Sequential Turnovers of Sex Chromosomes in African Clawed Frogs (*Xenopus*) Suggest Some Genomic Regions Are Good at Sex Determination

Benjamin L. S. Furman and Ben J. Evans[1]
Department of Biology, McMaster University, Hamilton, Ontario L8S 4K1, Canada

ABSTRACT Sexual differentiation is fundamentally important for reproduction, yet the genetic triggers of this developmental process can vary, even between closely related species. Recent studies have uncovered, for example, variation in the genetic triggers for sexual differentiation within and between species of African clawed frogs (genus *Xenopus*). Here, we extend these discoveries by demonstrating that yet another sex determination system exists in *Xenopus*, specifically in the species *Xenopus borealis*. This system evolved recently in an ancestor of *X. borealis* that had the same sex determination system as *X. laevis*, a system which itself is newly evolved. Strikingly, the genomic region carrying the sex determination factor in *X. borealis* is homologous to that of therian mammals, including humans. Our results offer insights into how the genetic underpinnings of conserved phenotypes evolve, and suggest an important role for cooption of genetic building blocks with conserved developmental roles.

KEYWORDS
sex chromosomes
Xenopus
rapid evolution
sex chromosome turnover
SOX3
Genetics of sex

For nearly all vertebrates, two sexes are needed to secure the benefits of genetic recombination associated with sexual reproduction (Barton and Charlesworth 1998). It is, therefore, not surprising that the genetic control of sexual differentiation is tightly regulated, and has remained unchanged for millions of years in several lineages (Graves and Peichel 2010; O'Meally *et al.* 2012; Veyrunes *et al.* 2008; Matsubara *et al.* 2006). However, genetic control of sexual differentiation has diversified in some groups. For example, nonhomologous sex chromosomes have been detected in several closely related species or populations of stickleback (Ross *et al.* 2009), medaka (Myosho *et al.* 2015), and cichlid (Roberts *et al.* 2009) fish, and rampant turnover of the sex chromosomes occurred over a broader phylogenetic scope in fish (Devlin and Nagahama 2002; Mank *et al.* 2006), gecko lizards (Gamble *et al.* 2015), and amphibians (Evans *et al.* 2012).

Among these turnover events, common elements have been independently coopted for sex determination in several instances. For example, one syntenic block of genes independently became sex-linked in a lizard (*Gekko hokouensis*) and birds (Kawai *et al.* 2009), and another separately became sex-linked in a frog (*Rana rugosa*) and therian mammals (Wallis *et al.* 2007; Uno *et al.* 2008, 2013). In addition, individual genes with sex-related function have repeatedly evolved into the trigger for sexual differentiation. Examples include homologs of *doublesex and mab-3 related transcription factor 1* (*DMRT-1*), an important sex-related gene in vertebrates (Zarkower 2001), which are triggers for sex determination in medaka fish, *Oryzias latipes* (Kondo *et al.* 2003, 2004), the African clawed frog *Xenopus laevis* (Yoshimoto *et al.* 2008), probably the Chinese half-smooth tongue sole (Chen *et al.* 2014), and all birds (Smith *et al.* 2009; but see Zhao *et al.* 2010). Similarly, homologs of *SOX3*, which is another important sex-related gene (Weiss *et al.* 2003), independently became triggers for sexual differentiation in the fish *O. dancena* (Takehana *et al.* 2014) and in the ancestor of therian mammals (Koopman *et al.* 1991). Turnover of sex chromosomes and the genes involved with sex determination provide opportunities to study how tightly regulated systems evolve, and in particular the extent to which this involves convergence, reversion to an ancestral state, or origin of genetic novelty.

In addition to being model organisms for biology (Cannatella and de Sá 1993; Hellsten *et al.* 2010; Harland and Grainger 2011), African clawed frogs (genus *Xenopus*) offer a promising system with which to study sex chromosomes. At least two species, *X. laevis* (Daudin 1802) and *X.* (*Silurana*) *tropicalis* (Gray 1864), have a nonhomologous trigger for sex determination (Yoshimoto *et al.* 2008; Olmstead *et al.* 2010; Roco *et al.* 2015). These two species are members of different subgenera

doi: 10.1534/g3.116.033423
Manuscript received July 11, 2016; accepted for publication September 4, 2016; published Early Online September 7, 2016.

Supplemental material is available online at www.g3journal.org/lookup/suppl/doi:10.1534/g3.116.033423/-/DC1.
[1]Corresponding author: Department of Biology, Life Sciences Building Room 328, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada. E-mail: evansb@mcmaster.ca

that are distinguished from each other by the number of chromosomes (x) carried by the gametes of their respective diploid ancestors, *i.e.*, x = 10 for subgenus *Silurana* and x = 9 for subgenus *Xenopus* (Evans *et al.* 2015). All extant species in subgenus *Xenopus* are polyploid, but with disomic chromosomal inheritance, and tetraploids in this subgenus have $4x$ = 36 chromosomes. In *X. laevis*, a gene called *DM-W* is the master sex regulator of sex determination (Yoshimoto *et al.* 2008); this gene appeared in an ancestor of *X. laevis* after divergence from the ancestor of *X. tropicalis*, and is present in many close relatives of *X. laevis* (Bewick *et al.* 2011). In subgenus *Silurana*, *X. tropicalis* has a complex trigger for sex determination that resides on Y, W, and Z chromosomes (Roco *et al.* 2015). This system in *X. tropicalis* produces distorted sex ratios in some crosses (Roco *et al.* 2015). Thus, African clawed frogs use at least two systems for sex determination, and at least one of them evolved during the diversification of this group.

Within subgenus *Xenopus*, species in a clade including *X. borealis* (Parker 1936), *X. muelleri* (Peters 1844), and *X. fischbergi* (Evans *et al.* 2015) appear to lack *DM-W* (Bewick *et al.* 2011), hinting at additional diversity of sex chromosomes in this group. The phylogenetic placement of this clade within *Xenopus* remains uncertain, making unclear the evolutionary histories of potentially diverse triggers for sex determination.

To further explore sex-related innovations in these frogs, we (i) used whole transcriptome information from several species to further resolve phylogenetic relationships within subgenus *Xenopus*. We (ii) tested whether *DM-W* is sex-linked in the most distantly related species from *X. laevis* that is known to carry *DM-W*, *i.e.*, *X. clivii* (Peracca 1898). Then, we (iii) used reduced representation genome sequencing and Sanger sequencing to identify the sex-linked region in *X. borealis*, and (iv) established homology between the genes on the sex chromosomes of *X. borealis* and several other distantly related species. Our results identify a new sex determination system in *X. borealis* that evolved after the *DM-W*-based system was already in place in an ancestor. Interestingly, the genomic regions involved in sex determination of *X. borealis* and therian mammals (including humans) are homologous. Rapid evolution of *Xenopus* sex chromosomes highlights a central role for cooption of genes with conserved developmental roles in the evolution of important genetic pathways.

MATERIALS AND METHODS

Exploring the origin of DM-W

Nuclear data: In order to infer evolutionary relationships among representative *Xenopus* species that do and do not carry *DM-W*, we performed phylogenetic analyses on nuclear sequence data obtained from two sources. For the tetraploid species *X. laevis* and the diploid outgroup species *X. tropicalis*, we used Unigene databases (downloaded November 2015). These datasets had 31,306 and 36,839 unique sequences for *X. laevis* and *X. tropicalis*, respectively. For the tetraploid species *X. borealis*, *X. clivii*, *X. allofraseri*, and *X. largeni*, we extracted RNA from liver tissue using the RNAEasy extraction kit (Qiagen Inc.). These four transcriptomes were multiplexed on two thirds of one lane of an Illumina HiSeq 2000 machine, with 100 bp paired end sequencing and using libraries that were prepared with the Illumina TruSeq RNA Sample Preparation Kit v2. This produced 18–20 million paired reads for each sample (data are deposited in the NCBI short read archive with accession numbers: *X. borealis* PRJNA318484, *X. clivii* PRJNA318394, *X. allofraseri* PRJNA318474, and *X. largeni* PRJNA318404).

Low quality reads and bases were removed using Trimmomatic version 0.30 (Bolger *et al.* 2014). We discarded the first and last 3 bp and then required the average Phred-scaled quality scores of retained sequences to be at least 15 in a sliding window of 4 bp. After imposing these requirements, we discarded all reads that were shorter than 36 bp. Across the samples, 88–95% of paired reads passed these filters. We then assembled the transcriptomes for each species with Trinity (version 2013_08_14), using default values for all settings including, for example, a kmer size of 25 and a minimum contig length of 200 (Grabherr *et al.* 2011; Haas *et al.* 2013). The resulting assemblies had 72,000–97,000 unique transcripts (*X. borealis* = 81,696, *X. clivii* = 72,019, *X. allofraseri* = 96,832, and *X. largeni* = 82,695) and N50 values (the minimum length, in bp, for the longest 50% of reads) ranging from 885–1176 bp (*X. borealis* = 1078, *X. clivii* = 885, *X. allofraseri* = 1176, and *X. largeni* = 1000). Additional information on Illumina sequencing is presented in Supplemental Material, Table S1.

We used a reciprocal BLAST (Altschul *et al.* 1997) approach between each tetraploid transcriptome (or Unigene database in the case of *X. laevis*) and the *X. tropicalis* Unigene database to collect sets of homologous sequences for phylogenetic analysis (Figure S1). These sets of sequences included orthologous gene sequences (sequences in different species whose divergence was triggered by speciation), homeologous gene sequences (sequences in the same or different species whose divergence was triggered by genome duplication), and included splice variants, segmental duplicates, and assembly errors generated by Trinity (Grabherr *et al.* 2011). We performed a quality control step, retaining only those alignments whose ungapped length was above an arbitrary cutoff of 299 bps, and that contained sequences from at least three ingroup species with at least one species having at least two sequences. The need for the requirement that at least one species have two (possibly homeologous) sequences is discussed next.

Because our ingroup species are tetraploid, it was crucial for our phylogenetic analyses to distinguish orthologous from homeologous gene sequences. Since speciation occurred more recently than whole genome duplication in subgenus *Xenopus*, orthologous genes are expected to be more closely related to one another than they are to homeologous genes. In a gene tree with only one sequence from each species, it was therefore a concern that the relationships among the sequences could be orthologous or homeologous. Therefore, we developed a phylogeny-based bioinformatic filter that identified alignments whose estimated phylogeny allowed us to distinguish orthologous from homeologous gene sequences (Figure S1). Importantly, we did not make any assumptions about how the orthologous sequences were related to one another. This filter involved three rounds of tree building, with each followed by assessment of sequence relationships using a script and functions from the R packages Ape, Phytools, and Phangorn (R Core Team 2015; Revell 2012; Schliep 2011; Paradis *et al.* 2004; this script is available at Dryad repository; see *Data availability*). The resulting alignments each included at least one species with two homeologous sequences, which diverged prior to speciation of extant tetraploids in subgenus *Xenopus*. Additionally, each alignment had at least three representative orthologous sequences. Similar BLAST and phylogenetic-based filtering approaches have been used in other studies to distinguish orthologous from homeologous gene sequences (Dehal and Boore 2005; Inoue *et al.* 2015). See File S1, section S1.1 for full details.

