Putative Independent Evolutionary Reversals from Genotypic to Temperature-Dependent Sex Determination are Associated with Accelerated Evolution of Sex-Determining Genes in Turtles

Robert Literman1 · Alexandria Burrett1 · Basanta Bista1 · Nicole Valenzuela1

Abstract

The evolutionary lability of sex-determining mechanisms across the tree of life is well recognized, yet the extent of molecular changes that accompany these repeated transitions remain obscure. Most turtles retain the ancestral temperature-dependent sex determination (TSD) from which multiple transitions to genotypic sex determination (GSD) occurred independently, and two contrasting hypotheses posit the existence or absence of reversals back to TSD. Here we examined the molecular evolution of the coding regions of a set of gene regulators involved in gonadal development in turtles and several other vertebrates. We found slower molecular evolution in turtles and crocodilians compared to other vertebrates, but an acceleration in Trionychia turtles and at some phylogenetic branches demarcating major taxonomic diversification events. Of all gene classes examined, hormone signaling genes, and Srd5a1 in particular, evolve faster in many lineages and especially in turtles. Our data show that sex-linked genes do not follow a ubiquitous nor uniform pattern of molecular evolution. We then evaluated turtle nucleotide and protein evolution under two evolutionary hypotheses with or without GSD-to-TSD reversals, and found that when GSD-to-TSD reversals are considered, all transitional branches irrespective of direction, exhibit accelerated molecular evolution of nucleotide sequences, while GSD-to-TSD transitional branches also show acceleration in protein evolution. Significant changes in predicted secondary structure that may affect protein function were identified in three genes that exhibited hastened evolution in turtles compared to other vertebrates or in transitional versus non-transitional branches within turtles, rendering them candidates for a key role during SDM evolution in turtles.

Keywords Genotypic sex determination · Sex determination mechanisms · Substitution rate · Temperature-dependent sex determination · Turtles

Introduction

Among vertebrates, various sex-determining mechanisms (SDMs) direct the developmental fate of individuals towards the male or female condition (Beukeboom and Perrin 2014), controlled primarily through the interaction of two major components: (1) the genes, whose products in a given developmental or genotypic background direct the morphological differentiation of bipotential gonads into testes or ovaries, and (2) the environmental cues that individuals experience throughout development, which may in some cases affect the deployment of those gene products. In many species, including all studied mammals, birds, snakes, and amphibians, the sex determination trigger is the individual’s genomic content [genotypic sex determination (GSD)] (Valenzuela and Lance 2004). Most commonly among vertebrates, these sex-specific genotypic differences are found in sex chromosomes that contain the sex-determining regions, which are present in one sex while absent in the other, or present in different doses between males and females. Sex chromosome systems often include male heterogamety (e.g., XX females, XY males) and female heterogamety (e.g., ZZ males, ZW females). In other species, including some fishes and squamates, most turtles, and all known crocodilians, no sex-limited genomic content exists and instead, sex is determined...
by environmental cues experienced during development (environmental sex determination), the most common cue of which is temperature [temperature-dependent sex determination (TSD)] (Ashman et al. 2014; Valenzuela and Lance 2004). Components of both the genotype and the environment may work within some taxa with intermediate mechanisms between the GSD and TSD extremes (Sarre et al. 2004; Valenzuela et al. 2003), such as taxa with GSD susceptible to environmental influence (Lagomarsino and Conover 1993; Quinn et al. 2007; Shine et al. 2002); while in other taxa, populations may possess contrasting GSD SDMs (Uno et al. 2008).

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

Most vertebrates (GSD or TSD) share a core suite of genes which are integral for the development of the sexual phenotype (Beukeboom and Perrin 2014; Valenzuela and Lance 2004). Individual genes can be classified as male-promoting or female-promoting based on their role in the developing gonads and whether these roles are fairly conserved among species (e.g., Aromatase [Cyp19a1], Dmrt1, Sox9 genes) or not (e.g., Sf1 [Nr5a1]) (Beukeboom and Perrin 2014; da Silva et al. 1996; Raymond et al. 2000; Valenzuela et al. 2013). Indeed, the relative functions and deployment of these genes have evolved among vertebrate lineages, with different genes attaining the top-most position within the sex determination network in different systems (Beukeboom and Perrin 2014; Valenzuela and Lance 2004). Network changes include shifts in the timing of gene activation (Valenzuela et al. 2013) or genes shifting positions in the regulatory pathways between roles as master regulators or downstream actors (Cutting et al. 2013). Alternatively, temperature may become the key regulator of one or more genes at a crucial developmental period [thermosensitive period (TSP)], tipping the balance towards a male or female fate in TSD taxa. As SDMs evolve among lineages and genes and proteins shuffle their task within the sex determination network, parallel molecular changes could be expected to accrue in two major genomic regions. First, molecular changes in the cis- and trans- regulatory regions, which define the expression landscape for a given gene, can lead to changes in the timing and conditions under which a gene may be expressed or silenced. Second, nucleotide changes in the functional domains of protein coding sequences themselves may alter the molecular activity of a gene product including changes in enzymatic rates, co-factor binding identity, and efficiency, as well as temperature’s effect on 3D protein conformation (Bond et al. 1998; Kumar et al. 2000; Marnell et al. 2003).

SDM evolutionary transitions occurred in fishes, geckos and agamid lizards, and turtles (Sabath et al. 2016; Sarre et al. 2011; Valenzuela and Lance 2004). Molecularly, TSD-to-GSD transitions that involved sex chromosome evolution would be characterized by the appearance (or translocation) of a master sex-determining gene onto a proto-sex chromosome, whose overall effect on the sex determination network would outweigh the ancestral effect of incubation temperature, as described in many taxa. Conversely, transitions from GSD-to-TSD would involve either (1) molecular changes that lead to temperature sensitivity on gene expression or on the activity of gene products at key points in this network or (2) the cooption of a novel thermosensitive element or pathway into the sex determination network. Fewer transitions from a GSD ancestor towards TSD are documented in vertebrates (Pokorna and Kratochvíl 2009; Sabath et al. 2016), perhaps because sex chromosomes are an ‘evolutionary trap’ (Pokorna and Kratochvíl 2009) given that during the transition to TSD, YY, or WW individuals would be produced that may be suboptimal or lethal. Alternatively, GSD-to-TSD transitions are prevented perhaps because more challenging changes may be required to de-differentiate sex chromosomes back into autosomes, a rare phenomenon that can also have high fitness cost (Vicoso and Bachtrog 2013). Finally, certain life history traits such as longevity may render TSD-to-GSD transitions more likely in some lineages as shorter life-spans accentuate the negative effects induced by TSD on population dynamics, such as sex ratio skews (Sabath et al. 2016).

