GENES AND GENOMES OF REPTILES

By

JENA LIND CHOJNOWSKI

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To my mother and father and to all those who have overcome their own setbacks, big or small, to succeed in their happiness

ACKNOWLEDGMENTS

I would like to thank my emotional support system, which includes my mother (my best cheerleader), my friends who had to deal with my venting, and my soccer friends who helped with my frustrations. My undergraduate assistants, especially Vipa Bernhardt, Jenessa Graham, and Rachel Seibert, were crucial in the beginning stages of my work. I would also like to thank my committee members: Mike Miyamoto, Lou Guillette, Jr., Mike Fields, Rebecca Kimball, and my advisor Ed Braun. I would like to thank Mike Miyamoto for always being positive, Lou Guillette for his invaluable technical help and pushing me to the next level, and Mike Fields for showing me there is life outside of the lab. I would like to give special thanks to Rebecca Kimball for helping me every step of the way by not only providing emotional support but also technical and psychological support. Without her expertise I would not be where I am today. Ed Braun, my advisor, is my teacher, my mentor, and my friend, all things necessary for a successful and fruitful pairing, which constitutes our union.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

GENES AND GENOMES OF REPTILES
By

Jena Lind Chojnowski

August 2010

Chair: Edward L. Braun
Major: Zoology

The dynamic aspects of reptilian genomes are just starting to be discovered. Looking at reptilian genomes on two different scales, the whole genome and individual genes, allows us to examine different aspects of reptilian evolution. The genome can supply novel information about evolutionary trends in amniotes including guanine-cytosine (GC) content, ultimately called isochore structure. Also, in a more reductionist view, newly discovered genes involved in temperature-dependent sex determination (TSD) in a reptile can provide information about the evolution of sex determination. While using expressed sequence tag libraries from a turtle and an alligator, GC content was shown to increase with the evolution of amniotes. An increase in GC content is thought to promote the thermal stability of the genome; specifically, better allowing the genome to deal with overt thermal pressures. Since both turtles and alligators are poikilotherms the expectation for their GC content would be that they more closely resemble amphibians which have a similar temperature regulation system rather than birds and mammals which are homeotherms. The isochore structure of the turtle genome was intermediate between that observed for amphibians and mammals; the isochore structure of the alligator genome was very similar to that observed for birds and mammals. This suggests that an organisms' overall temperature regulation system

is not a driving force for GC content but a closer look at an organism's overall temperature regime including its average body temperature throughout its life is necessary. Another use for the turtle library was to determine novel genes involved in the process of TSD. Surprisingly, a non-coding RNA (ncRNA) called metastasis-associated lung adrenocarcinoma transcript 1 (*MALAT1*) was found to exhibit sexual dimorphism during embryonic development, the first ncRNA to exhibit this pattern of expression. *MALAT1* is a long ncRNA (~7kilobases) that has two variants after cleavage, it is upregulated in many human carcinomas and is correlated with cancer progression, and it is differentially expressed in mammalian gonads (specifically, it exhibits increased accumulation in ovaries). When information about the pattern of expression *MALAT1* exhibits in mammals is combined with this new finding of sexually dimorphic expression in turtles a regulatory function in vertebrate sexual development is suggested.

CHAPTER 1
INTRODUCTION

Genomes and Evolution

Genomes are shaped over time by multiple evolutionary trends. Different constraints can act upon individual nucleotides and over time the whole genome can reflect the changes. One example of an evolutionary trend of genomes at the nucleotide level is that Guanine-Cytosine (GC) content promotes genome stability; meaning, that a high GC content for a genomic region decreases mutation rates due to thermal instability in that region.

A high GC content within a region of the genome promotes genomic stability in the presence of overt thermal pressures. The three hydrogen bonds that connect Guanine and Cytosine are stronger and harder to break than the two hydrogen bonds connecting Adenine (A) and Thymine (T) in a volatile environment like high temperatures. Breaking the connection between paired nucleotides can lead to mutation due to a faulty repair system. Therefore, high temperatures can lead to higher mutation rates in regions with less hydrogen bonding or more AT content.

The vertebrate genome is divided into long (>100 kilobase [kb]) regions with relatively homogenous base composition with sharp boundaries producing distinct patterns of AT-rich regions and GC-rich regions. The genomic divisions, called isochores or collectively called isochore structure, are hypothesized to be a reflection of evolutionary aspects such as mean body temperature, historical environmental temperature, and gene stability. When comparing across genomes, amphibians and fishes (cold-blooded or poikilothermic) contain relatively homogenous genomes that are AT-rich and mammals and birds (warm-blooded or homeothermic) have more

heterogeneity across the genome with more GC-rich regions, with some regions as low as 30% GC and others as high as 60% GC. The GC-rich regions of mammals and birds are not only more frequent across the genome but also contain a higher GC percentage when compared to the more limited GC-rich regions of amphibians and fishes (e.g., most of the *Xenopus* genome has <45% GC; see Bernardi 1995). These data led to the assumption that AT-rich isochores are ancestral and conserved among vertebrates.

Many hypotheses propose to explain isochore evolution. Two different classes of models have been proposed for the origin of GC-rich isochores: selection based models that postulate functional biases and mutation patterns correlated with selectively neutral regional changes. The selection and mutation models can further be placed into a variety of mechanistic categories like thermal or nonthermal biases (Figure 1-1; modified from Chojnowski et al. 2007). Though there are many hypotheses for isochore structure this dissertation will expound upon hypotheses that relate to only selection and thermal biases.

Bernardi (2000) proposed the prototypical thermal selectionist model which hypothesizes that deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and proteins that are encoded in GC-rich regions have a higher thermodynamic stability and thus, are better protected against deleterious mutations in volatile temperatures. In other words, GC-rich isochore structure is an adaptation to high body temperature. Evidence for this hypothesis is the observation that homeothermic birds and mammals have more and higher percentage GC-rich regions than the poikilothermic amphibians and fish. Furthermore, about 90% of all genes in humans are in GC-rich regions.

If the hypothesis that an adaptation to homeothermy for isochore structure is correct then all poikilothermic vertebrates should have the homogenous isochore structure with relatively high amounts of AT-rich regions similar to amphibians and fishes. However, Chojnowski et al. (2007) showed using 3rd codon position (GC3) data obtained from American alligator expressed sequence tags (ESTs) that alligators have a similar isochore structure to birds and mammals even though they are poikilothermic. GC3 was used as a surrogate for isochore structure because strong correlations have been repeatedly shown between GC3 and their surrounding DNA (the isochores in which genes are embedded) for both poikilotherm (Bernardi and Bernardi 1991) and homeotherm genomes (Bernardi 2000; Musto et al. 1998). Though alligators are poikilothermic they are able to maintain a mean body temperature similar to homeotherms through behavioral means. Thus, thermal hypotheses need to reflect overall maintenance of body temperature and changes in thermal biology that occurred during the origin of amniotes. A broader survey across reptiles will help to further examine how different organisms' body temperatures can affect isochore structure evolution.

Isochores and Turtles

Mammals and living archosaurs (birds and crocodilians) have heterogeneous genomes that include very GC-rich isochores. In sharp contrast, the genomes of amphibians and fishes are more homogeneous and they have a lower overall GC content. Because DNA with higher GC content is more thermostable, the elevated GC content of mammalian and archosaurian DNA has been hypothesized to be an adaptation to higher body temperatures. This hypothesis can be tested by examining structure of isochores across the reptilian clade, which includes the archosaurs,

testudines (turtles), and lepidosaurs (lizards and snakes), because reptiles exhibit diverse body sizes, metabolic rates and patterns of thermoregulation. The study (Chojnowski and Braun 2008) focuses on a comparative analysis of a new set of expressed genes of the Red-eared slider turtle and orthologs of the turtle genes in mammalian (human, mouse, dog, and opossum), archosaurian (chicken and alligator) and amphibian (Western clawed frog) genomes. EST data from a turtle cDNA (complementary DNA) library enriched for genes that have specialized functions (developmental genes) revealed using the GC content of the third codon position to examine isochore structure requires careful consideration of the types of genes examined. The more highly expressed genes (e.g., housekeeping genes) are more likely to be GC-rich than are genes with specialized functions. However, the set of highly expressed turtle genes demonstrated that the turtle genome has a GC content that is intermediate between the GC-poor amphibians and the GC-rich mammals and archosaurs. There was a strong correlation between the GC content of all turtle genes and the GC content of other vertebrate genes, indicating that the isochore structure of turtles is intermediate between that of amphibians and other amniotes. These data are consistent with some thermal hypotheses of isochore evolution, but we believe that the credible set of models for isochore evolution still includes a variety of models. These data expand the amount of genomic data available from reptiles upon which future studies of reptilian genomics can build.

Temperature-dependent Sex Determination

In many turtle species, sex is determined by a process known as temperature-dependent sex determination (TSD)(Crews et al. 1994). In contrast to genotypic sex determination (GSD), which is exemplified by the mammalian system that uses the sex-

determining region Y (*SRY*) gene (mammals expressing *SRY* develop as males)(Sinclair et al. 1990), the developmental cascade of TSD leading to gonadal differentiation in turtles is relatively poorly characterized. Incubation temperature provides the signal that leads to sex determination in turtles that exhibit TSD by altering the expression of specific genes, such as those encoding for steroidogenic enzymes and steroid hormone receptors, during embryogenesis (Crews and Bergeron 1994). However, the complete set of genes that are regulated by the environmental cue of incubation temperature remains a mystery.

Physiological changes that are equivalent to those caused by temperature can be elicited by the administration of exogenous estradiol (generating females) or nonaromatizable androgen (generating males) during incubation (Crews et al. 1991). Thus, the application of these hormones topically to the developing embryo before the critical sex committal stage can reverse the course of TSD. This system clearly has the capability to readily lend itself to experimental manipulation. In fact, the identification of turtle genes that are differentially expressed during TSD has the potential to provide a unique model system for sexual development in vertebrates, given the ease of manipulating the triggers for gonadal differentiation within the turtle system.

There are many genes conserved across taxa that are involved in GSD. It is clearly desirable to examine their role in TSD. Turtle orthologs of genes involved in mammalian GSD, such as cytochrome P450 (*CYP19*), steriodogenic factor (*SF1*), wilms' tumor (*WT1*), SRY-box 8 (*SOX8*),SRY-box 9 (*SOX9*), doublesex and mab-3 related transcription factor 1 (*DMRT1*), and nuclear receptor subfamily 0, group B, member (*DAX1*), have been identified (Fleming et al. 1999; Kettlewell et al. 2000;

Murdock and Wibbels 2003a; Schmahl et al. 2003; Takada et al. 2004; Torres Maldonado et al. 2002). All of these genes, with the exception of *SOX8*, have been found to have differential expression patterns in turtle embryos, or specifically gonads, based on changes in temperature. However, the total sets of genes involved in GSD remains unknown, limiting the use of homology to identify genes responsible for the TSD cascade in turtles. A more fundamental limitation is imposed by the fact that there must be differences between GSD and TSD, especially near the top of the GSD and TSD cascades since one is triggered by the expression of one or more specific genes and the other is triggered by environmental stimuli. Although, it is possible that the genes near the top of the GSD cascade could simply be subject to regulation by different incubation temperatures for TSD, or other environmental stimuli involved in environmental sex determination (ESD) systems, but it has been shown that downstream genes are more highly conserved than upstream genes across taxa with different sexual systems. Thus, there is empirical evidence that upstream regulators have exhibited more variation than either drift or, more likely, selection has been able to act upon (Western et al. 2000). This provides yet another reason why the use of genes involved in GSD is limiting (Modi and Crews 2005). So, the use of genes known from the mammalian sex determination pathway to identify those involved in turtle TSD only has the potential to identify downstream genes. This project aims to identify a set of genes that are greatly involved in up- and downstream events of the turtle TSD pathway.

Further understanding of the molecular mechanisms underlying TSD in reptiles like the Red-eared slider turtle (*Trachemys scripta)* will provide useful information on a variety of topics ranging from the evolution of sex chromosomes to the generation of

probes. A method to establish candidate genes for the process of TSD in a manner that is less biased by work in mammals has the potential to greatly improve and speed the process of establishing the genetic cascade of sex determination in *Trachemys scripta*, and ultimately in other reptiles. Establishing candidate genes and characterizing them in the developing embryo has major implications for the advancement of research in this field.

Genotypic Sex Determination versus Temperature-dependent Sex Determination

Genotypic sex determination, defined by genotype, typically reflects the presence of heteromorphic sex chromosomes, although the early stages of sex chromosome evolution may be characterized by sex chromosomes that are indistinguishable (homomorphic). Sex chromosomes and specific genes with a role in GSD have arisen independently in amphibians, reptiles, birds, insects, and mammals (Miller et al. 2004). The heteromorphic chromosomes are known to operate differently in various groups. Two known mechanisms for GSD are responses to dosage of the sex chromosome present in both sexes and the presence or absence of the dominant heterogamete (Marin et al. 2000; Marshall Graves and Shetty 2001). One of the best studied systems is that used by eutherians, where the product of the *SRY* causes gonadal differentiation into testes and subsequent secondary sex-specific features for males. Absence of the Y chromosome, which contains *SRY,* results in the default pathway for female development. But, the *SRY* gene is found only in mammals and not in any other vertebrate classes that use GSD. In fact, there have been a number of independent transitions between GSD and TSD (Valenzuela and Lance 2004).

The downstream molecular mechanisms responsible for sexual differentiation of gonads appear to be conserved over evolutionary time despite differences in upstream

triggers among taxa (Johnston et al. 1995). GSD is ultimately controlled by the presence of one gene or a suite of genes, usually unaffected by the external environment. In contrast, ESD reflects control of sexual development by the environment, and temperature is only one of many potential environmental triggers, including factors like pH, crowding, and water potential. Temperature must alter the production of sex factors (e.g. growth factors, transcription factors, hormones, steroids) that in turn over a sufficient time and rate initiate a genetic sex determination cascade (Wibbels et al. 1991). Whatever specific molecule acts as a sex factor and a trigger must be a gene product (a polypeptide or, in principle, an RNA) or the product of a biochemical reaction mediated by an enzyme. Therefore, environmental modulation of gene products represents the ultimate basis of TSD, and finding the genes that are modulated is of paramount interest. New genes determined to be important for TSD have the potential to be involved in GSD, not as a master switch but instead filling in some missing pieces in the overall scheme.

How Temperature-dependent Sex Determination Works

TSD is unknown in snakes, birds, mammals, and amphibians; has been found infrequently in fish and lizards; is prevalent in turtles; and is the only pattern in tuataras and crocodiles. For turtles, 79 species have been assayed by incubation at controlled temperatures, revealing that 64 have TSD and 15 have GSD. However, the majority (~70%) of turtle species remain untested (Valenzuela and Lance 2004).

Over time three different modes of TSD have arisen that are characterized by sex ratios as a function of incubation temperature. TSDIa, or male-female (MF), results in males at low temperatures and females at high temperatures. TSDIb, or female-male (FM), results in males at high temperatures and females at low temperatures. TSDII, or

female-male-female (FMF) results in females at both high and low temperatures and males at intermediate temperatures. A population wide 1:1 sex ratio is termed the pivotal temperature and mixed sex ratio temperatures are termed the transitional range, intersexes never having been reported in nature. During embryonic development, sex is established within a temperature-sensitive period (TSP), which contains a critical stage for sex commitment. It is unclear whether the development of different sexes at the same temperatures reflects genetic variation in study populations, stochastic aspects of gene expression, maternal effects, or a combination of these. But it will be impossible to ascertain the important factors without a thorough understanding of TSD.

Trachemys scripta – A Model System to Study Temperature Affects of Turtles

Trachemys scripta determines sex by TSDIa. In the laboratory the constant incubation temperature to produce 100% females is 31°C, to produce 100% males is 26°C and the pivotal temperature is 29.2°C. For this species TSP is set between developmental stages 14 and 21, stage 17 (the critical stage), and sex is determined by stage 21 (Figure 1-2) (Crews et al. 1994; Wibbels et al. 1991).

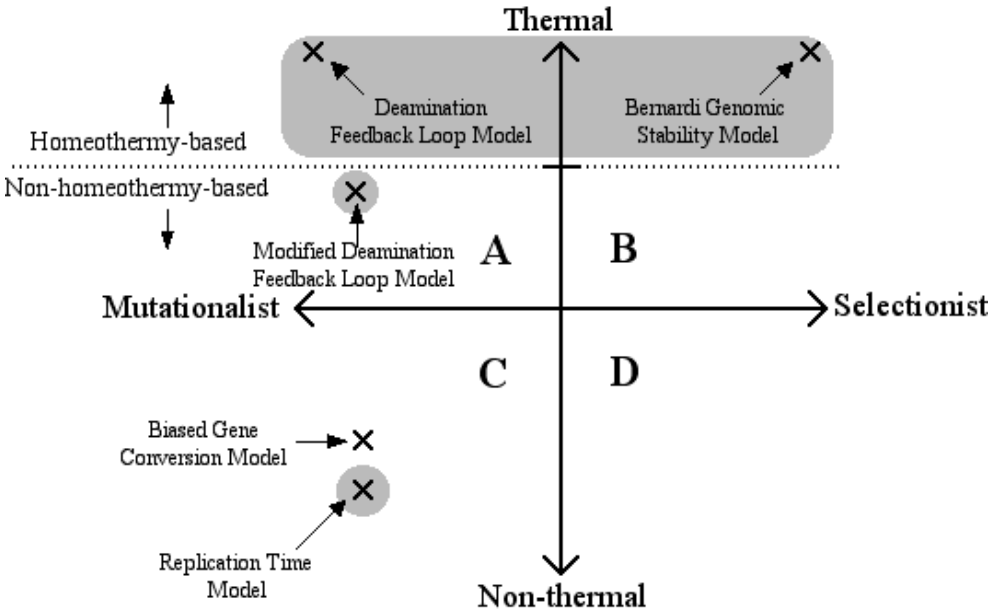


Figure 1-1. A unified framework for models of isochore evolution.

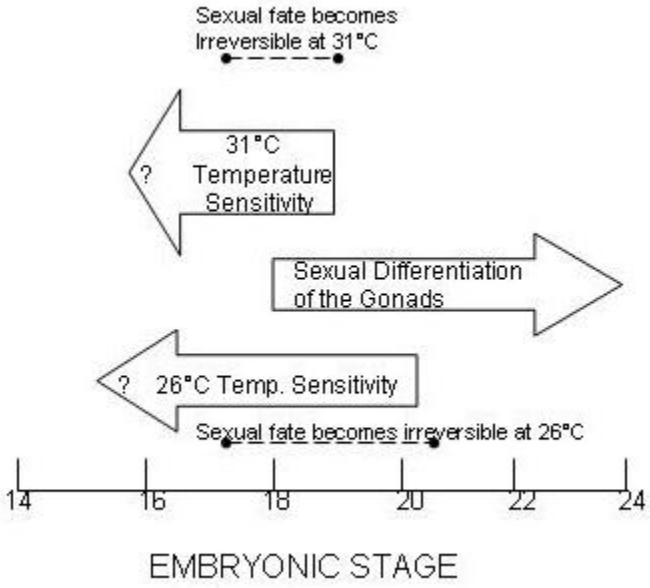


Figure 1-2. Chronology of temperature sensitivity. Temperature is on the y-axis and embryonic stage is on the x-axis. The female-producing temperature is at 31°C and the male-producing temperature is at 26°C. Modified from (Wibbels et al. 1991).

CHAPTER 2
PATTERNS OF VERTEBRATE ISOCHORE EVOLUTION REVEALED BY COMPARISON OF EXPRESSED MAMMALIAN, AVIAN AND CROCODILIAN GENES.