Phylogenetic analyses of nuclear DNA: After filtering these alignments, we performed several phylogenetic analyses on these data including: (i) individual gene tree analyses for each alignment (Beast; Drummond and Rambaut 2007), (ii) concatenated Bayesian analyses (Beast), (iii) concatenated maximum likelihood analyses (RAxML; Stamatakis 2014), (iv) a gene tree to species tree analysis using MPest

(Liu *et al.* 2010), and (v) a multi-species coalescent analysis using *Beast (Heled and Drummond 2010). For Analysis (i), a model of evolution was selected for each gene alignment using the Akaike Information Criterion MrModelTest2 (Nylander 2004). We set the root height to be 65 MY, with a SD of 4.62 MY (Bewick *et al.* 2012) and assumed a strict clock, and ran two chains, for at least 75 million generations. 197 files failed to converge with substitution model selected by MrModelTest, so we instead used the HKY +Γ model. For all analyses, we assessed convergence of the posterior distribution using loganalyser (part of the Beast package), and removed a 25% burn-in from each chain. For Analysis (i), we summarized relationships across the combined postburn-in posterior distribution of all individual gene analyses using an approach described in File S1, section S2.2. We analyzed two datasets for Analyses (ii) and for (iii). The first dataset was a concatenation of all gene alignments. The second dataset had all sites with gaps or missing data removed from the concatenated alignment. For Analysis (ii), for both datasets, we set a GTR + I + Γ substitution model (as selected by MrModelTest using AIC) and a strict clock with an exponential distribution for the rate with a mean rate of 1.0 and a SD of 0.33 (default settings in Beauti). The root height was set to 65 my (±4.62) as detailed above (Bewick *et al.* 2012). For each dataset, we ran four independent chains, for 50 million generations, and tested for convergence by inspecting the plots of parameter estimates and calculating ESS values using Tracer. Based on this inspection, we removed a 25% burn-in from each chain and constructed a consensus tree using TreeAnnotator. For Analysis (iii), we used the GTR + Γ model and performed 500 bootstrap replicates to assess support.

For Analysis (iv), we used the individually constructed Beast consensus chronograms that were generated from Analysis (i). We selected a random sample of 250 trees from the postburn-in posterior sample of tree topologies from each gene tree analysis to act as the "bootstrap" replicates, which MPest uses to assess support (Seo 2008). These trees were uploaded to the Straw server (Shaw *et al.* 2013) to run the MPest analysis.

To perform Analysis (v), we used only those gene alignments that had orthologous sequence data for all species (*i.e.*, five aligned orthologs within one homeologous lineage), and retained only the longest sequence in the other homeologous lineage (or a randomly selected sequence if there were multiple equally long sequences). Because the homeologous sequences are equivalently diverged from a set of orthologs, it did not matter from which species this latter homeologous sequence was derived. The result was a dataset that had gene sequence for all taxa, and minimizing missing data to only incomplete sequencing of a gene and insertion deletion mutations. We ran *Beast with a strict clock that was linked across all partitions. The GTR + Γ model of evolution was used and was linked across partitions. The tree topology, however, was free to vary among genes (*i.e.*, it was unlinked). We ran two independent chains for 500 million generations each. Convergence was assessed using effective sample size values calculated with Tracer. Based on this, we removed a 25% burn-in from each chain. This analysis did not include calibration points because all attempts to set one failed to converge on the posterior distribution. Instead, in order to assign dates to the nodes, trees in the resulting posterior distribution were rescaled using an *R* script that used functions from the phytools library (Revell 2012). As above, the root node age was drawn from a normal distribution with a mean of 65 and a SD of 4.62 (Bewick *et al.* 2012), and the rest of the nodes were assigned based on branch length from the root.

Phylogenetic analysis of mitochondrial DNA: We downloaded the previously sequenced mitochondrial genomes for *X. tropicalis* (direct GenBank submission: NC_006839.1), *X. borealis* (GenBank accession no. X155859; Lloyd *et al.* 2012), and *X. laevis* (GenBank accession no. HM991335; Irisarri *et al.* 2011). We used the *X. borealis* mitochondrial genome as a Blast query to recover matches from the transcriptomes of *X. clivii*, *X. allofraseri*, and *X. largeni*, retaining hits with less than an e^{-10} match. Then, using these assembled mitochondrial DNA (mtDNA) sequences and the previously sequenced mtDNA genomes, a multispecies alignment was performed using Mafft (Katoh and Standley 2013) followed by manual adjustment. In order to remove sections that were poorly aligned or had ambiguous homology, GBlocks (Castresana 2000) was used with default parameters. We then performed a Beast analysis of these data, a root node age set to 65 MY and a SD of 4.62 (Bewick *et al.* 2012), a GTR + I + Γ substitution model (as determined by AIC with MrModelTest2), and ran 13 chains. For comparative purposes, we ran this analysis with a relaxed clock and with a strict clock, and the suitability of each clock model was assessed by comparing the harmonic means of the postburn-in likelihood values. We also performed a RAxML analysis with a GTR + Γ model and 1000 bootstrap replicates to assess support.

Assessing sex specificity of DM-W in X. clivii

The phylogenetic results (discussed below in DM-W originated before speciation of *X. laevis*, *X. clivii*, *X. borealis*, and other 4x = 36 tetraploids) suggests that *X. clivii* is the most distantly related species to *X. laevis* that carries *DM-W*. Therefore, we tested whether *DM-W* is found only in *X. clivii* females by attempting to amplify *DM-W* in several wild-caught individuals for which sex was inferred based on external morphology (Evans *et al.* 2011a). We designed primers from a sequenced clone of *DM-W* from this species (Bewick *et al.* 2011; Table S2) and attempted to amplify this gene in 12 females and 13 males.

The sex determining region of X. borealis

X. borealis and X. laevis families: We generated *X. borealis* and *X. laevis* families from adults obtained from *Xenopus* Express (Brooksville, FL). To promote mating, parents each received 50 U of Human chorionic gonadotropic (HCG) followed by 200 and 50 U for the female and male, respectively, 6 hr later. The *X. borealis* offspring were reared to sexual maturity, killed with an overdose of MS222, and dissected to determine sex based on presence of testis or ovary. For *X. laevis*, tadpoles were reared for 4 wk and then killed with MS222. Sex of the *X. laevis* tadpoles was determined based on amplification or lack of amplification of a portion of *DM-W*; amplification of *DMRT-1* was used as a positive control (Yoshimoto *et al.* 2008; Bewick *et al.* 2011). For both families, DNA was extracted using DNEasy kits (Qiagen, Inc.) from either fresh liver tissue (*X. borealis*) or tadpole tail tissue (*X. laevis*).

Genotype by sequencing (GBS) sequencing: To identify the sex determining region of *X. borealis*, we performed GBS (Elshire *et al.* 2011) on parents and offspring of the *X. borealis* cross. DNA was extracted for 23 male and 24 female siblings, and both parents using DNEasy extraction kits (Qiagen, Inc.). For the mother and father, we sequenced multiple technical replicates to increase coverage 10-fold for each parent compared to each offspring. Library preparation using the *Eco*T22I restriction enzyme and sequencing was performed at Cornell University Institute of Biotechnology Genome Diversity Facility. Sequencing (100 bp, single end) was performed using an Illumina Hi-Sequation 2500 machine; 96 samples, of which 67 were *X. borealis* samples for this study, were repeated on two Illumina lanes at 96-plex each; the resulting sequence files were merged prior to processing.

We then used Tassel v.3.0 (Glaubitz *et al.* 2014), employing the UNEAK pipeline (Lu *et al.* 2013), to perform SNP calling of GBS data

without the use of a reference genome sequence. TASSEL also does demultiplexing, quality checking, and barcode trimming of sequences. During the process, reads were truncated to a maximum of 64 bp, and high quality reads with < 64 bp were padded with "A" nucleotides to bring them to the 64 bp length. We set the minimum number of times a read must be present (-c option) to five, and set the error tolerance rate (*i.e.*, the number of mismatched base pairs between reads) to 0.03 when forming groups of homologous sequences. The minimum and maximum allele frequencies of SNPs were set to 0.05 and 0.5, respectively, and the minimum and maximum call rate (*i.e.*, the proportion of all individuals that must have a sequence to call a SNP for a stack of reads) was set to 0.0 and 1.0, respectively. We then trimmed the dataset to only sequence tags that had SNP calls for at least 90% of individuals.

One concern we encountered was "under calling" of heterozygous sites, wherein sites that are actually heterozygous were called as homozygous. For instance, if the parental genotype calls were A/T and A/A, and an offspring was T/T, then it is likely the offspring was actually T/A because the coverage of the parents was ~10 × higher. To cope with this, we used a Perl script (deposited in Dryad, see above) to compare offspring genotype calls to those of parent genotype calls for each locus in order to identify biologically implausible genotypes. If < 10% of offspring had a biologically implausible genotype call, then the implausible genotype calls were changed to missing genotypes. If more than 10% of the offspring had implausible calls, then the site was discarded. With this Perl script, we then identified completely sex biased inheritance of parental SNPs, and used this information to determine whether such sites had inheritance consistent with a female heterogametic (ZZ/ZW) or male heterogametic (XX/XY) sex determining system. We limited our search to loci that were completely sex biased (*i.e.*, only daughters or only sons were heterozygous).

Comparative analysis of the X. borealis sex determining region: We used BLAST with the consensus sequences (64 bp long) surrounding the sex-linked SNPs (hereafter "tags") from *X. borealis*, generated by TASSEL, as a query to find matches in the *X. laevis* genome assembly v.7.1 (Bowes *et al.* 2008). Matching *X. laevis* scaffolds were then aligned to the reconstructed *X. tropicalis* chromosomes in the v.9.0 genome, using the program NUCMER (part of the MUMMER package; Delcher *et al.* 2002). Settings for NUCMER included a minimum length of a maximal exact match of 50 (-l 50), gaps between cluster of matching sequence was set to 500 (-g 500), match separation was set at 0.08 (-d 0.08), and the minimum cluster length was set to 150 (-c 150).

As discussed below, this analysis indicated that a genomic region containing three sex-related genes—*sex determining region Y-box 3* (*SOX3*), *androgen receptor* (*AR*), and *fragile X mental retardation 1* (*FMR1*)—might be sex-linked in *X. borealis*. To test this, we amplified and sequenced portions of these three genes in our *X. borealis* family using Sanger sequencing. Primers for both homeologs of *SOX3* and *FMR1* were designed from the *X. laevis* v.7.1 genome or from unpublished *X. borealis* genome sequence data. For *AR*, we used the primers detailed in Evans *et al.* (1998), which target the hypervariable region of one homeolog of the *AR* gene. Primer sequences are reported in Table S2. Using BLAST, we identified chromosomes or scaffolds in the *X. laevis* genome v.9 that are orthologous to these homeologous sequences in *X. borealis*. We also sequenced these genes in wild-caught *X. borealis*, including individuals of both sexes and from multiple localities.

As an additional independent test of whether the sex-determining regions of *X. laevis* and *X. borealis* reside in nonhomologous genomic regions, we evaluated sex linkage of a *RAB6A* homeolog that is located near *DM-W* (Uno *et al.* 2013), in both the *X. laevis* and *X. borealis* families. We designed primers for both homeologs using *X. laevis* genome v.7.1 (Table S2) and amplified in parents and offspring of both crosses, followed by Sanger sequencing.

Data availability

Representative individuals from the sex-linked alignments and wild samples were deposited in Genbank (accession *SOX3*:KX765742–KX765751; *FMR1*:KX765752–KX765762; and *AR*:KX765731–KX765741), and transcriptome and GBS sequences in the NCBI short read archive (accessions PRJNA318484, PRJNA318394, PRJNA318474, PRJNA318404, and PRJNA319044). The phylogenetic trees, gene sequence alignments, BEAST XML files for final gene trees, important scripts used in this study, and full alignments of sex-linked genes are deposited in Dryad (doi: 10.5061/dryad.00db7).

RESULTS AND DISCUSSION

DM-W originated before speciation of X. laevis, X. clivii, X. borealis, and other 4x = 36 tetraploids

The gene *DM-W* triggers female sexual differentiation in the African clawed frog *X. laevis* and is located on the female-specific portion of the W sex chromosome (Yoshimoto *et al.* 2008). This gene is carried by several other *Xenopus* species, but has not been detected in *X. borealis* (Bewick *et al.* 2011). The most distantly related species from *X. laevis* known to carry *DM-W* is *X. clivii*; however, phylogenetic relationships among these three species remain unresolved. If *X. borealis* does indeed lack *DM-W*, two possibilities exist: either (i) *DM-W* arose after divergence of *X. borealis* from the most recent common ancestor (MRCA) of species that carry this gene, including *X. laevis* and *X. clivii*, or (ii) *DM-W* evolved prior to this in the MRCA of species that do and do not carry *DM-W*, and was subsequently lost in a more recent ancestor of *X. borealis*. Analyses of partial mtDNA sequences support the former hypothesis (Evans *et al.* 2004, 2011b, 2015) and analysis of two linked nuclear DNA (nDNA) genes supports the latter (Evans *et al.* 2005, 2015; Evans 2007). Therefore, we estimated phylogenetic relationships among tetraploid species that represent the major *Xenopus* clades, in which *DM-W* has and has not been detected, using new and publicly available sequence data from nuclear and mitochondrial DNA from *X. largeni*, *X. allofraseri*, *X. borealis*, *X. clivii*, and *X. laevis*, and the diploid outgroup species *X. tropicalis*.

