

# Evolution of RAG-1 in Polyploid Clawed Frogs

Ben J. Evans,\* Darcy B. Kelley,† Don J. Melnick,‡ and David C. Cannatella§

\*Department of Biology, McMaster University, Hamilton, Ontario, Canada; †Department of Biological Sciences, Columbia University; ‡Center for Environmental Research and Conservation and Department of Ecology, Evolution and Environmental Biology, Columbia University; and §Section of Integrative Biology and Texas Memorial Museum, The University of Texas

Possible genetic fates of a gene duplicate are silencing, redundancy, subfunctionalization, or novel function. These different fates can be realized at the DNA, RNA, or protein level, and their genetic determinants are poorly understood. We explored molecular evolution of duplicated RAG-1 genes in African clawed frogs (*Xenopus* and *Silurana*) (1) to examine the fate of paralogs of this gene at the DNA level in terms of recombination, positive selection, and gene degeneration and in the absence of extensive recombination among alleles at different paralogs, (2) to test phylogenetic hypotheses about the origins of polyploid species. We found that recombination between different RAG-1 paralogs is infrequent, that degeneration of some paralogs has occurred via stop codons and frameshift mutations, and that this degeneration occurred in paralogs inherited from only one diploid progenitor species. Simulations and phylogenetic analyses of RAG-1 and mitochondrial DNA support one origin of extant tetraploids in *Xenopus* and at least one origin in *Silurana*, five allopolyploid origins of extant octoploids, and two allopolyploid origins of extant dodecaploids. In allopolyploid species, which inherit a complete genome from two different ancestors, genes inherited from the same ancestor have a longer period of coevolution than genes inherited from different ancestors. Because of this, gene ancestry could potentially influence gene fate: interacting paralogs derived from the same lower ploidy ancestor might have similar genetic destinies.

## Introduction

Gene duplication is an important driving force for biological innovation because it liberates one gene copy from purifying selection, permitting it to digress genetically and potentially to evolve a new function. Major episodes of gene duplication occurred at the base of the tree of life (Doolittle and Brown 1994; Pace 1997), in the ancestor of vertebrates, and again in the ancestor of teleost fishes (Holland and Garcia-Fernández 1996; Holland 1997; Vandepoele et al. 2002), in plants (Thompson and Lumaret 1992; Masterson 1994), and in other eukaryotes such as yeast (Goffeau 2004). Although genome duplication is relatively scarce in animals as compared to plants (Otto and Whitton 2000; Mable 2004), it plays a prominent role in speciation of the African clawed frogs of the pipid subfamily Xenopodinae (*Xenopus* and *Silurana*). Ploidy levels in this group range from diploid (2n) to dodecaploid (12n). Clawed frogs thus provide a powerful natural system for examining the genetic consequences of large-scale gene duplication.

Polyploidization can occur by spontaneous genome duplication (autopolyploidization) or by hybridization between species (allopolyploidization). Because allopolyploid clawed frogs have been produced by successively backcrossing unreduced hybrid eggs with sperm from each parental species (reviewed by Kobel 1996b), this mechanism appears more probable. Thus, phylogenetic lineages of clawed frogs may reticulate. To recover these reticulate patterns, information from nuclear genes is needed (fig. 1).

Estimation of phylogenetic relationships among allopolyploid species is more straightforward for disomic polyploids, in which each chromosome has only one homolog, than for polysomic polyploids, in which chromosomes form multivalents or random bivalents (Osborn et al. 2003), because recombination jumbles the phylogenetic signal of

gene duplicates in the latter configurations. Of course, a polyploid species may initially be polysomic and then become disomic over time. Recombination may also occur among different mitochondrial DNA (mtDNA) molecules within a cell (Ladoukakis and Zouros. 2001; Kraysberg et al. 2004), and the degree to which recombination might jumble the phylogenetic signal of this molecule should depend on the frequency of recombination and the extent of heteroplasmy. If one of these factors is low in mtDNA, then comparison of mtDNA and nuclear DNA (nDNA) genealogies could offer insight into the degree to which the phylogenetic signal in nDNA has been obscured by recombination among duplicated nDNA genes, and potentially could identify which paralogs were maternally versus paternally inherited in allopolyploids (fig. 1).

## The RAG-1 Gene

To better understand the evolution of clawed frogs, we has gathered sequence data from the RAG-1 gene, which has proved informative for phylogenetic studies (Groth and Barrowclough 1999; Hoegg et al. 2004). RAG-1 forms a heterodimer with a linked partner gene, RAG-2, that is essential for *V(D)J* recombination of DNA (Roth and Craig 1998; Brandt and Roth 2002). Both of these genes probably were once components of a transposable element that integrated into the genome of the ancestor of jawed vertebrates (Agrawal, Eastman, and Schatz 1997), marking the genesis of an adaptive immune system with somatic rearrangement of antigen receptor genes (DuPasquier, Zucchetti, and Santis 2004). All jawed vertebrates studied so far have adjacent RAG-1 and RAG-2 genes and immunoglobulin and T-cell receptor genes that generally require somatic recombination to be expressed (Litman 1993; Rast 1997). Using the RAG-mediated process of *V(D)J* recombination, B and T lymphocytes produce an almost limitless diversity of antigen receptors from a fixed number of genetic precursors (Brandt and Roth 2002). RAG-1 protein is encoded by only one exon in *Xenopus laevis*, and a previous study identified

Key words: reticulate evolution, genome duplication, polyploidization, ortholog, paralog, *Xenopus*, *Silurana*, Pipidae.

E-mail: evansb@mcmaster.ca.

