Studies on the molecular underpinnings of sex determination mechanism evolution and molecular sexing tools in turtles

by

Robert Literman

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Ecology and Evolutionary Biology

Program of Study Committee:
Maria N Valenzuela-Castro, Major Professor
Jeff Essner
John Nason
Michael Shogren-Knaak
Jonathan Wendel

Iowa State University

Ames, Iowa

2017

DEDICATION

This work is dedicated to my family: To my parents who supported me in all endeavors leading here, and to my wife Krystal and our son Rubin, who kept me going through it all.

TABLE OF CONTENTS

NOMENCLATURE

ESD	Environmental Sex Determination
gDNA	Genomic DNA
GSD	Genotypic Sex Determination
NGS	Next-Generation Sequencing
rDNA	Ribosomal DNA
SDM	Sex Determination Mechanism
TSD	Temperature-Dependent Sex Determination
TSP	Thermosensitive Period

ABSTRACT

Sex determination mechanisms (SDMs) direct the development of individuals towards a male or female fate, and in vertebrates they are typically controlled by an individual's genotypic content (genotypic sex determination, GSD) or through an environmental cue experienced during development, mainly temperature (temperature-dependent sex determination, TSD). Among vertebrates, SDMs are surprisingly labile, transitioning between different forms of TSD and GSD in some lineages more than others. Turtles represent a model clade to study SDM evolution, as multiple independent transitions between TSD and GSD have occurred throughout their evolution and a growing number of genomic datasets have become available.

This dissertation examines the molecular underpinnings of SDM evolution in turtles while also providing tools that enable studies of sex determination across taxa and of sex-specific traits. In Chapter 2, I examine the molecular evolution of a suite of vertebrate sex determining genes in turtles, contrasting their evolutionary rates to those of other major vertebrate clades and also among turtle lineages. Furthermore, I compare the evolutionary rates of turtle lineages which have undergone SDM transitions versus those that have not. I then compare the relative evolutionary rates of turtles that have transitioned from TSD-to-GSD against lineages which have possibly transitioned from GSD-to-TSD. Finally, I discuss amino acid substitutions which occur in the functional domains of key sex determining genes along transitional branches, providing targets for future research. In Chapter 3, I present an analytical pipeline which can diagnose sex with 100% accuracy in the ZZ/ZW spiny softshell turtle *Apalone spinifera* by leveraging a previously described sex-biased copy number

variation of rDNA clusters between the sexes. The pipeline is also applied to a previously published dataset of circulating hormonal concentrations in the snapping turtle *Chelydra serpentina*, where it shows greater than 85% accuracy in sex diagnosis. In Chapter 4, I lay out a bioinformatics pipeline which details the sample collection, sequencing, and analysis of sex-specific DNA libraries, which can be used to identify sex-linked DNA sequences in any taxon with sufficient genetic differentiation between the sexes. This information is leveraged to create sex-diagnostic PCR primers which are 100% accurate at diagnosing sex in the focal taxa, the ZZ/ZW *A. spinifera* and the XX/XY wood turtle *Glyptemys insculpta*. Furthermore, primers designed against the focal taxa data are also applied successfully to related species, expanding the utility of the pipeline, while simultaneously providing the first definitive evidence that the bog turtle *Glyptemys muhlenbergii* has an XX/XY sex chromosome system. Together these chapters provide data about the proximate mechanisms of SDM evolution in turtles, which are necessary to begin to understand the ultimate explanations for why SDM evolution is so labile across taxa.

CHAPTER 1: INTRODUCTION

Sex-determining mechanisms (SDMs) direct the development of individuals towards a male or female fate (Beukeboom and Perrin 2014), and the nature of the sex-determining trigger that tips that balance broadly categorizes these mechanisms into two major groups: genotypic sex determination (GSD) and environmental sex determination (ESD) (Valenzuela and Lance 2004). Sex in GSD species is determined by the content of an individual's genotype at conception, typically via sex-determining genes which reside in sex chromosomes. Alternatively in species with ESD, sex is determined not at conception but rather further along in development, triggered by some environmental cue which directs gonadal development towards the formation of either testes or ovaries (Charnier 1965, Charnov and Bull 1977, Valenzuela and Lance 2004). In vertebrates, the most common environmental cue for ESD is temperature (temperature-dependent sex determination, TSD) (Valenzuela and Lance 2004). Among vertebrates, all studied amphibians, mammals, snakes, and birds have GSD, while all crocodilians, most turtles, the tuataras, and some squamates and fishes exhibit TSD (Valenzuela and Lance 2004, Sarre *et al.* 2011). Given the phylogenetic distribution of SDMs across the vertebrate phylogeny it appears that despite the paramount importance of ensuring that males and females are produced at appropriate frequencies within a population, that SDM evolution among vertebrates is surprising labile, transitioning between different forms of GSD and TSD throughout the evolutionary history of vertebrates (Valenzuela and Lance 2004, Sarre *et al.* 2011, Beukeboom and Perrin 2014).

While ultimate explanations for the diversity in SDMs among vertebrates have been proposed (Charnov and Bull 1977, Valenzuela and Lance 2004, Warner and Shine 2008, Sabath *et al.* 2016), much of the underlying molecular architecture underlying such

transitions remains elusive. Turtles represent a model clade for the study of SDM evolution for a number of reasons. First, the ancestral state of turtle sex determination can be fairly reliably reconstructed as being TSD, and from that common ancestor multiple transitions from TSD-to-GSD have occurred, along with potential 'reversals' back to TSD from a GSD ancestor (Valenzuela and Adams 2011, Sabath *et al.* 2016). Second, genomic resources including time-course transcriptomic analyses (Czerwinski *et al.* 2016, Radhakrishnan *et al.* 2017, Zhang *et al.* 2017), full genome sequencing of TSD and GSD species (Shaffer *et al.* 2013, Wang *et al.* 2013), and cytogenetic analyses (Janes *et al.* 2008, Badenhorst *et al.* 2015, Montiel *et al.* 2016a, Montiel *et al.* 2016b) allow for a careful dissection of the sex-determining network and chromosomal evolution across many turtle species which span SDMs. Finally, many key players in the turtle sex determination network have been elucidated through candidate gene expression analyses (Maldonado *et al.* 2002, Valenzuela 2008b, Barske and Capel 2010, Valenzuela *et al.* 2013), which examine the temperature-sensitivity of gene expression across developmental stages for genes known to contribute to sex determination in more well-studied vertebrate clades, such as the therian mammals (Valenzuela 2008a). With such a wealth of information, turtles have emerged as a model clade to examine the molecular underpinnings of SDM evolution.

Despite valuable strides to unravel the mysteries of turtle sex determination and its evolution, numerous remain unanswered. What properties of turtle genomes facilitate their evolutionary lability with respect to SDM transitions? How do members of the sex determination gene network evolve as turtle lineages transition between TSD and GSD, shifting the role of sex-determining trigger from egg incubation temperature to a genotypic component, or vice versa? The sex of TSD turtle species can be predicted and controlled with

high repeatability through the manipulation of incubation temperatures (Valenzuela 2009b), facilitating sex-specific studies of embryonic gene expression and epigenetic modifications throughout the developmental period during embryogenesis when sex is being determined. However, as sex in GSD turtles is determined primarily by the genotype and because turtle embryos lack any discernable sexual dimorphism before and during the sex-determining period, how can we reliably diagnose the genotypic sex of embryonic GSD turtles to facilitate complementary studies of their underlying sex determination network and the potential relic effects of incubation temperature on sex-determining gene expression? How can we generate robust datasets to distinguish sex-specific DNA in GSD species that not only enable the development of sex-diagnostic PCR assays, but also illuminate the evolutionary dynamics of sex-linked DNA across taxa to test theoretical models? These are the questions that I address in this dissertation.

In the following chapters, I examine the molecular evolution of core vertebrate sex determination genes and what they tell us about SDM evolution in turtles (Chapter 2), discuss a method to diagnose sex in softshell turtles by leveraging sex-specific copy number variation of rDNA clusters contained within their sex chromosomes (Chapter 3), and present a Next-Generation Sequencing (NGS) pipeline that permits the design of sex-diagnostic PCR primers for any GSD species with sufficient genetic differentiation between the sexes (Chapter 4).

More specifically, in Chapter 2, I investigate the molecular evolution of fifteen sex determination genes in ten turtle species, including five TSD and five GSD turtles. The evolutionary rate of these genes and their associated proteins is compared between turtles and other major vertebrate clades, as well as among turtle lineages. Furthermore, the correlation

between SDM transition and evolutionary rate among turtles is discussed under two contrasting evolutionary hypotheses which reconstruct turtle SDM evolution with and without SDM evolutionary reversals. Finally, key amino acid substitutions which lead to predicted secondary protein structure changes in functional domains of key sex determination genes are identified that are potential targets for future functional studies. In Chapter 3 (Literman *et al.* 2014), I describe a qPCR-based assay which is capable of 100% accurate genotypic sex identification for the ZZ/ZW softshell turtle *Apalone spinifera*. This method can be applied to test for sexual dimorphism in any continuous and multivariate dataset (such as shape or behavior) more reliably that previously proposed methods. Practical experimental design and analytical elements which affect data interpretation are discussed. In Chapter 4 (Literman *et al.* 2017), I lay out a bioinformatics pipeline which takes as input two lanes of DNA-sequencing data derived from genomic DNA of one male and one female and yields 100% accurate sex-diagnostic primer sets. We apply this method to two focal species, *A. spinifera* and the wood turtle *Glyptemys insculpta*. Primers designed for these focal species were also successfully applied to closely related species, which provide the first reliable evidence for the existence an XX/XY SDM in the critically-endangered bog turtle *Glyptemys muhlenbergii*, a species for which egg incubation experiments are largely precluded. Data from this chapter also provide nucleotide sequence information which is used in Chapter 1, expanding the utility of the pipeline beyond sex-diagnostic primer design. Taken together, these chapters both advance our knowledge of the molecular underpinnings of SDM evolution in turtles, while also providing new molecular tools which can be used broadly across taxa (and data types) to bring the SDMs of non-model organisms out of the shadows,

facilitating sex-specific studies in species or developmental stages which were previously precluded due to a lack of reliable sex markers.

CHAPTER 2: PUTATIVE INDEPENDENT EVOLUTIONARY REVERSALS FROM GENOTYPIC TO TEMPERATURE DEPENDENT SEX DETERMINATION ARE ASSOCIATED WITH ACCELERATED EVOLUTION OF SEX-DETERMINING GENES IN TURTLES

Robert Literman[1], Alexandria Burrett[1], Omar E. Hernández[2], and Nicole Valenzuela[1]

Author Contributions: RL collected bioinformatics data, designed and performed analyses, and wrote the manuscript. NV assisted in data analysis and manuscript editing. AB assisted with sequence data collection. OH assisted in specimen procurement.

Author Affiliations:

[1]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA.

[2]Fundación para el Desarrollo de las Ciencias Físicas, Matemáticas y Naturales, FUDECI, Av. Palacio de Las Academias, Edf. Anexo, Piso 2, Caracas, Venezuela.

2.1 Abstract

The evolutionary lability of sex determination across the tree of life is well recognized, yet the extent of molecular changes that accompany the repeated transitions in sex determination remain obscure. Most turtles retain the ancestral temperature-dependent sex determination (TSD) from which multiple transitions to genotypic sex determination (GSD) occurred independently, and two contrasting hypotheses differentially posit the existence of two reversals back to TSD. Here we examined the molecular evolution of the coding regions of a set of gene regulators of gonadal development in turtles and several other vertebrates. We

found slower molecular evolution in turtles and crocodilians compared to other vertebrates, but an acceleration in Trionychia turtles and at some phylogenetic branches demarcating major taxonomic diversification events. Of all gene classes examined, hormone signaling genes and *Srd5a1* in particular, evolve faster in many lineages and especially in turtles. Our data show that sex-linked genes do not follow a ubiquitous nor uniform pattern of molecular evolution. We then evaluated turtle nucleotide and protein evolution under both evolutionary hypotheses with or without GSD-to-TSD reversals, and found that when GSD-to-TSD reversals are allowed, all transitional branches irrespective of direction, exhibit accelerated molecular evolution of nucleotide sequences, while GSD-to-TSD transitional branches also show acceleration in protein evolution. Significant changes in predicted secondary structure that may affect protein function were identified in three genes that exhibited accelerated evolution in turtles compared to other vertebrates or in transitional versus non-transitional branches within turtles, rendering them candidates for a key role during SDM evolution in turtles.

2.2 Introduction

Among vertebrates various sex-determining mechanisms (SDMs) direct the developmental fate of individuals towards the male or female condition (Beukeboom and Perrin 2014), controlled primarily through the interaction of two major components: (1) the genes, whose products in a given developmental or genotypic background direct the morphological differentiation of bipotential gonads into testes or ovaries, and (2) the environmental cues that individuals experience throughout development which may in some cases affect the deployment of those gene products. In many species, including all studied

mammals, birds, snakes, and amphibians, the sex determination trigger is the individual's genomic content (genotypic sex determination or GSD) (Valenzuela and Lance 2004). Most commonly among vertebrates, these sex-specific genotypic differences are found in sex chromosomes that contain the sex-determining regions which are present in one sex while absent in the other, or present in different doses between males and females. Sex chromosome systems include male heterogamety (e.g. XX females, XY males) and female heterogamety (e.g. ZZ males, ZW females). In other species, including some fishes and squamates, most turtles, and all known crocodilians, no sex-specific genomic content exists and instead, sex is determined by environmental cues experienced during development (environmental sex determination or ESD), the most common cue of which is temperature (temperature-dependent sex determination; TSD) (Valenzuela and Lance 2004, Ashman *et al.* 2014). Components of both genotype and the environment may work within some taxa with intermediate mechanisms between the GSD and TSD extremes (Valenzuela *et al.* 2003, Sarre *et al.* 2004), such as taxa with GSD + environment influence (Lagomarsino and Conover 1993, Shine *et al.* 2002, Quinn *et al.* 2007); while in other taxa, populations may possess contrasting SDMs (Uno *et al.* 2008).

Sex determination is fundamental to the life history of any species. Yet, the extent to which evolutionary changes in the master sex-determining factor(s) trigger concerted changes in other elements of the sex determination gene network to maintain proper development of males and females remains an open question. Thus, understanding the molecular evolution of core gene regulators of sexual development across vertebrates could help uncover associations between SDM transitions and changes in the underlying molecular machinery that may be candidate key contributors to the evolution of sex determination itself.

Most vertebrates (GSD or TSD) share a core suite of genes which are integral for the development of the sexual phenotype (Valenzuela and Lance 2004, Beukeboom and Perrin 2014). Individual genes can be classified as male-promoting or female-promoting based on their role on the developing gonads and whether these roles are fairly conserved among species (e.g. *Aromatase [Cyp19a1], Dmrt1, Sox9* genes) or not (e.g.*Sf1 [Nr5a1]*) (da Silva *et al.* 1996, Raymond *et al.* 2000, Valenzuela *et al.* 2013, Beukeboom and Perrin 2014). Indeed, the relative roles and deployment of these genes has evolved among vertebrate lineages, with different genes attaining the top-most role within the sex determination network in different systems (Valenzuela and Lance 2004, Beukeboom and Perrin 2014). Network changes include shifts in the timing of gene activation (Valenzuela et al. 2013), genes shifting positions in the regulatory pathways between being master regulators or downstream actors (Cutting *et al.* 2013), or temperature becoming the key regulator of one or more genes at a crucial developmental period (thermosensitive period or TSP), tipping the balance towards a male or female fate in TSD taxa.

SDM evolutionary transitions occurred in fishes, geckos and agamid lizards, and turtles (Valenzuela and Lance 2004, Sarre *et al.* 2011, Sabath *et al.* 2016). Molecularly, TSD-to-GSD transitions that involved sex chromosome evolution would be characterized by the appearance (or translocation) of a master sex-determining gene onto a proto-sex chromosome, whose overall effect on the sex determination network would outweigh the ancestral effect of incubation temperature, as described in many taxa. Conversely, transitions from GSD-to-TSD would involve either (1) molecular changes that lead to temperature sensitivity on the expression or activity of gene products at key points in this network or (2) the cooption of a novel thermosensitive element or pathway into the sex determination

network. Fewer transitions from a GSD ancestor towards TSD are documented in vertebrates (Pokorna and Kratochvíl 2009, Sabath *et al.* 2016), perhaps because sex chromosomes are an 'evolutionary trap' (Pokorna and Kratochvíl 2009) given that during the transition to TSD, YY or WW individuals would be produced that may be suboptimal or lethal, or alternatively, perhaps because less likely events may be required to de-differentiate sex chromosomes back into autosomes, a rare phenomenon that can also have high fitness cost (Vicoso and Bachtrog 2013). Alternatively, certain life history traits such as longevity may render TSD-to-GSD transitions more likely in some lineages as shorter life-spans accentuate the negative effects induced by TSD on population dynamics such as sex ratios skews (Sabath *et al.* 2016).

An ultimate explanation for the evolutionary lability in SDM remains elusive. Turtles represent a model clade to study SDM evolution as they possess TSD and GSD. TSD has been reconstructed as the ancestral state for turtles, from which multiple independent transitions to GSD appeared to have occurred, as well as reversals back to TSD (Valenzuela and Lance 2004, Valenzuela and Adams 2011, Sabath *et al.* 2016). Interestingly, transitions in turtle SDM are accompanied by drastic cytogenetic reshuffling, where lineages that experienced an SDM transition exhibit a ~20 fold higher rate of chromosome number evolution (Valenzuela and Adams 2011). This raises the question of whether an increased rate of molecular evolution at the gene level accompanies SDM transitions in turtles, as does a greater rate of chromosome evolution in this group (Valenzuela and Adams 2011). Addressing this and related questions is facilitated by the growing number of turtle genomic resources including a BAC library (Janes *et al.* 2008, Badenhorst *et al.* 2015), candidate-gene expression analyses (Maldonado *et al.* 2002, Barske and Capel 2010, Valenzuela *et al.* 2013), transcriptomics (Czerwinski *et al.* 2016, Radhakrishnan *et al.* 2017, Zhang *et al.* 2017),

methylation profiling (Matsumoto *et al.* 2016, Venegas *et al.* 2016), as well as sequenced genomes of both TSD and GSD species (Shaffer *et al.* 2013, Wang *et al.* 2013).

Here we examine the molecular evolution of a subset of fifteen genes in the vertebrate sex determination network (transcription factors, hormone-signaling genes, WNT-signaling genes, and temperature-sensing genes), using turtles as a focal group to test whether lineages which have undergone SDM transitions (TSD-to-GSD or GSD-to-TSD) are characterized by higher rates of nucleotide or amino acid substitution rates in the target sex determination genes compared to turtle and other major vertebrate lineages where no SDM transition occurred. This comparative approach permits us to analyze turtle evolutionary rates in a broader phylogenetic context among major vertebrate clades to illuminate the molecular underpinnings of SDM evolution. We then examine changes in the predicted protein secondary structure within functional domains which may alter protein activity for genes that exhibited exceptional evolution in turtles.

2.3 Methods

Sample and data collection

Coding sequence data from fifteen genes involved in vertebrate sex determination (Table 2.1) were collected for twenty-five vertebrate species from various public databases (Figure 2.1, Table 2.2), complemented with sequences for a number of turtle species we obtained during parallel studies (Shaffer *et al.* 2013, Literman *et al.* 2017, Radhakrishnan *et al.* 2017), including both RNA-Seq and DNA-Seq data (Table 2.2). Samples of *Apalone spinifera, Chrysemys picta, Staurotypus triporcatus, Glyptemys insculpta,* and *Carettochelys insculpta* derive from privately owned, pet trade or wild individuals as described elsewhere

(Montiel *et al.* 2016a, Literman *et al.* 2017, Radhakrishnan *et al.* 2017). *Emydura macquarii* tissues were collected as part of other studies conducted at the University of Canberra under appropriate permits. *Podocnemis expansa* samples were collected by FUDECI from the wild in 2003 in Venezuela as part of another study approved by local authorities, and imported under CITES permits. All procedures were approved by the IACUC of Iowa State University, University of Northern Iowa, and University of Canberra.

Data was available for all genes in all taxa with the exception of the *Rspo1* gene, which could not be found for *Ophiophagus hannah* in any available datasets. In order to improve the quality of the sequence data and to ensure that similar isoforms were being compared for each of these genes, all coding sequences were manually extracted from publicly available genomes (when available) rather than from pre-existing gene annotations or other computer-based gene predictions. For species with additional publicly available sequence data, any existing data gaps in the target genes were filled in via the NCBI Short Reach Archive (SRA) or through published mRNA sequences from the target taxa downloaded from Genbank. A dated pruned phylogeny of all species used in this study was generated using the TimeTree database (Hedges *et al.* 2006) (Figure 2.1).

Data processing

For each gene, two datasets were generated: one dataset included sequence alignments from all 25 species ('All Species'), and a second dataset included alignments from the ten turtle taxa exclusively ('Turtles'). In all cases, nucleotide sequences were translationally aligned using MUSCLE with default parameters as implemented in Geneious v.9 (Kearse *et al.* 2012). In order to minimize the impact of taxa-specific indels, all alignments were visually assessed to ensure that homologous regions were aligned, and

misalignments were corrected manually. The lengths and sequence identity of the final alignments for the 'All Species' and 'Turtles' datasets after gaps were removed differ from each other (Table B.1). Alignments were generated for both nucleotide and amino acid sequences.

For each alignment, maximum likelihood was applied to determine the most likely model of substitution using jModelTest v.2.1.4 (Darriba *et al.* 2012) for nucleotide alignments and ProtTest v.2.4 (Abascal *et al.* 2005) for amino acid alignments. Alignments were imported into MEGA v.7 (Kumar *et al.* 2008) and using the estimated model parameters, the number of substitutions-per-site was estimated using maximum likelihood against a well-supported species phylogeny (Figure 2.1) with any sites containing gaps eliminated from the analysis. Substitutions-per-site-per-million years (SSM) for each branch were then calculated by dividing the number of substitutions-per-site by the TimeTree divergence time estimates, and this rate per million years was used in all downstream statistical tests. To estimate neutral nucleotide substitution rates for different taxonomic groups, the nucleotide data was also analyzed using only the third codon position data.

In order to detect amino acid substitutions that could potentially alter protein function of a few genes identified as undergoing exceptional molecular evolution in turtles in the above analyses, the secondary structure of proteins was predicted from the amino acid alignments using the EMBOSS plugin as implemented in Geneious, and functional domains were annotated via the UniProt database.

Data analysis

Parallel analyses were performed for the nucleotide and the amino acid alignments. Due to unequal variances in rates among groups, non-parametric statistical tests (described

below) were performed to compare the differences among groups. For both the 'All Species' and 'Turtle' datasets, pairwise Steel-Dwass tests were performed to test whether substitution rate changes could be explained by phylogenetic history, and to test whether genes classes differed in their substitution rates within taxonomic groups.

For each gene, in order to identify outlier phylogenetic branches experiencing faster substitution rates relative to all other branches ('Fast Branches'), a branch-specific Z score was calculated per gene as:

$$Z = \frac{SSM(Rate\ of\ branch) - \overline{SSM}\ (Rate\ mean\ for\ gene)}{\sigma\ SSM\ (Rate\ STDEV\ for\ gene)}$$

Within each taxonomic group, a separate Z-score analysis was carried out to identify outlier genes experiencing higher substitution rates relative to all genes for that group ('Fast Genes'), where a gene-specific Z score was calculated as:

$$Z = \frac{SSM(Rate\ mean\ for\ Gene) - \overline{SSM}\ (Rate\ mean\ for\ all\ genes)}{\sigma\ SSM\ (Rate\ STDEV\ for\ all\ genes)}$$

Both sets of Z-scores were assessed at α=0.05 (threshold $Z > 1.644$). This analysis was implemented using Microsoft Excel 2010.

For the 'Turtle' dataset, Wilcoxon/Mann-Whitney U tests (2-way comparisons) and Steel-Dwass Tests (3-way comparisons) were performed to investigate whether phylogenetic branches characterized by a sex determination transition (TSD-to-GSD, or GSD-to-TSD) differed in substitution rate relative to non-transitional branches or to each other. Each branch on the turtle phylogeny was scored as transitional or non-transitional based on two proposed hypotheses of the evolution of sex determination in turtles: (1) A hypothesis which reconstructs five evolutionary transitions in turtles, all from the ancestral TSD condition to a derived GSD condition (Sabath *et al.* 2016); and (2) a hypothesis that reconstructs five

transitions from TSD-to-GSD (two of which overlap with the previous hypothesis) plus two reversals back to TSD from GSD in the lineages represented by *C. insculpta* and *P. expansa* (Valenzuela and Adams 2011) (Figure 2.2). Considering both hypotheses, the focal taxa included in this study encompass representative species for four of the five TSD-to-GSD transitions predicted under each hypothesis. These tests were implemented using JMP, Version 12.0.1 (SAS Institute Inc., Cary, NC, 2015).

2.4 Results

Neutral substitution rates (at third codon position) are slowest in turtles and crocodilians among vertebrates, and fastest in Trionychia among turtles

For the 'All Species' dataset, the analysis across all genes showed significant differences in neutral substitution rates at the third codon position among taxonomic groups (Table 2.3, full statistical results in Table B.2). Pairwise comparisons using the Steel-Dwass test revealed a higher substitution rate in mammals than in birds, turtles and crocodilians ($p<0.003$), while crocodilians and turtles exhibited a similar and lower rate than all other groups (Table 2.3).

Turtle branches were also analyzed separately using the 'Turtles' dataset which are more complete alignments than the "All Species" dataset (they contains fewer gap positions), and because mammals, squamates, and birds exhibited significantly faster overall substitution rates relative to turtles which could obscure biologically important shits among turtles. Species-level comparisons revealed that turtle sub-groups also differed in neutral substitution rates , which were slower in Emydidae (*Glyptemys, Chrysemys,* plus *Trachemys)* than in other turtles, while rates in Trionychia (*Apalone, Pelodiscus,* plus *Carettochelys*) were faster

than in Americhelydia (*Staurotypus* plus *Chelonia*) but similar to Pleurodiran turtles (*Emydura* plus *Podocnemis*) (Table 2.4, full statistical results in Table B.3).

Overall nucleotide and amino acid substitution rates resemble neutral rates

Considering all codon positions, for the 'All Species' dataset the differences in nucleotide substitution rate among taxonomic groups resemble those for the third codon position, except that across all codon positions turtle genes evolved at a significantly faster rate than in crocodilians ($p=0.0487$) (Table 2.3, full statistical results in Table B.4). At the amino acid level, mammals, birds, and squamates formed a group with substitution rates ~3X faster than turtles and crocodilians ($p<0.005$) (Table 2.3, full statistical results in Table B.5).

Likewise for the 'Turtles' dataset, overall nucleotide substitution rate mimicked the neutral substitution rate differences, which was slower overall in Emydidae than in other turtles, whereas Trionychia exhibited a faster rate than Americhelydia (Table 2.5, full statistical results in Table B.6). At the amino acid level, the rates in Trionychia and Pleurodiran turtles were faster than in Emydidae, and no other differences were significant (Table 2.5, full statistical results in Table B.7).

"Fast branches" experience higher substitution rates in more genes than other branches

For each gene alignment, certain branches on the phylogenetic tree (Figure 2.1) display significantly faster nucleotide substitution rates relative to the overall average rate for that gene (Table B.8). This was true for several root branches across many genes (12 of 15 genes at the root of placental mammals and Iguania lizard group, and 10 of 15 genes at the root of reptiles), whereas in turtles this was true for a smaller subset of genes. For the 'All Species' dataset, 11 of 48 branches in the phylogenetic tree exhibited a significantly faster nucleotide substitution rate relative to other branches for at least one gene (Table B.8).

Despite the lower nucleotide substitution rate observed in turtles as a whole, two turtle branches display a significantly faster nucleotide substitution rate for one gene each (*Lhx9* in *Carettochelys insculpta*; *Sox9* in the Trionychia root branch) (Table B.8).

Somewhat similar results were obtained at the protein level, where 16 of 48 branches had at least a single gene with a faster than average amino acid substitution rate (Table B.9). This list was topped by the root of placental mammals and Iguania lizards (7 of 15 genes each), and the reptile root branch (6 of 15 genes). In turtles, additional fast branches were detected at the amino acid level than at the nucleotide level. Indeed, four turtle branches exhibit faster amino acid substitution rates for at least one gene: the Trionychia root branch (*Esr1, Hsf2, Sox9*), the turtle root branch (*Cirbp*), *Carettochelys inscultpta* (*Lhx9*), and the Americhelydia root branch (*Wt1*) (Table B.9).

Genes on Z (but not X) sex chromosomes evolve faster at the amino acid level

Three target genes in this study, *Ar, Dmrt1,* and *Ctnnb1,* are sex-linked (located in the X or Z chromosomes) in certain focal taxa, and their protein sequences evolve faster when they are Z-linked than on taxa where they are autosomal. For instance, *Dmrt1* is Z-linked in birds (Nanda *et al.* 1999), and three of the five avian branches (*Taeniopygia guttata, Falco peregrinus*, and the Neoaves root branch) have a significantly faster amino acid substitution rate than all other branches (although interestingly, this faster rate is absent in chicken [*Gallus gallus*] or the bird root branch) (Table B.9). Second, *Ctnnb1* is Z-linked in snakes (Vicoso *et al.* 2013), and *O. hannah* snakes show a significantly faster amino acid substitution rate relative to all other species, including the other snake examined, *Python molurus*. In contrast, *Ar* is X-linked in mammals (Ross *et al.* 2005), yet no mammalian branch exhibits faster amino acid substitution rates (Table B.9).

Turtle stand-alone analysis revealed additional evolutionary differences in rates of molecular evolution

Faster than average nucleotide substitution rates were found in 6 of 18 branches relative to other turtle branches for at least one gene (Table B.10). Indeed, all genes except for *Lhx9* evolved at a faster than average rate in the Trionychia root branch, while this was true only for up to 2 genes in any other branch. At the protein level, 9 of 18 branches exhibited a faster amino acid substitution rate for at least one gene (Table B.11). The Trionychia root branch showed this pattern for 11 of 15 genes, followed by *Podocnemis expansa* and *Carettochelys insculpta* for 3 of 15 genes (*P. expansa: Cirbp*, *Hsf2*, and *Wt1*; *C. insculpta*: *Ar*, *Lhx9*, and *Wnt4*).

Hormone-signaling genes evolved faster in most vertebrates and *Srd5a1* in mammals and some turtles

Most gene classes display similar rates of substitution at the nucleotide and amino acid levels with some noticeable exceptions. Namely, for the 'All Species' dataset nucleotide sequences of hormone signaling genes evolve faster than transcription factors in squamates (Table B.12), whereas their amino acid sequences evolved faster than all other gene classes in squamates and turtles, and only faster than transcription factors in crocodilians (Table B.13). For the 'Turtles' dataset, the nucleotide sequences of hormone signaling genes evolve faster than temperature signaling genes in Emydidae (Table B.14), and their amino acid sequences evolve faster than the WNT signaling genes in the Americhelydia, Emydidae and Trionychia turtles, and also faster than the temperature signaling genes in Trionychia (Table B.15).

When broken down to the level of individual genes, few genes stand out as experiencing faster than average substitution rates and this varied by dataset and sequence

type. For instance, no such gene was detected using the 'All Species' nucleotide dataset (Table B.16), whereas for the 'Turtles' dataset, *Srd5a1* showed faster evolution in the Americhelydia turtles relative to other genes for that group ($p<0.02$) (Table B.17). At the amino acid level, DMRT1 evolved faster in birds ($p<0.01$) and SRD5A1 in mammals ($p<0.05$) (Table B.16), whereas for the 'Turtles' dataset, SRD5A1 evolved faster in both the Americhelydia and Trionychia groups ($p<0.025$) (Table B.17).

GSD-to-TSD reversal hypothesis helps explain molecular evolution in turtles

We examined sequence data for five turtle species with temperature-dependent sex determination (TSD) and five turtles with genotypic sex determination (GSD) under two contrasting evolutionary hypotheses that explain the evolutionary history of turtle SDM transitions (Figure 2.2). No nucleotide or amino acid substitution rate differences were detected between transitional and non-transitional branches (Prob > ChiSq > 0.4) under the hypothesis that turtles underwent exclusively transitions from TSD-to-GSD (Sabath *et al.* 2016), (Figure 2.3). However, higher nucleotide and amino acid substitution rates were detected in transitional branches (Prob > ChiSq < 0.0001) under the hypothesis that *Carettochelys insculpta* and *Podocnemis expansa* represent lineages where a GSD-to-TSD reversals occurred (Valenzuela and Adams 2011) (Figure 2.3). Broken down by individual genes, nucleotide sequences of *Dmrt1*, *Hsf2*, *Sox9*, and *Wnt4* evolved faster in transitional branches relative to non-transitional branches ($p<0.05$), and the same was true for the amino acid sequences of HSF2, RSPO1, and SOX9 ($p<0.05$). Moreover, the GSD-to-TSD reversal branches exhibit significantly faster nucleotide and amino acid substitution rates relative to non-transitional and to TSD-to-GSD transitional branches ($p<0.04$) across all genes (Figure 2.3). Notably, no significant differences in the amino acid substitution rate were detected

across genes between non-transitional branches and TSD-to-GSD branches (p=0.0862, Figure 2.3), suggesting that the trend in the transition-dependent substitution rate at the protein level is driven by the GSD-to-TSD reversal branches. Further gene-specific comparisons were precluded because there are only two GSD-to-TSD branches and three TSD-to-GSD branches.

Potentially functional evolutionary changes accrued in turtle proteins

We searched for target proteins in our dataset that either (1) evolve faster in a transitional turtle branch compared to all other species or, (2) evolve faster in transitional versus non-transitional turtle branches, and subjected them to secondary protein structure predictions in order to identify amino acid substitutions in transitional turtle branches that may induce functional changes. Structural changes within UniProt functional domains were detected in the HSF2 protein, which evolved faster in the Trionychia turtles relative to all other species, and in turtle transitional branches relative to non-transitional branches. Likewise, we detected structural changes within functional domains of the RSPO1 protein which evolved faster in transitional branches relative to non-transitional branches, and in the LHX9 protein, which evolved faster in *C. insculpta* relative to all other species (Figure 2.4).

Specifically, a series of amino acid substitutions that change the secondary protein structure were identified in turtle transitional branches in both the positive and negative transactivation domains of the HSF2 protein, which form its C-terminal domain and contribute to transcriptional activation of target genes (Yoshima *et al.* 1998). In the positive transactivation domain 1 which binds co-regulators during cellular stress (Figure 2.4A), *P. expansa* has a single substitution predicted to cause the loss of two helices present in all other species, while in a separate site *C. insculpta* has a single substitution that is predicted to join

two shorter helices found in all other species into a single longer helix. In the associated negative regulator domain, which keeps HSF2 inactive during non-stress, *E. macquarii*, *P. expansa*, and *C. insculpta* share a unique predicted secondary structure characterized by the addition of two short alpha helices and notably, these result from three different and independent mutations in each lineage. In the positive transactivation domain 2 (Figure 2.4B), *S. triporcatus* accrued several amino acid substitutions predicted to drastically change its secondary structure from that of other turtles by inducing multiple alpha helices, while *G. insculpta* has a single substitution which also induces a novel helix among turtles in this domain. In the corresponding negative regulator domain *P. expansa*, *A. spinifera*, and *P. sinensis* experienced a reduction or loss of alpha helices that exist in all other turtle species.