From these data, we recovered 1585 sets of homologous nuclear gene sequences (File S1, section S1.1). Each set consisted of at least one species with two homeologous sequences (*i.e.*, generated from tetraploidization), at least 300 bp for all species, and a minimum of three ingroup taxa for at least one set of orthologs. When combined, these data included 2,696,030 bp. Data from a given ingroup species were missing from the gene alignments as rarely as 14% of the gene alignments (for *X. laevis*) to as much as 64% of the gene alignments (for *X. clivii*, File S1, section S1.1). These data formed the basis of Analyses (i–iv). Analyses with gapped sites removed [alternate Analysis (ii) and (iii)], included a total of 788,627 aligned bp. The *BEAST analysis [Analysis (v)] included 151 gene alignments (238,606 bp, 70,233 sites sequenced for all taxa, with some gaps due to insertion-deletion mutations or incomplete gene sequences).

All of the multigene analyses (ii–v) strongly supported, with posterior probabilities of 1.0 (or bootstrap support of 100%), two reciprocally monophyletic clades, with the first including *X. borealis* and *X. clivii* and the second including *X. laevis*, *X. largeni*, and *X. allofraseri* (Figure 1, Figure S2, and File S1, section S2.1). Similar to previous studies (Evans *et al.* 2004, 2005, 2007, 2011b, 2015), these analyses failed to resolve relationships among *X. laevis*, *X. largeni*, and *X. allofraseri* with

strong support (Figure 1, Figure S2, and File S1, section S2.1). Analyses of individual genes (i) identified substantial gene tree discordance among chronograms estimated from each gene (Table S3 and File S1, section S2.2). Despite this discordance, in the pooled postburn-in posterior distribution of these chronograms, a sister relationship between *X. borealis* and *X. clivii* was at least twice as common as any other relationship with either of these species (Table S3 and File S1, section S2.2).

Because previous phylogenetic inferences from mtDNA and nDNA differed with respect to the placement of *X. clivii*, we reexamined mtDNA relationships with additional data from the liver transcriptome sequences of *X. clivii*, *X. largeni*, and *X. allofraseri*, and complete mtDNA genome sequences from *X. tropicalis*, *X. laevis*, and *X. borealis*. After gaps and ambiguously aligned portions were removed, the alignment length was 8318 bp, which spans about 50% of the complete mtDNA genomes of *X. laevis*, *X. borealis*, and *X. tropicalis*. When analyzed with a relaxed molecular clock Bayesian analysis, or with a no clock maximum likelihood analysis, a phylogeny that was topologically consistent with the nDNA analyses was recovered. This topology included a clade containing *X. borealis* and *X. clivii*, although support for this clade was lower than the multigene analyses of nDNA described above (posterior probability was 0.75 and bootstrap support was 66%; Figure S3). Analysis with manual removal of ambiguously aligned sequences instead of GBLOCKS (16,260 bp aligned) recovered the same topology for both analyses and with similar levels of support (results not shown).

Analysis with a strict molecular clock supported an alternative mtDNA topology, with *DM-W* containing species forming a monophyletic group, as was found by previous studies (Evans *et al.* 2004, 2015). However, Bayes factors calculated following Nylander *et al.* (2004), indicate that a relaxed clock model is strongly preferred over the strict clock (BF = 9.3; Kass and Raftery 1995). An important difference between this and previous mtDNA analyses is that this study is based on a ~sixfold larger dataset. Similar to the nDNA analyses, mtDNA analyses failed to confidently resolve the relationships of *X. laevis*, *X. allofraseri*, and *X. largeni* (Figure S3).

Although the support for a sister relationship of *X. borealis* and *X. clivii* is lower in the mtDNA analysis than in the nDNA analyses, this relationship has more support than any alternative. Thus, using the most favored models of evolution we considered, the most strongly supported phylogenetic relationships among nDNA and among mitochondrial DNA are both consistent with an origin of *DM-W* prior to the diversification of the most recent common ancestor of all of our ingroup taxa (*X. laevis*, *X. largeni*, *X. allofraseri*, *X. clivii*, and *X. borealis*). Results from mtDNA and nDNA, thus, both suggest that *DM-W* originated before the diversification of extant ($4x = 36$) tetraploids in subgenus *Xenopus*.

DM-W is sex-linked in X. clivii

Our phylogenetic results indicate that *X. clivii*, a species that carries *DM-W*, is closely related to several species in which *DM-W* has not been detected, including *X. borealis* (Bewick *et al.* 2011). *DM-W* was previously amplified in one female *X. clivii* individual, but it is not clear whether this gene is also sex-linked in this species. Put another way, although *DM-W* arose before *X. laevis* and *X. clivii* diverged from one another, it is possible that *DM-W* acquired its role as a trigger for sexual differentiation (and thus its female-specific mode of inheritance) in an ancestor of *X. laevis* after divergence from an ancestor of *X. clivii*. Therefore, we tested whether *DM-W* is found only in *X. clivii* females, including in our assay males and females from the populations on each side of the Ethiopian Rift Valley (Evans *et al.* 2011a). We were able to amplify *DM-W* in a subset of females (8 of 12 females) from both sides

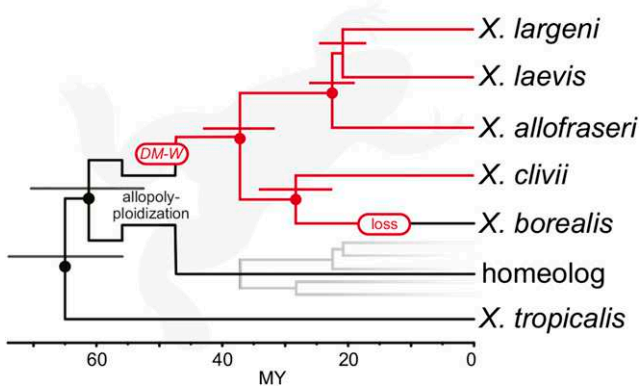

Figure 1 Phylogenetic relationships inferred from representative species in subgenus *Xenopus* suggests *DM-W* was gained before diversification of ($4x = 36$) tetraploids and then lost in an ancestor of *X. borealis*. This phylogeny was recovered from *BEAST analysis of transcriptome data and is topologically consistent with those recovered from other analyses of nuclear DNA and of mitochondrial DNA. Dots over nodes indicate 1.0 posterior probability; bars above nodes indicate the 95% credible intervals for divergence time in millions of years (MY). All species depicted are tetraploids except the outgroup species, *X. tropicalis*, which is diploid. For this analysis, one homeolog from any one of the tetraploid species was included for each gene, and is indicated by the gray subtree (File S1, section S1.1). The timing of the origin of *DM-W* with respect to the allopolyploidization event (whether before or after) is unclear. *Xenopus* silhouette from Phylopic by Sarah Werning, CC04 license.

of the Rift Valley, but no males (0 out of 12 males; a 13th male also failed to amplify in a positive control; Figure S5). The failure of *DM-W* to amplify in four female samples, which were also from both sides of the Ethiopian Rift Valley, could be due to divergence at our primer sites or misidentification of the sex of these individuals when sampled in the field (specimens of these individuals were not available for examination). It is also possible that additional sex determining systems may also be present in *X. clivii*, as is the case in *X. tropicalis* (Roco *et al.* 2015). Either way, female-specific amplification is consistent with the hypothesis that *DM-W* is found only in female *X. clivii*, that this gene triggers female sexual differentiation in at least some *X. clivii* individuals, and (more broadly) that *DM-W* was the ancestral trigger for female differentiation in subgenus *Xenopus*.

The sex determining region of X. borealis is different from that of X. laevis and that of X. tropicalis

Our inability to detect *DM-W* in *X. borealis* could be because this gene is not present, or because divergence at primer sites prevented amplification with the polymerase chain reaction. To find the sex-linked region of *X. borealis*, we examined patterns of inheritance of SNPs identified in our GBS data from the *X. borealis* family. Of the ~89,000 SNPs identified by Tassel (Table S1), ~21,000 were successfully genotyped in at least 90% of the offspring, and 15,632 of these passed our filter because they had "undercalled" genotypes in < 10% of the offspring (*Materials and Methods*). Of these, variation in 25 SNPs had a completely sex-linked pattern of inheritance (in offspring one sex is completely homozygous and the other completely heterozygous). By inspecting the genotypes of the parents, we could then distinguish ZZ/ZW from XX/XY systems (Figure S4). All 25 tags were consistent with female heterogamy. In 24 of them, the mother and daughters were heterozygous and the father and sons were homozygous; a pattern best explained by a SNP on the W chromosome. In one of the 25 tags,

the mother and sons were heterozygous and the father and daughters were homozygous; a pattern consistent with a SNP on the Z chromosome of the mother that was not present in either Z chromosome of the father. Overall, these results support genetic sex determination and female heterogamy in *X. borealis*, at least in the strain we examined, which is also the case in *X. laevis* (Mikamo and Witschi 1966) and possibly all other *DM-W*-containing *Xenopus* species.

To evaluate homology of the sex determining regions of *X. borealis*, *X. laevis*, and *X. tropicalis*, we aligned the *X. borealis* tags to the *X. laevis* genome assembly. This resulted in tags matching either (i) one region in *X. laevis*, (ii) two regions, (iii) multiple regions, or (iv) no regions. Scenario (ii) is likely the result of the short tags matching both homeologs in the *X. laevis* genome with similar strength. Scenarios (iii) and (iv) are not surprising given the short length of the tags and the divergence between *X. laevis* and *X. borealis* (Figure 1), and we discarded these tags. Ten of the 25 tags had only one or two *X. laevis* scaffold matches below our BLAST threshold ($<e^{-5}$). Six of these 10 scaffolds (either the single match or a randomly retained scaffold if there were two matches) aligned to *X. tropicalis* chromosome XTR8, two scaffolds had a split alignment with portions of each matching two different *X. tropicalis* chromosomes (XTR1 and XTR5 or XTR3 and XTR6, respectively), one matched *X. tropicalis* chromosome XTR4, and one matched *X. tropicalis* chromosome XTR7.

Most of the tags mapped to the XTR8 chromosome, suggesting that the sex chromosomes in *X. borealis* might be homologous to this *X. tropicalis* chromosome. To test this, we designed homeolog-specific primers based on *X. laevis* sequences, to amplify and sequence three genes (*SOX3*, *AR*, and *FMR1*) in our *X. borealis* family that are known to reside on chromosome XTR8 in *X. tropicalis* (Uno *et al.* 2013). This effort identified sex-linked polymorphisms in *X. borealis* in one homeolog of each gene, and each was consistent with a female heterogametic (ZZ/ZW) sex chromosome system. For *SOX3*, *AR*, and *FMR1*, we successfully amplified and genotyped 93, 41, and 54 offspring, respectively, including 47, 24, and 30 daughters, respectively. For all three of these genes, we identified at least one heterozygous site in the mother of the cross that allowed us to confirm sex linkage and female heterogamy (Figure S4; alignments of all sequences are deposited in Dryad and representative sequences are deposited in GenBank; see *Data availability*). For the *AR* amplification, the father appeared to have a null allele, but importantly, this did not compromise our ability to assess sex linkage and female heterogamy, which was based on patterns of inheritance of a heterozygous SNP from the mother (Figure S4), resulting in completely sex associated genotypes in the offspring. The top BLAST hit of the sex-linked *X. borealis SOX3* and *FMR1* homeologs to the *X. laevis* genome indicated that these sequences were orthologous to *X. laevis* chromosome XLA8L (and thus homeologous to XLA8S); *AR* was orthologous to an unplaced scaffold (scaffold 37), but fluorescent *in situ* hybridization studies place this gene on XLA8L (Uno *et al.* 2013).