*Mol. Biol. Evol.* 22(5):1193–1207. 2005

doi:10.1093/molbev/msi104

Advance Access publication February 9, 2005

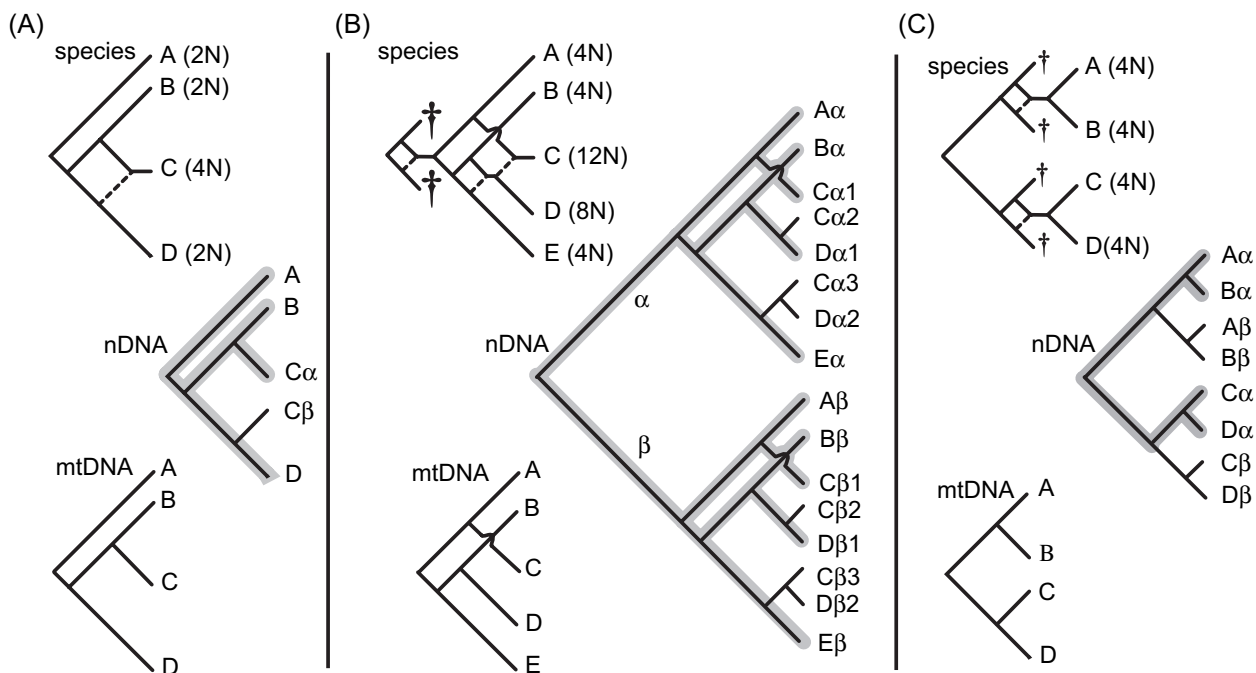


FIG. 1.—Species, mtDNA, and nDNA phylogenies for examples of allopolyploid evolution, assuming no recombination, gene conversion, or ancestral polymorphism. Species relationships reticulate; a solid line indicates the maternal relationship, and a dashed line indicates the paternal relationship. (A) Two diploids hybridize to produce a tetraploid. The nDNA phylogeny indicates that the tetraploid carries two gene paralogs, each derived from a different most recent common ancestor. The mtDNA phylogeny reveals only half of the ancestry of species C. (B) Diversification of a tetraploid derived from a single allopolyploidization event between two extinct diploid species that produces a tetraploid ancestor of all extant species. Subsequent diversification and allopolyploidization generate octoploid species D and dodecaploid species C. Tetraploids have two paralogous genes ( $\alpha$  and  $\beta$ ), octoploids have four ( $\alpha 1$ ,  $\alpha 2$ ,  $\beta 1$ ,  $\beta 2$ ), and dodecaploids have six ( $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$ ,  $\beta 1$ ,  $\beta 2$ ,  $\beta 3$ ). In this example, the most recent maternal ancestor of the dodecaploid is a tetraploid, and the mtDNA phylogeny reflects only a third of the ancestry of this species. (C) Multiple allopolyploidization events generate different tetraploid ancestors of extant species. Some intraspecific relationships among nDNA paralogs are closer than interspecific relationships. The full complexity of the evolutionary history of these species is not evident in the mtDNA phylogeny. In (A–C), relationships among maternally or biparentally inherited nDNA paralogs (shown as shadowed lines) are identical to mtDNA relationships.

only one copy of the RAG-1 and RAG-2 genes; these are physically linked by a 6-kb intergenic region (Greenhalgh, Olesen, and Steiner 1993).

In this study, we aim to evaluate the genetic consequences of gene duplication of the RAG-1 gene at the DNA level in terms of gene degeneration, positive selection, and recombination and—in the absence of extensive recombination among paralogs—to test hypotheses concerning reticulate relationships of clawed frogs.

## Materials and Methods

### Samples

Living pipoid frogs include the families Pipidae and Rhinophrynidae (Ford and Cannatella 1993). Within Pipidae, the subfamily Pipinae includes the New World genus *Pipa* and the African genera *Hymenochirus* and *Pseudhymenochirus*, and the subfamily Xenopodinae includes the African clawed frog genera *Xenopus* and *Silurana* (Cannatella and Trueb 1988b; de Sá and Hillis 1990; Trueb 1996; Evans et al. 2004). *Silurana* includes one diploid species with 20 chromosomes and three tetraploid species with 40 chromosomes. In *Xenopus*, tetraploids appear to have completely replaced diploids; this genus includes 10 tetraploid species with 36 chromosomes, 5 octoploid species with 72 chromosomes, and 2 dodecaploid species with 108 chromosomes (Kobel, Loumont, and Tinsley 1996);

no diploids (with 18 chromosomes) are known. Some species are not yet described (Evans et al. 2004). This study analyzed genetic samples from all described species of clawed frog, some undescribed ones, and other pipoids including *Pipa pipa*, *P. parva*, *Hymenochirus* sp., and *Rhinophrynus dorsalis* (table 1). The spadefoot toad *Scaphiopus huerterii* was used as an outgroup. Detailed information about sampling location and voucher specimens is in Evans et al. (2004).

## Amplification, Cloning, Sequencing, and Alignment of RAG-1 Genes

In *X. laevis*, RAG-1 is 1,045 amino acids long (Greenhalgh, Olesen, and Steiner 1993). Our data include DNA sequences within the open reading frame of RAG-1 from nucleotide positions 1685–2826; this region corresponds to 381 amino acids from positions 562–942 and spans almost the entire portion of RAG-1 necessary for heterodimerization with RAG-2 (Sadofsky et al. 1993; McMahan, Sadofsky, and Schatz 1997).

Pipid RAG-1 sequences were amplified with the polymerase chain reaction (PCR) and cloned with the TA cloning kit (Invitrogen, table 1). We did not sequence every possible paralog from some species. The following PCR primers were used for amplification: Xenrag1forward3: 5'-GGA TGA GTA TCC AGT AGA TAC AAT CTC