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

Here we examine the molecular evolution of a subset of 15 genes in the vertebrate sex determination network (transcription factors, hormone signaling genes, WNT signaling genes, and temperature-sensing genes), using turtles as a focal group to test whether lineages which have undergone SDM transitions (TSD-to-GSD or GSD-to-TSD) are characterized by higher rates of nucleotide or amino acid substitution in the target sex determination genes compared to turtle and other major vertebrate lineages where no SDM transition occurred. This comparative approach permits us to investigate turtle evolutionary rates in a broader phylogenetic context among major amniote clades to illuminate the molecular underpinnings of SDM evolution. We then examine changes in the predicted protein secondary structure within functional domains, which may alter protein activity for genes that exhibited exceptional evolution in turtles. The scarcity of comparable upstream sequences across the focal taxa precluded conducting the same analyses on regulatory regions and thus, we focus on protein coding sequences whose evolution could profoundly affect the molecular architecture of gonadal development during SDM transitions.

### Methods

#### Sample and Data Collection

Out of the ~60 known genes in the vertebrate gonadal development network (Mizoguchi and Valenzuela 2016) coding DNA sequences (CDS) from 15 genes (Table 1) were collected for 25 amniote species from various public databases for which CDS data were available (Fig. 1; Table 2), complemented with sequences for a number of turtle species we obtained during parallel studies (Literman et al. 2017; Radhakrishnan et al. 2017b; Shaffer et al. 2013), including both RNA-Seq and DNA-Seq data (Table 2). These amniotes were chosen because of the availability of well-annotated sequences of the genes of interest which enabled comparison to turtles (our focal clade of interest). Samples of *Apalone spinifera, Chrysemys picta, Staurotypus triporcatus, Glyptemys insculpta*, and *Carettochelys insculpta* derive from privately owned, pet trade or wild individuals as described elsewhere (Literman et al. 2017; Montiel et al. 2017; Radhakrishnan et al. 2017b). *Emydura macquarii* tissues were collected in the wild as part of other studies conducted at the University of Canberra (Australia) under appropriate permits. *Podocnemis expansa* samples were collected by FUDECI (Venezuela) from the wild as part of another study approved by local authorities, and imported under CITES permits. All procedures were approved by the IACUC of Iowa State University, University of Northern Iowa, and University of Canberra.

Data were available for all genes in all taxa with the exception of the *Rspo1* gene, which could not be found for the snake *Ophiophagus hannah* in any available datasets. In order to improve the quality of the sequence data and to

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Gene name</th>
<th>Gene class</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Ar</em></td>
<td>Androgen receptor</td>
<td>Hormone signaling</td>
</tr>
<tr>
<td><em>Cirbp</em></td>
<td>Cold-inducible RNA-binding protein</td>
<td>Temperature signaling</td>
</tr>
<tr>
<td><em>Ctnnb11</em></td>
<td>Beta-catenin</td>
<td>WNT signaling</td>
</tr>
<tr>
<td><em>Cyp19a1 (Aromatase)</em></td>
<td>Aromatase</td>
<td>Hormone signaling</td>
</tr>
<tr>
<td><em>Dmrt1</em></td>
<td>Doublesex and Mab-3-related transcription factor 1</td>
<td>Transcription factor</td>
</tr>
<tr>
<td><em>Esr1</em></td>
<td>Estrogen receptor 1</td>
<td>Hormone signaling</td>
</tr>
<tr>
<td><em>Esr2</em></td>
<td>Estrogen receptor 2</td>
<td>Hormone signaling</td>
</tr>
<tr>
<td><em>Hsf2</em></td>
<td>Heat-shock factor 2</td>
<td>Temperature signaling</td>
</tr>
<tr>
<td><em>Lhx9</em></td>
<td>LIM homeobox 9</td>
<td>Transcription factor</td>
</tr>
<tr>
<td><em>Nr5a1 (Sf1)</em></td>
<td>Steroidogenic factor 1</td>
<td>Transcription factor</td>
</tr>
<tr>
<td><em>Rspo1</em></td>
<td>R-spondin 1</td>
<td>WNT signaling</td>
</tr>
<tr>
<td><em>Sry9</em></td>
<td>SRY box 9</td>
<td>Transcription factor</td>
</tr>
<tr>
<td><em>Srd5a1</em></td>
<td>Steroid 5-alpha reductase 1</td>
<td>Hormone signaling</td>
</tr>
<tr>
<td><em>Wnt4</em></td>
<td>Wingless-type MMTV integration site family, member 4</td>
<td>WNT signaling</td>
</tr>
<tr>
<td><em>Wt1</em></td>
<td>Wilms tumor 1</td>
<td>Transcription factor</td>
</tr>
</tbody>
</table>
Fig. 1 Phylogenetic relationships of 25 vertebrate species analyzed in this study. Branch lengths are scaled to TimeTree consensus divergence times. Branch thicknesses and darkness are scaled to average nucleotide substitution rates across all genes. SDMs are denoted in brackets. Specific branches discussed in the text are numbered (1 = Placental mammal root branch; 2 = Iguania lizard root branch; 3 = Reptile root branch; 4 = Trionychia root branch; 5 = Neoaves root branch). Temp. = temperature sex reversal or TSD reports. (Color figure online)

Table 2 Species used in this study along with data sources

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

DNA-NGS = inhouse generated next-generation DNA sequencing data (short read Illumina HiSeq data)
ensure that similar isoforms were compared for each of these genes, all coding sequences were manually extracted from publicly available genomes (when present) rather than from pre-existing gene annotations or other computer-based gene predictions. For species with additional publicly accessible sequence data, any existing data gaps in the target genes were filled in via the NCBI Short Reach Archive (SRA) or through published mRNA sequences from the target taxa downloaded from Genbank. A dated pruned vertebrate phylogeny of all species examined in this study was generated using the median time divergence estimates from the TimeTree database, which are calculated from numerous published studies (Hedges et al. 2006) (Fig. 1).