In wild-caught *X. borealis*, we successfully sequenced amplifications from three females and three males for *SOX3*, and amplifications from the same individuals plus a fourth male for *AR* and *FMR1*. Two of three females tested had the same heterozygous genotypes in *SOX3* and *FMR1* as the females in our lab family; for *AR*, neither of these samples had the same sex-linked polymorphism as the lab family. The wild-caught females also had other polymorphic sites, some of which were shared with male wild samples. These results indicate either that these genes reside in the pseudoautosomal region in *X. borealis*, that there is variation in the sex determining system within *X. borealis*, or some combination of these possibilities. It is also possible that the sex of some of the wild-caught individuals was misidentified based on external morphology; unfortunately, specimens of these individuals were not available for examination. Examination of other wild-caught individuals whose sex is determined surgically is an important next step for further characterizing the sex-specific region of the sex chromosomes of *X. borealis*.

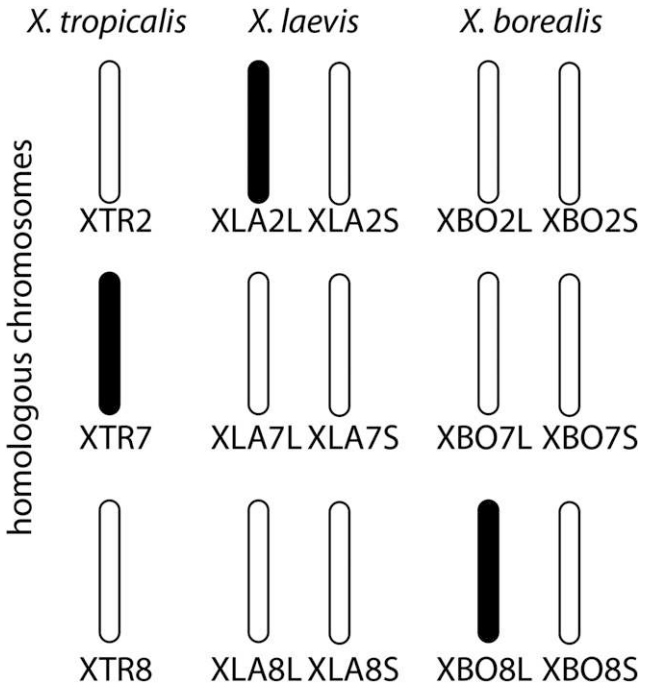

Figure 2 Sex chromosomes, indicated in black, in three species of African clawed frog are not homologous. For the tetraploid species, *X. laevis* and *X. borealis*, both homeologous (L and S) chromosomes are shown. Chromosome nomenclature for *X. tropicalis* and *X. laevis* follows (Matsuda *et al.* 2015).

Analysis of polymorphisms in Sanger sequences of homeologs of the *RAB6A* gene indicated that one homeolog is linked to *DM-W* in *X. laevis*, as indicated by sex-linked inheritance (File S1, section S2.3), in agreement with a finding from fluorescence *in situ* hybridization (Uno *et al.* 2013). This analysis also revealed that the ortholog of *RAB6A* that is sex-linked in *X. laevis* is not sex-linked in *X. borealis* (File S1, section S2.3). Overall, these results demonstrate that the genomic region containing the trigger for sex determination differs between the *X. borealis* strain we examined and *X. laevis*.

Some genomic regions are good at sex determination

Our results indicate that the sex chromosomes of *X. borealis* are homologous to *X. tropicalis* chromosome XTR8, orthologous to *X. laevis* chromosome XLA8L, and homeologous to *X. laevis* chromosome XLA8S (Figure 2). In the diploid species *X. tropicalis*, the gene that triggers sex determination is unknown, but resides on the distal end of the petite arm of chromosome XTR7 (Olmstead *et al.* 2010; Wells *et al.* 2011; Roco *et al.* 2015). XTR7 is homologous to *X. laevis* autosomes XLA7L and XLA7S (Matsuda *et al.* 2015; Uno *et al.* 2013). The sex chromosome of *X. laevis* is XLA2L; this chromosome and its homeologous chromosome XLA2S are homologous to XTR2 of *X. tropicalis* (Matsuda *et al.* 2015; Uno *et al.* 2013). Thus, at least three sets of nonhomologous sex chromosomes are present within the African clawed frogs (Figure 2). The sex chromosomes of *X. laevis* and *X. borealis* occur in orthologous subgenomes (*i.e.*, portions of their respective allotetraploid genomes that are derived from the same diploid ancestor). There is still the possibility that *DM-W* has been

translocated to the newer sex chromosomes in *X. borealis*, but all efforts to detected it have failed (Bewick *et al.* 2011 and here).

Another frog species, *R. rugosa*, has sex chromosomes that are at least partially homologous to those in *X. borealis* (and this may be true for two other *Rana* species; Miura 2008). In both species, *SOX3* and *AR* are located on the sex chromosomes (Fujii *et al.* 2014; Uno *et al.* 2015). Interestingly, this inference extends even farther: orthologs of *AR*, *SOX3*, and *FMR1* are also present on the X chromosome of therian mammals, including humans (Uno *et al.* 2013), and *SOX3* is a new trigger for sex determination in a fish (*O. dancena*; Takehana *et al.* 2014). Similarly, the Z chromosome of lacertid lizards is partially homologous to the X chromosomes of therian mammals (Rovatsos *et al.* 2016). The phylogenetic placement of these lineages with respect to other species that have different sex determining systems [specifically *X. laevis* and monotremes (Veyrunes *et al.* 2008)] strongly suggests several independent origins of sex linkage of these homologous regions, or minimally of regions containing *SOX3*. Similarly, another region containing *DMRT-1*, an ortholog of which is related by partial gene duplication (paralogous) to *DM-W*, independently became sex-linked in birds and a gecko lizard (Kawai *et al.* 2009). Taken together, these observations are consistent with the proposal that certain genomic regions contain blocks of genes that are particularly suited to perform the task of triggering sex determination (Graves and Peichel 2010; Brelsford *et al.* 2013).

Conclusions

Sex chromosomes carry the genetic trigger that initiates sexual differentiation, a crucial developmental phenomenon that is generally required for reproduction (Matzuk and Lamb 2008). Sex chromosome turnover could occur by translocation among chromosomes of a conserved genetic trigger, or via a novel mutation creating a new trigger on an autosome. That sex chromosomes in African clawed frogs and several other lineages have frequently turned over contrasts sharply with other lineages with ancient sex chromosomes, such as therian mammals. Indeed, transitions in sex chromosomes appear to be more frequent when sex chromosomes are cytologically homomorphic and/or nondifferentiated (Bachtrog *et al.* 2014), which is the case in *Xenopus*, including *X. borealis* (Tymowska 1991), but not therian mammals. However, the evolutionary dynamics of these systems are highlighted by loss of the Y chromosome in various therians (Just *et al.* 1995; Sutou *et al.* 2001) and duplication of *SRY*, an ancient trigger for sex determination in this group (Geraldes *et al.* 2010).

A lack of recombination in the genomic region carrying the trigger for sex determination causes sex chromosomes to diverge from one another (Rice 1987). If the region of suppressed recombination expands, as it did in therian mammals, genomic elaborations such as loss and dosage compensation of sex-linked genes may arise and act as "evolutionary traps" that impede evolutionary change or, more specifically, future sex chromosome turnover (Bull 1983; Pokorna and Kratochvíl 2009; Gamble *et al.* 2015). In theory, before such evolutionary traps evolve, genes with sexually antagonistic function could catalyze sex chromosome turnover by increasing the fixation probability of new sex-determining genes that arise on a linked autosomal region (van Doorn and Kirkpatrick 2007). Related to this, dosage compensation has not been detected in species with female heterogamy (Mank 2009; Vicoso and Bachtrog 2009) or in anurans (frogs) in general, (*e.g.*, Schmid *et al.* 1986), and is unlikely to exist in *Xenopus* species whose female heterogametic sex chromosomes are homomorphic at the cytological (Tymowska 1991) and molecular level (Bewick *et al.* 2013). An absence of dosage compensation may prevent sex chromosome divergence (Adolfsson and Ellegren 2013) leaving a permissive environment for sex chromosome turnover in the presence of maintained homomorphic sex chromosomes, thereby avoiding these evolutionary traps. However, this is not always the case, as some snakes have differentiated sex chromosomes despite a lack of global dosage compensation (Vicoso *et al.* 2013; Rovatsos *et al.* 2015). More information about the nature of the master trigger for sex determination in *X. borealis*, on sex-linked genes, and on sex-biased expression of genes elsewhere in the genome may cast additional light on the drivers of sex chromosome turnover in these frogs. The drivers could include the role of alternative mechanisms that could resolve sexual conflict, such as gene duplication (Gallach *et al.* 2011; Wyman *et al.* 2012), which is a potentially important factor in these tetraploid species.

The sex determination system we detected in *X. borealis* is set apart from most other rapidly evolving systems, in that it is derived from an ancestral trigger that itself was newly evolved (*i.e.*, *DM-W*), as opposed to groups with diverse mechanisms that are each potentially once-evolved (autapomorphic). Our results support the hypothesis that *X. borealis* and *X. clivii* are sister taxa, and that *DM-W* is restricted to female *X. clivii*. This suggests that female sexual differentiation was triggered by *DM-W* in the ancestor of all extant species of subgenus *Xenopus*. This also suggests that the new system we report in *X. borealis* is derived with respect to the *DM-W*-based system. Thus, the sex chromosomes of *Xenopus* are an example of multiple important biological novelties arising in rapid succession.

Perhaps most interesting, however, are the aspects of sex determination that convergently evolve in distantly related organisms in the context of frequent sex chromosome turnover. These aspects include the participation of key sex-related genes (*e.g.*, *DMRT-1* in *X. laevis* and in *O. latipes*) and the role of homologous genomic regions (*e.g.*, carrying *SOX3*, in *X. borealis* and in *O. dancena*). This study contributes to a growing body of evidence that, in lineages with rapidly changing sex chromosomes, the turnover is catalyzed by cooption of genetic building blocks that are already involved in the development and maintenance of sexual differentiation.

ACKNOWLEDGMENTS

We thank Brian Golding for access to computing resources and Adam Bewick for assistance with rearing tadpoles. Funding for this study was provided by the Natural Science and Engineering Research Council of Canada (RGPIN/283102-2012) and the Museum of Comparative Zoology, Harvard University.

LITERATURE CITED

Adolfsson, S., and H. Ellegren, 2013 Lack of dosage compensation accompanies the arrested stage of sex chromosome evolution in ostriches. Mol. Biol. Evol. 30: 808–810.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped blast and psi-blast: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389–3402.

Bachtrog, D., J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto *et al.*, 2014 Sex determination: why so many ways of doing it? PLoS Biol. 12: e1001899.

Barton, N. H., and B. Charlesworth, 1998 Why sex and recombination? Science 281: 1986–1990.

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

Bewick, A. J., F. J. J. Chain, J. Heled, and B. J. Evans, 2012 The pipid root. Sys. Biol. 61: 913–926.

Bewick, A. J., F. J. Chain, L. B. Zimmerman, A. Sesay, M. J. Gilchrist *et al.*, 2013 A large pseudoautosomal region on the sex chromosomes of the frog silurana tropicalis. Genome Biol. Evol. 5: 1087–1098.