**Table 1**  
**Species, Ploidy, Clones, Chimeras, and GenBank Accession Numbers of Sequences in this Study**

Species	Ploidy	Number of Clones Sequenced	Number of Expected Paralogs	Number of Divergent Paralogs	Number of Chimeras (putative PCR artifacts)	nDNA GenBank accession number	mtDNA GenBank accession number
Outgroup							
<i>Scaphiopus hurterii</i>		NC	1	1	NC	AY874301	AY581619
<i>Silurana</i>							
<i>Silurana tropicalis</i>	20	2	1	1	0	AY874306 ( $\alpha$ )	AY581670
<i>Silurana cf. tropicalis</i>	20	6	1	1	0	AY874307 ( $\alpha$ )	AY581666
<i>Silurana epitropicalis</i>	40	5	2	2	2	AY874308 ( $\alpha$ ), AY874311 ( $\beta$ )	AY581665
<i>Silurana</i> new tetraploid 1	40	6	2	2	1	AY874313 ( $\alpha$ ), AY874312 ( $\beta$ )	AY581664
<i>Silurana</i> new tetraploid 2	40	3	2	2	0	AY874309 ( $\alpha$ ), AY874310 ( $\beta$ )	AY581667
<i>Xenopus</i>							
<i>Xenopus amieti</i>	72	17	4	3	2	AY874332 ( $\alpha$ 1), AY874331 ( $\alpha$ 2), AY874345 ( $\beta$ 1), NS ( $\beta$ 2)	AY581634
<i>Xenopus cf. fraseri</i> 1	36	15	2	2	0	AY874322 ( $\alpha$ ), AY874346 ( $\beta$ )	AY581632
<i>Xenopus cf. fraseri</i> 2	36	9	2	2	1	AY874323 ( $\alpha$ ), AY874324 ( $\beta$ )	AY581631
<i>Xenopus</i> new tetraploid	36	5	2	2	0	AY874329 ( $\alpha$ ), AY874355 ( $\beta$ )	AY581652
<i>Xenopus andrei</i>	72	8	4	3	0	AY874338 ( $\alpha$ 1), AY874337 ( $\alpha$ 2), AY874351 ( $\beta$ 1), NS ( $\beta$ 2)	AY581627
<i>Xenopus borealis</i>	36	7	2	2	0	AY874328 ( $\alpha$ ), AY874357 ( $\beta$ )	AY581653
<i>Xenopus boumbaensis</i>	72	9	4	3	2	AY874321 ( $\alpha$ 1), AY874325 ( $\alpha$ 2), AY874348 ( $\beta$ 1), NS ( $\beta$ 2)	AY581633
<i>Xenopus clivii</i>	36	13	2	2	5	AY874326 ( $\alpha$ ), AY874354 ( $\beta$ )	AY581637
<i>Xenopus gilli</i>	36	9	2	2	0	AY874314 ( $\alpha$ ), AY874342 ( $\beta$ )	AY581649
<i>Xenopus laevis</i>	36	4	2	2	0	AY874315 ( $\alpha$ ), AY874341 ( $\beta$ )	AY581642
<i>Xenopus largeni</i>	36	8	2	2	2	AY874327 ( $\alpha$ ), AY874343 ( $\beta$ )	AY581643
<i>Xenopus longipes</i>	108	14	6	5	2	AY874333 ( $\alpha$ 1), AY874334 ( $\alpha$ 2), AY874335 ( $\alpha$ 3), AY874349 ( $\beta$ 1), NS ( $\beta$ 2), AY874350 ( $\beta$ 3)	AY581625
<i>Xenopus muelleri</i>	36	18	2	2	1	AY874320 ( $\alpha$ ), AY874356 ( $\beta$ )	AY581657
<i>Xenopus pygmaeus</i>	36	19	2	1	0	AY874330 ( $\alpha$ ), NS ( $\beta$ )	AY581626
<i>Xenopus ruwenzoriensis</i>	108	23	6	4	0	AY874336 ( $\alpha$ 1), AY874339 ( $\alpha$ 2), AY874340 ( $\alpha$ 3), AY874347 ( $\beta$ 1), NS ( $\beta$ 2), NS ( $\beta$ 3)	AY581624
<i>Xenopus vestitus</i>	72	12	4	4	0	AY874319 ( $\alpha$ 1), AY874318 ( $\alpha$ 2), AY874353 ( $\beta$ 1), AY874352 ( $\beta$ 2)	AY581636
<i>Xenopus wittei</i>	72	23	4	3	0	AY874317 ( $\alpha$ 1), AY874316 ( $\alpha$ 2), AY874344 ( $\beta$ 1), NS ( $\beta$ 2)	AY581628
Other pipoids							
<i>Rhinophrynus dorsalis</i>		NC	1	1	NC	AY874302	AY581620
<i>Hymenochirus sp.</i>		1	1	1	0	AY874305	AY581623
<i>Pipa parva</i>		2	1	1	0	AY874304	AY581622
<i>Pipa pipa</i>		3	1	1	0	AY874303	AY581621

NOTE.—There is no voucher for some samples; further geographic information for each sample is available in Evans et al. (2004). In *X. borealis*, the mtDNA GenBank sequence is not from the same individual as the nDNA sequences. NS means not sequenced, and NC means not cloned. RAG-1 does not support separate species status of *S. tropicalis* and *S. cf. tropicalis*.

CAA GAG-3', Xenrag1rev2: 5'-TTT CTG GGA CAT GTG CCA GGG TTT TGT G-3'. One paralog of *Xenopus* new tetraploid (paralog  $\beta$ ) was amplified using primers designed to amplify this lineage: RAG1BETAF3: 5'-CTG TGA TGG GAT GGG AGA TGT G-3' and RAG1BETAR3: 5'-TGG ACA GGA GCT CTG CAA AGC GCT GG-3'. Two hundred and forty-one clones of pipid RAG-1 paralogs were directly amplified from colonies using vector primers M13 forward and M13 reverse and sequenced with these primers and internal ones: RAG1F4: 5'-GCA AGC CTC TCT GNC TGA TGC-3' and RAG1R4: 5'-GTT TTT ATA GAA CTC CCC TAT-3'. Multiple clones were sequenced in diploid species (*Silurana tropicalis*, *P. pipa*, *P. parva*) to explore the possibility that the RAG-1 gene was duplicated in diploids (table 1). *Rhinophrynus dorsalis* and *S. hurterii* were sequenced directly from amplified genomic DNA (courtesy of T. Townsend). Sequences were run on ABI 3100 and ABI 3730XL automated sequencers and edited with Sequencher version 4.1 (Gene Codes Corp.). Alignment was performed with ClustalX (Thompson et al. 1997) and then edited manually with MacClade version 4.06 (D. R. Maddison and W. P. Maddison 2000) taking codon frame into consideration. No regions of ambiguous homology were encountered.

Some sequences recovered from different clones possessed 1–3 polymorphic base pairs (bp). These differences might represent different alleles of the same gene, or they could be a result of mutation during PCR amplification, sequencing, or cloning. When multiple closely related sequences were identified, a representative sequence with the least number of autapomorphies was selected for further analysis of recombination and phylogeny; other closely related sequences were excluded.

### Analyses of Recombination

Chimerical sequences potentially can arise from recombination between alleles on homologous chromosomes, recombination between alleles at paralogous genes, recombination between alleles of orthologous genes via hybridization, and/or from PCR in which a partially extended amplified product primes a different gene in the next round of amplification. To identify chimerical sequences, we cloned and sequenced multiple copies of the alleles and employed a variety of tests for recombination.

Initially, sequences were inspected using MacClade, and chimerical sequences were identified that were composed of fragments that were identical, or almost so, to nonoverlapping portions of other divergent conspecific sequences (table 1). We interpret all of these chimeras to be derived from PCR because without exception (1) the number of unique sequences between putative break points was equal to or less than the number of genes expected by the species ploidy level (table 1) and (2) there is evidence (detailed below) that the divergent nonchimerical sequences carry a phylogenetic signal consistent with low or nonexistent recombination. Thus, we deleted chimerical clones from further analysis. In *Xenopus* cf. *fraseri* 2, a chimerical clone included 317 bp of unique sequence joined to a larger fragment that was identical to all other conspecific clones. This

unique fragment was included as a separate taxonomic unit (*X. fraseri* 2 paralog  $\beta$ ).

Simulations and empirical evaluations of tests for detecting recombination suggest that the performance of different methods varies with the level of divergence, the amount of recombination, and rate variation among sites (Posada and Crandall 2001; Posada 2002). Because these variables are generally not known with certainty beforehand, conclusions about the presence of recombination should not be based on results of a single test (Posada 2002). For this reason, we tested the included sequences for recombination using several approaches, including the Informative Sites Test (IST), the Recombination Detection Program (RDP), Geneconv, Chimaera, Bootscan, Siscan, and a Bayesian multiple change point (BMCP) model, and explored a range of parameter settings for each method. Details of these methods can be found elsewhere (Maynard Smith 1992; Salminen et al. 1995; Padidam, Sawyer, and Fauquet 1999; M. J. Gibbs, Armstrong, and A. J. Gibbs 2000; Martin and Rybicki 2000; Posada and Crandall 2001; Worobey 2001; Suchard et al. 2002). We used the first six methods to test for recombination separately in *Silurana* and *Xenopus*, excluding other pipoids and the outgroup. The last method, which has the advantage of testing for recombination while simultaneously inferring parental heritage, was used to test for evidence of recombination among sets of paralogs in each species. Correction for multiple tests is built into the program or incorporated in the test (BMCP), or the Bonferroni correction was applied (IST).