The 3' end of the downstream LIM domain of LHX9 in turtles generally contains a 10-amino acid long alpha helix which is disrupted in *P. expansa* and *C. insculpta*, the representatives of the two putative GSD-to-TSD reversals in turtles (Figure 2.4C). In *P. expansa* an inserted short beta strand splits the helix, while in *C. inculpta* additional substitutions are predicted to yield a unique secondary structure among turtles with a highly truncated alpha helix and the expansion of the beta sheet predicted in *P. expansa*. Further upstream substitutions in *C. insculpta* cause the shifting of a beta sheet relative to all other turtles. Additionally, the 5' end of the upstream LIM domain in *C. insculpta* also contains a split helix, where in all other turtles the helix is unbroken.

Finally, the thromospondin-1 domain of the RSPO1 protein in two XX/XY turtle species (*E. macquarii* and *S. triporcatus*) shares a secondary structure unique among turtles, lacking an alpha helix in the C-terminal end of the domain. Notably, this pattern evolved through two independent amino acid substitutions at separate sites in this domain in each

lineage (Figure 2.4D). *S. triporcatus* has also accumulated substitutions in the second of two upstream Furin-like repeat domains which cause both the truncation of an alpha helix which is longer in all other species, as well as the insertion of a helix at the 3' end of the domain which is unique to *S. triporcatus.*

2.5 Discussion

Numerous independent transitions between GSD and TSD have occurred throughout vertebrate evolution (Valenzuela and Lance 2004, Beukeboom and Perrin 2014), yet, the gene network that regulates sex determination has remained remarkably similar in its composition, whereas the roles of this shared set of genes have shuffled as SDMs evolve (Cutting *et al.* 2013, Valenzuela *et al.* 2013). Here we advance our understanding of other changes that have accompanied SDM transitions, as we detected substantial molecular evolution among major vertebrate clades in the coding sequence of a subset of genes involved in sex determination. Among those, we highlight in particular the changes identified in turtle lineages which have experienced SDM transitions, including predicted structural changes of potential functional importance.

Genes in the sex determination network evolve slowly in turtles and crocodilians but show an acceleration in Trionychia turtles

Among the five major vertebrate clades examined, the target genes evolve at a significantly lower rate in the turtles and crocodilians, a pattern observed at both the nucleotide and amino acid levels and accentuated in crocodilians relative to turtles when considering all nucleotides (Table 2.3). These observations agree with recent genome-wide reports for turtles, Anolis lizards and crocodilians (Fujita *et al.* 2011, Shaffer *et al.* 2013, Green *et al.* 2014), but counter the expectation that the evolutionary lability of sex-

determining mechanisms in turtles when compared to mammals and birds might be associated with higher molecular evolution overall. The lower genome-wide substitution rates of crocodilians and turtles have been hypothesized to derive from a combination of their slower metabolic rates and longer generation times relative to endothermic mammals, birds, and also to most squamates (Shaffer *et al.* 2013, Green *et al.* 2014).

Despite the relatively low rates of molecular evolution in turtles, significant differences in evolutionary rates were detected among turtle lineages. For instance, both the nucleotide and amino acid substitution rates of the target genes in Emydidae (Figure 2.2, Table 2.2) were slower than in other turtle lineages (85%-244% depending on the dataset used, Table 2.4). Furthermore, the branch leading to the Trionychia turtles (softshell plus pig-nose turtles) (Figure 2.2, Table 2.2) stood out by its faster nucleotide and amino acid substitution rate relative to the turtle average for most genes (Tables B.10 + B.11), revealing that a major acceleration in molecular evolution accompanied the divergence of this morphologically-intriguing lineage. However, whether this hastened molecular evolution is functionally linked to SDM transitions and restricted to genes in the sex-determining network, or whether it is a more generalized phenomenon associated with other ecological, genomic, or morphological innovations in this clade remains to be tested. The higher neutral rate of evolution in Trionychia compared to other turtles supports the notion that this lineage follows a unique evolutionary trajectory. Notably, members of this turtle family underwent a cytogenetic revolution, as they contain the species with the highest numbers of chromosomes among turtles (*A. spinifera/P. sinensis*: 2n=66; *C. insculpta*: 2n=68) (Bickham and Legler 1983, Sato and Ota 2001, Badenhorst *et al.* 2013). It is tempting to speculate that perhaps the

same drivers might be responsible for the acceleration of molecular evolution and chromosomal fissions observed in this group.

Fast branches: Genes in the sex determination network evolve faster on root branches leading to major taxonomic diversification events

Major clade root branches, such as the branches leading to the placental mammals, Iguania lizards, and the root of all reptiles had the most genes with a faster than average substitution rate for both nucleotides and amino acids (Tables B.8 + B.9), suggesting that the sex determination network experienced major molecular changes during significant taxonomic diversification events. Although appealing, further research is warranted to test the biological relevance of this hypothesis under a more extensive taxonomic sampling than is currently possible, as genome level data for additional species becomes available. Our results on the evolutionary rate of sex determining genes among model and non-model vertebrates complement our knowledge of the evolution of sex-chromosome content and gene synteny that is well documented among diverse vertebrate clades (Montiel *et al.* 2016b, Rovatsos *et al.* 2016, Ezaz *et al.* 2017, Graves 2017).

Protein sequences evolve faster on sex chromosomes, but not always

At first glance, sex-linkage effects were observed at the amino acid level for a number of species. For example, the DMRT1 protein evolved faster in birds where this gene is Z-linked (Nanda *et al.* 1999), in particular in the branches for the zebra finch (*T. guttata*), the peregrine falcon (*F. peregrinus*), and their root branch (Neoaves) (Table B.9). Interestingly however, the root branch for all birds and the branch leading to chicken (*G. gallus*) have substitution rates more comparable to the their autosomal homologs in the other taxa examined, suggesting that DMRT1 evolution has been significantly accelerated in the Neoaves relative to the more basal birds, and that some driver specific to Neoaves other than

the evolution of avian sex determination might be causal. Second, beta-catenin (CTNNB1) is a highly conserved signaling protein, exhibiting the lowest average amino acid substitution rate of all target genes such that no substitutions were detected in 33 of 48 branches. However, *Ctnnb1* is Z-linked in snakes (Vicoso *et al.* 2013) and its protein sequence evolved faster in the king cobra (*O. hannah*) (Table B.9) than in the Indian python *P. molurus* which displayed only the substitutions accumulated during the diversification of the Serpentes as a group. These results in birds and snakes suggest that the differential molecular evolution is not uniform for sex-linked genes but that lineage-specific effects are also at play. Furthermore, the other sex-linked protein in the analysis, the mammalian X-linked AR (Ross *et al.* 2005), did not show evidence of 'Fast X'-type evolution (Table B.9). Combined, all these findings refute the generality even among closely-related taxa of the 'Fast X' and 'Fast Z' hypotheses, which state that genes residing in sex-limited (non-psuedoautosomal) portions of sex chromosomes should evolve faster than their autosomal counterparts due to their smaller effective population size relative to autosomes (Charlesworth *et al.* 1987, Mank *et al.* 2007, Bachtrog *et al.* 2011). Instead, our data suggest that gene-specific, lineage-specific, or other chromosome-specific effects may override the evolutionary dynamics driven by sex-linkage alone. Alterative drivers have been proposed, such as GC content and expression levels that may explain the difference in molecular evolution rate in mammalian sex-linked genes (Nguyen *et al.* 2015).

Fast genes: The hormone signaling gene *Srd5a1* evolves faster in more lineages than other genes

Hormone signaling genes were the only gene class exhibiting a consistent tendency for faster evolution at the nucleotide or amino acid levels (Tables B.12 – B.15). Notably, in squamates and turtles, hormone signaling protein sequences evolved faster than all other

gene classes (Table B.13), whereas in other groups this tendency was seen mostly when compared to the slowest gene class. On a gene by gene level, *Srd5a1* exhibited a faster nucleotide substitution rate in the Americhelydia turtles (Figure 2.2, Table 2.2, Table B.17) and a faster amino acid substitution rate in mammals along with the Americhelyida and Trionychia turtles relative to all other genes for those groups (Table B.17). *Srd5a1* encodes a 5-alpha reductase protein which converts weaker androgens to more potent androgen compounds (Chang *et al.* 2011), a critical step in sexual differentiation towards a male fate (Urbatzka *et al.* 2007). Deficiencies in human 5-alpha reductase activity are linked to partial or complete sex reversal of genotypic males (XY) at birth (exhibiting intersex or even female phenotype due to low levels of potent androgens) that is reverted partially at puberty when appropriate hormones are synthesized (Wilson *et al.* 1993). We hypothesize that the increased rate of molecular evolution of *Srd5a1* observed here may reflect more profound changes in the hormone signaling pathways or how androgens are utilized during sexual development among taxa, but this hypothesis requires further testing.

Contrasting hypotheses of turtle sex determination mechanism evolution affect the interpretation of transition-dependent shifts in substitution rates

Our 'Turtle' dataset permitted us to examine the influence of SDM transitions on molecular evolution but the same analyses were precluded for the 'All Species' dataset. The evolutionary history of SDM transitions among the turtles remains uncertain, and current contrasting evolutionary hypotheses (Figure 2.2) were evaluated here (Valenzuela and Adams 2011, Sabath *et al.* 2016). First, some studies reconstruct all transitions in turtle sex determination as occurring from a TSD ancestor to a derived GSD state (Sabath *et al.* 2016). When we examine molecular evolution under this hypothesis, no significant change in substitution rate was seen between the transitional and non-transitional branches at either the

nucleotide or amino acid levels. Such observations counter the notion that accelerated molecular evolution in the genes examined here is associated with transitions in sex determination. Whether genes in this network other than those analyzed here might exhibit such an association remains to be tested. Alternatively, miniscule genetic changes such as point mutations rather than more dramatic shifts in substitution rate may alter sex determination profoundly and affect SDM evolution. For instance, the sex determination network of *Takifugu* fish was hijacked by a single nucleotide polymorphism in the kinase domain of the gene *Amhr2* (Kamiya et al. 2012).

Second, another study reconstructed seven total transitions in turtles, five TSD-to-GSD transitions (three of which overlap with the previous hypothesis) plus two GSD-to-TSD 'reversals' in the lineages represented by *Carettochelys insculpta* and *Podocnemis expansa* (Valenzuela and Adams 2011). Further, these transition braches display ~20X faster rate of evolution of chromosome number (Valenzuela and Adams 2011), suggesting that although GSD-to-TSD transitions might be rare, turtle genomes are generally more labile in transition branches. Interestingly, under this hypothesis we found that transitional branches irrespective of their direction (TSD-to-GSD or GSD-to-TSD) had a faster nucleotide substitution rates than non-transitional branches (Figure 2.3), but this did not translate into faster protein sequence evolution in TSD-to-GSD branches. However, genes in the GSD-to-TSD 'reversal' branches evolved significantly faster at the nucleotide (~20% faster) and protein sequence levels (~10% faster) relative to the TSD-to-GSD transitional branches. This result is concordant with the idea that regaining TSD during GSD-to-TSD transitions requires a more dramatic rewiring of the sex determination network, perhaps reflecting the difficulty of escaping the evolutionary trap of sex chromosomes (Pokorna and Kratochvíl 2009), and

helping explain why these transitions are considerably rarer than TSD-to-GSD transitions at least in turtles.

When broken down by gene, the nucleotide sequences of *Dmrt1, Hsf2, Sox9,* and *Wnt4*, and the protein sequences of HSF2, RSPO1, and SOX9 evolved faster in transitional branches. HSF2 is a member of the heat-shock family of proteins which is able to transduce temperature signals into transcriptional responses (Sarge *et al.* 1991). RSPO1 is an integral signaling molecule in the WNT signaling pathway regulating expression of *Ctnnb1* and *Wnt4,* which helps direct proper ovarian development (Parma *et al.* 2006). SOX9 is a male-promoting transcription factor in most vertebrates and in many taxa it is critical for both the initiation and maintenance of the male developmental pathway (da Silva *et al.* 1996). Because of their identity and function, the accelerated amino acid evolution of these genes in transitional turtle branches renders them important candidates for further study, as they span the key processes of temperature-sensing, hormonal regulation, and gonadal differentiation that underlie sexual development.

Secondary protein structure change in transitional turtle branches in functionally important domains

Our results revealed structural changes that may play a role in SDM evolution in three proteins which showed an accelerated rate of amino acid evolution in transitional turtle branches relative to all species or in transitional relative to non-transitional branches within turtles. These findings should help guide future studies to determine the effect that these substitutions may have on protein function, if any.

First, the HSF2 protein is involved in stress response, temperature-signal transduction, and spermatogenesis (Yoshima *et al.* 1998, Garolla *et al.* 2013). The HSF2 C-terminal domain contains two pairs of transactivation domains that modulate transcription

during cell stress through co-regulatory ligand binding, plus two pairs of negative regulation domains that maintain HSF2 as inactive when cells are not stressed (Yoshima *et al.* 1998). These four sub-domains exhibit predicted secondary structure changes with potentially functional consequences specific to transitional turtle branches. For instance, the two representatives of the putative GSD-to-TSD transition lineages (*P. expansa* and *C. insculpta*), each have separate single amino acid substitutions in different sites of the HSF2 first transactivation domain that lead to structural changes that are unique among turtles (Figure 2.4A). Specifically, *P. expansa*'s single substitution deletes two short alpha helices present in all other turtles, while *C. insculpta*'s creates one longer alpha helix where all other taxa have two helices separated by a beta sheet. On the other hand, *E. macquarii* (representing a TSD-to-GSD transition), *P. expansa,* and *C. insculpta* share a similar secondary structure in the negative regulator domain of the this first transactivator domain compared to all other turtles, but caused by different substitutions in each lineage suggesting that they result from convergent evolution rather than shared ancestry. Additionally, the HFS2 second transactivation domain of *Staurotypus triporcatus* (representing a TSD-to-GSD transition) accumulated multiple substitutions which create several novel alpha helices not present in any other turtle (Figure 2.4B), and *Glyptemys insculpta* (representing a TSD-to-GSD transition) exhibits yet another novel alpha helix in this domain. Finally, *Apalone spinifera* and *Pelodiscus sinensis* (representing a TSD-to-GSD transition) along with *P. expansa* show a drastic reduction or loss of alpha helices in the negative regulator domain for HFS2's second transactivator that are present in all other turtle species. We hypothesize that because the regulatory role of HSF2 is self-modulated via the negative regulator domains, changing depending on the degree of cell stress (e.g. high temperature), the structural changes

described here could impact how HSF2 is deployed during bouts of high temperature (changes in transactivation domains), and perhaps also under more moderate temperatures (changes in negative regulator domains), a hypothesis which requires future functional testing.

Second, LHX9 is a transcription factor which is essential for the formation and general development of the bipotential gonad and typically shows equal levels of activity in both developing male and female embryos among vertebrates (Birk *et al.* 2000, Ottolenghi *et al.* 2001, Mazaud *et al.* 2002, Oréal *et al.* 2002, Barske and Capel 2010). It activates the NR5A1 protein (Shima *et al.* 2012) whose transcription is evolutionarily labile across vertebrates (Valenzuela *et al.* 2013). In order to bind co-regulator proteins, LIM-containing transcription factors possess tandem protein-binding LIM domains separated by a linker sequence (Rétaux *et al.* 1999). Here we found that *C. insculpta*'s LHX9 is characterized by two structural changes unique among vertebrates (Figure 2.4C). Namely, *C. insculpta*'s downstream LIM domain has a smaller conserved C-terminal alpha helix and beta strands shifted within the center of the domain, and the N-terminal helix in the upstream LIM domain is also disrupted. This raises the question as to whether such changes to the protein:protein ligand-binding domain might operate as part of a shifting regulatory network where old interactions are lost, and new molecular relationships may be forged.

Finally, RSPO1 is a critical co-regulator of the WNT/CTNNB1 pathway, primarily involved in ovarian development through *Ctnnb1* activation and stabilization (Parma *et al.* 2006). RSPO1 is also a SOX9 antagonist, working to both promote ovarian development while silencing male developmental pathways (Chassot *et al.* 2008). RSPO1, as other – spondin protein family members, contains a thrombospondin-1 repeat domain which appears

unnecessary to modulate *Ctnnb1* (Parma et al. 2006), and its precise role is still unclear. Two independent single amino acid substitutions in the TSP1 domain in *E. macquarii* and *S. triporcatus* cause the convergent loss of an alpha helix found in all other turtles (Figure 2.4D). While the role of the TSP1 domain remains elusive, the furin-like repeat domains appear integral to R-spondins' ability to modulate the *Ctnnb1*/*Wnt* pathway (Kim *et al.* 2006, Kim *et al.* 2008). Intriguingly, in *S. triporcatus* a single amino acid substitution in the second of two tandem Furin-like repeat domains truncates an alpha helix and adds a beta strand, a predicted structure unique among all vertebrates examined. We hypothesize that these changes may impact RSPO1-mediated regulation of the *Ctnnb1*/*Wnt* pathway in this XX/XY turtle, an important pathway in ovarian development, providing another target for future functional testing.

2.6 Conclusions

Evolutionary transitions in sex-determining mechanism require at least a partial rewiring of the vertebrate sex determination network, potentially involving the molecular evolution of protein-coding regions of genes in its regulatory network. Among the genes that experienced significant molecular evolution as identified in our analyses, hormonal signaling genes (and *Srd5a1* in particular) and HSF2 emerged as being dramatically altered during SDM shifts in turtles and worthy of further study. Our work revealed a significant acceleration of substitution rates during putative GSD-to-TSD reversals than during TSD-to-GSD transitions or SDM stasis. Evolutionary state reconstruction has a major influence on our understanding of SDM evolutionary history, as no differences on the molecular evolution of this regulatory network were detected when GSD-to-TSD reversals were ignored under

one of the evolutionary hypotheses evaluated. These discrepancies highlight the importance of reliable methods for ancestral state reconstruction, and the need for additional genomic and SDM data across the tree of life to enable robust analyses. Our data predicted changes in the functional domains of proteins in the sex determination network which may play a role in SDM evolution in turtles, providing targets for future research. No single substitution or structural change was present in all of the studied GSD turtle species, and some transitional branches appear to evolve much faster than others, a testament to the independent nature of turtle SDM evolution and the many trajectories taken by nature to produce males and females.

2.7 Acknowledgements

We greatly appreciate the assistance in specimen collection for *G. insculpta* (Jeff Tamplin), *E. macquarii* (Arthur Georges), and *C. insculpta* (Gabriel Rivera).

2.8 Table and Figures

Table 2.1: Target genes with varied roles in vertebrate sex determination examined in this study.

Gene Symbol	Gene Name	Gene Class
Ar	Androgen Receptor	Hormone signaling
Cirbp	Cold-inducible RNA Binding Protein	Temperature signaling
Ctnnb11	Beta Catenin	WNT signaling
Cyp19a1 (Aromatase)	Aromatase	Hormone signaling
Dmrt1	Doublesex And Mab-3 Related Transcription Factor 1	Transcription Factor
Esr1	Estrogen Receptor 1	Hormone signaling
Esr2	Estrogen Receptor 2	Hormone signaling
Hsf2	Heat Shock Factor 2	Temperature signaling
Lhx9	LIM Homeobox 9	Transcription factor
Nr5a1 (Sf1)	Steroidogenic Factor 1	Transcription factor
Rspo1	R-spondin 1	WNT signaling
Sox9	SRY Box 9	Transcription factor
Srd5a1	Steroid 5 Alpha-Reductase 1	Hormone signaling
Wnt4	Wingless-Type MMTV Integration Site Family, Member 4	WNT signaling
Wt1	Wilms Tumor 1	Transcription factor

Table 2.2: Species used in this study along with data sources. DNA-NGS = in house generated Next-Generation DNA sequencing data (short read Illumina HiSeq data).

Species	Taxonomic Group	NCBI Genome ID and Other Data Sources
Gallus gallus	Bird	Gallus_gallus-4.0
Falco peregrinus	Bird	F_peregrinus_v1.0 + SRA
Taeniopygia guttata	Bird	Taeniopygia_guttata-3.2.4 + SRA
Alligator mississippiensis	Crocodilian	AllMis0.2 + SRA
Alligator sinensis	Crocodilian	ASM45574v1 + SRA
Crocodylus porosus	Crocodilian	CroPor_comp1 + SRA
Gavialis gangeticus	Crocodilian	GavGan_comp1
Monodelphis domestica	Mammal (Marsupial)	MonDom5 + SRA
Homo sapiens	Mammal (Placental)	GRCh38
Mus musculus	Mammal (Placental)	GRCm38.p2
Anolis carolinensis	Squamate (Lizard)	AnoCar2.0 + SRA + TSA + Genbank
Gekko japonicas	Squamate (Lizard)	Gekko_japonicus_V1.1
Pogona vitticeps	Squamate (Lizard)	SRA
Ophiophagus hannah	Squamate (Snake)	OphHan1.0 + SRA
Python molurus	Squamate (Snake)	Python_molurus_bivittatus-5.0.2
Chelonia mydas	Turtle (Americhelydia)	CheMyd_1.0 + SRA
Staurotypus triporcatus	Turtle (Americhelydia)	DNA-NGS Data
Chrysemys picta	Turtle (Emydidae)	Chrysemys_picta_bellii-3.0.1 + RNA-Seq Data
Glyptemys insculpta	Turtle (Emydidae)	DNA-NGS Data
Trachemys scripta	Turtle (Emydidae)	SRA + Genbank mRNA
Emydura macquarii	Turtle (Pleurodira)	DNA-NGS Data
Podocnemis expansa	Turtle (Pleurodira)	RNA-Seq Data + SRA
Apalone spinifera	Turtle (Trionychia)	RNA-Seq + DNA-NGS Data
Carettochelys insculpta	Turtle (Trionychia)	DNA-NGS Data + Sanger Sequencing
Pelodiscus sinensis	Turtle (Trionychia)	PelSin_1.0 + SRA

Table 2.3: Overall substitution rates per million years of nucleotide and amino acid sequences across target genes in each taxonomic group, and relative rate (%) compared to turtles. For each comparison (third codon, all nucleotides and amino acids) statistically indistinguishable rates share bracketed letters.

	Mean of Substitutions/Site/MY across all Genes in All Species					
	Third Codon Position		All Nucleotides		Amino Acids	
	Substitutions per Myr	% from Turtles	Substitutions per Myr	% from Turtles	Substitutions per Myr	% from Turtles
Mammals	2.60E-03 [A]	387.5%	1.02E-03 [A]	328.4%	5.66E-04 [A]	241.9%
Squamates	2.26E-03 [AB]	323.3%	8.68E-04 [AB]	265.1%	5.24E-04 [A]	216.4%
Birds	1.73E-03 [B]	223.5%	6.64E-04 [B]	179.3%	3.76E-04 [A]	127.2%
Turtles	5.34E-04 [C]	-	2.38E-04 [C]	-	1.65E-04 [B]	-
Crocodilians	4.37E-04 [C]	-18.0%	1.67E-04 [D]	-29.6%	9.03E-05 [B]	-45.4%

Table 2.4: Overall substitution rates per million years of nucleotide and amino acid sequences for all genes examined in each focal turtle clade, and relative rate (%) compared to Emydidae. For each comparison (third codon, all nucleotides and amino acids), statistically indistinguishable rates share bracketed letters.

	Mean of Substitutions/Site/MY for all Genes in Turtles					
	Third Codon Position		All Nucleotides		Amino Acids	
	Substitutions per Myr	% from Emydidae	Substitutions per Myr	% from Emydidae	Substitutions per Myr	% from Emydidae
Trionychia	9.37E-04 [A]	244.5%	4.09E-04 [A]	219.5%	3.14E-04 [A]	231.6%
Pleurodira	5.34E-04 [AB]	96.3%	2.53E-04 [AB]	97.7%	1.70E-04 [A]	79.5%
Americhelydia	5.12E-04 [B]	88.2%	2.38E-04 [B]	85.9%	2.39E-04 [AB]	152.4%
Emydidae	2.72E-04 [C]	-	1.28E-04 [C]	-	9.47E-05 [B]	-

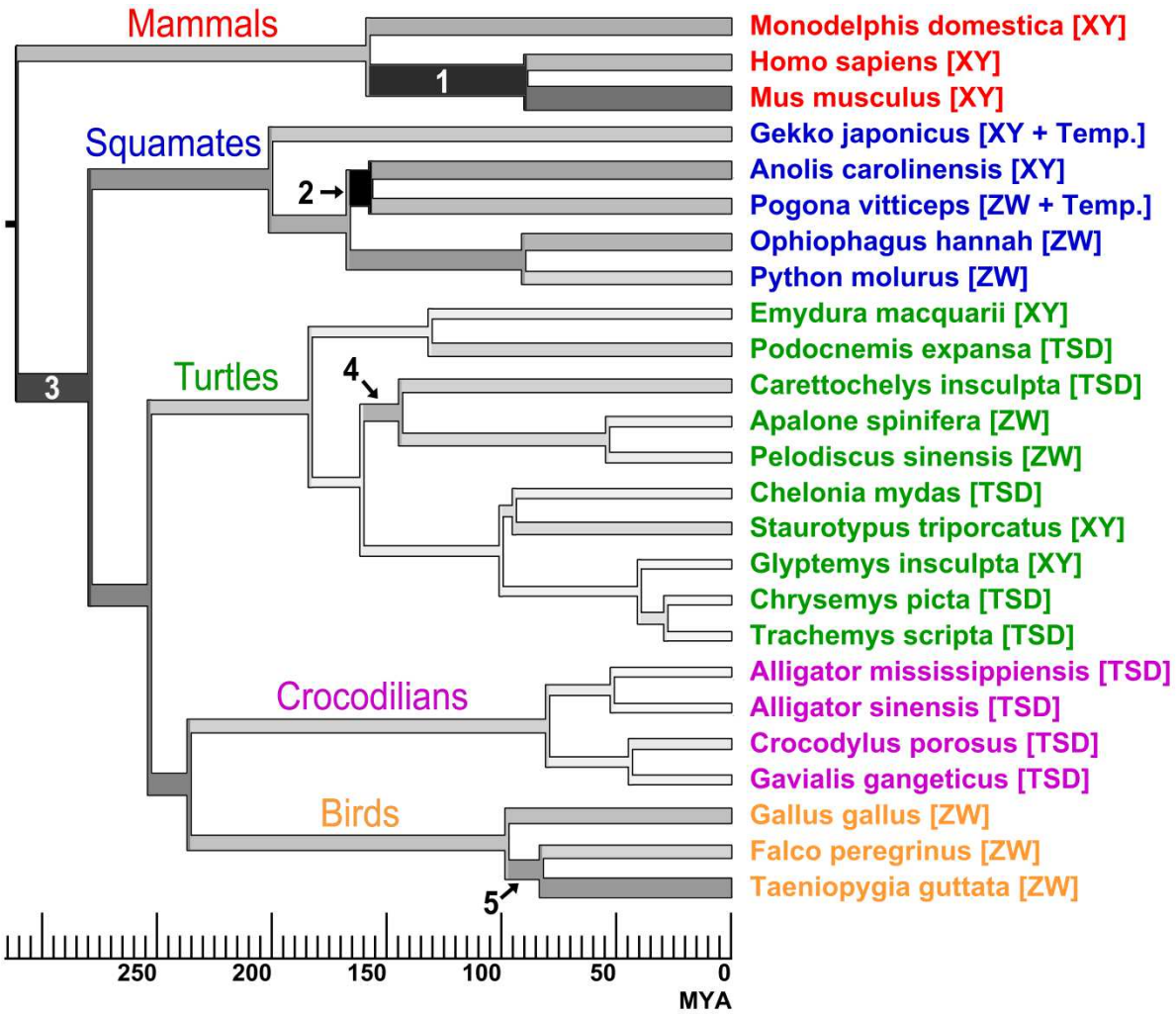

Figure 2.1: Phylogenetic relationships of twenty-five vertebrate species analyzed in this study. Branch lengths are scaled to TimeTree consensus divergence times. Branch thicknesses and darkness are scaled to average nucleotide substitution rates across all genes. Sex-determining mechanisms are denoted in brackets. Specific branches discussed in the text are numbered (1 = Placental mammal root branch; 2 = Iguania lizard root branch; 3 = Reptile root branch; 4 = Trionychia root branch; 5 = Neoaves root branch). Temp. = temperature sex reversal or TSD reports.

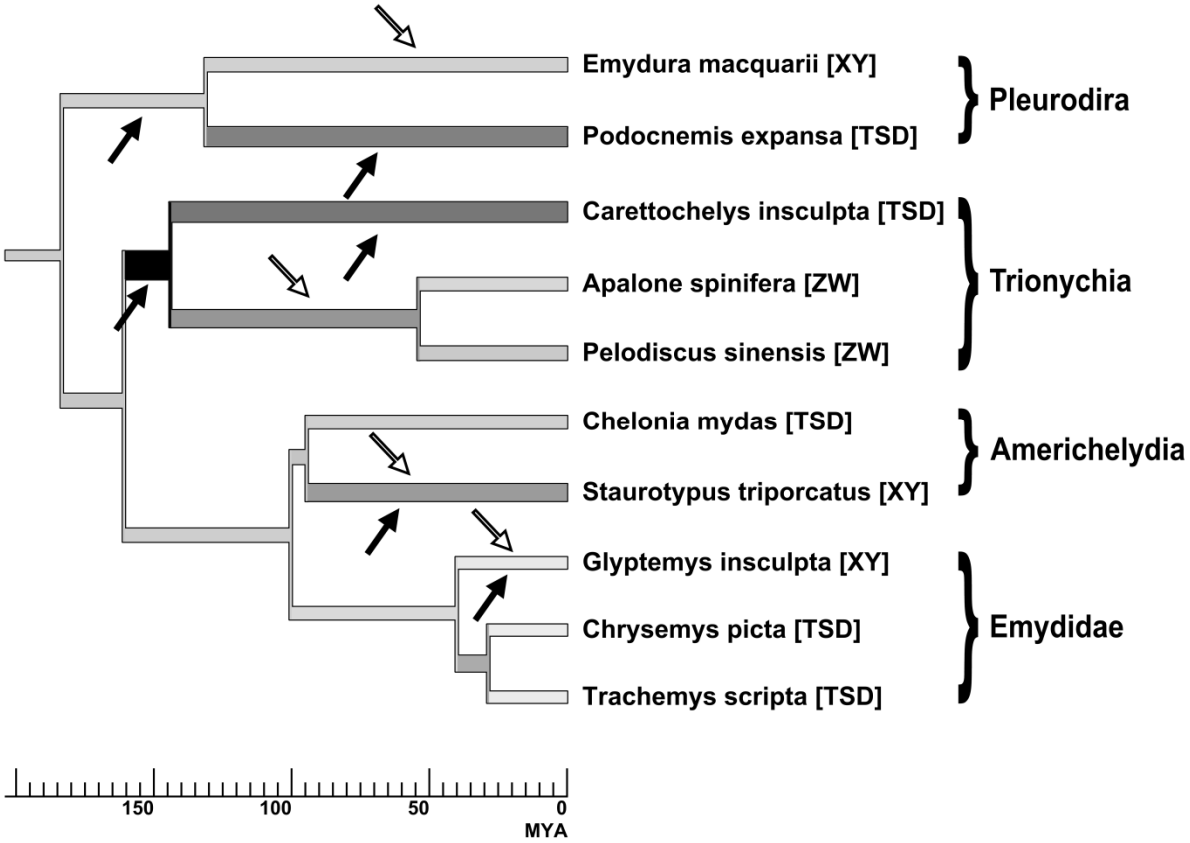

Figure 2.2: Phylogenetic relationships of ten turtle species analyzed in this study. Branch lengths, thicknesses and darkness as in Figure 2.1. Arrows indicate branches where a transition in sex determination mechanism was reconstructed to have occurred under two evolutionary hypotheses: (1) Open arrowheads above branches indicate transitions under Sabath et al. 2016 hypothesis where all transitions occur from TSD-to-GSD. (2) Closed arrowheads below branches indicate transitions under Valenzuela and Adams 2011 hypothesis, where *Carettochelys insculpta* and *Podocnmenis expansa* lineages underwent reversals from GSD to TSD.