**Data Processing**

For each gene, two datasets were generated: one dataset included sequence alignments from all 25 species (‘All Species’), and a second dataset included alignments from the ten turtle taxa exclusively (‘Turtles’). In all cases, nucleotide sequences were translationally aligned using MUSCLE with default parameters as implemented in Geneious v.9 (Kearse et al. 2012). In order to minimize the impact of taxa-specific indels, all alignments were visually assessed to ensure that homologous regions were aligned, and misalignments were corrected manually. The lengths and sequence identity of the final alignments for the ‘All Species’ and ‘Turtles’ datasets after gaps were removed differ from each other (Table S1). Alignments were generated for both nucleotide and amino acid sequences.

For each alignment, maximum likelihood was applied to determine the most likely model of substitution using jModelTest v.2.1.4 (Darriba et al. 2012) for nucleotide alignments and ProtTest v.2.4 (Abascal et al. 2005) for amino acid alignments. Alignments were imported into MEGA v.7 (Kumar et al. 2016) and using the estimated model parameters, the number of substitutions per site along each branch was estimated using maximum likelihood against a well-supported species phylogeny (Fig. 1). Substitutions per site per million years (SSM) for each branch were then calculated by dividing the number of substitutions per site (generated from MEGA) by the divergence time estimates from TimeTree (Ho et al. 2005; Smith and Donoghue 2008) to obtain a rate of molecular evolution that was used in all downstream statistical tests. This time-corrected approach accounts for the impact of long branches when comparing the extent of changes among lineages with differing time of divergence, because an older lineage could accumulate more changes even if the rate of substitutions is identical to that of a younger lineage. To estimate neutral nucleotide substitution rates for different taxonomic groups, the nucleotide data were also analyzed using only the third codon positions (note that this is an approximation since not all third-position substitutions are silent). Sites containing gaps were eliminated from the analysis.

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

**Data Analysis**

Parallel analyses were performed for the nucleotide and the amino acid alignments. Due to unequal variances in rates among groups, non-parametric statistical tests (described below) were performed to compare the differences among groups. For both the ‘All Species’ and ‘Turtle’ datasets, pairwise Steel–Dwass tests were performed to test whether substitution rate changes could be explained by phylogenetic history, and to test whether gene classes differed in their substitution rates within taxonomic groups. These tests were implemented using JMP, Version 12.0.1 (SAS Institute Inc., Cary, NC, 2015). Significance of the results from the Steel–Dwass test is reported at a Bonferroni-corrected alpha to account for multiple comparisons.

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

\[
Z = \frac{SSM(Rate of branch) - SSM(Rate mean for gene)}{\sigma SSM (Rate STDEV for gene)}.
\]

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

\[
Z = \frac{SSM(Rate mean for gene) - SSM (Rate mean for all genes)}{\sigma SSM (Rate STDEV for all genes)}.
\]

For the ‘Turtle’ dataset, Wilcoxon/Mann–Whitney U tests (2-way comparisons) and Steel–Dwass Tests (3-way comparisons) were performed to investigate whether phylogenetic branches characterized by a sex determination transition (TSD-to-GSD or GSD-to-TSD) differed in substitution rate relative to non-transitional branches or to each other (TSD-to-GSD versus GSD-to-TSD). Each branch on the turtle phylogeny was scored as transitional or non-transitional (TSD-to-GSD or GSD-to-TSD) differed in substitution rate relative to non-transitional branches or to each other (TSD-to-GSD versus GSD-to-TSD).
and (2) a hypothesis that reconstructs five transitions from TSD-to-GSD (two of which overlap with the previous hypothesis) plus two reversals back to TSD from GSD in the lineages represented by *C. insculpta* and *P. expansa* (Valenzuela and Adams 2011) (Fig. 2). Considering both hypotheses, the focal taxa included in this study encompass representative species for four of the five TSD-to-GSD transitions predicted under each hypothesis. All comparisons, including Z-score analyses, Mann–Whitney U tests, and Steel–Dwass tests were explored at $\alpha = 0.05$ (threshold $Z > 1.644$) to detect trends, but significance was assessed at a Bonferroni-corrected $\alpha = 7.14E^{-3}$ (threshold $Z > 2.45$) for ‘All Species’ tests, and at $\alpha = 5.56E^{-3}$ (threshold $Z > 2.54$) for the ‘Turtles’ dataset.

Although our main objective was to test whether transitions in SDM were accompanied by an acceleration in the evolution of the genes of interest, which requires the use of an explicit time component as described above, we also analyzed the accumulation of substitutions in these DNA sequences among taxonomic groups and SDM transition branches using CODEML as implemented in the program ete3 (Huerta-Cepas et al. 2016), where divergence time estimates are not incorporated explicitly. Using the same starting alignments, values for $dN$ and $dS$, and $dN/dS$ were calculated for each branch of the tree and comparisons between groups were performed using Mann–Whitney U tests for comparisons between two groups, or Steel–Dwass tests for comparisons among more than two groups. Higher values of $dN$ and $dS$ are indicative of increases in non-synonymous and synonymous substitution rates, respectively. Complete methods for this analysis are described in the Supplementary Material. Because results from CODEML were concordant with the findings of our first approach, below we detail only the CODEML results where differences were detected.