For IST, we used a likelihood ratio test and Modeltest version 3.06 (Posada and Crandall 1998) to select a model for phylogeny estimation and data simulation. In *Silurana*, 367 third-position sites were analyzed with the KG80 +  $\Gamma$  model. For *Xenopus* sequences, after removing the short fragment from *X. fraseri* 2 paralog  $\beta$ , other gaps were present; so only 214 third-position sites were analyzed with the HKY +  $\Gamma$  model. A successive approximation approach (Swofford et al. 1996) was used to estimate the most likely tree under each model; parameters used for PIST version 1.0 (Rambaut and Worobey 2001) were estimated from these trees.

For RDP (Martin and Rybicki 2000), internal and external references were used to determine the phylogenetic significance of sites for *Silurana*, as recommended for analysis of less than 10 sequences, using version 1.045. For *Xenopus*, only internal references were used, as recommended for data sets with more than 30 sequences (Martin and Rybicki 2000). Tests were carried out with window sizes of 10–100 variable sites. For Geneconv, sequences were scanned as triplets, and adjacent inserted or deleted characters were treated as a single polymorphism. Values for the mismatch penalty parameter ranging were varied from 1 to 5. The minimum aligned fragment length, minimum polymorphism in fragments, minimum pairwise fragment score, and maximum number of overlapping fragments were set to 1, 2, 2, and 1, respectively. For Chimaera, windows from 10 to 100 variable sites were explored, and 1,000 permutations were used to assess significance.

For Bootscan, we used a bootstrap cutoff of 95% and a sliding window of 200 bp, as in Posada and Crandall (2001), but increased the step size to 20 bp because lower step sizes produced an unreasonable level of false positives. To calculate distance matrices from replicated alignments,

the HKY model (chosen because it was one of the most complex models implemented by RDP) was used with base frequencies estimated from the data. The transition/transversion ratio for *Xenopus* was set at 2.0 and that for *Silurana* was set at 3.0; these values were estimated from a neighbor-joining tree using PAUP\* (Swofford 2002). For Siscan, we used the same window and step sizes as for Bootscan. Gaps were not considered, only variable sites within each triplet were examined, and permutations were used to assess significance.

We also used a BMCP model that permits crossover points along the length of a sequence to test for recombination, as implemented by Oh Brother (Suchard et al. 2002). For pairs of tetraploid species (each of which has two RAG-1 paralogs), we evaluated the probabilities of each of the three possible phylogenies among their paralogs along the length of the sequence. For octoploids and dodecaploids, more paralogs are present and many more topologies are possible. To narrow down candidate topologies, we followed Haake et al. (2004) in using multiple overlapping 100 nucleotide windows, estimating phylogenies of each window with MrBayes version 3.0b4 (Huelsenbeck and Ronquist 2001) and using the set of the five most probable trees from each window for analysis.

#### Gene Degeneration and Positive Selection at the DNA Level

Each paralog was screened for stop codons and frame-shift mutations by translating sequences using MacClade. Reading frame was obtained from the complete *X. laevis* RAG-1 gene (GenBank accession number L19324), which corresponds to *X. laevis* paralog  $\beta$  in this study. When deletions or insertions were present, we did not change the reading frame when determining the number of stop codons.

To test for positive selection, we compared the likelihood of the data under nested models of evolution using PAML version 3.14 (Yang 1997). One model allows the ratio of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site ( $\omega$ ) for each codon position to vary between 0 and 1 according to a beta distribution (approximated by 10 categories) or to have  $\omega$  higher than 1, with this value of  $\omega$  estimated from all sites under positive selection (model M8 in Yang and Nielsen 2002). The likelihood of this model was compared to the likelihood of another model in which  $\omega$  for all sites is less than or equal to 1 (model M7 in Yang and Nielsen 2002) using a likelihood ratio test, with two degrees of freedom. We performed tests for positive selection only on sequences that did not have inframe stop codons or frame-shift mutations and which thus potentially code for fully functional proteins.

#### Phylogenetic Analysis

Because we did not recover convincing evidence of interspecific or intraspecific recombination (see below), we performed phylogenetic analysis on all unique, divergent, and nonchimerical sequences using Bayesian analysis with MrBayes. We chose from different models of evolution based on Bayes factors as in Nylander et al. (2004). In

**Table 2**  
Harmonic Mean of Post-burn-in Likelihood Under Different Evolutionary Models

Model	Free Parameters	−ln(L)
F81	3	8,826.51
F81 + I	4	8,574.06
F81 + $\Gamma$	4	8,555.62
F81 + I + $\Gamma$	5	8,553.54
HKY	4	8,503.94
HKY + I	5	8,243.29
HKY + $\Gamma$	5	8,220.81
HKY + I + $\Gamma$	6	8,218.52
GTR	8	8,477.57
GTR + I	9	8,232.35
GTR + $\Gamma$	9	8,207.51
GTR + I + $\Gamma$	10	8,206.02
Separate GTR for each codon position, linked base frequencies, and I + $\Gamma$	20	8,193.49
Separate GTR and base frequencies for each codon position, linked I + $\Gamma$	26	8,183.46
Same GTR and base frequencies for all sites, separate I + $\Gamma$ for each codon position	14	8,084.85
GTR + I + $\Gamma$ , base frequencies positions 1 + 2, GTR + I + $\Gamma$ base frequencies position 3	20	8,080.77
Separate GTR + I + $\Gamma$ for each codon position, linked base frequencies	24	8,071.69
Separate GTR + I + $\Gamma$ and base frequencies for each codon position	30	8,058.04

NOTE.—Free parameters do not include branch length and topology parameters.

this approach, the harmonic means of the post-burn-in tree likelihoods of various models were compared using Bayes factors, with an interpretation of these values taken from Kass and Raftery (1995).

Eighteen nonpartitioned and partitioned models were explored (table 2). Nonpartitioned models included model F81 (Felsenstein 1981), HKY (Hasegawa, Kishino, and Yano 1985), and the general-time-reversible model (GTR; Tavaré 1986; Rodríguez et al. 1990) and each of these models with a proportion of invariable sites (I), a gamma distribution of rate heterogeneity among sites ( $\Gamma$ ), or both of these parameters. Partitioned models allowed the I parameter,  $\Gamma$  parameter, GTR rate matrix, and/or base frequencies to vary among codon positions (table 2). For each model, we ran four independent Metropolis-coupled Markov chain Monte Carlo analyses each for 2 million generations. Each run began with a random tree for each of four simultaneous chains, with flat Dirichlet prior distributions set to 1.0 for each rate substitution type and base frequency, and the differential heating parameter set to 0.2. Postrun analysis indicated that parameter estimates and tree likelihoods of all runs reached stationarity before 200,000 generations; to be conservative, these generations were discarded from each analysis as burn-in. These analyses were carried out on the SHARCNET cluster at McMaster University.

#### Testing Hypotheses About Origins of Ploidy

Phylogenetic hypotheses concerning the origins of polyploid species were evaluated using parametric bootstrap tests. A nonpartitioned GTR + I +  $\Gamma$  model of evolution was employed for simulations, using Seq-Gen

version 1.2.7 (Rambaut and Grassly 1997) and PAUP\* for analysis. Other details of the parametric bootstrap tests are in Evans et al. (2004). The sequential Bonferroni correction was applied (Rice 1989).

