Pleurodeles poireti*. *Chromosoma*, 90:261–264.

Dournon, C., Houillon, C., and Pieau, C. (1990). Temperature sex-reversal in amphibians and reptiles. *The International Journal of Developmental Biology*, 34(1):81–92.

Drummond, A. J. and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7:214.

Duarte, T. C., Veiga-Menoncello, A. C. P., Lima, J. R. F., Strüssmann, C., Grande, M. L. D., Giaretta, A. A., Pereira, E. G., Rossa-Feres, D. C., and Recco-Pimentel, S. M. (2010). Chromosome analysis in *Pseudopaludicola* (Anura, Leiuperidae), with description of sex chromosomes XX/XY in *P. saltica*. *Hereditas*, 147:43–52.

Duellman, W. E. and Trueb, L. (1994). *Biology of amphibians*. Baltimore (MD): The Johns Hopkins University Press.

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5:113.

Engel, W. and Schmid, M. (1981). H-Y antigen as a tool for the determination of the heterogametic sex in amphibia. *Cytogenetics and Cell Genetics*, 30(3):1306–136.

Engen, S., Ringsby, T. H., ther, B.-E. S., Lande, R., Jensen, H., rd, M. L., and Ellegren, H. (2007). Effective size of fluctuating populations with two sexes and overlapping generations. *Evolution*, 61(8):1873–1885.

Estes, R. (1977). Relationships of the South African fossil frog *Eoxenopoides reuningi* (Anura, Pipidae). *Annals of South African Museum*, 73:49–80.

Estes, R., Spinar, Z. V., and Nevo, E. (1978). Early Cretaceous pipid tadpoles from Israel (amphibia: Anura). *Herpetologica*, 34(4):374–393.

Estoup, A., Jarne, P., and Cornuet, J.-M. (2002). Invited review: Homoplasy and mutation model at microsatellite loci and their consequences for population genetic analysis. *Molecular Ecology*, 11:1591–1604.

Evans, B. J. (2007). Ancestry influences the fate of duplicated genes millions of years after duplication in allopolyploid clawed frogs (*Xenopus*). *Genetics*, 176(2):1119–1130.

Evans, B. J. (2008). Genome evolution and speciation genetics of clawed frogs (*Xenopus* and *Silurana*). *Frontiers in Bioscience*, 13:4697–4706.

Evans, B. J., Bliss, S. M., Mendel, S. A., and Tinsley, R. C. (2011a). The Rift Valley is a major barrier to dispersal of African clawed frogs (*Xenopus*) in Ethiopia. *Molecular Ecology*, 20(20):4216–4230.

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 Asian-Australian faunal zone interface. *Systematic Biology*, 52(6):794–819.

Evans, B. J., Greenbaum, E., Kusamba, C., Carter, T. F., Tobias, M. L., Mendel, S. A., and Kelley, D. B. (2011b). 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., Morales, J. C., Picker, M. D., Kelley, D. B., and Melnick, D. J. (1997). Comparative molecular phylogeography of two *Xenopus* species, *X. gilli* and *X. laevis*, in the south-western Cape Province, South Africa. *Molecular Ecology*, 6(4):333–343.

Evans, B. J., Morales, J. C., Picker, M. D., Melnick, D. J., and Kelley, D. B. (1998). Absence of extensive introgression between *Xenopus gilli* and *Xenopus laevis laevis* (Anura: Pipidae), in southwestern Cape Province, South Africa. *Copeia*, 1998(2):504–509.

Evans, B. J., Pyron, R. A., and Wiens, J. J. (2012). *Polyploidy and Genome Evolution*, chapter Polyploidization and sex chromosome evolution in amphibians, pages 385–410. Springer, Berlin.

Ewing, B. and Green, P. (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, 8:186–194.

Filatov, D. A. and Charlesworth, D. (2002). Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Molecular Biology and Evolution*, 19(6):898–907.

Filatov, D. A., Laporte, V., Vitte, C., and Charlesworth, D. (2001). DNA diversity in sex-linked and autosomal genes of the plant species *Silene latifolia* and *Silene dioica*. *Molecular Biology and Evolution*, 18(8):1442–1454.

Ford, L. S. and Cannatella, D. C. (1993). The major clades of frogs. *Herpetological Monographs*, 7:94–117.

Fridolfsson, A.-K., Cheng, H., Copeland, N. G., Jenkins, N. A., Liu, H.-C., Raudsepp, T., Woodage, T., Chowdhary, B., Halverson, J., and Ellegren, H. (1998). Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proceedings of National Academy of Sciences of the United States of America*, 95(14):8147–8152.

Fridolfsson, A.-K. and Ellegren, H. (2000). Molecular evolution of the avian *CHD1* genes on the Z and W sex chromosomes. *Genetics*, 155:1903–1912.

Frost, D. R. (2011). Amphibian species of the world: An online reference. Version 5.5. New York: American Museum of Natural History.

Frost, D. R., Grant, T., Faivovich, J., Bain, R. H., Haas, A., Haddad, C. F. B., de Sá, R. O., Cahnning, A., Wilkinson, M., Donnellan, S. C., Raxworthy, C. J., Campbe;;, J. A.,

Blotto, B. L., Moler, P., Drewes, R. C., Nussbaum, R. A., Lynch, J. D., Green, D. M., and Wheeler, W. C. (2006). The amphibian tree of life. *Bulletin of the American Museum of Natural History*, 297:1–291.

Gamble, T. and Zarkower, D. (2012). Sex determination. *Current Biology*, 22:R257–R252.

Gheerbrant, E. and Rage, J.-C. (2006). Paleobiogeography of Africa: How distinct from Gondwana and Laurasia? *Palaeogeography, Palaeoclimatology, and Palaeoecology*, 241(2):224–246.

Graf, J.-D. (1989). Genetic mapping in *Xenopus laevis*: Eight linkage groups established. *Genetics*, 123:389–398.