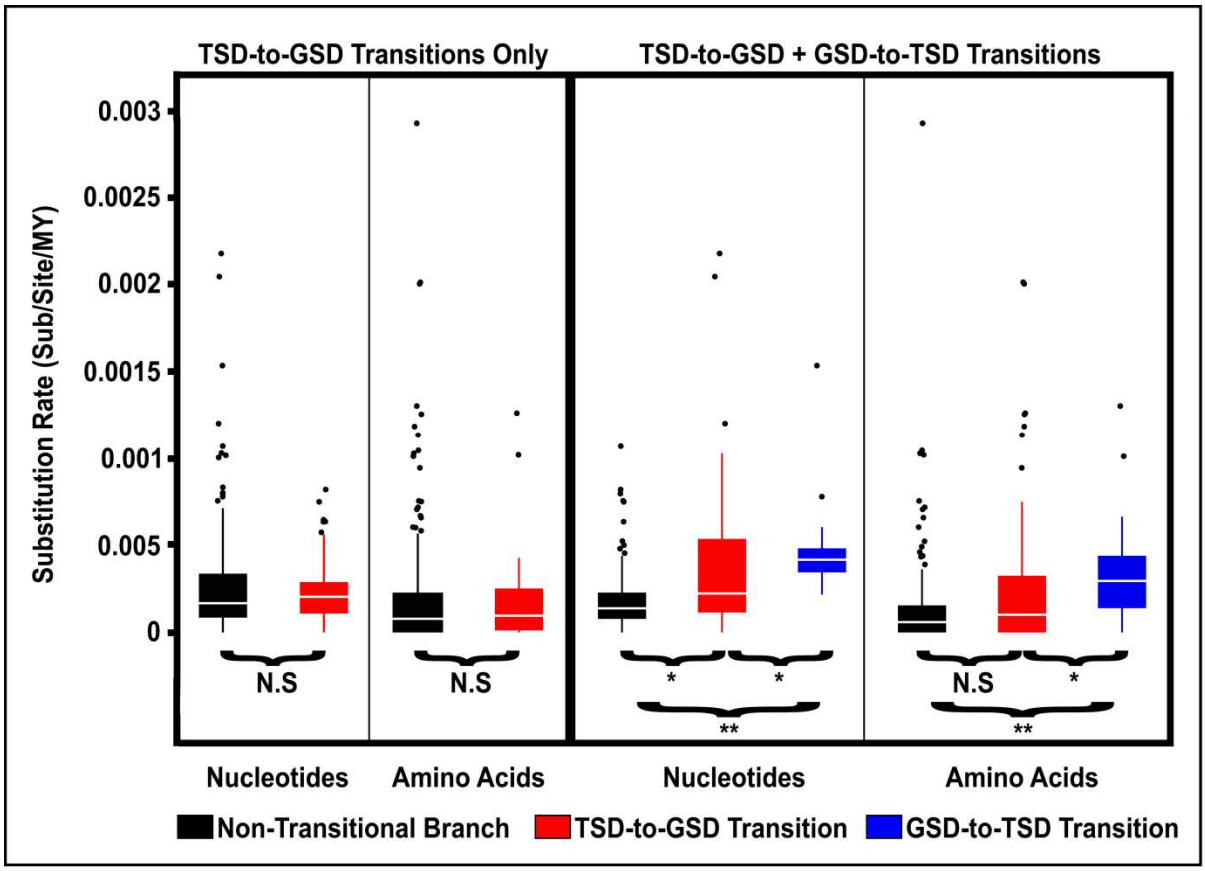

Figure 2.3: Acceleration in nucleotide and amino acid substitution rates across all genes are not correlated with turtle SDM transitions under Sabath et al. 2016 hypothesis, where all transitions occur from TSD-to-GSD, but are correlated at the nucleotide level under Valenzuela and Adams 2011 hypothesis, where *Carettochelys insculpta* and *Podocnmenis expansa* lineages underwent reversals back to TSD from a GSD ancestor. Under this hypothesis, proteins evolve faster in GSD-to-TSD branches than in both non-transitional and TSD-to-GSD branches, while proteins on TSD-to-GSD branches do not evolve significantly faster than non-transitional branches. (* = $p < 0.05$; ** = $p < 0.001$; N.S = Not Significant)

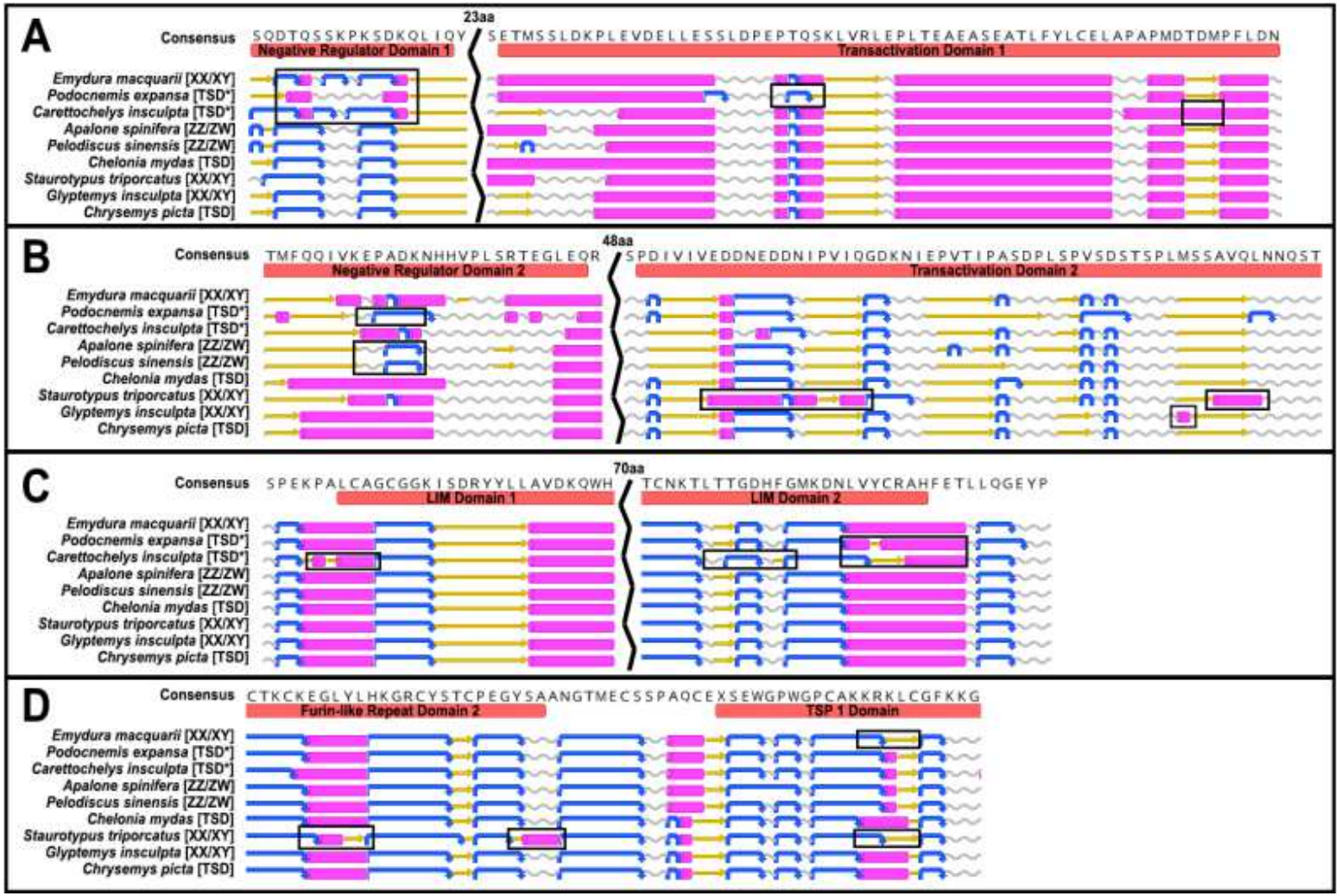

Figure 2.4: Turtle lineages which have undergone transitions in sex determination have accrued amino acid substitutions in function domains of sex determination genes. Boxes highlight focal regions with substitutions. Sex-determining mechanisms for each species noted in brackets. TSD* = GSD-to-TSD reversal branch under Valenzuela and Adams 2011. (A,B) Transactivation and negative regulatory domains of HSF2.. (C) LIM domains of LHX9. (D) FU (Furin-like repeat domains) and TSP1 domains of RSPO1. Zigzags denote trimmed sequence between domains. aa = Amino acids trimmed

CHAPTER 3: QPCR-BASED MOLECULAR SEXING BY COPY NUMBER VARIATION IN rRNA GENES AND ITS UTILITY FOR SEX IDENTIFICATION IN SOFT-SHELL TURTLES

Robert Literman, Daleen Badenhorst, Nicole Valenzuela

Modified from a paper published in *Methods in Ecology and Evolution* (Literman et al. 2014)

Author Contributions: RL conceived the project, collected the *Apalone spinifera* data, performed analyses on *A. spinifera* data, and co-wrote the manuscript with NV. NV performed analyses for *Chelydra serpentina* data. DB performed initial cytogenetic work.

Author Affiliations:

Department of Ecology, Evolution and Organismal Biology, 251 Bessey Hall, Iowa State University, Ames IA 50011

3.1 Abstract

Sex diagnosis is important in ecology, evolution, conservation biology, medicine, and food production. However, sex diagnosis is difficult in species without conspicuous sexual dimorphism or at life stages before such differences develop. This problem is exacerbated when the diagnostic trait is a continuous (non-discrete) variable to which general analytical methods are not commonly applied. Here we demonstrate the use of copy-number variation between males and females of the nucleolar organizing region (NOR) in the genome of *Apalone spinifera* softshell turtles, which we quantify by real-time PCR. We analyze these continuous data using mixture models that can be applied either in discriminant analysis when a subset of individuals of known sex is used as a training set, or in clustering

procedures when all individuals are of unknown sex. Using individuals of known sex, the discriminant analysis exhibited 100% accurate classification rate for both the training set and the test set. Classification rates were also 100% when using the clustering procedure to identify groups and classify individuals in the absence of sex information. Standard curves using only male DNA provided better discrimination than using mixed-sex DNA during qPCR. NOR copy number is an effective sex diagnostic for *A. spinifera* turtles. Our sexing approach using qPCR of 18S genes should prove useful for other taxa that also possess dimorphic NORs, as is known in some vertebrates and insects. While the 18S copy numbers in our dataset exhibited a non-overlapping binomial distribution, this may not always be the case in future studies of *A. spinifera* or for other taxa. Importantly however, our sex-typing approach using mixture models provides an attractive alternative under overlapping distributions of these and of other continuous data such as hormone levels, gene expression levels, shape or behavior. We present an example using overlapping distributions of hormone levels in *Chelydra serpentina* turtles, to demonstrate the broader utility of mixture models for sex-typing, and obtain a high correct classification of 90%.

3.2 Introduction

Accurate and early identification of the sex of animals is imperative in fields spanning from medicine to evolutionary and conservation biology. For instance, sex assessment is required prior to embryo implantation in humans and domestic animals in order to diagnose diseases or to appraise embryo quality during in-vitro fertilization (Hamilton *et al.* 2012). Likewise, fisheries and other animal industries benefit from early sex identification to select the most desirable gender for commercial purposes (Singh 2013). In conservation

biology, sex information is crucial to implement sex-specific management strategies or sex ratio monitoring of endangered species (Mrosovsky 1982, Korstian *et al.* 2013). Finally, studying the ecology and evolution of sex allocation and sex-specific traits also requires reliable sex identification (Ellegren and Sheldon 1997, Griffiths 2000).

Obtaining information on individual sex is simple for species or life stages that exhibit obviously dimorphic phenotypes. However, difficulties emerge for organisms or life stages where no diagnosable external dimorphism exists that is detectable by visual inspection. Several techniques have been devised to sex individuals in such cases and applied to diverse taxonomic groups. However some direct techniques are destructive, such as the observation of gonadal morphology or histology in dead animals (Yntema and Mrosovsky 1980), while others are invasive, such as laparoscopic inspection of gonadal morphology in live animals (Wood *et al.* 1983, Rostal *et al.* 1994). Less intrusive sex diagnosis can be accomplished by detecting the presence/absence of a sex-linked trait using molecular approaches, such as the cytogenetic detection of sex chromosomes (Ezaz *et al.* 2005, Badenhorst *et al.* 2013), PCR amplification of a sex-specific marker (Griffiths 2000, Morinha *et al.* 2012, Korstian *et al.* 2013), or quantitative PCR (qPCR) of genes linked to the sex chromosomes that are present in two copies in one sex and one copy in the other (Phillips and Edmands 2012, Alasaad *et al.* 2013, Ballester *et al.* 2013). The molecular techniques mentioned above represent examples of discrete traits. Alternatively, sex assessment may rely on the indirect measurement of some continuous feature that is sexually dimorphic such as hormone levels (Owens *et al.* 1978, Akyuz *et al.* 2010), gene expression (Hamilton *et al.* 2012), or multivariate data such as shape (Valenzuela *et al.* 2004, Ceballos and Valenzuela 2011, Ceballos *et al.* 2014).

Turtles are a lineage exemplifying the need and difficulty of sex diagnosis. While many turtle species display sexually dimorphic characters as adults such as size or shape differences (Ceballos *et al.* 2012), hatchlings and juveniles usually lack early sexual dimorphism that is visually diagnosable. Yet, sex information of embryonic or young turtles is crucial to monitor sex ratios and to study sex-specific traits that may influence fitness [e.g. (Janzen 1993, Ceballos *et al.* 2014)]. Consequently, multiple sexing techniques have been developed for turtles, including gonadal inspection or histology (Yntema and Mrosovsky 1980), laparoscopy (Wood *et al.* 1983, Rostal *et al.* 1994), radioimmunoassay of circulating hormones in blood (Owens *et al.* 1978, Lance *et al.* 1992, Rostal *et al.* 1994, Valenzuela 2001), or chorioallantoic/amniotic fluid of the egg (Gross *et al.* 1995). The least invasive sexing method for juveniles utilizes geometric morphometric quantification of subtle dimorphism in the turtle carapace of several species (Valenzuela *et al.* 2004) or in the anal region of the plastron in others (Ceballos *et al.* 2014). However, because geometric morphometrics quantifies shape by the relative position of carapace scutes which serve as homologous landmarks, it cannot be applied to softshell turtles since their shells lack carapace scutes altogether, and their sexual size dimorphism is not evident prior to sexual maturity at 8-10 years of age (Ernst and Lovich 2009). Moreover, tests of circulating hormones are expensive and cumbersome.

Apalone spinifera softshell turtles exhibit a ZZ/ZW sex chromosome mechanism of genotypic sex determination (Badenhorst *et al.* 2013). Unfortunately, molecular cytogenetic techniques are costly and highly specialized, such that ZZ/ZW detection for sex-typing large numbers of individuals in population-level studies is precluded. Importantly, fluorescent in situ hybridization (FISH) of an 18S rRNA gene probe, revealed that the nucleolar-organizing

region (NOR) in *A. spinifera* is located on the sex chromosomes and exhibits a much greater copy number on the W than on the Z (Figure 3.1), making it a promising dimorphic marker for sex identification (Badenhorst *et al.* 2013). The NOR contains genes for the three major ribosomal RNA subunits (18S, 5.8S, 28S) repeated in tandem to permit sufficient transcription to supply cellular demands for ribosomes (Shaw and McKeown 2011). When NORs are located in the non-recombining region of sex chromosomes, the number of repeats may become sexually dimorphic, as in *A. spinifera* turtles (Badenhorst *et al.* 2013).

When using continuous traits for sex-typing the analytical methods to assign individuals as male or female fall into two main categories. The first category uses a set of individuals of known sex to train an algorithm that is then used to assign the sex of unknown samples as male or female (Valenzuela *et al.* 2004, Ceballos and Valenzuela 2011, Ceballos *et al.* 2014). The second category relies on the bimodality of the continuous variable in the absence of any a priori sex information from any individual, and then assignment of an individual as male or female based on how close its value is to one or the other group mean. This latter assignment however, is usually done in an *ad hoc* fashion rather than using standardized statistical procedures, especially for individuals with intermediate values that approach the area of overlap in the bimodal distribution [e.g. (Valenzuela 2001, Weissmann *et al.* 2013)]. Thus, while a variety of molecular sexing techniques have been widely used to assign individuals to sexes, a general approach for the use of any continuous dimorphic molecular data as a sex diagnostic tool is not commonly applied, particularly when the cutoff between males and females in the binomial distribution is not as evident. Mixture models provide such a framework (Fraley and Raftery 2002).

Here we use the novel 18S genomic region for sexing *A. spinifera* turtles. The 18S copy number variation among individuals represents a continuous variable that can be quantified via qPCR and analyzed using mixture models and univariate discrimination (Fraley and Raftery 2002) for sex-typing. Our approach offers an attractive alternative for the fast, accurate and reliable sex diagnosis in softshell turtles. Our molecular method is applicable to broader taxa that possess sexually dimorphic NORs (Goodpasture and Bloom 1975, Hsu *et al.* 1975, Schmid *et al.* 1983, Bickham and Rogers 1985, Schmid *et al.* 1993, Born and Bertollo 2000, Kawai *et al.* 2007, Abramyan *et al.* 2009b, Monti *et al.* 2011, Takehana *et al.* 2012, Badenhorst *et al.* 2013), and our analytical approach is appropriate for any other bimodal continuous variables and multivariate traits with overlapping distributions. We provide such an example using hormonal data from snapping turtles, *Chelydra serpentina*.

3.3 Materials and methods

Sample collection

Apalone spinifera eggs were incubated at 26°C, 28°C, or 31°C as previously described (Valenzuela 2010). Hatchlings were housed in a temperature-controlled facility and were given access to UV light, burrowing substrate, water, and a dry basking area to ensure healthy growth. At approximately three months of age gonadal differentiation was advanced to the point that the sex of 89 hatchlings could accurately be determined by visual gonadal inspection. At this age ovaries displayed clear ovarian ducts and prominent follicles, while testes exhibit substantial seminiferous tubule development and are smaller than the ovaries.

DNA extraction and quality control

DNA was extracted from muscle tissue using Gentra Puregene DNA extraction kit (Gentra) following the manufacturer instructions and was quantified and quality checked using a Nanodrop ND-1000 Spectrophotometer and gel electrophoresis (0.8% agarose). Then, a subset of 40 male and 40 female hatchlings with high molecular weight DNA was selected for further analysis. DNA was diluted to 1.25 ng/ul for use in the quantitative PCR (qPCR) assay. This DNA concentration produced qPCR amplification profiles with similar fluorescence levels for both the 18S and GAPDH genes during a pilot test.

Quantification of 18S rRNA repeat copy number

Copy number of the 18S rRNA repeats was quantified in each individual using qPCR and normalized against GAPDH, a single copy gene used as endogenous control. Using data from an *A. spinifera* transcriptome (Radhakrishnan *et al.* 2017), qPCR primers were designed to amplify a 144bp fragment of 18S rRNA (forward 5'-GAGTATGGTTGCAAAGCTGAAA-3'; reverse 5'-CGAGAAAGAGCTATCAATCTGT-3') and a 129bp fragment of GAPDH (forward 5'-GGAGTGAGTATGACTCTTCCT'-3'; reverse 5'-CAGCATCTCCCCACTTGA-3'). Standard curves were generated by pooling equimolar amounts of five high-quality genomic DNA (gDNA) samples. Pooled DNA was diluted to 100 ng/ul, and then serially diluted from 1:10 to 1:640 for final concentrations of 10, 5, 2.5, 1.25, 0.625, 0.3125, and 0.1563 ng/ul. Two different standard curves were tested in this study: (1) a mixed-sex standard curve containing DNA from three male and two female samples chosen at random (to simulate conditions where individual sex is unknown), and (2) a male-only standard curve made by pooling the DNA of five known males (to test if

a standard curve made with DNA from the sex that has smaller 18S blocks provides better discrimination of 18S copy number between males and females). qPCR was performed using Brilliant II Sybr Green qPCR Master Mix (Agilent) in an Mx3000P thermocycler (Agilent), with ROX as the reference dye for background correction. qPCR was performed in 25 ul reactions containing 2ul of sample DNA (2.5 ng) or standard DNA, and a final primer concentration of 400nM. qPCR cycling conditions were as follows: 1 cycle at 95°C for 10 min; 45 cycles of 95°C for 30 sec, 58°C for 1 min, 72°C for 1 min; and a dissociation-curve cycle of 95°C for 1 min, 55°C for 30 sec, taking readings at 0.5°C increments until reaching 95°C for 1 min, to test for unspecific amplification. Samples and standards were run in duplicate in each qPCR plate. Threshold fluorescent values for each qPCR plate were first automatically assigned by the MxPro software, and an overall average threshold value was manually chosen which was appropriate for all genes and plates. Any samples whose replicates exhibited non-specific amplification or a C_T deviation greater than 0.5 between duplicates were excluded from further analysis. Negative, no-template controls were also run in duplicate to test for primer dimers or contamination. The efficiency of each qPCR reaction was calculated from the standards as:

$$Eff = 10^{-(1/\text{slope})}$$

Copy number of the 18S gene was normalized against GAPDH using the comparative C_T method of normalization (Livak and Schmittgen 2001):

$$Ratio\ (\frac{18S}{GAPDH}) = 2^{-\Delta C_T} = 2^{(C_T\ GAPDH - C_T\ 18S)}$$

Other normalization methods are compared in Appendix B.1.

General analytical method for sex identification

The goal of any sexing technique is to assign individuals to groups (males and females). Using a single continuous trait, the first step in this process is to visualize a histogram of the data which should be bimodal with respect to sex (Appendix B.2). A test is then carried out to validate the sexual dimorphism of the trait in question and its efficacy for accurate sex-typing of individuals as described below. Here we use mixture models which consider the data as containing combinations of two or more distributions, with each mixture component corresponding to a group whose parameters can then be estimated (Baudry *et al.* 2010). The most common component is typically a combination of multiple normal distributions. Analytically, parameter estimates of mixture models may be calculated using an expectation maximization (EM) procedure in a likelihood framework [see (Fraley and Raftery 2002)]. To implement the procedure described above two conceptual approaches are possible, which depend on the data available (Appendix B.2). R-code and data for an implementation example are found in Appendix B.3.

Procedure 1 – Discriminant Analysis

If the sex of a subsample of individuals is known (determined by other techniques such as gonadal inspection), this subsample is first used as a training set to find the parameters for each group's distribution (means and standard deviations for males and females). The conditional probabilities of each sample belonging to each of the groups given the parameters of the data (z) is calculated and individuals are assigned to the group that minimizes the uncertainty ($1 - z$). When applied to the training set, the training classification rates measure the fit of the model to the data. Second, conditional probabilities are calculated for each unknown sample in the test dataset and individuals are assigned to groups in the

same fashion using the parameters calculated from the training set. An additional test can be carried out by dividing the subsample of individuals of known sex into two groups, one to be used as a smaller training set and the other to be used as a test set by ignoring the known sex information. In this case the parameters of the male and female groups are calculated as described above using the smaller training set, and then used to classify the test set individuals as male or female. Thus, the classification for the test set serves as cross-validation for the sex-typing approach (since the true sex of individuals in the test set is actually known). The classification error rate for the test-set provides the level of confidence that can be expected for the sex-typing of unknowns using this approach (Fraley and Raftery 2002).

Procedure 2 – Clustering Analysis

If the sex of all individuals is unknown, mixture models are first used to find the distributions (groups) that best fit the data, and to estimate the means and standard deviations of each group thus identified. The conditional probabilities of each sample belonging to each of the groups given the parameters of the data (z) is calculated and individuals are assigned to the group that minimizes the uncertainty ($1 - z$), in the same manner as for Procedure 1. The uncertainty provides a measure of the quality of the classification by subtracting from 1 the probability of the most likely group for each individual (Fraley *et al.* 2012).

Potential Data Complications

The method described above is straightforward when the variation for both sexes is similar (standard deviations are comparable), and when there are no outlier values. To ensure this, some additional steps should be followed. First, deviant values are identified using the EM algorithm in the mixture model, where outliers are classified into their own cluster [see

(Fraley and Raftery 2002)]. This classification is inspected visually to determine the sex of the group(s) that corresponds to the outlier values. For instance, if the biological expectation is that males have low values while females have high values, samples with deviant high numbers will denote females with extreme values at the upper tail of their distribution, while deviant low values would correspond to males at the low tail of their distribution. After classification, assignments could be visually inspected with respect to the distribution. Second, if the variation is not uniform between the sexes the mixture model procedure will favor a model with unequal variance, as this provides the best fit to the data. That model is then implemented for parameter estimation and classification.

Another complication emerges when the distributions of male and female values overlap. In the case that sex information is available for a subset of individuals, an estimate of the overlap and the classification error that it generates can be obtained using Procedure 1. Additionally, for this case and for the case when sex information is unavailable for a subsample of individuals, the uncertainty levels calculated as described for both procedures can be used to remove from the dataset individuals that cannot be classified with an acceptable confidence level as defined by the researcher (e.g. >80%, >90%, >95%) (Appendix B.5).

Here we tested both analytical methods (Procedures 1 and 2) in *A. spinifera*. First, because sex information was available for all our samples, we tested the classification rates employing Procedure 1 (discriminant analysis) when two thirds of the samples (46 hatchlings) were used as a training set to generate the discriminant model and the other third (22 hatchlings) as a testing set for cross-validation. Second, we tested the classification rates employing Procedure 2 (clustering analysis) by treating all the samples as if they were

unknowns (ignoring the gonadal sex information available), and allowing the mixture model to identify the groups in the absence of any prior sex information. We then examined the concordance of the estimated and true sex information to assess the performance of Procedure 2. Statistical analyses were carried out using the mclust v4.2 package (Fraley *et al.* 2012) in R using the MclustDA function for Procedure 1 and the Mclust function for Procedure 2.

3.4 Results

Gonadal inspection, qPCR quality control, and 18S normalization

The sex ratio of the 89 hatchlings was 46 males: 43 females (Table 3.1), which did not differ from 1:1 and was not influenced by temperature (Chi-Square, p=0.75), as expected for a species with genotypic sex determination (Bull and Vogt 1979).

Of these 89 individuals, 40 males and 40 females with high-quality DNA were used for qPCR. Four male and eight female samples were removed from the analysis as their C_T values differ by >0.5 cycles between replicates. The final dataset contained 68 individuals. Dissociation curves after qPCR detected no secondary products or primer dimers, and negative controls were clean. Plate efficiencies and quality of the standard curve as determined from the coefficient of determination (R^2) are summarized in Table 3.2. The qPCR efficiencies (*Eff*) per plate ranged between 1.973 -2.034 (97.3%-103.4%), and the R^2 values are all > 0.99. Thus, amplification reactions for GAPDH and 18S were highly efficient, comparable, and appropriate to predict the sample C_T values by linear regression. Alternative methods of normalization were also tested and our results were robust across all methods (Appendix B.1).

To assess the similarity of the qPCR reaction efficiencies between the 18S and GAPDH amplification, we run a regression analysis of the ΔC_T values ($C_{T,GADPH} - C_{T,18S}$) of the standard curve samples against the Log_2(DNA amount) for each standard curve dilution, following Livak and Schmitgen (2001). The slope of the regression of ΔC_T versus Log_2(DNA template amount) was less than 0.1 for both standard curve types (Figure 3.2) indicating that the qPCR efficiencies were similar for the amplification of 18S and GAPDH, and thus, that the use of the comparative C_T method of normalization was appropriate for our data (Livak and Schmittgen 2001). This is important because the C_T method of normalization is only applicable if the qPCR reaction efficiencies for the gene of interest and endogenous control are both around 100% (Eff~2) and comparable between genes.

Values of the ratio of 18S rRNA to GAPDH copy number exhibited a bimodal distribution with no overlapping values between males and females (Figure 3.3). The highest male 18S/GAPDH ratio was 298, while the lowest female 18S/GAPDH ratio was 429 (Table 3.3, Figure 3.3). On average females had approximately four times as many copies of the 18S rRNA gene than males (Table 3.3). This result is concordant with the cytogenetic evidence which shows that the W chromosome in female *A. spinifera* contains an extended NOR region which houses many more copies of 18S rRNA than males (Badenhorst *et al.* 2013)(Figure 3.1). Additionally, the variance in 18S copy number among females (Coefficient of Variation = 42.3-45.5%) was greater than the variance among males (Coefficient of Variation = 26.3-29%) (Figure 3.4, Table 3.3). These differences were caused mainly by a subset of females that had relatively higher 18S/GAPDH values than the rest, and which are identified as an outlier group using mixture models (Figure 3.3G). When those outlier females are removed the coefficient of variation is similar for males and females. Our

results were robust to using two alternative methods for normalization of 18S copy number, using the same samples run with the male-only and mixed-sex standard curves (Appendix B.1). An ANOVA did not detect any effect of the standard curve type (mixed-sex versus only-male standards) on the mean 18S/GAPDH ratio within-sex (Figure 3.4, Table 3.3).

Individual sex assessment using mixture models of clustering

Results from the analytical mixture models using the discriminant analysis (Procedure 1), and treating 46 individuals as a training set and 22 individuals as a test set, resulted in a classification error rate of 0% for the training set (Figure 3.3B,C) and for the test-set during cross-validation (Figure 3.3D,E). Using the mixture models and treating all individuals as unknowns identified three clusters corresponding to the male, female, and female outliers groups from the bimodal distribution (Figure 3.3G). As expected, uncertainty values increased in the areas between two groups (Figure 3.3H). However, the classification rate treating all individuals as unknown was 100% accurate. Namely, all individuals in the lowest group (group 1) were males, and all individuals in groups 2 and 3 were females.

Sex assessment using mixture models of clustering when distributions overlap

To test our approach for cases where the distribution of the values for males and females overlap, we carried out an additional analysis (Appendix B.5) using a testosterone radioimmunoassay dataset of snapping turtles for which gonadal sex information was available from laparoscopic examination (Ceballos & Valenzuela 2011). Procedure 1 (Discriminant Analysis), dividing the data set of 136 individuals into a training set of 46 turtles and a test set of 90 turtles, resulted in a classification error of 11% for the training set and of 13% for the test set (Appendix B.5). Procedure 2 (Clustering Analysis), treating individuals as if their sex information was unknown, resulted in the misclassification of 21

out of the 136 individuals (classification error = 15%). Removing individuals from the dataset whose classification uncertainty exceeded 0.05 (for whom the sex-typing was less than 95% certain according to the mixture model), improved the classification rate to 90% (error rate = 10%).

3.5 Discussion

qPCR quantification

If genomic DNA is to be used to create the qPCR standard curve, our results indicate that the comparative C_T method ($2^{-\Delta Ct}$) is the simplest method to apply and perhaps preferable to alternative methods of normalization (see Appendix B.1 for a comparison of the merits and results of alternative normalization methods) for *A. spinifera*, because once the qPCR reaction is optimized, it requires no pre-knowledge of the sex of any individual. However, using samples of known sex would still be beneficial for validation. Additionally, using standard curves is important to evaluate if qPCR conditions are similar and optimal for the gene of interest and the endogenous control gene. Our approach can distinguish between male and female *A. spinifera* with as little as 5ng of high quality genomic DNA (for duplicate reactions of 2.5ng each) which could be extracted from a blood draw or a small tissue clip, and would permit the sexing of embryos, hatchlings, or juveniles in a variety of studies. For instance, sexing *A. spinifera* embryos would enable tests of the effect of temperature, sex, and their interaction in developmental studies of gene expression, which were precluded in previous studies of sex determination in this species (Valenzuela *et al.* 2006, Valenzuela and Shikano 2007, Valenzuela 2008b, Valenzuela 2008a, Valenzuela *et al.* 2013). Though not tested directly, template quality (DNA degradation) should have a minor effect given that the

amplicon from the qPCR is a small product of only ~150bp for both genes and should amplify even if the DNA is degraded. However, a quality check should be carried out after DNA extraction to test the integrity of the DNA for qPCR. The presence of PCR inhibitors would affect both the 18S and housekeeping genes, but it would be expected that their ratio (and thus our method) should remain unaffected. The 18S primers used in this study were designed in a highly conserved region such that they should work across a wide gamut of animals from insects to vertebrates. However, GAPDH DNA sequences are more variable among taxa such that species-specific primers need to be designed for other studies.

Analytical method for sex identification

Our test using *A. spinifera* softshell turtles demonstrate the utility of our analytical approach to sex-type individuals under two possible scenarios: (1) when sex information is available for a subset of individuals, and (2) when all individuals are of unknown sex. Results indicated that when applied to the sexually dimorphic NOR region of the *A. spinifera* genome, the use of mixture models and univariate discrimination exhibited high classification rates (100%), low error rates during cross-validation (0%), and high discrimination power even when individuals were treated as unknowns (100%). While this is not surprising since our data set contained values with a non-overlapping binomial distribution, our findings corroborate that the distributions estimated by the mixture models did not create an artificial overlap of values between males and females where none existed.

Although when testing the *C. serpentina* dataset whose male and female hormonal values overlap (Appendix B.5) the error rate was higher than our results for *A. spinifera*, our approach compares well with those of previous studies in other turtles using continous traits with overlapping distributions, such as to those in *Podocnemis expansa* using geometric

morphometrics [e.g. 75-90% correct crossvalidation (Valenzuela *et al.* 2004, Ceballos *et al.* 2014). These findings are important because there is no warranty that further sampling or data generated by other researchers from softshell turtles or from other species with sexually dimorphic NORs, will not contain overlapping values of 18S copy number between males and females. Thus, it is important to have in place an analytical method that is flexible in its application for all possible potential circumstances. Additionally, the level of overlap of the male and female distributions is also affected by the normalization method and composition of the standard curve (whether containing DNA from both sexes or only the sex with the lower values of the continous trait) (Appendix B.1). Our results in *C. serpentina* demonstrate that our analytical method is efficient for sex-typing when distributions overlap.

3.6 Conclusions

Since the ZZ/ZW sex chromosome system present in *A. spinifera* has remained virtually unchanged since the split of *Apalone* and *Pelodiscus* ~95million years ago (mya; (Kawai *et al.* 2007, Badenhorst *et al.* 2013), our sexing technique should be widely applicable to other *Apalone* and *Pelodiscus* species. Furthermore, our approach should also apply to a wide variety of species that exhibit sexually dimorphic NORs. Among those there are species where the NORs also differ in size between the two sex chromosomes [*Hoplias malabaricus* fish (Born and Bertollo 2000); medaka fish *Oryzias hubbsi* and *O. javanicus* (Takehana *et al.* 2012); and *Bufo marinus* toads (Abramyan *et al.* 2009a)]. In some other taxa the NOR is present in the X chromosome and absent in the Y [*Staurotypus salvini* turtles (Bickham and Rogers 1985), *Gastrotheca riobambae* frogs (Schmid *et al.* 1983), long-nosed potoroo *Potorous tridactylia* and *Carollia perspicillata* bats (Goodpasture and Bloom 1975,

Hsu *et al.* 1975)]; or present in both X in diploid females and in the single X of haploid males [potato aphids *Macrosiphum euphorbiae* (Monti *et al.* 2011)]. In yet others, the NOR is present in the Z but not in the W [*Buergeria buergeri* frogs (Schmid *et al.* 1993)]. Additionally, the use of digital PCR (dPCR) (Vogelstein and Kinzler 1999) is likely to make our approach even more powerful.

Notably, the use of mixture models is an alternative to identify individual sex based on any continuous variable such as circulating hormone levels which have been used to identify sex in chickens (Weissmann *et al.* 2013) and reptiles with temperature-dependent sex determination such as *Chelonia mydas* sea turtles (Owens *et al.* 1978), *Gopherus agassizii* desert tortoises (Rostal *et al.* 1994), *C. serpentina* (Ceballos and Valenzuela 2011), and Amazonian giant river turtles *P. expansa* (Lance *et al.* 1992, Valenzuela 2001). Our additional example analysis on *C. serpentina* demonstrates its utility for sex-typing using hormonal data and under overlapping distributions (Appendix B.5). Similarly, gene expression levels are also a continuous trait amenable to analysis by this approach and have been used to sex-type bovine blastocysts (Hamilton *et al.* 2012). Even behavior, such as the *fee* glissando components of songs in male black-capped chickadee *Poecile atricapillus* provide a continuous trait for sex-typing (Hahn *et al.* 2013) that could be analyzed by mixture models. Importantly, mixture models are not restricted to univariate discrimination but can be applied equally to multivariate data such as shape which can be quantified by geometric morphometrics as done to sex-type the giant Amazonian river turtle *P. expansa* and painted turtle *Chrysemys picta* hatchlings (Valenzuela *et al.* 2004, Ceballos and Valenzuela 2011, Ceballos *et al.* 2014), or by linear measurements as in *Lepidochelys olivacea* sea turtles (Michel-Morfin *et al.* 2001). Thus, both the genomic region and the

analytical approach proposed here should be broadly applicable for sex-typing beyond softshell turtles.

3.7 Acknowledgements

All procedures were approved by the IACUC of Iowa State University. We thank D.C. Adams, and the members of the D.C. Adams and J. Serb labs at Iowa State for their comments. This work was partially funded by grant MCB 1244355 to NV.

3.8 Tables and Figures

Table 3.1: Sex ratios of *Apalone spinifera* hatchlings determined by visual sexing of gonads. p-values represent results of Chi-Square analyses that test if the sex ratio is skewed from 1:1. 28°C/31°C corresponds to hatchlings from 28°C or 31°C whose incubation information was lost.

Egg Incubation Temperature	26°C	28°C	31°C	Unknown	Overall
Number of Males	8	13	20	5	46
Number of Females	7	13	17	6	43
Chi-Square p-value	.7963	1.0	.6219	.7630	.7505

Table 3.2: Efficiency of the qPCR and quality of the standard curve for all plates run in this study. Two 96-well plates were needed for each gene given our sample size.