Bolger, A. M., M. Lohse, and B. Usadel, 2014 Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics 30: 2114–2120.
Bowes, J. B., K. A. Snyder, E. Segerdell, R. Gibb, C. Jarabek *et al.*, 2008 Xenbase: a xenopus biology and genomics resource. Nucleic Acids Res. 36: D761–D767.
Brelsford, A., M. Stöck, C. Betto-Colliard, S. Dubey, C. Dufresnes *et al.*, 2013 Homologous sex chromosomes in three deeply divergent Anuran species. Evolution 67: 2434–2440.
Bull, J. J., 1983 Evolution of Sex Determining Mechanisms, The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA
Cannatella, D. C., and R. O. de Sá, 1993 *Xenopus laevis* as a model organism. Sys. Biol. 42: 476–507.
Castresana, J., 2000 Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17: 540–552.
Chen, S., G. Zhang, C. Shao, Q. Huang, G. Liu *et al.*, 2014 Whole-genome sequence of a flatfish provides insights into zw sex chromosome evolution and adaptation to a benthic lifestyle. Nat. Genet. 46: 253–260.
Daudin, F. M., 1802 *Histoire Naturelle des Rainettes, des Grenouilles et des Crapauds. Ouvrage Orné de 38 Planches Représentant 54 Espèces Peintes D'après Nature*. Levrault, Paris.
Dehal, P., and J. L. Boore, 2005 Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol. 3: e314.
Delcher, A. L., A. Phillippy, J. Carlton, and S. L. Salzberg, 2002 Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 30: 2478–2483.
Devlin, R. H., and Y. Nagahama, 2002 Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191–364.
Drummond, A. J., and A. Rambaut, 2007 Beast: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7: 214.
Elshire, R. J., J. C. Glaubitz, Q. Sun, J. A. Poland, K. Kawamoto *et al.*, 2011 A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.
Evans, B. J., 2007 Ancestry influences the fate of duplicated genes millions of years after polyploidization of clawed frogs (Xenopus). Genetics 176: 1119–1130.
Evans, B., J. Morales, M. Picker, D. Melnick, and D. Kelley, 1998 Absence of extensive introgression between *Xenopus gilli* and *Xenopus laevis laevis* (Anura: Pipidae) in southwestern Cape Province, South Africa. Copeia 1998: 504–509.
Evans, B. J., D. B. Kelley, R. C. Tinsley, D. J. Melnick, and D. C. Cannatella, 2004 A mitochondrial DNA phylogeny of African clawed frogs: phylogeography and implications for polyploid evolution. Mol. Phylogenet. Evol. 33: 197–213.
Evans, B. J., D. B. Kelley, D. J. Melnick, and D. C. Cannatella, 2005 Evolution of RAG-1 in polyploid clawed frogs. Mol. Biol. Evol. 22: 1193–1207.
Evans, B. J., S. M. Bliss, S. A. Mendel, and R. C. Tinsley, 2011a The Rift Valley is a major barrier to dispersal of African clawed frogs (*Xenopus*) in Ethiopia. Mol. Ecol. 20: 4216–4230.
Evans, B. J., E. Greenbaum, C. Kusamba, T. F. Carter, M. L. Tobias *et al.*, 2011b Description of a new octoploid frog species (Anura: Pipidae: *Xenopus*) from the Democratic Republic of the Congo, with a discussion of the biogeography of African clawed frogs in the Albertine Rift. J. Zool. (Lond.) 283: 276–290.
Evans, B. J., R. A. Pyron, and J. J. Weins, 2012 Polyploidy and sex chromosome evolution in amphibians, pp. 385–410 in Polyploidy and Genome Evolution, edited by Soltis, P. S., and D. E. Soltis. Springer, Berlin.
Evans, B. J., T. F. Carter, E. Greenbaum, V. Gvoždík, D. B. Kelley *et al.*, 2015 Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (*Xenopus*, pipidae) from west and central Africa. PLoS One 10: e0142823.
Fujii, J., M. Kodama, A. Oike, Y. Matsuo, M.-S. Min *et al.*, 2014 Involvement of androgen receptor in sex determination in an amphibian species. PLoS One 9: e93655.
Gallach, M., S. Domingues, and E. Betrán, 2011 Gene duplication and the genome distribution of sex-biased genes. Int. J. Evol. Biol. 2011: 989438.
Gamble, T., J. Coryell, T. Ezaz, J. Lynch, D. P. Scantlebury *et al.*, 2015 Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. Mol. Biol. Evol. 32: 1296–1309.
Geraldes, A., T. Rambo, R. A. Wing, N. Ferrand, and M. W. Nachman, 2010 Extensive gene conversion drives the concerted evolution of paralogous copies of the *SRY* gene in European rabbits. Mol. Biol. Evol. 27: 2437–2440.
Glaubitz, J. C., T. M. Casstevens, F. Lu, J. Harriman, R. J. Elshire *et al.*, 2014 TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS One 9: e90346.
Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson *et al.*, 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29: 644–652.
Graves, J. A. M., and C. L. Peichel, 2010 Are homologies in vertebrate sex determination due to shared ancestry or to limited options? Genome Biol. 11: 205.
Gray, J. E., 1864 Notice of a new genus (*Silurana*) of frogs from West Africa. J. Nat. Hist. 14: 315–316.
Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood *et al.*, 2013 *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8: 1494–1512.
Harland, R. M., and R. M. Grainger, 2011 *Xenopus* research: metamorphosed by genetics and genomics. Trends Genet. 27: 507–515.
Heled, J., and A. J. Drummond, 2010 Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27: 570–580.
Hellsten, U., R. M. Harland, M. J. Gilchrist, D. Hendrix, J. Jurka *et al.*, 2010 The genome of the western clawed frog *Xenopus tropicalis*. Science 328: 633–636.
Inoue, J., Y. Sato, R. Sinclair, K. Tsukamoto, and M. Nishida, 2015 Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. Proc. Natl. Acad. Sci. USA 112: 14918–14923.
Irisarri, I., M. Vences, D. San Mauro, F. Glaw, and R. Zardoya, 2011 Reversal to air-driven sound production revealed by a molecular phylogeny of tongueless frogs, family Pipidae. BMC Evol. Biol. 11: 114.
Just, W., W. Rau, W. Vogel, M. Akhverdian, K. Fredga *et al.*, 1995 Absence of *Sry* in species of the vole *Ellobius*. Nat. Genet. 11: 117–118.
Kass, R. E., and A. E. Raftery, 1995 Bayes factors. J. Am. Stat. Assoc. 90: 773–795.
Katoh, K., and D. M. Standley, 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30: 772–780.
Kawai, A., J. Ishijima, C. Nishida, A. Kosaka, H. Ota *et al.*, 2009 The ZW sex chromosomes of *Gekko hokouensis* (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. Chromosoma 118: 43–51.
Kondo, M., I. Nanda, U. Hornung, S. Asakawa, N. Shimizu *et al.*, 2003 Absence of the candidate male sex-determining gene *dmrt1b(Y)* of medaka from other fish species. Curr. Biol. 13: 416–420.
Kondo, M., I. Nanda, U. Hornung, M. Schmid, and M. Schartl, 2004 Evolutionary origin of the medaka Y chromosome. Curr. Biol. 14: 1664–1669.
Koopman, P., J. Gubbay, N. Vivian, P. Goodfellow, and R. Lovell-Badge, 1991 Male development of chromosomally female mice transgenic for *sry*. Nature 351: 117–121.
Liu, L., L. Yu, and S. V. Edwards, 2010 A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10: 302.
Lloyd, R. E., P. G. Foster, M. Guille, and D. T. J. Littlewood, 2012 Next generation sequencing and comparative analyses of *Xenopus* mitogenomes. BMC Genomics 13: 496.
Lu, F., A. E. Lipka, J. Glaubitz, R. Elshire, J. H. Cherney *et al.*, 2013 Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet. 9: e1003215.
Mank, J. E., 2009 The W, X, Y and Z of sex-chromosome dosage compensation. Trends Genet. 25: 226–233.

Mank, J. E., D. E. Promislow, and J. C. Avise, 2006 Evolution of alternative sex-determining mechanisms in teleost fishes. Biol. J. Linn. Soc. Lond. 87: 83–93.
Matsubara, K., H. Tarui, M. Toriba, K. Yamada, C. Nishida-Umehara *et al.*, 2006 Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. Proc. Natl. Acad. Sci. USA 103: 18190–18195.
Matsuda, Y., Y. Uno, M. Kondo, M. J. Gilchrist, A. M. Zorn *et al.*, 2015 A new nomenclature of *Xenopus laevis* chromosomes based on the phylogenetic relationship to *Silurana/Xenopus tropicalis*. Cytogenet. Genome Res. 145: 187–191.
Matzuk, M. M., and D. J. Lamb, 2008 The biology of infertility: research advances and clinical challenges. Nat. Med. 14: 1197–1213.
Mikamo, K., and E. Witschi, 1966 The mitotic chromosomes in *Xenopus laevis* (Daudin): normal, sex reversed and female WW. Cytogenetics 5: 1–19.
Miura, I., 2008 An evolutionary witness: the frog *Rana rugosa* underwent change of heterogametic sex from XY male to ZW female. Sex Dev. 1: 323–331.
Myosho, T., Y. Takehana, S. Hamaguchi, and M. Sakaizumi, 2015 Turnover of sex chromosomes in celebensis group medaka fishes. G3 (Bethesda) 5: 2685–2691.
Nylander, J., 2004 *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Center, Uppsala University, Sweden.
Nylander, J. A., F. Ronquist, J. P. Huelsenbeck, and J. Nieves-Aldrey, 2004 Bayesian phylogenetic analysis of combined data. Sys. Biol. 53: 47–67.
Olmstead, A. W., A. Lindberg-Livingston, and S. J. Degitz, 2010 Genotyping sex in the amphibian, *Xenopus (Silurana) tropicalis*, for endocrine disruptor bioassays. Aquat. Toxicol. 98: 60–66.
O'Meally, D., T. Ezaz, A. Georges, S. D. Sarre, and J. A. Graves, 2012 Are some chromosomes particularly good at sex? Insights from amniotes. Chromosome Res. 20: 7–19.
Paradis, E., J. Claude, and K. Strimmer, 2004 APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20: 289–290.
Parker, H. W., 1936 Reptiles and amphibians collected by the Lake Rudolf Rift Valley expedition, 1934. J. Nat. Hist. 18: 594–609.
Peracca, M. G. 1898. Descrizione di una nuova specie di amfibio del gen. *Xenopus* Wagl. dell'Eritrea. Bollettino dei Musei di Zoologia e Anatomia Comparata della R. Universita di Torino 13 (321): 1–4.
Peters, W. C. H., 1844 Über einige neue fische und amphibien aus Angola und Mozambique. Monatsberichte der Königlichen Preussische Akademie des Wissenschaften zu Berlin 1844: 32–37.
Pokorna, M., and L. Kratochvíl, 2009 Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? Zool. J. Linn. Soc. 156: 168–183.
R Core Team, 2015 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria.
Revell, L. J., 2012 phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3: 217–223.
Rice, W. R., 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes on JSTOR. Evolution 41: 911–914.
Roberts, R. B., J. R. Ser, and T. D. Kocher, 2009 Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326: 998–1001.
Roco, A. S., A. W. Olmstead, S. J. Degitz, T. Amano, L. B. Zimmerman *et al.*, 2015 Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. Proc. Natl. Acad. Sci. USA 112: E4752–E4761.
Ross, J. A., J. R. Urton, J. Boland, M. D. Shapiro, and C. L. Peichel, 2009 Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). PLoS Genet. 5: e1000391.
Rovatsos, M., J. Vukić, P. Lymberakis, and L. Kratochvíl, 2015 Evolutionary stability of sex chromosomes in snakes. Proc. Biol. Sci. 282: 20151992.
Rovatsos, M., J. Vukić, and L. Kratochvíl, 2016 Mammalian X homolog acts as sex chromosome in lacertid lizards. Heredity 117: 8–13.
Schliep, K. P., 2011 phangorn: phylogenetic analysis in R. Bioinformatics 27: 592–593.
Schmid, M., S. Sims, T. Haaf, and H. Macgregor, 1986 Chromosome banding in amphibia. X: 18S and 28S ribosomal RNA genes, nucleolus organizers and nucleoli in Gastrotheca riobambae. Chromosoma 94: 139–145.
Seo, T.-K., 2008 Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25: 960–971.
Shaw, T. I., Z. Ruan, T. C. Glenn, and L. Liu, 2013 STRAW: species TRee analysis web server. Nucleic Acids Res. 41: W238–W241.
Smith, C. A., K. N. Roeszler, T. Ohnesorg, D. M. Cummins, P. G. Farlie *et al.*, 2009 The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. Nature 461: 267–271.
Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.
Sutou, S., Y. Mitsui, and K. Tsuchiya, 2001 Sex determination without the Y chromosome in two Japanese rodents *Tokudaia osimensis osimensis* and *Tokudaia osimensis* spp. Mamm. Genome 12: 17–21.
Takehana, Y., M. Matsuda, T. Myosho, M. L. Suster, K. Kawakami *et al.*, 2014 Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. Nat. Commun. 5: 4157.
Tymowska, J., 1991 Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*, pp. 259–297 in Amphibian Cytogenetics and Evolution, edited by Green, D. M., and S. K. Sessions. Academic Press San Diego, San Diego, CA.
Uno, Y., C. Nishida, Y. Oshima, S. Yokoyama, I. Miura *et al.*, 2008 Comparative chromosome mapping of sex-linked genes and identification of sex chromosomal rearrangements in the Japanese wrinkled frog (*Rana rugosa*, Ranidae) with ZW and XY sex chromosome systems. Chromosome Res. 16: 637–647.
Uno, Y., C. Nishida, C. Takagi, N. Ueno, and Y. Matsuda, 2013 Homoeologous chromosomes of *Xenopus laevis* are highly conserved after whole-genome duplication. Heredity 111: 430–436.
Uno, Y., C. Nishida, C. Takagi, T. Igawa, N. Ueno *et al.*, 2015 Extraordinary diversity in the origins of sex chromosomes in anurans inferred from comparative gene mapping. Cytogenet. Genome Res. 145: 218–229.
van Doorn, G. S., and M. Kirkpatrick, 2007 Turnover of sex chromosomes induced by sexual conflict. Nature 449: 909–912.
Veyrunes, F., P. D. Waters, P. Miethke, W. Rens, D. McMillan *et al.*, 2008 Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. 18: 965–973.
Vicoso, B., and D. Bachtrog, 2009 Progress and prospects toward our understanding of the evolution of dosage compensation. Chromosome Res. 17: 585–602.
Vicoso, B., J. Emerson, Y. Zektser, S. Mahajan, and D. Bachtrog, 2013 Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. PLoS Biol. 11: e1001643.
Wallis, M. C., P. D. Waters, M. L. Delbridge, P. J. Kirby, A. J. Pask *et al.*, 2007 Sex determination in platypus and echidna: autosomal location of *SOX3* confirms the absence of *SRY* from monotremes. Chromosome Res. 15: 949–959.
Weiss, J., J. J. Meeks, L. Hurley, G. Raverot, A. Frassetto *et al.*, 2003 *Sox3* is required for gonadal function, but not sex determination, in males and females. Mol. Cell. Biol. 23: 8084–8091.
Wells, D. E., L. Gutierrez, Z. Xu, V. Krylov, J. Macha *et al.*, 2011 A genetic map of *Xenopus tropicalis*. Dev. Biol. 354: 1–8.
Wyman, M. J., A. D. Cutter, and L. Rowe, 2012 Gene duplication in the evolution of sexual dimorphism. Evolution 66: 1556–1566.
Yoshimoto, S., E. Okada, H. Umemoto, K. Tamura, Y. Uno *et al.*, 2008 A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. Proc. Natl. Acad. Sci. USA 105: 2469–2474.
Zarkower, D., 2001 Establishing sexual dimorphism: conservation amidst diversity? Nat. Rev. Genet. 2: 175–185.
Zhao, D., D. McBride, S. Nandi, H. McQueen, M. McGrew *et al.*, 2010 Somatic sex identity is cell autonomous in the chicken. Nature 464: 237–242.