Introduction

Vertebrate nuclear genomes are characterized by distinct biases in nucleotide composition, with strongly (guanine –cytosine [GC]) and weakly (adenine-thymine [AT]) base-pairing nucleotides clustering in long (>100 kilobases [kb]) regions of relatively homogenous base composition called isochores. Neighboring isochores appear to have relatively sharp boundaries, a feature that originally made it possible to separate and characterize isochores using density gradient ultracentrifugation (Bernardi 2000). The availability of draft genome sequences for model systems (e.g. mouse, chicken, *Arabidopsis*) has made it possible to define isochores computationally using methods like the traditional sliding window approach and the Z-curve method (Zhang et al. 2001). The regional GC content of genomes is correlated with many important features, like the distribution of genes and repetitive elements (Bernardi 2000; Hackenberg et al. 2005), chromosomal banding (Saccone 1997), and patterns of C—phosphate—G (CpG) methylation (Caccio et al. 1997).

The genomes of warm-blooded (more properly, homeothermic) amniotes (birds and mammals) exhibit a distinct genomic heterogeneity in GC content, with some regions as low as 30% GC and others as high as 60% GC (Bernardi 1995). Conversely, cold-blooded (poikilothermic) vertebrates have been proposed to have a more homogeneous genome (Bernardi and Bernardi 1991), and the poikilotherms that are best characterized from a genomic standpoint (fish and amphibians) generally have lower average GC content than birds and mammals (e.g., most of the *Xenopus* genome has < 45% GC; see Bernardi 1995). These data led to the assumption that GC-poor

isochores are ancestral and conserved among vertebrates. Two different classes of models have been proposed to explain the origin of GC-rich isochores: those based on selection and those based on mutational patterns combined with the fixation of selectively neutral changes (Figure 2-1). The former postulate that GC-rich isochores arose for functional reasons (i.e. natural selection) and the latter postulate the existence of regional mutation biases.

Selectionist and mutationalist models of GC-rich isochore evolution can be placed into a variety of mechanistic categories. Previous analyses of reptilian genome data have addressed the role of thermal factors in isochore evolution (Hughes et al. 1999; Eyre-Walker and Hurst 2001; Varriale and Bernardi 2006), and this study will focus on the division between thermal and non-thermal models (Figure 2-1). Bernardi proposed the prototypical thermal selectionist model, postulating that the selective advantage of GC-rich isochores stems from higher thermodynamic stability of deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and important proteins (e.g., housekeeping proteins) encoded in GC rich regions – "Bernardi genomic stability" hypothesis. Alternatively, Fryxell and Zuckerkandl (2000) proposed a thermal mutationalist model known as "deamination feedback loop" that postulates a higher rate of cytosine deamination (leading to C→T transitions) in GC-poor regions at high temperatures, further reducing their GC content and strengthening the feedback loop. This model has a unidirectional GC→AT mutational bias, so it is only able to explain the origin of a heterogeneous genome if the homogeneous ancestral genome was GC-rich.

Although there has been substantial focus on thermal models, non-thermal models based upon selection or mutation have also been proposed (reviewed in Li

1997, p. 407-411; Eyre-Walker and Hurst 2001). One of the best studied models involves GC-biased gene conversion (Galtier et al. 2001), and this model has the benefit of explaining the observed correlation between recombination frequency and GC-content (for details see Eyre-Walker 1993; Duret et al. 2002). However, these different models form a continuum where selection and mutation can either generate similar compositional patterns or act in opposition to each other. Placing all of these models into this common framework (Figure 2-1) simplifies the examination of different models.

Bernardi (2000) cites several observations as support for the thermal selectionist model. First, homeotherms have strong GC-rich isochore structure while poikilotherms do not. Second, density gradient centrifugation data indicating a number of non-avian reptiles (poikilothermic amniotes) lack the strong GC-rich isochore structure that is characteristic of birds and mammals. Those data suggest the GC-rich isochore structure arose independently in birds and mammals from the ancestral "cold-blooded" isochore structure. Third, there is evidence for selection on GC-rich (but not GC-poor) mammalian genes (see Cacciò et al. 1995). Finally, there is population genetic data suggesting that the mammalian major histocompatibility complex (*MHC*) genes have been subject to selection favoring AT → GC mutations or that GC-biased gene conversion has occurred (see Eyre-Walker 1999; note that a "biased gene conversion" model could be a non-thermal mutationalist alternative to models based on selection; Figure 2-1). If a thermal model is assumed, combining the data highlighted by Bernardi (2000) and the difficulty reconciling the Fryxell and Zuckerkandl (2000) deamination feedback loop model with a GC-poor ancestral genome creates a framework in which

the preponderance of evidence points to a selectionist model. Testing thermal selectionist models requires a rigorous examination of the independent origins of GC-rich isochores in birds and mammals, which can be accomplished using comparative genomics.

The apparent phylogenetic support for independent origins of a GC-rich isochore structure in birds and mammals has been proposed to be compelling evidence for a thermal selectionist model (e.g., Bernardi 2000), but two critical assumptions should be examined before accepting such a model. First, the strength of the evidence that the GC-rich isochore structure of birds and mammals was derived independently from a GC-poor ancestral condition should be carefully examined. Second, it is unclear whether increased GC content provides enhanced thermal stability that is advantageous for homeotherms.

Several studies suggest that poikilothermic amniotes have GC rich isochores. Hughes et al. (1999) reported that 3rd codon position (GC3) for two poikilothermic amniotes, the Red-eared slider turtle (*Trachemys scripta*; six genes) and the Nile crocodile (*Crocodylus niloticus*; ten genes), resembles the GC3 of the chicken. Hamada et al. (2002) expanded this small dataset to eleven genes by adding α-globin genes from three snake species, also supporting the existence of GC-rich isochore structure in non-avian reptiles. Hughes et al. (1999) and Hamada et al. (2002) both suggest the common ancestor of amniotes may have had a GC-rich isochore structure, but the limited number of sequences used makes it difficult to view those studies as definitive evidence for the presence of GC-rich isochores in non-avian reptiles.

This study tests critical aspects of the current thermal selectionist model by comparing GC3 for a set of American alligator (*Alligator mississippiensis*) expressed sequence tag (EST assemblies to the GC3 of their orthologs in homeothermic amniotes (the chicken, human, and mouse) and a poikilothermic vertebrate (the western clawed frog *Xenopus tropicalis*). We use GC3 variation as a surrogate for isochore structure because previous analyses have shown that GC3 is correlated with intronic GC content (GCi) and the GC content of the larger genomic regions (isochores) in which the genes are located (Bernardi 2000). Our analyses strongly support the existence of a GC-rich isochore structure in the alligator, providing evidence that the shift to GC-rich avian isochores occurred prior to their divergence from crocodilians. This result has profound implications for the set of models that have been used to explain the evolution of isochore structure, and we were able to falsify a subset of these models.

Methods

EST Collection

A total of 6,732 reads (accession numbers are ES316475 to ES321899), ranging in length from 76 to 812 nucleotides (mean 484.4), were obtained from three alligator complementary DNA (cDNA) libraries using either Applied Biosystems (ABI 377 or ABI 3100 Avant) or Amersham (MegaBACE) automated sequencers. The libraries were from juvenile liver, adult liver, and adult testis, and all reads were assembled essentially as described in Liang et al. (2000), yielding a total of 3,064 assemblies. The overall GC-content for these assemblies is provided in the supplementary information (Figure 2-2), along with the correlation between the overall GC-content and the GC3 based upon these EST assemblies (Figure 2-2).

EST Assembly and GC3 Analyses

EST assemblies were used as tBLASTx (Nucleotide 6-frame translation-nucleotide 6-frame translation Basic Local Alignment Search Tool) (Altschul et al. 1997) queries to search cDNA sequences from the EnsEMBL database (Birney et al. 2006). Human, mouse, chicken and frog sequences identified by the alligator query were aligned using Clustal W (Chenna et al. 2003) and alignments that appeared largely anomalous were discarded. The anomalous alignments included those that did not have an ortholog in one of the three amniote species (mouse, chicken, or human) and those that were difficult to align for other reasons (e.g. due to misannotations). The remaining alignments (a total of 366, 98 of which include the frog) were optimized by eye using MacClade (Maddison and Maddison 2002). The segment of the alignment present in all organisms was assigned to a CHARSET (character set). Codon positions were identified and third position base composition was calculated using pylogenetic analysis using parsimony* (PAUP*) 4.0b10 (Swofford 2003). Base composition values were extracted from the PAUP logfile using a script and imported into Microsoft Excel.

Since strong correlations have been repeatedly shown between GC3 of protein-coding genes, their introns, and their surrounding DNA (the isochores in which genes are embedded) for both poikilotherm (Bernardi and Bernardi 1991) and homeotherm genomes (Bernardi 2000; Musto et al. 1999), the correlation between GC3 values for orthologous coding sequences will be used to study isochore structure. This follows other studies that have examined vertebrate isochore structure (Zoubak et al. 1996) and evolution (Bernardi et al. 1997; Galtier and Mouchiroud 1998). This correlation was examined using the open source of the R software (The R Project for Statistical Computing, http://www.r-project.org/ 1997) and the equations were fit in R using

orthogonal regression (Isobe et al. 1990). Lines for the best-fitting equation and unity were added to all plots.

Results

Histograms of GC3 content for 366 assembled alligator ESTs that could be aligned with their human, mouse and chicken orthologs revealed a remarkable similarity between the alligator and the homeothermic amniotes (Figure 2-3). A subset of these assemblies that could also be aligned with their orthologs in the western clawed frog revealed a striking difference in GC3 content between the frog and all of the amniotes, including the alligator (Figure 2-3). When our EST assembly data is viewed in light of the evidence for a correlation between GC3 and the GC content of flanking regions (Bernardi 2000), it suggests that alligators show a degree of genomic heterogeneity due to isochore structure similar to that exhibited by homeothermic amniotes. Only the mouse histogram stands out within the amniotes as having a narrower distribution, and this observation is consistent with previous studies (Bernardi 2000).

Examining the relationships among amniotes in GC3 content revealed a strong positive correlation for all comparisons (Figure 2-4), although comparisons of more distantly related organisms showed a lower correlation coefficient than those of more closely related organisms, as expected. GC3 values for amniotes and the frog, used as a representative poikilotherm in previous analyses (e.g., Bernardi 2000), were also correlated (Figure 2-5), although the correlation was weaker than that observed within amniotes. The slopes of lines describing the relationship between GC3 values in different amniotes were close to one, with only the human-mouse comparison exceeding unity by a large amount (Figure 2-4). In contrast, comparison of either human or alligator with that of the frog revealed a slope even greater than that evident in the

human-mouse comparison (Figure 2-5A), making it clear that GC3 values of both homeothermic and poikilothermic amniotes show the same relationship to the frog (Figure 2-5B).

Discussion

These GC3 data indicate that the alligator isochore structure is quite similar to the avian and mammalian isochore structure and different from the amphibian isochore structure. This indicates that strong GC-rich isochore structure, suggested to be present only in homeotherms, actually arose prior to the origin of homeothermy (Figure 2-6). Like Duret et al. (2002), we propose that the strong GC-rich isochore structure arose during the origin of amniotes.

The previous work based on a small number of reptilian sequences (Hughes et al. 1999; Hamada et al. 2002) is easy to reconcile with a model in which strong GC-rich isochore structure is a feature of all amniotes and is not limited to homeotherms. In fact, the slope (*m*) of the line relating chicken and alligator GC3 observed here (*m*=0.88) is quite similar to that observed by Hughes et al. (1999) in a comparison of Nile crocodile and chicken genes (*m*=0.85), suggesting that their results were not unreasonably biased by the small number of genes used. Furthermore, Hughes et al. (1999) observed a similar slope (*m*=0.77) in another comparison of six Red-eared slider turtle and chicken genes. However, the much broader survey of genes in our EST-based study clearly provides greater confidence in the conclusion that GC-rich isochore structure was present in the common ancestor of archosaurs (birds and crocodilians). A logical extension of this conclusion that considers the previous work (e.g., Hughes et al. 1999; Duret et al. 2002; Hamada et al. 2002) would be a model placing the origin of strong GC-rich isochore structure in the common ancestor of all amniotes, but further

corroboration of that model will require broader surveys of many additional non-avian reptiles.

The existence of a strong GC-rich isochore structure in the alligator has profound implications for selecting the plausible models for the evolution of isochore structure from the larger set of potential models (Figure 2-1). Both the Bernardi (2000) genomic stability model and the deamination feedback loop model (Fryxell and Zuckerkandl 2000) predict that homeotherms will have a strong GC-rich isochore structure and that poikilotherms will not (hence the “warm-blooded isochore” terminology typically used in the literature). Since our data indicate a poikilotherm (the alligator) has a strong GC-rich isochore structure similar to homeotherms both of these models should be considered falsified. However, basing thermal models upon the division between homeotherms and poikilotherms may be misleading, since there are poikilotherms that are able to maintain a high and stable body temperature similar to homeotherms through behavioral mechanisms (Seebacher et al. 1999; Seebacher and Shine 2004). Thus, the basis for thermal models can reflect changes in thermal biology that occurred during the origin of amniotes (Figure 2-6).

The evidence that the origin of GC-rich isochore structure was independent of a shift from poikilothermy to homeothermy (Figure 2-6) places an important novel constraint on thermal models of isochore evolution. Non-homeothermic thermal models must consider multiple body temperature parameters (e.g., both the maximum and mean temperature). For example, a non-homeothermic version of the Bernardi (2000) genomic stability model might be coupled to the maximum body temperature. In contrast, a non-homeothermic variant of the deamination feedback loop model is more

problematic (noted in Figure 2-1 as Modified Deamination Feedback Loop Model), since the accumulation of thermally-driven mutations would be expected to reflect the mean body temperature. Thus, less change from the ancestral state (high GC) is predicted in poikilotherms which should lead to higher GC content in alligators relative to homeotherms. Our data contradicts this prediction, as we found that low GC3 genes in chicken also have low GC3 in the alligator (Figure 2-4D).

Genomic comparisons similar to this study that focus on organisms from different thermal environments and with different mechanisms for physiological and behavioral thermoregulation may further constrain thermal models of isochore evolution. For example, the tropical pufferfish *Tetraodon nigroviridis* has a ~4% higher mean GC content than its temperate relative *Fugu rubripes* (Jabbari and Bernardi 2004), which is consistent with a non-homeothermic thermal model of isochore evolution. However, when these types of comparisons are conducted on a large scale it will be important to acknowledge that isochore structure may be more reflective of the historical environment of organisms' ancestors than the present-day environment. In fact, there are aspects of evolutionary history that may have an impact upon the present study, like the possibility that the ancestors of extant crocodilians were homeotherms (Seymour et al. 2004; but see Hillenius and Ruben 2004). Thus, the evaluation of genomic data should be conducted in a phylogenetic framework that also considers the paleoenvironment.

The phylogenetic evidence suggesting a single origin of the GC-rich isochore structure of amniotes indicates that non-thermal models (regions C and D of Figure 2-1) should receive additional attention. It is easy to imagine specific genetic changes at the

ancestor of modern amniotes that might establish either a mutational bias or a selective environment that would generate a genomic structure uniting the amniotes (like the strong GC-rich isochore structure). For example, a change in chromatin proteins could lead either to regional mutational biases or to selection that favors GC-rich alleles in genic regions. The accumulation of evidence suggesting that isochore evolution may not have a thermal basis should open the field to novel routes of inquiry.

Both thermal and non-thermal models should be examined using a historical framework. For example, the strong GC-rich isochore structure of mammals has been degrading (Duret et al. 2002; Webster et al. 2006), suggesting that analyses of mammalian genomes alone may not reveal the processes that led to the strong GC-rich isochore structure of amniotes. This is illustrated by Eyre-Walker's (1992) test of the Wolfe et al. (1989) replication time model (a non-thermal mutationalist model). The replication time model postulates that genes replicating early in S (synthesis) phase of the cell cycle show different mutational patterns (e.g., AT→GC) than those replicating late in S phase. Using 44 mammalian (primate and rodent) genes, Eyre-Walker (1992) did not find the correlation between replication time and GC-content expected if this model were correct (Figure 2-1). Although this result is compelling, it remains possible that birds will exhibit this correlation because GC-rich isochore structure appears to be strengthening in birds (Webster et al. 2006). If so, the absence of this correlation in mammals might reflect the mechanistic basis for the degradation of GC-rich isochores reported by Duret et al. (2002). Thus, a broader phylogenetic and historical perspective, including data from birds and other reptiles, would greatly increase the power of such analyses.

Our analysis of alligator EST data provides a framework that can be expanded by the acquisition of EST or genomic sequence data from additional organisms, especially reptiles. The present work provides a conclusive answer regarding the existence of a strong GC-rich isochore structure in a poikilotherm. However, a broader survey would have the potential either to further constrain the set of credible models for isochore evolution or provide novel information necessary to define additional models that are better able to explain isochore evolution.

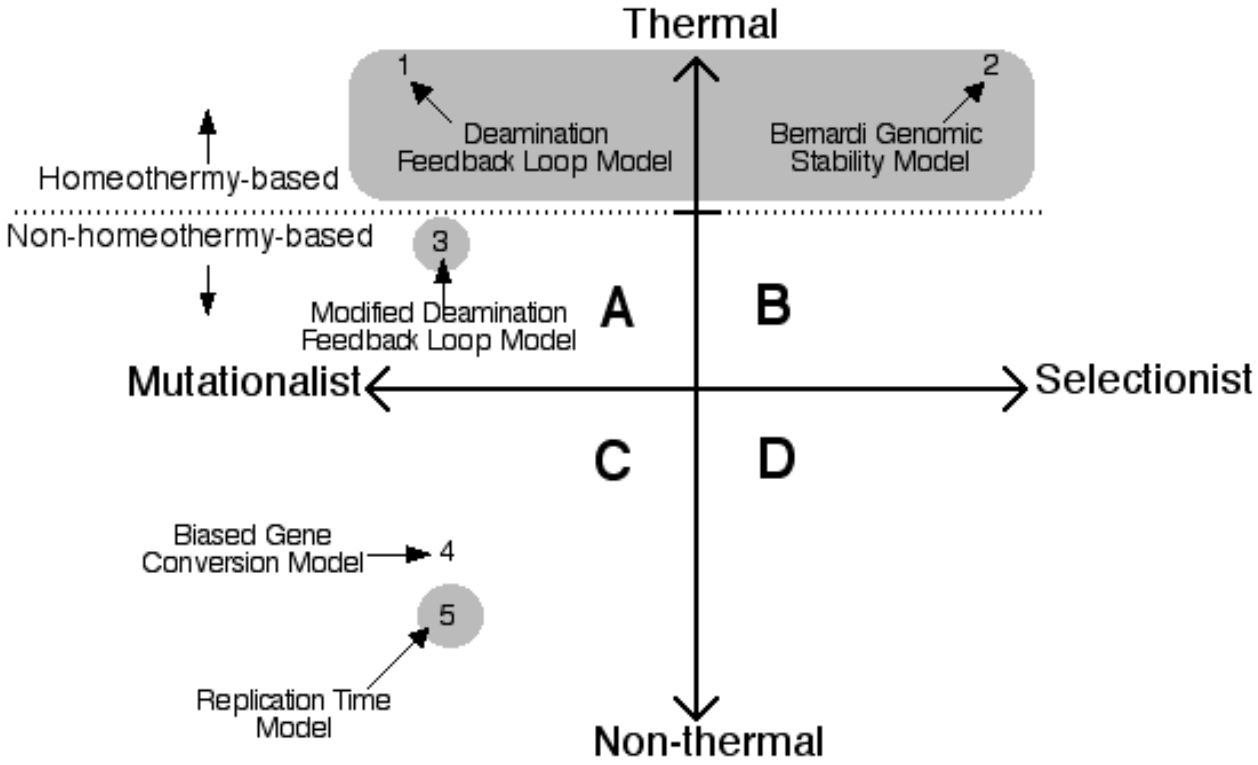


Figure 2-1. A unified framework for models of isochore evolution. The axes define mutationalist and selectionist models and emphasize that both types of models can be thermal or non-thermal in nature. Several proposed models are represented on the graph by a number in the relevant position (1. Fryxell and Zuckerkandl 2000; 2. Bernardi 2000; 3. proposed here; 4. Eyre-Walker 1999; 5. Wolfe et al.1989). This framework emphasizes that models of isochore evolution can be a continuum. The gray areas represent regions of "model space" that can be excluded, based either upon our data or Eyre-Walker (1992), who examined the replication time model (Wolfe et al. 1989). Rationale for excluding models is detailed in the discussion.