Graf, J. D. (1996). Molecular approaches to the phylogeny of *Xenopus*.

Graves, J. A. M. (2004). The degenerate Y chromosome - can conversion save it? *Reproduction, Fertility and Development*, 16(5):527–534.

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 Disteche, C. M. (2007). Does gene dosage really matter? *Journal of Biology*, 6(1):doi:10.1186/jbiol55.

Graves, J. A. M. and Peichel, C. L. (2010). Are homologies in vertebrate sex determinination due to shared ancestry or to limited options? *Genome Biology*, 11(205):doi:10.1186/gb–2010–11–4–205.

Green, D. M. (1988). Heteromorphic sex chromosomes in the rare and primitive frog *Leiopelma hamiltoni* from New Zealand. *Journal of Heredity*, 79:165–169.

Green, D. M. (2002). Chromosome polymorphism in Archey's frog (*Leiopelma archeyi*) from New Zealand. *Copeia*, 1:204–207.

Green, D. M., Kezer, J., and Nussbaum, R. A. (1984). Triploidy in Hochstetter's frog, *Leiopelma hochstetteri*, from New Zealand. *New Zealand Journal of Zoology*, 11:457–460.

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.

Gregory, T. R., editor (2005). *The Evolution of the Genome*, chapter Polyploidy in animals, pages 428–517. Elsevier Academic Press, Burlington.

Grützer, F., Rens, W., Tsend-Ayush, E., El-Mogharbel, N., O'Brien, P. C. M., Jones, R. C., Ferguson-Smith, M. A., and Graves, J. A. M. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosome. *Nature*, 432:913–917.

Guerrero, R. F., Kirkpatrick, M., and Perrin, N. (2012). Cryptic recombination in the every-oung sex chromosomes of hylid frogs. *Journal of Evolutionary Biology*, 25(10):1947–1954.

Haag, E. S. and Doty, A. V. (2005). Sex determination across evolution: Connecting the dots. *PLoS Biology*, 3(1):e21.

Hanada, H. (2002). G and C banding show structural differences between Z and W chromosomes in the frog *Buergeria buergeri*. *Hereditas*, 136:151–154.

Hayes, T. B. (1998). Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. *Journal of Experimental Zoology*, 281(5):373–399.

Heinicke, M. P., Duellman, W. E., and Hedges, S. B. (2007). Major Caribbean and Central American frog faunas originated by ancient oceanic dispersal. *Proceedings of the National Academy of Sciences*, 104(24):10092–10097.

Heled, J. and Drummond, A. J. (2010). Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution*, 27(3):570–580.

Heled, J. and Drummond, A. J. (2011). Calibrated tree priors for relaxed phylogenetics and divergence time estimation. *Systematic Biology*, 61(1):138–149.

Hellsten, U., Harland, R. M., Gilchrist, M. J., Hendrix, D., Jurka, J., Kapitonov, V., Ovcharenko, I., Putnam, N. H., Shu, S., Taher, L., Blitz, I. L., Blumberg, B., Dichmann, D. S., Dubchak, I., Amaya, E., Detter, J. C., Fletcher, R., Gerhard, D. S., Goodstein, D., Graves, T., Grigoriev, I. V., Grimwood, J., Kawashima, T., Lindquist, E., Lucas, S. M., Mead, P. E., Mitros, T., Ogino, H., Ohta, Y., Poliakov, A. V., Pollet, N., Robert, J., Salamov, A., Sater, A. K., Schmutz, J., Terry, A., Vize, P. D., Warren, W. C., Wells, D., Wills, A., Wilson, R. K., Zimmerman, L. B., Zorn, A. M., Grainger, R., Grammer, T., Khokha, M. K., Richardson, P. M., and Rokhsar, D. S. (2010). The genome of the Western clawed frog *Xenopus tropicalis*. *Science*, 328:633–636.

Hellsten, U., Khokha, M. K., Grammer, T. C., Harland, R. M., Richardson, P., and Rokhsar, D. S. (2007). Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biology*, 5:37.

Henrici, A. C. (1998). A new pipoid anuran from the Late Jurassic Morrison Formation at Dinosaur National Monument, Utah. *Journal of Vertebrate Paleontology*, 18(2):321–332.

Henrici, A. C. and Báez, A. M. (2001). First occurrence of *Xenopus* (Anura: Pipidae) on the Arabian Peninsula: A new species from the Late Oligocene of the Republic of Yemen. *Journal of Paleontology*, 75(4):870–882.

Hillis, D. M. and Green, D. M. (1990). Evolutionary changes of heterogametic sex in the phylogenetic history of amphibians. *Journal of Evolutionary Biology*, 3:49–64.

Hillis, D. M., Huelsenbeck, J. P., and Cunningham, C. W. (1994). Application and accuracy of molecular phylogenies. *Science*, 264(5159):671–677.

Hsü, C.-Y., Yü, N.-W., and Liang, H.-M. (1971). Induction of sex reversal in female tadpoles of *Rana catesbeiana* by temperature treatment. *Endocrinologia Japonica*, 18(3):243–251.

Huang, H. and Knowles, L. L. (2009). What is the danger of the anomaly zone for empirical phylogenetics? *Systematic Biology*, 58(5):527–536.

Hudson, R. R., Kreitman, M., and Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116(1):153–159.

Huelsenbeck, J. P. and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755.

Inoue, H., Fukumori, Y., and Hiroyoshi, T. (1983). Mapping of autosomal male-determining factors of the housefly, *Musca domestica* L., by means of sex-reversal. *Idengaku Zasshi*, 58(5):451–461.

Ioerger, T. R., Clark, A. G., and Kao, T.-H. (1990). Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences*, 87:9732–9735.

Irisarri, I., Vences, M., Mauro, D. S., Glaw, F., and Zardoya, R. (2011). Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. *BMC Evolutionary Biology*, 11:114.