**Results**

**Substitution Rates are Slowest in Turtles and Crocodilians and Fastest in Mammals Among Amniotes**

For the ‘All Species’ dataset containing 25 amniotes, the analysis across all 15 genes showed significant differences in nucleotide substitution rates at the third codon positions among taxonomic groups indicative of differences in the neutral rates of evolution (Table 3, detailed statistical results in Table S2). Pairwise comparisons using the Steel–Dwass test revealed a higher substitution rate in mammals than in any other group ($p < 0.002$), while crocodilians and turtles exhibited a similar and lower rate than all other groups (Table 3). A similar observation was made when considering all codon positions, although in this case turtle genes exhibited a significantly faster rate than crocodilians ($p < 0.001$) (Table 3, detailed statistical results in Table S3). At the amino acid level, mammals, birds, and squamates formed a group with substitution rates ~3× faster than turtles and crocodilians ($p < 0.005$) (Table 3, detailed statistical results in Table S4). Results from the CODEML analysis were concordant with the time-dependent analysis in terms of $dS$ (silent) and $dN$ (replacement) substitutions (Table S5).
Among Turtles, Substitution Rates are Slowest in the Emydidae Species

Turtle branches were also analyzed separately using the ‘Turtles’ dataset (N = 10), which are more complete alignments than the “All Species” dataset (turtle sequences were longer because they contained fewer gap positions that needed to be removed), and because mammals, squamates, and birds exhibited significantly faster overall substitution rates relative to turtles which could obscure biologically important shifts among turtles. Clade-level comparisons revealed that turtle sub-groups differed in third-position substitution rates, which were slower in Emydidae (Glyptemys, Chrysemys, plus Trachemys) than in other turtles, while no rate differences were detected between Trionychia (Apalone, Pelodiscus, plus Carettochelys), Americhelydia (Staurotypus plus Chelonia), or Pleurodiran turtles (Emydura plus Podocnemis) (Table 4, detailed statistical results in Table S6). Overall nucleotide substitution rate mimicked the third-position substitution rate differences qualitatively, which was slower overall in Emydidae than in Trionychia and Pleurodira (Table 4, detailed statistical results in Table S7). At the amino acid level, the rates in Trionychia were faster than in Emydidae (Table 4, detailed statistical results in Table S8). Emydidae also exhibited a lower dN and dS (Table S9). No other comparisons were significant.

Root Branches Tend to Experience Higher Substitution Rates in More Genes Than Other Branches

For each gene alignment, certain branches on the phylogenetic tree (Fig. 1) display significantly faster nucleotide substitution rates (“fast branches”) relative to the overall average rate for that gene (Table S10). This was true for several root branches across many genes (9 out of 15 genes at the root of the Iguania lizards, 6 out of 10 genes at the root of reptiles, and 7 out of 10 genes for the placental mammal root), whereas in turtles this was true for a smaller subset of genes. For the ‘All Species’ dataset, 8 of 48 branches in the phylogenetic tree exhibited a significantly faster nucleotide substitution rate relative to other branches for at least one gene (Table S10). Despite the lower nucleotide substitution rate observed in turtles as a whole, one turtle branch displayed a significantly faster nucleotide substitution rate for one gene each (Sox9 in the Trionychia root branch), while

Table 3 Overall substitution rates per million years of nucleotide and amino acid sequences across target genes in each taxonomic group, and relative rate (%) compared to turtles

<table>
<thead>
<tr>
<th>Taxonomic Group</th>
<th>Third Codon Position</th>
<th>All Nucleotides</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Substitutions per Myr</td>
<td>% from Turtles</td>
<td>Substitutions per Myr</td>
</tr>
<tr>
<td>Mammals</td>
<td>2.60E−03 [A]</td>
<td>387.5</td>
<td>1.02E−03 [A]</td>
</tr>
<tr>
<td>Squamates</td>
<td>2.28E−03 [B]</td>
<td>326.7</td>
<td>8.79E−04 [B]</td>
</tr>
<tr>
<td>Birds</td>
<td>1.73E−03 [B]</td>
<td>223.5</td>
<td>6.64E−04 [B]</td>
</tr>
<tr>
<td>Crocodilians</td>
<td>4.37E−04 [C]</td>
<td>−18.0</td>
<td>1.67E−04 [D]</td>
</tr>
</tbody>
</table>

For each comparison (third codon, all nucleotides, and amino acids), statistically indistinguishable rates share bracketed letters. Myr=million years

Table 4 Overall substitution rates per million years of nucleotide and amino acid sequences for all genes examined in each focal turtle clade, and relative rate (%) compared to Emydidae

<table>
<thead>
<tr>
<th>Taxonomic Group</th>
<th>Third Codon Position</th>
<th>All Nucleotides</th>
<th>Amino Acids</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Substitutions per Myr</td>
<td>% from Emydidae</td>
<td>Substitutions per Myr</td>
</tr>
<tr>
<td>Pleurodira</td>
<td>5.34E−04 [A]</td>
<td>96.3</td>
<td>2.53E−04 [A]</td>
</tr>
<tr>
<td>Americhelydia</td>
<td>5.12E−04 [A]</td>
<td>88.2</td>
<td>2.38E−04 [AB]</td>
</tr>
</tbody>
</table>

For each comparison (third codon, all nucleotides, and amino acids), statistically indistinguishable rates share bracketed letters. Myr=million years
the trend for \( Lhx9 \) in \( C. insculpta \) was not significant at the Bonferroni-adjusted alpha (Table S10).

Somewhat similar results were obtained at the protein level, where 12 of 48 branches had at least a single gene with a faster than average amino acid substitution rate (Table S11). This list was topped by the root of Iguania lizards (5 of 15 genes), placental mammals (5 of 15 genes), and the reptile root branch (4 of 15 genes). In turtles, additional fast branches were detected at the amino acid level beyond those detected at the nucleotide level. Indeed, four turtle branches exhibited faster amino acid substitution rates for at least one gene: \( C. insculpta \) (\( Lhx9 \)), and the Americhelydia root branch (\( Wt1 \)) (Table S11), while the tendency detected in the Trionychia root branch (\( Esr1, Hsf2, Sox9 \)) and the turtle root branch (\( Cirbp \)) is not significant after Bonferroni correction. Using the turtle only dataset also revealed faster than average nucleotide substitution rates in 2 of 18 branches relative to other turtle branches for at least one gene (Table S12). Indeed, 10 of 15 genes evolved at a faster than average rate in the Trionychia root branch, while this was true only for a single gene in any other branch. At the protein level, 6 of 18 branches exhibited a faster amino acid substitution rate for at least one gene (Table S13). The Trionychia root branch showed this pattern for 5 of 15 genes.