Phylogenetic hypotheses that we tested are divided into hypotheses concerning the origin of octoploids (Hypotheses 1A–C) and hypotheses concerning the origin of dodecaploids (Hypotheses 2A–C). Hypothesis 1A postulates two origins of octoploids, the minimum number suggested by mtDNA (Evans et al. 2004), with *Xenopus vestitus* originating from one instance of allopolyploidization and the other four octoploids originating from another (fig. 2). Under Hypothesis 1A, if *Xenopus amieti*, *Xenopus andrei*, *Xenopus wittei*, and *Xenopus boumbaensis* were all derived from one octoploid ancestor, barring recombination and gene conversion, each paralog in each of these species should have a closer relationship to an interspecific paralog from each of the other three species than to the other intraspecific paralogs (fig. 2). As shown in examples in figure 1, relationships are expected to be closer among some orthologous genes than among intraspecific paralogous genes when allopolyploid speciation precedes speciation of polyploids without a change in genome size. Hypothesis 1B postulates three separate origins of octoploids with *X. amieti* and *X. boumbaensis* sharing a most recent common octoploid ancestor, *X. wittei* and *X. andrei* sharing another most recent common octoploid ancestor, and *X. vestitus* being independently evolved. Hypothesis 1C postulates four separate origins of octoploids with *X. amieti* and *X. boumbaensis* sharing a most recent common octoploid ancestor and each of the other octoploids being independently evolved. Close relationships between *X. amieti* and *X. boumbaensis* mtDNA (Evans et al. 2004) and also between the  $\alpha 1$  paralog of these species (fig. 3) provided a rationale for uniting these taxa in Hypotheses 1B and 1C. A backbone constraint was employed for Hypotheses 1A–C to permit dodecaploid paralogs to have either an octoploid or a tetraploid maternal ancestor and to have any relationship with respect to octoploid paralogs.

Hypothesis 2A postulates a single origin of dodecaploids (*Xenopus ruwenzoriensis* and *Xenopus longipes*) and monophyly of sets of RAG-1 paralogs from a common octoploid and a common tetraploid ancestor (fig. 2). Hypothesis 2B is less constrained than 2A, and postulates only a single octoploid ancestor for both dodecaploids but permits each of the dodecaploids to have a different tetraploid ancestor. Hypothesis 2C postulates recent common ancestry of *X. amieti* and an octoploid maternal ancestor of dodecaploids. In Hypothesis 2C, *X. amieti* was selected as a putative close relative to the dodecaploids because the mtDNA sequences of these species are closely related (Evans et al. 2004) and, similar to Hypothesis 2B, this hypothesis also allows the tetraploid ancestors of each dodecaploid to be different.

## Results

### The Number of RAG-1 Paralogs Corresponds to Expectations from Ploidy

If duplication of the RAG-1 gene occurred only by genome duplication, diploids would be expected to have

one copy, tetraploids two, octoploids four, and dodecaploids six. After deleting putative PCR chimeras, we found that the number of differentiated sequences (table 1) was less than or equal to the number predicted by the ploidy of each species, under the assumption of a single copy of RAG-1 in diploids. This result is consistent with the assertion that the duplicate copies of RAG-1 are derived from entire genome duplication and not from duplication of this gene alone, although we did not sequence enough clones to statistically demonstrate this in all species. A search of available sequences from the *S. tropicalis* genome project (genome.jgi-psf.org/xenopus0/xenopus0.home.html) also found only one copy of RAG-1, as expected in a diploid. For some of the predicted genes, no alleles were found in the clones (table 1). This may have occurred by chance because we sequenced few clones per species, because these alleles did not amplify well with the primers that we used, or because some duplicated RAG-1 genes were deleted in polyploids. It is also possible that directional gene conversion homogenized a subset of the duplicate genes.

### Recombination Among RAG-1 Paralogs Is Infrequent

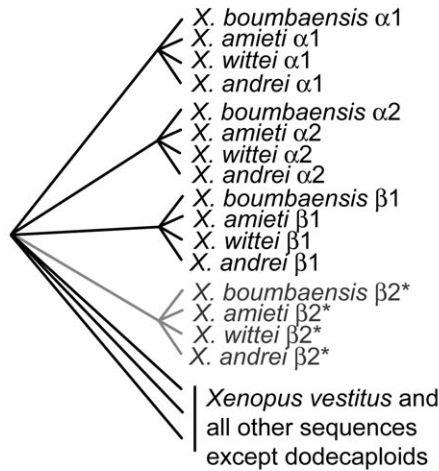
Most tests of recombination (RDP, Geneconv, Chimera, IST, and BMCP) did not detect significant evidence of intraspecific or interspecific recombination in *Xenopus* or *Silurana*. The only methods that detected putative recombinants were Bootscan and Siscan, and these methods did not identify any of the same putatively recombined regions at the settings we used.

One of the putative recombinants that Bootscan identified was *X. vestitus* paralog  $\beta 1$  with sister sequence *X. vestitus* paralog  $\beta 2$  as the major parent and *X. longipes* paralog  $\beta 3$  as a possible minor parent. However, because *X. vestitus* paralogs  $\beta 1$  and  $\beta 2$  are sister sequences (fig. 3), they are expected to be each others' major parent in a putative recombination event identified by a phylogenetic test of recombination such as Bootscan. No other putative recombinant identified by Bootscan or Siscan included two sequences from the same species.

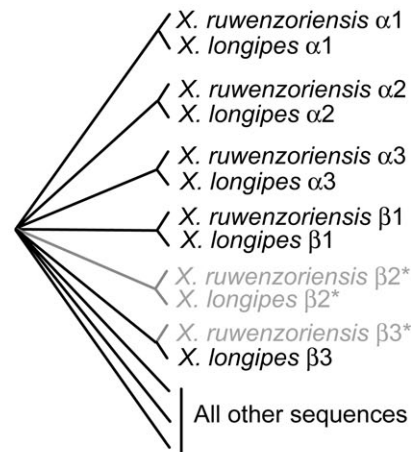
The results of Bootscan and Siscan analysis appear dubious because the putatively recombined regions are small (15–437 bp) and because none of the parent sequences are from the same species, except in *X. vestitus*, in which conspecific sequences are sister to one another. For these reasons, we suspect that the results of Bootscan and Siscan are actually a result of phylogenetic noise or variation in evolutionary rates along the length of the sequence. Indeed, a lack of congruence among different methods may suggest false positives in tests for recombination (Posada and Crandall 2001; Posada 2002). Comparison of the RAG-1 and mtDNA genealogies supports this assertion because well-supported orthologous relationships are similar in both genealogies (although not identical, see below).

In the absence of convincing evidence of recombination and barring gene conversion and ancestral polymorphism, genealogical relationships at RAG-1 should provide insights into species phylogeny. For further discussion, we treat duplicate intraspecific copies of RAG-1 as independent genes on separate and nonhomologous chromosomes.