Iturra, E. and Veloso, A. (1989). Further evidence for early sex chromosome differentiation of Anura species. *Genetica*, 25(1):25–31.

Iturra, P. and Veloso, A. (1981). Evidence for heteromorphic sex chromosomes in male amphibians (Anura: Leptodactylidae). *Cytogenetics and cell genetics*, 31(2):108–110.

Just, W., Rau, W., Vogel, W., Akhverdian, M., Fredga, K., Graves, J. A. M., and Lyapunova, E. (1995). Absence of Sry in species of the vole *Ellobius*. *Nature Genetics*, 11:117–118.

Kaiser, V. B. (2010). Gene loss on Y chromosomes: A consequence of purifying selection? *Fly*, 4(4):270–272.

Kaiser, V. B. and Charlesworth, B. (2010). Muller's Ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. *Genetics*, 185(1):339–348.

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., Tohari, S., Brenner, S., Miyadai, T., Venkatesh, B., Suzuki, Y., and Kikuchi, K. (2012). A trans-species missense SNP in amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genetics*, 8(7):e1002798.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kluge, A. G. and Farris, J. S. (1969). Quantitative phyletics and evolution of the Anurans. *Systematic Biology*, 18(1):1–32.

Kobel, H. R. (1985). Sex determination in polyploid *Xenopus*. *South African Journal of Science*, 81:205–206.

Kobel, H. R. (1996). *The Biology of Xenopus*, chapter Allopolyploid speciation, pages 391–401. Clarendon Press, Oxford.

Kobel, H. R., Barandun, B., and Thiebaud, C. H. (1998). Mitochondrial rDNA phylogeny in *Xenopus*. *Herpetological Journal*, 8(1):13–17.

Kobel, H. R. and Pasquier, L. D. (1986). Genetics of polyploid *Xenopus*. *Trends in Genetics*, 2:310–315.

Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., Asakawa, S., Hori, H., Schmid, M., Shimizu, N., and Schartl, M. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Research*, 16(7):815–826.

Kondo, M., Nanda, I., Hornung, U., Asakawa, S., Shimizu, N., Mitani, H., Schmid, M., Shima, A., and Schartl, M. (2003). Absence of the candidate male sex-determining gene *dmrt1b(Y)* of medaka from other fish species. *Current Biology*, 13(5):416–420.

Kondo, M., Nanda, I., Hornung, U., Schmid, M., and Schartl, M. (2004). Evolutionary origin of the medaka Y chromosome. *Current Biology*, 14(18):1664–1669.

Korte, C., Hesselbo, S. P., Jenkyns, H. C., Rickaby, R. E., and Spötl, C. (2009). Palaeoenvironmental significance of carbon- and oxygen-isotope stratigraphy of marine Triassic Jurassic boundary sections in SW Britain. *Journal of the Geological Society*, 166:431–445.

Kozielska, M., Pen, I., Beukeboom, L. W., and Weissing, F. J. (2996). Sex ratio selection and multi-factorial sex determination in the housefly: A dynamic model. *Journal of Evolutionary Biology*, 19(3):879–888.

Kubatko, L. S. and Degnan, J. H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology*, 56(1):17–24.

Kuroiwa, A., Ishiguchi, Y., Yamada, F., Shintaro, A., and Matsuda, Y. (2010). The process of a Y-loss event in an XO/XO mammal, the Ryukyu spiny rat. *Chromosoma*, 119(5):519–526.

Lande, R. and Barrowclough, G. F. (1987). *Viable populations for conservation*, chapter Effective population size, genetic Effective population size, genetic variation, and their use in population management, pages 87–124. Cambridge University Press, Cambridge.

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., and Hartl, D. L. (2005). Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics*, 171:1813–1822.

Leder, E. H., Cano, J. M., Leinonen, T., O'Hara, R. B., Nikinmaa, M., Primmer, C. R., and Merilä, J. (2010). Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. *Molecular Biology and Evolution*, 27(7):1495–1503.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map (SAM) format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Liu, L. and Pearl, D. K. (2007). Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Systematic Biology*, 56(3):504–514.

Lourenço, L. B., Recco-Pimentel, S. M., and ao J. Cardoso, A. (1999). Two karyotypes and heteromorphic sex chromosomes in *Physalaemus petersi* (Anura, Leptodactylidae). *Canadian Journal of Zoology*, 77:624–631.

Lynch, J. D. (1973). The transition from archaic to advanced frogs.

Lynch, M. and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science*, 290:1151–1155.

Lynn, A., Schrump, S., Cherry, J., Hassold, T., and Hunt, P. (2005). Sex, not genotype, determines recombination levels in mice. *The American Journal of Human Genetics*, 77:670–675.

Mable, B. K. (2004). 'Why polyploidy is rarer in animals than in plants': Myths and mechanisms. *Biological Jounal of the Linnean Society*, 82(4):453–466.

Mable, B. K., Alexandrou, M. A., and Taylor, M. I. (2011). Genome duplication in amphibians and fish: An extended synthesis. *Journal of Zoology*, 284(3):151–182.

Maddison, D. R. and Maddison, W. P. (2000). MacClade 4: Analysis of phylogeny and character evolution. Version 4.0. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3):523–536.

Mahony, M. J. (1991). Heteromorphic sex chromosomes in the Australian frog *Crinia bilingua* (Anura: Myobatrachidae). *Genome*, 34:334–337.

Maisey, J. G. (2000). Continental break up and the distribution of fishes of Western Gondwana during the Early Cretaceous. *Cretaceous Research*, 21(2-3):281–314.

Makoto, Y. (1963). Influence of high temperature on the development of gonads of thiourea-treated frog tadpoles. *Kumamoto Journal of Science Ser. B. Sect. 2, Biology*, 6(2):79–101.

Mank, J. E. (2009). The W, X, Y and Z of sex-chromosome dosage compensation. *Trends in Genetics*, 25(5):226–233.