**Genes on Z (But Not X) Sex Chromosomes Evolve Faster at the Amino Acid Level**

Three target genes in this study, \( Ar, Dmrt1 \), and \( Cttnb1 \), are sex-linked (located in the X or Z chromosomes) in certain focal taxa, and their protein sequences evolve faster when they are Z-linked than on taxa where they are autosomal or X-linked. For instance, \( Dmrt1 \) is Z-linked in birds (Nanda et al. 1999), and three of the five avian branches (\( Taeniopygia guttata, Falco peregrinus \), and the Neoves root branch) have a significantly faster amino acid substitution rate than all other branches (although interestingly, this faster rate is absent in chicken [\( Gallus gallus \)] or the bird root branch) (Table S11). Second, \( Cttnb1 \) is Z-linked in snakes (Vicoso et al. 2013), and \( O. hannah \) snakes show a significantly faster amino acid substitution rate relative to all other species, including the other snake examined, \( Python molurus \). In contrast, \( Ar \) is X-linked in mammals (Ross et al. 2005), yet no mammalian branch exhibits faster amino acid substitution rates (Table S11).

**Hormone Signaling Genes Evolved Faster in Most Vertebrates and \( Srd5a1 \) in Mammals and Some Turtles**

Most gene classes display similar rates of substitution at the nucleotide and amino acid levels across our target taxa with some noticeable exceptions (Tables S14–S17). Namely, for
relative to TSD-to-GSD transitional branches albeit not significantly \((p = 0.0406)\). Notably, no significant differences in the amino acid substitution rate were detected across genes between non-transitional branches and TSD-to-GSD branches \((p = 0.0862, \text{Fig. 3})\), suggesting that the trend in the transition-dependent substitution rate at the protein level is driven by the GSD-to-TSD reversal branches. This pattern was also observed using CODEML (Table S21). Namely, GSD-to-TSD reversal branches showed higher dN and dS rates relative to non-transitional branches (> 500 and > 300%, respectively) as well as to TSD-to-GSD branches (> 400 and > 250%, respectively), whereas TSD-to-GSD transitional branches and non-transitional branches displayed similar rates. Broken down by individual genes, nucleotide sequences of \textit{Dmrt1}, \textit{Hsf2}, \textit{Sox9}, and \textit{Wnt4} exhibited a non-significant tendency to evolve faster in transitional branches relative to non-transitional branches \((p < 0.05)\), and the same was true for the amino acid sequences of HSF2, RSPO1, and SOX9 \((p < 0.05)\). Further transition type-specific comparisons were precluded because there are only two GSD-to-TSD branches and four TSD-to-GSD branches.

**Potentially Functional Evolutionary Changes Accrued in Turtle Proteins**

We searched for target proteins in our dataset that either (1) evolve faster in a transitional turtle branch compared to all other species or, (2) tended to evolve faster in transitional versus non-transitional turtle branches, and subjected them to secondary protein structure predictions in order to identify amino acid substitutions in transitional turtle branches that may induce functional changes. Structural changes within UniProt functional domains were detected in the HSF2 protein, which evolved faster in the Trionychia turtles relative to all other turtle clades, and displayed a similar tendency in turtle transitional branches relative to non-transitional branches. Likewise, we detected structural changes within functional domains of the RSPO1 protein, which tended to evolve faster in transitional branches relative to non-transitional branches, and in the \textit{LHX9} protein, which evolved faster in \textit{C. insculpta} relative to all other species (Fig. 4).

Specifically, a series of amino acid substitutions that change the secondary protein structure were identified in turtle transitional branches in both the positive and negative transactivation domains of the HSF2 protein, which form its C-terminal domain and contribute to transcriptional activation of target genes (Yoshima et al. 1998). In the positive transactivation domain 1, which binds co-regulators during cellular stress (Fig. 4a), \textit{P. expansa} has a single substitution predicted to cause the loss of two helices present in all other species, while in a separate site, \textit{C. insculpta} has a single substitution that is predicted to join two shorter helices found in all other species into a single, longer helix. In the associated negative regulator domain, which keeps...
HSF2 inactive during non-stress conditions, E. macquarii, P. expansa, and C. insculpta share a unique predicted secondary structure characterized by the addition of two short alpha helices and notably, these result from three different and independent mutations in each lineage. In the positive transactivation domain 2 (Fig. 4b), S. triporcatus accrued several amino acid substitutions predicted to drastically change its secondary structure from that of other turtles by inducing multiple alpha helices, while G. insculpta has a single substitution which also induces a novel helix among turtles in this domain. In the corresponding negative regulator domain, P. expansa, A. spinifera, and P. sinensis experienced a reduction or loss of alpha helices that exist in all other turtle species.

The 3′ end of the downstream LIM domain of LHX9 in turtles generally contains a 10-amino acid long alpha helix which is disrupted in P. expansa and C. insculpta, the representatives of the two putative GSD-to-TSD reversals in turtles (Fig. 4c). In P. expansa, an inserted short beta strand splits the helix, while in C. insculpta additional substitutions are predicted to yield a unique secondary structure among turtles with a highly truncated alpha helix and the expansion of the beta sheet predicted in P. expansa. Further upstream substitutions in C. insculpta cause the shifting of a beta sheet relative to all other turtles. Additionally, the 5′ end of the upstream LIM domain in C. insculpta also contains a split helix, where in all other turtles the helix is unbroken.

Finally, the thrombospondin-1 domain of the RSPO1 protein in two XX/XY turtle species (E. macquarii and S. triporcatus) shares a secondary structure unique among turtles, lacking an alpha helix in the C-terminal end of the domain. Notably, this pattern evolved through two independent amino acid substitutions at separate sites in this domain in each lineage (Fig. 4d). S. triporcatus has also accumulated substitutions in the second of two upstream Furin-like repeat domains which cause both the truncation of an alpha helix, which is longer in all other species, as well as the insertion of a helix at the 3′ end of the domain, which is unique to S. triporcatus.