Communicating editor: S. I. Wright

Supplemental Information:

Sequential turnovers of sex chromosomes in African clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination

Benjamin L. S. Furman* **and Ben J. Evans***,[1]

*Biology Department, Life Sciences Building Room 328, McMaster University, 1280 Main Street West, Hamilton, ON L8S4K1

ABSTRACT Sexual differentiation is fundamentally important for reproduction, yet the genetic triggers of this developmental process can vary, even between closely related species. Recent studies have uncovered, for example, variation in the genetic triggers for sexual differentiation within and between species of African clawed frogs (genus *Xenopus*). Here, we extend these discoveries by demonstrating that yet another sex determination system exists in *Xenopus*, specifically in the species *X. borealis*. This system evolved recently in an ancestor of *X. borealis* that had the same sex determination system as *X. laevis*, a system which itself is newly evolved. Strikingly, the genomic region carrying the sex determination factor in *X. borealis* is homologous to that of therian mammals, including humans. Our results offer insights into how the genetic underpinnings of conserved phenotypes evolve, and suggest an important role for co-option of genetic building blocks with conserved developmental roles.

KEYWORDS

sex chromosomes
Xenopus
rapid evolution
sex chromosome turnover
SOX3

S1. SUPPLEMENTAL METHODS

S1.1. Distinguishing orthologous and homeologous sequences

Our phylogenetic analyses involved tetraploid species, and it was crucial to distinguish orthologous sequences (those that diverged from one another due to speciation events) from homeologous sequences (those that diverged from one another due to genome duplication). To accomplish this, we used a phylogenetic approach. Because genome duplication preceded speciation of the tetraploid species in our ingroup, orthologs of the different species are more closely related to one another than to homeologs of any of the species. In other words, a species will have a gene sequence that is more closely related to the gene sequence of another species, than the duplicate copy of that same gene within its own genome. Thus, we retained for analysis only those genes that had two deeply diverged lineages in at least one species (corresponding to the two sets of homeologs for that species) and assumed relationships within each of these lineages were orthologous.

To generate sequence alignments, we extracted putative coding sequence for every gene in each transcriptome assembly using Get_orfs_or_cds.py, which is part of the Galaxy Tool Shed (Cock *et al.* 2009; Blankenberg *et al.* 2014). We then used a modified reciprocal BLAST (Altschul *et al.* 1997) approach to collect homologous sequences. We BLASTed each transcriptome assembly and the *Xenopus laevis* Unigene database against the *X. tropicalis* Unigene database, and saved the top hit below a threshold *e-value* ($\leq e^{-}10$). We then blasted the *X. tropicalis* Unigene database back against each transcriptome assembly, and saved all hits below this threshold *e-value* (resulting in 15,109 *X. tropicalis* sequences with matching transcriptome sequence from the other species). We retained multiple sequences to ensure that we would capture both homeologs in the tetraploids when present, although this then required us to filter these data based on phylogenetic relationships to remove redundant sequences that were generated during transciptome assembly (see below). We retained for analysis only those BLAST matches that had at least three ingroup species with at least one species that had two sequences present (i.e. potential homeologs), leaving 7,794 sets of homologous sequences. Using a Perl script, we then generated individual sequence files for each best-BLAST result. Each file had one *X. tropicalis* ortholog and sequences from each tetraploid transcriptome matching that ortholog. We used MAFFT v.7 (Katoh and Standley 2013) to align each set of homologous sequences.

To ensure that we had enough data to make robust phylogenetic conclusions, we filtered the data sets to those that had at least 300 bp of ungapped alignment. However, some files did not meet this requirement because some sequences were short. Thus, we used a Perl script that, for alignments with less than 300 bp of ungapped alignment, would test for subsets of the data that a) match the previous taxon requirement (three ingroup + at least one species with two sequences), and b) meet the bp requirement of 300 ungapped sites. This script began by trying all combinations of taxa in an alignment that was one less than the total number of sequences (e.g. if 10 sequences in the total alignment, then all combinations of nine taxa would be tested). If none of these combinations met the taxa requirement, the file was discarded. If none of the combinations met the bp requirement (but could meet the taxa requirement), then all combinations of taxa that were two less than the total number of sequences were inspected (e.g. all combinations of eight taxa in the above example). The script would continue in this fashion, testing smaller and smaller combinations of taxa, until either both requirements were met (at which point a new alignment file would be generated with the particular combination of taxa), or the requirements could not be met and the file was discarded. The only exception to this execution were files with >15 sequences present. For these files only the total number of sequences minus one, and, if necessary, the total number minus two were tested; further combinations were deemed too computationally intensive to explore. If these files did not meet the requirements by the total number of sequences minus two, then they were discarded. If, at a given combination of taxa, multiple sets met the requirements, the script would select the combination that produced the longest ungapped alignment length. By doing this, we were able to salvage datasets that did not initially meet the bp requirement. This process left us with 3,781 gene alignments.

In addition to potential homeologs and orthologs, these alignments frequently included multiple slightly diverged but closely related sequences, which probably stemmed from allelic differences, sequence errors, misassembled transcripts, segmental duplications, and splice variants. For each alignment, we therefore used a phylogenetic approach to select representative orthologous and homeologous sequences. First, maximum likelihood trees were constructed using RAxML v.8.0.25 (Stamatakis 2014) using the GTRGAMMA model for all alignments, setting *X. tropicalis* as the outgroup, and with other parameters at the default settings. We did not perform model testing for this step because it was performed on each of two downstream phylogenetic analyses for alignments that passed an initial filter. The resulting phylogenies were then parsed using a script written in R (R Core Team 2015) and functions available in the ape (Paradis *et al.* 2004), phytools (Revell 2012), and phangorn (Schliep 2011) libraries (available with the Dryad repository doi:10.5061/dryad.00db7). This R script checked whether a tree had only two deeply diverged (putatively homeologous) sequences for at least one ingroup species. The script did this by comparing the age of the most recent common ancestors (MRCAs) of sequences within and between species; homeologs were expected to have a deeper MRCA than that of the orthologs. By identifying and comparing the MRCA of pairwise comparisons within and between species, we identified gene alignments in which there were only two lineages of sequences within the ingroup and where at least one species had one (putatively homeologous) sequence in each lineage. We then assumed that these homeologs were generated by one whole genome duplication (WGD) event prior to radiation of extant *Xenopus* species.

After building the maximum likelihood trees and filtering with the R script, we were left with 1,644 alignments with two deeply diverged *Xenopus* lineages for at least one species that we assumed were homeologous. We then built chronograms with each of these gene alignments using BEAST (Drummond *et al.* 2012). Each alignment was tested for a model of evolution using MRMODELTEST v.2 (Nylander 2004), and one MCMC chain was run for 25 million generations, using an estimated strict clock. We set the age of the root to 65 million years (my), with a standard deviation of 4.62 following (Bewick *et al.* 2012). Input files were prepared for BEAST using a Perl script. After inspecting the posterior distributions of parameter values from all analyses to ensure that stationarity had been reached for parameter estimates, we applied a conservative 25% burnin to all analyses and generated consensus trees using TREEANNOTATOR (part of the BEAST package). We then reinspected the trees with our R script, to confirm that there was still two deeply diverged lineages and identifiable homeologs, which left 1,600 alignments. At this point, because our best BLAST approach retained all matching sequences below a threshold e-value, some orthologs within each homeologous lineage were still represented by multiple sequences. We then selected the longest sequence for each species within each homeologous lineage using a Perl script. We then rebuilt the BEAST trees using two

chains run for 75 million generations (substitution models selected as before) and inspected all posterior distributions for convergence of parameter estimates (using the R package mcmcse; Flegal *et al.* 2016), and re-inspected the trees with our R script to ensure homeologs were still present.

The final data set consisted of 1,585 genes, with between five and 10 taxa (each ingroup species had two, one, or zero sequences in each alignment, depending on whether there was missing data from neither, one, or both homeologs). The individual gene alignments were 382–13,911 bp (first and third quartiles = 978–1,981 bp). The total aligned length of the data set had 2,696,030 bp, and an ungapped alignment had a total length of 788,627 bp. The proportion of missing genes for each homeolog is *X. borealis* = 0.55/0.58, *X. clivii* = 0.64/0.63, *X. allofraseri* = 0.52/0.51, *X. largeni* = 0.56/0.58, *X. laevis* = 0.14/0.14.

S2. SUPPLEMENTAL RESULTS AND DISCUSSION

S2.1. Multigene Phylogenetic Analyses of Nuclear DNA

Hereafter we refer to each set of orthologous sequences within subgenus *Xenopus* as the alpha and beta orthologs. The assignment of a particular set of orthologous sequences to each of these categories was arbitrary. The chronograms and maximum likelihood trees recovered from both concatenated data sets and the phylogeny recovered from the MPEST analysis all strongly supported a sister relationship of *X. borealis* and *X. clivii* in both the alpha and beta orthologs (Fig. S2; maximum likelihood results not shown). Similarly, the multigene *BEAST analysis also recovered strong support for this sister relationship (Fig. 1). This relationship had posterior/bootstrap support (chronograms and MPEST, respectively) of 1.0/100% in all analyses, which, in the case of the BEAST and RAXML analyses, may in part reflect overconfidence associated with analysis of concatenated data (Kubatko and Degnan 2007).