Mank, J. E. and Ellegren, H. (2009). Are sex-biased genes more dispensable? *Biology Letters*, doi:10.1098/rsbl.2008.0732.

Mank, J. E., Hosken, D. J., and Wedell, N. (2011). Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. *Evolution*, 65(8):2133–2144.

Matson, C. K. and Zarkower, D. (2012). Sex and the singular DM domain: Insights into sexual regulation, evolution and plasticity. *Nature Reviews Genetics*, 13(3):163–174.

Matsuba, C., Alho, J. S., and Merilä, J. (2010). Recombination rate between sex chromosomes depends on phenotypic sex in the common frog. *Evolution*, 64:3634–3637.

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. *PNAS*, 103(48):18190–18195.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C. E., Shibata, N., Asakawa, S., Shimizu, N., Horik, H., Hamaguchi, S., and Sakaizumi, M. (2002). *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature*, 417:559–563.

Matsuda, M., Shinomiya, A., Kinoshita, M., Suzuki, A., Kobayashi, T., Paul-Prasanth, B., lieng Lau, E., Hamaguchi, S., Sakaizumi, M., and Nagahama, Y. (2007). *DMY* gene induces male development in gentically female (XX) medaka fish. *Proceedings of the National Academy of Sciences*, 104(10):3865–3870.

McHone, J. G. (2000). Non-plume magmatism and rifting during the opening of the central Atlantic Ocean. *Tectonophysics*, 316(3-4):287–296.

McHone, J. G. and Butler, J. R. (1984). Mesozoic igneous provinces of New England and the opening of the North Atlantic Ocean. *Geological Society of America Bulletin*, 95(7):757–765.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20:1297–1303.

McLoughlin, S. (2001). The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany*, 49(3):371–300.

Measey, G. J. and Channing, A. (2003). Phylogeography of the genus *Xenopus* in southern Africa. *Amphibia-Reptilia*, 24(3):321–330.

Mikamo, K. and Witschi, E. (1963). Functional sex-reversal in genetic females of *Xenopus laevis*, induced by implanted testes. *Genetics*, 48:1411–1421.

Mikamo, K. and Witschi, E. (1964). Masculinization and breeding the WW in *Xenopus*. *Experientia*, 20:622–623.

Miller, S. W., Hayward, D. C., Bunch, T. A., Miller, D. J., Ball, E. E., Bardwell, V. J., Zarkower, D., and Brower, D. L. (2003). A DM domain protein from a coral, *Acropora millepora*, homologous to proteins important for sex determination. *Evolution and Development*, 5(3):251–258.

Miura, I. (1994). Sex chromosome differentiation in the Japanese brown frog, *Rana japonica*. I. Sex-related heteromorphism of the distribution pattern of constitutive heterochromatin in chromosome no. 4 of the Wakuya population. *Zoological Science*, 11:797–806.

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.

Miura, I., Ohtami, H., Nakamura, M., Ichikawa, Y., and Saitoh, K. (1998). The origin and differentiation of the heteromorphic sex chromosomes Z, W, X, and Y in the frog *Rana rugosa*, inferred from the sequences of a sex-linked gene, ADP/ATP translocase. *Molecular Biology and Evolution*, 15(12):1612–1619.

Miura, I., Ohtani, H., and Ogata, M. (2011). Independent degeneration of W and Y sex chromosomes in frog *Rana rugosa*. *Chromosome Research*, doi:10.1007/s10577-011-9258-8.

Morescalchi, A. (1964). Ii corredo cromosomico di *Discoglossus pictus* Otth.: Cromosomi sessuali, spiralizzazione cromosomica e zone eterocromatiche. *Caryologia (Firenze)*, 17:327–345.

Morin, R. D., Chang, E., Petrescu, A., Liao, N., Griffith, M., Kirkpatrick, R., Butterfield, Y. S., Young, A. C., Stott, J., Barber, S., Babakaiff, R., Dickson, M. C., Matsuo, C., Wong, D., Yang, G. S., Smailus, D. E., Wetherby, K. D., Kwong, P. N., Grimwood, J., III, C. P. B., Brown-John, M., Reddix-Dugue, N. D., Mayo, M., Schmutz, J., Beland, J., Park, M., Gibson, S., Olson, T., Bouffard, G. G., Tsai, M., Featherstone, R., Chan, S., Siddiqui, A. S., Jang, W., Lee, E., Klein, S. L., Blakesley, R. W., Zeeberg, B. R., Narasimhan, S., Weinstein, J. N., Pennacchio, C. P., Myers, R. M., Green, E. D., Wagner, L., Gerhard, D. S., Marra, M. A., Jones, S. J., and Holt, R. A. (2006). Sequencing and analysis of 10,967 full-length cDNA clones from *Xenopus laevis* and *Xenopus tropicalis* reveals post-tetraploidization transcriptome remodeling. *Genome Research*, 16:796–803.

Muller, H. J. (1925). Why polyploidy is rarer in animals than in plants. *The American Naturalist*, 59(663):346–353.

Muyle, A., Zemp, N., Deschamps, C., Mousset, S., Widmer, A., and Marais, G. A. B. (2012). Rapid *de novo* evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biology*, 10(4):e1001308.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., Schmid, M., and Schartl, M. (2002). A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proceedings of the National Academy of Sciences*, 99(18):11778–11783.

Nascimento, J., Quinderé, Y. R. S. D., Recco-Pimentel, S. M., Lima, J. R. F., and Lourenço, L. B. (2010). Heteromorphic Z and W sex chromosomes in *Physalaemus ephippifer* (Steindachner, 1864) (Anura, Leiuperidae). *Genetica*, 138:1127–1132.

Nei, M. and Takahata, N. (1993). Effective population size, genetic diversity and coalescence time in subdivided populations. *Journal of Molecular Evolution*, 37:240–244.