---

**Fig. 4** Amino acid substitutions in functional domains of sex determination genes in turtle lineages experiencing shifts in sex determination. Boxes highlight focal regions with substitutions. SDMs for each species noted in brackets. TSD* = GSD-to-TSD reversal branch under Valenzuela and Adams (2011). a, b Transactivation and negative regulatory domains of HSF2. c LIM domains of LHX9. d FU and TSP1 domains of RSPO1. Zigzags denote trimmed sequence between domains. aa Amino acids trimmed.
Discussion

Numerous independent transitions between GSD and TSD have occurred throughout vertebrate evolution (Beukeboom and Perrin 2014; Valenzuela and Lance 2004), yet, the gene network that regulates sex determination has remained remarkably similar in its composition, whereas the roles of this shared set of genes have shuffled as SDMs evolve (Cutting et al. 2013; Valenzuela et al. 2013). Here we advance our understanding of other changes that have accompanied SDM transitions, as we detected substantial molecular evolution among major vertebrate clades in the coding sequence of a subset of 15 genes involved in sex determination, for which adequate CDS data were available across the focal taxa (25 selected amniotes). Among those, we highlight in particular the changes identified in turtle lineages which have experienced SDM transitions, including predicted structural changes of potential functional importance. Results using dN and dS analysis were mostly concordant with our time-explicit approach and thus we discuss CODEML results only for cases were differences were observed. Analysis of the upstream regulatory regions associated with gene activation and silencing of these genes was precluded due to lack of data in the low-coverage genomes available for non-model taxa. We note also that data scarcity prevented a broader taxonomic sampling such that our conclusions should be interpreted as working hypotheses to guide future research facilitated by growing genomic information of non-model organisms. Additionally, future genome-wide analyses are warranted to test whether the patterns detected here with a limited set of genes related to sexual development might be more general and present in other regulatory networks.

Genes in the Sex Determination Network Evolve Slowly in Turtles and Crocodilians But Show an Acceleration in Trionychia Turtles

Among the five major vertebrate clades examined, the target genes evolve at a significantly lower rate in the turtles and crocodilians, a pattern observed at both the nucleotide and amino acid levels and accentuated in crocodilians relative to turtles when considering all nucleotides (Table 3). These observations agree with recent genome-wide reports for turtles, Anolis lizards, and crocodilians (Fujita et al. 2011; Green et al. 2014; Shaffer et al. 2013), but counter the expectation that the evolutionary lability of SDMs in turtles, when compared to mammals and birds, might be associated with higher molecular evolution overall. The lower genome-wide substitution rates of crocodilians and turtles have been hypothesized to derive from a combination of their slower metabolic rates and longer generation times relative to endothermic mammals, birds, and to most squamates (Green et al. 2014; Shaffer et al. 2013). Whether population sizes of turtles and crocodilians might also affect their molecular evolution alone or in combination with other life history traits (Bromham 2002) remains unknown.

Despite the relatively low rates of molecular evolution in turtles, significant differences in evolutionary rates were detected among turtle lineages. For instance, both the nucleotide and amino acid substitution rates of the target genes in Emydidae (Fig. 2; Table 2) were slower than in other turtle lineages (85–244% depending on the dataset used, Table 4). Furthermore, the branch leading to the Trionychia turtles (softshell plus pig-nose turtles) (Fig. 2; Table 2) stood out by its faster nucleotide and amino acid substitution rate relative to the turtle average for most genes (Tables S10 and S11), revealing that a major acceleration in molecular evolution accompanied the divergence of this morphologically intriguing lineage. However, whether this hastened molecular evolution is functionally linked to SDM transitions and restricted to genes in the sex-determining network, or whether it is a more generalized phenomenon associated with other ecological, genomic, or morphological innovations in this clade remains to be tested. The higher neutral rate of evolution in Trionychia compared to other turtles supports the notion that this lineage follows a unique evolutionary trajectory. Notably, members of this turtle family underwent a cytogenetic revolution, as they contain the species with the highest numbers of chromosomes among turtles (A. spinifera, P. sinensis: 2n = 66; C. insculpta: 2n = 68) (Badenhorst et al. 2013; Bickham and Legler 1983; Sato and Ota 2001). It is tempting to speculate that perhaps the same drivers might be responsible for the acceleration of molecular evolution and chromosomal fissions observed in this group (Montiel et al. 2016).

Faster Branches Are Mostly Root Branches

Major clade root branches, such as the branches leading to the placental mammals, Iguania lizards, and the root of all reptiles had the most genes with a faster than average substitution rate for both nucleotides and amino acids (Tables S8 and S9), suggesting that the sex determination network experienced major molecular changes during significant taxonomic diversification events. Although appealing, further research is warranted to test the biological relevance of this hypothesis under a more extensive taxonomic sampling than is currently possible, as genome-level data for additional species become available. Our results on the evolutionary rate of sex-determining genes among model and non-model vertebrates complement our knowledge of the evolution of sex chromosome content and gene synteny that...
is well documented among diverse vertebrate clades (Ezaz et al. 2017; Graves 2017; Montiel et al. 2016; Rovatsos et al. 2016) by illuminating the evolution of protein coding sequences that accompanies shifts in SDM.

**Protein Sequences Evolve Faster on Sex Chromosomes, But Not Always**

At first glance, sex-linkage effects were observed at the amino acid level for a number of species. For example, the DMRT1 protein evolved faster in birds where this gene is Z-linked (Nanda et al. 1999), in particular in the branches of the zebra finch (T. guttata), falcon (F. peregrinus), and their root branch (Neoaves) (Table S10). Interestingly however, the root branch for all birds and the branch leading to chicken (G. gallus) have substitution rates more comparable to their autosomal homologs in the other taxa examined, suggesting that DMRT1 evolution has been significantly accelerated in the Neoaves relative to the more basal birds, and that some driver specific to Neoaves, other than the evolution of avian sex determination, might be causal. Second, beta-catenin (CTNNB1) is a highly conserved signaling protein, exhibiting the lowest average amino acid substitution rate of all target genes such that no substitutions were detected in 33 of 48 branches. Moreover, Ctnnb1 is the only gene that exhibited a lower dN/dS ratio in all taxa compared to all other genes (results not shown) indicating that it is under strong purifying selection. However, Ctnnb1 is Z-linked in snakes (Vicoso et al. 2013) and its protein sequence evolved faster in the king cobra (O. hannah) (Table S10) than in the Indian python P. molurus, which displayed only the substitutions accumulated during the diversification of the Serpentes as a group. These results in birds and snakes suggest that the differential molecular evolution is not uniform for sex-linked genes, but that important lineage-specific effects are also at play. Furthermore, the other sex-linked protein in the analysis, the mammalian X-linked AR (Ross et al. 2005), did not show evidence of ‘Fast X’-type evolution (Table S10). Combined, all these findings refute the absolute generality, even among closely related taxa, of the ‘Fast X’ and ‘Fast Z’ hypotheses, which state that genes residing in sex-limited (non-sexual autosomal) portions of sex chromosomes should evolve faster than their autosomal counterparts due to their smaller effective population size relative to autosomes (Bachtrog et al. 2011; Charlesworth et al. 1987; Mank et al. 2007). Instead, our data suggest that gene-specific, lineage-specific, or other chromosome-specific effects may override the evolutionary dynamics driven by sex-linkage alone. Similarly, a recent study detected chromosomal and lineage effects on molecular evolution when contrasting sex-linked sequences to the same sequences linked to autosomes in turtles and other amniotes (Radhakrishnan et al. 2017b). Alternative drivers of molecular evolution have been proposed, such as GC content and expression levels that may explain the difference in the rate of molecular evolution in mammalian sex-linked genes (Nguyen et al. 2015).