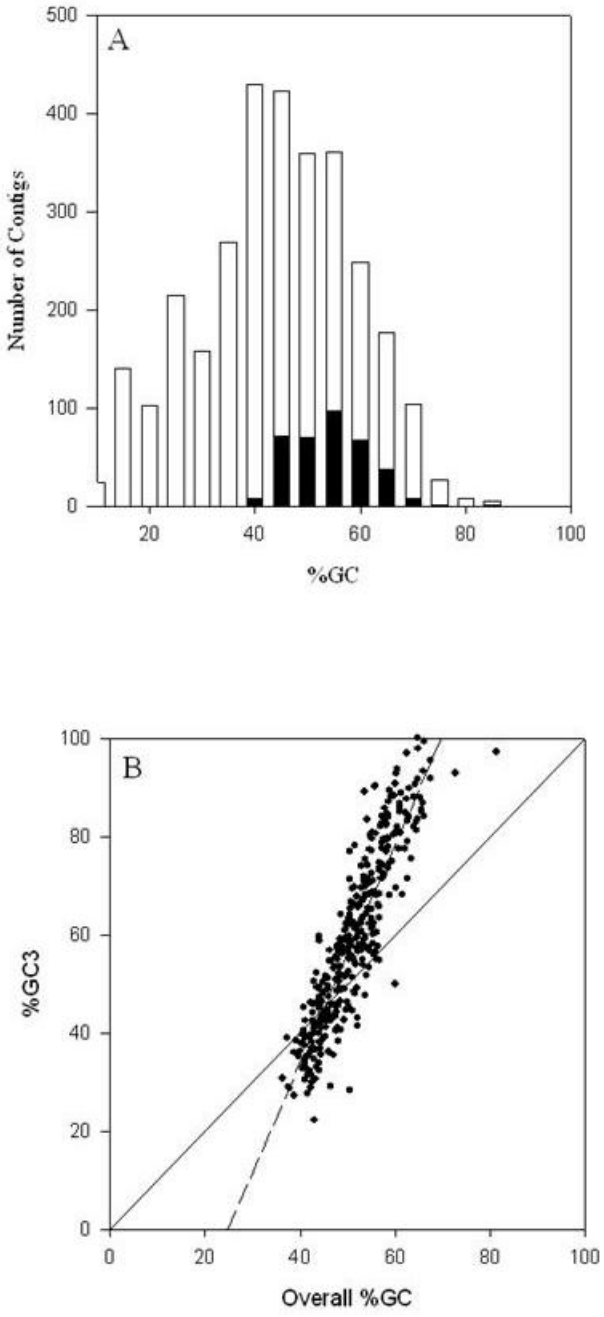


Figure 2-2. GC-content for the complete Alligator dataset. A) Histogram showing numbers of contigs (EST assemblies) in each GC-content bin with a width of 5%. The contigs that were included in the alignments are indicated in black. B) Correlation between GC3 and overall GC-content for the contigs that were included in the alignments.

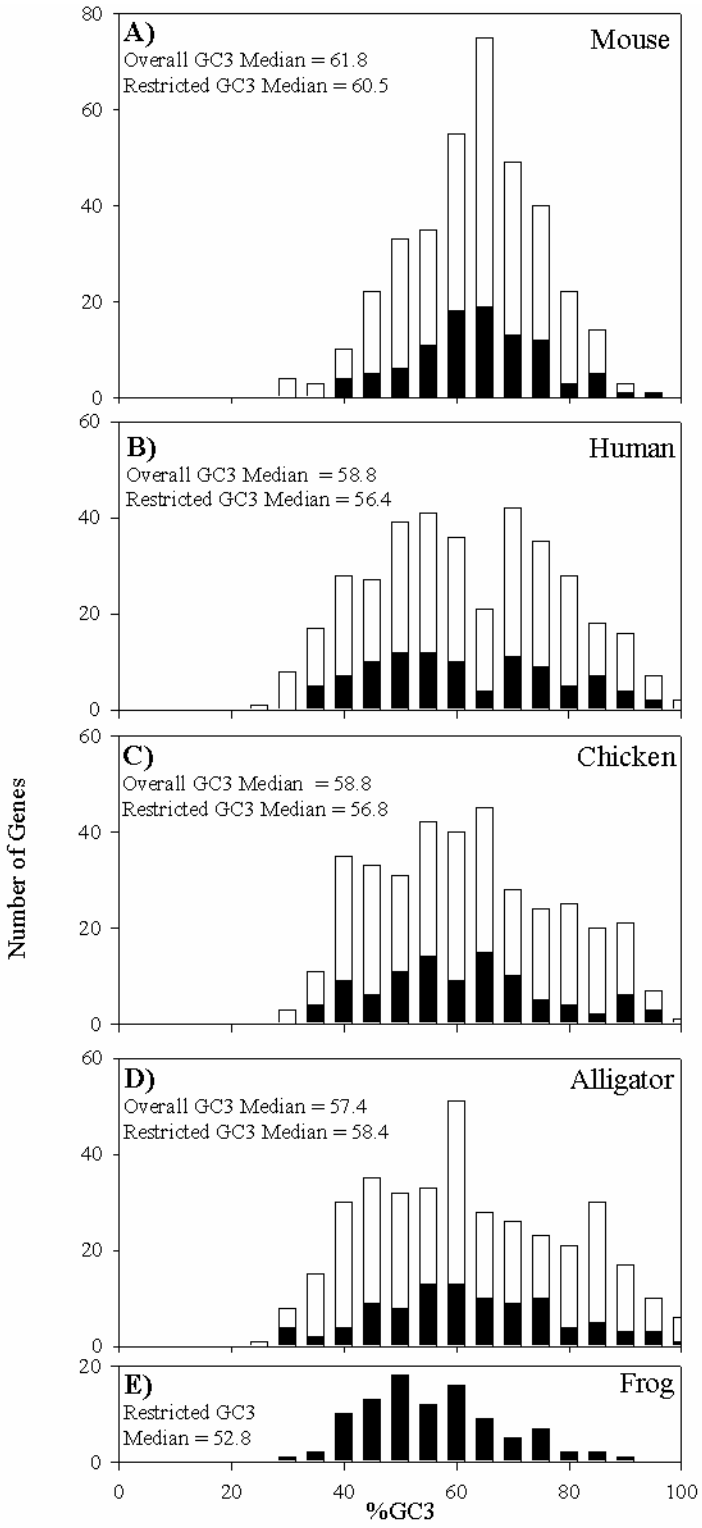


Figure 2-3. Histogram showing GC3 content. Genes were placed in bins with a width of 5%. Black bars indicate the subset of 98 alignments that include all four amniotes and the frog (the “restricted set”); while the white portions of the bars indicate the remaining 268 alignments that include the four amniotes (all 366 genes form the “overall set”).

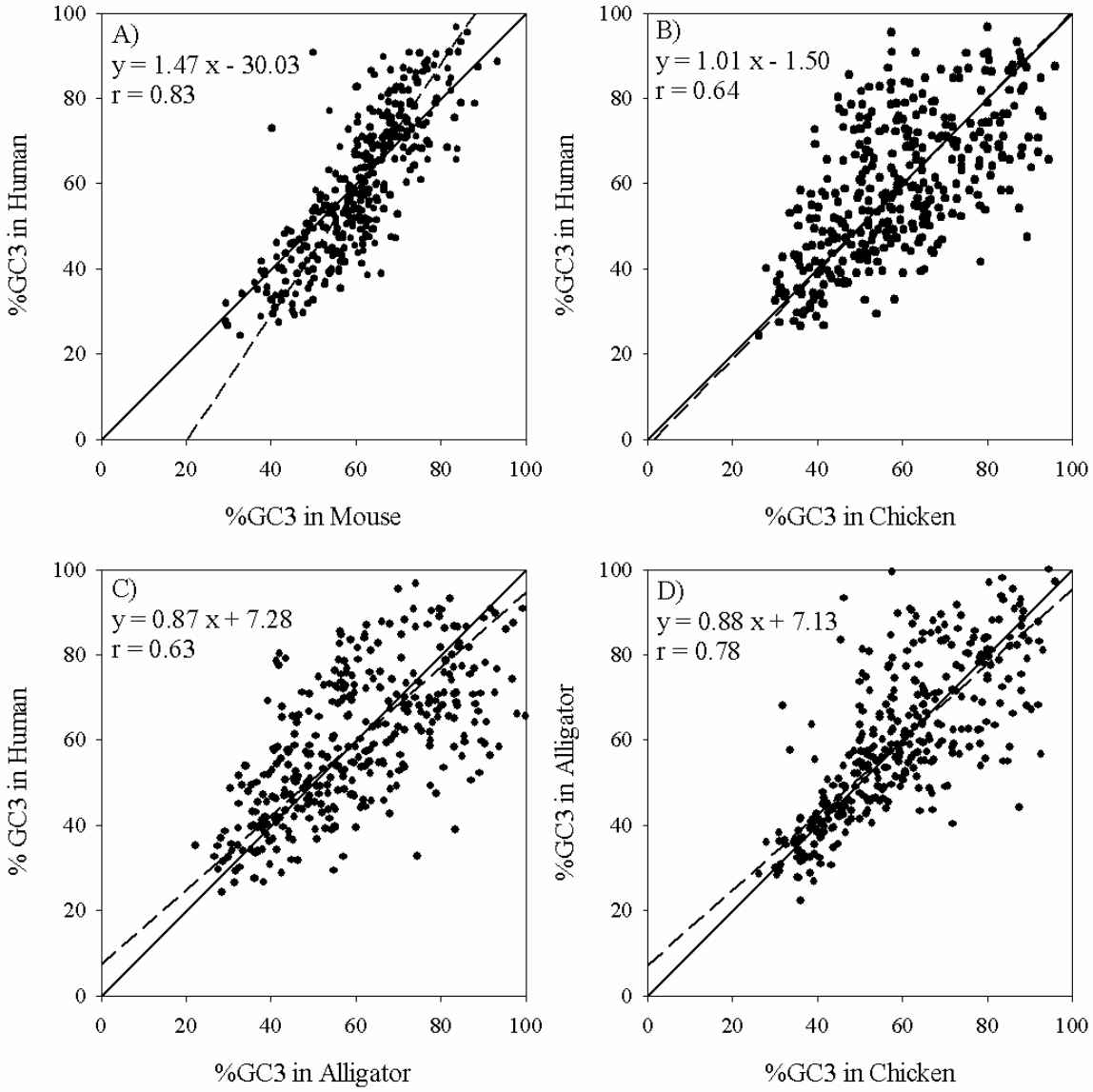


Figure 2-4. Correlations among amniotes in GC3 content. The graphs correspond to comparisons between A) human and mouse; B) human and chicken; C) human and alligator; D) chicken and alligator. The complete set of 366 genes was included. Comparisons within amniotes that are not presented here can be found in Appendix A, and confidence intervals for the slopes of lines relating GC3 for different organisms can be found in Appendix B.

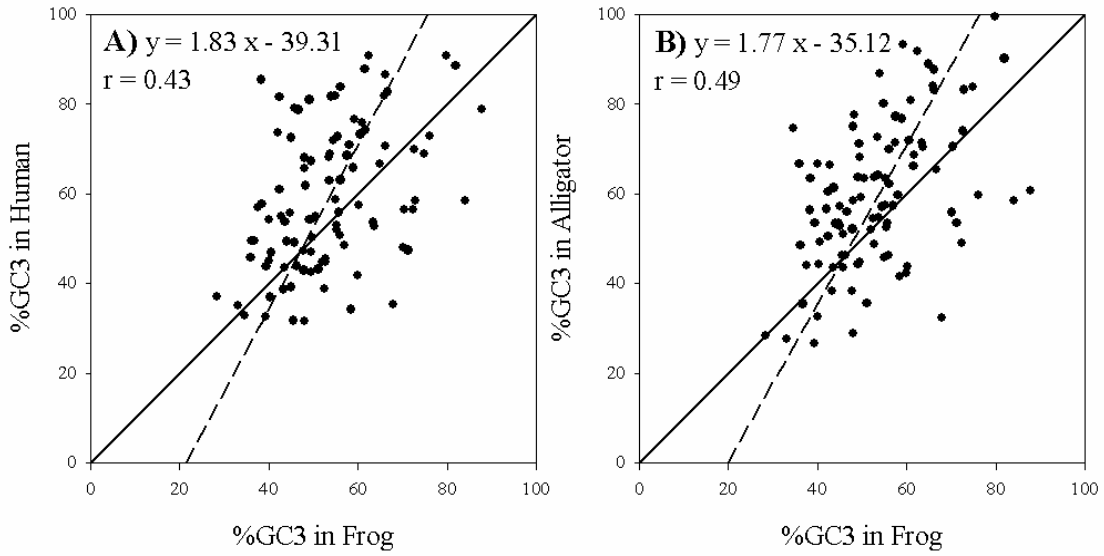


Figure 2-5. Correlations between an amphibian and amniotes in GC3 content. The graphs correspond to comparisons between A) human and frog; B) alligator and frog. The smaller 98 gene set was included here. Comparisons between the frog and amniotes that are not presented here can be found in the Appendix A, and confidence intervals for the slopes of lines relating GC3 for different organisms can be found in Appendix B.

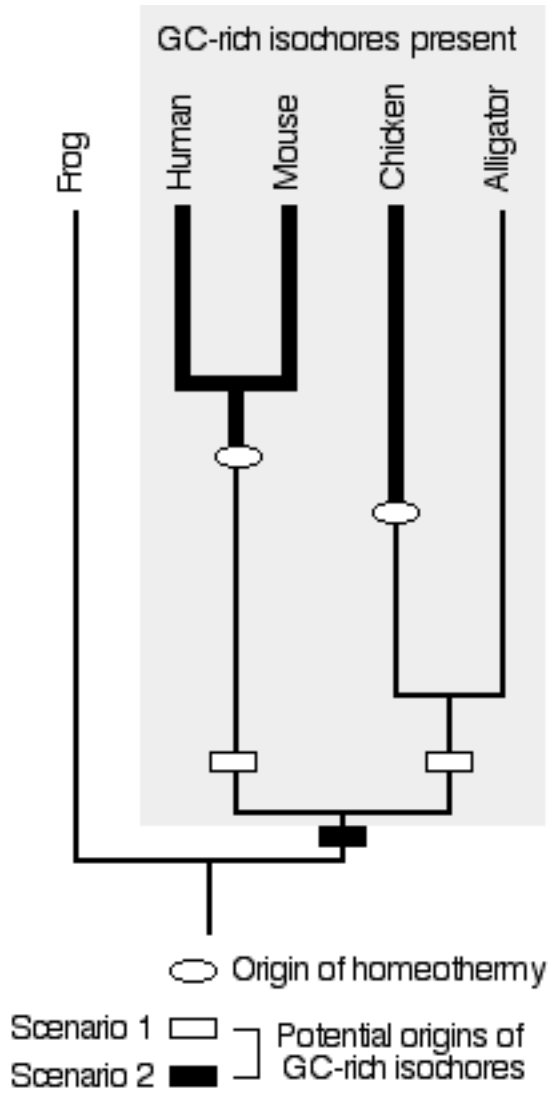


Figure 2-6. Strong GC-rich isochore structure appears to unite amniotes. Tetrapod phylogeny showing multiple origins of homeothermy and the distribution of strong GC-rich isochore structure (indicated using the gray region). The most parsimonious reconstruction of character states is a single origin of strong GC-rich isochore structure at the base of the amniotes (scenario 1). However, the limited taxon sampling available does not allow us to exclude two independent origins of strong GC-rich isochore structure (scenario 2), one in mammals and one in archosaurs (birds and crocodilians). Neither scenario is consistent with a correlation between the origin of homeothermy and the origin of GC-rich isochore structure. The ancestral GC-poor state for tetrapod isochores reflects information on amphibians (e.g., our data) and other outgroup taxa (e.g., Bernardi 2000). Branch lengths reflect divergence times based upon the fossil record (summarized in Benton and Donoghue 2007).

CHAPTER 3
TURTLE ISOCHORE STRUCTURE IS INTERMEDIATE BETWEEN AMPHIBIANS AND OTHER AMNIOTES

Introduction

Vertebrate genomes are mosaics of isochores, defined as long (>100 kilobases [kb]) regions of relatively homogenous base composition (either guanine-cytosine [GC]-rich or adenine-thymine [AT]-rich) that have sharp boundaries with neighboring regions (Bernardi 2000). One motivation for the study of isochore evolution is the observation that “warm-blooded” (hereafter homeothermic) mammals and birds exhibit substantial differences from “cold-blooded” (hereafter poikilothermic) amphibians and fish in heterogeneity of their genomic GC content (Bernardi 2000). Mammals and birds have isochores with as much as 60% GC and other genomic regions with as little as 30% GC, while amphibians and fish have more homogenous and often less GC-rich genomes. The best-characterized isochore structures include the human (Costantini et al. 2006), mouse (Zhang and Zhang 2004), and chicken (Costantini et al. 2007a; Costantini et al. 2007b; Gao and Zhang 2006) genomes, while amphibian (Fortes et al. 2007), crocodilian (Chojnowski et al. 2007), fish (Costantini et al. 2007a) and marsupial (Gu et al. 2007) genomes are somewhat less well-characterized. Many fundamental biological properties (e.g., gene density, recombination) are correlated with isochore structure (reviewed by Bernardi 2000; Eyre-Walker and Hurst 2001). In fact, the map of the human isochore revealed that the number and distribution of isochores is consistent with the patterns of chromosome banding during prophase (Costantini et al. 2006), emphasizing that isochores are fundamental to vertebrate genomic organization and evolution.

Models of isochore evolution have invoked either mutational biases (leading to patterns of neutral evolution that generate isochores) or natural selection (Bernardi 2007; Eyre-Walker and Hurst 2001). However, a second major division exists between models with a thermal basis and those that are non-thermal (Chojnowski et al. 2007). The most complete thermal hypotheses are also based upon selection, and they suggest that the evolution of GC-rich isochores is largely based on the greater stability of GC-rich regions at high temperature (Bernardi 2000; Bernardi 2007). However, there may also be selection for specific types of protein encoded by genes in GC-rich isochores, since those proteins differ structurally from proteins encoded by genes in GC-poor isochores (Chiusano et al. 1999; D'Onofrio et al. 1999). This can be called the genomic stability hypothesis as it is largely based upon the notion that GC-rich isochores are an adaptation to constantly high body temperatures. The deamination feedback loop (Fryxell and Zuckerkandl 2000) is a neutral (mutational) thermal hypothesis, since it postulates that GC-poor regions have a higher rate of cytosine deamination (which results in C→T transitions) at high temperatures, further reducing their GC content. However, the deamination feedback loop predicts a unidirectional bias that explains homeothermic isochore structures only if the ancestral condition was GC-rich, which is not suggested by the relatively homogeneous GC-poor genomes in amphibians and fishes. The observation that the poikilothermic American alligator has a GC-rich isochore structure similar to that of mammals and birds (Chojnowski et al. 2007), adds additional emphasis that GC-rich isochores cannot be an adaptation to homeothermy *per se*. However, Chojnowski et al. (2007) pointed out that some poikilotherms maintain a body temperature similar to that of many homeotherms for

relatively long periods (Seebacher and Shine 2004), and suggested that thermal models of isochore evolution should consider this fact.

A model unrelated to thermal factors and based upon natural selection is the "epigenetic optimization hypothesis" of Vinogradov (2005), which postulates that base composition is optimized for active transcription of GC-rich regions and suppression of GC-poor regions. It is unclear, however, how the epigenetic optimization hypothesis explains the transition from the more homogeneous genomes of amphibians and fish to the more heterogeneous mammalian and archosaurian genomes. One non-thermal mutational hypothesis invokes distinct biases for regions with different replication times; early and late replicating sequences are proposed to exhibit different patterns of mutation that reflect changes in the free nucleotide pools (Wolfe et al. 1989). Although one study reported that isochore structure and replication time are uncorrelated in somatic mammalian cells (Eyre-Walker 1992), replication times can change during differentiation (Hiratani et al. 2004) and other studies have found a good correlation (Schmegner et al. 2007). Different patterns of deoxyribonucleic acid (DNA) repair might also result in changes to the overall base composition (Boulikas 1992), although it is unclear whether these biases extend over the length of typical isochore structure. However, the mutational model invoked most commonly is biased gene conversion (Duret et al. 2006; Li et al. 2007), a process linked to recombination that results in a higher probability of GC alleles converting AT alleles than is true of the reverse.