Plate	Standard Curve	qPCR Efficiency	Standard Curve R^2
18S 1	Male-Only	1.9966	.996
18S 2	Male-Only	2.0058	.993
GAPDH 1	Male-Only	1.9864	.997
GAPDH 2	Male-Only	1.9731	.996
18S 1	Mixed Sex	2.0336	.994
18S 2	Mixed Sex	2.0080	.996
GAPDH 1	Mixed Sex	1.9975	.993
GAPDH 2	Mixed Sex	2.0126	.994

Table 3.3: Normalized 18S copy number (ratio of 18S rRNA to GAPDH) in the *Apalone spinifera* genome as measured by qPCR. All samples were run with both a male-only and mixed-sex standard curve, and 18S values normalized with three alternative methods (equations 2-4) as detailed in the text. C.V. = Coefficient of Variation. Avg = average. MaxFem= maximum female value. MinMale = minimum male value. Significant P values are denoted in bold. F = F statistic; df = degrees of freedom, P = P value.

	Relative Standard Curve Method		**Pfaffl Calibrator Method**		**Comparative Ct Method (2ΔCt)**	
Normalized 18S	Male-Only Standards	Mixed Sex Standards	Male-Only Standards	Mixed Sex Standards	Male-Only Standards	Mixed Sex Standards
Male Minimum	0.420	0.295	0.306	0.295	66.028	85.036
Male Average	0.955	0.667	0.832	0.670	177.698	192.544
Male Maximum	1.423	1.039	1.243	1.039	266.871	298.172
Female Minimum	2.320	1.361	2.247	1.236	429.049	435.039
Female Average	4.095	2.431	3.973	2.193	760.853	740.707
Female Maximum	9.494	5.481	9.184	4.917	1764.447	1640.591
Avg. Female:Male	4.288	3.645	4.775	3.273	4.282	3.847
C.V. Male	26.364	28.95	26.963	28.921	26.974	28.751
C.V. Female	42.355	45.482	42.372	45.365	42.563	44.419
Gap between sexes (MinFem-MaxMale)	0.897	0.322	1.004	.197	162.178	136.867
Total Range (MaxFem-MinMale)	9.075	5.186	8.878	4.622	1698.419	1555.555
ANOVA test of Effect of Standards on Male Average	F= 29.57 df=71 **P <0.05**		F= 10.69 df=71 **P <0.05**		F= 1.48 df=71 P >0.05	
ANOVA test of Effect of Standards on Female Average	F= 20.94 df=63 **P <0.05**		F= 26.52 df=63 **P <0.05**		F= 0.0609 df=63 P >0.05	

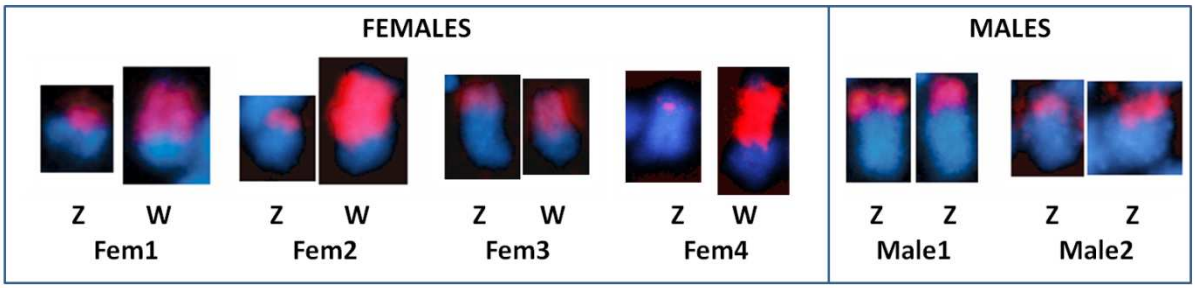

Figure 3.1: ZZ/ZW sex chromosomes of *Apalone spinifera* (modified from Badenhorst et al. 2013). Red color corresponds to the fluorescent in situ hybridization of an 18S rRNA gene probe revealing a larger block of 18S repeats in the W than in the Z chromosomes.

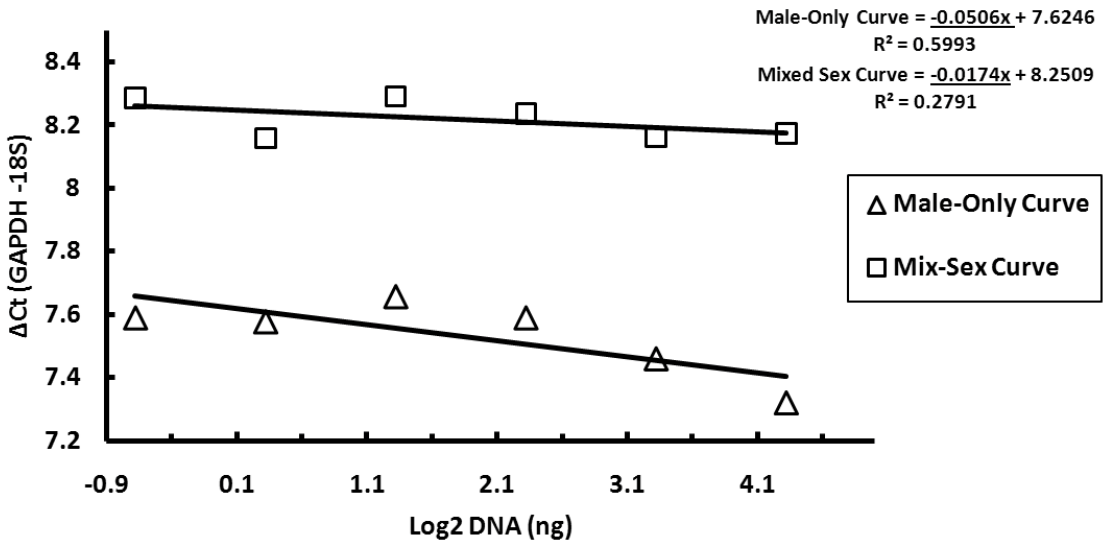

Figure 3.2: Assessment of the qPCR efficiencies for the gene of interest (GOI) and endogenous control (EC). The regression of ΔC_T against template amount (Log_2 DNA) revealed a slope close to zero (less than 0.1), indicating that the qPCR reaction efficiencies for the GOI and EC are similar enough to use the comparative C_T method (Livak and Schmittgen 2001). Both standard curve types employed in this study meet this requirement. Slope values are underlined.

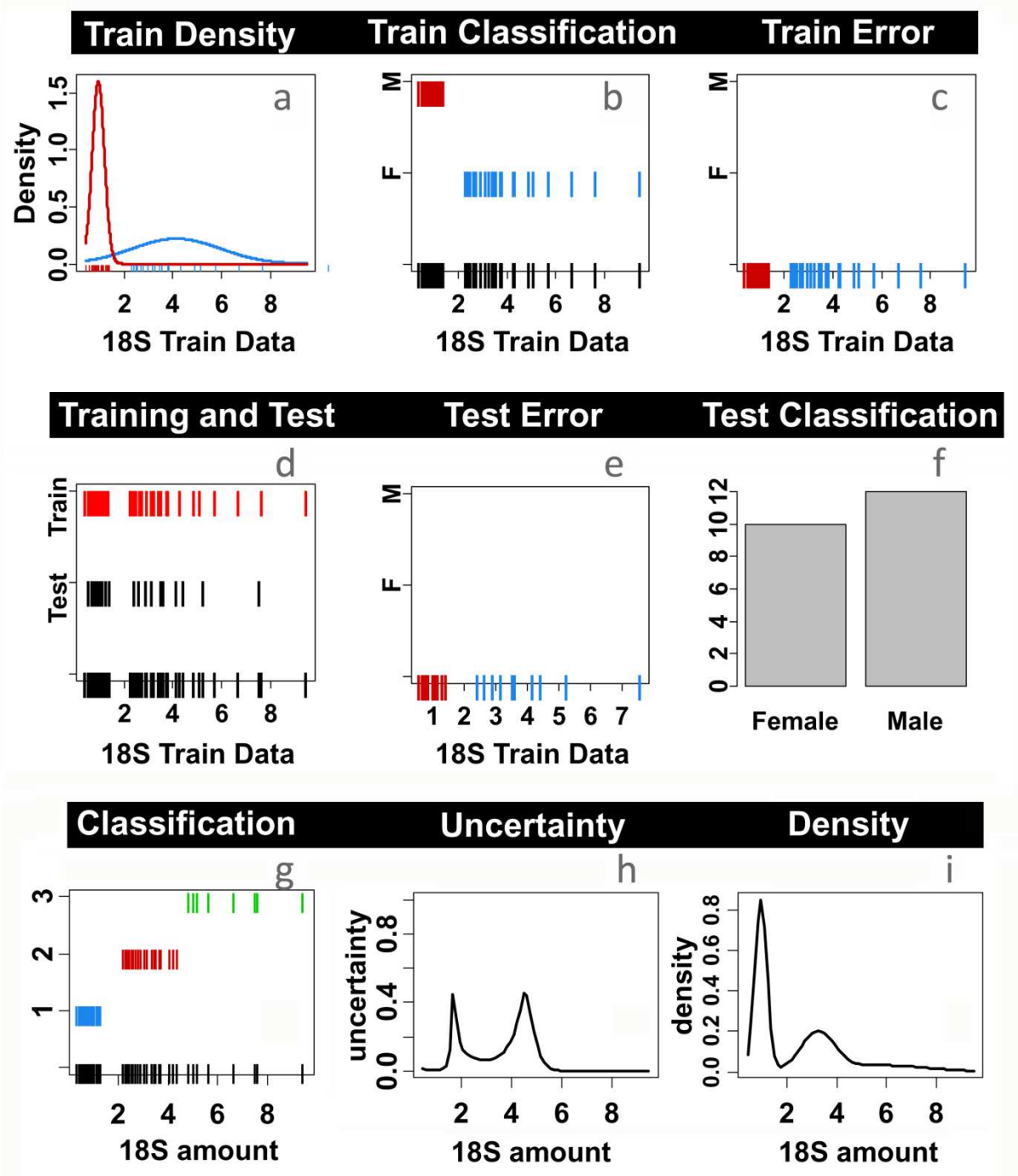

Figure 3.3. Results of the use of mixture models for discriminant analysis (a-f panels) for sex-typing of *Apalone spinifera* turtles when sex information is available [including distribution density, classification and error rates for the training and test datasets], and for clustering (g-i panels) in the absence of a priori sex information. In panel g, note the identification of three groups: typical males (blue), typical females (red), and "outlier" high-value females (green).

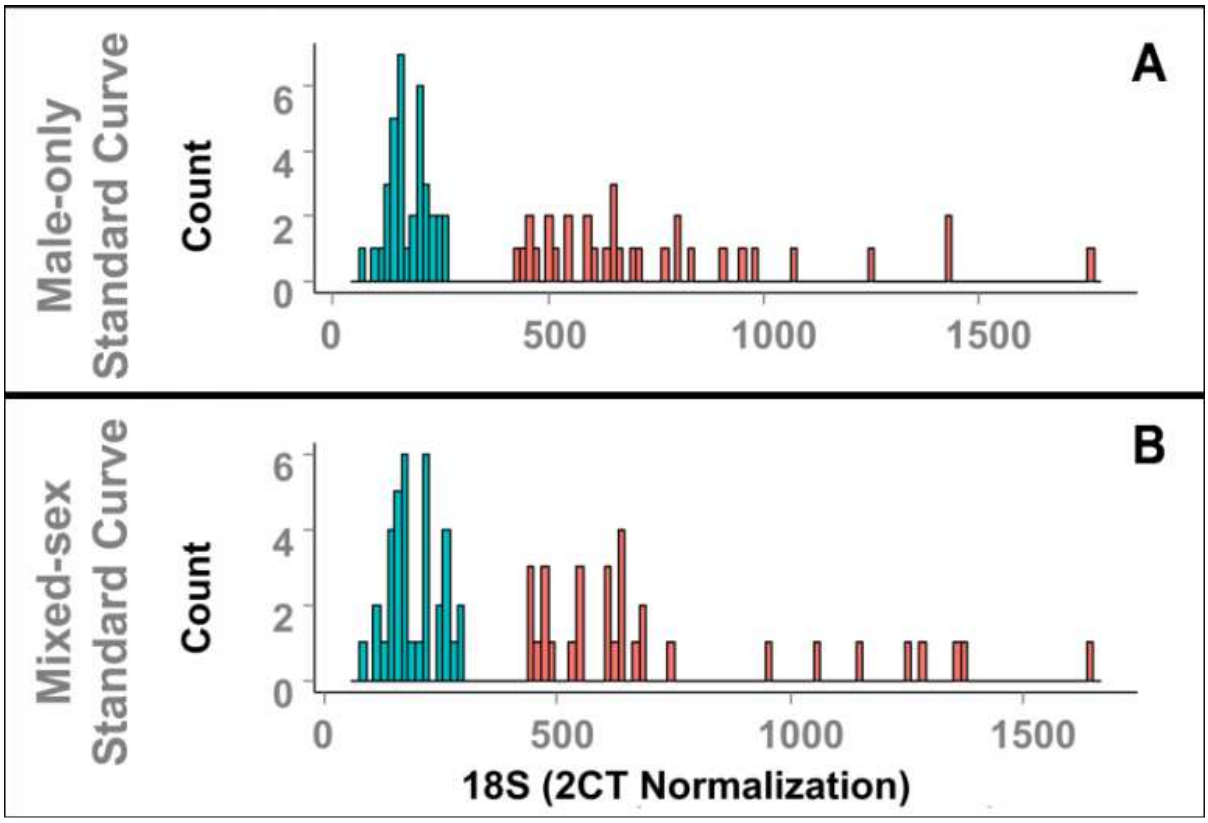

Figure 3.4: Histograms of the distribution of 18S normalized to GAPDH in *Apalone spinifera* using the comparative $2C_T$ normalization method, and two types of standards: male-only DNA (A) or mixed-sex DNA (B).

CHAPTER 4: DEVELOPMENT OF SEXING PRIMERS IN *GLYPTEMYS INSCULPTA* AND *APALONE SPINIFERA* TURTLES UNCOVERS AN XX/XY SEX-DETERMINING SYSTEM IN THE CRITICALLY-ENDANGERED BOG TURTLE *GLYPTEMYS MUHLENBERGII*

Robert Literman[1], Srihari Radhakrishnan[1], Jeff Tamplin[2], Russell Burke[3], Cassie Dresser[4], Nicole Valenzuela[1]

Modified from a paper published in *Conservation Genetics Resources* (Literman et al. 2017)

Author Contributions: RL conceived study, developed bioinformatics pipeline, ran PCR reactions, analyzed data, and wrote manuscript. SR assisted with bioinformatics. JT, RB, CB assisted with sample procurement. NV assisted in manuscript editing.

Author Affiliations:

1: Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, USA

2: Department of Biology, University of Northern Iowa, Cedar Falls, USA

3: Department of Biology, Hofstra University, Hempstead, USA

4: Department of Ecology and Evolutionary Biology, University of Tennessee Knoxville, Knoxville, USA

4.1 Abstract

In species or developmental stages where the sex of an individual cannot be reliably identified through external morphology, molecular markers can provide a critical tool to study sex-specific traits that are elusive otherwise. Here we generated two sets of sex-diagnostic PCR primers for each of two focal turtle species with contrasting genotypic sex

determination (GSD) systems: the wood turtle, *Glyptemys insculpta* (XX/XY), and the spiny softshell turtle, *Apalone spinifera* (ZZ/ZW). These markers identified males and females with 100% accuracy as validated with numerous individuals of known sex. Notably, one of the markers developed for *G. insculpta* permitted the successful diagnosis of individual sex in the critically-endangered bog turtle, *Glyptemys muhlenbergii,* also with 100% accuracy. This cross-species application provided the first evidence that *G. muhlenbergii* shares an XX/XY sex-determining mechanism with *G. insculpta*, a finding with important evolutionary and conservation implications. Similarly, the markers from *A. spinifera* were successful in identifying the sex of two individuals (one male and one female) of the Chinese softshell turtle, *Pelodiscus sinensis* (ZZ/ZW). These cross-species observations highlight the potential applicability of these types of markers on closely related taxa that share a sex-determining mechanism, which should be tested in a case-by-case basis.

4.2 Introduction

The sex of an individual has important effects on a large number of traits such as growth, migration pattern, and mortality. These effects are well documented in turtles (Rhen and Lang 1995, Aresco 2005, Chaloupka and Limpus 2005, Steen *et al.* 2006), the most endangered vertebrate group (Hoffmann *et al.* 2010). While conservation strategies are improved when individual sex can be quickly and reliably determined non-lethally (Faust and Thompson 2000, Korstian *et al.* 2013), sex diagnosis is hindered in animal species or life stages lacking readily discernable sexual dimorphisms. While many turtles are visibly sexually dimorphic as adults (Ceballos and Valenzuela 2011), few are sufficiently dimorphic early in life. Molecular assays offer a non-lethal and accurate alternative to invasive or

destructive sexing methods. The simplest molecular sexing technique is by the presence/absence of a sex-linked molecular marker, detectable using PCR amplification (Bredbacka *et al.* 1995, Clinton *et al.* 2001, Gamble and Zarkower 2014). Here we report on the development of two sex-diagnostic PCR primer sets for each of two focal turtle species (*Glyptemys insculpta* and *Apalone spinifera*) with contrasting genotypic sex determination (GSD) whose heterogamety is well characterized cytogenetically. We then test these markers in closely related species to each of the focal taxa.

This work has important conservation implications. The wood turtle *G. insculpta* (GIN hereafter) is Endangered (IUCN 2016) due to slow sexual maturity, nest predation and anthropogenic effects (Levell 2000, Saumure *et al.* 2007, Ernst and Lovich 2009). Its congener the bog turtle, *Glyptemys muhlenbergii* (GMU hereafter) is Critically Endangered (IUCN 2016) due partly to habitat disruption, disease, anthropogenic effects, and low genetic diversity (Bury 1979, Tryon and Herman 1990, USFWS 2001, Rosenbaum *et al.* 2007, Tesauro and Ehrenfeld 2007, Ernst and Lovich 2009). GIN possesses slightly heteromorphic XX/XY macro-sex chromosomes (Montiel *et al.* 2016a), while GMU's sex-determining mechanism remains unconfirmed. Current head-starting programs for GMU (USFWS 2001), include artificial egg incubation, and knowing the sex-determining mechanism of artificially-incubated species is critical to avoid producing biased sex ratios. This is a concern in species with temperature-dependent sex determination (TSD) (Morreale *et al.* 1982), but not in GSD species.

The spiny softshell turtle *A. spinifera* (ASP hereafter) and its congeners (*A. ferox* and *A. mutica*) are classified as of Least Concern, however the subspecies *A. spinifera atra* is Critically Endangered, and *A. mutica*'s status may change to Near Threatened once sufficient

data become available (IUCN 2016). ASP has a heteromorphic ZZ/ZW micro-sex chromosome system (Badenhorst *et al.* 2013) which shares homology to the sex chromosome system of the Chinese softshell turtle, *Pelodiscus sinensis* (PSI hereafter) (Kawai *et al.* 2007, Kawagoshi *et al.* 2009). PSI is classified as vulnerable with declining populations (IUCN 2016).

4.3 Methods

Sample sources and DNA extraction

Genomic DNA was sequenced for each focal species (GIN and ASP) to generate low-coverage whole genome data to identify sex-linked markers, using a pipeline modified from Vicoso *et al.* (2013) as summarized below and detailed in the Appendix C. The GIN samples were obtained from a pair of adults (one male, one female) confiscated by Iowa's DNR and local Humane Society as described in Montiel *et al.* (2016a). The ASP samples (one male, one female) were obtained from known-sex adults obtained from a turtle farm.

In order to confirm the sex-specificity of PCR reactions, additional validation DNA was extracted from individuals of known-sex from GIN (20 males and 20 females), ASP (40 males and 40 females), GMU (20 males and 22 females), and PSI (one male and one female) from various sources. In all cases, high molecular weight DNA was extracted using standard protocols and following manufacturer's instructions. GIN DNA samples were obtained via Phenol-Chloroform extraction from Longmire-stabilized blood draws stored at 10°C, taken from free-ranging adults that are part of a separate study. ASP DNA was extracted from ethanol-stabilized muscle tissue of hatchlings stored at -20C following Valenzuela (2009a) using the Gentra Puregene DNA extraction kit (Gentra). GMU DNA samples were obtained

from (a) blood from 2 male and 2 female adults donated by the Virginia Zoo via Phenol-Chloroform extraction and (b) tissue or toe clips from wild-caught and captive-bred adults using the Qiagen DNA extraction kit (Qiagen) as part of a separate study. PSI DNA samples were extracted from cultured cells of a single male and single female specimen as described in Badenhorst *et al.* (2013) using the Gentra Puregene DNA extraction kit (Gentra). DNA quantity and quality was assessed using a NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) prior to sequencing or PCR.

For all species, the sex of individuals was determined from secondary sexual characteristics for adults, and by gonadal inspection three months post-emergence for ASP hatchlings. All protocols were approved by the IACUC of Iowa State University (ASP, PSI), the University of Northern Iowa (GIN), and the University of Tennessee – Knoxville (GMU).

DNA sequencing and identification of sex-linked loci using comparative read mapping (CRM)

DNA of one male and one female individual of each focal species (GIN and ASP, see above) was used to generate four 400bp-insert DNA-Seq libraries, each of which was then sequenced on a single lane of Illumina's HiSeq 2000 platform. Raw DNA-Seq reads were trimmed by removing sequencing adapters and low quality bases (<Q5) using BBDuk from the BBMap software package (Bushnell B. - sourceforge.net/projects/bbmap/). Reads shorter than 35bp were discarded.

To identify potential sex-diagnostic markers linked to the heterogametic sex chromosome (Y- and W-linked loci in GIN and ASP, respectively), we adapted the bioinformatics pipeline of Vicoso *et al.* (2013) for *in silico* comparative read mapping (technical bioinformatics details provided in Appendix B). Briefly, the DNA-seq reads from both sexes were pooled and assembled together, and then the male and female reads were

mapped back onto the *de novo* assembly separately. The ratio of the mapped read coverage between the sexes onto *de novo* genome scaffolds helps identify loci in sex chromosomes in two ways: (1) Reads from X-linked loci should be twice as abundant in females (XX) than in males (XY) (and vice-versa for sex-limited Z-linked genes) and (2) reads from loci in the sex-limited portion of the Y should be present in males and absent in females (and vice-versa for sex-limited W-linked genes). Autosomal or pseudo-autosomal loci should be present in equal copy numbers between the sexes, resulting in a between-sex mapping ratio of approximately one.

Implementing this logic conservatively, scaffolds showing a heterogametic:homogametic (XY/XX, or ZW/ZZ) read mapping ratio greater than 10 were classified as putatively Y- or W-linked, while scaffolds with a homogametic:heterogametic (XX/XY, or ZZ/ZW) ratio of 1.5 - 2.5 were classified as putatively X- or Z-linked. Scaffolds with ratios outside those ranges were classified as "Other" and represent potentially autosomal loci or technical artifacts (Table C.1)

Due to the relatively small size of our assembled *de novo* scaffolds, to reduce false positives such as autosomal indels specific to our sequenced specimens, well-assembled reference genomes from the closest available relatives to our focal species were used to map the putative sex-linked sequences. Larger genomic scaffolds with many clustered Y- or W-like signals were more likely to represent true sex chromosome regions (see Supplementary Materials). Our GIN Y-scaffolds were BLASTed against the *Chrysemys picta* (CPI hereafter) genome (Shaffer *et al.* 2013) (Chrysemys_picta_bellii-3.0.1, Genbank Accession: GCA_000241765.1), a TSD species from the same family (Emydidae). Our ASP W-scaffolds

were BLASTed against the ZW female PSI genome (Wang *et al.* 2013) (PelSin_1.0, Genbank Accession: GCA_000230535.1), a species from the same family (Trionychidae).

Primer design and PCR

Genes and genomic scaffold windows from the reference genomes with the highest density of GIN Y-like or ASP W-like BLAST hits were manually inspected using Geneious (Kearse *et al.* 2012) to detect regions suitable for PCR primer design (e.g. 16-25bp DNA stretches with 40-60% GC content that share melting temperature estimate). At minimum, primers were designed to sit in genomic loci which were present in the heterogametic sex yet absent in the homogametic sex. Ideally a single primer set would yield a distinguishable set of amplicons from each sex to facilitate straightforward data analysis. In the event that such a locus could not be identified, a heterogametic-specific primer set was designed to be multiplexed with an autosomal control primer set.

First, we tested the sex-specificity of putative diagnostic markers using validation DNA from known-sex individuals of the species from which the DNA-seq data were derived. Secondly, we tested the cross-species applicability of the GIN-derived primers on male and female GMU, and of the ASP-derived primers in 1 male and 1 female PSI samples we had available. All validation DNA was diluted to ~30ng/ul prior to PCR. PCR amplification was conducted in 15ul reactions containing ~30-50ng template DNA, 1X Taq buffer, 1.5mM $MgCl_2$, 0.2mM dNTPs, 0.5uM of each primer, and 1U recombinant Taq (Invitrogen). PCR conditions for all reactions were similar: one cycle at 94°C for 10 min, followed by 35 cycles of 94°C for 30 s, 60°C for 30s, and 72°C for 30-120s depending on the expected amplicon size, followed by a product extension step of 72°C for 7m. PCR products were visualized on 1% agarose gels and the presence/absence of expected bands was scored (Table 4.1). In order

to avoid misdiagnosis of sex due to PCR failure, Y-linked or W-linked fragments were only scored as present or absent when the control loci amplified clearly.

4.4 Results

Primer design and PCR

Two diagnostic primer sets were designed for each species (Table 4.1); two Y-linked male-specific target loci for GIN and two W-linked female-specific target loci for ASP. Each ASP primer set was also designed to simultaneously amplify a Z-linked control locus in both sexes. No X-linked controls were identifiable around the Y-linked loci in GIN, so instead, a positive control primer set was designed to amplify an autosomal locus in both sexes in a multiplexed reaction with the diagnostic loci.

For GIN, we designed two pairs of Y-specific diagnostic primers, with one at the 5' end of the CPI locus LOC101949910 (olfactory receptor 1009-like; GIN_OLF hereafter) and the other in an unannotated region of the CPI genomic scaffold NW_004848975 (NW_X8975 hereafter) (Table 4.1). Because these primers were not predicted to amplify any control sequence in females (i.e. few to no female reads mapped to these loci or any homologous gene), a pair of control primers were included in the PCR cocktail to simultaneously amplify a fragment of the TEX15 gene, which was known to amplify in both males and females (Table 4.1). Both pairs of diagnostic primers were designed for multiplexing with the control primers with minimal cross-reactivity.

The control TEX15 locus amplified in all validation GIN individuals (n=40), whereas the two sex-diagnostic primers amplified exclusively in males (n=20) and not in females (n=20) (Figure 4.1A-B). These primers were then cross-applied in the GMU samples of

known sex. All GMU samples show clear TEX15 amplification (n=42), whereas only males (n=20) and no females (n=22) showed GIN_OLF amplification, consistent with the XX/XY pattern from GIN (Figure 4.1C). However, the NW_X8975 locus failed to amplify in any GMU (results not shown).

For ASP, we designed two pairs of sex-diagnostic primers corresponding to the PSI-annotated loci LOC102460627 (histone-lysine N-methyltransferase SETD1B-like) and LOC102454141 (tyrosine-protein phosphatase non-receptor type 11-like), which represent variants of the canonical genes SETD1B and PTPN11, respectively. Results from the CRM analysis suggested that these loci were Z/W-linked. We detected *de novo* ASP scaffolds containing the canonical SETD1B and PTPN11 sequences whose read mapping ratios suggested Z-linkage, plus scaffolds containing variant sequences (putative paralogs) whose ratios suggested W-linkage. In both SETD1B and PTPN11, the canonical and variant loci contain conserved regions as well as portions that differ between the putative Z and W copies. Therefore, we designed a single diagnostic primer pair in the conserved regions, but which flanked the variable areas, in order to simultaneously amplify diagnostic W-linked fragments and control Z-linked fragments which are easily discernable by their amplicon size.

Z-linked fragments of SETD1B and PTPN11 amplified in all validation ASP individuals (n=80), whereas W-linked variants amplified exclusively in females (n=40) and not in males (n=40) (Figure 4.1D-E). These primers were tested in one male and one female PSI, and both showed clear amplification of the Z-linked SETD1B and PTPN11 fragments, whereas the W-linked variants amplified only in the female (Figure 4.1F).

4.5 Discussion

Identification and validation of sex markers

We used two lanes of Illumina Hi-Seq data per focal species to assemble a low coverage *de novo* genome for two distantly related GSD turtle species, *Glyptemys insculpta* and *Apalone spinifera*, by modifying the comparative read mapping pipeline of Vicoso *et al.* (2013). From these data, we then designed two pairs of diagnostic PCR primers per species that were 100% accurate in assessing the sex for all individuals tested as described below (no individual was misclassified, Figure 4.1A-B,D-E,). Importantly for species of conservation concern, sex diagnosis by PCR utilized less than 100ng DNA, which is easily obtainable non-lethally from a small blood or tissue sample.

Robust validation using numerous individuals of known sex is essential when developing sexing techniques (Ceballos *et al.* 2014, Literman *et al.* 2014, Gómez-Saldarriaga *et al.* 2016), as well as proper PCR controls (Robertson and Gemmell 2006). We applied two controls here. First, to account for PCR failure we designed each PCR reaction to amplify one control fragment in every individual (autosomal- or X/Z-locus), and an additional fragment that should be present in heterogametic individuals exclusively (Y/W-locus). Second, we used two sex-diagnostic loci per focal species to avoid misdiagnosis due to null alleles (Robertson and Gemmell 2006). Results from both loci were consistent in every individual from the focal species, and there were no individuals from which we could not conclusively determine their sex.

Cross-species applicability of these primer sets can also be tested *in silico* if male- and female-specific genomics data are available. However, due to potential sample-specific artifacts in genomics data (e.g. differential sequencing depth, non-sex-linked variation

between samples), it is necessary to validate any primers with a robust PCR experiment as implemented here.

Leveraging reference genomes (Shaffer *et al.* 2013, Wang *et al.* 2013) permitted the identification of larger genomic regions from close relatives characterized by contiguous sex-diagnostic signals from our focal species, which were more suggestive of true sex linkage versus specimen-specific indels that may exist among individuals within populations. Notably, sex markers can also be identified without a reference genome using only the *de novo* genome data (albeit with a higher probability of false positives) by focusing on longer *de novo* scaffolds with heterogametic-only read coverage to develop PCR markers (see Appendix C).

Cross-species application and the sex-determining system of *Glyptemys muhlenbergii*

Sex-diagnostic markers can be tested for their applicability across species that share the same GSD mechanism (Gamble *et al.* 2014, Rovatsos *et al.* 2015). For instance, our results from 42 GMU individuals of known sex using the GIN_OLF primers showed a consistent pattern with that observed in GIN (Figure 4.1C). These observations provide the first reliable evidence that GMU possesses an XX/XY GSD system which likely arose at the split of *Glyptemys* from other Emydidae turtle lineages ~20Mya (Valenzuela and Adams 2011). Interestingly, the GIN_X8975 primers did not amplify in any GMU samples, suggesting that molecular evolution has accrued in the sex chromosomes of these species since their split from each other, at least at one of the primer binding sites. Incubation experiments with GMU are precluded by its Critically Endangered status, such that until now, only one inconclusive datum existed suggestive of GSD (Ewert and Nelson 1991). *Glyptemys'* XY system represents the youngest known turtle sex chromosomes (Montiel *et*

al. 2016a). Future cytogenetic characterization of GMU's XX/XY system is warranted to test if morphological evolution has also accrued in this sex chromosome pair between GIN and GMU.

Second, ASP and PSI share a ZZ/ZW system that is undistinguishable cytogenetically (Badenhorst *et al.* 2013). Our preliminary test with only one male and one female PSI indicated that ASP markers are suitable for sex identification in PSI (Figure 4.1F), but validation with additional samples remains necessary. Based on the observed conservation of the ASP diagnostic loci in PSI after ~95 My of divergence (Valenzuela and Adams 2011), we hypothesize that our ASP primers may also identify sex in the critically-endangered subspecies *A. spinifera atra* and the potentially endangered *A. mutica,* but tests with individuals of known sex are required in each case. We note that this PCR method for ASP is simpler than the qPCR sexing technique developed previously based rDNA copy number variation (Literman *et al.* 2014), because the sex of a single individual may be assessed by PCR whereas classification by qPCR requires a data distribution obtained from many other individuals of known sex (Literman *et al.* 2014)

Conservation implications and conclusions

Due to the declining populations of GIN and GMU, conservation efforts attempting to bolster populations include active management. Our newly-developed markers now enable sex-specific studies of hatchling and juvenile behavior (Tuttle and Carroll 2005, Tamplin 2006), migration (Castellano *et al.* 2008, Curtis and Vila 2015) and survival (Walde *et al.* 2007, Paterson *et al.* 2012). Because many *Glyptemys* populations are isolated (Rosenbaum *et al.* 2007, Spradling *et al.* 2010, Shoemaker and Gibbs 2013) genetic diversity and population connectivity are of concern, and sex information is important if captive breeding

(Williams and Osentoski 2007) and head-starting programs (Michell and Michell 2015) involve transplanting hatchling or juvenile turtles between adjacent populations or active sex ratio manipulation to enhance population growth. On the other hand, confirming that GMU has GSD means that head-start programs need not worry about causing biased sex ratios by incubation temperature, as occurs with TSD species (Morreale *et al.* 1982). Instead, incubation conditions for GMU should be optimized to maximize hatchling fitness. Importantly, the diagnostic markers developed here can also be used to detect sex reversals (i.e., the mismatch between phenotypic and genotypic sex) both in nature or captivity, as occurs in the ZZ/ZW bearded dragon (*Pogona vitticeps*) in response to extreme temperatures (Quinn *et al.* 2007), or in species where environmental contaminants may cause sex reversal in GSD species (Mizoguchi and Valenzuela 2016, Tamschick *et al.* 2016). In conclusion, we identified sex-diagnostic primers using a pipeline that should be applicable to any species with sex chromosomes or sex-specific loci, either XX/XY or ZZ/ZW, including non-model organisms of conservation concern lacking detectable external sexual dimorphism. This could enhance conservation programs by enabling the accurate assessment and management of population sex ratios. These methods permit quick and inexpensive sexing by PCR and we demonstrate its use in turtles, the most endangered vertebrate group (Hoffmann *et al.* 2010). Importantly our cross-species application revealed that the critically-endangered bog turtle, *Glyptemys muhlenbergii*, possesses an XX/XY GSD system shared with its congener *G. insculpta.*

4.6 Acknowledgements

This work was funded in part by a Bern W. Tryon Bog Fund grant from the Knoxville Zoological Gardens to NV and RB, and National Science Foundation grant MCB 1244355 to NV. We greatly appreciate the assistance of Mike Ogle at Zoo Knoxville and Amanda Guthrie at the Virginia Zoo, and Steve DeSimone at Cold Spring Harbor Fish Hatchery and Aquarium for their assistance in this project.