**Fast Genes: the Hormone Signaling Gene Srd5a1 Evolves Faster in More Lineages Than all Other Genes**

Hormone signaling genes were the only gene class exhibiting a consistent tendency for faster evolution at the nucleotide or amino acid levels (Tables S14–S17). On a gene by gene level, Srd5a1 exhibited a trend towards a faster nucleotide substitution rate in the Americhelydia turtles (Fig. 2; Table 2, Table S18) as well as a faster amino acid substitution rate (a trend also observed in mammals and Trionychia turtles, Tables S17 and S18) relative to all other genes for those groups. Srd5a1 is also the only gene that exhibits greater dN/dS ratio in all taxa compared to all other genes (results not shown) indicating that it is under relaxed purifying or positive selection. Srd5a1 encodes a 5-alpha reductase protein which converts weaker androgens to more potent androgen compounds (Chang et al. 2011), a critical step in sexual differentiation towards a male fate (Urbatzka et al. 2007). Deficiencies in human 5-alpha reductase activity are linked to partial or complete sex reversal of genotypic males (XY) at birth (exhibiting intersex or even female phenotype due to low levels of potent androgens) that is reverted partially at puberty when appropriate hormones are synthesized (Wilson et al. 1993). We hypothesize that the increased rate of molecular evolution of Srd5a1 observed here may reflect more profound changes in the hormone signaling pathways or how androgens are utilized during sexual development among taxa, but this hypothesis requires further testing.

**Contrasting Hypotheses of the Evolution of Turtle Sex Determination Affect the Interpretation of Transition-Dependent Shifts in Substitution Rates**

Our ‘Turtle’ dataset permitted us to examine the influence of SDM transitions on molecular evolution, but the same analyses were precluded for the ‘All Species’ dataset. The evolutionary history of SDM transitions among the turtles remains uncertain, and current contrasting evolutionary hypotheses (Fig. 2) were evaluated here (Sabath et al. 2016; Valenzuela and Adams 2011). First, some research reconstructs all transitions in turtle sex determination as occurring from a TSD ancestor to a derived GSD state (Sabath et al. 2016). Under this scenario, no significant change in substitution rate was seen between the transitional and non-transitional branches at either the nucleotide or amino acid levels. However, CODEML detected higher dN and dS rates in transitional branches (Table S20), underscoring that the accumulation of substitutions alone can provide a
misleading estimate of the tempo of evolution when divergence time is not taken into account. The time-dependent rates observed counter the notion that accelerated molecular evolution in the genes examined here is associated with overall turnovers in sex determination. Whether genes in this network, other than those analyzed here, might exhibit such an association remains to be tested. Alternatively, minuscule genetic changes such as point mutations, rather than more dramatic shifts in substitution rate, may alter sex determination profoundly and affect SDM evolution. For instance, the sex determination network of Takifugu fish was hijacked by a single nucleotide polymorphism in the kinase domain of the gene Amhr2 (Kamiya et al. 2012).

Second, another study reconstructed seven total transitions in turtles, five TSD-to-GSD transitions (three of which overlap with the previous hypothesis) plus two GSD-to-TSD ‘reversals’ in the lineages represented by C. insculpta and P. expansa (Valenzuela and Adams 2011). Further, these transition branches display ~20x faster rate of evolution of chromosome number (Valenzuela and Adams 2011), suggesting that although GSD-to-TSD transitions might be rare, turtle genomes are generally more evolutionarily labile in transitional branches. Interestingly, under this hypothesis, we found that transitional branches, irrespective of their direction (TSD-to-GSD or GSD-to-TSD), had a faster nucleotide and amino acid substitution rate than non-transitional branches (Fig. 3), but this did not translate into accelerated protein sequence evolution in TSD-to-GSD branches. However, genes in the GSD-to-TSD ‘reversal’ branches evolved significantly faster at the nucleotide level (~20% faster) relative to the TSD-to-GSD transitional branches, a pattern supported by the dN and dS results (Table S21). This finding is concordant with the idea that regaining TSD during GSD-to-TSD transitions requires a more dramatic rewiring of the sex determination network, perhaps reflecting the difficulty of escaping the evolutionary trap of sex chromosomes (Pokorná and Kratochvíl 2009), and helping explain why these transitions are considerably rarer than TSD-to-GSD transitions, at least in turtles.

When broken down by gene, the nucleotide sequences of Dmrt1, Hsf2, Sox9, and Wnt4, and the protein sequences of HSF2, RSP01, and SOX9 tended to evolve faster in transitional branches. HSF2 is a member of the heat-shock family of proteins, which is able to transduce temperature signals into transcriptional responses (Sarge et al. 1991). RSP01 is an integral signaling molecule in the WNT signaling pathway regulating expression of Ctnnb1 and Wnt4, which helps direct proper ovarian development (Parma et al. 2006). SOX9 is a male-promoting transcription factor in most vertebrates, and in many taxa it is critical for both the initiation and maintenance of the male developmental pathway (da Silva et al. 1996). Because of their identity and function, the accelerated amino acid evolution of these genes in transitional turtle branches renders them important candidates for further study, as they span the key processes of temperature-sensing, hormonal regulation, and gonadal differentiation that underlie sexual development.