Despite their potentially critical role for understanding isochore evolution, there is a surprising lack of information about reptilian isochore structure. There are three major lineages of living reptiles: lepidosaurs (lizards, snakes and tuatara), testudines (turtles),

and archosaurs (crocodilians and birds) (Hugall et al. 2007; Iwabe et al. 2005; Rest et al. 2003). This emphasizes that it is inappropriate to consider reptiles homogeneous. In fact, reptiles exhibit diverse body sizes, metabolic rates and patterns of thermoregulation (Zug et al. 2001), so it is important to learn more about reptilian isochore structures to examine the relationship between GC content and thermal biology. The large-scale study by Chojnowski et al. (2007) extended and strongly supported earlier studies based upon a few genes (Hughes et al. 1999) by demonstrating that crocodilians and birds have similar isochore structures. Thus, the GC-rich isochore structure of birds is actually an archosaurian phenomenon. Archosaurs, in turn, have an isochore structure similar to that of mammals. It remains unclear, however, whether GC-rich isochore structure is a feature of amniote genomes or the product of convergent evolution between archosaurs and mammals. To define the isochore structure of another reptilian lineage, we examined the GC content of multiple expressed genes in the Red-eared slider turtle (*Trachemys scripta*) and compared this to results from organisms with known isochore structures.

Methods

Collection and Assembly of Expressed Sequence Tags (ESTs)

Three subtraction libraries of turtles were generated using the suppression subtraction hybridization method (Diatchenko et al. 1996) with the polymerase chain reaction (PCR)-Select™ cDNA Subtraction Kit (Clontech, Mountain View, CA). The source of the messengerRNAs (mRNAs) used to generate the libraries was whole turtle embryos that were incubated at different temperatures or with estradiol (manuscript in preparation; for additional details regarding the library, contact J.L.C.). The use of suppression subtraction hybridization is expected to enrich the library for conditionally

expressed transcripts relative to housekeeping genes. A total of 1983 ESTs (deposited in database of expressed sequence tags [dbEST] with accession numbers FG341000 to FG341832), were obtained from the turtle subtraction cDNA libraries using an Applied Biosystems (ABI 3100 Avant) automated sequencer. ESTs were assembled as described by Liang et al. (2000).

EST assemblies were used as Nucleotide 6-frame translation-nucleotide 6-frame translation Basic Local Alignment Search Tool (tBLASTx) (Altschul et al. 1997) queries to search cDNA sequences from the EnsEMBL database (Birney et al. 2006) and the alligator EST assemblies from Chojnowski et al. (2007). The EnsEMBL sequences were annotations of the genomic sequences of the human, mouse, dog, gray short-tailed opossum, chicken and western clawed frog and the set of these sequences that were identified by the turtle query were aligned using Clustal W (Chenna et al. 2003). Alignments that appeared largely anomalous, due to issues such as misannotations were discarded, as were individual sequences that appeared unlikely to be orthologs of the other sequences in the alignment or which exhibited anomalies. This increased the overall number of sequences analyzed but resulted in an unequal representation of species for each gene. Between 192 and 274 genes were included from each EnsEMBL species while 59 alligator genes were included from the dataset of Chojnowski et al. (2007). Alignments were optimized by eye using MacClade (Maddison and Maddison 2002) and the alignment segment present in all organisms (after anomalous sequences were discarded) was assigned to a character set (CHARSET). Codon positions were then identified and pylogenetic analysis using parsimony* (PAUP*) 4.0b10 (Swofford 2003) was used to calculate third-position base composition. Base composition data

were extracted from the PAUP logfile using a shell script and imported into Microsoft Excel.

3rd Codon Position (GC3) Analyses

Strong correlations between GC3 of protein-coding genes, their introns, and their surrounding DNA (the isochores in which genes are embedded) have been found in both poikilotherms (Bernardi and Bernardi 1991) and homeotherms (Bernardi 2000; Musto et al. 1999). Thus, GC3 values for orthologous coding sequences were used as a surrogate to compare turtle isochore structure to that in other organisms, as in other studies (e.g., Bernardi et al. 1997; Galtier and Mouchiroud 1998; e.g., Zoubak et al. 1996). The open-source R statistical software (R Development Core Team 2007) was used to fit linear equations to the data using orthogonal regression (Isobe et al. 1990) and to calculate correlations. Lines for the best-fitting equation and for unity (y = x) are included in all plots.

Results and Discussion

Types of Genes Analyzed

The genes obtained from the three embryonic turtle-subtraction libraries and their orthologs in other organisms exhibited a striking difference in GC3 from the set of genes obtained from a standard library (neither normalized nor subtracted) examined by Chojnowski et al. (2007). Median GC3 values from all organisms were approximately 10% lower than in that study (Figure 3-1) and the frog fell within the diversity of values for amniotes (albeit toward the lower end). This can be explained on the basis of the differences between the types of genes we expect to sample from the two different libraries, since there is strong evidence that the GC content of the genes is highly reflective of their type (e.g., housekeeping, specialized). The set of genes housed in

GC-rich regions is greatly enriched for house-keeping genes and other highly expressed genes (Arhondakis et al. 2008; Kudla et al. 2006). Therefore, genes from GC-rich regions will be overrepresented in a standard library and genes from GC-poor regions are likely to be underrepresented. The opposite is expected for a subtracted library, where highly expressed genes are removed.

To determine whether the difference in the types of genes represented in the libraries is a reasonable explanation for the lower overall median GC3 of the turtle ESTs, we calculated the median GC3 after restricting our analysis to the set of genes that are also present in the alligator (called “restricted” in Figure 3-1). As we predicted, the restricted median GC3 values were almost identical to those reported in our previous study (Chojnowski et al. 2007). Based upon the restricted median GC3 content, the turtle was the lowest of any amniote, indicating that the isochore structure of the turtle is intermediate between that of the frog and that seen in other amniotes.

The Distribution of GC3 across Organisms

The distribution of GC3 values, regardless of whether we consider the full set or the restricted set of sequences, also reveals substantial variation that is not captured by the median (Figure 3-2). For example, the restricted median GC3 values for the frog and turtle differ by <1% but the proportion of genes in the highest GC3 category (90 - 100%) is different. In fact, the turtle GC3 distribution has a much greater skewness than does that of the frog (data not shown). These data are consistent with previous results (e.g., Bernardi 2000), indicating that amphibian genomes have a lower GC content than do amniote genomes. However, Bernardi (2000) also suggest that third-codon-position GC-content distributions for genes in GC-poor regions, likely to have been enriched in our turtle libraries, show extensive overlap in different organisms.

Turtle Isochore Structure is Intermediate between those of Amphibians and Other Amniotes

The relationship between the GC3 values for different organisms can be examined using orthogonal regression (Chojnowski et al. 2007), and the slopes of lines obtained in this way reveal the degree of similarity between the isochore structures of the organisms analyzed. A slope of unity would indicate identical isochore structure (at least from the standpoint of GC3 content) and slopes with larger deviations from unity also correspond to organisms with greater differences in their isochore structures. Remarkably, the slopes of lines describing the relationships between organisms were very similar to those in our previous study (Chojnowski et al. 2007), despite using all set of turtle sequences rather than the restricted set. Thus, the genes sampled from the subtraction library do not differ from the genes sampled by Chojnowski et al. (2007) from the standpoint of the slope of these comparisons. In fact, comparisons between the turtle and various mammals revealed slopes that were both lower than unity and lower than in comparison between either turtle and archosaurs or between turtles and frogs (Figure 3-3). Although the slopes of the comparisons between turtles and archosaurs were less than unity, the slope of those between frogs and turtles exceeds one. These results are consistent with a model in which the isochore structure of the turtle is intermediate between that of the frog and that of other amniotes.

The correlation coefficient provides another way of examining the relationship between the isochore structures of two different organisms, since the number of genes that undergo changes in GC3 content may differ in distinct lineages. The turtle shows a higher correlation with the archosaurs than with any other organisms (Table 3-1). The lowest correlation coefficients involved comparisons with the frog, consistent with the

divergent phylogenetic position of this organism (Figure 3-4). However, there was also evidence for complex patterns of change in amniote isochore structures. For example, comparisons involving the opossum also show consistently low correlation coefficients (Table 3-1), with the exception of the comparison between opossums and mice. This is consistent with other studies showing a decrease in the GC-content of the opossum (Gu and Li 2006; Gu et al. 2007). Comparisons involving the opossum reveal a pattern that contrasts sharply with those involving the turtle. The latter, especially the comparisons between turtles and archosaus, have high correlation coefficients but slopes that are less than unity. This suggests that turtle genes and archosaur genes tend to be similar in categories of GC3-content although the GC-content of those categories differs, especially for the most GC-rich genes. Thus, few genes in the turtle-archosaur clades have undergone changes in their GC-content category but the slopes of comparisons between turtles and archosaurs indicate that the most GC-rich genes have increased their GC-content in archosaurs, decreased their GC-content in turtles, or that both changes have occurred.

A Phylogenetic Framework for Reptilian Isochore Evolution

A single origin for GC-rich isochore structure at the base of the amniotes was the most parsimonious interpretation of the data on alligator isochore structure provided by Chojnowski et al. (2007). This “amniote hypothesis” was suggested by a previous study that used a smaller set of turtle genes and crocodilian genes (Hughes et al. 1999), and it was embraced by studies that advocated non-thermal hypotheses for the origin of GC-rich isochores (Duret et al. 2002). The amniote hypothesis is inconsistent with the thermal models since they predict two independent origins of GC-rich isochore structure coincident with the origin of homeothermy in birds and mammals (Bernardi 2000;

Bernardi 2007). A simple convergent origin of GC-rich isochore structure in mammals and birds is excluded by the existence of a GC-rich isochore structure in alligators (Chojnowski et al. 2007). The analyses presented here indicate that the isochore structure of turtles is intermediate between that of frogs and the GC-rich isochore structures of archosaurs and mammals. Thus, there are two hypotheses regarding the evolution of GC-rich isochores in amniotes, a convergent origin in mammals and archosaurs (combined with a modest increase in the GC content of the turtle genome) or an origin in the common ancestor of amniotes along with the loss of this GC-rich structure in turtles.

Chojnowski et al. (2007) suggested a modified thermal hypothesis, the “non-homeothermic thermal hypothesis”, that is consistent with the GC-rich isochore structure of the alligator. This hypothesis postulates that GC-rich isochores reflect the maintenance of relatively high body temperatures either due to endothermy (in mammals and birds) or behavior and thermal inertial (in crocodilians) (see Seebacher et al. 1999). Alternatively, extant crocodilians may have had homeothermic ancestors (Seymour et al. 2004). If this was the case, the GC-rich isochore structure of the alligator may be retained from an ancestral condition. However, there are a number of arguments against the hypothesis that ancestors of extant crocodilians were homeotherms (Hillenius and Ruben 2004), and we contend that behavioral thermoregulation is sufficient to explain the isochore structure of crocodilians if GC-rich isochore structure is related to thermal factors.

This study is consistent with a modified (non-homeothermic) thermal hypothesis, since most turtles have less thermal inertia than do crocodilians. Indeed, the typical

body temperature of the turtle species used in this study is lower than typical body temperatures of crocodilians (Gatten 1974; Seebacher et al. 1999), making it reasonable to postulate that turtles would have an isochore structure intermediate between those of amphibians and archosaurs were the thermal hypothesis correct. However, turtle thermoregulation is complex (e.g., Thomas et al. 1999) and it is likely to have undergone multiple changes over evolutionary times. Given the challenges associated with modeling behavioral thermoregulation for extant organisms (e.g., Christian et al. 2006) it will probably be very difficult to develop successful models able to be applied to the evolutionary history of extant organisms. Despite the challenges, it may be desirable to extend these analyses beyond the examination of reptiles to include the examination of isochore structures in homeothermic lineages that undergo torpor or hibernation, to determine whether the GC content underwent a decrease correlated with their fluctuating body temperatures.

Regardless of the specific relationship between isochore structure and thermal biology, or even whether any relationship of this type exists, expanding the amount of sequence data available for reptiles will be helpful. Varriale and Bernardi (2006) measured the GC-contents of reptilian genomes using HPLC and analytical ultracentrifugation and they found only a modest difference between turtles and crocodilians. Shedlock et al. (2007) also reported similarities between the genomes of turtles and archosaurs. Varriale and Bernardi (2006) also noted substantial variation within lizards and snakes, suggesting that detailed analyses of the genomes of lizards and snakes will prove fruitful. A draft genome sequence for the lizard *Anolis carolinensis* will soon be available, so it will be possible to conduct detailed analyses of isochore

structure for that organism. Our observation that there is a continuum of isochore structures within the reptiles indicates that exploring the variation in isochore structures within reptiles further will prove illuminating.

Conclusions

These analyses of turtle EST data, especially when combined with previous analyses of sequence data from reptiles (Chojnowski et al. 2007; Fortes et al. 2007; Hamada et al. 2002; Hughes et al. 1999), provides surprising information about the variance in the isochore structures of reptilian genomes. The turtle appears to have a relatively GC-rich isochore structure, stronger than the isochore structure of the frog but weaker than that of mammals or archosaurs. Our previous work emphasized the need to consider a complete set of models, which we defined based upon whether the models invoke patterns of mutation or selection and whether or not they involved thermal factors. If the complete set of models that are plausible *a priori* are considered, it should be possible to further constrain the credible set of models for isochore evolution or to provide the novel information necessary to define additional models that are better able to explain isochore evolution.

Table 3-1. Slopes and correlation coefficients for all combinations of organisms.

	Human	*Mouse*	*Dog*	*Opossum*	*Chicken*	*Alligator*	*Turtle*	*Frog*
Human	--	0.83	1.00	1.00	0.80	0.90	0.62	0.38
Mouse	0.69	--	1.19	1.30	0.86	0.19	0.69	0.47
Dog	0.92	0.71	--	1.00	0.82	0.93	0.61	0.41
Opossum	0.49	0.75	0.51	--	0.75	1.35	0.55	0.33
Chicken	0.61	0.55	0.66	0.50	--	1.04	0.73	0.52
Alligator	0.67	0.39	0.61	0.42	0.72	--	0.80	0.43
Turtle	0.66	0.56	0.65	0.59	0.74	0.76	--	0.82
Frog	0.40	0.40	0.43	0.36	0.52	0.63	0.47	--

The upper right cells are slopes and the taxon in the leftmost column are on the x-axis. The lower left cells are correlation coefficients.

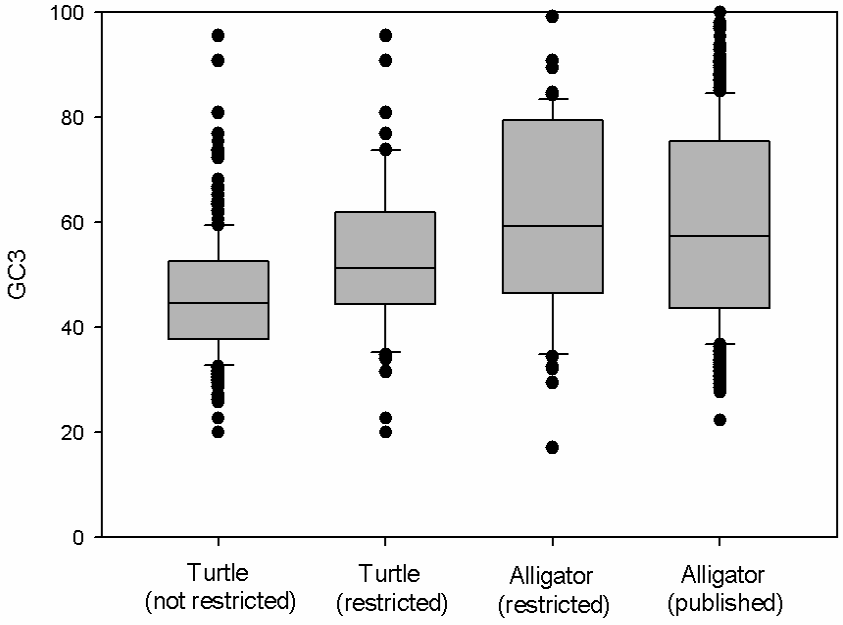


Figure 3-1. Third-codon-position GC content of turtle genes and alligator genes. This boxplot shows GC3 values for the complete set of turtle EST assemblies generated using the subtraction libraries and the "restricted" subset of those EST assemblies that have orthologs in the alligator libraries used by Chojnowski et al. (2007). For comparison, the complete set of alligator EST assemblies analyzed by Chojnowski et al. (2007) and the subset of the alligator EST assemblies that have turtle orthologs are shown.

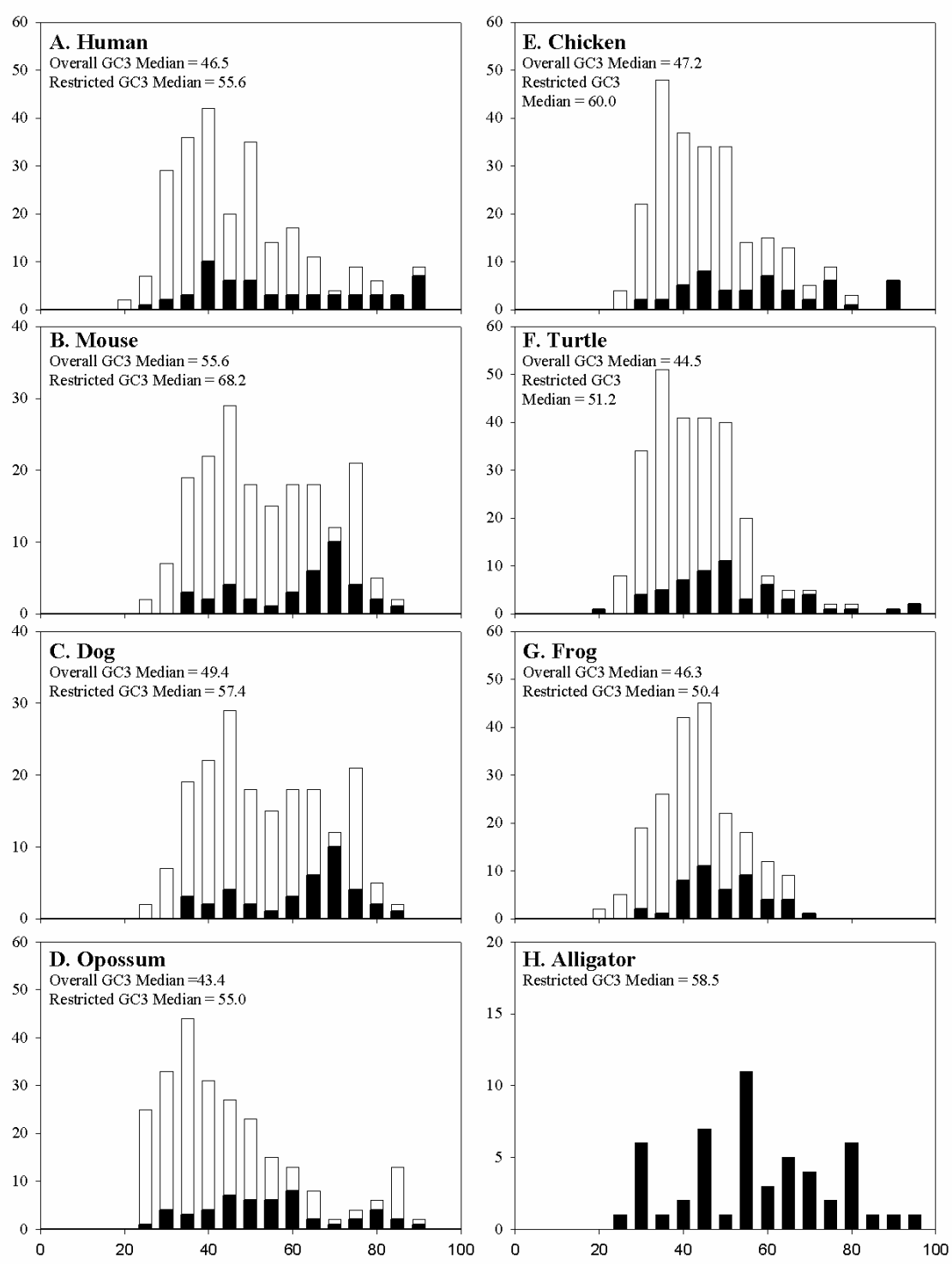


Figure 3-2. GC3 content of the focal genes in different vertebrate lineages. Genes were placed in bins with a width of 5%. Black bars indicate the subset of 59 alignments that include the alligator (the “restricted set”); the white portions of the bars indicate the remaining 192 to 274 alignments (each organism has a different number of alignments).