Nevo, E. (1969). Pipid frogs from the Early Cretaceous of Israel and pipid evolution. *Bulletin of the Museum of Comparative Zoology*, 136:255–218.

Nylander, J. A. A. (2004). MrModeltest v2. Program distributed by the author. *Evolutionary Biology Centre, Uppsala University*.

Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P., and Nieves-Aldrey, J. L. (2004). Bayesian phylogenetic analysis of combined data. *Systematic Biology*, 53(1):47–67.

Odierna, G., Aprea, G., Capriglione, T., Castellano, S., and Balletto, E. (2007). Cytological evidence for population-specific sex chromosome heteromorphism in Palaearctic green toads (amphibia, Anura). *Journal of Biosciences*, 32(4):763–768.

Ogawa, A., Murata, K., and Mizuno, S. (1998). The loaction of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *PNAS*, 95:4415–4418.

Ohno, S. (1967). *Sex chromosomes and sex-linked genes*. Springer, Berlin.

Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag.

Ohta, S. (1986). Sex determining mechanism in *Buergeria buergeri* (Schlegel). I: Heterozygosity of chromosome pair no 7 in the female. *Scientific report of the Laboratory for Amphibian Biology, Hiroshima University*, 8:29–43.

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*, 98:60–66.

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*, 20:7–19.

Orr, H. A. (1990). "Why polyploidy is rarer in animals than in plants" revisted. *The American Naturalist*, 136(6):759–770.

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

Osawa, N., Oshima, Y., and Nakamura, M. (2005). Molecular cloning of *Dmrt1* and its expression in the gonad of *Xenopus*. *Zoological Science*, 22(6):681–687.

Otto, S. P. and Barton, N. H. (1997). The evolution of recombination: Removing the limits to natural selection. *Genetics*, 147:879–906.

Otto, S. P., Pannell, J. R., Peichel, C. L., Ashman, T.-L., Charlesworth, D., Chippindale, A. K., Delph, L. F., Guerrero, R. F., Scarpino, S. V., and McAllister, B. F. (2011). About PAR: The distinct evolutionary dynamics of the pseudoautosomal region. *Trends in Genetics*, 27(9):358–367.

Otto, S. P. and Whitton, J. (2000). Polyploid incidence and evolution. *Annual Review of Genetics*, 34:401–437.

Pagel, M. and Meade, A. (2006). Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist*, 167(6):808–825.

Pease, J. B. and Hahn, M. W. (2012). Sex chromosomes evolved from independent ancestral linkage groups in winged insects. *Molecular Biology and Evolution*, page doi:10.1093/molbev/mss010.

Perrin, N. (2009). Sex reversal: A foutain of youth for sex chromosomes? *Evolution*, 63(12):3043–3049.

Pischedda, A. and Chippindale, A. K. (2006). Intralocus sexual conflict dimishes the benefits of sexual selection. *PLoS Biology*, 4(11):e356.

Pitman, W. I., Cande, S., LaBrecque, J., and Pindell, J. (1993). *Biological relationships between Africa and South America*, chapter Fragmentation of Gondwana: The separation of Africa from South America, pages 15–34. New Haven (CT): Yale University Press.

Pomiankowski, A., Nöthiger, R., and Wilkins, A. (2004). The evolution of the *Drosophila* sex-determination pathway. *Genetics*, 166(4):1761–1773.

Popov, P. and Dimitrov, B. (1999). Karyotype study of *Rana camerani* and comparisons with the other 26-chromosome European brown frog species (amphibia, Anura). *Cytobios*, 97(384):13–22.

Pyron, R. A. and Wiens, J. J. (2011). A large-scale phylogeny of amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, 61(2):543–583.

Rambaut, A. and Drummond, A. J. (2007). Tracer v1.4. Available from http://beast.bio.ed.ac.uk/Tracer.

Raymond, C. S., Kettlewell, J. R., Hirsch, B., Bardwell, V. J., and Zarkower, D. (1999). Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. *Developmental Biology*, 215(2):208–220.

Raymond, C. S., Murphy, M. W., O'Sullivan, M. G., Bardwell, V. J., and Zarkower, D. (2000). *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes and Development*, 14:2587–2595.

Rice, W. R. (1986). On the instability of polygenic sex determination: The effect of sex-specific selection. *Evolution*, 40(3):633–639.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43:223–225.

Rice, W. R. (1996). Evolution of the Y sex chromosome in animals. *BioScience*, 46(5):331–343.

Roelants, K. and Bossuyt, F. (2005). Archaeobatrachian paraphyly and Pangean diversification of crown-group frogs. *Systematic Biology*, 54(1):111–126.

Roelants, K., Gower, D. J., Wilkinson, M., Loader, S. P., Biju, S. D., Guillaume, K., Moriau, L., and Bossuyt, F. (2007). Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences*, 104(3):887–892.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12):1572–1574.

Rosenberg, N. A. (2002). The probability of topological concordance of gene trees and species trees. *Theoretical Population Biology*, 61(2):225–247.

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 Genetics*, 5(2):e1000391.

Ryuzaki, M., Hanada, H., Okumoto, H., Takizawa, N., and Nishioka, M. (1999). Evidence for heteromorphic sex chromosomes in males of *Rana tagoi* and *Rana sakuraii* in Nishitama district of Tokyo (Anura: Ranidae). *Chromosome Research*, 7:31–42.

Salzmann, U. and Hoelzmann, P. (2005). The Dahomey Gap: An abrupt climatically induced rain forest fragmentation in West Africa during the late Holocene. *The Holocene*, 15(2):190–199.

Sanderson, M. J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution*, 19:101–109.

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times in the absence of rate consistency. *Molecular Biology and Evolution*, 19:1218–1231.

Sassoon, D. and Kelley, D. B. (1986). The sexually dimorphic larynx of *Xenopus laevis*: Development and androgen regulation. *American Journal of Anatomy*, 177:457–472.