The relationships among *X. laevis*, *X. allofraseri*, and *X. largeni*, remain unresolved in all analyses. In fact, there was even conflict in the resolution of these three taxa between the alpha and beta orthologs within the same analysis (Fig. S2). For instance, in the concatenated BEAST analysis with all data, in the alpha orthologs supported a sister relationship of *X. largeni* and *X. laevis*, while the beta orthologs supported *X. largeni* and *X. allofraseri* as sister (Fig. S2). For the concatenated BEAST analysis with gapped sites removed, the alpha orthologs had the same resolution as the alpha orthologs in the full data set, but the beta orthologs supported *X. laevis* and *X. allofraseri* as sister taxa. For both concatenated chronograms, these various resolutions had posterior support 1.0 and bootstrap values >80% (and up to 100%). The MPEST analysis had the same results as the concatenated analysis with gapped sites removed, with alpha orthologs supporting *X. laevis* and *X. largeni* (93%) and beta orthologs supporting *X. laevis* and *X. allofraseri* (55%; Fig. S2). The maximum likelihood analyses also had different resolutions between the alpha and beta orthologs (results not shown). The *BEAST analysis also did not confidently resolve relationships among these three taxa, having only 0.48 posterior support for a sister relationship of *X. laevis* and *X. largeni* (Fig. 1).

S2.2. Phylogenetic discordance of individual gene trees

Using an R function, sister clade relationships were counted across the concatenated post burnin posterior distribution of each individual gene analysis (1,585 alignments; Analysis (i) in Methods). A sister relationship was counted only when at least one additional orthologous sequence was also present (to avoid inferring support when only two species were present and limit the search to topologically interesting relationships). Additionally, we collected the age of divergence for any inferred sister relationships for each tree in the posterior distribution of each gene tree analysis.

This analysis revealed considerable discordance among gene trees (Table S3). The strongest support for any clade was a monophyletic group of *X. borealis* and *X. clivii*, found in 52% and 47% of the alpha and beta orthologs, respectively, across the combined post-burnin posterior distribution of trees (with each gene having the same number of trees in the combined posterior distribution). The mean node age this group was 19.79 and 21.19 my for alpha and beta, respectively, and a large range (Table S3). A similar number of trees in the posterior distribution placed either *X. borealis* or *X. clivii* as the earliest branching 4*x*=36 tetraploid *Xenopus* species. Topologies that did not support *X. borealis* and *X. clivii* as sister taxa, generally estimated an older age of divergence (Table S3), which is consistent with the hypothesis that incomplete lineage sorting contributes to the gene tree discordance. No strongly supported resolution of relationships within the clade containing *X. laevis*, *X. allofraseri*, and *X. largeni* was recovered across the gene trees, with each of the three possible sister relationships being nearly equally supported, and with similar divergence times of about 13 my (Table S3). Less than 10% of the trees in the combined posterior distribution of trees supported alternate topologies that lacked the clade containing *X. laevis*, *X. allofraseri*, and *X. largeni* and the clade containing *X. borealis* and *X. clivii*.

S2.3. The sex determining region of *X. laevis* is not homologous to that of *X. borealis*.

The Genotype By Sequencing (GBS) results indicated that the sex determining regions of *X. borealis* and *X. laevis* are non-homologous. To ensure that the region containing *DM-W* is not sex linked in *X. borealis*, we amplified a gene linked to *DM-W* (as previous attempts to amplify *DM-W* in *X. borealis* have been unsuccessful; Bewick *et al.* 2011). The gene *RAB6A* has one homolog located physically close to *DM-W* (Uno *et al.* 2013). We designed primers for both homeologs using the *X. laevis* genome v7.1 and amplified portions of both homeologs in our *X. laevis* cross. Molecular polymorphism in the homeolog of *RAB6A*, located on the unplaced scaffold 68,908, exhibited a pattern of inheritance that was consistent with sex linkage in the *X. laevis* family, with an insertion-deletion mutation in the mother inherited by all daughters and no sons (11 daughters, seven sons; Fig. S4). Conversely, parental polymorphisms in both homeologs of *RAB6A* in the *X. borealis* cross did not exhibit sex-linked inheritance; a maternal SNP on the *X. borealis* ortholog of *X. laevis* scaffold 68,908 copy was shared among daughters and sons including seven out of 11 daughters and four out of 10 sons (the father appeared to carry a null allele but, similar to *AR*, this did not affect out ability to determine a lack of sex linkage). A maternal SNP on the *X. borealis* ortholog of the other *RAB6A* homeolog in *X. laevis* (located on scaffold 19,8991) was detected in three out of six daughters and four out of eight sons, indicating that this *RAB6A* homeolog is not sex-linked. A neighbor-joining tree of these sequences confirmed the orthology and homeology of these sequences, with one homeolog in *X. borealis* grouping with the sex-linked *RAB6A* sequences of *X. laevis*, and the other *X. borealis* sequences forming their own clade.

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, 1997 Gapped blast and psi-blast: a new generation of protein database search programs. Nucleic Acids Res. **25**: 3389–3402.

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

Bewick, A. J., F. J. J. Chain, J. Heled, and B. J. Evans, 2012 The Pipid Root. Sys. Biol. **61**: 913–926.

Blankenberg, D., G. Von Kuster, E. Bouvier, D. Baker, E. Afgan, N. Stoler, J. Taylor, A. Nekrutenko, *et al.*, 2014 Dissemination of scientific software with galaxy toolshed. Genome Biol. **15**: 403.

Cock, P. J., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, *et al.*, 2009 Biopython: freely available python tools for computational molecular biology and bioinformatics. Bioinformatics **25**: 1422–1423.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut, 2012 Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. **29**: 1969–73.

Flegal, J. M., J. Hughes, and D. Vats, 2016 *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA and Minneapolis, MN, R package version 1.2-1.

Katoh, K. and D. M. Standley, 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. **30**: 772–80.

Kubatko, L. S. and J. H. Degnan, 2007 Inconsistency of phylogenetic estimates from concatenated data under coalescence. Sys. Biol. **56**: 17–24.

Nylander, J., 2004 MrModeltest v2 distributed by author. evolutionary biology center, uppsala university.

Paradis, E., J. Claude, and K. Strimmer, 2004 {APE}: analyses of phylogenetics and evolution in R language. Bioinformatics **20**: 289–290.

R Core Team, 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Revell, L. J., 2012 phytools: An r package for phylogenetic comparative biology (and other things). Method Ecol. Evol. **3**: 217–223.

Schliep, K. P., 2011 phangorn: Phylogenetic analysis in r. Bioinformatics **27**: 592–593.

Stamatakis, A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**: 1312–3.

Uno, Y., C. Nishida, C. Takagi, N. Ueno, and Y. Matsuda, 2013 Homoeologous chromosomes of Xenopus laevis are highly conserved after whole-genome duplication. Heredity **111**: 430–6.

SUPPLEMENTAL FIGURES AND TABLES

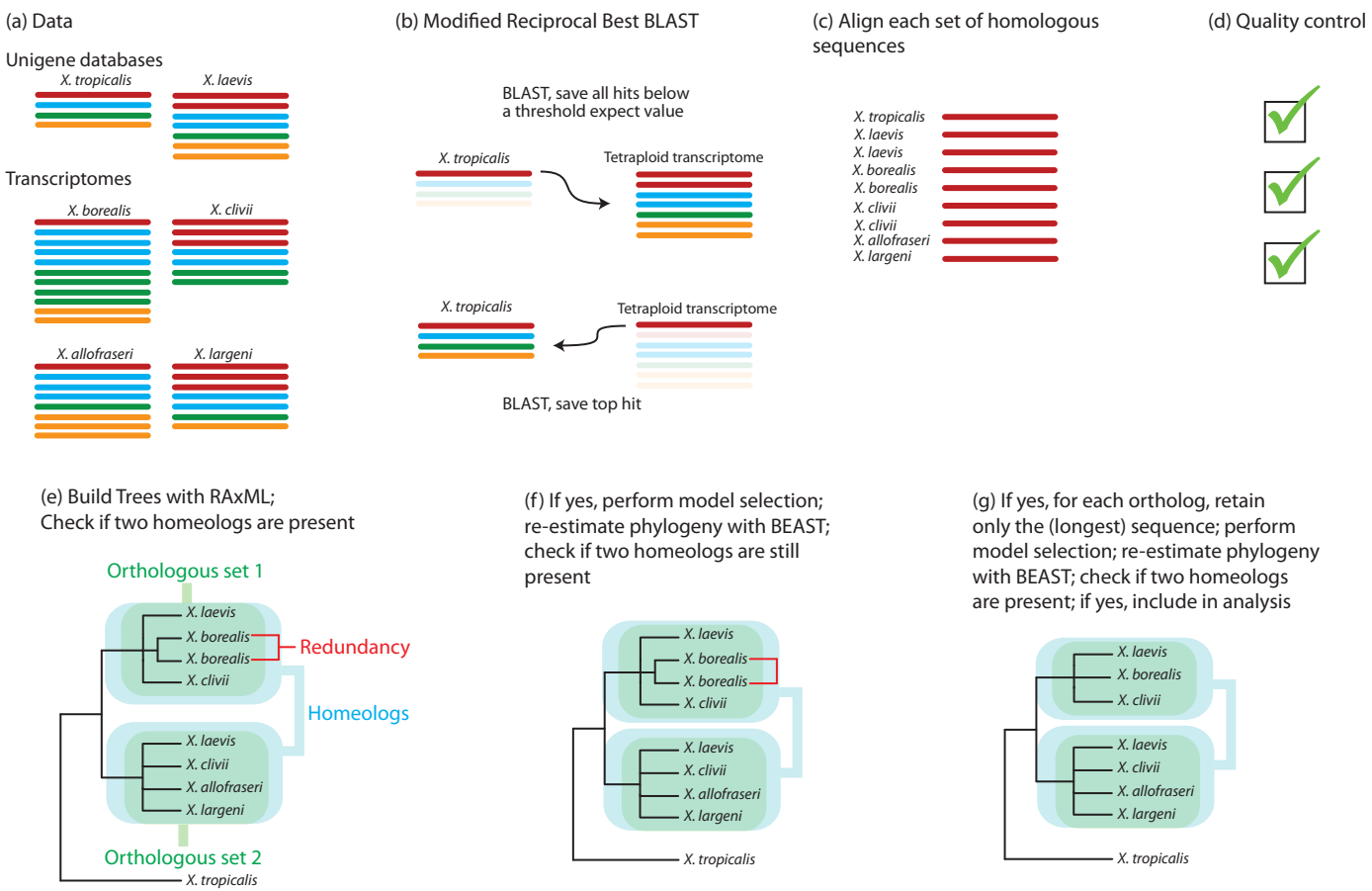

Figure S1 Our bioinformatics pipeline for identifying orthologous sequences used (a) Unigene and *de novo* transcriptome assemblies and (b) a modified reciprocal best BLAST hit approach to generate (c) sets of homologous sequences which were subjected to (d) quality control to ensure ungapped alignment length >300 bp, at least three ingroup species present, with at least one ingroup species with two (possibly homeologous) sequences present. We then used (e) RAxML to estimate a preliminary phylogeny from several thousand alignments. Phylogenies were parsed for homeologs and if present (f) model selection and BEAST analysis was performed. If homeologs were still present, the longest sequence from each ortholog was retained, and (g) another model selection and BEAST analysis was performed. If two homeologs were still present the alignment was included in downstream analyses. In (a-c) colors represent different genes; paralogs have the same color. In (a-f), redundancy includes allelic variants, splice variants, non-overlapping and overlapping gene fragments, and assembly errors. For some genes, one or both homeologs were not sequenced for some individuals.