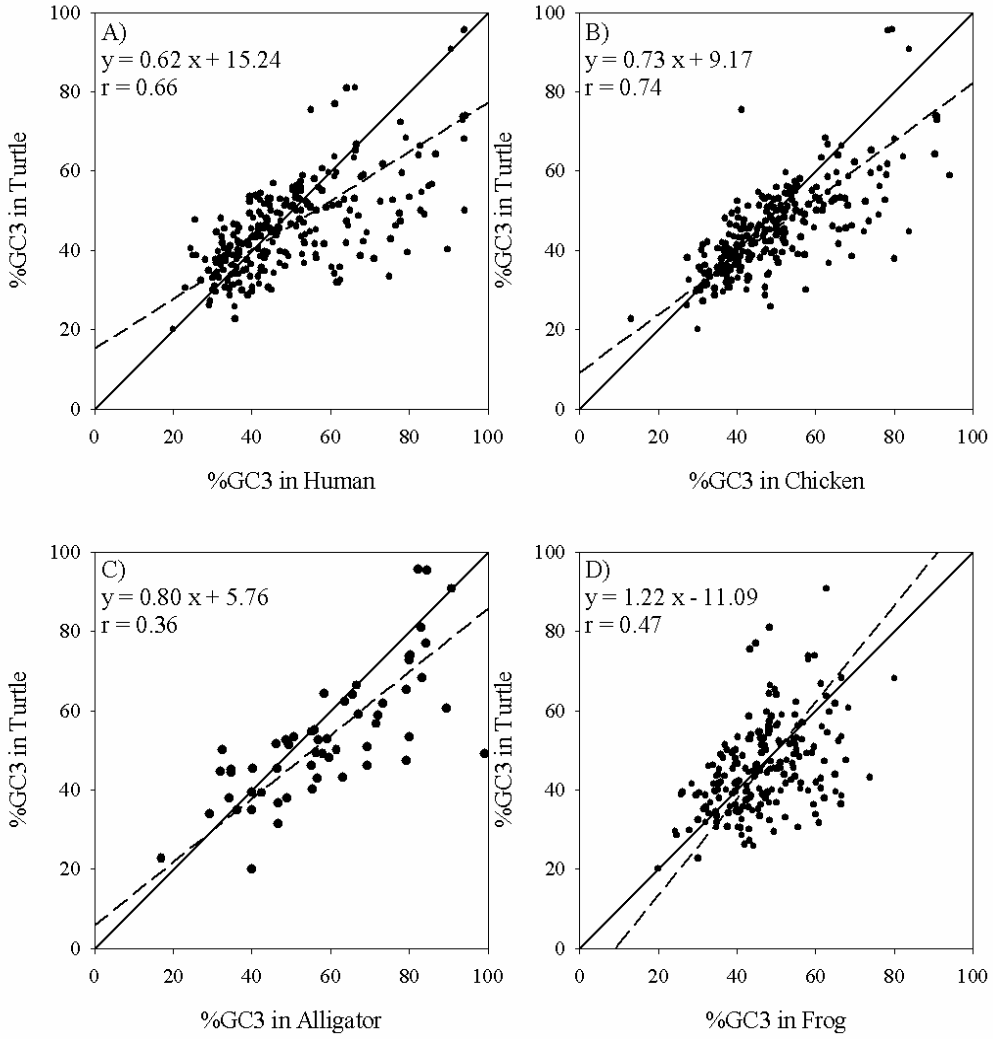


Figure 3-3. GC3 content of turtle genes is strongly correlated with the GC3 content in other organisms. The graphs correspond to comparisons between A) turtles and humans; B) turtles and chickens; C) turtles and alligators; D) turtles and frogs. The complete set of genes between turtles and each other organism was included (255, 261, 59, and 211, respectively). Comparisons of turtles with mice, dogs, and opossums are not shown but their graphs were very similar to the comparison between turtles and humans. Slopes from orthogonal regression and correlation coefficients for all comparisons can be found in Table 1.

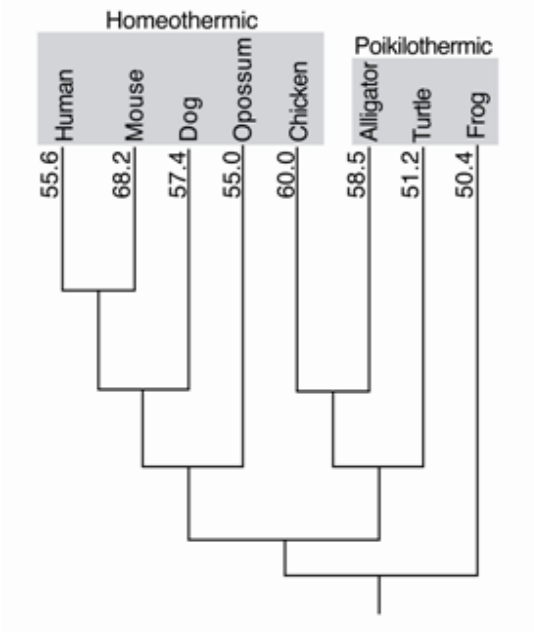


Figure 3-4. Phylogeny of organisms used in this study. Numbers are median GC3 values for each organism. Patterns of thermoregulation (homeothermy or poikilothermy) are indicated. Lengths of branches are arbitrary.

CHAPTER 4
IDENTIFICATION OF GENES SHOWING SEXUALLY DIMORPHIC EXPRESSION IN A TURTLE WITH TEMPERATURE-DEPENDENT SEX DETERMINATION

Introduction

Many reptilian taxa, including the majority of turtle species, exhibit temperature-dependent sex determination (TSD) (Crews et al. 1994; Janzen and Phillips 2006). Incubation temperature is the initial cue for sexual development in TSD, in sharp contrast to genetic sex determination (GSD) that is evident in a number of vertebrate groups such as amphibians, snakes, birds, and mammals (reviewed in Ezaz et al. 2006). GSD is best characterized in therian mammals and is initiated by the *Sex-determining region Y* (*SRY*) gene located on the Y chromosome that causes organisms expressing the gene to develop as males (Sinclair et al. 1990). However, *SRY* orthologs have not been identified in other groups of vertebrates regardless of whether they express TSD or GSD, suggesting that *SRY* is an innovation unique to therian mammals. In fact, a "trigger" gene for sexual development has only been identified in one other vertebrate taxon, a fish with GSD (medaka; see Volff et al. 2003; Matsuda et al. 2007). Indeed, it is unclear whether a trigger gene exists in organisms that exhibit TSD, since there are several models that can explain TSD. For example, TSD may reflect regulation of a trigger gene (or set of trigger genes) by incubation temperature, it may reflect the impact of temperature upon the activity of specific enzymes that have a role in signaling, or it may reflect a combination of both phenomena (Ramsey and Crews 2007; Shoemaker et al. 2007). Regardless, it is clear that TSD in turtles and other organisms is unlikely to be regulated by a gene homologous to a known trigger.

Although there is a lack of conservation for the trigger gene(s), a number of genes involved in gonadal differentiation and other aspects of sexual development have been

found to be conserved among vertebrates, including differing sex determining systems (Johnston et al. 1995; Ramsey and Crews 2009; Western and Sinclair 2001). A number of orthologs of genes first identified in mammals have been identified and characterized in different vertebrate groups (Guan, Kobayashi, and Nagahama 2000; Shibata, Takase, and Nakamura 2002; Clinton 1998; Smith and Sinclair 2001; Western et al 2000), including turtles (Table 4-1). Studies focused on orthologs of known genes have provided valuable information, although this approach has limits. The complete set of genes involved in mammalian GSD remains unknown. Furthermore, it seems clear that it will not reveal the trigger gene(s), if any, at the top of the regulatory cascade. A complementary approach is to identify candidate genes for the turtle TSD cascade by identifying genes with specific patterns of gene expression.

This study used the red-eared slider turtle (*Trachemys scripta*), a turtle with TSD, to identify genes that exhibit sexually dimorphic expression during the temperature-sensitive period (TSP), which contains the critical stage for commitment to a specific sex. Red-eared slider turtles produce only females when eggs are incubated at 31degrees Celsius (°C) and only males when eggs are incubated at 26°C. They have a population-wide 1:1 sex ratio at 29.2°C, called the pivotal temperature for this species (Crews et al. 1994, Wibbels et al. 1991). It is unclear whether the development of different sexes at the same temperature reflects genetic variation in study populations, stochastic aspects of gene expression, maternal effects, or some combination of these factors. Temperature acts to establish sex during the TSP, which begins near stage 14 (using the developmental stages described by Yntema [1968]) and extends through stages 19 or 20 (the TSP ends slightly earlier at higher temperatures). Switching red-

eared slider turtle eggs between warmer and cooler temperatures during the TSP results in sex reversal (Wibbels et al. 1991). Sex reversal can also be elicited by the administration of exogenous estrogen (generating females) or nonaromatizable androgen (generating males) during incubation. In fact, exposing eggs incubated at 26°C to exogenous estrogen before the TSP will result in the production of 100% females (Crews et al. 1991). These observations regarding the sensitivity of developing red-eared slider turtles to temperature and steroid hormones make them well-suited for further examination of differential gene expression during the TSP.

The goal of this study is to identify genes that exhibit sexually dimorphic expression in the red-eared slider turtle during the TSP. To complement this search we also identified genes that show increased messenger ribonucleic acid (mRNA) accumulation in response to estrogen exposure and stage effect within the TSP. To accomplish this, we produced three subtraction libraries. Two of these libraries were enriched for genes that show higher mRNA accumulation during the TSP in one specific temperature regime (i.e., genes that exhibit greater mRNA accumulation at the female-producing temperature [31°C] than at the male-producing temperature [26°C] and vice versa). The third library was enriched for genes that show increased mRNA accumulation during the TSP in sex-reversed embryos produced at the male-producing temperature after treatment with exogenous estrogen. Subsets of the complimentary deoxyribonucleic acids (cDNAs) from these libraries were examined more thoroughly by macroarray hybridization and semi-quantitative polymerase chain reaction (semiQ-PCR) or quantitative real-time PCR (qRT-PCR). This approach has the potential to identify

genes that exhibit sexually dimorphic expression during the TSP in the red-eared slider turtle without making a priori assumptions about the identity of the genes.

Methods

Experimental set-up

Freshly laid *Trachemys scripta* eggs (500) were purchased from Kliebert Turtle Farms in Hammond, Louisiana in 2004 and 2006. They were kept at room temperature for less than 48 hours until they were established as viable by candling. Those viable were randomly separated equally into four experimental groups in containers with moistened vermiculite (1:1 vermiculite to water). The experimental groups were 31°C (female), 26°C (male), 26°C painted with exogenous estradiol-17β (E_2) in 1microgram per microliter (µg/µl) in 95% ethanol (non-denatured) (female), and 26°C painted with exogenous 95% ethanol (non-denatured) as the vehicle control (male). Application of E_2 and vehicle occurred at stage 14. The egg boxes were rotated daily within the incubators and a random selection of eggs were checked periodically for developmental stage according to Yntema's staging guidelines (1968). The temperature was monitored daily with HOBO data loggers and supplemented with in-incubator thermometers. Sex was determined for each experimental group through a visual inspection by two independent researchers at hatching (gonads are visually distinct at hatching but not before) of 10 embryos per experimental group for relevant gross anatomy by two independent researchers.

Isolation of RNA

Whole embryos were taken between stages 17 and 20 from each experimental group and quickly frozen in liquid nitrogen and stored at -80°C. A subset of whole embryos was also collected at time = 0, 6, and 24 hours after E_2 and vehicle application

at stage 14. Conducting the E_2 time trial at stage 14 before the TSP removes E_2 effects within gonadal differentiation and leaves just those from the trial. Total RNA was extracted for each embryo by homogenization in TRIzol, chloroform extraction, and isopropanol precipitation according to Sambrook and Russell (2001). Total RNA yield and quality were assessed with the ND-1000 Nanodrop spectrophotometer (NanoDrop Technologies, Thermo Fisher Scientific, Wilmington, DE 19810, USA), and the integrity was verified by running samples on a 1% agarose gel.

Suppression subtractive hybridization (SSH)

Three libraries were selectively induced for female against male, male against female, and E_2 against vehicle. Testers and drivers were made from pooled RNA from stages 17-20 (two individuals per stage) from each experimental group from 2004 (stated above). cDNA synthesis was performed with the BD SMART™ PCR cDNA Synthesis Kit (Clontech, Mountain View, CA) according to the manufacturer's protocol. Three subtraction libraries were constructed with the Clontech PCR–Select™ cDNA Subtraction Kit (Clontech, Mountain View, CA) according to the manufacturer's protocol except a PolyEthylene Glycol (PEG) precipitation followed by an ethanol wash was used to purify the PCR products after cDNA synthesis instead of the column chromatography. The resulting cDNA was ligated into a pGEM®-T Easy vector and transformed into electrocompetent cells (Lucigen) by the manufacturer's protocol. Individual colonies were picked and stored in 96-well plates with 50% glycerol at -80°C. Plasmid inserts were purified using a modified 96-well Perfectprep® Plasmid protocol (5Prime), according to Sambrook and Russell (2001) or a TempliPhi Amplification kit (as recommended by manufacturer; Amersham Biosciences). Single-pass sequencing was

conducted on an ABI Prism™ 3100-Avant genetic analyzer (PE Applied Biosystems) using the ABI BigDye® Terminator v.3.1 chemistry.

Analysis of SSH results

Sequences from the libraries with redundancy were aligned and all sequences (both individual and aligned) were edited in Sequencher™ 4.1 (Gene Codes Corp.). Sequences were put into FASTA format and run in GOanna from the AgBase v.2.0 database to determine the top Basic Local Alignment Search Tool (BLAST) hit for each sequence and to simultaneously determine Gene Onotology (GO) terms for each hit (McCarthy et al. 2006). GOanna uses protein-protein BLAST (BLASTX: determines gene products from sequences) therefore any sequences that did not have a hit were run through nucleotide-nucleotide BLAST (BLASTN: determines all aspects of RNA transcripts including untranslated regions and non-protein coding RNAs) on the National Center for Biotechnology Information (NCBI) server (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

GeneMerge categorized the human homologs of genes that resulted from all three SHH libraries with over-represented GO terms from the biological processes category given a human background set of genes (Castillo-David and hartl 2003). The significance cutoff was set at $p<0.05$.

The genes that have human homologs from the SSH libraries were clustered into functionally related groups within the subset of biological processes by Database for Annotation, Visualization and Integrated Discovery (DAVID) v.6.7 (Dennis et al. 2003; Huang et al. 2009). The DAVID tool for functional annotation clustering uses GO terms and the term enrichment score was used at high stringency (based on $p<0.05$).

Macroarray preparation and analyses

Three hundred and seventy four clones (including 322 knowns [BLAST hits] and 61 unknowns) from the combined result of three SSH libraries (discussed above) as well as positives (Arabidopsis thaliana *RCA* [X14212], *CAB* [X56062], and *RBCL* [U91966]) and negative controls were spotted onto membranes (Pall Biodyne B Nylon, Nunc) using 100 nanoliter pins on a Biomek 2000 (Beckman Coulter, USA). All samples were spotted in duplicate with four replicates per experimental group (only looked at female and male). The controls provided information about cDNA labeling efficiency, blocking at the prehybridization step, and nonspecific binding. Total RNA was collected (as stated above) from 2 embryos from stage 17 and 2 embryos from stage 19. All collected embryos from the male experimental group in 2004 were pooled as well as for the female experimental group. Pooled total RNA was mixed with control cDNAs and then reverse transcribed before they were labeled with α33P-2'-Deoxyadenosine-5'-Triphosphate (dATP) as previously described (Blum et al. 2008). After hybridization (Blum et al. 2008), the membranes were rinsed and exposed to a phosphor imager and scanned (Molecular Devices Typhoon Scanner).

Signal intensities were quantified using ImageQuant 5.1 (Molecular Dynamics) and intensity differences were calculated as described below. Positive controls used to standardize across arrays for hybridization efficiency were chosen based on their coefficient of variation across all 8 arrays (the coefficient of variance could not exceed 0.2). Each array was normalizated to the geometric mean for all positive controls chosen for that array (Pfaffl et al. 2004). For each array, the range (maximum-minimum) of the negative controls and other blank sites was divided by the median and that number was added to the median to determine the “floor value” for each array. The

highest value for all the arrays was set as the floor value, meaning, any value from each array that is equal to or below that number is considered zero. After normalizing across arrays and discarding values equal to or below the floor value, the average of the duplicates within arrays, the estimated standard deviation across replicate arrays, the median across replicate arrays, and the fold change between treatment groups were calculated. If the standard deviation across replicates was greater than or equal to 2 then those spots were not reliable for further examination and if the fold change between treatment groups was too similar (between 1.65 and .65) then those genes were not considered to have definitive differential expression based on the macroarray experiment, and therefore, not considered for further examination.

Semi-quantitative PCR preparation and analysis

Embryos were collected and total RNA was extracted as stated above from the 31°C experimental group at stage 17, the 26°C experimental group at stage 17, and from the E_2 time trial experimental group at stage 14 from 2006. Four embryos from each group were pooled and cDNA was made using Invitrogen's Superscript III Reverse Transcription kit (Invitrogen, USA) following the manufacturer's instructions.

Pilot experiments were conducted to determine optimum PCR conditions for the candidate genes and a control gene (Protein Phosphatase 1 [*PP1*], Shoemaker et al. 2007). PCR primers were designed using Primer 3 (frodo.wi.mit.edu). Controls were systematically run in each set of semi-quantitative assays: 1) a cDNA positive control for a known sample; 2) a PCR positive control; and 3) a negative control. An internal exogenous standard (cDNA synthesis with no reverse transcriptase) was also run separately for each cDNA mixture. PCR products were loaded onto a 1.5% TBE gel with

ethidium bromide and a 1 kbp DNA ladder molecular weight marker (Minnesota Molecular) and electrophoresed at 90V for 45 minutes.

Analysis of gel images was conducted using ImageJ (available at http://rsb.info.nih.gov/ij; developed by WayneRasband, National Institutes of Health, Bethesda, MD). Each experimental gene was standardized to the control gene, *PP1*, and then normalized to the female group.

Quantitative Real-time PCR (qRT-PCR) preparation and analysis

Five individual whole embryos were collected from stages 17 and 5 from stage 19 from two experimental groups, male and female, and total RNA was extracted as stated above. cDNA was made using ImProm-II™ Reverse Transcriptase and random primers following the manufacturer's instructions. Relative gene expression levels were quantified using an ABI StepOnePlus™ Real-time PCR cycler (StepOne™ Software v2.1) with the following cycling parameters: initial denaturing for 10 minute at 95°C, followed by 40 cycles of 35 seconds (s) at 95°C, 30 s at 60°C, and 30 s at 72°C. The final cycle was followed by a melting curve analysis to verify the amplification of a single product in each well. Specificities of all primer pairs were also verified by sequencing PCR products. Repeating the above procedures on RNA samples (prior to reverse transcription) verified that no products were amplified from contaminating genomic DNA. All samples were run in duplicate and included 3.75 ul of a 1:100 diluted sample, 1 micromolar (uM) of each primer, and 2x SYBR Green Master Mix (Applied Biosystems) in a total of 15ul. PCR efficiencies were calculated from a gene-specific standard curve from a 10-fold dilution series. Relative transcript abundance was normalized to the expression of *PP1* by using the relative standard curve method (Larionov et al. 2005). To determine if expression differed between experimental groups a two-tailed Student's

t-test and a standard error analysis were performed. Primers used to assay gene expression were designed using Primer 3 (frodo.wi.mit.edu) and Amplify (engels.genetics.wisc.edu/amplify/).