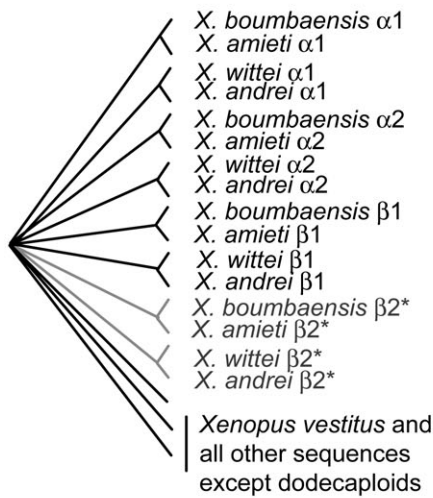
1A: two octoploid origin



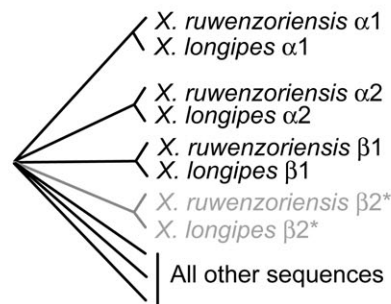
2A: single dodecaploid origin



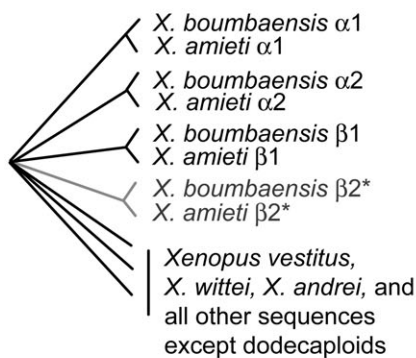
1B: three octoploid origins



2B: single dodecaploid maternal ancestor



1C: four octoploid origins



2C: dodecaploid most recent common maternal ancestry with X. amieti

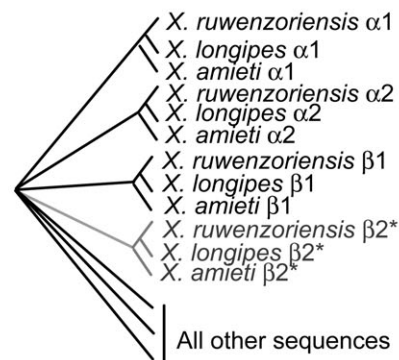


FIG. 2.—Hypotheses for the origin of polyploid species. Gray lineages depict expected relationships under each hypothesis among paralogs that were not sequenced.

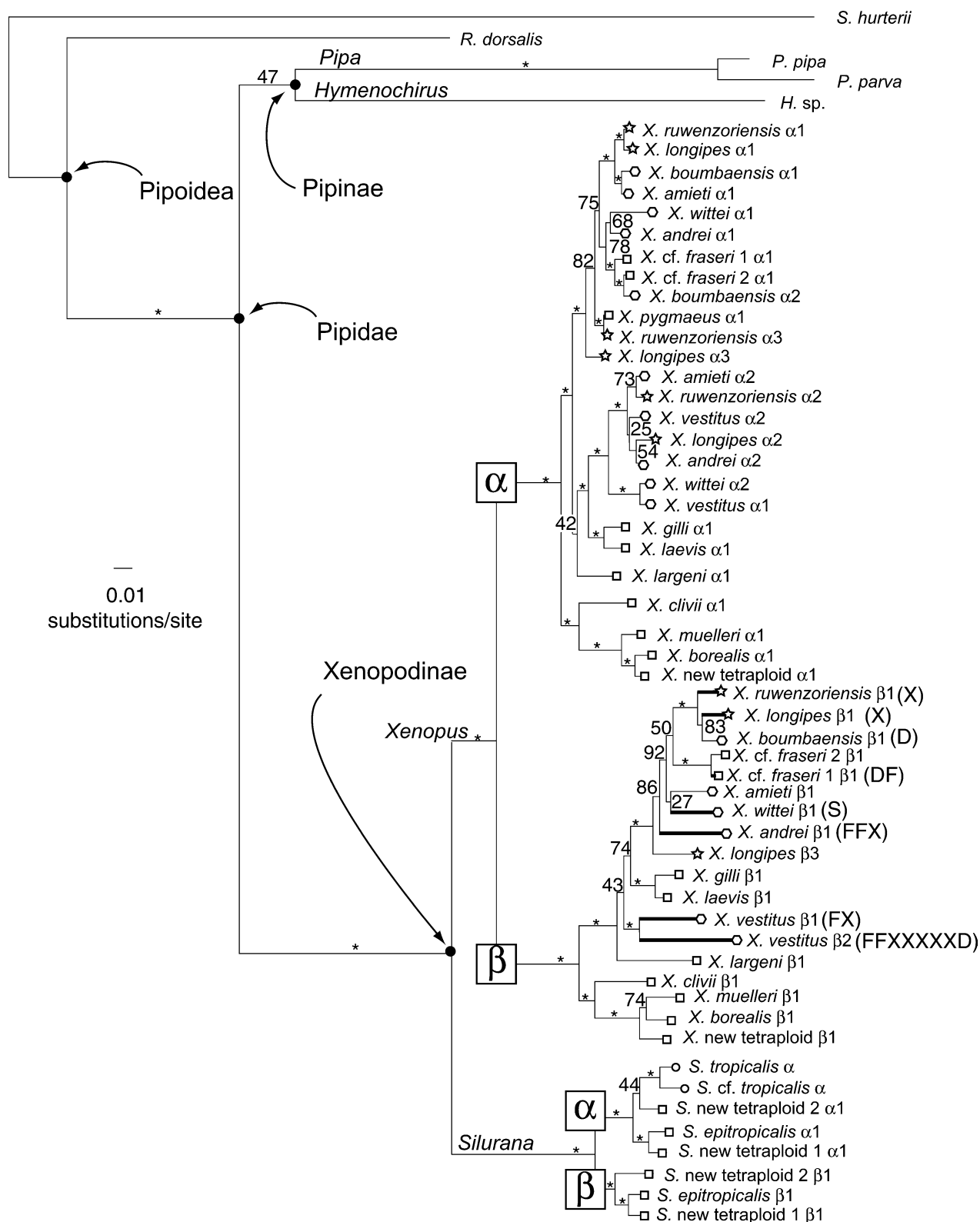


FIG. 3.—Phylogenetic relationships among pipoid nDNA inferred from partitioned Bayesian analysis partitioned using a different GTR + I +  $\Gamma$  model and separate base frequencies for each codon position. Posterior probabilities are shown as percentages, or, for values greater than or equal to 95, as an asterisk. The symbols  $\alpha$  and  $\beta$  indicate the major duplicated RAG-1 lineages in *Silurana* and *Xenopus*; if tetraploids originated via allopolyploidization, each of these lineages is derived from a different diploid ancestor. Branches terminate with a circle, square, polygon, or star to denote a paralog from a diploid, tetraploid, octoploid, or dodecaploid, respectively. Sequences with degenerate code due to stop codons or frameshift mutations have a thick branch. To the right of some sequences, the letters X, F, S, and D indicate the presence of a stop codon, frameshift deletion, frameshift insertion, and inframe deletion, respectively.