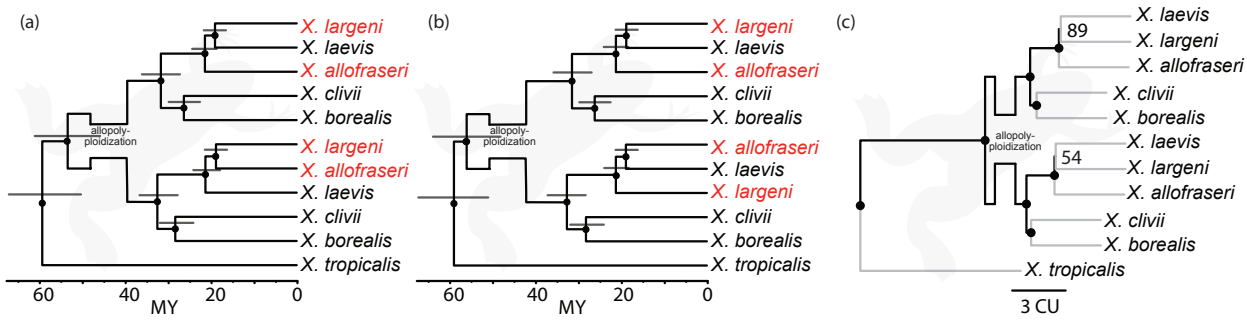

Figure S2 Analysis of nuclear data using BEAST with either (a) all gene alignments concatenated together or (b) concatenated gene alignment with gapped sites removed. Individual nuclear gene trees were also analyzed with (c) MPEST as described in the methods. In (c), grey lineages have arbitrary branch lengths, CU indicates coalescent units, and numbers indicate bootstrap support. Taxa with conflicting phylogenetic placement within each homeologous lineage are highlighted in red; other labeling follows Fig. 1

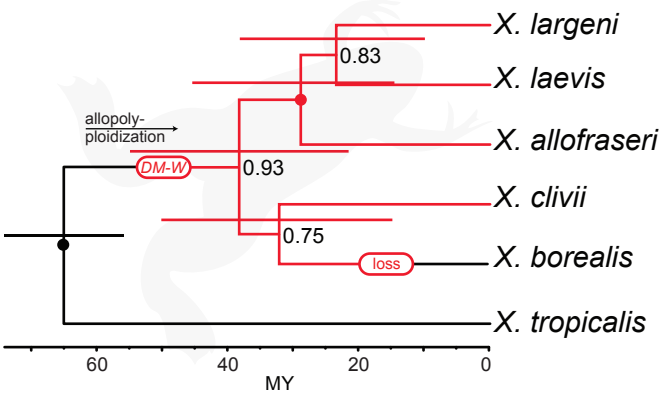

Figure S3 Bayesian analysis of mitochondrial DNA (mtDNA) alignments (after removing poorly aligned regions, see methods) with a relaxed molecular clock. This analysis produced an identical topology to that of the *BEAST analysis using nuclear DNA (nDNA) (Fig. 1). Labeling follows Fig. 1.

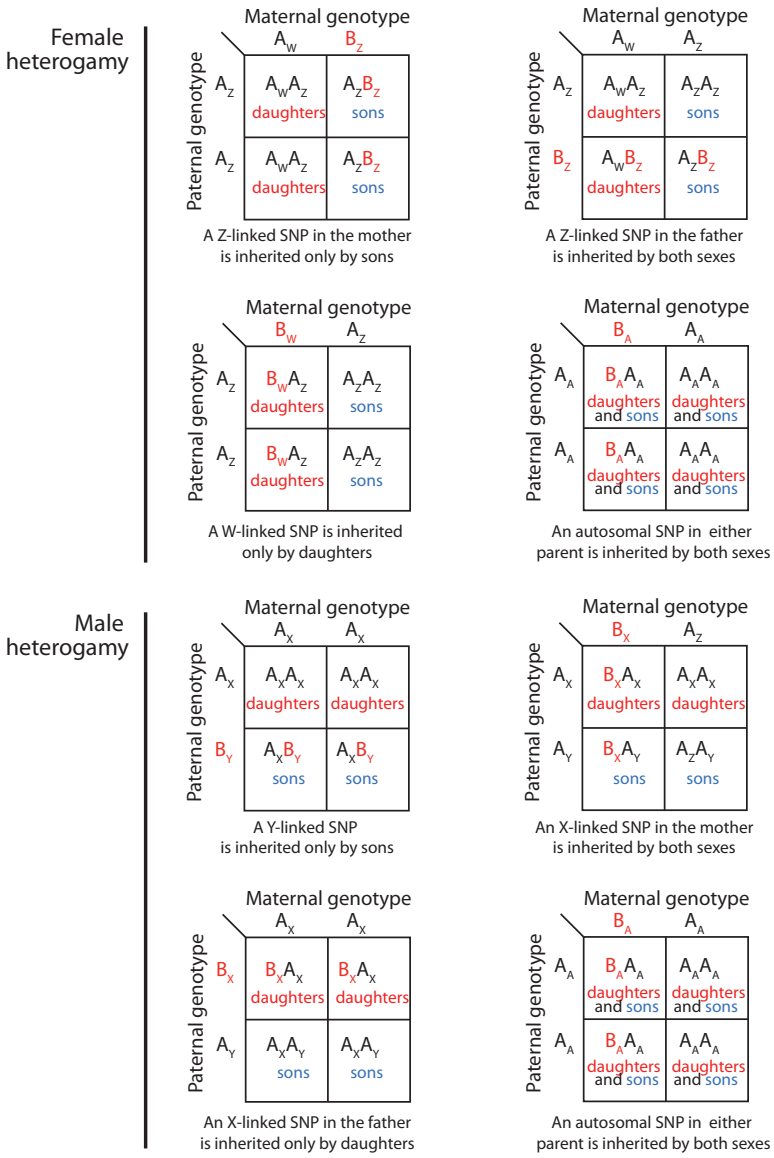

Figure S4 Not all SNPs are informative with respect to male versus female heterogamy. Diagnosis of female heterogamy requires a sex-linked SNP in the mother, and diagnosis of male heterogamy requires a sex-linked SNP in the father. For each parent, genotypes include nucleotides that are found in both parents (A) or only one (B) and that are linked to the W, Z, X, or Y chromosomes (W, Z, X, or Y subscripts respectively) or an autosome (A subscript)

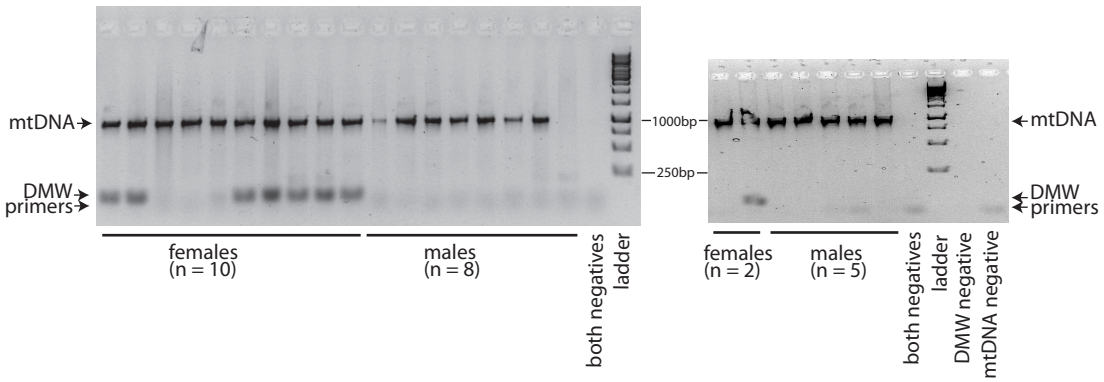

Figure S5 Attempts to amplify *DM-W* in wild *X. clivii* were only successful in females. mtDNA was used as a positive control and failed to amplify in one male.

Table S1 Transcriptome and GBS sequencing statistics. Raw sequence is number of reads; Trimmomatic is number of reads that passed our filter; Trinity is number of number of transcripts (N50 bp in brackets); Tassel values reflect reads for both lanes of sequencing (merged) and the total number of snps.

Species	Raw Sequence	Trimmomatic	Trinity	Tassel
X. allofraseri	20213893	18940455	96832 (1176)	-
X. borealis	19593759	19296751	81696 (1078)	-
X. largeni	19512126	19258973	82695 (1000)	-
X. clivii	18056373	17441776	72019 (885)	-
GBS	652154864	-	-	88996

Table S2 The primers used throughout the paper. Reverse primers provided in reverse compliment of the aligned sequence. The * denotes the additional primer combination used in the wild population samples (in text referred to as the alternate set of primers).

Gene	Primer name	Direction	Sequence
AR	begin	forward	ATGGCGGTGCACATAGGG
	down2	reverse	CGGGGGTCTCTTCGCTCT
SOX3	beta_for3	forward	GGTTTGGTCCCGGGGGCAGCGC
	beta_rev1	reverse	CTGAAGGGAAGAATGGTCGCC
	HiSeq_For5	forward	TAGGGAAGTTTTGTGCCGGGA
	HiSeq_Rev1	reverse	ACTCTGAAGGGAWGAGGGGTCG
	HiSeq_Rev5	internal reverse	GGGGTGGAAGCATTGCCCTTA
FMR1	alpha_f1_int	forward	TTTGTATTATGGTCTTTCAGGTGTATT
	alpha_r1	reverse	TTATGAATGAGCTTTTTGTTGCTGG
	alpha_f2*	forward	TGCCAACATACAACAGGCTAGGAAG
	alpha_r3*	reverse	CTAAGTTTCGTGGAACTTGTATAACATT
X. clivii DM-W	clivii_for2	forward	AATGAGGAACCATACAGCCCCAGGC
	clivii_rev2	reverse	GATTTCTGCATCGGGCACACCG
RAB6A	RAB6A_alpha_exon9_for2	forward	GCTCCTGTTAATGGCGCCCGTC
	RAB6A_alpha_exon9_rev1	reverse	CTGCTTATATATTAACAAGCCT
RAB6A	RAB6A_beta_exon9_for2	forward	GCTCCTGTTAATGTCGCCCATG
	RAB6A_beta_exon9_rev1	reverse	CTGCTTATATATTAACAAGCCC

Table S3 Conditional support and median ages with the 95% bounds for various sister relationships clades summarized across the combined post-burnin posterior distributions of individual gene tree analyses (Analysis (i), methods). The inferred sister relationship is dependent on the presence of the "condition" taxa being present (i.e. a sister relationship of *X. laevis* and *X. allofraseri* would be counted only if *X. largeni* was also in the alignment). Alpha and beta referred to the individual homeologous lineages generated by the WGD that preceded the speciation of extant *Xenopus* 4x=36 tetraploids.

homeolog	clade	condition	support	present	proportion	age	0.025	0.975
alpha	(*X. largeni*, *X. laevis*)	*X. allofraseri*	2769633	7320610	0.38	15.18	6.11	31.20
	(*X. allofraseri*, *X. laevis*)	*X. largeni*	2569313	7320610	0.35	15.10	4.46	28.47
	(*X. largeni*,*X. allofraseri*)	*X. laevis*	1806136	7320610	0.25	14.91	4.69	35.59
	(*X. clivii*, *X. borealis*)	any	3752122	7200600	0.52	22.21	7.37	40.52
	(*X. borealis*, any)	*X. clivii*	1428625	7200600	0.20	34.08	17.40	55.74
	(*X. clivii*, any)	*X. borealis*	1749065	7200600	0.25	33.64	10.54	57.36
beta	(*X. largeni*, *X. laevis*)	*X. allofraseri*	2700763	7344612	0.37	14.26	5.25	29.15
	(*X. allofraseri*, *X. laevis*)	*X. largeni*	2602804	7344612	0.35	14.90	4.87	28.99
	(*X. largeni*,*X. allofraseri*)	*X. laevis*	1899217	7344612	0.26	14.13	3.43	29.58
	(*X. clivii*, *X. borealis*)	any	3020271	6552546	0.46	23.52	6.06	41.69
	(*X. borealis*, any)	*X. clivii*	1602842	6552546	0.24	34.83	14.01	63.30
	(*X. clivii*, any)	*X. borealis*	1638245	6552546	0.25	32.65	8.85	62.56