4.7 Tables and Figures

Table 4.1: Primer sequences used in this study for each focal species (*Apalone spinifera* – ASP, and *Glyptemys insculpta* - GIN) along with the expected amplicon number and size for each reaction depending on genotypic sex of the test individual

			Expected PCR Band Count		
Primer Pair	**Forward Sequence (5'-3')**	**Reverse Sequence (5'-3')**	**XX or ZZ**	**XY or ZW**	**Amplicon Size**
TEX15 (GIN)	CAGGAATCTGGATGGAAGTTT	GGTATGGATATGGTGGTGATTAG	1	1	285bp
GIN_OLF (GIN)	GAGGATGAAGCCAGTCACT	GTATCAGGGAGTTCAGAAAGTT	0	1	800bp
NW_X8975 (GIN)	AGAGAGTACGTGGCAGTTCA	ACTCCTTGTGCAGCTGTGA	0	1	840bp
PTPN11 (ASP)	GCTCATGACTATACGCTAAGAGA	ACCTAACACTCTCCCATCCTT	1	2	Z: 860bp W: 1500bp
SETD1B (ASP)	GATCGAATTACATCCTGCCT	TAAATTAGGACTGGAAGACACC	1	2	Z: 1050bp W: 2700bp

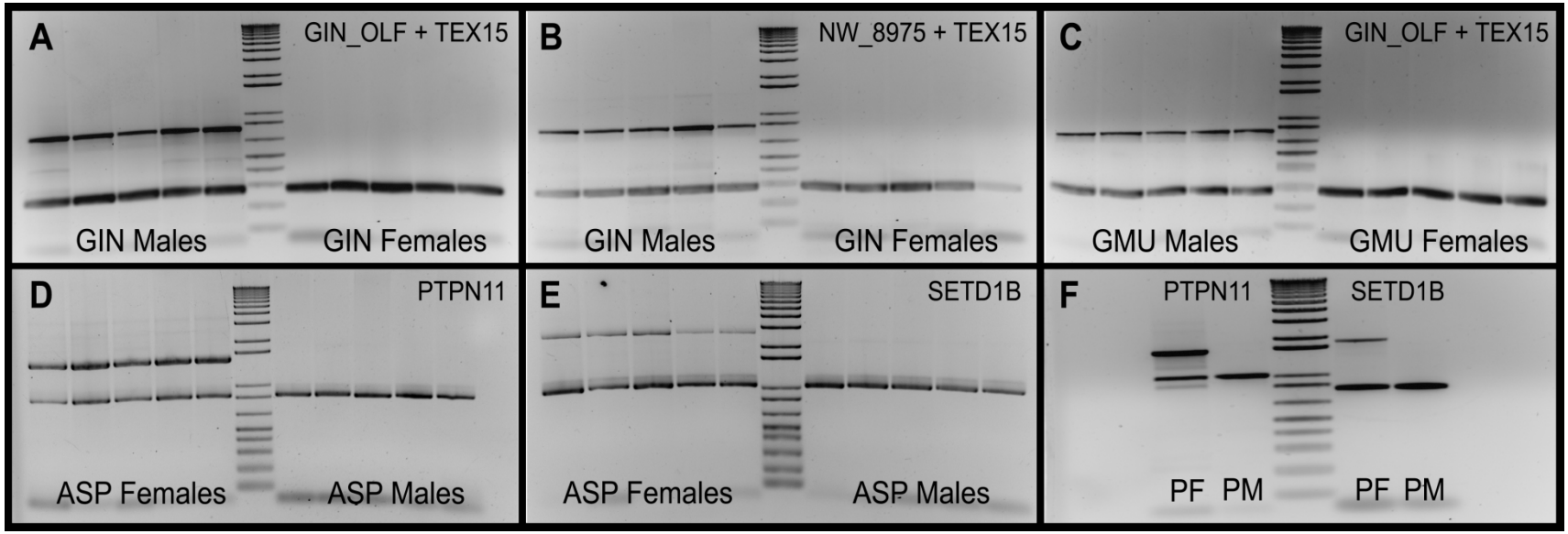

Figure 4.1: Sex-diagnostic PCR results in *Glyptemys insculpta* (GIN), *Glyptemys muhlenbergii* (GMU), *Apalone spinifera* (ASP), and *Pelodiscus sinensis* (PF: female, PM: male). Lower bands correspond to the TEX15 autosomal control (A-C) or Z-linked control loci (D-F). Higher bands correspond to Y- or W-limited sex-diagnostic loci. Ladder = 1kb+ (Invitrogen).

CHAPTER 5: SUMMARY AND CONCLUSIONS

5.1 Summary

The proper development of individuals into males and females is fundamental to the life history of sexually reproducing species, yet the evolution of sex determination mechanisms (SDMs) among vertebrates is surprisingly labile within certain lineages (Valenzuela and Lance 2004, Sarre *et al.* 2011, Beukeboom and Perrin 2014). Turtles represent a model clade to study SDM evolution as among their ~300 species (Mittermeier *et al.* 2015) there have been multiple independent transitions between the turtle ancestral condition of temperature-dependent sex determination (TSD) and genotypic sex determination (GSD) (Valenzuela and Lance 2004, Sabath *et al.* 2016). Additionally, a rapidly growing body of genomic, transcriptomic, and cytogenetic data for the group have accelerated the discoveries on the molecular underpinnings of turtle SDM evolution (Shaffer *et al.* 2013, Wang *et al.* 2013, Czerwinski *et al.* 2016, Montiel *et al.* 2016a, Montiel *et al.* 2016b, Radhakrishnan *et al.* 2017). This growing body of research across molecular scales opens the door to answer some of the fundamental questions about the very nature of SDM evolution. In this dissertation I contribute to advance our knowledge in this active field of research by providing insights into the molecular evolution of the turtle sex determination network (Chapter 2), by developing tools for the accurate diagnosis of sex in GSD species that enable studies of sex-specific traits and the monitoring of sex ratios for conservation (Chapters 3 and 4), by discovering the SDM of a critically-endangered turtle (Chapter 4), as well as by identifying the content of sex chromosomes that enable the study of their idiosyncratic evolutionary dynamics (Chapters 3 and 4). I first describe key steps in SDM evolution to provide the context for the broader implications of my findings.

5.2 SDM Evolution 101

Major transitions in SDM occur when a new sex-determining trigger is introduced into the sex determination network whose impact on sex determination outweighs that of the previous trigger. During shifts from TSD to GSD or between GSD systems, a new genotypic trigger can evolve in a number of different ways. For instance, the sex-determining factor in therian mammals, including the marsupials and placental mammals, is the Y-linked gene *Sry* which has been proposed to have evolved through a duplication of the *Sox3* gene, followed by neofunctionalization (Foster and Graves 1994, Waters *et al.* 2007). In other cases the sex determination network may be hijacked by a small, sex-determining substitution in a gene that already acts in the sex determination network, such as in the *Takifugu rubipes* fish where the sex-determining factor evolved through a single nucleotide polymorphism in the gene *Amhr2* (Kamiya et al. 2012). In the specific case of TSD-to-GSD transitions, individuals in the ancestral TSD population share the full complement of genes required to produce either males or females, and the only required molecular change to achieve a transition towards GSD is a mutation whose presence or absence in a genotype is capable of directing sexual development independent of temperature, thus creating a novel sex-determining locus. These transitions need not erase all temperature-sensitivity from every step in the developmental pathway, but only at a few critical steps, which is evident from the relic thermosensitive expression of sex determination genes that is seen in some GSD organisms, such as the *Wt1* gene in the ZZ/ZW *Apalone spinifera* (Valenzuela 2008b), a genus where temperature has no effect on sex determination (Janzen 1993, Literman *et al.* 2014). As the new GSD system becomes established in the population, selection will then favor the evolution or translocation of sexually antagonistic genes onto the heterogametic sex chromosome (Y or W

chromosomes) which enhance the fitness of one sex over the other (Parker *et al.* 1972, Trivers 1972, Cox and Calsbeek 2009), establishing a block of lower recombination that protects the linkage between sexually antagonistic genes and the sex-determining region. On the other hand, a transition from an established GSD system towards TSD would be expected to have additional complications. For instance, during the first stages of this transition, temperature would induce sex reversal of GSD individuals, some of which may inherit dysfunctional chromosome pairings (e.g. YY or WW individuals) which would likely have negative fitness effects (Bull 1983). Furthermore, depending on the extent of sexually-antagonistic content on the heterogametic sex chromosome, thermally-reversed individuals may express sexually-antagonistic phenotypes that are mismatched to their genotypic sex (e.g. females with male-beneficial genotypes that are detrimental to females). The negative fitness effects of such events have been shown in experiments where genotypic sex is reversed through hormonal or environmental treatments (Warner and Shine 2008, Cotton and Wedekind 2009). Thus, successful GSD to TSD transitions would require the de-differentiation or the partial/total loss of sex chromosomes at the same time that the sexually-antagonistic pathways of gonadal development in the ancestral GSD population would need to be rewired to respond to temperature, all of which may involve more drastic shuffling of genes within the sex determination network relative to TSD-to-GSD transitions.

5.3 Conclusions

GSD and TSD represent extreme ends of a sex determination continuum (Sarre *et al.* 2004), and understanding the molecular causes and effects of SDM transitions among lineages provides vital clues towards unlocking the mystery of both the proximate and ultimate explanations for the evolutionary lability of SDMs among species. Overall, I found

that the evolutionary rate of turtle genes and proteins in the sex determination network was significantly lower than that of mammals, squamates, and birds, and was comparable if not faster than that of crocodilians. Despite their relatively slow molecular evolution as a clade, turtles show remarkable lability in SDM evolution (Valenzuela and Lance 2004), suggesting that a slower basal substitution rate in sex determination network genes hardly precludes SDM transitions, and that the molecular changes required to shift from one SDM to another can occur even in the face of high overall genetic conservation. As the master sex-determining trigger changes among lineages, depending on where in the sex determination cascade that change occurs, the molecular network responsible for proper testicular or ovarian development must accommodate other changes in its circuitry, as in order produce males and females the genes and their associated proteins must work in concert to prevent infertile or unviable offspring (Vicoso and Bachtrog 2013). My findings in Chapter 2 support this idea, as despite their slower overall evolutionary rate relative to other vertebrates, sex determination network genes in turtle lineages which have undergone an SDM transitions exhibit faster nucleotide evolution when compared to lineages which have retained their ancestral SDM. When considering transitions between the extremes of TSD and GSD, I found that as hypothesized, significant differences in the evolutionary dynamics are evident when contrasting transitions either to or from a TSD condition. GSD-to-TSD lineages evolve faster at the nucleotide and protein level relative to both non-transitional and TSD-to-GSD transitional lineages, supporting the hypothesis that GSD-to-TSD transitions require a greater overall adjustment in the sex determination network. Notably, while the evolutionary rate of nucleotides for TSD-to-GSD lineages was faster than that of non-transitional branches, the evolutionary rate of focal proteins was not significantly different from that of non-transitional

lineages. I interpret this finding as further support for the hypothesis, as while TSD-to-GSD transitions are predicted to correlate with changes in the molecular machinery underlying sex determination, these changes need not be as drastic as those required for GSD-to-TSD transitions, because transforming TSD into GSD just needs the evolution of a single sex-determining locus which acts as a trigger to activate pre-established sex developmental pathways without further modification. Nucleotide substitutions in the TSD-to-GSD branches lead to fewer amino acid substitutions than those of GSD-to-TSD transitions, suggesting that the proteins themselves retain their ancestral roles within the network, perhaps changing more in their regulation than in their functional role, a hypothesis worthy of future research. In contrast, the increased rate in protein evolution in GSD-to-TSD transitional lineages supports the hypothesis that as temperature takes over the role as sex-determining factor, the overturn of a pre-existing GSD system necessitates substantial network rewiring.

Whether the acceleration of molecular evolutionary rate in transitional turtle lineages represents a causal agent which facilitates SDM transition, or alternatively, if instead what we are seeing is a snapshot of compensatory molecular evolution after the SDM transition occurred remains to be tested. Adding sequence data across a wider distribution of turtles would allow for higher precision with respect to identifying where and when evolutionary rate changes occurred across the phylogeny, which would further elucidate the evolutionary processes underlying SDM transition. Importantly, under an alternative evolutionary hypothesis where only TSD-to-GSD transitions were reconstructed among turtles (Sabath *et al.* 2016), no effect of SDM transition on evolutionary rate was found, a result which leads to a fundamentally different conclusion with respect to the proposed hypotheses. Under this evolutionary hypothesis, the evolutionary rates of the focal sex determination network genes

and proteins in transitional and non-transitional turtle lineages mirror each other, which suggest that transitions between SDMs are not characterized by changes in the protein-coding portion of the sex determination network, but rather at some level not analyzed as part of this dissertation. If this evolutionary hypothesis turns out to be a more accurate reflection of turtle SDM evolution, future research should investigate other factors which play a role in the sex determination network, with transcription-regulatory loci serving as an ideal candidate for future study. The discrepancy in results under the contrasting evolutionary hypotheses highlights the importance of accurate ancestral state reconstruction when analyzing such datasets.

Scaling down from the evolutionary to the molecular scale, while the overall rate of protein evolution in TSD-to-GSD turtle lineages was not significantly different from that of non-transitional branches, we were still able to identify amino acid substitutions in functional protein domains in these turtles, along with GSD-to-TSD lineages, which may represent key evolutionary steps underlying SDM evolution. Substitutions within functional domains of the proteins HSF2 and RSPO1 were detected in turtle lineages which have transitioned both from TSD-to-GSD and vice versa, while substitutions in the LHX9 protein were detected in *Carettochelys insculpta* and *Podocnemis expansa*, both independent representatives of GSD-to-TSD lineages under the SDM evolutionary hypothesis of Valenzuela and Adams (2011). RSPO1 is a critical regulator of the ovarian developmental pathway (Parma *et al.* 2006), LHX9 is involved in the formation and development of the bipotential gonad (Shima *et al.* 2012), while HSF2 is both integrated into the androgen signaling pathway (Wang *et al.* 2003) and the thermal-transduction pathway (Yoshima *et al.* 1998). Based on their roles in the sex determination and thermal-sensing pathways, substitutions in the functional domains of these

proteins provide tantalizing targets for future studies on the evolution to and from TSD systems. The substitutions described in this study occur in domains which are typically involved in protein:protein interactions as opposed to DNA binding sites, and thus, co-precipitation experiments investigating these relationships may be fruitful. To date, few studies have functionally examined specific genetic mutations which lead to a transition to or from TSD, and the results presented in this dissertation provide *a priori* candidate proteins worthy of further examination.

Our ability to examine and understand the molecular mechanisms underlying vertebrate sex determination relies on accurate sex diagnosis of the study organism, which can be precluded in certain species or developmental stages which lack obvious sexual dimorphism. In Chapters 3 and 4 of this dissertation, I provide analytical pipelines to facilitate molecular sexing of individuals which each can be applied with only nanograms of DNA, opening the door to the development of molecular sexing markers for many GSD taxa from very little starting genetic material, facilitating their use with species of conservation concern. The use of mixture models and univariate discriminant analysis (Fraley *et al.* 2012) in Chapter 3 (Literman *et al.* 2014) permits the accurate sex diagnosis of the ZZ/ZW *A. spinifera* by leveraging sex-specific copy number variation of 18S rDNA between males and females. Sex-biased rDNA distribution can be found in other taxa, including the XX/XY Mexican musk turtle *Staurotypus triporcatus* (Bickham and Rogers 1985) along with some fishes (Born and Bertollo 2000, Takehana *et al.* 2012), frogs (Schmid *et al.* 1983, Schmid *et al.* 1993, Abramyan *et al.* 2009a), bats (Goodpasture and Bloom 1975), and even potato aphids (Monti *et al.* 2011). The prevalence for sex-biased rDNA copy number among diverse species expands the utility of this assay far beyond that of a single turtle taxon. Furthermore,

the analytical pipeline is applicable for any type of continuous variable with an *a priori* known sex-biased distribution, as we show in this study through its application for sex diagnosis using circulating testosterone levels in the TSD snapping turtle *Chelydra serpentina*. Thus, this method could be applied to other sexually-dimorphic continuous dataset from univariate or multivariate traits, such as shape or behavior. The bioinformatics pipeline presented in Chapter 4 (Literman *et al.* 2017) is presented as a blueprint for the development of PCR-diagnostic primers in any GSD species with sufficient genetic differentiation between the sexes. Notably, this pipeline does not require any *a priori* knowledge about the genotypic nature of the GSD system, as it identifies *de novo* sex-linked DNA and thus can be used even in the absence of a reference genome. Sex identification of embryonic turtle samples, among other taxa, is nearly impossible without the use of molecular markers, as morphological differences between the sexes rarely manifest themselves during the developmental periods when sex is being determined. By enabling sex diagnosis from small amounts of template DNA, the studies presented in this dissertation open the door to unprecedented sex-specific studies of sex determination pathways throughout embryogenesis which were previously precluded due to lack of reliable sex markers, particularly in non-model systems. The tools described here can also be used in a wider range of applications where sex identification of monomorphic individuals can be leveraged, from laboratory work to field ecology studies. In the case of juvenile turtles, sex-specific studies of early mortality, growth, migration, and behavior have been historically precluded as clear sexual dimorphisms which are adequate for sex diagnosis often do not emerge until years after hatching, and at these young life stages sex diagnosis typically requires either invasive or destructive methods (e.g. visual inspection of gonads). Using the

pipelines presented in Chapter 4 of this dissertation, molecular markers can now be developed for any suitable GSD taxa, which will allow for sex-specific studies of any number of traits without the need for destructive methods, thus complementing other such methods developed for TSD taxa. Furthermore, with an *a priori* knowledge about sex-biased continuous variables (e.g. gene expression patterns, methylation patterns, geometric morphometrics), the sex of TSD species could also be diagnosed using the pipeline from Chapter 3, with sex diagnosis accuracy varying based on the nature of the distribution overlap between sexes.

The second critical factor that is needed to understand the molecular dynamics underlying SDM evolution is the identity of the sex-determining trigger. While the sex determination network may include sex-linked and autosomal genes, sex determination in GSD species is set off by genotypic triggers which exist in different copy numbers between males and females, either through a presence-absence mechanism (e.g. the Y-linked *Sry* gene in therian mammals) or through a dosage mechanism (e.g. the Z-linked *Dmrt1* gene in birds). As SDM transitions are characterized by a change in the sex-determining trigger, and as sexually antagonistic loci may also cluster on sex chromosomes, understanding the genomic content of sex chromosomes is critical to our understanding of SDM evolution. *De novo* identification of sex-linked DNA (Chapter 4) allows for preliminary characterization of the sex chromosome content in GSD taxa, a pipeline with many applications. While this study was performed using a single lane of 150bp HiSeq DNA-sequencing per sex, the pipeline itself is fully amenable to more complex sequencing strategies which could yield even more complete genomic assemblies and allow for more complete annotation. Aside from sex-diagnostic PCR design, which was the focus of Chapter 4, the applications of identifying sex-

linked DNA are numerous. Identification of sex-linked DNA in the XX/XY wood turtle *Glyptemys insculpta* facilitated the first definitive evidence that its congener the critically-endangered bog turtle *G. muhlenbergii* had XX/XY sex chromosomes, adding a critical piece of data to our knowledge about SDM evolution and distribution among turtles. From a conservation perspective, knowing that a turtle species has GSD also has functional implications for artificial egg-rearing, as in TSD species egg incubation temperature must be carefully managed to ensure that hatchling sex ratios are in line with desired results (Morreale *et al.* 1982). Once GSD is confirmed within a species, egg incubation temperatures can instead be set to those that maximize hatchling survival and fitness.

While full annotation of the sex-linked DNA from the focal species in this study is forthcoming, the data that has already been analyzed sheds light on some notable differences in the evolutionary dynamics between these two independent SDM transitions. Comparing the W-linked sequences from *A. spinifera* to the published genome of *Pelodiscus sinensis*, a turtle from the same family (Trionychidae) with homologous sex chromosomes (Badenhorst *et al.* 2013), we found that the W chromosome of *A. spinifera* contains many full-length, in-frame protein coding genes, many of which have Z-linked homologs. Furthermore, the two genes selected as candidate sex-diagnostic loci in *A. spinifera*, *Setd1b* and *Ptpn11*, reside on chromosome 15 of chicken which previous studies have shown to be homologous to the ZW chromosome pairs of *A. spinifera* and *P. sinensis* (Kawagoshi et al. 2009, Badenhorst et al. 2013). Through *de novo* identification of sex-linked DNA, we are able to both confirm suspected chromosomal homologies, but also identify homologous chromosomal breakpoints and novel relationships between the sex chromosomes and autosomes among taxa where cytogenetic information is available. Contrary to the results found in *A. spinifera*, comparing

the Y-linked sequences from *G. insculpta* to the published genome of the painted turtle *Chrysemys picta*, also a turtle from the same family (Emydidae), revealed almost no full-length genes that could be assigned to the male-specific regions of the Y chromosome, and no X-linked homologs to Y-linked loci were identified, suggesting a different evolutionary trajectory relative to that of *A. spinifera*. Although it is possible that male-specific Y-linked genes in *G. insculpta* evolved *de novo* in that lineage which would preclude their discovery via homology searching, cytogenetic studies of the *G. insculpta* Y chromosome indicate that the male-specific regions on the Y chromosome are characterized by chromosomal inversion events relative to the X chromosome, and are limited in size while containing high levels of repeat content, both factors which suggests that those regions may be gene poor (Montiel *et al.* 2016a). Results from Chapter 4 provide support for that hypothesis, although increased sequencing effort and full annotation studies should be performed to test for the presence of protein-coding genes in those regions. In the case of *G. insculpta,* perhaps it is instead the case that non-protein components play a role in sex determination such as small-RNAs, which are known to accumulate near repeat-heavy genomic areas (Reinhart and Bartel 2002), and are involved in sex determination in insects and plants (Akagi *et al.* 2014, Kiuchi *et al.* 2014). The XY sex chromosome system of *G. insculpta* and *G. muhlenbergii* represents the youngest known sex chromosome system to have evolved within turtles, dated to ~20 million years old (Valenzuela and Adams 2011), and the lack of significant accumulation of sexually-antagonistic genes onto the male-specific regions of Y chromosome may be indicative of their young age, as this system may be only at the first steps in the evolution of the sex chromosomes in this clade.

In conclusion, the studies presented in this dissertation add to a growing body of research investigating the putative causes, effects, and molecular mechanisms underlying SDM transitions, using turtles as a model clade. By describing pipelines with unrestricted taxonomic utility, I truly hope that the work described here can facilitate a flurry of new research in a diverse range of non-model organisms, greatly expanding our overall knowledge about the distribution, evolution, and genetic components of SDMs across vertebrates. The more we know about the molecular nature of sex determination from a wider distribution of species, which accomplish a shared goal via unique and independent pathways, the more completely we will be able to understand how these systems have evolved in the past, and continue to evolve into the future.

REFERENCES

Abascal, F., R. Zardoya and D. Posada (2005). "ProtTest: selection of best-fit models of protein evolution." Bioinformatics **21**(9): 2104-2105.

Abramyan, J., T. Ezaz, J. A. M. Graves and P. Koopman (2009a). "Z and W sex chromosomes in the cane toad (Bufo marinus)." Chromosome Research **17**(8): 1015-1024.

Abramyan, J., C.-W. Feng and P. Koopman (2009b). "Cloning and Expression of Candidate Sexual Development Genes in the Cane Toad (Bufo marinus)." Developmental Dynamics **238**(9): 2430-2441.

Akagi, T., I. M. Henry, R. Tao and L. Comai (2014). "A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons." Science **346**(6209): 646-650.

Akyuz, B., O. Ertugrul, M. Kaymaz, H. C. Macun and D. Bayram (2010). "The effectiveness of gender determination using polymerase chain reaction and radioimmunoassay methods in cattle." Theriogenology **73**(2): 261-266.

Alasaad, S., J. Fickel, R. C. Soriguer, Y. P. Sushitsky and G. Chelomina (2013). "Quantitative sexing (Q-sexing) technique for animal sex-determination based on X chromosome-linked loci: empirical evidence from the Siberian tiger." African Journal of Biotechnology **12**(1): 14-18.

Aresco, M. J. (2005). "The effect of sex-specific terrestrial movements and roads on the sex ratio of freshwater turtles." Biological Conservation **123**(1): 37-44.

Ashman, T., D. Bachtrog, H. Blackmon, E. Goldberg, M. Hahn, M. Kirkpatrick*, et al.* (2014). "Tree of Sex: A database of sexual systems." Scientific Data **1**.

Bachtrog, D., M. Kirkpatrick, J. E. Mank, S. F. McDaniel, J. C. Pires, W. Rice*, et al.* (2011). "Are all sex chromosomes created equal?" TRENDS in Genetics **27**(9): 350-357.

Badenhorst, D., L. W. Hillier, R. Literman, E. E. Montiel, S. Radhakrishnan, Y. Shen*, et al.* (2015). "Physical mapping and refinement of the painted turtle genome (*Chrysemys picta*) inform amniote genome evolution and challenge turtle-bird chromosomal conservation." Genome biology and evolution **7**(7): 2038-2050.

Badenhorst, D., R. Stanyon, T. Engstrom and N. Valenzuela (2013). "A ZZ/ZW microchromosome system in the spiny softshell turtle, *Apalone spinifera*, reveals an intriguing sex chromosome conservation in Trionychidae." Chromosome Research **21**(2): 137-147.

Ballester, M., A. Castello, Y. Ramayo-Caldas and J. M. Folch (2013). "A Quantitative Real-Time PCR Method Using an X-Linked Gene for Sex Typing in Pigs." Molecular Biotechnology **54**(2): 493-496.

Barske, L. A. and B. Capel (2010). "Estrogen represses SOX9 during sex determination in the red-eared slider turtle *Trachemys scripta*." Developmental biology **341**(1): 305-314.

Baudry, J. P., A. E. Raftery, G. Celeux, K. Lo and R. Gottardo (2010). "Combining Mixture Components for Clustering." Journal of Computational and Graphical Statistics **19**(2): 332-353.

Beukeboom, L. W. and N. Perrin (2014). The evolution of sex determination, Oxford University Press, USA.

Bickham, J. W. and J. M. Legler (1983). "Karyotypes and evolutionary relationships of trionychoid turtles." Cytologia **48**(1): 177-183.

Bickham, J. W. and D. S. Rogers (1985). "Structure and variation of the nucleolus organizer region in turtles." Genetica **67**(3): 171-184.

Birk, O. S., D. E. Casiano, C. A. Wassif, T. Cogliati, L. Zhao, Y. Zhao*, et al.* (2000). "The LIM homeobox gene Lhx9 is essential for mouse gonad formation." Nature **403**(6772): 909-913.

Boisvert, S., F. Laviolette and J. Corbeil (2010). "Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies." Journal of computational biology **17**(11): 1519-1533.

Born, G. G. and L. A. C. Bertollo (2000). "An XX/XY sex chromosome system in a fish species, Hoplias malabaricus, with a polymorphic NOR-bearing x chromosome." Chromosome Research **8**(2): 111-118.

Bredbacka, P., A. Kankaanpää and J. Peippo (1995). "PCR-sexing of bovine embryos: a simplified protocol." Theriogenology **44**(2): 167-176.

Bull, J. J. (1983). Evolution of sex determining mechanisms, The Benjamin/Cummings Publishing Company, Inc.

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

Bury, R. B. (1979). "Review of the ecology and conservation of the bog turtle, *Clemmys muhlenbergii*." Special Scientific Report - Wildlife **219**: 1-9.

Bustin, S. A. (2000). "Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays." Journal of molecular endocrinology **25**(2): 169-193.

Castellano, C. M., J. L. Behler and G. R. Ultsch (2008). "Terrestrial Movements of Hatchling Wood Turtles (*Glyptemys insculpta*) in Agricultural Fields in New Jersey." Chelonian Conservation and Biology **7**(1): 113-118.

Ceballos, C. P., D. C. Adams, I. J.B. and N. Valenzuela (2012). "Phylogenetic patterns of sexual size dimorphism in turtles and their implications for Rensch´s rule." Evolutionary Biology **40**: 194-208.

Ceballos, C. P., O. E. Hernández and N. Valenzuela (2014). "Divergent Sex-specific Plasticity in Long-lived Vertebrates with Contrasting Sexual Dimorphism." Evolutionary Biology **41**(1): 81-98.

Ceballos, C. P. and N. Valenzuela (2011). "The Role of Sex-specific Plasticity in Shaping Sexual Dimorphism in a Long-lived Vertebrate, the Snapping Turtle Chelydra serpentina." Evolutionary Biology **38**(2): 163-181.

Chaloupka, M. and C. Limpus (2005). "Estimates of sex-and age-class-specific survival probabilities for a southern Great Barrier Reef green sea turtle population." Marine Biology **146**(6): 1251-1261.

Chang, K.-H., R. Li, M. Papari-Zareei, L. Watumull, Y. D. Zhao, R. J. Auchus*, et al.* (2011). "Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer." Proceedings of the National Academy of Sciences **108**(33): 13728-13733.

Charlesworth, B., J. Coyne and N. Barton (1987). "The relative rates of evolution of sex chromosomes and autosomes." The American Naturalist **130**(1): 113-146.

Charnier, M. (1965). "Action of temperature on the sex ratio in the *Agama agama* (Agamidae, Lacertilia) embryo." Comptes Rendus des Seances de la Societe de Biologie et de ses Filiales **160**(3): 620-622.

Charnov, E. L. and J. Bull (1977). "When is sex environmentally determined?" Nature **266**(5605): 828-830.

Chassot, A.-A., F. Ranc, E. P. Gregoire, H. L. Roepers-Gajadien, M. M. Taketo, G. Camerino*, et al.* (2008). "Activation of β-catenin signaling by Rspo1 controls differentiation of the mammalian ovary." Human molecular genetics **17**(9): 1264-1277.

Clinton, M., L. Haines, B. Belloir and D. McBride (2001). "Sexing chick embryos: a rapid and simple protocol." British Poultry Science **42**(1): 134-138.

Cotton, S. and C. Wedekind (2009). "Population consequences of environmental sex reversal." Conservation Biology **23**(1): 196-206.

Cox, R. M. and R. Calsbeek (2009). "Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict." The American Naturalist **173**(2): 176-187.

Curtis, J. and P. Vila (2015). "The Ecology of the Wood Turtle (*Glyptemys insculpta*) in the Eastern Panhandle of West Virginia." Northeastern Naturalist **22**(2): 387-402.

Cutting, A., J. Chue and C. A. Smith (2013). "Just how conserved is vertebrate sex determination?" Developmental Dynamics **242**(4): 380-387.

Czerwinski, M., A. Natarajan, L. Barske, L. L. Looger and B. Capel (2016). "A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*." Developmental Biology **420**(1): 166-177.

da Silva, S. M., A. Hacker, V. Harley, P. Goodfellow, A. Swain and R. Lovell-Badge (1996). "Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds." Nature genetics **14**(1): 62-68.

Darriba, D., G. L. Taboada, R. Doallo and D. Posada (2012). "jModelTest 2: more models, new heuristics and parallel computing." Nature methods **9**(8): 772-772.

Ellegren, H. and B. C. Sheldon (1997). "New tools for sex identification and the study of sex allocation in birds." Trends in Ecology & Evolution **12**(7): 255-259.

Ernst, C. H. and J. E. Lovich (2009). Turtles of the United States and Canada, JHU Press.

Ewert, M. A. and C. E. Nelson (1991). "Sex determination in turtles: diverse patterns and some possible adaptive values." Copeia **1991**(1): 50-69.

Ezaz, T., A. E. Quinn, I. Miura, S. D. Sarre, A. Georges and J. A. M. Graves (2005). "The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes." Chromosome Research **13**(8): 763-776.

Ezaz, T., K. Srikulnath and J. A. M. Graves (2017). "Origin of Amniote Sex Chromosomes: An Ancestral Super-Sex Chromosome, or Common Requirements?" Journal of Heredity **108**(1): 94-105.

Faust, L. J. and S. D. Thompson (2000). "Birth sex ratio in captive mammals: patterns, biases, and the implications for management and conservation." Zoo Biology **19**(1): 11-25.

Foster, J. W. and J. Graves (1994). "An SRY-related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene." Proceedings of the National Academy of Sciences **91**(5): 1927-1931.

Fraley, C., A. Raftery and L. Scrucca (2012). "mclust version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation." Technical Report No. 597 **Department of Statistics, University of Washington.**

Fraley, C. and A. E. Raftery (2002). "Model-based clustering, discriminant analysis, and density estimation." Journal of the American Statistical Association **97**(458): 611-631.

Fujita, M. K., S. V. Edwards and C. P. Ponting (2011). "The Anolis lizard genome: an amniote genome without isochores." Genome biology and evolution **3**: 974-984.

Gamble, T., A. J. Geneva, R. E. Glor and D. Zarkower (2014). "Anolis sex chromosomes are derived from a single ancestral pair." Evolution **68**(4): 1027-1041.

Gamble, T. and D. Zarkower (2014). "Identification of sex-specific molecular markers using restriction site-associated DNA sequencing." Molecular Ecology Resources **14**(5): 902-913.

Garolla, A., M. Torino, B. Sartini, I. Cosci, C. Patassini, U. Carraro*, et al.* (2013). "Seminal and molecular evidence that sauna exposure affects human spermatogenesis." Human Reproduction **28**(4): 877-885.

Gómez-Saldarriaga, C., N. Valenzuela and C. P. Ceballos (2016). "Effects of incubation temperature on sex determination in the endangered magdalena river turtle, *Podocnemis lewyana*." Chelonian Conservation and Biology **15**(1): 43-53.

Goodpasture, C. and S. E. Bloom (1975). "Visualization of nucleolar organizer regions in mammalian chromosomes using silver staining." Chromosoma **53**(1): 37-50.

Graves, J. A. M. (2017). "How Australian mammals contributed to our understanding of sex determination and sex chromosomes." Australian Journal of Zoology **64**(4): 267-276.

Green, R. E., E. L. Braun, J. Armstrong, D. Earl, N. Nguyen, G. Hickey*, et al.* (2014). "Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs." Science **346**(6215): 1254449.

Griffiths, R. (2000). "Sex identification in birds." Seminars in Avian and Exotic Pet Medicine **9**(1): 14-26.

Gross, T. S., D. A. Crain, K. A. Bjorndal, A. B. Bolten and R. R. Carthy (1995). "Identification of Sex in Hatchling Loggerhead Turtles (Caretta-Caretta) by Analysis of Steroid Concentrations in Chorioallantoic/Amniotic Fluid." General and Comparative Endocrinology **99**(2): 204-210.

Hahn, A. H., A. Krysler and C. B. Sturdy (2013). "Female song in black-capped chickadees (Poecile atricapillus): Acoustic song features that contain individual identity information and sex differences." Behavioural Processes **98**: 98-105.

Hamilton, C. K., A. Combe, J. Caudle, F. A. Ashkar, A. D. Macaulay, P. Blondin*, et al.* (2012). "A novel approach to sexing bovine blastocysts using male-specific: gene expression." Theriogenology **77**(8): 1587-1596.

Hedges, S. B., J. Dudley and S. Kumar (2006). "TimeTree: a public knowledge-base of divergence times among organisms." Bioinformatics **22**(23): 2971-2972.