Sato, T., Endo, T., Yamahira, K., Hamaguchi, S., and Sakaizumi, M. (2005). Induction of female-to-male sex reversal by high temperature treatment in medaka, *Oryzias latipes*. *Zoological Science*, 22(9):985–988.

Schempp, W. and Schmid, M. (1981). Chromosome banding in amphibia. VI. BrdU-replication patterns in Anura and demonstration of XX/XY sex chromosomes in *Rana esculenta*. *Chromosoma*, 83(5):697–710.

Schmid, M. (1978). Chromosome banding in amphibia. I. Constitutive heterochromatin and nucleolus organizer regions in *Bufo* and *Hyla*. *Chromosoma*, 66:361–388.

Schmid, M. (1980). Chromosome banding in amphibia V. Highly differentiated ZW/ZZ sex chromosomes and exceptional genome size in *Pyxicephalus adspersus* (Anura, Ranidae). *Chromosoma*, 80:69–96.

Schmid, M. and Bachmann, K. (1981). A frog with highly evolved sex chromosomes. *Experientia*, 37(3):243–245.

Schmid, M., Feichtinger, W., Steinlein, C., Garcia, R. V., and Badillo, A. F. (2003). Chromosome banding in amphibia: XXVIII. Homomorphic XY sex chromosomes and a derived Y-autosome translocation in *Eleutherodactylus riveroi* (Anura, Leptodactylidae). *Cytogenetic and Genome Research*, 101(1):62–73.

Schmid, M., Feichtinger, W., Steinlein, C., Nanda, I., Mais, C., Haaf, T., Garcia, R. V., and Badillo, A. F. (2002a). Chromosome banding in amphibia. XXII. Atypical Y chromosomes in *Gastrotheca walkeri* and *Gastrotheca ovifera* (Anura, Hylidae). *Cytogenetic and Genome Research*, 96:228–238.

Schmid, M., Feichtinger, W., Steinlein, C., Rupprecht, A., Haff, T., and Kaiser, H. (2002b). Chromosome banding in amphibia. XXIII. Giant W sex chromosomes and extremely small genomes in *Eleutherodactylus euphronidies* and *Eleutherodactylus shrevei*. *Cytogenetic and Genome Research*, 97(1-2):81–94.

Schmid, M., Haaf, T., Geile, B., and Sims, S. (1983). Chromosome banding in amphibia VIII. An unusual XY/XX-sex chromosome system in *Gastrotheca riobambae* (Anura, Hylidae). *Chromosoma*, 88(1):69–82.

Schmid, M. and Klett, R. (1994). Chromosome banding in amphibia. XX. DNA replication patterns in *Gastrotheca riobambae* (Anura, Hylidae). *Cytogenetics and Cell Genetics*, 65(1-2):122–126.

Schmid, M., Ohta, S., Steinlein, C., and Guttenbach, M. (1993). Chromosome banding in amphibia. XIX. primitive ZW/ZZ sex chromosomes in *Buergeria buergeri* (Anura, Rhacophoridae). *Cytogenetics and Cell Genetics*, 62(4):238–246.

Schmid, M., Sims, S. H., Haaf, T., and Macgregor, H. C. (1986). Chromosome banding in amphibia. X: 18S and 28S ribosomal RNA genes, nucleolus organizers and nucleoli in *Gastrotheca riobambae*. *Chromosoma*, 94(2):139–145.

Schmid, M. and Steinlein, C. (2001). *Genes and mechanisms in vertebrate sex determination*, chapter Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class amphibia, pages 143–176. Birkhäuser Verlag, Basel-Boston-Berlin.

Schmid, M. and Steinlein, C. (2003). Chromosome banding in amphibia. XXIX. the primitive XY/XX sex chromosomes of *Hyla femoralis* (Anura, Hylidae). *Cytogenetic and Genome Research*, 101:74–79.

Schmid, M., Steinlein, C., Bogart, J. P., Feichtinger, W., León, P., Marca, E. L., Díaz, L. M., Sanz, A., Chen, S. H., and Hedges, S. B. (2010). The chromosomes of terraranan frogs. *Cytogenetic and Genome Research*, 130-131:1–568.

Schmid, M., Steinlein, C., and Feichtinger, W. (1989). Chromosome banding in amphibia. XIV. the karyotype of *Centrolenella antisthenesi* (Anura, Centrolenidae). *Chromosoma*, 97:434–438.

Schmid, M., Steinlein, C., and Feichtinger, W. (1992). Chromosome banding in amphibia. XVII. First demonstration of multiple sex chromosomes in amphibians: *Eleutherodactylus maussi* (Anura, Leptodactylidae). *Chromosoma*, 101:284–292.

Schmid, M., Steinlein, C., Feichtinger, W., de Almeida, C. G., and Duellman, W. (1988). Chromosome banding in amphibia. XIII. Sex chromosomes, heterochromatin and meiosis in marsupial frogs. *Chromosoma*, 97:33–42.

Schmid, M., Steinlein, C., Friedl, R., de Almeida, C. G., Haaf, T., Hillis, D. M., and Duellman, W. E. (1990). Chromosome banding in amphibia. XV. Two types of Y chromosomes and heterochromatin hypervariabilty in *Gastrotheca pseustes* (Anura, Hylidae). *Chromosoma*, 99:413–423.

Seifertova, E., Zimmerman, L. B., Gilchrist, M., Macha, J., Kubickova, S., Zarsky, V., Owens, N. D. L., Sesay, A., Tlapakova, T., and Krylov, V. (2013). Efficient high-throughput sequencing of laser microdissected chromosome arm. *submitted*.

Sémon, M. and Wolfe, K. H. (2008). Preferential subfunctionalization of slow-evolving genes after allopolyploidization in *Xenopus laevis*. *PNAS*, 105(24):8333–8338.