Results and Discussion

Suppression subtraction hybridization (SSH) libraries

SSH was used to construct libraries enriched for cDNAs that correspond to mRNAs that exhibit different levels of accumulation during the TSP. Three subtracted cDNA libraries were constructed: one enriched for mRNAs that accumulate at higher levels at the female-producing temperature than the male-producing temperature (hereafter called the “female library”); another enriched for mRNAs that accumulate at higher levels at the male-producing temperature than the female-producing temperature (hereafter called the “male library”); and a third enriched for mRNAs that accumulate at higher levels at the male-producing temperature with exogenous estradiol-17β (sufficient for sex reversal) than in similar embryos treated with the vehicle alone (hereafter called the “E_2 library”). A total of 767 sequences were obtained and were previously deposited in dbEST (FG341000:FG341832). The SSH ESTs were processed as described (Chojnowski et al. 2007; Chojnowski and Braun 2008), yielding a total of 581 contigs and singletons (unigenes) after assembly using CAP3 (Huang and Madan 1999).

SSH libraries typically contain some housekeeping genes (Diatchenko et al. 1996; Luo and Lai 2001; Chen at al. 2009) since it is difficult to completely eliminate genes that do not exhibit differential expression between the two experimental conditions. However, analyses of GC-content (Chojnowski and Braun 2008) provide a line of evidence that the proportion of cDNAs that correspond to housekeeping gene

transcripts is greatly reduced in our SSH libraries; housekeeping genes tend to have a higher GC-content than genes that exhibit lower levels of expression (Kudla et al. 2006; Arhondakis et al. 2008). We found that the GC-content of the turtle transcripts was lower than expected for other reptilian EST efforts. Thus, the SSH method did appear to enrich for genes with lower levels of mRNA accumulation despite being unable, as expected, to eliminate all housekeeping gene cDNAs.

Genes found in the SSH libraries

GeneMerge was used to test for over-represented GO terms signifying biological processes from genes with human homologs found in all three SSH libraries. A total of 34 over-represented GO terms were significant ($p<0.05$) and they represent a broad range of biological processes. A few umbrella categories that a number of over-represented GO terms are anatomical structure morphogenesis (GO:0009653; includes face morphogenesis [GO:0060325] and skeletal system morphogenesis [GO:0048705]), cellular processing (GO:0009987; includes ribosomal small subunit biogenesis [GO:0042274], T cell differentiation in the thymus [GO:0033077], cellular membrane organization [GO:0016044], DNA packaging [GO:0006323], regulation of cell cycle [GO:0051726], negative regulation of apoptosis [GO:0043066]), and metabolic processing (GO:0044267; includes translation [GO:0006412], transcription [GO:0006350], protein folding [GO:0006457], translational initiation [GO:0006413], and translational elongation [GO:0006414]). These categories show that the genes found in the SSH libraries involve active cell differentiation and processing, as expected for mRNAs expressed in developing embryos.

In addition, genes that have human homologs from the SSH libraries were clustered into functionally related groups within the subset of biological processes by

the DAVID (Database for Annotation, Visualization and Integrated Discovery) tool for functional annotation clustering with high stringency. The different groups represent the diversity of the libraries' genes. A pertinent cluster to our study that emerged is one enriched for genes involved in developmental processes (Figure 4-1). Though these genes are identified as being associated with human developmental processes, this study offers a chance to determine if they have been co-opted for similar functions in the turtle. One of the genes from this cluster matrix metalloproteinase 2 (*MMP2*), starred in figure 4-1, is of particular interest because it is one of the first Müllerian inhibiting substance (*MIS*)-target genes involved in Müllerian duct regression and is involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development, reproduction, and tissue remodelling. *MMP2*'s involvement in mammalian development leads us to believe it has potential to be a candidate gene for TSD.

A number of distinct genes (7) in the SSH libraries encode temperature responsive proteins or regulatory genes involved in the heat shock response. Ten temperature responsive cDNAs were found in the female library (two of which exhibited within-library redundancy) whereas only one each was found in the male and E_2 libraries (Table 4-2). Since the female library was enriched for genes expressed at a temperature 5ºC higher than either the male or the E_2 libraries the larger number of heat shock cDNAs could simply reflect a temperature effect. However, specific temperature responsive mRNAs accumulate differentially during gonadal differentiation in another reptile with TSD (*Alligator mississippiensis*; Kohno et al. 2010). Furthermore, specific heat shock proteins play a critical role in the transcriptional complex of steroid hormone receptors

and their corresponding chaperones and cofactors (Picard 2006). Given that temperature is the initial signal in TSD temperature responsive genes represent good candidates for involvement in the TSD cascade.

Differential expression revealed by macroarray analyses

A macroarray assay was used to refine the set of genes identified by sequencing the SSH libraries for sexual dimorphism and place our analysis of transcript accumulation under different experimental conditions in a quantitative framework (Figure 4-2). A total of 29 signals were detected as having differential expression patterns: 19 female biased signals and 10 male biased signals. However, the degree of differential expression revealed by the macroarray analyses was typically <2-fold. Thus, our macroarray analyses were able to show that a number of cDNAs present in the SSH libraries do exhibit sexual dimorphic patterns of expression under the conditions we tested, although the differences in the amount of mRNA present was typically limited. Given that experiments were conducted on whole embryos to ensure a full scale approach for candidate genes, those found to have significant results are underrepresented assuming a specific tissue or subset of tissues drive mRNA production of any given gene at a given time during development (Ramsey and Crews 2007).

The genes that emerged from the macroarray as being differentially expressed have a mixture of biological roles in humans based on DAVID. Some genes overlap in their biological roles while others have more distinct roles. For example, 10 genes (*GTPBP4*, *HSP90AA1*, *ARID4A*, *RAN*, *HBZ*, *SERPINA3*, *BRIP1*, *NFE2L1*, *CDK6*, and *NFIB*) are involved in the regulation of metabolic processing and 6 genes (*GTPBP4*, *BRIP1*, *RAN*, *KATNA1*, *CDK6*, and *NFIB*) are involved in cell division and proliferation.

Moreover, genes like *AFP* and *LAPTM4A* are independently involved in reproduction and transport respectively. Though the macroarray genes that are differentially expressed are not all specifically involved in sexual development of a human, their diversity provides a way to look at turtle development from different perspectives.

Semi quantitative (semiQ) PCR validation of TSD candidate genes

The semiQ-PCR experiments were split into three categories: sexual dimorphic expression at stage 17, differences between stages 14 and 17, and an E_2 time trial conducted during stage 14 including fast (6 hours) and slow (24 hours) responses (Figure 4-3). Five candidate genes (chromosome 16 open reading frame 62 [*C16ORF62*], chaperonin containing TCP1, subunit 3 [*CCT3*], *MMP2*, nuclear factor I/X [*NFIX*], and Notch homolog 2 [*NOTCH2*]) were tested for all three categories.

Only the turtle ortholog of *C16ORF62*, a gene of unknown function, showed evidence of sexually dimorphic expression at stage 17. The transcript of this gene showed greater accumulation in males than in females (~2.5 fold increase), the same trend that was evident in the macroarray results. Since *C16ORF62* is a gene of unknown function it is truly a novel candidate for a gene involved in turtle TSD and potentially in sexual development in other vertebrates as well.

Four genes exhibited increased accumulation during stage 17 relative to stage 14 in embryos incubated at the male producing temperature (26ºC). This stage effect was evident for *C16ORF62*, *CCT3*, *MMP2*, and *NFIB*; the most striking is a ~8-fold increase in mRNA accumulation between stage 14 and 17 for *C16ORF62*. The others showed a range of relative increase in mRNA accumulation of 1.2-fold to 4.8-fold with stage progression. *CCT3*, *MMP2*, and *NFIB* have been implicated in gonad development (*MMP2*: Robinson et al. 2001) or other aspects of development (*NFIB* and *CCT3*:

Chaudhry et al. 1997; Walkley et al. 1996) in mammals. When this information is combined with our observation that turtle orthologs exhibited increased mRNA accumulation as development proceeded from stage 14 to stage 17 (early in TSP) it is reasonable to speculate that these genes play a role in turtle development, potentially sexual development in the case of *MMP2*.

All five genes show a rapid (6 hours) response to E_2 exposure. *C16ORF62*, *CCT3*, and *MMP2* all show a downregulation of mRNA expression and *NFIB* and *NOTCH2* show an upregulation. Four of the genes (*CCT3*, *MMP2*, *NFIB*, and *NOTCH2*) exhibited similar mRNA accumulation both 6 hours and 24 hours after E_2 exposure; accumulation of the *C16ORF62* mRNA almost returned to pre-exposure levels after 24 hours. Of all these genes, only *MMP2* has been previously shown to be affected by E_2 while the others have not had any previous experimentation done with E_2. Mahmoodzadeh et al (2010) showed that E_2 inhibits *MMP2* gene expression in rat fibroblasts and those results corroborate our findings for *MMP2*'s involvement with E_2.

Expression of a long noncoding RNA (ncRNA) is sexually-dimorphic

A number of cDNAs on the macroarray (61) could not be identified using BLASTX, suggesting that they correspond either to cDNAs for which only untranslated region was included in the EST read or non-coding RNAs (ncRNA). To identify some of these cDNAs we conducted BLASTN (Nucleotide-nucleotide BLAST) searches and revealed that one of the cDNAs that expresses sexual dimorphism is a ncRNA, metastasis associated lung adenocarcinoma transcript 1 (*MALAT1*).

MALAT1 is a long (~7kb) ncRNA that undergoes a cleavage that produces two RNAs, a smaller transferRNA (tRNA)-like cytoplasmic RNA (~61nucleotides [nt]) and a 6.7kb RNA that localize to two different subcellular compartments, cytoplasm and

nuclear speckles respectively (Wilusz et al. 2008). Characteristically, it has short blocks of high conservation across the entire transcript, especially in 3' half of RNA, and lacks repetitive elements except for a short interspersed nucleotide element (SINE) and long interspersed nucleotide element (LINE) element near its 5' end. The smaller (~61nt) transcript after cleavage is highly conserved across many species, including mouse, human, dog, lizard, frog, and stickleback. It has not yet been found in chicken though this could be a database annotation error and not necessarily a negative result.

MALAT1 shows a broad distribution of expression in normal human and mouse tissues but its misregulation is correlated with the progression of cancers and it is upregulated in many human carcinomas (Ji et al. 2003; Guffanti et al. 2009; Koshimizu et al. 2010; Guo et al. 2010). More importantly for this study, *MALAT1* accumulation is higher in adult mammalian ovaries than adult testes (Hutchinson et al. 2007; Wilusz et al. 2008). However, the pattern of differential expression for *MALAT1* in adult mammalian gonads is distinct from the pattern we observed using the macroarray assay, in which the mRNA accumulation appeared 1.6-fold higher in whole male turtle embryos. Since *MALAT1* is a ncRNA that shows dimorphic expression in the TSP we felt it was an excellent candidate for a gene involved in TSD so we used qRT-PCR to verify the pattern of expression suggested by the macroarray.

We used qRT-PCR to examine *MALAT1* RNA accumulation because it represents a rigorous test of differential expression. *MALAT1* RNA accumulation was examined independently for multiple individuals (n=5) and the two stages during the TSP (stages 17 and 19) rather than using pooled samples. This analysis revealed a slight but significant sexual dimorphism (about 1.4-fold higher in males) in the amount of *MALAT1*

RNA during both stages we examined (Figure 4-4). *MALAT1* RNA expression also shows a modest increase as development progresses from stage 17 to stage 19 in both males and females. These observations are consistent with the hypothesis that *MALAT1* plays a role in turtle TSD.

Conclusions

Little is known about the genes involved in vertebrate TSD, regardless of whether they are protein-coding genes or ncRNAs. Indeed, most of the available information on TSD reflects studies that have focused on the orthologs of genes involved in GSD in mammals (Table 4-1). Here we report a survey of genes identified based upon their patterns of mRNA accumulation during sexual development in the Red-eared slider turtle. This strategy is complementary to the analysis of the orthologs of genes involved in mammalian sexual development, and it revealed two genes (*MALAT1* and *C16ORF62*) that show greater accumulation in males than in females during the TSP in the red-eared slider turtle. Four genes that exhibited increased mRNA expression as development proceeded from stage 14 to stage 17 were identified, as were a set of genes that responded to E_2 exposure. This survey focused on changes in mRNA accumulation in whole embryos. Thus, it remains possible that some or all of these genes exhibit sexually dimorphic expression in specific tissues (e.g., the developing gonad or brain). However, the genes we identified are likely to be significant since differential expression at the whole embryo level is expected to be a conservative way to examine gene expression during sexual development.

*MMP*s (matrix metalloproteinases) are involved in the breakdown of extracellular matrices in physiological processes, including cancer (Bourboulia and Stetler-Stevenson 2010). *MMP2* is sexually dimorphic in developing male mice because it

functions as a paracrine death factor in Müllerian duct regression downstream of the *MIS* cascade (Roberts et al. 2002). In addition, *MMP2* was found to be sexually dimorphic and regulated by testosterone in songbirds in relation to the vocal control center during adult neurogenesis (Kim et al. 2008). Furthermore, estrogen affects the MMP pathway in humans by increasing MMP2 enzymatic activity (Grandas et al. 2009). It is unclear if the increase in MMP2 is through an increase in mRNA accumulation or potentially through binding affinity changes. Though *MMP2* was not found to be sexually dimorphic in turtles it was found to be inhibited by E_2, opposite of mammals but potentially similar to birds (Kim et al. 2008). Together with the prior knowledge of its involvement in mammalian and avian development, *MMP2* is a novel candidate gene for development of a turtle.

ncRNAs are believed to play a large number of biological roles (reviewed in Ponting et al. 2009), but their role in sexual development remains poorly characterized (McFarlane and Wilhelm 2010). Although there is some evidence that ncRNAs have roles in sexual development in both mammals (McFarlane and Wilhelm 2010) and birds (Zhao et al. 2010), this is the first evidence that a ncRNA may have a role in sexual development for an organism with TSD. This hypothesis is corroborated by the fact that *MALAT1* exhibits differential expression in mammalian gonads (expression is higher in adult ovaries than in testes). However, the pattern of sexual dimorphism reported for mammals is distinct from that evident in turtles (expression is higher in male embryos than in female embryos). Our findings highlight the importance of examining ncRNAs when investigating vertebrate sexual development.

Table 4-1. A general overview of sexually dimorphic gene expression in turtles with TSD.

	stage 17		late in TSP		after TSP		
Gene	Testis	Ovary	Testis	Ovary	Testis	Ovary	Reference
SF1	+		+		+		Fleming et al. 1999
WT1	same	same	same	same	same	same	Schmahl et al. 2003; Spotila et al. 1998; Valenzuela 2008
DAX1	same	same	same	same	same	same	Shoemaker et al. 2007
SOX9	same	same	+		+		Spotila et al. 1998; Shoemaker et al. 2007
DMRT1	+		+		+		Murdock and Wibbels 2003b; Torres Maldonado et al. 2002; Shoemaker et al. 2007)
CYP19	same	same		+		+	Murdock and Wibbels 2003a; Murdock and Wibbels 2003b; Ramsey et al. 2007
SOX8	same	same	same	same	same	same	Takada et al. 2004;
FOXL2	same	same		+		+	Shoemaker et al. 2007
MIS	same	same	+		+		Takada et al. 2004; Shoemaker et al. 2007
RSPONDIN	same	same		+		+	
WNT4	same	same	same	same		+	Shoemaker et al. 2007

Table 4-2. Temperature responsive genes found in SSH.

Name	Library where found	Redundancy within library
HSPA8	Male	1
HSP90B1	E_2	1
CIRBP	Female	2
HSBP1	Female	3
HSP90AA1	Female	1
HSPD1	Female	1
SERPINH1	Female	1

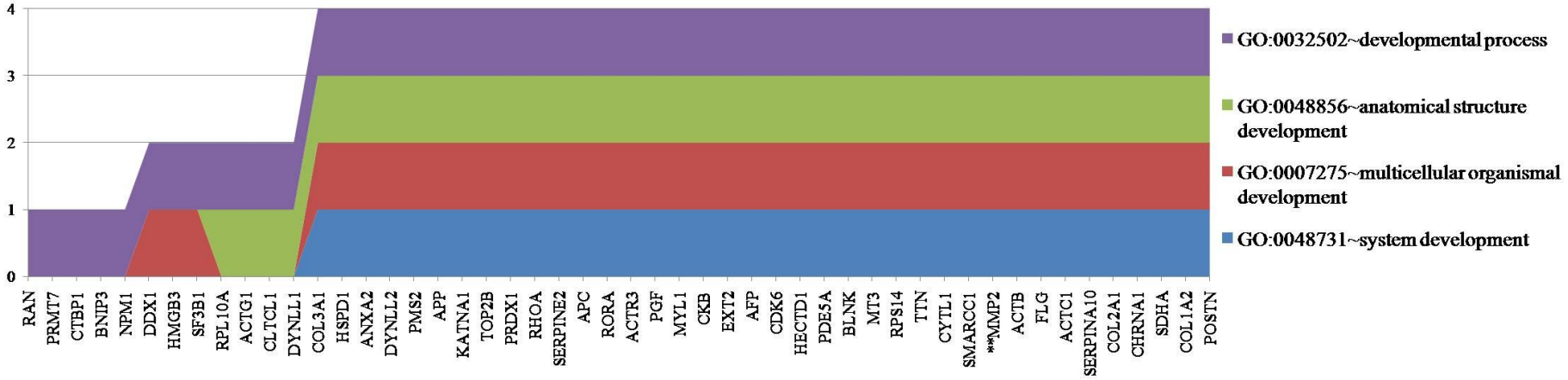


Figure 4-1. Development Cluster from DAVID functional annotation clustering with high stringency. High stringency was used to determine this functional cluster of developmental genes and GO terms from the resulting known genes from the SSH libraries.