Heid, C. A., J. Stevens, K. J. Livak and P. M. Williams (1996). "Real time quantitative PCR." Genome research **6**(10): 986-994.

Hoffmann, M., C. Hilton-Taylor, A. Angulo, M. Böhm, T. M. Brooks, S. H. M. Butchart*, et al.* (2010). "The Impact of Conservation on the Status of the World's Vertebrates." Science **330**(6010): 1503-1509.

Hsu, T. C., S. E. Spirito and M. L. Pardue (1975). "Distribution of 18+28S ribosomal genes in mammalian genomes." Chromosoma **53**(1): 25-36.

IUCN (2016). "The IUCN red list of threatened species." **Version 2016-3**.

Janes, D. E., C. Organ and N. Valenzuela (2008). "New resources inform study of genome size, content, and organization in nonavian reptiles." Integrative and comparative biology **48**(4): 447-453.

Janzen, F. J. (1993). "The influence of incubation temperature and family on eggs, embryos, and hatchlings of the smooth softshell turtle (*Apalone mutica*)." Physiological Zoology **66**(3): 349-373.

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

Kawagoshi, T., Y. Uno, K. Matsubara, Y. Matsuda and C. Nishida (2009). "The ZW micro-sex chromosomes of the Chinese soft-shelled turtle (*Pelodiscus sinensis*, Trionychidae, Testudines) have the same origin as chicken chromosome 15." Cytogenetic and Genome Research **125**(2): 125-131.

Kawai, A., C. Nishida-Umehara, J. Ishijima, Y. Tsuda, H. Ota and Y. Matsuda (2007). "Different origins of bird and reptile sex chromosomes inferred from comparative mapping of chicken Z-linked genes." Cytogenetic and Genome Research **117**(1-4): 92-102.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock*, et al.* (2012). "Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data." Bioinformatics **28**(12): 1647-1649.

Kim, K.-A., M. Wagle, K. Tran, X. Zhan, M. A. Dixon, S. Liu*, et al.* (2008). "R-Spondin family members regulate the Wnt pathway by a common mechanism." Molecular biology of the cell **19**(6): 2588-2596.

Kim, K.-A., J. Zhao, S. Andarmani, M. Kakitani, T. Oshima, M. E. Binnerts*, et al.* (2006). "R-Spondin proteins: a novel link to β-catenin activation." Cell cycle **5**(1): 23-26.

Kiuchi, T., H. Koga, M. Kawamoto, K. Shoji, H. Sakai, Y. Arai*, et al.* (2014). "A single female-specific piRNA is the primary determiner of sex in the silkworm." Nature **509**(7502): 633-636.

Korstian, J. M., A. M. Hale, V. J. Bennett and D. A. Williams (2013). "Advances in sex determination in bats and its utility in wind-wildlife studies." Molecular Ecology Resources **13**(5): 776-780.

Kumar, S., M. Nei, J. Dudley and K. Tamura (2008). "MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences." Briefings in bioinformatics **9**(4): 299-306.

Lagomarsino, I. V. and D. O. Conover (1993). "Variation in environmental and genotypic sex-determining mechanisms across a latitudinal gradient in the fish, *Menidia menidia*." Evolution: 487-494.

Lance, V. A., N. Valenzuela and P. von Hildebrand (1992). "A hormonal method to determine sex of hatchling giant river turtles, *Podocnemis expansa*: application to endangered species." J.Exp.Zool. **270**: 16A.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nature methods **9**(4): 357-359.

Levell, J. (2000). "Commercial exploitation of Blanding's Turtle, *Emydoidea blandingii*, and the Wood Turtle, *Clemmys insculpta*, for the live animal trade." Chelonian Conservation and Biology **3**(4): 665-674.

Literman, R., D. Badenhorst and N. Valenzuela (2014). "qPCR-based molecular sexing by copy number variation in rRNA genes and its utility for sex identification in soft-shell turtles." Methods in Ecology and Evolution **5**(9): 872-880.

Literman, R., S. Radhakrishnan, J. Tamplin, R. Burke, C. Dresser and N. Valenzuela (2017). "Development of sexing primers in *Glyptemys insculpta* and *Apalone spinifera* turtles uncovers an XX/XY sex-determining system in the critically-endangered bog turtle *Glyptemys muhlenbergii*." Conservation Genetics Resources: 1-8.

Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method." Methods **25**(4): 402-408.

Maldonado, L. T., A. L. Piedra, N. M. Mendoza, A. M. Valencia, A. M. Martınez and H. M. Larios (2002). "Expression profiles of Dax1, Dmrt1, and Sox9 during temperature sex determination in gonads of the sea turtle *Lepidochelys olivacea*." General and comparative endocrinology **129**(1): 20-26.

Mank, J. E., E. Axelsson and H. Ellegren (2007). "Fast-X on the Z: rapid evolution of sex-linked genes in birds." Genome Research **17**(5): 618-624.

Matsumoto, Y., B. Hannigan and D. Crews (2016). "Temperature Shift Alters DNA Methylation and Histone Modification Patterns in Gonadal Aromatase (cyp19a1) Gene in Species with Temperature-Dependent Sex Determination." PloS one **11**(11): e0167362.

Mazaud, S., E. Oreal, C. Guigon, D. Carre-Eusebe and S. Magre (2002). "Lhx9 expression during gonadal morphogenesis as related to the state of cell differentiation." Gene expression patterns **2**(3): 373-377.

Michel-Morfin, J. E., V. M. Gómez Muñoz and C. Navarro Rodríguez (2001). "Morphometric model for sex assessment in hatchling olive ridley sea turtles." Chelonian Conservation and Biology **4**(1): 53-58.

Michell, K. and R. G. Michell (2015). "Use of Radio-Telemetry and Recapture to Determine the Success of Head-Started Wood Turtles (*Glyptemys insculpta*) in New York." Herpetological Conservation and Biology **10**(1): 525-534.

Mittermeier, R. A., P. P. van Dijk, A. G. Rhodin and S. D. Nash (2015). "Turtle hotspots: An analysis of the occurrence of tortoises and freshwater turtles in biodiversity hotspots, high-biodiversity wilderness areas, and turtle priority areas." Chelonian Conservation and Biology **14**(1): 2-10.

Mizoguchi, B. A. and N. Valenzuela (2016). "Ecotoxicological Perspectives of Sex Determination." Sexual Development **10**(1): 45-57.

Monti, V., G. C. Manicardi and M. Mandrioli (2011). "Cytogenetic and molecular analysis of the holocentric chromosomes of the potato aphid Macrosiphum euphorbiae (Thomas, 1878)." Comparative Cytogenetics **5**(3): 163-172.

Montiel, E., D. Badenhorst, J. Tamplin, R. Burke and N. Valenzuela (2016a). "Discovery of the youngest sex chromosomes reveals first case of convergent co-option of ancestral autosomes in turtles." Chromosoma: 1-9.

Montiel, E. E., D. Badenhorst, L. S. Lee, R. Literman, V. Trifonov and N. Valenzuela (2016b). "Cytogenetic insights into the evolution of chromosomes and sex determination reveal striking homology of turtle sex chromosomes to amphibian autosomes." Cytogenetic and Genome Research **148**(4): 292-304.

Morinha, F., J. A. Cabral and E. Bastos (2012). "Molecular sexing of birds: A comparative review of polymerase chain reaction (PCR)-based methods." Theriogenology **78**(4): 703-714.

Morreale, S. J., G. J. Ruiz and E. Standora (1982). "Temperature-dependent sex determination: current practices threaten conservation of sea turtles." Science **216**(4551): 1245-1247.

Morrison, T. B., J. J. Weis and C. T. Wittwer (1998). "Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification." Biotechniques **24**(6): 954-958, 960, 962.

Mrosovsky, N. (1982). "Sex-ratio bias in hatchling sea turtles from artificially incubated eggs." Biological Conservation **23**(4): 309-314.

Nanda, I., Z. Shan, M. Schartl, D. W. Burt, M. Koehler, H.-G. Nothwang*, et al.* (1999). "300 million years of conserved synteny between chicken Z and human chromosome 9." Nature genetics **21**(3): 258-259.

Nguyen, L.-P., N. Galtier and B. Nabholz (2015). "Gene expression, chromosome heterogeneity and the fast-X effect in mammals." Biology letters **11**(2): 20150010.

Oréal, E., S. Mazaud, J. Y. Picard, S. Magre and D. Carré-Eusèbe (2002). "Different patterns of anti-Müllerian hormone expression, as related to DMRT1, SF-1, WT1, GATA-4, Wnt-4, and Lhx9 expression, in the chick differentiating gonads." Developmental Dynamics **225**(3): 221-232.

Ottolenghi, C., C. Moreira-Filho, B. B. Mendonça, M. Barbieri, M. Fellous, G. D. Berkovitz*, et al.* (2001). "Absence of mutations involving the LIM homeobox domain gene LHX9 in 46, XY gonadal agenesis and dysgenesis." The Journal of Clinical Endocrinology & Metabolism **86**(6): 2465-2469.

Owens, D. W., J. R. Hendrickson, V. Lance and I. P. Callard (1978). "Technique for determining sex of immature chelonia-mydas using a radioimmunoassay." Herpetologica **34**(3): 270-273.

Parker, G. A., R. Baker and V. Smith (1972). "The origin and evolution of gamete dimorphism and the male-female phenomenon." Journal of theoretical biology **36**(3): 529-553.

Parma, P., O. Radi, V. Vidal, M. C. Chaboissier, E. Dellambra, S. Valentini*, et al.* (2006). "R-spondin1 is essential in sex determination, skin differentiation and malignancy." Nature genetics **38**(11): 1304-1309.

Paterson, J. E., B. D. Steinberg and J. D. Litzgus (2012). "Revealing a cryptic life-history stage: differences in habitat selection and survivorship between hatchlings of two turtle species at risk (*Glyptemys insculpta* and *Emydoidea blandingii*)." Wildlife Research **39**(5): 408-418.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT–PCR." Nucleic acids research **29**(9): e45-e45.

Phillips, B. C. and S. Edmands (2012). "Does the speciation clock tick more slowly in the absence of heteromorphic sex chromosomes?" Bioessays **34**(3): 166-169.

Pokorna, M. and L. Kratochvíl (2009). "Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap?" Zoological Journal of the Linnean Society **156**(1): 168-183.

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

Quinn, A. E., A. Georges, S. D. Sarre, F. Guarino, T. Ezaz and J. A. M. Graves (2007). "Temperature Sex Reversal Implies Sex Gene Dosage in a Reptile." Science **316**(5823): 411-411.

Radhakrishnan, S., R. Literman, J. Neuwald, A. Severin and N. Valenzuela (2017). "Transcriptomic responses to environmental temperature by turtles with temperature-dependent and genotypic sex determination assessed by RNAseq inform the genetic architecture of embryonic gonadal development." PloS one **12**(3): e0172044.

Raymond, C. S., M. W. Murphy, M. G. O'Sullivan, V. J. Bardwell and D. Zarkower (2000). "Dmrt1, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation." Genes & Development **14**(20): 2587-2595.

Reinhart, B. J. and D. P. Bartel (2002). "Small RNAs correspond to centromere heterochromatic repeats." Science **297**(5588): 1831-1831.

Rétaux, S., M. Rogard, I. Bach, V. Failli and M.-J. Besson (1999). "Lhx9: a novel LIM-homeodomain gene expressed in the developing forebrain." Journal of Neuroscience **19**(2): 783-793.

Rhen, T. and J. W. Lang (1995). "Phenotypic Plasticity for Growth in the Common Snapping Turtle - Effects of Incubation-Temperature, Clutch, and Their Interaction." American Naturalist **146**(5): 726-747.

Robertson, B. C. and N. J. Gemmell (2006). "PCR-based sexing in conservation biology: Wrong answers from an accurate methodology?" Conservation Genetics **7**(2): 267-271.

Rosenbaum, P. A., J. M. Robertson and K. R. Zamudio (2007). "Unexpectedly low genetic divergences among populations of the threatened bog turtle (*Glyptemys muhlenbergii*)." Conservation Genetics **8**(2): 331-342.

Ross, M. T., D. V. Grafham, A. J. Coffey, S. Scherer, K. McLay, D. Muzny*, et al.* (2005). "The DNA sequence of the human X chromosome." Nature **434**(7031): 325-337.

Rostal, D. C., V. A. Lance, J. S. Grumbles and A. C. Alberts (1994). "Seasonal reproductive cycle of the desert tortoise (Gopherus agassizii) in the eastern Mojave Desert." Herpetological Monographs **8**: 72-82.

Rovatsos, M., J. Vukić and L. Kratochvíl (2016). "Mammalian X homolog acts as sex chromosome in lacertid lizards." Heredity.

Rovatsos, M., J. Vukić, P. Lymberakis and L. Kratochvíl (2015). Evolutionary stability of sex chromosomes in snakes. Proceedings of the Royal Society of London B, The Royal Society.

Sabath, N., Y. Itescu, A. Feldman, S. Meiri, I. Mayrose and N. Valenzuela (2016). "Sex determination, longevity, and the birth and death of reptilian species." Ecology and Evolution **6**(15): 5207-5220.

Sarge, K. D., V. Zimarino, K. Holm, C. Wu and R. I. Morimoto (1991). "Cloning and characterization of two mouse heat shock factors with distinct inducible and constitutive DNA-binding ability." Genes & Development **5**(10): 1902-1911.

Sarre, S. D., T. Ezaz and A. Georges (2011). "Transitions between sex-determining systems in reptiles and amphibians." Annual review of genomics and human genetics **12**: 391-406.

Sarre, S. D., A. Georges and A. Quinn (2004). "The ends of a continuum: genetic and temperature-dependent sex determination in reptiles." Bioessays **26**(6): 639-645.

Sato, H. and H. Ota (2001). "Karyotype of the Chinese soft-shelled turtle, *Pelodiscus sinensis*, from Japan and Taiwan, with chromosomal data for *Dogania subplana*." Current herpetology **20**(1): 19-25.

Saumure, R. A., T. B. Herman and R. D. Titman (2007). "Effects of haying and agricultural practices on a declining species: The North American wood turtle, *Glyptemys insculpta*." Biological Conservation **135**(4): 565-575.

Schmid, M., T. Haaf, B. Geile and S. Sims (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., S. Ohta, C. Steinlein and M. Guttenbach (1993). "Chromosome-banding in Amphibia .19. Primitive ZW/ZZ sex-chromosomes in *Buergeria buergeri* (Anura, Rhacophoridae)." Cytogenetics and Cell Genetics **62**(4): 238-246.

Shaffer, B. H., P. Minx, D. E. Warren, A. M. Shedlock, R. C. Thomson, N. Valenzuela*, et al.* (2013). "The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage." Genome Biology **14**(3): R28.

Shaw, P. J. and P. C. McKeown (2011). The Structure of rDNA Chromatin. The Nucleolus. M. O. J. Olson, Springer New York. **15:** 43-55.

Shima, Y., K. Miyabayashi, S. Haraguchi, T. Arakawa, H. Otake, T. Baba*, et al.* (2012). "Contribution of Leydig and Sertoli cells to testosterone production in mouse fetal testes." Molecular endocrinology **27**(1): 63-73.

Shine, R., M. Elphick and S. Donnellan (2002). "Co-occurrence of multiple, supposedly incompatible modes of sex determination in a lizard population." Ecology Letters **5**(4): 486-489.

Shoemaker, K. T. and J. P. Gibbs (2013). "Genetic Connectivity among Populations of the Threatened Bog Turtle (*Glyptemys muhlenbergii*) and the Need for a Regional Approach to Turtle Conservation." Copeia **2013**(2): 324-331.

Singh, A. K. (2013). "Introduction of modern endocrine techniques for the production of monosex population of fishes." General and Comparative Endocrinology **181**: 146-155.

Spradling, T. A., J. W. Tamplin, S. S. Dow and K. J. Meyer (2010). "Conservation genetics of a peripherally isolated population of the wood turtle (*Glyptemys insculpta*) in Iowa." Conservation Genetics **11**(5): 1667-1677.

Steen, D., M. Aresco, S. Beilke, B. Compton, E. Condon, C. Kenneth Dodd*, et al.* (2006). "Relative vulnerability of female turtles to road mortality." Animal Conservation **9**(3): 269-273.

Takehana, Y., K. Naruse, Y. Asada, Y. Matsuda, T. Shin-I, Y. Kohara*, et al.* (2012). "Molecular cloning and characterization of the repetitive DNA sequences that comprise the constitutive heterochromatin of the W chromosomes of medaka fishes." Chromosome Research **20**(1): 71-81.

Tamplin, J. (2006). "Response of hatchling wood turtles (*Glyptemys insculpta*) to an aquatic thermal gradient." Journal of Thermal Biology **31**(5): 400-405.

Tamschick, S., B. Rozenblut-Kościsty, M. Ogielska, A. Lehmann, P. Lymberakis, F. Hoffmann*, et al.* (2016). "Sex reversal assessments reveal different vulnerability to endocrine disruption between deeply diverged anuran lineages." Scientific Reports **6**.

Tesauro, J. and D. Ehrenfeld (2007). "The effects of livestock grazing on the bog turtle [*Glyptemys* (= *Clemmys*) *muhlenbergii*]." Herpetologica **63**(3): 293-300.

Trivers, R. (1972). "Parental investment and sexual selection." Sexual Selection & the Descent of Man, Aldine de Gruyter, New York: 136-179.

Tryon, B. W. and D. W. Herman (1990). Status, conservation, and management of the bog turtle, *Clemmys muhlenbergii*, in the southeastern United States. Proceedings of the First International Symposium on Turtles and Tortoises: Conservation and Captive Husbandry., Chapman University, Orange California.

Tuttle, S. E. and D. M. Carroll (2005). "Movements and Behavior of Hatchling Wood Turtles (*Glyptemys insculpta*)." Northeastern Naturalist **12**(3): 331-348.

Uno, Y., C. Nishida, Y. Oshima, S. Yokoyama, I. Miura, Y. Matsuda*, et al.* (2008). "Comparative chromosome mapping of sex-linked genes and identification of sex chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems." Chromosome Research **16**(4): 637-647.

Urbatzka, R., I. Lutz and W. Kloas (2007). "Aromatase, steroid-5-alpha-reductase type 1 and type 2 mRNA expression in gonads and in brain of *Xenopus laevis* during ontogeny." General and comparative endocrinology **153**(1): 280-288.

USFWS (2001). Bog turtle (*Clemmys muhlenbergii*), northern population recovery plan. Hadley, MA.

Valenzuela, N. (2001). "Constant, shift and natural temperature effects on sex determination in *Podocnemis expansa* turtles." Ecology **82**(11): 3010-3024.

Valenzuela, N. (2008a). "Evolution of the gene network underlying gonadogenesis in turtles with temperature-dependent and genotypic sex determination." Integrative and Comparative Biology **48**(4): 476-485.

Valenzuela, N. (2008b). "Relic thermosensitive gene expression in genotypically-sex-determined turtles." Evolution **62**(1): 234-240.

Valenzuela, N. (2009a). "Egg incubation and collection of painted turtle embryos." Cold Spring Harbor Protocol **4**(7).

Valenzuela, N. (2009b). "The painted turtle, *Chrysemys picta:* a model system for vertebrate evolution, ecology, and human health." Cold Spring Harbor Protocols **2009**(7): pdb. emo124.

Valenzuela, N. (2010). "Multivariate expression analysis of the gene network underlying sexual development in turtle embryos with temperature-dependent and genotypic sex determination " Sexual Development **4**(1-2): 39-49.

Valenzuela, N. and D. C. Adams (2011). "Chromosome Number and Sex Determination Coevolve in Turtles." Evolution **65**(6): 1808-1813.

Valenzuela, N., D. C. Adams, R. M. Bowden and A. C. Gauger (2004). "Geometric morphometric sex estimation for hatchling turtles: A powerful alternative for detecting subtle sexual shape dimorphism." Copeia(4): 735-742.

Valenzuela, N., D. C. Adams and F. J. Janzen (2003). "Pattern does not equal process: exactly when is sex environmentally determined?" The American Naturalist **161**(4): 676-683.

Valenzuela, N. and V. Lance (2004). Temperature-dependent sex determination in vertebrates, Smithsonian Books Washington, DC.

Valenzuela, N., A. LeClere and T. Shikano (2006). "Comparative gene expression of steroidogenic factor 1 in *Chrysemys picta* and *Apalone mutica* turtles with temperature-dependent and genotypic sex determination." Evolution and Development **8**(5): 424-432.

Valenzuela, N., J. L. Neuwald and R. Literman (2013). "Transcriptional evolution underlying vertebrate sexual development." Developmental Dynamics **242**(4): 307-319.

Valenzuela, N. and T. Shikano (2007). "Embryological ontogeny of Aromatase gene expression in *Chrysemys picta* and *Apalone mutica* turtles: comparative patterns within and across temperature-dependent and genotypic sex-determining mechanisms." Development, Genes and Evolution **217**: 55-62.

Venegas, D., A. Marmolejo-Valencia, C. Valdes-Quezada, T. Govenzensky, F. Recillas-Targa and H. Merchant-Larios (2016). "Dimorphic DNA methylation during temperature-dependent sex determination in the sea turtle *Lepidochelys olivacea*." General and Comparative Endocrinology **236**: 35-41.

Vicoso, B. and D. Bachtrog (2013). "Reversal of an ancient sex chromosome to an autosome in Drosophila." Nature **499**(7458): 332-335.

Vicoso, B., J. J. Emerson, Y. Zektser, S. Mahajan and D. Bachtrog (2013). "Comparative Sex Chromosome Genomics in Snakes: Differentiation, Evolutionary Strata, and Lack of Global Dosage Compensation." PLoS Biology **11**(8).

Vogelstein, B. and K. W. Kinzler (1999). "Digital PCR." Proceedings of the National Academy of Sciences of the United States of America **96**(16): 9236-9241.

Walde, A. D., J. R. Bider, D. Masse, R. A. Saumure and R. D. Titman (2007). "Nesting ecology and hatching success of the wood turtle, *Glyptemys insculpta*, in Quebec." Herpetological Conservation and Biology **2**(1): 49-60.

Wang, G., J. Zhang, D. Moskophidis and N. F. Mivechi (2003). "Targeted disruption of the heat shock transcription factor (hsf)-2 gene results in increased embryonic lethality, neuronal defects, and reduced spermatogenesis." Genesis **36**(1): 48-61.

Wang, Z., J. Pascual-Anaya, A. Zadissa, W. Li, Y. Niimura, Z. Huang*, et al.* (2013). "The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan." Nature Genetics **45**(6): 701-706.

Warner, D. and R. Shine (2008). "The adaptive significance of temperature-dependent sex determination in a reptile." Nature **451**(7178): 566-568.

Waters, P. D., M. C. Wallis and J. A. M. Graves (2007). Mammalian sex—origin and evolution of the Y chromosome and SRY. Seminars in cell & developmental biology, Elsevier.

Weissmann, A., S. Reitemeier, A. Hahn, J. Gottschalk and A. Einspanier (2013). "Sexing domestic chicken before hatch: A new method for in ovo gender identification." Theriogenology **80**(3): 199-205.

Williams, D. A. and M. F. Osentoski (2007). "Genetic considerations for the captive breeding of tortoises and freshwater turtles." Chelonian Conservation and Biology **6**(2): 302-313.

Wilson, J. D., J. E. Griffin and D. W. Russell (1993). "Steroid 5α-reductase 2 deficiency." Endocrine Reviews **14**(5): 577-593.

Wood, J. R., F. E. Wood, K. H. Critchley, D. E. Wildt and M. Bush (1983). "Laparoscopy of the green sea turtle, *Chelonia mydas*." British Journal of Herpetology **6**(9): 323-327.

Yntema, C. L. and N. Mrosovsky (1980). "Sexual-differentiation in hatchling loggerheads (*Caretta caretta*) incubated at different controlled temperatures." Herpetologica **36**(1): 33-36.

Yoshima, T., T. Yura and H. Yanagi (1998). "Novel testis-specific protein that interacts with heat shock factor 2." Gene **214**(1): 139-146.

Zhang, H., X. Xu, Z. He, T. Zheng and J. Shao (2017). "De novo transcriptome analysis reveals insights into different mechanisms of growth and immunity in a Chinese soft-shelled turtle hybrid and the parental varieties." Gene **605**: 54-62.

APPENDIX A: SUPPLEMENTAL TABLES FOR CHATPER 2

Table A.1: Length and percent identify of gene alignments used in this study before and after sequence gaps were removed. bp = nucleotide base pairs , aa = amino acids

				Gene							
				AR	**CIRBP**	**CTNNB1**	**CYP19A1**	**DMRT1**	**ESR1**	**ESR2**	**HSF2**
Nucleotide Alignment	**All Species**	**Full Alignment**	**Length (bp)**	2985	534	2343	1530	1188	1800	1684	1719
			Percent Identity	77.9%	86.5%	90.0%	81.7%	78.1%	83.4%	81.0%	81.8%
		Gaps Removed	**Length (bp)**	1119	510	2343	1479	714	1719	1524	1458
			Percent Identity	87.3%	87.5%	90.0%	82.1%	82.0%	84.2%	82.2%	84.3%
	Turtles	**Full Alignment**	**Length (bp)**	2391	522	2343	1512	1119	1764	1668	1683
			Percent Identity	91.5%	99.3%	99.9%	91.3%	94.6%	94.9%	93.3%	94.5%
		Gaps Removed	**Length (bp)**	1245	513	2343	1497	1101	1761	1635	1656
			Percent Identity	94.5%	96.0%	96.6%	93.5%	94.6%	94.4%	94.2%	94.3%
Amino Acid Alignment	**All Species**	**Full Alignment**	**Length (aa)**	989	178	781	511	392	600	561	574
			Percent Identity	81.0%	95.7%	99.7%	83.0%	78.7%	87.0%	84.4%	84.7%
		Gaps Removed	**Length (aa)**	373	170	781	494	238	577	508	488
			Percent Identity	94.0%	97.0%	99.7%	83.5%	91.8%	97.8%	86.3%	87.6%
	Turtles	**Full Alignment**	**Length (aa)**	797	174	781	504	373	588	556	561
			Percent Identity	91.5%	99.3%	99.9%	91.3%	94.6%	94.9%	93.3%	94.5%
		Gaps Removed	**Length (aa)**	415	171	781	499	367	587	545	552
			Percent Identity	95.7%	99.8%	99.9%	91.6%	95.3%	95.1%	93.7%	95.1%

Table A.1 continued: Length and percent identify of gene alignments used in this study before and after sequence gaps were removed. bp = nucleotide base pairs , aa = amino acids

				Gene						
				LHX9	**NR5A1**	**RSPO1**	**SOX9**	**SRD5A1**	**WNT4**	**WT1**
Nucleotide Alignment	**All Species**	**Full Alignment**	**Length (bp)**	1218	1578	831	1668	849	1071	1368
			Percent Identity	90.8%	78.4%	79.7%	82.3%	73.0%	86.6%	84.8%
		Gaps Removed	**Length (bp)**	1149	831	723	1263	483	975	741
			Percent Identity	92.1%	85.3%	82.2%	87.8%	78.5%	86.7%	88.9%
	Turtles	**Full Alignment**	**Length (bp)**	1191	1398	780	1491	798	1053	1251
			Percent Identity	98.2%	95.3%	93.8%	94.6%	84.7%	98.5%	93.7%
		Gaps Removed	**Length (bp)**	1161	1350	753	1446	798	975	1086
			Percent Identity	96.2%	92.5%	94.4%	92.1%	90.0%	93.3%	95.6%
Amino Acid Alignment	**All Species**	**Full Alignment**	**Length (aa)**	406	525	277	554	281	357	458
			Percent Identity	96.4%	82.3%	79.7%	88.9%	68.6%	94.4%	90.2%
		Gaps Removed	**Length (aa)**	387	294	246	422	161	325	249
			Percent Identity	97.5%	88.3%	82.3%	95.1%	77.0%	95.2%	95.4%
	Turtles	**Full Alignment**	**Length (aa)**	397	466	260	497	266	351	417
			Percent Identity	98.2%	95.3%	93.8%	94.6%	84.7%	98.5%	93.7%
		Gaps Removed	**Length (aa)**	387	450	251	483	266	325	362
			Percent Identity	98.7%	95.2%	94.9%	95.5%	84.7%	98.8%	96.9%

Table A.2: Steel-Dwass test results comparing third codon position substitution rates among major vertebrate clades. Statistically significant differences are denoted in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Squamata	Crocodilia	94.194	8.988	10.480	3.568E-13	1.226E-03	9.189E-04	1.544E-03
Mammalia	Crocodilia	79.749	7.877	10.124	3.568E-13	1.594E-03	1.308E-03	2.157E-03
Mammalia	Aves	25.093	7.095	3.537	3.711E-03	6.353E-04	1.438E-04	1.218E-03
Testudines	Crocodilia	24.566	12.863	1.910	3.119E-01	7.120E-05	-2.983E-05	1.718E-04
Squamata	Aves	10.009	8.691	1.152	7.789E-01	1.787E-04	-2.495E-04	6.146E-04
Squamata	Mammalia	-22.458	8.691	-2.584	7.331E-02	-4.633E-04	-9.724E-04	3.282E-05
Crocodilia	Aves	-68.114	7.877	-8.647	3.568E-13	-1.025E-03	-1.339E-03	-7.669E-04
Testudines	Aves	-129.752	13.501	-9.611	3.568E-13	-9.600E-04	-1.217E-03	-7.307E-04
Testudines	Mammalia	-155.335	13.502	-11.505	3.568E-13	-1.538E-03	-1.932E-03	-1.270E-03
Testudines	Squamata	-157.231	12.683	-12.397	3.568E-13	-1.155E-03	-1.411E-03	-9.110E-04

Table A.3: Steel-Dwass test results comparing third codon position substitution rates among turtle clades. Statistically significant differences are in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Trionychia	Emydidae	49.707	7.094	7.007	1.488E-11	3.631E-04	2.386E-04	5.193E-04
Pleurodira	Emydidae	29.760	6.557	4.539	3.349E-05	2.207E-04	9.858E-05	3.811E-04
Trionychia	Americhelydia	16.853	6.559	2.570	4.993E-02	2.023E-04	0.000E+00	3.900E-04
Trionychia	Pleurodira	13.778	6.559	2.101	1.529E-01	1.413E-04	-3.845E-05	3.360E-04
Pleurodira	Americhelydia	3.689	5.506	0.670	9.085E-01	5.110E-05	-1.514E-04	2.542E-04
Emydidae	Americhelydia	-23.076	6.556	-3.520	2.438E-03	-2.009E-04	-3.382E-04	-5.485E-05

Table A.4: Steel-Dwass test results comparing nucleotide substitution rates for all codon positions among major vertebrate clades. Statistically significant differences are in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Squamata	Crocodilia	102.716	8.988	11.428	3.568E-13	4.964E-04	4.061E-04	6.083E-04
Mammalia	Crocodilia	81.897	7.878	10.396	3.568E-13	6.509E-04	5.136E-04	9.071E-04
Testudines	Crocodilia	35.214	12.866	2.737	4.874E-02	4.229E-05	0.000E+00	8.389E-05
Mammalia	Aves	25.307	7.095	3.567	3.322E-03	2.693E-04	6.387E-05	5.017E-04
Squamata	Aves	15.889	8.691	1.828	3.572E-01	1.047E-04	-5.547E-05	2.672E-04
Squamata	Mammalia	-20.810	8.691	-2.394	1.168E-01	-1.553E-04	-3.616E-04	2.574E-05
Crocodilia	Aves	-68.800	7.878	-8.734	3.568E-13	-4.012E-04	-5.358E-04	-2.933E-04
Testudines	Aves	-118.063	13.502	-8.744	3.568E-13	-3.553E-04	-4.686E-04	-2.537E-04
Testudines	Mammalia	-151.444	13.503	-11.216	3.568E-13	-6.037E-04	-7.954E-04	-4.824E-04
Testudines	Squamata	-160.330	12.685	-12.639	3.568E-13	-4.498E-04	-5.395E-04	-3.685E-04

Table A.5: Steel-Dwass test results comparing amino acid substitution rates among major vertebrate clades. Statistically significant differences are in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Squamata	Crocodilia	66.100	8.923	7.408	1.637E-12	2.065E-04	9.340E-05	3.478E-04
Mammalia	Crocodilia	56.217	7.803	7.205	6.158E-12	2.636E-04	1.211E-04	3.927E-04
Testudines	Crocodilia	28.189	12.527	2.250	1.614E-01	0.000E+00	0.000E+00	5.173E-05
Mammalia	Aves	17.013	7.088	2.400	1.152E-01	9.500E-05	-1.282E-05	2.754E-04
Squamata	Aves	15.347	8.681	1.768	3.924E-01	5.847E-05	-3.348E-05	1.921E-04
Squamata	Mammalia	-8.508	8.689	-0.979	8.647E-01	-3.421E-05	-1.899E-04	7.513E-05
Crocodilia	Aves	-39.520	7.745	-5.103	3.323E-06	-1.115E-04	-2.207E-04	-4.590E-05
Testudines	Aves	-52.556	13.278	-3.958	7.202E-04	-7.560E-05	-1.620E-04	-1.039E-05
Testudines	Squamata	-88.833	12.550	-7.078	1.496E-11	-1.619E-04	-2.583E-04	-7.742E-05
Testudines	Mammalia	-89.474	13.332	-6.711	1.934E-10	-1.956E-04	-3.425E-04	-9.500E-05

Table A.6: Steel-Dwass test results comparing nucleotide substitution rates for all codon positions among turtle clades. Statistically significant differences are in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Trionychia	Emydidae	48.773	7.094	6.875	3.734E-11	1.619E-04	1.009E-04	2.581E-04
Pleurodira	Emydidae	31.076	6.558	4.739	1.279E-05	1.130E-04	5.199E-05	1.873E-04
Trionychia	Americhelydia	16.924	6.558	2.581	4.849E-02	1.008E-04	0.000E+00	2.039E-04
Trionychia	Pleurodira	11.822	6.559	1.802	2.721E-01	5.992E-05	-2.437E-05	1.690E-04
Pleurodira	Americhelydia	4.933	5.506	0.896	8.070E-01	3.510E-05	-6.689E-05	1.319E-04
Emydidae	Americhelydia	-19.911	6.554	-3.038	1.273E-02	-9.309E-05	-1.552E-04	-1.550E-05

Table A.7: Steel-Dwass test results comparing amino acid substitution rates among turtle clades. Statistically significant differences are in red.

Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Trionychia	Emydidae	29.733	6.945	4.281	1.094E-04	7.836E-05	1.893E-05	1.656E-04
Pleurodira	Emydidae	16.498	6.359	2.594	4.673E-02	4.400E-05	0.000E+00	9.851E-05
Trionychia	Americhelydia	10.809	6.484	1.667	3.413E-01	4.601E-05	-1.007E-05	1.380E-04
Trionychia	Pleurodira	9.351	6.522	1.434	4.781E-01	3.994E-05	-2.290E-05	1.380E-04
Pleurodira	Americhelydia	1.644	5.415	0.304	9.903E-01	0.000E+00	-8.330E-05	7.010E-05
Emydidae	Americhelydia	-12.124	6.253	-1.939	2.118E-01	-6.401E-06	-1.189E-04	0.000E+00

Table A.8: Z-score analysis identifying genes for each phylogenetic branch with significantly faster than average nucleotide substitution rates relative to other branches. Significant Z-scores (>1.644) are red. Number of fast genes for each branch indicated by number in parentheses after branch name.