Sereno, P. C., Wilson, J. A., and Conrad, J. L. (2004). New dinosaurs link southern landmasses in the Mid-Cretaceous. *Proceedings of the Royal Society B*, 271:1325–1330.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P. J., Cordum, H. S., Hillier, L., Brown, L. G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., Chinwalla, A., Delehaunty, A., Delehaunty, K., Du, H., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hou, S.-F., Latrielle, P., Leonard, S., Mardis, E., Maupin, R., McPherson, J., Miner, T., Nash, W., Nguyen, C., Ozersky, P., Pepin, K., Rock, S., Rohlfing, T., Scott, K., Schultz, B., Strong, C., Tin-Wollam, A., Yang, S.-P., Waterston, R. H., Wilson, R. K., Rozen, S., and Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, 423:825–837.

Smith, A. G., Smith, D. G., and Funnell, B. M. (1994). *Atlas of Mesozoic and Cenozoic coastlines*. Cambridge: Cambridge University Press.

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., nigo Martinez-Solano, I., and Perrin, N. (2011). Ever-young sex chromosomes in European tree frogs. *PLoS Biology*, 9(5):e1001062.

Stöck, M., Moritz, C., Hickerson, M., Frynta, D., Dujsebayeva, T., Eremchenko, V., Macey, J. R., Papenfuss, T. J., and Wake, D. B. (2006). Evolution of mitochondrial relationships and biogeography of Palearctic green toads (*Bufo viridis* subgroup) with insights in their genomic plasticity. *Molecular Phylogenetics and Evolution*, 41:663–689.

Stöck, M., Savary, R., Betto-Collard, C., Biollay, S., Jourdan-Pineau, H., and Perrin, N. (2013). Low rates of X-Y recombination, not turnovers, account for homomorphic sex chromosomes in several diploid species of Palearctic grren toads (*Bufo viridis* subgroup). *Journal of Evolutionary Biology*, page doi:10.1111/jeb.12086.

Stöck, M., Steinlein, C., Lamatsch, D. K., Schartl, M., and Schmid, M. (2005). Multiple origins of tetraploid taxa in the Eurasian *Bufo viridis* subgroup. *Genetica*, 124:255–272.

Sukumaran, J. and Linkem, C. W. (2009). Choice of topology estimators in Bayesian phylogenetic analysis. *Molecular Biology and Evolution*, 26(1):1–3.

Swofford, D. L. (2022). Phylogenetic analysis using parsimony (* and other methods). version 4. Sunderland (MA): Sinauer Associates.

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.

Trueb, L. (1996). *The Biology of Xenopus*, chapter Historical constraints and morphological novelties in the evolution of the skeletal system of pipid frogs (Anura: Pipidae), pages 349–377. Clarendon Press, Oxford.

Trueb, L. and Báez, A. M. (1997). Redescription of the Paleogene *Shelania pascuali* from Patagonia and its bearing on the relationships of fossil and recent pipoid frogs. *Scientific Papers, Natural History Museum, Univerisity of Kansas*, 4:1–41.

Trueb, L. and Báez, A. M. (2006). Revision of the Early Cretaceous *Cordicephalus* from Israel and an assessment of its relationships among pipoid frogs. *Journal of Vertebrate Paleontology*, 26(1):44–59.

Trueb, L. and Cannatella, D. C. (1986). Systematics, morphology, and phylogeny of genus *Pipa* (Anura: Pipidae). *Herpetologica*, 42(4):412–449.

Trueb, L., Ross, C. F., and Smith, R. (2005). A new pipoid anuran from the Late Cretaceous of South Africa. *Journal of Vertebrate Paleontology*, 25(3):533–547.

Tsuda, Y., Nishida-Umehara, C., Ishijima, J., Yamada, K., and Matsuda, Y. (2007). Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma*, 116(2):159–173.

Tucker, P. K., Adkins, R. M., and Rest, J. S. (2003). Differential rates of evolution for the *ZFY*-related zinc finger genes, *Zfy*, *Zfx*, and *Zfa* in the mouse genus *Mus*. *Molecular Biology and Evolution*, 20(6):999–1005.

Tucker, P. K. and Lundrigan, B. L. (1993). Rapid evolution of the sex determining locus in old world mice and rats. *Nature*, 364(6439):715–717.

Tymowska, J. (1991). *Polyploidy and cytogenetic variation in frogs of the genus Xenopus*. Amphibian cytogenetics and evolution. Academic Press., San Diego, CA.

Uchida, T. (1937a). Studies on the sexuality of amphibia. II. Sexual induction in a sexually semi-differentiated salamander. *Journal of the Faculty of Science Hokkaido Imperial University Series VI. Zoology*, 6(1):35–38.

Uchida, T. (1937b). Studies on the sexuality of amphibia: III. sex transformation in *Hynobius retardatus* by the function of high temperature. *Journal of the Faculty of Science Hokkaido Imperial University Series VI. Zoology*, 6(1):59–70.

Uno, Y., Nishida, C., Takagi, C., Ueno, N., and Matsuda, Y. (2013). Homoeologous chromosomes of *Xenopus laevis* are highly conserved after whole genome duplication. *Journal of Heredity*, in review.

Uno, Y., Nishida, C., Yoshimoto, S., Ito, M., Oshima, Y., Yokoyama, S., Nakamura, M., and Matsuda, Y. (2008). Diversity in the origins of sex chromosomes in Anurans inferred from comparative mapping of sexual differentiation genes for three species of the Raninae and Xenopodinae. *Chromosome Research*, 16:999–1011.

van Doorn, G. S. and Kirkpatrick, M. (2007). Turnover of sex chromosomes induced by sexual conflict. *Nature*, 449:909–912.

Vences, M., Vieites, D. R., Glaw, F., Brinkmann, H., Kosuch, J., Veith, M., and Meyer, A. (2003). Multiple overseas dispersal in amphibians. *Proceedings of the Royal Society B*, 270(1532):2435–2442.