**Secondary Protein Structure Changes in Transitional Turtle Branches in Functionally Important Domains**

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

First, the HSF2 protein is involved in stress response, temperature-signal transduction, and spermatogenesis (Garolla et al. 2013; Yoshima et al. 1998). The HSF2 C-terminal domain contains two pairs of transactivation domains that modulate transcription during cell stress through co-regulatory ligand binding, plus two pairs of negative regulation domains that maintain HSF2 inactive when cells are not stressed (Yoshima et al. 1998). These four sub-domains exhibit predicted secondary structure changes with potentially functional consequences specific to transitional turtle branches. For instance, the two representatives of the putative GSD-to-TSD transition lineages (P. expansa and C. insculpta), each have separate single amino acid substitutions in different sites of the HSF2 first transactivation domain that lead to structural changes unique among turtles (Fig. 4a). Specifically, P. expansa’s single substitution deletes two short alpha helices present in all other turtles, while C. insculpta’s creates one longer alpha helix where all other taxa have two helices separated by a beta sheet. On the other hand, E. macquarii (representing a TSD-to-GSD transition lineage), P. expansa, and C. insculpta share a similar secondary structure in the negative regulator domain of the first transactivator domain compared to all other turtles, but caused by different substitutions in each lineage, suggesting that they result from somewhat convergent evolution rather than from shared ancestry. Additionally, the HFS2 second transactivation domain of S. tripuratus (representing a TSD-to-GSD transition) accumulated multiple substitutions, which create several novel alpha helices not present in any other turtle (Fig. 4b), and G. insculpta (representing a TSD-to-GSD transition) exhibits yet another novel alpha helix in this domain. Finally, A. spinifera and P. sinensis, (representing a TSD-to-GSD transition) along with P. expansa, show a drastic reduction or loss of alpha helices in the negative regulator domain for HFS2’s second transactivator present in all other turtle species. We hypothesize that because the regulatory role of HSF2 is
self-modulated via the negative regulator domains, which varies depending on the degree of cell stress (e.g., high temperature), the structural modifications described here could impact how HSF2 is deployed during bouts of high temperature (changes in transactivation domains), and perhaps also under more moderate temperatures (changes in negative regulator domains). This hypothesis requires future functional testing.

Second, LHX9 is a transcription factor which is essential for the formation and general development of the bipotential gonad and typically shows equal levels of activity in both developing male and female embryos among vertebrates (Barske and Capel 2010; Birk et al. 2000; Mazaud et al. 2002; Oréal et al. 2002; Ottolenghi et al. 2001). It activates the NR5A1 protein (also known as SF1) (Shima et al. 2012) whose transcription is evolutionarily labile across vertebrates (Valenzuela et al. 2013). In order to bind co-regulator proteins, LIM-containing transcription factors possess tandem protein-binding LIM domains separated by a linker sequence (Rétaux et al. 1999). Here we found that C. insculpta’s LHX9 is characterized by two predicted structural changes unique among vertebrates. Namely, C. insculpta’s downstream LIM domain has a smaller conserved C-terminal alpha helix and beta strands shifted within the center of the domain, and the N-terminal helix in the upstream LIM domain is also disrupted. This raises the question as to whether such predicted modifications to the protein:protein ligand-binding domain might operate as part of a shifting regulatory network where old interactions are lost, and new molecular relationships may be forged.

Finally, RSPO1 is a critical co-regulator of the WNT/CTNNB1 pathway, primarily involved in ovarian development through Ctnnb1 activation and stabilization (Parma et al. 2006). RSPO1 is also a SOX9 antagonist, working to both promote ovarian development while silencing male developmental pathways (Chassot et al. 2008). RSPO1, as other R-spondin protein family members, contains a thrombospondin-I repeat domain which appears unnecessary to modulate Ctnnb1 (Parma et al. 2006), and its precise role is still unclear. Two independent single amino acid substitutions in the TSP1 domain in E. macquarii and S. triporcatus are predicted to cause the convergent loss of an alpha helix found in all other turtles (Fig. 4d). While the role of the TSP1 domain remains elusive, the furin-like repeat domains appear integral to R-spondins’ ability to modulate the Ctnnb1/Wnt pathway (Kim et al. 2008, 2006). Intriguingly, in S. triporcatus a single amino acid substitution in the second of two tandem Furin-like repeat domains truncates an alpha helix and adds a beta strand, a predicted structure unique among all vertebrates examined. We hypothesize that these modifications may impact RSPO1-mediated regulation of the Ctnnb1/Wnt pathway in this XX/XY turtle, an important pathway in ovarian development, providing another target for future functional testing.

Conclusions

Evolutionary transitions in sex-determining mechanism require at least a partial reorganization of the regulatory network involving, among other possibilities, the molecular evolution of protein coding regions of its gene members. Our work revealed a significant acceleration of substitution rates during putative GSD-to-TSD reversals than during TSD-to-GSD transitions or SDM stasis. Evolutionary state reconstruction has a major influence on our understanding of SDM evolutionary history, as no differences on the molecular evolution of this regulatory network were detected when GSD-to-TSD reversals are not considered by one of the evolutionary hypotheses evaluated. These discrepancies highlight the importance of reliable methods for ancestral state reconstruction, and the need for additional genomic and SDM data across the tree of life to enable robust analyses. Our data predicted modifications in the functional domains of proteins in the sex determination network which may play a role in SDM evolution in turtles, providing targets for future research. Among the genes that experienced significant molecular evolution, hormonal signaling genes (Srd5a1 in particular) and HSF2 emerged as being dramatically altered during SDM shifts in turtles. No single substitution or structural change was present in all of the studied GSD turtles, and some transitional branches appear to evolve much faster than others, a testament to the independent nature of turtle SDM evolution and the many trajectories taken by nature to produce males and females.

Acknowledgements We greatly appreciate the assistance in specimen collection for G. insculpta (Jeff Tamplin), E. macquarii (Arthur Georges), P. expansa (Omar Hernandez), and C. insculpta (Gabriel Rivera). This work was funded in part by NSF Grants DEB 1310793 to NV and RL, and MCB 1244355 and IOS 1555999 to NV.

References


Montiel EE, Badenhorst D, Tamplin J, Burke RL, Valenzuela N (2017) Discovery of the youngest sex chromosomes reveals first case of...
convergent co-option of ancestral autosomes in turtles. Chromosoma 126:105