Branch	AR	CIRBP	CTNNB1	CYP19A1	DMRT1	ESR1	ESR2	HSF2
Placental Root (12)	1.122	2.536	2.348	1.716	2.016	2.858	2.790	2.459
Iguania Root (12)	4.735	5.247	-0.505	2.736	-0.197	1.215	1.837	2.584
Reptile Root (10)	-0.301	-0.388	4.269	2.362	2.432	3.988	1.756	0.716
Archosaur Root (6)	2.545	-0.585	0.823	1.004	1.509	0.114	-0.073	1.763
Archelosaur Root (5)	1.801	0.111	0.671	1.941	1.804	1.382	2.434	1.362
Mus musculus (5)	0.449	1.683	2.669	2.567	1.408	1.686	1.088	0.932
Neoaves Root (3)	0.774	-0.615	0.755	0.225	2.375	-0.132	1.933	0.807
Taeniopygia guttata (2)	1.074	0.805	0.639	1.148	1.787	-0.213	1.771	1.183
Squamate Root (2)	0.135	0.253	0.667	1.167	1.067	1.144	1.303	2.901
Carettochelys insculpta (1)	-0.109	-0.265	-0.342	-0.245	-0.483	-0.468	-0.431	-0.207
Trionychia Root (1)	0.194	0.003	0.122	-0.176	-0.193	0.675	0.145	0.447

Branch	LHX9	NR5A1	RSPO1	SOX9	SRD5A1	WNT4	WT1
Placental Root (12)	4.441	3.612	1.581	1.759	1.377	2.565	2.101
Iguania Root (12)	2.768	2.627	4.961	2.406	4.190	3.123	1.685
Reptile Root (10)	-0.206	2.493	1.950	-0.448	2.932	2.817	3.740
Archosaur Root (6)	0.324	1.805	1.177	2.078	-0.392	1.896	2.594
Archelosaur Root (5)	0.383	-0.487	1.261	3.219	-0.864	0.901	0.069
Mus musculus (5)	0.655	0.941	0.885	0.450	1.902	0.566	0.711
Neoaves Root (3)	-0.127	0.834	0.184	-1.003	0.739	2.021	0.450
Taeniopygia guttata (2)	-0.363	-0.473	-0.109	1.159	0.542	0.265	1.480
Squamate Root (2)	2.583	0.373	-0.055	-0.102	0.500	-0.009	0.873
Carettochelys insculpta (1)	1.869	-0.320	-0.344	-0.363	-0.285	-0.331	-0.401
Trionychia Root (1)	-0.618	-0.058	0.278	2.808	-0.085	1.238	-0.581

Table A.9: Z-score analysis identifying proteins for each phylogenetic branch with significantly faster than average amino acid substitution rates relative to other branches. Significant Z-scores (>1.644) are red. Number of fast proteins for each branch indicated by number in parentheses after branch name.

Branch	AR	CIRBP	CTNNB1	CYP19A1	DMRT1	ESR1	ESR2	HSF2
Placental Root (7)	-0.466	0.319	1.110	1.122	1.393	1.872	2.857	3.509
Iguania Root (7)	-0.530	4.338	-0.460	1.250	0.577	4.686	2.368	-0.678
Reptile Root (6)	0.701	0.173	0.546	4.136	1.393	1.552	2.040	1.026
Archosaur Root (4)	6.182	-0.522	-0.460	2.114	0.155	0.748	0.502	0.100
Squamate Root (3)	0.003	0.911	-0.460	-0.009	1.294	0.427	1.665	2.485
Trionychia Root (3)	0.465	-0.522	-0.460	0.584	-0.497	1.894	1.557	1.800
Taeniopygia guttata (3)	0.096	0.048	-0.460	-0.259	2.760	-0.418	1.430	2.361
Mus musculus (2)	0.121	2.774	0.740	2.963	1.246	1.037	0.586	0.573
Neoaves Root (2)	-0.282	1.539	-0.460	-0.984	3.191	-0.558	1.801	0.983
Lacertoidea Root (1)	-0.254	-0.522	-0.460	0.030	0.531	1.346	1.535	0.079
Ophiophagus hannah (1)	-0.246	-0.522	5.508	0.219	-0.392	0.188	0.274	0.378
Archelosaur Root (1)	0.140	2.270	-0.460	0.700	0.890	1.280	-0.197	0.193
Testudines Root (1)	-0.091	1.686	-0.460	0.081	-0.514	-0.353	-0.583	1.023
Carettochelys insculpta (1)	1.310	-0.522	0.290	0.083	-0.282	-0.487	-0.282	0.117
Americhelydia Root (1)	0.987	-0.522	-0.460	-0.984	-0.809	0.200	-1.151	-0.970
Falco peregrinus (1)	-0.422	0.050	-0.460	-0.298	2.502	-0.353	-0.360	-0.363

Branch	LHX9	NR5A1	RSPO1	SOX9	SRD5A1	WNT4	WT1
Placental Root (7)	3.672	3.813	1.707	0.440	0.329	3.177	0.914
Iguania Root (7)	-0.450	-0.582	5.253	3.936	5.795	0.417	1.650
Reptile Root (6)	-0.450	4.621	0.099	4.027	1.142	4.376	1.846
Archosaur Root (4)	0.917	-0.378	1.283	-0.256	-0.656	3.231	3.324
Squamate Root (3)	3.629	0.070	-0.010	-0.324	0.080	-0.124	0.993
Trionychia Root (3)	-0.450	0.072	0.339	1.838	0.055	-0.486	1.365
Taeniopygia guttata (3)	-0.167	-0.334	-0.066	1.662	0.271	-0.194	-0.235
Mus musculus (2)	1.225	0.037	0.976	-0.259	0.624	-0.102	0.852
Neoaves Root (2)	-0.450	-0.336	0.241	-0.587	1.100	-0.486	0.714
Lacertoidea Root (1)	-0.371	2.125	-0.083	-0.173	0.096	-0.486	0.235
Ophiophagus hannah (1)	0.066	0.902	-	-0.374	0.547	1.016	0.903
Archelosaur Root (1)	-0.450	0.129	1.371	0.387	-0.656	-0.147	-0.779
Testudines Root (1)	-0.450	-0.309	-0.293	-0.479	-0.070	-0.486	-0.779
Carettochelys insculpta (1)	3.161	-0.247	-0.313	-0.230	-0.131	0.130	-0.034
Americhelydia Root (1)	-0.450	-0.582	0.188	-0.587	1.441	-0.486	3.117
Falco peregrinus (1)	-0.450	-0.055	-0.461	-0.215	-0.616	-0.486	-0.779

Table A.10: Z-score analysis identifying genes for each turtle branch with significantly faster than average nucleotide substitution rates relative to other turtle branches. Significant Z-scores (>1.644) are red. Number of fast genes for each branch indicated by number in parentheses after branch name

Branch	AR	CIRBP	CTNNB1	CYP19A1	DMRT1	ESR1	ESR2	HSF2
Trionychia Root (14)	3.040	2.637	2.115	2.033	1.727	3.739	3.255	2.961
Podcnmeis expansa (2)	1.887	1.221	1.123	1.332	1.555	0.848	0.902	1.420
Americhelydia Root (2)	0.226	-1.108	0.096	-1.593	-0.758	-0.303	-1.286	-0.153
Deirochelyinae Root (2)	-0.416	-1.108	2.180	0.790	1.769	0.327	1.107	-0.713
Carettochelys insculpta (1)	1.106	0.849	0.846	1.543	1.587	0.133	0.695	0.960
Trionychidae Root (1)	0.210	1.005	0.163	1.206	0.325	0.235	-0.174	0.492

Branch	LHX9	NR5A1	RSPO1	SOX9	SRD5A1	WNT4	WT1
Trionychia Root (14)	-0.448	2.790	2.993	3.620	2.012	3.760	2.681
Podcnmeis expansa (2)	0.763	0.247	0.348	0.294	0.424	0.173	2.008
Americhelydia Root (2)	-0.448	-1.043	1.995	-0.899	2.193	-0.092	-1.057
Deirochelyinae Root (2)	-0.015	-0.050	-0.679	0.852	0.084	-0.490	0.795
Carettochelys insculpta (1)	3.955	0.445	0.644	-0.082	1.197	0.281	0.215
Trionychidae Root (1)	0.002	1.982	-0.263	0.496	0.322	0.751	0.181

Table A.11: Z-score analysis identifying proteins for each turtle branch with significantly faster than average amino acid substitution rates relative to other turtle branches. Significant Z-scores (>1.644) are red. Number of fast proteins for each branch indicated by number in parentheses after branch name

Branch	AR	CIRBP	CTNNB1	CYP19A1	DMRT1	ESR1	ESR2	HSF2
Trionychia Root (11)	2.293	-0.243	-0.483	2.269	1.746	3.929	3.689	2.993
Podcnmeis expansa (3)	1.357	4.123	0.854	1.243	0.258	0.321	0.559	1.723
Carettochelys insculpta (3)	1.981	-0.243	0.736	0.872	0.829	-0.244	0.175	0.827
Americhelydia Root (2)	1.455	-0.243	-0.483	-1.236	-0.899	0.456	-1.051	-0.967
Deirochelyinae Root (2)	-0.789	-0.243	-0.483	1.381	2.969	0.074	0.221	-0.967
Cryptodiran Root (1)	-0.789	-0.243	-0.483	0.581	-0.196	-0.571	0.387	0.320
Pelodiscus sinensis (1)	-0.789	-0.243	2.758	1.203	-0.626	-0.406	-0.174	-0.089
Americhelydia/Emydiade Root (1)	0.054	-0.243	2.420	-0.233	-0.420	-0.523	-0.625	-0.967
Staurotypus triporcatus (1)	-0.389	-0.243	-0.483	-0.285	0.854	0.144	-0.134	0.593

Branch	LHX9	NR5A1	RSPO1	SOX9	SRD5A1	WNT4	WT1
Trionychia Root (11)	-0.371	2.161	3.809	3.256	1.648	-0.698	1.730
Podcnmeis expansa (3)	0.648	-0.251	0.016	-0.166	0.292	-0.698	2.472
Carettochelys insculpta (3)	3.966	0.994	0.812	-0.140	0.688	2.366	0.412
Americhelydia Root (2)	-0.371	-1.212	-0.660	-0.791	2.881	-0.698	2.137
Deirochelyinae Root (2)	-0.371	0.244	-0.660	1.894	-0.593	-0.698	-0.069
Cryptodiran Root (1)	-0.371	-1.212	-0.660	-0.336	-0.548	2.061	-0.735
Pelodiscus sinensis (1)	-0.371	-0.238	-0.404	-0.791	0.315	0.454	-0.426
Americhelydia/Emydiade Root (1)	-0.371	-0.296	-0.017	-0.572	-0.762	0.354	-0.171
Staurotypus triporcatus (1)	-0.021	2.156	0.265	0.381	0.632	0.617	-0.019

Table A.12: Steel-Dwass test results comparing nucleotide substitution rates among gene classes within major vertebrate clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Birds	WNT Signaling	Transcription Factor	0.000	3.818	0.000	1.000	-4.56E-06	-4.94E-04	3.01E-04
Birds	Transcription Factor	Temperature	-0.630	3.834	-0.164	0.998	-4.25E-05	-4.75E-04	5.14E-04
Birds	WNT Signaling	Temperature	-2.417	3.005	-0.804	0.852	-1.31E-04	-5.25E-04	3.56E-04
Birds	Temperature	Hormone	-2.450	3.834	-0.639	0.919	-8.96E-05	-5.68E-04	2.80E-04
Birds	Transcription Factor	Hormone	-5.520	4.123	-1.339	0.538	-1.85E-04	-4.80E-04	2.16E-04
Birds	WNT Signaling	Hormone	-7.253	3.818	-1.900	0.228	-1.89E-04	-4.78E-04	6.92E-05
Crocodilians	WNT Signaling	Transcription Factor	8.229	4.502	1.828	0.260	5.92E-05	-2.33E-05	1.52E-04
Crocodilians	WNT Signaling	Hormone	2.095	4.502	0.465	0.967	1.11E-05	-7.65E-05	1.08E-04
Crocodilians	Temperature	Hormone	2.000	4.518	0.443	0.971	1.44E-05	-8.64E-05	1.10E-04
Crocodilians	WNT Signaling	Temperature	-0.119	3.535	-0.034	1.000	-2.17E-06	-1.15E-04	1.38E-04
Crocodilians	Transcription Factor	Temperature	-8.550	4.518	-1.892	0.231	-5.85E-05	-1.50E-04	3.48E-05
Crocodilians	Transcription Factor	Hormone	-8.657	4.865	-1.780	0.283	-4.47E-05	-1.11E-04	1.68E-05
Mammals	WNT Signaling	Transcription Factor	1.280	3.818	0.335	0.987	3.88E-05	-4.03E-04	6.69E-04
Mammals	WNT Signaling	Temperature	0.750	3.005	0.250	0.995	4.24E-05	-7.11E-04	8.16E-04
Mammals	Transcription Factor	Temperature	-0.350	3.834	-0.091	1.000	-6.44E-06	-6.87E-04	5.85E-04

Table A.12 continued: Steel-Dwass test results comparing nucleotide substitution rates among gene classes within major vertebrate clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Mammals	WNT Signaling	Hormone	-2.987	3.818	-0.782	0.863	-1.54E-04	-7.24E-04	4.35E-04
Mammals	Temperature	Hormone	-4.970	3.834	-1.296	0.565	-2.18E-04	-8.83E-04	3.57E-04
Mammals	Transcription Factor	Hormone	-6.880	4.123	-1.669	0.340	-2.40E-04	-7.11E-04	1.56E-04
Squamates	WNT Signaling	Transcription Factor	3.484	5.076	0.686	0.902	6.93E-05	-2.50E-04	2.85E-04
Squamates	WNT Signaling	Temperature	2.246	3.881	0.579	0.939	6.37E-05	-3.31E-04	3.63E-04
Squamates	Transcription Factor	Temperature	-0.506	5.112	-0.099	1.000	-1.27E-05	-2.84E-04	2.86E-04
Squamates	WNT Signaling	Hormone	-9.707	5.076	-1.912	0.223	-2.04E-04	-5.23E-04	7.48E-05
Squamates	Temperature	Hormone	-10.383	5.112	-2.031	0.177	-2.53E-04	-6.03E-04	7.88E-05
Squamates	Transcription Factor	Hormone	-15.956	5.508	-2.897	0.020	-2.59E-04	-4.95E-04	-2.82E-05
Turtles	WNT Signaling	Transcription Factor	8.421	7.370	1.143	0.663	2.87E-05	-3.91E-05	9.87E-05
Turtles	WNT Signaling	Temperature	5.789	5.772	1.003	0.748	3.09E-05	-5.30E-05	1.22E-04
Turtles	Transcription Factor	Temperature	0.811	7.388	0.110	1.000	0.00E+00	-7.06E-05	7.62E-05
Turtles	WNT Signaling	Hormone	-5.333	7.375	-0.723	0.888	-1.56E-05	-8.37E-05	5.00E-05
Turtles	Temperature	Hormone	-13.263	7.396	-1.793	0.276	-4.87E-05	-1.21E-04	2.50E-05
Turtles	Transcription Factor	Hormone	-17.158	7.975	-2.151	0.137	-4.84E-05	-1.02E-04	7.29E-06

Table A.13: Steel-Dwass test results comparing amino acid substitution rates among gene classes within major vertebrate clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Birds	Transcription Factor	Temperature	-0.210	3.829	-0.055	1.000	-2.70E-06	-2.21E-04	4.84E-04
Birds	Temperature	Hormone	-4.410	3.834	-1.150	0.658	-9.74E-05	-3.26E-04	1.67E-04
Birds	WNT Signaling	Temperature	-4.583	2.956	-1.551	0.407	-7.56E-05	-2.72E-04	1.81E-04
Birds	Transcription Factor	Hormone	-6.200	4.120	-1.505	0.434	-8.38E-05	-2.80E-04	1.09E-04
Birds	WNT Signaling	Transcription Factor	-6.293	3.752	-1.677	0.336	-6.97E-05	-2.91E-04	7.22E-05
Birds	WNT Signaling	Hormone	-9.600	3.797	-2.529	0.056	-1.73E-04	-3.96E-04	0.00E+00
Crocodilians	WNT Signaling	Transcription Factor	2.133	4.103	0.520	0.954	0.00E+00	-2.51E-05	5.94E-05
Crocodilians	WNT Signaling	Temperature	0.000	3.327	0.000	1.000	0.00E+00	-4.70E-05	7.15E-05
Crocodilians	Transcription Factor	Temperature	-1.750	4.168	-0.420	0.975	0.00E+00	-3.99E-05	4.42E-05
Crocodilians	WNT Signaling	Hormone	-9.257	4.427	-2.091	0.156	-4.31E-05	-1.41E-04	0.00E+00
Crocodilians	Temperature	Hormone	-10.050	4.476	-2.245	0.111	-6.43E-05	-1.67E-04	0.00E+00
Crocodilians	Transcription Factor	Hormone	-16.171	4.723	-3.424	0.003	-6.20E-05	-1.36E-04	0.00E+00
Mammals	Transcription Factor	Temperature	-1.610	3.833	-0.420	0.975	-3.91E-05	-3.55E-04	3.22E-04
Mammals	WNT Signaling	Temperature	-1.750	3.005	-0.582	0.937	-6.68E-05	-4.28E-04	7.78E-04
Mammals	WNT Signaling	Transcription Factor	-1.760	3.816	-0.461	0.967	-3.84E-05	-3.14E-04	4.61E-04

Table A.13 continued: Steel-Dwass test results comparing amino acid substitution rates among gene classes within major vertebrate clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Mammals	Temperature	Hormone	-6.650	3.834	-1.734	0.306	-3.34E-04	-9.91E-04	1.33E-04
Mammals	WNT Signaling	Hormone	-8.107	3.818	-2.123	0.146	-2.64E-04	-9.35E-04	1.35E-04
Mammals	Transcription Factor	Hormone	-10.000	4.123	-2.426	0.072	-2.91E-04	-8.75E-04	2.82E-05
Squamates	Transcription Factor	Temperature	2.683	5.107	0.525	0.953	2.27E-05	-1.69E-04	2.42E-04
Squamates	WNT Signaling	Temperature	-1.194	3.864	-0.309	0.990	-5.19E-06	-3.14E-04	2.23E-04
Squamates	WNT Signaling	Transcription Factor	-6.284	5.071	-1.239	0.602	-5.48E-05	-2.40E-04	1.13E-04
Squamates	Temperature	Hormone	-14.583	5.111	-2.853	0.022	-2.79E-04	-5.50E-04	-2.20E-05
Squamates	WNT Signaling	Hormone	-15.524	5.075	-3.059	0.012	-2.96E-04	-5.17E-04	-5.78E-05
Squamates	Transcription Factor	Hormone	-18.222	5.507	-3.309	0.005	-2.35E-04	-4.60E-04	-5.07E-05
Turtles	Transcription Factor	Temperature	8.695	7.084	1.227	0.609	0.00E+00	0.00E+00	5.10E-05
Turtles	WNT Signaling	Temperature	0.417	5.308	0.078	1.000	0.00E+00	-8.62E-06	9.80E-06
Turtles	WNT Signaling	Transcription Factor	-9.011	7.053	-1.278	0.577	0.00E+00	-4.20E-05	0.00E+00
Turtles	Temperature	Hormone	-33.397	7.329	-4.557	0.000	-1.36E-04	-2.21E-04	-4.55E-05
Turtles	WNT Signaling	Hormone	-38.372	7.286	-5.267	0.000	-1.35E-04	-2.04E-04	-6.18E-05
Turtles	Transcription Factor	Hormone	-42.400	7.902	-5.366	0.000	-1.20E-04	-1.87E-04	-5.16E-05

Table A.14: Steel-Dwass test results comparing nucleotide substitution rates among gene classes within major turtle clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Americhelydia	WNT Signaling	Temperature	2.250	2.668	0.843	0.834	6.34E-05	-1.88E-04	4.68E-04
Americhelydia	WNT Signaling	Transcription Factor	2.083	2.884	0.722	0.888	5.43E-05	-1.99E-04	3.17E-04
Americhelydia	Transcription Factor	Temperature	-0.125	2.664	-0.047	1.000	-3.32E-06	-2.39E-04	3.67E-04
Americhelydia	WNT Signaling	Hormone	-1.125	3.072	-0.366	0.983	-1.88E-05	-2.41E-04	1.97E-04
Americhelydia	Temperature	Hormone	-3.383	2.993	-1.130	0.671	-8.79E-05	-4.39E-04	1.03E-04
Americhelydia	Transcription Factor	Hormone	-3.825	3.069	-1.246	0.597	-8.71E-05	-2.71E-04	1.08E-04
Emydidae	Transcription Factor	Temperature	4.125	3.396	1.215	0.617	1.93E-05	-5.99E-05	1.18E-04
Emydidae	WNT Signaling	Temperature	2.175	3.405	0.639	0.919	6.75E-06	-5.33E-05	8.35E-05
Emydidae	WNT Signaling	Transcription Factor	-2.050	3.691	-0.555	0.945	-1.25E-05	-8.37E-05	5.82E-05
Emydidae	Transcription Factor	Hormone	-5.805	3.939	-1.474	0.453	-3.74E-05	-1.12E-04	3.27E-05
Emydidae	WNT Signaling	Hormone	-9.045	3.940	-2.296	0.099	-5.14E-05	-1.14E-04	5.00E-06
Emydidae	Temperature	Hormone	-9.940	3.833	-2.593	0.047	-5.79E-05	-1.76E-04	0.00E+00

Table A.14 continued: Steel-Dwass test results comparing nucleotide substitution rates among gene classes within major turtle clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Pleurodira	Transcription Factor	Temperature	1.375	2.669	0.515	0.956	5.05E-05	-2.53E-04	3.15E-04
Pleurodira	WNT Signaling	Temperature	1.375	2.669	0.515	0.956	1.88E-05	-2.40E-04	2.54E-04
Pleurodira	Transcription Factor	Hormone	0.525	3.074	0.171	0.998	1.58E-05	-1.85E-04	2.28E-04
Pleurodira	WNT Signaling	Hormone	-0.225	3.074	-0.073	1.000	-4.73E-06	-2.33E-04	1.72E-04
Pleurodira	WNT Signaling	Transcription Factor	-1.083	2.887	-0.375	0.982	-3.17E-05	-2.49E-04	1.76E-04
Pleurodira	Temperature	Hormone	-1.983	2.997	-0.662	0.911	-2.89E-05	-3.54E-04	2.22E-04
Trionychia	WNT Signaling	Transcription Factor	-0.550	3.697	-0.149	0.999	-3.53E-06	-2.94E-04	1.50E-04
Trionychia	WNT Signaling	Temperature	-2.325	3.410	-0.682	0.904	-6.55E-05	-1.96E-04	2.86E-04
Trionychia	Temperature	Hormone	-3.150	3.834	-0.822	0.844	-7.13E-05	-3.25E-04	1.31E-04
Trionychia	Transcription Factor	Hormone	-4.005	3.940	-1.016	0.740	-7.05E-05	-2.79E-04	1.98E-04
Trionychia	WNT Signaling	Hormone	-6.075	3.940	-1.542	0.412	-8.32E-05	-3.01E-04	6.88E-05
Trionychia	Transcription Factor	Temperature	-0.225	3.410	-0.066	1.000	-2.40E-05	-2.21E-04	4.10E-04

Table A.15: Steel-Dwass test results comparing amino acid substitution rates among gene classes within major turtle clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Americheldia	Transcription Factor	Temperature	2.375	2.499	0.951	0.777	1.14E-05	-1.75E-04	3.28E-04
Americheldia	WNT Signaling	Temperature	0.375	2.347	0.160	0.999	0.00E+00	-2.29E-04	2.59E-04
Americheldia	WNT Signaling	Transcription Factor	-2.000	2.701	-0.740	0.881	0.00E+00	-2.42E-04	9.20E-05
Americheldia	Transcription Factor	Hormone	-6.525	3.048	-2.141	0.140	-1.95E-04	-6.15E-04	2.80E-05
Americheldia	Temperature	Hormone	-6.883	2.963	-2.323	0.093	-2.19E-04	-1.16E-03	1.85E-05
Americheldia	WNT Signaling	Hormone	-8.175	3.017	-2.709	0.034	-2.02E-04	-5.23E-04	0.00E+00
Emydidae	Transcription Factor	Temperature	3.375	3.190	1.058	0.715	0.00E+00	-4.16E-05	1.23E-04
Emydidae	WNT Signaling	Temperature	-2.025	2.764	-0.733	0.884	0.00E+00	-6.25E-05	1.05E-05
Emydidae	Transcription Factor	Hormone	-5.535	3.867	-1.431	0.480	-4.37E-05	-1.37E-04	3.79E-05
Emydidae	WNT Signaling	Transcription Factor	-6.700	3.274	-2.047	0.171	-6.50E-07	-1.17E-04	0.00E+00
Emydidae	Temperature	Hormone	-9.030	3.756	-2.404	0.076	-6.50E-05	-2.55E-04	0.00E+00
Emydidae	WNT Signaling	Hormone	-13.005	3.735	-3.482	0.003	-8.74E-05	-1.50E-04	0.00E+00

Table A.15 continued: Steel-Dwass test results comparing amino acid substitution rates among gene classes within major turtle clades. Statistically significant differences are in red.

Clade	Group A	Group B	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL
Pleurodira	Transcription Factor	Temperature	0.875	2.655	0.330	0.988	1.48E-05	-3.68E-04	1.75E-04
Pleurodira	WNT Signaling	Temperature	-0.875	2.591	-0.338	0.987	0.00E+00	-4.37E-04	2.02E-04
Pleurodira	WNT Signaling	Transcription Factor	-2.583	2.851	-0.906	0.802	-4.40E-05	-1.56E-04	1.09E-04
Pleurodira	Temperature	Hormone	-2.917	2.987	-0.976	0.763	-5.92E-05	-4.33E-04	2.18E-04
Pleurodira	Transcription Factor	Hormone	-3.375	3.069	-1.100	0.690	-7.26E-05	-3.63E-04	9.38E-05
Pleurodira	WNT Signaling	Hormone	-5.325	3.048	-1.747	0.299	-8.86E-05	-4.02E-04	6.95E-05
Trionychia	Transcription Factor	Temperature	3.750	3.378	1.110	0.683	5.01E-05	-1.29E-04	2.80E-04
Trionychia	WNT Signaling	Temperature	0.375	3.325	0.113	0.999	0.00E+00	-1.66E-04	1.38E-04
Trionychia	WNT Signaling	Transcription Factor	-5.250	3.676	-1.428	0.482	-5.04E-05	-2.55E-04	4.44E-05
Trionychia	Transcription Factor	Hormone	-9.495	3.938	-2.411	0.075	-1.96E-04	-6.00E-04	3.20E-06
Trionychia	Temperature	Hormone	-10.150	3.819	-2.658	0.039	-2.59E-04	-8.88E-04	0.00E+00
Trionychia	WNT Signaling	Hormone	-13.635	3.929	-3.470	0.003	-2.46E-04	-6.31E-04	-7.34E-05

Table A.16: Z-score analysis identifying genes within each phylogenetic clade with a significantly faster than average nucleotide or amino acid substitution rates relative to other genes. Significant Z-scores (>1.644) are red.

	Nucleotide Data					Amino Acid Data				
	Birds	**Crocodilians**	**Mammals**	**Squamates**	**Turtles**	**Birds**	**Crocodilians**	**Mammals**	**Squamates**	**Turtles**
AR	0.047	-0.323	-0.459	-0.126	-0.133	-0.405	-0.005	-0.697	-0.409	-0.186
CIRBP	-0.101	-0.100	0.127	0.177	-0.306	-0.494	-0.384	-0.616	-0.394	-0.627
CTNNB1	-0.559	-0.285	-0.560	-0.615	-0.372	-0.676	-0.392	-0.867	-0.502	-0.698
CYP19A1	0.040	-0.170	0.191	0.159	0.144	-0.150	0.011	0.609	0.067	0.947
DMRT1	1.550	-0.091	0.098	-0.303	-0.237	2.407	-0.132	0.487	0.061	-0.167
ESR1	-0.483	-0.262	0.055	-0.004	0.039	-0.362	-0.054	0.101	0.222	0.189
ESR2	0.856	-0.081	0.276	0.108	0.084	0.395	-0.008	0.212	0.048	0.437
HSF2	-0.126	0.090	-0.356	-0.188	-0.051	0.165	-0.143	0.036	-0.080	0.198
LHX9	-1.170	-0.599	-0.494	-0.341	-0.420	-0.646	-0.382	-0.672	-0.423	-0.563
NR5A1	-0.036	0.173	0.361	0.173	0.433	-0.263	-0.247	0.282	-0.005	-0.023
RSPO1	-0.339	0.838	0.363	0.957	-0.021	0.235	0.723	1.016	1.111	0.349
SOX9	-0.199	-0.269	-0.375	-0.281	0.537	-0.222	-0.356	-0.555	-0.231	-0.079
SRD5A1	0.608	1.216	0.961	0.796	0.163	1.116	1.624	1.730	1.440	1.179
WNT4	0.346	0.365	0.196	0.076	0.572	-0.618	-0.322	-0.461	-0.380	-0.554
WT1	-0.432	-0.502	-0.385	-0.373	-0.431	-0.482	0.067	-0.608	-0.278	-0.402

Table A.17: Z-score analysis identifying genes within each turtle clade with a significantly faster than average nucleotide or amino acid substitution rates relative to other genes. Significant Z-scores (>1.644) are red.

	Nucleotide Data				Amino Acid Data			
	Americhelydia	**Emydidae**	**Pleurodira**	**Trionychia**	**Americhelydia**	**Emydidae**	**Pleurodira**	**Trionychia**
AR	-0.091	0.351	-0.050	-0.087	-0.035	-0.194	-0.055	-0.137
CIRBP	-0.563	-0.613	-0.465	-0.357	-0.493	-0.726	-0.655	-0.721
CTNNB1	-0.381	-0.674	-0.681	-0.530	-0.493	-0.726	-0.779	-0.706
CYP19A1	-0.454	0.551	0.120	-0.096	-0.183	0.435	0.723	0.519
DMRT1	-0.294	-0.199	0.037	-0.380	-0.171	-0.448	-0.028	-0.261
ESR1	-0.050	0.138	0.092	0.048	0.179	-0.007	0.013	0.440
ESR2	-0.317	0.348	-0.130	-0.189	-0.122	1.276	0.240	0.225
HSF2	-0.185	0.170	0.061	-0.110	-0.225	0.267	0.419	-0.015
LHX9	-0.869	-1.376	-0.455	-0.084	-0.473	-0.726	-0.667	-0.525
NR5A1	0.187	-0.065	0.145	0.269	-0.118	0.200	-0.318	-0.123
RSPO1	0.881	-0.158	0.001	-0.104	-0.256	0.074	0.121	0.174
SOX9	0.201	0.716	0.683	0.650	-0.210	0.489	0.027	0.059
SRD5A1	2.177	1.229	0.279	0.515	3.095	1.250	1.445	2.013
WNT4	0.362	-0.067	0.572	0.830	-0.448	-0.726	-0.704	-0.577
WT1	-0.604	-0.351	-0.210	-0.375	-0.048	-0.437	0.219	-0.366

APPENDIX B: SUPPLEMENTAL MATERIAL FOR CHAPTER 3

B.1 Copy Number Quantification by Real-Time qPCR

Real-time qPCR is a commonly used method to quantify the initial template amount present in a PCR reaction (Morrison *et al.* 1998) and has been used for sex diagnosis [e.g.(Phillips and Edmands 2012, Alasaad *et al.* 2013, Ballester *et al.* 2013)]. C_T values are used for relative quantification among samples (Heid *et al.* 1996). Using samples that contain known amounts of DNA the efficiency of the qPCR reactions is calculated as

$$Eff = 10^{-(1/\text{slope})} \qquad \textbf{(1)}$$

An endogenous control (EC) such as a known single copy gene is used to normalize the copy number of the gene of interest (GOI) used for sex diagnosis. Different methods exist to quantify the normalized copy number using the qPCR data and the following three normalization approaches were compared in this study.

(1) Relative Standard Curve Quantification (Bustin 2000): Here, the C_T of each unknown sample is quantified by linear regression using the standard curve, and the ratio of the gene of interest to the EC is obtained directly by dividing the initial template quantities.

$$(GOI/EC) \qquad \textbf{(2)}$$

(2) Pfaffl Calibrator Method (Pfaffl 2001): Here, the ratio of GOI to EC is calculated as:

$$\text{Ratio } \left(\frac{GOI}{EC}\right) = \frac{(Eff_{EC})^{C_T\ EC\ Sample}}{(Eff_{GOI})^{C_T\ GOI\ Sample}} \div \frac{(Eff_{EC})^{C_T\ EC\ Calibrator}}{(Eff_{GOI})^{C_T\ GOI\ Calibrator}} \qquad \textbf{(3)}$$

where *Eff* is the gene- and plate-specific qPCR efficiency calculated from the standard curves (equation 1 above), and the calibrator is a standard sample diluted to the same concentration of the unknown DNA samples and which is amplified in all plates . This

provides a second form of control and standardization for plate-to-plate variation of each gene.

(3) Comparative C_T Method (Livak and Schmittgen 2001): This method is only applicable if the qPCR reaction efficiencies for the gene of interest and endogenous control are both around 100% (Eff~2) and comparable between genes. In this case the copy number ratio is calculated as:

$$Ratio\ (\frac{GOI}{EC}) = 2^{-\Delta C_T} = 2^{(C_T\ EC - C_T\ GOI)} \qquad \textbf{(4)}$$

Results from alternative normalization methods (equations 2-4), and standard types.

Our results were robust to using the alternative methods for normalization of 18S copy number described by equations 2-3 (where GOI = 18S and EC = GAPDH, using the 1:80 dilution standard as calibrator), using the same samples run with the male-only and mixed-sex standard curves (Figure B.1).

If gDNA is used to create the qPCR standard curve, the comparative C_T method ($2^{-\Delta CT}$) is the simplest method and perhaps preferable to alternative methods of normalization for *A. spinifera*, as once the qPCR reaction is optimized it requires no pre-knowledge of the sex of any individual. However, samples of known sex would still be beneficial as benchmarks for validation. Additionally, because standard curves permit assessing that qPCR reaction conditions are indeed similar and optimal between genes, they should be included in the qPCR even when using the comparative C_T method. Thus, when the qPCR efficiencies are very similar for both primer sets, the comparative C_T method is the easiest and least laborious method to implement. However, if the qPCR reaction efficiencies vary between primer sets or plates, the Pfaffl method and the comparative standard curve method provide good alternatives (equations 2,3).