Vicoso, B. and Bachtrog, D. (2009). Progress and prospects toward our understand of the evolution of dosage compensation. *Chromosome Research*, 17:585–602.

Vicoso, B. and Bachtrog, D. (2011). Lack of global disage compensation in *Schistosoma mansoni*, a female-heterogametic parasite. *Genome Biology and Evolution*, 3:230–235.

Vicoso, B., Kaiser, V. B., and Bachtrog, D. (2013). Sex-biased gene expression at homomorphic sex chromosomesin emus and its implications for sex chromosome evolution. *PNAS*, page doi:10.1073/pnas.1217027110.

Viulleumier, S., Lande, R., van Alphen, J. J. M., and Seehausen, O. (2007). Invasion and fixation of sex-reversal genes. *Journal of Evolutionary Biology*, 20(3):913–920.

Wallace, H., Wallace, B. M. N., and Badawy, G. M. I. (1997). Lampbrush chromosomes and chiasmata of sex-reversed crested newts. *Chromosoma*, 106(8):526–533.

Warren, W. C. (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, 453:175–183.

Watson, J. T. and Kelley, D. B. (1992). Testicular masculinization of vocal behavior in juvenile female *Xenopus laevis* reveals sensitive periods for song duration, rate, and frequency spectra. *Journal of Comparative Physiology*, 171:343–350.

Wells, D. E., Gutierrez, L., Xu, Z., Krylov, V., Macha, J., Blankenburg, K. P., Hitchens, M., Bellot, L. J., Spivey, M., Stemple, D. L., Kowis, A., Ye, Y., Pasternak, S., Owen, J., Tran, T., Slavikova, R., Tumova, L., Tlapakova, T., Seifertova, E., Scherer, S. E., and Sater, A. K. (2011). A genetic map of *Xenopus tropicalis*. *Developmental Biology*.

Werren, J. H. and Beukeboom, L. W. (1998). Sex determination, sex ratios, and genetic conflict. *Annual Review of Ecology and Systematics*, 29:233–261.

Whitfield, L. S., Lovell-Badge, R., and Goodfellow, P. N. (1993). Rapid sequence evolution of the mammalian sex-determining gene *SRY*. *Nature*, 364:713–715.

Wiens, J. J. (2011). Re-evolution of lost mandibular teeth in frogs after more than 200 million years, and re-evaluating Dollo's law. *Evolution*, 65(5):1283–1296.

Wiley, J. E. (2003). Replication banding and FISH analysis reveal the origin of the *Hyla femoralis* karyotype and XY/XX sex chromosomes. *Cytogenetic and Genome Research*, 101(1):80–83.

Wilkins, A. S. (1995). Moving up the hierarchy: A hypothesis on the evolution of a genetic sex determination pathway. *BioEssays*, 17(1):71–77.

Withjack, M. O., Schlische, R. W., and Olsen, P. E. (1998). Diachronous rifting, drifting, and inversion on the passive margin of Central Eastern North America: An analog for other passive margins. *AAPG Buletin*, 82(5A):817–835.

Witschi, E. (1929). Studies on sex differentiation and sex determinationin amphibians. III. Rudimentary hermaphroditism and Y chromosome in *Rana temporaria*. *Journal of Experimental Zoology*, 54:157–223.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. *Nature*, 430:85–88.

Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2008). Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genetics*, 40(3):346–350.

Wolfe, K. H. (2001). Yesterday's polyploids and the mystery of diploidization. *Nature Reviews Genetics*, 2:333–341.

Wong, K. M., Suchard, M. A., and Huelsenbeck, J. P. (2008). Alignment uncertainty and genomic analysis. *Science*, 319(5862):473–476.

Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114–138.

Wright, S. I. and Charlesworth, B. (2004). The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics*, 168(2):1071–1076.

Wyckoff, G. J., Li, J., and Wu, C.-I. (2002). Molecular evolution of functional genes on the mammalian Y chromosome. *Molecular Biology and Evolution*, 19(9):1633–1636.

Yager, D. D. (1996). *The Biology of Xenopus*, chapter Sound production and acoustic communication in *Xenopus borealis*, pages 121–141. Clarendon Press, Oxford.

Yang, Z. (1997). PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS*, 13:555–556.

Yang, Z. (1998). Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15(5):568–573.

Yang, Z., Wong, W. S., and Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22(4):1107–1118.

Yi, W. and Zarkower, D. (1999). Similarity of DNA binding and transcriptional regulation by *Caenorhabditis elegans* MAB-3 and *Drosophila melanogaster* DSX suggests conservation of sex determining mechanisms. *Development*, 126:873–881.

Yoshida, T. H. (1957). Sex chromosomes of the tree frog, *Hyla arborea japonica*. *Journal of the Faculty of Science Hokkaido University Series VI. Zoology*, 13(1-4):352–358.

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*, 137:2519–2526.

Yoshimoto, S., Okada, E., Oishi, T., Numagami, R., Umemoto, H., Tamura, K., Kanda, H., Shiba, T., Takamatsu, N., and Ito, M. (2006). Expression and promoter analysis of *Xenopus DMRT1* and functional characterization of the transactivation property of its protein. *Development, Growth and Differentiation*, 48:597–603.

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*. *PNAS*, 105(7):2469–2474.

Zha, X., Xia, Q., Duan, J., Wang, C., He, N., and Xiang, Z. (2009). Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. *Insect Biochemistry and Molecular Biology*, 39:315–321.

Zhang, J. (2004). Evolution of *DMY*, a newly emergent male sex-determination gene of medaka fish. *Genetics*, 166(4):1887–1895.

Zhou, Q. and Bachtrog, D. (2012). Chromosome-wide gene silencing initiates Y degeneration in *Drosophila*. *Current Biology*, 22(6):522–525.