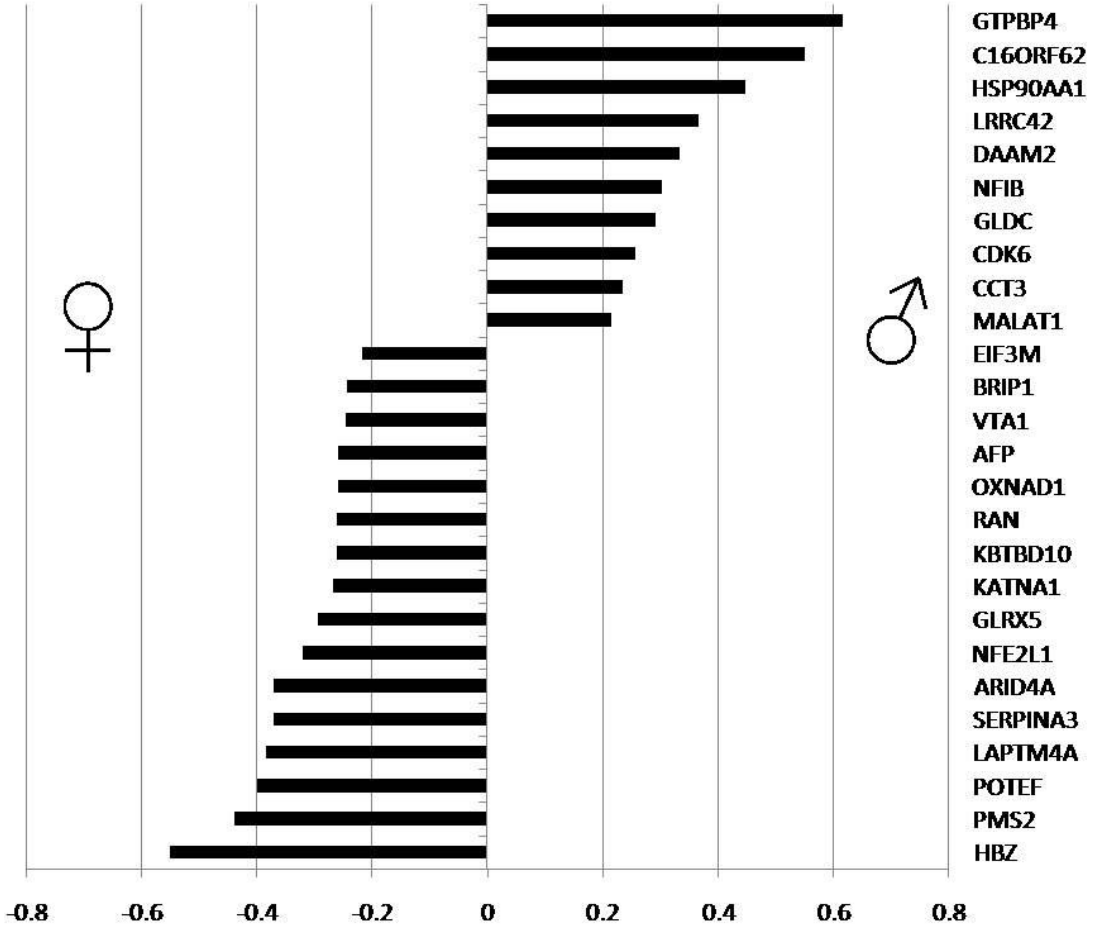
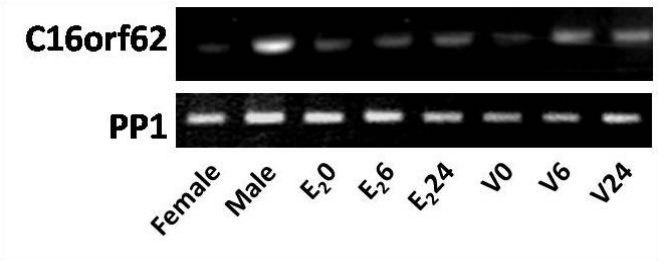
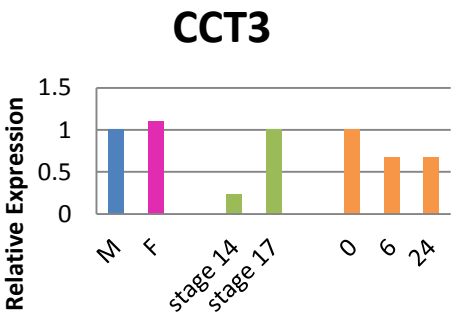


Figure 4-2. Selected macroarray results showing sexually dimprophic patterns. Log view of the fold change between female and male expression patterns determined from the macroarray.

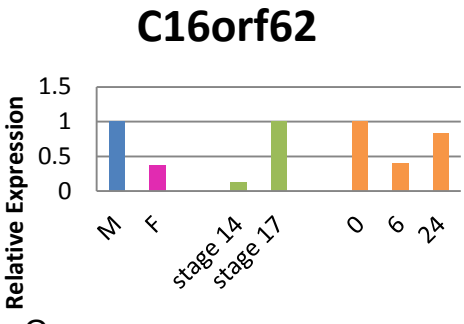
A.



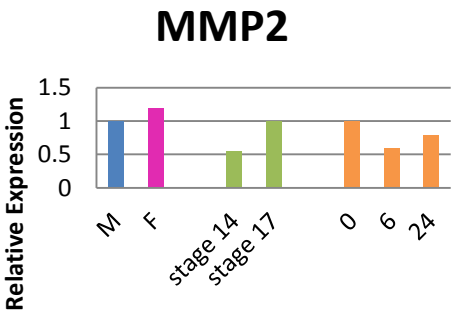
B.



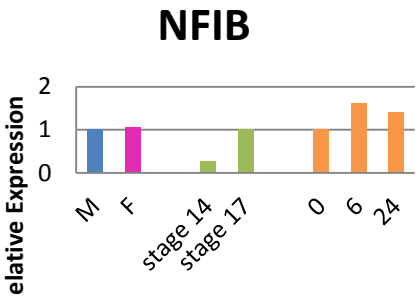
C.



D.



E.



F.

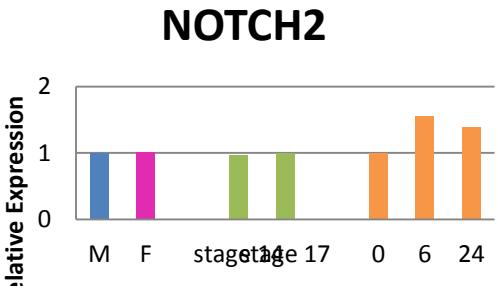


Figure 4-3. Semi-quantitative PCR Results. A. An example of a semi-quantitative gel image. B-F. Semi-quantitative results for sexual dimorphism (M=male, F=female), stage effect between males(green bars), and the E2 time trial (orange bars).

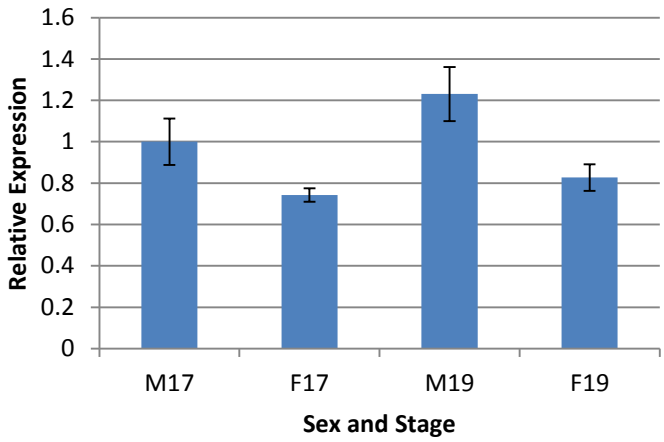


Figure 4-4. Quantitative RT-PCR showing sexual dimorphic expression at stage 17 and 19. The expression of *MALAT1* (n=5) consists of sex and stage (e.g. M17 = male, stage 17) and is relative to the control gene (*PP1*). The relative expression for M17 was set as 1 to allow easier comparison between groups. Gene expression differences were analyzed between sexes at the same stage (stage 17, p=0.04 and stage 19, p=0.01) and between stages of the same sex (male, p=0.18 and female, p=0.24) with a two-tailed TTest.

CHAPTER 5
CONCLUSIONS

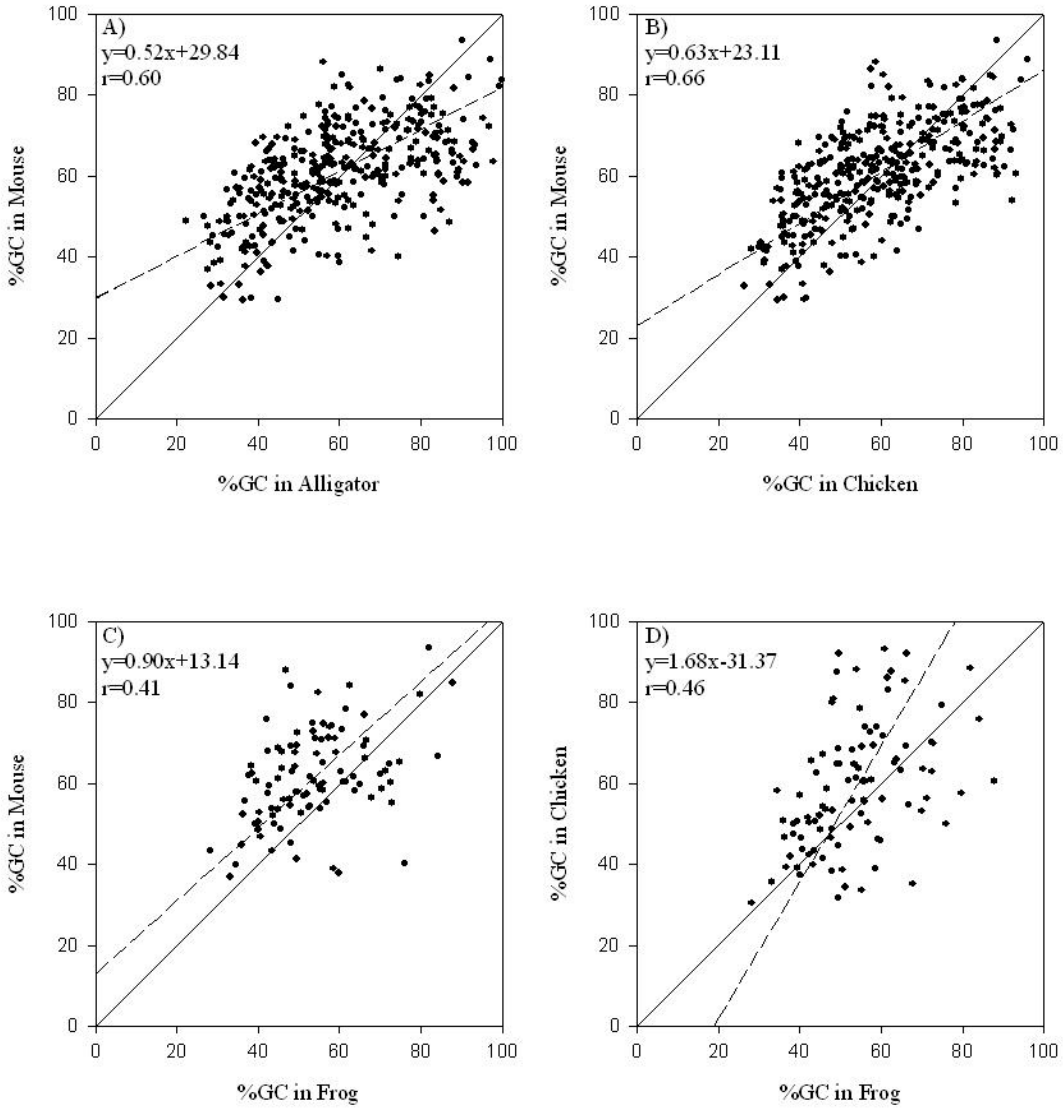
Vertebrate genomes are mosaics of isochores, defined as long (> 100 kb) regions with relatively homogeneous within-region base composition. Birds and mammals have more GC-rich isochores than amphibians and fish, and the GC-rich isochores of birds and mammals have been suggested to be an adaptation to homeothermy. If this hypothesis is correct, all poikilothermic (cold-blooded) vertebrates, including the non-avian reptiles, are expected to lack a GC-rich isochore structure. Previous studies using various methods to examine isochore structure in crocodilians, turtles, and squamates have led to different conclusions. We collected more than 6,000 ESTs from the American alligator to overcome sample size limitations suggested to be fundamental problems in the previous reptilian studies. The alligator ESTs were assembled and aligned with their human, mouse, chicken and western clawed frog orthologs, resulting in 366 alignments. Analyses of third codon position GC-content provided conclusive evidence that the poikilothermic alligator has GC-rich isochores, like homeothermic birds and mammals. We placed these results in a theoretical framework able to unify available models of isochore evolution. The data collected for this study allowed us to reject the models that explain the evolution of GC-content using changes in body temperature associated with the transition from poikilothermy to homeothermy. Falsification of these models places fundamental constraints upon the plausible pathways for the evolution of isochores.

Vertebrate genomes are comprised of isochores that are relatively long (>100 kb) regions with a relatively homogenous (either GC-rich or AT-rich) base composition and with rather sharp boundaries with neighboring isochores. Mammals and living

archosaurs (birds and crocodilians) have heterogeneous genomes that include very GC-rich isochores. In sharp contrast, the genomes of amphibians and fishes are more homogeneous and they have a lower overall GC content. Because DNA with higher GC content is more thermostable, the elevated GC content of mammalian and archosaurian DNA has been hypothesized to be an adaptation to higher body temperatures. This hypothesis can be tested by examining structure of isochores across the reptilian clade, which includes the archosaurs, testudines (turtles), and lepidosaurs (lizards and snakes), because reptiles exhibit diverse body sizes, metabolic rates and patterns of thermoregulation. This study focuses on a comparative analysis of a new set of expressed genes of the Red-eared slider turtle and orthologs of the turtle genes in mammalian (human, mouse, dog, and opossum), archosaurian (chicken and alligator) and amphibian (Western clawed frog) genomes. EST data from a turtle cDNA library enriched for genes that have specialized functions (developmental genes) revealed using the GC content of the third codon position to examine isochore structure requires careful consideration of the types of genes examined. The more highly expressed genes (e.g., housekeeping genes) are more likely to be GC-rich than are genes with specialized functions. However, the set of highly expressed turtle genes demonstrated that the turtle genome has a GC content that is intermediate between the GC-poor amphibians and the GC-rich mammals and archosaurs. There was a strong correlation between the GC content of all turtle genes and the GC content of other vertebrate genes, with the slope of the line describing this relationship also indicating that the isochore structure of turtles is intermediate between that of amphibians and other amniotes. These data are consistent with some thermal hypotheses of isochore

evolution, but we believe that the credible set of models for isochore evolution still includes a variety of models. These data expand the amount of genomic data available from reptiles upon which future studies of reptilian genomics can build.

APPENDIX A
CORRELATIONS AMONG VERTEBRATES IN GC3 CONTENT



The graphs correspond to the comparisons not found in Figures 3 or 4. Comparisons A) alligator and mouse and B) chicken and mouse used the complete set of 366 genes and comparison C) frog and mouse and D) frog and chicken used the smaller 98 gene set.

APPENDIX B

VALUES OF THE SLOPE OF LINES RELATING GC3 IN DIFFERENT ORGANISMS

Slope values correspond to the listed figures. 95% confidence intervals for each of the comparisons are provided.

Comparison (X-Y)	*Figure*	*Slope*	*95% Confidence Interval*
mouse-human	3A	1.47	0.15
chicken-human	3B	1.01	0.10
alligator-human	3C	0.87	0.09
chicken-alligator	3D	0.88	0.09
frog-human	4A	1.83	0.36
frog-alligator	4B	1.77	0.35
alligator-mouse	S2A	0.52	0.05
chicken-mouse	S2B	0.63	0.06
frog-mouse	S2C	0.9	0.18
frog-chicken	S2D	1.68	0.33

LIST OF REFERENCES

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-402

Arhondakis S, Clay O, Bernardi G (2008) GC level and expression of human coding sequences. Biochem Biophys Res Commun 367:542-5

Benton MJ and Donoghue PC. (2007) Paleontological evidence to date the tree of life.

Mol Biol Evol. 24:26-53

Bernardi G. (1995) The human genome: organization and evolutionary history. Annu. Rev. Genet. 29:445–447

Bernardi G (2000) Isochores and the evolutionary genomics of vertebrates. Gene 241:3-17

Bernardi G (2007) The neoselectionist theory of genome evolution. Proc Natl Acad Sci U S A 104:8385-90

Bernardi G, Bernardi G (1991) Compositional Properties of Nuclear Genes from Cold-Blooded Vertebrates. Journal of Molecular Evolution 33:57-67

Bernardi G, Hughes S, Mouchiroud D (1997) The major compositional transitions in the vertebrate genome. J Mol Evol 44 Suppl 1:S44-51

Birk OS, Casiano DE, Wassif CA, Cogliati T, Zhao L, Zhao Y, Grinberg A, Huang S, Kreidberg JA, Parker KL, Porter FD, Westphal H (2000) The LIM homeobox gene Lhx9 is essential for mouse gonad formation. Nature 403:909-13

Birney E, Andrews D, Caccamo M, Chen Y, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Flicek P, Graf S, Hammond M, Herrero J, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Kokocinski F, Kulesha E, London D, Longden I, Melsopp C, Meidl P, Overduin B, Parker A, Proctor G, Prlic A, Rae M, Rios D, Redmond S, Schuster M, Sealy I, Searle S, Severin J, Slater G, Smedley D, Smith J, Stabenau A, Stalker J, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Hubbard TJ (2006) Ensembl 2006. Nucleic Acids Res 34:D556-61

Blum JL, Prucha MS, Patel VJ, Denslow ND (2008) Use of cDNA macroarrays and gene profiling for detection of effects of environmental toxicants. Methods Mol Biol 410:43-53

Boulikas T (1992) Evolutionary consequences of nonrandom damage and repair of chromatin domains. J Mol Evol 35:156-80

Bourboulia D, Stetler-Stevenson WG (2010) Matrix metalloproteinases (MMPs) and tissue inhibitors of metalloproteinases (TIMPs): Positive and negative regulators in tumor cell adhesion. Semin Cancer Biol [Epub ahead of print]

Cacciò S, Jabbari K, Matassi G, Guermonprez F, Desgres J, and Bernardi G. (1997) Methylation patterns in the isochores of vertebrate genomes. Gene 205:119-124.