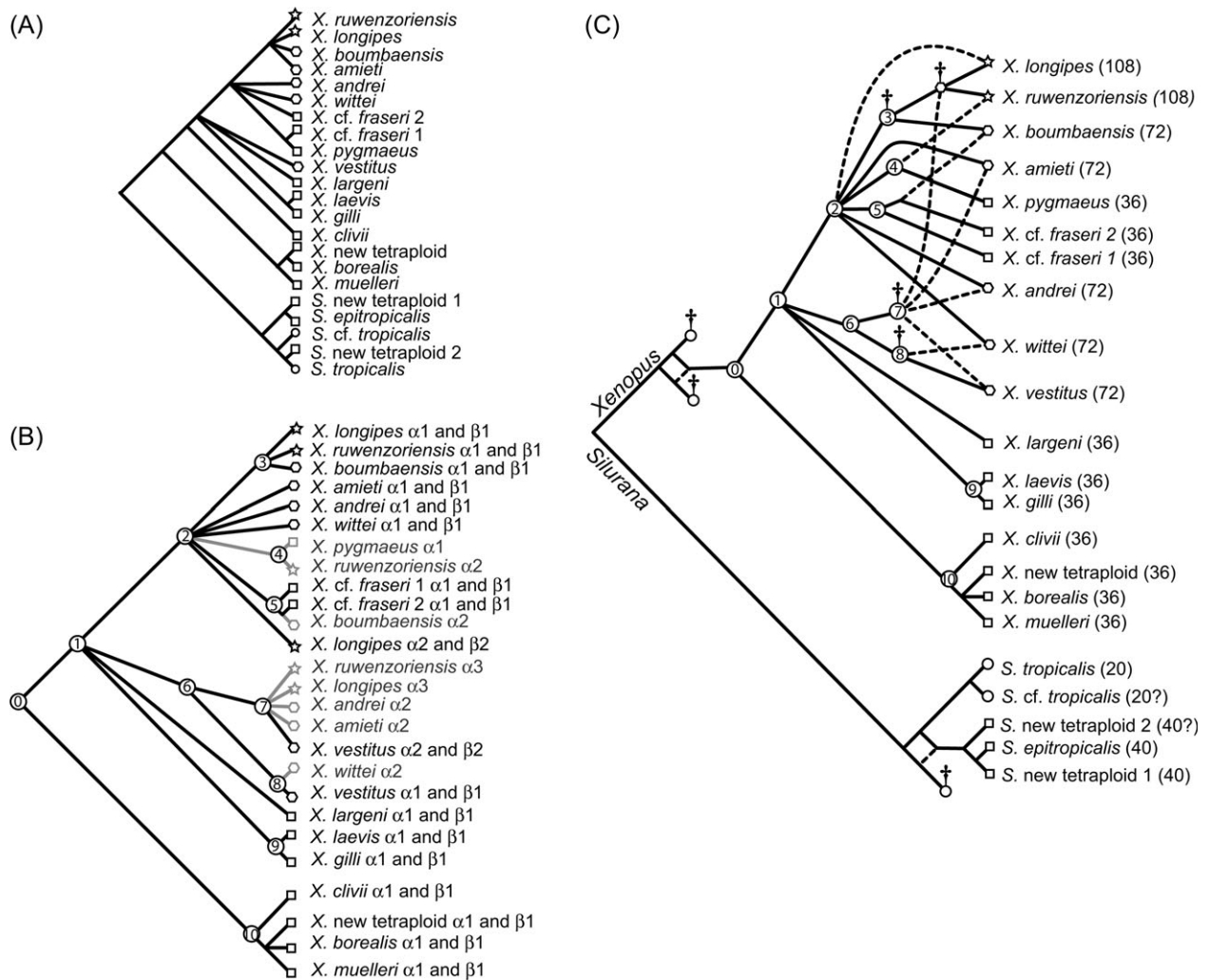


FIG. 4.—Reticulate relationships among clawed frogs as inferred from mtDNA and RAG-1 consensus phylogenies. (A) A mtDNA phylogeny from Evans et al. (2004) with branches with less than 95% posterior probability collapsed. (B) A consensus of the  $\alpha$  and  $\beta$  lineages of *Xenopus* with branches with less than 95% posterior probability collapsed and branches with conflicting topology also collapsed (*Silurana* paralogs are not shown). Paralogs that were found only in the  $\alpha$  lineage are in gray, and some nodes are numbered to facilitate comparison with (C). (C) Reticulate relationships among extant clawed frogs inferred from mtDNA and RAG-1. The numbers of chromosomes are indicated in parentheses with question marks following inferred ploidy levels of species that have not been karyotyped. Solid lines indicate maternal or biparental relationships, dashed lines indicate paternal relationships; nodes are numbered as in (B). The maternal and paternal ancestry of *X. vestitus* paralogs might also be the reverse of those depicted. Some aspects of this phylogeny are not consistent with the mtDNA phylogeny, possibly as a result of ancestral polymorphism. Dagggers ( $\dagger$ ) indicate inferred ancestral species for which extant descendants with the same ploidy are not known.

#### A Complex Model Is Preferred, but Other Models Recover Similar Trees

Bayes factors favored the most parameterized model which uses separate GTR + I +  $\Gamma$  and separate base frequencies for each codon position (table 2). Two times the logarithm of the Bayes factors of this model and all other models is greater than 10 and is indicative of a “very strong” improvement (Kass and Raftery 1995; Nylander et al. 2004).

To explore the concern that overparameterization may affect the consensus topology recovered by the favored model, we compared this topology to that recovered from a much simpler model that was still quite likely ( $-\ln L = 8,084.85$ ): this model partitioned the I and  $\Gamma$  parameters across codon positions but used one GTR rate matrix

and one set of base frequencies across all sites (table 2). Compared to the favored model, this simpler model uses about half of the free parameters (not including branch length and topology parameters). The consensus topology of the less parameterized model was exactly the same as that of the more parameterized model, with two exceptions. First, *Pipa* and *Hymenochirus* formed a paraphyletic assemblage, rather than a clade, as recovered by morphology and mtDNA (Cannatella and Trueb 1988a; Trueb and Báez 1997; Evans et al. 2004), whereas the topology recovered from the most parameterized model provides weak support for the currently accepted relationship with a posterior probability of 47% uniting *Pipa* and *Hymenochirus*. Second, an alternative relationship exists between *X. amieti* paralog  $\beta 1$  and *X. wittei* paralog  $\beta 1$ , with a posterior

probability of only 27% under the most parameterized model (fig. 3). Moreover, the well-supported relationships are identical in both analyses. Other simpler models also produced consensus trees that were very similar to that of the most complex model.

#### Mitochondrial DNA and RAG-1 Genealogies Are Similar

There is a high degree of congruence between the  $\alpha$  and  $\beta$  genealogies of RAG-1 (fig. 3) and between each of these genealogies and an mtDNA tree (Evans et al. 2004). Mitochondrial DNA and major lineages of RAG-1 all support the monophyly of *Xenopus largeni*, *X. laevis*, *Xenopus gilli*, *X. fraseri*, *Xenopus pygmaeus*, all octoploids, and both dodecaploids, but none provide resolution for the placement of *X. largeni* within this clade. Both data sets support monophyly of *X. laevis* and *X. gilli*, a close relationship between maternally inherited genes of *X. ruwenzoriensis*, *X. longipes*, and *X. boumbaensis*, and a sister relationship between *Silurana epitropicalis* and *Silurana* new tetraploid 1 (figs. 3 and 4; Evans et al. 2004). All support the monophyly of *Xenopus muelleri*, *Xenopus borealis*, and *Xenopus* new tetraploid, and mtDNA and the  $\alpha$  lineage of RAG-1 both provide strong support for a sister relationship between *X. borealis* and *Xenopus* new tetraploid.

Well-supported differences between RAG-1 and mtDNA genealogies could stem from ancestral polymorphism, phylogenetic noise, or undetected recombination. Some well-supported differences are present between mtDNA and both RAG-1 lineages, but few are present between the  $\alpha$  and  $\beta$  RAG-1 genealogies. In both major RAG-1 genealogies, a *Xenopus clivii* paralog is strongly allied to paralogs of *X. muelleri*, *X. borealis*, and *Xenopus* new tetraploid, which is consistent with the *muelleri* subgroup (Kobel, Loumont, and Tinsley 1996). However, mtDNA strongly supports the *X. clivii* haplotype as being more closely related to other *Xenopus* (fig. 4; Evans et al. 2004). Second, *X. fraseri* 1 is sister to *X. pygmaeus* in the mtDNA genealogy, whereas *X. cf. fraseri* 1 appears more closely related to *X. cf. fraseri* 2 in both RAG-1 genealogies. Third, the mtDNA grouping of *S. tropicalis* and *S. cf. tropicalis* is paraphyletic, but RAG-1 genes of these taxa form a clade (figs. 3 and 4).