Using standard curve or calibrator-based quantification methods still resulted in non-overlapping 18S/GAPDH ratios between males and females, but the resulting absolute values are specific to the 18S content of that specific set of standard curves and calibrators and as such are not directly comparable between research studies. Instead, these methods can be implemented by examining the bimodal distribution of dataset-specific ratios to assign individuals as male or female.

If using gDNA as a standard curve or calibrator these methods provide greater separation of male and female groups when a male-specific standard curve or calibrator is used, as using a mixed sex standard curve led to a compression of the bimodal distribution as seen in Figure B.1.

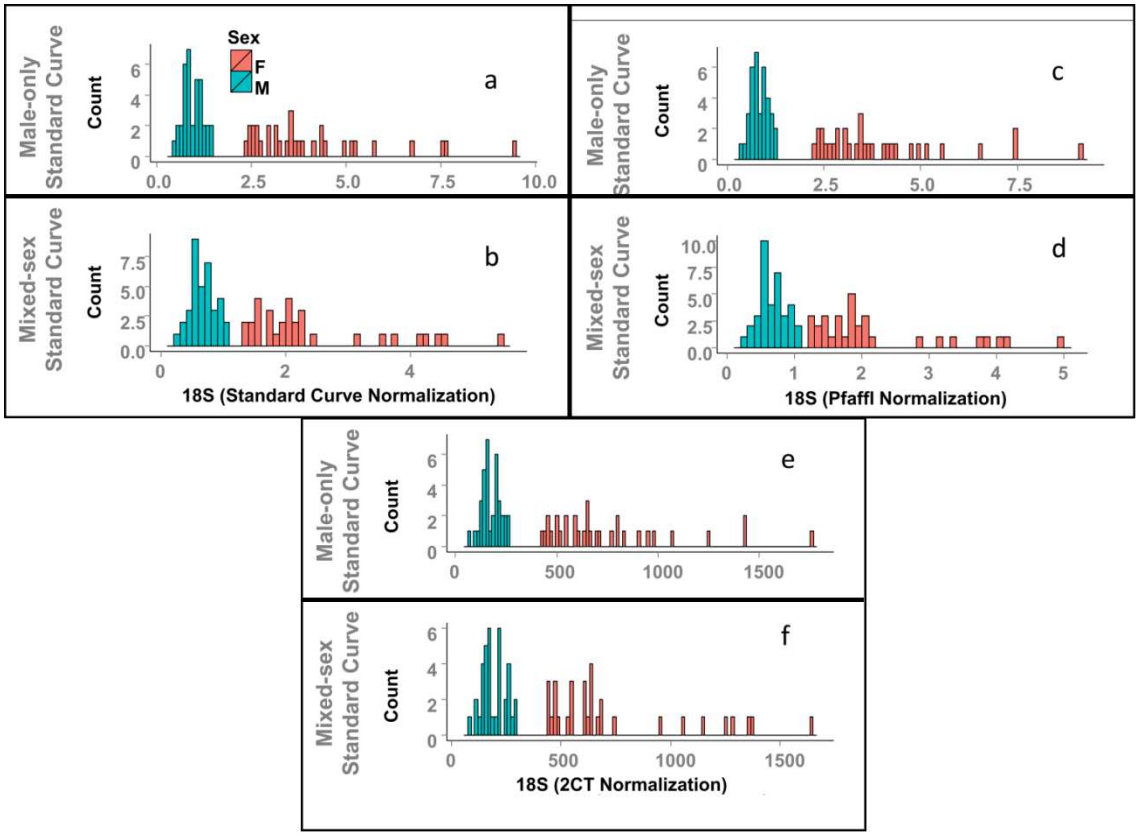

Figure B.1: Histograms of 18S/GAPDH ratios using different data normalization strategies. (A) Standard Curve Method; (B) Pfaffl Method; (C) $2C_T$ Method

B.2 Analytical Flow Chart

The goal of any sexing technique is to assign individuals to groups (males and females). Using a single continuous trait, the first step in this process is to visualize a histogram of the data which should be bimodal with respect to sex. The choice of which continuous variable to use must be based on some empirical observation that guides the researcher to hypothesize that the trait might be sexually dimorphic. A test is then carried out to validate the sexual dimorphism of the trait in question and its efficacy for accurate sex-typing of individuals as described in the text.

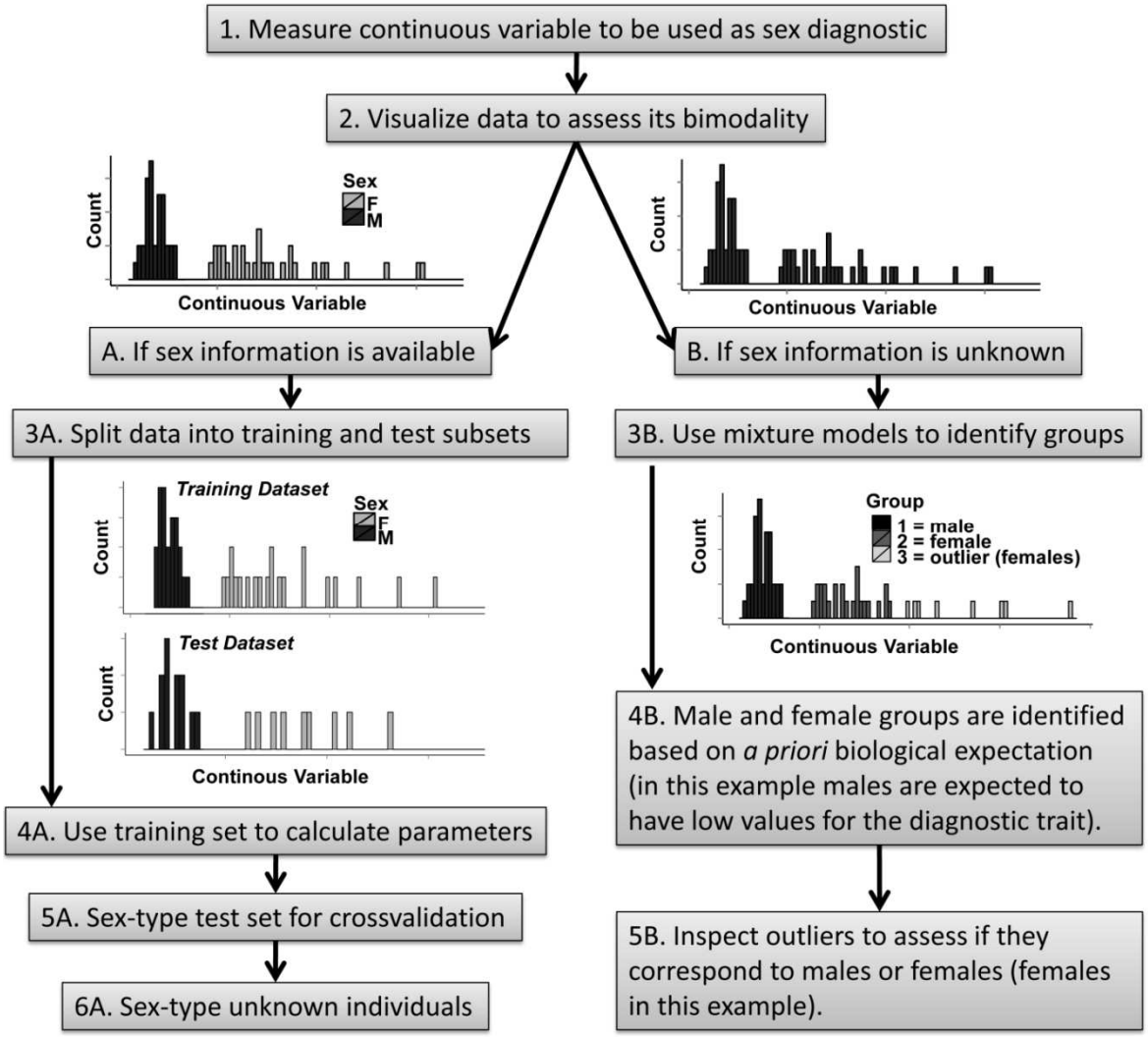

Figure B.2: Flow chart describing data analysis of continuous variables under different scenarios

B.3 R Code

This is an example of R Code using *Apalone spinifera* 18S copy number quantified by qPCR using GAPDH as normalizer, a male-only DNA standard curve, and the standard curve normalization method of 18S quantification. An example dataset is provided in Appendix B.4

```
#########################################################
###INITIAL VISUALIZATION OF DATA TO INSPECT BIMODALITY###
#########################################################
rm(list=ls())
library(ggplot2)

###READ DATA###
mydata<-read.csv(file="MyDataFile.csv",header=T)
sex<-mydata$Gsex

###Adjust binwidth as needed (large bins may mask bimodality)
myBinwidth<-0.1

###Single color histogram if individual sex is unknown
qplot(mydata$SCratio,binwidth=myBinwidth,xlab="18S/GAPDH Ratio",ylab="Count") +
geom_histogram(binwidth=myBinwidth,color="black")+theme(axis.line = element_line(colour = "black"),panel.grid.major =
element_blank(),panel.grid.minor = element_blank(),panel.background = element_blank())

###Histogram by sex if individual sex is known
qplot(mydata$SCratio,binwidth=myBinwidth,fill=sex,xlab="18S/GAPDH Ratio",ylab="Count") +
geom_histogram(binwidth=myBinwidth,color="black")+theme(axis.line = element_line(colour = "black"),panel.grid.major =
element_blank(),panel.grid.minor = element_blank(),panel.background = element_blank())

#############################################################################
###  UNIVARIATE DISCRIMINANT ANALYSIS WHEN SEX OF SOME INDIVIDUALS IS KNOWN  ##
#############################################################################

rm(list=ls())
library(mclust)

###READ DATA###
mydata<-read.csv(file="MyDataFile.csv",header=T)

###DIVIDE DATASET INTO TRAINING SET AND TESTING SET FOR CROSSVALIDATION###
trainData <- mydata[1:46,]   #MAKE TRAINING SET (ROW RANGE OF INDIVIDUALS WITH KNOWN SEX)
trainClass <- trainData$Gsex  #READ TRUE SEX INFORMATION FOR TRAINING SET
testData<- mydata[-(1:46),]  #MAKE TEST SET (THE REST OF THE DATA THAT IS NOT TRAINING SET)
testClass <- testData$Gsex    #READ TRUE SEX INFORMATION FOR TESTING SET FOR CROSSVALIDATION

###DISCRIMINANT ANALYSIS###
###CHECK MODEL TO USE ("E" MODEL = EQUAL VARIANCE; "V" MODEL = DIFFERENT VARIANCE##
cv1EMtrain(trainData$SCratio,labels=trainClass)          #COMPARE UNIVARIATE MODELS BY LEAVE-ONE-OUT
CROSSVALIDATION
bicEMtrain(trainData$SCratio,labels=trainClass) #COMPARE UNIVARIATE MODELS BY BIC
###Run Discriminant Analysis and Cross-validation test
modV <- MclustDA(trainData$SCratio, trainClass, modelType = "EDDA", modelName = "V") #RUN DA WITH BEST MODEL
DASCmale <- summary(modV, newdata = testData$SCratio, newclass = testClass)
DASCmale

###GRAPH DISCRIMINANT ANALYSIS##
par(mfrow = c(2,3), mar = c(4,4,2,1))
plot(modV)
plot(modV, what = "classification")
plot(modV, what = "error")
```

```
plot(modV, what = "train&test", newdata = testData$SCratio)
plot(modV, what = "error", newdata = testData$SCratio,newclass = testClass)

###PREDICT SEX OF UNKNOWN SAMPLES BASED ON TRAINING SET##
pred <- predict(modV, testData$SCratio)  ###IF TEST DATA CONSISTS OF INDIVIDUALS OF UNKNOWN SEX
pred
sex<-pred$classification            ###EXTRACTS SEX CLASSIFICATION FOR UNKNOWN INDIVIDUALS
testData$"DAsexSC"<-NA
testData$DAsexSC<-sex                        ##APPENDS COLUMN WITH SEX CLASSIFICATION
testData
plot(sex)

###SAVE FILE WITH SEX IDENTIFICATION###
write.csv(testData,file="SexUnknowns.csv")

#############################################################################
### CLUSTERING USING MIXTURE MODELS WHEN ALL INDIVIDUALS ARE OF UNKNOWN SEX ###
#############################################################################

rm(list=ls())
library(mclust)

###READ DATA###
mydata<-read.csv(file="MyDataFile.csv",header=T)   ###READS FILE AND USES COLUMN HEADING FOR LABELS.

###CLUSTERING###
SC<-Mclust(mydata$SCratio)   ###READS COLUMN WITH DATA FOR SEXING
summary(SC, parameters=TRUE)
par(mfrow=c(2,2))
plot(SC)                            ###PLOTS GRAPHS (SAVE OR EXPORT GRAPHS AS NEEDED)
class<-SC$classification            ###EXTRACTS CLASSIFICATION OF INDIVIDUALS INTO GROUPS
mydata$"SexSC"<-NA
mydata$SexSC<-class                          ###APPENDS COLUMN WITH GROUP CLASSIFICATION (MALE, FEMALE AND
OUTLIER GROUPS)
mydata

uncerSC <- 1 - apply( SC$z, 1, max)  ###CALCULATES UNCERTAINTY OF CLASSIFICATION TO BEST GROUP PER
INDIVIDUAL
uncerSC
mydata$"unSC"<-NA
mydata$unSC<-uncerSC                         ###APPENDS COLUMN WITH UNCERTAINTY
mydata
quantile(uncerSC)                   ###SHOWS QUANTILES FOR UNCERTAINTY

###SAVE FILE WITH SEX IDENTIFICATION AND UNCERTAINTY DATA###
write.csv(mydata,file="SexClassified.csv")

###PLOT HISTOGRAM WITH GROUP ASSIGNMENT ###
library(ggplot2)
myBinwidth<-0.1
qplot(mydata$SCratio,binwidth=myBinwidth,fill=as.factor(class),xlab="18S/GAPDH Ratio",ylab="Count") +
geom_histogram(binwidth=myBinwidth,color="black")+theme(axis.line = element_line(colour = "black"),panel.grid.major =
element_blank(),panel.grid.minor = element_blank(),panel.background = element_blank())
```

B.4 Sample Dataset of *Apalone spinifera* 18S Copy Number Data

Table B.1: Example dataset of *A. spinifera* 18S copy number ("mydata" in Appendix B.2) quantified by qPCR using GAPDH as normalizer, a male-only standard curve, and the standard curve normalization method of 18S quantification; Gsex = gonadal sex. SCratio = 18S copy number quantified by the standard curve method of normalization. Set = sampled used in train or test set for discriminant analysis. SexSC = sex classification using mclust function in the absence of known sex information. DAsexSC = sex classification from discriminant analysis.

Sample	Gsex	SCratio	Set	SexSC	DAsexSC
Sample10	F	2.319923	train	2	F
Sample3	F	2.412821	train	2	F
Sample34	F	2.511757	train	2	F
Sample23	F	2.538339	train	2	F
Sample32	F	2.688288	train	2	F
Sample18	F	2.791118	train	2	F
Sample35	F	2.985106	train	2	F
Sample31	F	3.155844	train	2	F
Sample29	F	3.276169	train	2	F
Sample4	F	3.458333	train	2	F
Sample14	F	3.559099	train	2	F
Sample13	F	3.564437	train	2	F
Sample24	F	3.77533	train	2	F
Sample5	F	3.837209	train	2	F
Sample39	F	4.302966	train	2	F
Sample22	F	4.336232	train	2	F
Sample25	F	4.912381	train	3	F
Sample16	F	5.117225	train	3	F
Sample26	F	5.742553	train	3	F
Sample11	F	6.727627	train	3	F
Sample7	F	7.674374	train	3	F
Sample6	F	9.494475	train	3	F
Sample80	M	0.419836	train	1	M
Sample58	M	0.557123	train	1	M
Sample50	M	0.656997	train	1	M
Sample41	M	0.684857	train	1	M
Sample72	M	0.741722	train	1	M
Sample78	M	0.75	train	1	M
Sample74	M	0.762274	train	1	M
Sample77	M	0.799595	train	1	M
Sample79	M	0.829032	train	1	M
Sample76	M	0.839286	train	1	M
Sample73	M	0.857143	train	1	M
Sample49	M	0.891786	train	1	M
Sample66	M	0.923077	train	1	M
Sample62	M	0.977376	train	1	M
Sample65	M	1.082524	train	1	M
Sample64	M	1.091644	train	1	M
Sample48	M	1.107728	train	1	M
Sample47	M	1.150492	train	1	M
Sample67	M	1.150852	train	1	M

Table B.1 continued: Example dataset of *A. spinifera* 18S copy number ("mydata" in Appendix B.2) quantified by qPCR using GAPDH as normalizer, a male-only standard curve, and the standard curve normalization method of 18S quantification; Gsex = gonadal sex. SCratio = 18S copy number quantified by the standard curve method of normalization. Set = sampled used in train or test set for discriminant analysis. SexSC = sex classification using mclust function in the absence of known sex information. DAsexSC = sex classification from discriminant analysis.

Sample	Gsex	SCratio	Set	SexSC	DAsexSC
Sample56	M	1.245902	train	1	M
Sample46	M	1.289513	train	1	M
Sample61	M	1.350299	train	1	M
Sample70	M	1.408724	train	1	M
Sample27	F	2.446875	test	2	F
Sample30	F	2.655308	test	2	F
Sample15	F	2.917012	test	2	F
Sample19	F	3.166667	test	2	F
Sample33	F	3.542857	test	2	F
Sample2	F	3.637216	test	2	F
Sample21	F	4.186869	test	2	F
Sample17	F	4.446154	test	2	F
Sample9	F	5.27044	test	3	F
Sample8	F	7.595808	test	3	F
Sample45	M	0.583461	test	1	M
Sample71	M	0.702222	test	1	M
Sample75	M	0.762208	test	1	M
Sample44	M	0.820132	test	1	M
Sample69	M	0.854067	test	1	M
Sample43	M	0.892626	test	1	M
Sample54	M	1.028351	test	1	M
Sample55	M	1.089286	test	1	M
Sample42	M	1.117333	test	1	M
Sample59	M	1.162741	test	1	M
Sample63	M	1.319767	test	1	M
Sample51	M	1.422652	test	1	M

B.5 Application of Pipeline for Sex Diagnosis of the TSD *Chelydra serpentina*

Example of discriminant and clustering analyses using a dataset of *Chelydra serpentina* circulation testosterone levels measured by radioimmunoassay in individuals with reliable information of gonadal sex diagnosed by laparoscopy from Ceballos and Valenzuela (2011). Testosterone values correspond to gamma counter readings (counts per minute = CPM) for 136 individuals 4 hrs after FSH challenge following (Lance *et al.* 1992). It should be noted that it is females (and not males) which show elevated testosterone in response to the FSH challenge, while male levels are lower. CPM values are bimodal and the distributions for males and females are overlapping (Figure B.3a,b).

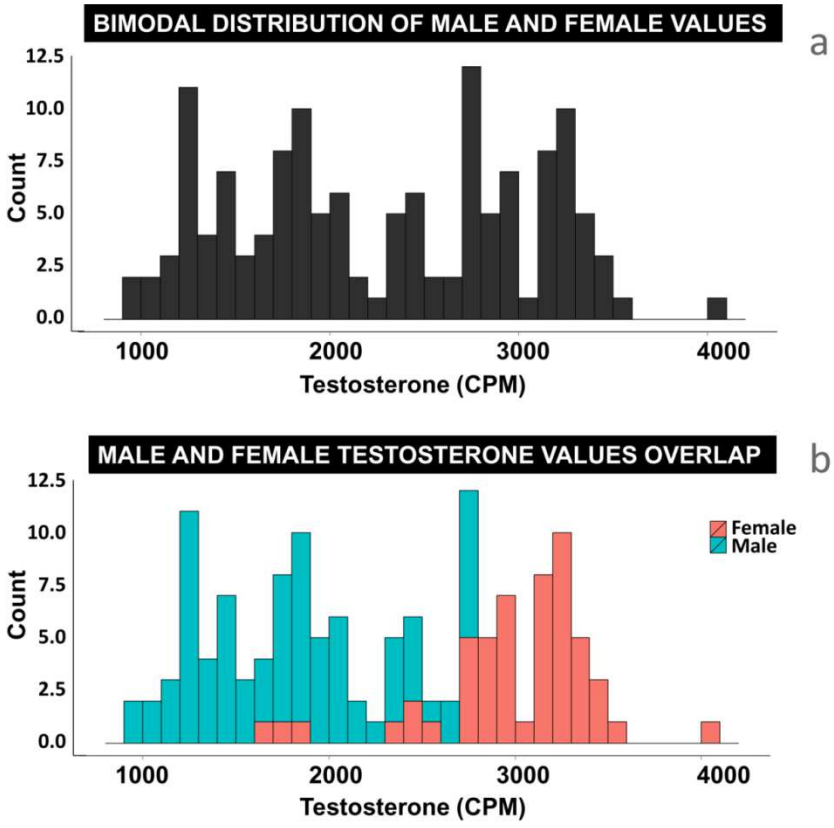

Figure B.3: Histogram distributions of circulating testosterone levels in *Chelydra serpentina* turtles following FSH challenge.

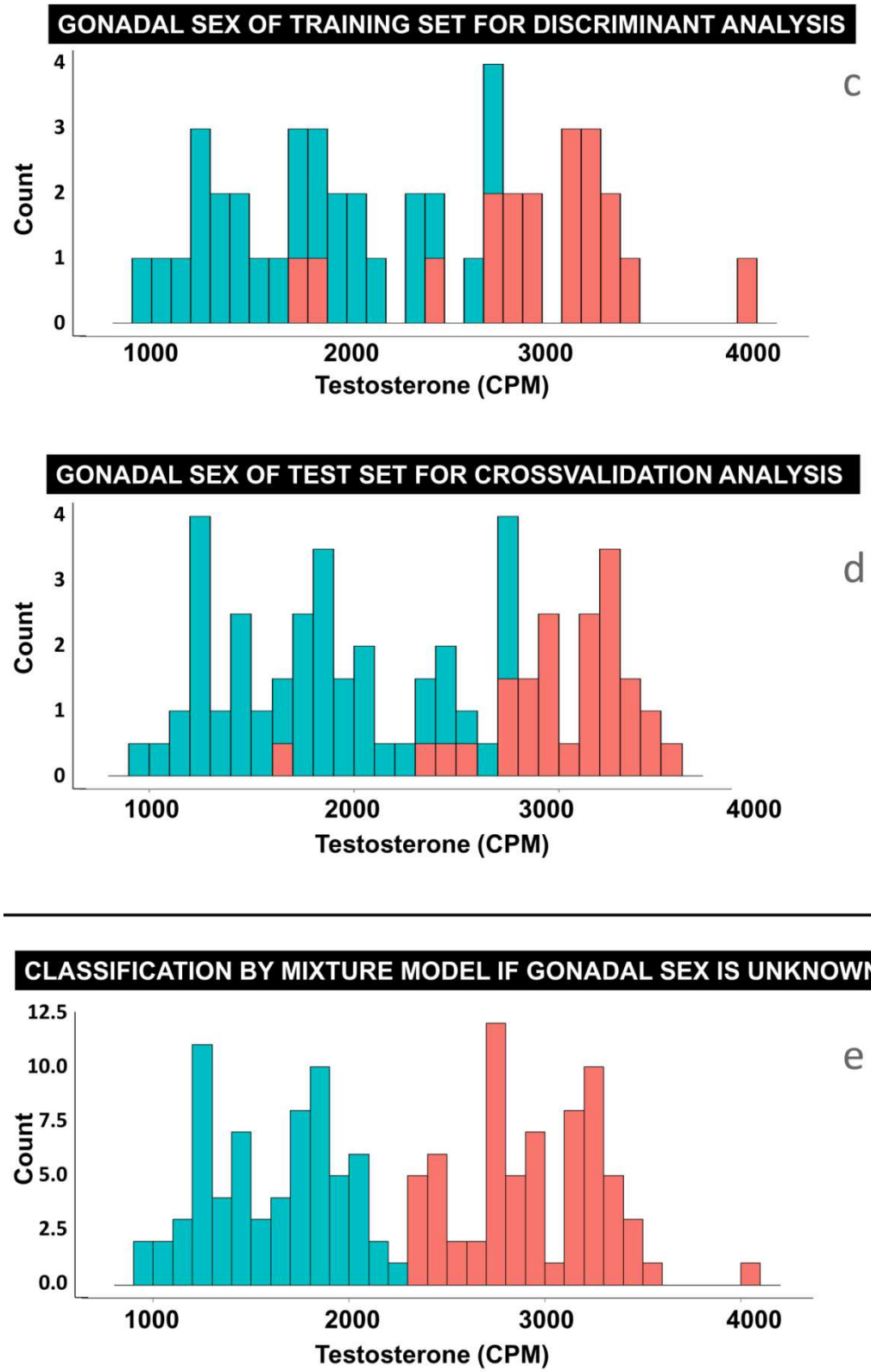

Figure B.3 continued: Histogram distributions of circulating testosterone levels in *Chelydra serpentina* turtles following FSH challenge.

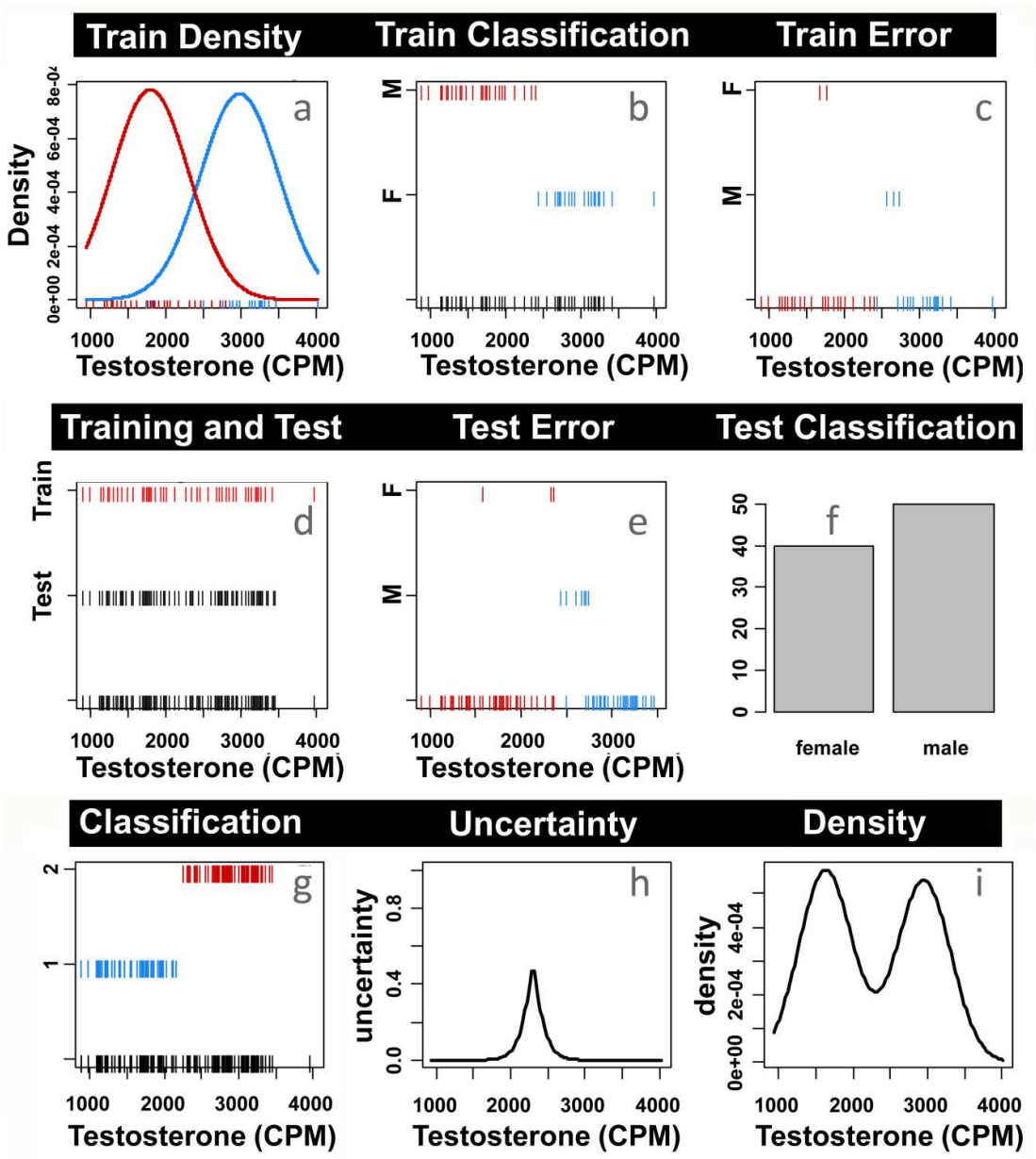

Figure B.4: Results from discriminant analysis for sex-typing using the sex information available from laparoscopy [including distribution density, classification and error rates for the training and test datasets]. The overlapping distribution of testosterone values causes some error in the classification of the training and test sets (c,e). Panels g-i illustrate the classification based on clustering in the absence of a priori sex information (ignoring the sex information from laparoscopy).

APPENDIX C: SUPPLEMENTAL MATERIAL FOR CHAPTER 4

C.1 Detailed Bioinformatics Methods

Identifying sex-linked loci using comparative read mapping (CRM)

As the first step in this protocol, we assembled a fragmentary *de novo* genome for each species using Ray (Boisvert *et al.* 2010), pooling the reads from both sexes and assembling them with default parameters and an arbitrary k-value of 31. Second, we mapped the DNA-seq reads obtained from the male and female samples against the assembled genome separately using Bowtie2 (Langmead and Salzberg 2012) with the 'sensitive' setting. Third, the mapping densities of the reads were compared between the sexes as follows: The RPKM (Reads Per Kilobase of scaffold per Million mapped reads) for each *de novo* Ray scaffold was calculated for each sex using the pileup.sh module from the BBMap package and the RPKM ratios of male to female reads onto each scaffold were assessed for *Glyptemys insculpta*, and of female to male reads for *Apalone spinifera* (See Methods for description of scaffold designation justification). Fragmentary genome assembly statistics and *de novo* scaffold classification information can be seen in Table C.1.

Leveraging existing genomic data to find sex-linked loci

Because each *de novo* genome was assembled from low-coverage sequencing data, scaffold lengths were shorter overall than those of an average eukaryotic gene, with the putative W- and Y-linked scaffolds being the shortest of all (Table C.1). This is expected because they represent a smaller portion of the genome and consequently of the sequencing data relative to X-linked, Z-linked, or autosomal sequences. Therefore, to distinguish whether potential Y/W-linked *de novo* scaffolds represented isolated indels that may be

specific to our heterogametic sequenced individual versus longer, contiguous DNA stretches sharing similar signals which might be more indicative of real sex-linkage, we leveraged existing high-coverage turtle genomes as reference (Shaffer *et al.* 2013, Wang *et al.* 2013). BLAST searches were performed using the Megablast program as implemented in the Geneious software package (v.9.1.2) (Kearse *et al.* 2012). Namely, our *G. insculpta* Y-scaffolds (GINY) were BLASTed against the genes and the genome of the painted turtle *Chrysemys picta* (Chrysemys_picta_bellii-3.0.1, Genbank Accession: GCA_000241765.1) which belongs to the same family (Emydidae), while our *A. spinifera* W-scaffolds (ASPW) were BLASTed against genes from the Chinese softshell turtle *Pelodiscus sinensis* (PelSin_1.0, Genbank Accession: GCA_000230535.1) which also belongs to the same family (Tryonichidae).

To identify protein coding genes that may exist on the *G. insculpta* Y or *A. spinifera* W, the potentially Y- and W-linked sequences were first BLASTed against a truncated version of all annotated genes from each reference genome. Because the introns and UTRs of genes are typically much longer than the coding region and often contain abundant repeat elements that could obscure downstream analysis, the exons of all annotated reference genes were extracted along with 400bp of surrounding intronic sequence. This intronic buffer was retained to allow for adequate BLAST results from shorter exons that otherwise could not be mapped. While many protein coding genes in the *P. sinensis* genome appeared to be Z- and W-linked in *A. spinifera*, including the two loci identified in the main text, many fewer *C. picta* genes appeared to by X- or Y-enriched in the GIN data. Because the BLAST results from the *G. insculpta* Y scaffolds lacked gene targets in the *C. picta* reference genome, the

G. insculpta Y *de novo* scaffolds were also BLASTed directly against the entire *C. picta* genome.

We ranked the BLAST results based on the percentage of the truncated gene (exons plus intronic buffer) that was covered by *G. insculpta* Y or *A. spinifera* W BLAST hits, and greater coverage over longer stretches was interpreted as stronger evidence of potential sex linkage. To analyze the mapping of the *G. insculpta* Y scaffolds against the *C. picta* reference genome, the *C. picta* genome was divided into 10kb windows and each window was scored for *G. insculpta* Y BLAST hit coverage using the makewindows and coverage modules from the BEDTools suite (Quinlan and Hall 2010). The genes and genomic scaffold windows with the highest BLAST coverage of *G. insculpta* Y or *A. spinifera* W hits were then manually inspected to detect regions suitable for PCR primer design.

Using de novo genome assemblies exclusively

Sex markers can also be identified in the absence of a reference genome using only the *de novo* genome data albeit with a higher probability of false positives even if a similar logic is followed and preference is given to longer *de novo* scaffolds. This can be accomplished by inspecting the read mapping ratios of male and female reads onto each *de novo* scaffold, as scaffolds that contain none or very few mapped reads from the homogametic sex can be considered potentially Y- or W-linked. Longer scaffolds lacking any mapped reads from the homogametic sex are potentially more informative. To investigate this in a *post-hoc* framework, the five longest *de novo* scaffolds to which no homogametic reads mapped from *G. insculpta* and *A. spinifera* were BLASTed against their respective reference genomes. The fourth largest heterogametic-only scaffold of *A. spinifera* BLASTed to the SETD1B variant identified through the reference genome method described

in the text, while the second largest scaffold of *G. insculpta* BLASTed to the sex-specific region of the *Chrysemys picta* NW_004848975 locus. In the absence of a reference genome one could therefore design primer sets within the largest *de novo* heterogametic-only scaffolds until a sex-specific region is identified by trial and error, although this approach has a higher potential for false positives. Nonetheless, for species of interest lacking an appropriate reference genome this modification may be a suitable approach allowing for marker discovery.

C.2 Supplemental Table

Table C.1: Putative sex-linkage classification scheme and numerical breakdown of *de novo* genome assemblies for *Apalone spinifera* and *Glyptemys insculpta*

	Apalone spinifera		*Glyptemys insculpta*	
RPKM Read Mapping Ratio (Heterozygote:Homozygote)	Scaffold Count	N50 (bp)	Scaffold Count	N50 (bp)
≥10 (putative W-/Y-linked)	35,277 (2.94%)	205	224,440 (7.97%)	215
0.40 – 0.667 (putative Z-/X-linked)	92,214 (7.68%)	1,255	267,787 (9.51%)	819
"Other" (autosomal/artifacts)	1,072,854 (89.38%)	2,807	2,323,362 (82.52%)	920
Total Scaffold Count (Pooled Sex Assembly)	1,200,345	2,715	2,815,589	860