Cacciò S, Zoubak S, D'Onofrio G, and Bernardi G. (1995) Nonrandom frequency patterns of synonymous substitutions in homologous mammalian genes. J. Mol. Evol. 40:280–292

Castillo-Davis CI, Hartl DL (2003) GeneMerge--post-genomic analysis, data mining, and hypothesis testing. Bioinformatics 19:891-92

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, and Thompson JD. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31:3497-3500

Chaudhry AZ, Lyons GE, Gronostajski RM (1997) Expression patterns of the four nuclear factor I genes during mouse embryogenesis indicate a potential role in development. Dev Dyn 208:313-25

Chen AQ, Wang ZG, Xu ZR, Yu SD, Yang ZG (2009) Analysis of gene expression in granulosa cells of ovine antral growing follicles using suppressive subtractive hybridization. Anim Reprod Sci 115:39-48

Chiusano ML, D'Onofrio G, Alvarez-Valin F, Jabbari K, Colonna G, Bernardi G (1999) Correlations of nucleotide substitution rates and base composition of mammalian coding sequences with protein structure. Gene 238:23-31

Chojnowski JL, Braun EL (2008) Turtle isochore structure is intermediate between amphibians and other amniotes. Integrative and Comparative Biology 48:454-462

Chojnowski JL, Franklin J, Katsu Y, Iguchi T, Guillette LJ, Jr., Kimball RT, Braun EL (2007) Patterns of vertebrate isochore evolution revealed by comparison of expressed mammalian, avian, and crocodilian genes. J Mol Evol 65:259-66

Christian KA, Tracy CR, Tracy CR (2006) Evaluating thermoregulation in reptiles: An appropriate null model. American Naturalist 168:421-430

Clinton M (1998) Sex determination and gonadal development: a bird's eye view. J Exp Zool 281:457-65

Costantini M, Auletta F, Bernardi G (2007a) Isochore patterns and gene distributions in fish genomes. Genomics 90:364-71

Costantini M, Clay O, Auletta F, Bernardi G (2006) An isochore map of human chromosomes. Genome Res 16:536-41

Costantini M, Di Filippo M, Auletta F, Bernardi G (2007b) Isochore pattern and gene distribution in the chicken genome. Gene 400:9-15

Crews D, Bergeron JM (1994) Role of reductase and aromatase in sex determination in the red-eared slider (Trachemys scripta), a turtle with temperature-dependent sex determination. J Endocrinol 143:279-89

Crews D, Bergeron JM, Bull JJ, Flores D, Tousignant A, Skipper JK, Wibbels T (1994) Temperature-dependent sex determination in reptiles: proximate mechanisms, ultimate outcomes, and practical applications. Dev Genet 15:297-312

Crews D, Bull JJ, Wibbels T (1991) Estrogen and sex reversal in turtles: a dose-dependent phenomenon. Gen Comp Endocrinol 81:357-64

Crews D, Cantu AR, Bergeron JM, Rhen T (1995) The relative effectiveness of androstenedione, testosterone, and estrone, precursors to estradiol, in sex reversal in the red-eared slider (Trachemys scripta), a turtle with temperature-dependent sex determination. Gen Comp Endocrinol 100:119-27

Cummings AM, Kavlock RJ (2004) Function of sexual glands and mechanism of sex differentiation. J Toxicol Sci 29:167-78

D'Onofrio G, Jabbari K, Musto H, Bernardi G (1999) The correlation of protein hydropathy with the base composition of coding sequences. Gene 238:3-14

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol 4:3

Diatchenko L, Lau YF, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 93:6025-30

Duret L, Eyre-Walker A, Galtier N (2006) A new perspective on isochore evolution. Gene 385:71-4

Duret L, Semon M, Piganeau G, Mouchiroud D, Galtier N (2002) Vanishing GC-rich isochores in mammalian genomes. Genetics 162:1837-47

Eisenberg E, Levanon EY (2003) Human housekeeping genes are compact. Trends Genet 19:362-5

Eyre-Walker A (1992) Evidence that both G + C rich and G + C poor isochores are replicated early and late in the cell cycle. Nucleic Acids Res 20:1497-501

Eyre-Walker A. (1993) Recombination and genome evolution in mammals. Proc. R. Soc., B. 252:237-243

Eyre-Walker A. (1999) Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics 152:675–683

Eyre-Walker A, Hurst LD (2001) The evolution of isochores. Nat Rev Genet 2:549-55

Ezaz T, Stiglec R, Veyrunes F, Marshall Graves JA (2006) Relationships between vertebrate ZW and XY sex chromosome systems. Curr Biol 16:R736-43

Fleming A, Wibbels T, Skipper JK, Crews D (1999) Developmental expression of steroidogenic factor 1 in a turtle with temperature-dependent sex determination. Gen Comp Endocrinol 116:336-46

Fortes GG, Bouza C, Martinez P, Sanchez L (2007) Diversity in isochore structure among cold-blooded vertebrates based on GC content of coding and non-coding sequences. Genetica 129:281-9

Fryxell KJ, Zuckerkandl E (2000) Cytosine deamination plays a primary role in the evolution of mammalian isochores. Mol Biol Evol 17:1371-83

Galtier N, Mouchiroud D (1998) Isochore evolution in mammals: a human-like ancestral structure. Genetics 150:1577-84

Galtier N, Piganeau G, Mouchiroud D, and Duret L. (2001) GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics 159:907-911

Gao F, Zhang CT (2006) Isochore structures in the chicken genome. Febs J 273:1637-48

Gatten R (1974) Effect of nutritional status on the preferred body temperature of the turtles *Pseudyms scripta* and *Terrapene ornate*. Copeia 1974:912-917

Goth A, Booth D (2004) Temperature-dependent sex ratio in a bird. Biology Letters

Goth A, Booth DT (2005) Temperature-dependent sex ratio in a bird. Biology Letters 1:31-33

Grandas OH, Mountain DH, Kirkpatrick SS, Cassada DC, Stevens SL, Freeman MB, Goldman MH (2009) Regulation of vascular smooth muscle cell expression and function of matrix metalloproteinases is mediated by estrogen and progesterone exposure. J Vasc Surg 49:185-91

Gu J, Li WH (2006) Are GC-rich isochores vanishing in mammals? Gene 385:50-6

Gu W, Ray DA, Walker JA, Barnes EW, Gentles AJ, Samollow PB, Jurka J, Batzer MA, Pollock DD (2007) SINEs, evolution and genome structure in the opossum. Gene 396:46-58

Guan G, Kobayashi T, Nagahama Y (2000) Sexually dimorphic expression of two types of DM (Doublesex/Mab-3)-domain genes in a teleost fish, the Tilapia (Oreochromis niloticus). Biochem Biophys Res Commun 272:662-6

Guffanti A, Iacono M, Pelucchi P, Kim N, Solda G, Croft LJ, Taft RJ, Rizzi E, Askarian-Amiri M, Bonnal RJ, Callari M, Mignone F, Pesole G, Bertalot G, Bernardi LR, Albertini A, Lee C, Mattick JS, Zucchi I, De Bellis G (2009) A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC Genomics 10:163

Guo F, Li Y, Liu Y, Wang J, Li Y, Li G (2010) Inhibition of metastasis-associated lung adenocarcinoma transcript 1 in CaSki human cervical cancer cells suppresses cell proliferation and invasion. Acta Biochim Biophys Sin (Shanghai) 42:224-9

Hackenberg M, Bernaola-Galvan P, Carpena P, and Oliver JL. (2005) The biased distribution of Alus in human isochores might be driven by recombination. J Mol Evol. 60:365-377

Hamada K, Horiike T, Kanaya S, Nakamura H, Ota H, Yatogo T, Okada K, Nakamura H, Shinozawa T (2002) Changes in body temperature pattern in vertebrates do not influence the codon usages of alpha-globin genes. Genes Genet Syst 77:197-207

Harry JL, Williams KL, Briscoe DA (1990) Sex determination in loggerhead turtles: differential expression of two hnRNP proteins. Development 109:305-12

Hillenius WJ, Ruben JA (2004) Getting warmer, getting colder: reconstructing crocodylomorph physiology. Physiol Biochem Zool 77:1068-72; discussion 1073-5

Hiratani I, Leskovar A, Gilbert DM (2004) Differentiation-induced replication-timing changes are restricted to AT-rich/long interspersed nuclear element (LINE)-rich isochores. Proc Natl Acad Sci U S A 101:16861-6

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res 9:868-77

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44-57

Hugall AF, Foster R, Lee MS (2007) Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Syst Biol 56:543-63

Hughes S, Zelus D, Mouchiroud D (1999) Warm-blooded isochore structure in Nile crocodile and turtle. Mol Biol Evol 16:1521-7

Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A (2007) A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics 8:39

Isobe T, Feigelson ED, Akritas MG, Babu GJ (1990) Linear-Regression in Astronomy.1. Astrophysical Journal 364:104-113

Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K (2005) Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA-coded proteins. Mol Biol Evol 22:810-3

Jabbari K and Bernardi G. (2004) Body temperature and evolutionary genomics of vertebrates: a lesson from the genomes of Takifugu rubripes and Tetraodon nigroviridis. Gene 333:179-188

Janzen FJ, Phillips PC (2006) Exploring the evolution of environmental sex determination, especially in reptiles. J Evol Biol 19:1775-84

Ji P, Diederichs S, Wang W, Boing S, Metzger R, Schneider PM, Tidow N, Brandt B, Buerger H, Bulk E, Thomas M, Berdel WE, Serve H, Muller-Tidow C (2003) MALAT-1, a novel noncoding RNA, and thymosin beta4 predict metastasis and survival in early-stage non-small cell lung cancer. Oncogene 22:8031-41

Johnston CM, Barnett M, Sharpe PT (1995) The molecular biology of temperature-dependent sex determination. Philos Trans R Soc Lond B Biol Sci 350:297-303; discussion 303-4

Katoh-Fukui Y, Tsuchiya R, Shiroishi T, Nakahara Y, Hashimoto N, Noguchi K, Higashinakagawa T (1998) Male-to-female sex reversal in M33 mutant mice. Nature 393:688-92

Kettlewell JR, Raymond CS, Zarkower D (2000) Temperature-dependent expression of turtle Dmrt1 prior to sexual differentiation. Genesis 26:174-8

Kim DH, Lilliehook C, Roides B, Chen Z, Chang M, Mobashery S, Goldman SA (2008) Testosterone-induced matrix metalloproteinase activation is a checkpoint for neuronal addition to the adult songbird brain. J Neurosci 28:208-16

Kohno S, Katsu Y, Urushitani H, Ohta Y, Iguchi T, Guillette LJ, Jr. (2010) Potential contributions of heat shock proteins to temperature-dependent sex determination in the American alligator. Sex Dev 4:73-87

Koshimizu TA, Fujiwara Y, Sakai N, Shibata K, Tsuchiya H (2010) Oxytocin stimulates expression of a noncoding RNA tumor marker in a human neuroblastoma cell line. Life Sci 86:455-60

Kreidberg JA, Sariola H, Loring JM, Maeda M, Pelletier J, Housman D, Jaenisch R (1993) WT-1 is required for early kidney development. Cell 74:679-91

Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M (2006) High guanine and cytosine content increases mRNA levels in mammalian cells. PLoS Biol 4:e180

Larionov A, Krause A, Miller W (2005) A standard curve based method for relative real time PCR data processing. BMC Bioinformatics 6:62

Li MK, Gu L, Chen SS, Dai JQ, Tao SH (2007) Evolution of the isochore structure in the scale of chromosome: insight from the mutation bias and fixation bias. J Evol Biol 21:173-182

Li WH. (1997) Molecular Evolution. Sunderland, MA: Sinauer Associates. p. 407-411

Liang F, Holt I, Pertea G, Karamycheva S, Salzberg SL, Quackenbush J (2000) An optimized protocol for analysis of EST sequences. Nucleic Acids Res 28:3657-65

Loffler K, Zarkower D, Koopman P (2003) Etiology of ovarian failure in blepharophimosis ptosis epicanthus inversus syndrome: FOXL2 is a conserved, early-acting gene in vertebrate ovarian development. Endocrinol 144:3237-43

Luo MJ, Lai MD (2001) Identification of differentially expressed genes in normal mucosa, adenoma and adenocarcinoma of colon by SSH. World J Gastroenterol 7:726-31

Luo X, Ikeda Y, Parker KL (1994) A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. Cell 77:481-90

Maddison D, Maddison W (2002) MacClade. Sinauer Associates, Sunderland, Mass

Mahmoodzadeh S, Dworatzek E, Fritschka S, Pham TH, Regitz-Zagrosek V (2010) 17beta-Estradiol inhibits matrix metalloproteinase-2 transcription via MAP kinase in fibroblasts. Cardiovasc Res 85:719-28

Marin I, Siegal ML, Baker BS (2000) The evolution of dosage-compensation mechanisms. Bioessays 22:1106-14

Marshall Graves JA, Shetty S (2001) Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. J Exp Zool 290:449-62

Matsuda M, Shinomiya A, Kinoshita M, Suzuki A, Kobayashi T, Paul-Prasanth B, Lau EL, Hamaguchi S, Sakaizumi M, Nagahama Y (2007) DMY gene induces male development in genetically female (XX) medaka fish. Proc Natl Acad Sci U S A 104:3865-70

McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP et al (2006) AgBase: a functional genomics resource for agriculture. BMC Genomics 7:229

McFarlane L, Wilhelm D (2009) Non-coding RNAs in mammalian sexual development. Sex Dev 3:302-16

Miller D, Summers J, Silber S (2004) Environmental versus genetic sex determination: a possible factor in dinosaur extinction? Fertil Steril 81:954-64

Miyamoto N, Yoshida M, Kuratani S, Matsuo I, Aizawa S (1997) Defects of urogenital development in mice lacking Emx2. Development 124:1653-64

Modi WS, Crews D (2005) Sex chromosomes and sex determination in reptiles. Curr Opin Genet Dev 15:660-5

Murdock C, Wibbels T (2003a) Cloning and expression of aromatase in a turtle with temperature-dependent sex determination. Gen Comp Endocrinol 130:109-19

Murdock C, Wibbels T (2003b) Expression of Dmrt1 in a turtle with temperature-dependent sex determination. Cytogenet Genome Res 101:302-8

Murdock C, Wibbels T (2006) Dmrt1 expression in response to estrogen treatment in a reptile with temperature-dependent sex determination. J Exp Zool B Mol Dev Evol 306:134-9

Musto H, Romero H, Zavala A, Bernardi G (1999) Compositional correlations in the chicken genome. J Mol Evol 49:325-9

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP (2004) Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol Lett 26:509-15

Picard D (2006) Chaperoning steroid hormone action. Trends Endocrinol Metab 17:229-35

Ponting CP, Oliver PL, Reik W (2009) Evolution and functions of long noncoding RNAs. Cell 136:629-41

Ramsey M, Crews D (2007) Adrenal-kidney-gonad complex measurements may not predict gonad-specific changes in gene expression patterns during temperature-dependent sex determination in the red-eared slider turtle (Trachemys scripta elegans). J Exp Zool A Ecol Genet Physiol 307:463-70

Ramsey M, Crews D (2009) Steroid signaling and temperature-dependent sex determination-Reviewing the evidence for early action of estrogen during ovarian determination in turtles. Semin Cell Dev Biol 20:283-92

Rest JS, Ast JC, Austin CC, Waddell PJ, Tibbetts EA, Hay JM, Mindell DP (2003) Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. Mol Phylogenet Evol 29:289-97

Roberts LM, Visser JA, Ingraham HA (2002) Involvement of a matrix metalloproteinase in MIS-induced cell death during urogenital development. Development 129:1487-96

Robinson LL, Sznajder NA, Riley SC, Anderson RA (2001) Matrix metalloproteinases and tissue inhibitors of metalloproteinases in human fetal testis and ovary. Mol Hum Reprod 7:641-8

Saccone S, Cacciò S, Perani P, Andreozzi L, Rapisarda A, Motta S, and Bernardi G. (1997) Compositional mapping of mouse chromosomes and identification of the gene-rich regions. Chromosome Res. 5:293-300

Sambrook J, Russell D (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schmahl J, Yao HH, Pierucci-Alves F, Capel B (2003) Colocalization of WT1 and cell proliferation reveals conserved mechanisms in temperature-dependent sex determination. Genesis 35:193-201

Schmegner C, Hameister H, Vogel W, Assum G (2007) Isochores and replication time zones: a perfect match. Cytogenet Genome Res 116:167-72

Seebacher F, Grigg GC, Beard LA (1999) Crocodiles as dinosaurs: behavioural thermoregulation in very large ectotherms leads to high and stable body temperatures. J Exp Biol 202:77-86

Seebacher F, Shine R (2004) Evaluating thermoregulation in reptiles: the fallacy of the inappropriately applied method. Physiol Biochem Zool 77:688-95

Seymour RS, Bennett-Stamper CL, Johnston SD, Carrier DR, Grigg GC (2004) Evidence for endothermic ancestors of crocodiles at the stem of archosaur evolution. Physiol Biochem Zool 77:1051-67

Shawlot W, Behringer RR (1995) Requirement for Lim1 in head-organizer function. Nature 374:425-30

Shedlock AM, Botka CW, Zhao S, Shetty J, Zhang T, Liu JS, Deschavanne PJ, Edwards SV (2007) Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. Proc Natl Acad Sci U S A 104:2767-72

Shibata K, Takase M, Nakamura M (2002) The Dmrt1 expression in sex-reversed gonads of amphibians. Gen Comp Endocrinol 127:232-41

Shoemaker CM, Queen J, Crews D (2007) Response of candidate sex-determining genes to changes in temperature reveals their involvement in the molecular network underlying temperature-dependent sex determination. Mol Endocrinol 21:2750-63

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-4

Smith CA, Sinclair AH (2001) Sex determination in the chicken embryo. J Exp Zool 290:691-9

Smith CA, Smith MJ, Sinclair AH (1999) Gene expression during gonadogenesis in the chicken embryo. Gene 234:395-402

Spotila LD, Spotila JR, Hall SE (1998) Sequence and expression analysis of WT1 and Sox9 in the red-eared slider turtle, Trachemys scripta. J Exp Zool 281:417-27

Swofford D (2003) PAUP*: pylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, Mass

Takada S, DiNapoli L, Capel B, Koopman P (2004) Sox8 is expressed at similar levels in gonads of both sexes during the sex determining period in turtles. Dev Dyn 231:387-95

Thomas RB, Vogrin N, Altig R (1999) Sexual and seasonal differences in behavior of Trachemys scripta (Testudines: Emydidae). Journal of Herpetology 33:511-515

Torres Maldonado LC, Landa Piedra A, Moreno Mendoza N, Marmolejo Valencia A, Meza Martinez A, Merchant Larios H (2002) Expression profiles of Dax1, Dmrt1, and Sox9 during temperature sex determination in gonads of the sea turtle Lepidochelys olivacea. Gen Comp Endocrinol 129:20-6

Valenzuela N, Lance V (2004) Temperature-Dependent Sex Determination in Vertebrates. Smithsonian Institution, Washington DC

Varriale A, Bernardi G (2006) DNA methylation in reptiles. Gene 385:122-7

Vinogradov AE (2005) Noncoding DNA, isochores and gene expression: nucleosome formation potential. Nucleic Acids Res 33:559-63

Volff JN, Kondo M, Schartl M (2003) Medaka dmY/dmrt1Y is not the universal primary sex-determining gene in fish. Trends Genet 19:196-9

Walkley NA, Demaine AG, Malik AN (1996) Cloning, structure and mRNA expression of human Cctg, which encodes the chaperonin subunit CCT gamma. Biochem J 313 (Pt 2):381-9

Webster MT, Axelsson E, and Ellegren H. (2006) Strong regional biases in nucleotide substitution in the chicken genome. Mol. Biol. Evol. 23:1203-1216

Western PS, Harry JL, Marshall Graves JA, Sinclair AH (2000) Temperature-dependent sex determination in the American alligator: expression of SF1, WT1 and DAX1 during gonadogenesis. Gene 241:223-32

Western PS, Sinclair AH (2001) Sex, genes, and heat: triggers of diversity. J Exp Zool 290:624-31

Wibbels T, Bull JJ, Crews D (1991) Chronology and morphology of temperature-dependent sex determination. J Exp Zool 260:371-81

Wibbels T, Crews D (1994) Putative aromatase inhibitor induces male sex determination in a female unisexual lizard and in a turtle with temperature-dependent sex determination. J Endocrinol 141:295-9

Wibbels T, Gideon P, Bull JJ, Crews D (1993) Estrogen- and temperature-induced medullary cord regression during gonadal differentiation in a turtle. Differentiation 53:149-54

Wilusz JE, Freier SM, Spector DL (2008) 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. Cell 135:919-32

Wolfe KH, Sharp PM, Li WH (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283-5

Yntema C (1968) A series of stages in the embryonic development of Chelydra serpentina. J Morph 125:219-252

Zarkower D (2001) Establishing sexual dimorphism: conservation amidst diversity? Nat Rev Genet 2:175-85

Zhang CT, Wang J, and Zhang R. (2001) A novel method to calculate the G+C content of genomic DNA sequences. J. Biomol. Struct. Dyn. 19:333-341

Zhang CT, Zhang R (2004) Isochore structures in the mouse genome. Genomics 83:384-94

Zhao D, McBride D, Nandi S, McQueen HA, McGrew MJ, Hocking PM, Lewis PD, Sang HM, Clinton M (2010) Somatic sex identity is cell autonomous in the chicken. Nature 464:237-42

Zoubak S, Clay O, Bernardi G (1996) The gene distribution of the human genome. Gene 174:95-102

Zug G, Vitt L, Caldwell J (2001) Herpetology: An Introductory Biology of Amphibians and Reptiles. Academic Press, San Deigo, CA

BIOGRAPHICAL SKETCH

Jena Lind Chojnowski is the biological daughter of JoAnn Langer and Gary Chojnowski and the step daughter of Edward Langer. She grew up in Florida with her mother, stepfather, and older brother, Gary Alan Chojnowski. While excelling academically in high school, two subjects caught her interest: biology and history. Since she didn't know exactly what she wanted to do with her educational passions she pursued both at the University of Florida where she was a zoology major and a history minor. She combined her favorite subjects in her first internship which was at the Florida Museum of Natural History in the department of Zooarcheology while continuing her education. Learning that archeology was not for her and realizing that molecular questions fascinated her after taking a Genetics course she did her next internship in the Department of Anthropology looking at the molecular side of human history. After graduation in 2002, Jena took a job as a full time technician looking at avian phylogenetics in the Department of Zoology and left her historical pursuits behind. History was still a fascinating subject to her but it has in time become a hobby instead of a career. Though she was working on birds as a lab technician, a subject not very dear to her heart, she continued to fall in love with the genetics and molecular side of things. Narrowing her interests to molecular biology, she decided to continue in research and attend graduate school for a PhD. As time passed Jena's interests have shifted to and fro but always continued on the pathway of molecular biology. She graduates with her PhD in August 2010 and will continue with her pursuits. Every single one.