#### A Reticulate Phylogeny of Clawed Frogs

To better understand speciation of clawed frogs, we synthesized bifurcating genealogies from mtDNA (Evans et al. 2004) and RAG-1 (fig. 3) into a reticulate phylogeny that reflects bifurcating speciation events without change in genome size and reticulating speciation events via allopolyploidization, as in figure 1. In both data sets, we first collapsed branches with less than 95% posterior probability (fig. 4A and B). In the RAG-1 data set, we also collapsed branches that conflicted between each of the  $\alpha$  and  $\beta$  lineages of RAG-1 (fig. 4B). A reticulate phylogeny was constructed from the remaining topologies, but allowing for some discrepancies with the mtDNA phylogeny (Fig. 4C).

The phylogeny suggests that extant tetraploids evolved once in *Xenopus* and at least once in *Silurana*.

Parametric bootstrap tests support separate origins of each extant octoploid species. Hypotheses 1A–C, which postulate fewer octoploid origins, are all rejected at  $P < 0.01$  (figs. 2–4) and were, respectively, 29, 33, and 19 steps longer than the unconstrained tree. Dodecaploids are also multiply evolved. Hypothesis 2A, which posits a single origin of dodecaploids, is rejected ( $P < 0.01$ ) and is four steps longer than the unconstrained tree. Simulations do not reject Hypothesis 2B, that the dodecaploids share recent common ancestry with the same octoploid species ( $P = 0.21$ ); this hypothesis is one step longer than the unconstrained tree. However, Hypothesis 2C, dodecaploid recent common ancestry with *X. amieti*, is rejected ( $P < 0.01$ ) and is nine steps longer.

Comparison of mtDNA and RAG-1 genealogies suggests that either the maternal or paternal ancestor of *X. vestitus* shared recent common ancestry with the paternal ancestor of *X. wittei*. This partial common ancestry corresponds with the *vestitus-wittei* group, defined on the basis of morphology (Kobel, Loumont, and Tinsley 1996), even though these species do not share closely related mtDNA (Evans et al. 2004) or recent ancestry of the other half of their genomes. The other ancestor of *X. vestitus* was closely related to some ancestors of *X. amieti*, *X. andrei*, *X. longipes*, and *X. ruwenzoriensis*. The maternal and paternal ancestors of *X. vestitus* were closely related, and the maternal and paternal ancestors of *X. boumbaensis* were closely related.

This phylogeny shares elements of previous phylogenetic hypotheses based on other types of data (immunological, karyological, electrophoretic, osteological), including the close relationship between *X. wittei* and *X. vestitus*, between *X. laevis* and *X. gilli*, and between *X. borealis*, *X. muelleri*, and *Xenopus* new tetraploid (Mann et al. 1982; Bürki and Fischberg 1985; Graf and Fischberg 1986; Tymowska 1991; Graf 1996; Kobel, Barandun, and Thiebaud 1998).

#### Evolutionary Fate of RAG-1 Duplicates at the DNA Level

Gene truncation or degeneration was indicated by inframe stop codons, frameshift deletions, or frameshift insertions in some paralogs that were carried by octoploids or dodecaploids and by one paralog carried by a tetraploid, *X. fraseri* (fig. 3). Both *X. ruwenzoriensis* paralog  $\beta 1$  and *X. vestitus* paralog  $\beta 2$  have the same stop codon at the same position. However, 11 more closely related sequences do not share this mutation, and therefore the most parsimonious explanation is that these stop codons evolved independently in each species. All other examples of observed gene degeneration are unique and thus also evolved independently (fig. 3). An inframe deletion was also present in the  $\beta 1$  paralog of the octoploid *X. boumbaensis*, but this paralog appears otherwise functional in the region we sequenced. All examples of gene degeneration or length modification are in paralogs in the  $\beta$  genealogy (fig. 3).

Our analyses recovered significant evidence of weak positive selection at 12 sites. The likelihood of the model M8 ( $-\ln L = -7,063.078$ ) was significantly better than that of model M7 ( $-\ln L = -7,073.830$ ,  $P < 0.001$ ,  $df = 2$ ). The parameters estimated for the beta distribution were

$p = 0.35061$  and  $q = 4.32445$ . The proportion of sites under positive selection was 0.034 and  $\omega = 1.13228$ . According to model M8, the individual amino acid positions and the probability of positive selection were as follows: 586 (0.825), 593 (0.996), 638 (0.999), 643 (0.9230), 707 (0.905), 730 (0.574), 757 (0.845), 821 (0.560), 825 (0.999), 878 (0.518), 883 (0.924), and 886 (0.864). Two of the three amino acid positions that are active sites for *V(D)J* recombinase (positions D600 and D708 in Kim et al. 1999) were spanned by these data. Each of these codons has silent polymorphisms at the third position in the nondegenerate sequences, but both positions are completely conserved at the amino acid level. A third residue thought to be important for catalytic activity (position R713 in Kim et al. 1999) was also conserved at the amino acid level in coding sequences, but this position was changed from arginine to glycine in *X. andrei* paralog  $\beta 1$  and missing due to a gap in *X. vestitus* paralog  $\beta 1$ . Both of these paralogs also had stop codons (fig. 3).

## Discussion

A nuclear gene, RAG-1, together with previous findings using mtDNA, was used to examine reticulate relationships among polyploid clawed frogs. In this clade, speciation is believed to occur through allopolyploidization. Species thus formed should contain both a maternal and a paternal nuclear genome that are each derived from an ancestor with fewer chromosomes. At a given locus, as many as six gene copies could be present in the case of dodecaploids, though a smaller number might be detected if genes are deleted, difficult to amplify, or lost, or if gene conversion homogenizes alleles at different loci. Indeed, we distinguished differentiated copies of the RAG-1 gene whose number was equal to or less than the number of copies expected on the basis of species ploidy (table 1).

One potential caveat in the use of duplicated genes to estimate phylogenetic relationships among these species is that recombination could occur either between different copies in one species or between species. Five of seven tests found no significant evidence for either intraspecific or interspecific recombination. In the remaining cases (Bootsan and Siscan), we attribute suspected examples of recombination to phylogenetic noise. If recombination occurs between genes with different evolutionary histories, conflicting phylogenetic signal is expected within the recombined gene (Schierup and Hein 2000). In contrast, we find that the major genealogies of RAG-1 are similar to one another and that each is also similar, though not identical, to the mtDNA genealogy. Together, these results suggest that recombination among duplicated copies of RAG-1 is infrequent. A low level of recombination among different paralogs is also consistent with studies that suggest that multivalents rarely form in polyploid clawed frogs (Tymowska 1991) but does not rule out the possibility that recombination occurs more frequently among other duplicated genes in clawed frogs.

The RAG-1 genealogy (figs. 3 and 4) suggests that at least 10 individual polyploidization events occurred in clawed frogs and most of them are definitively by allopolyploidization. *Xenopus* tetraploids originated once (by auto-

or allopolyploidization), and *Silurana* tetraploids originated once by allopolyploidization (fig. 3). Pairs of octoploids such as *X. vestitus* and *X. wittei*, and *X. amieti* and *X. boumbaensis*, share recent common ancestry of half of their nuclear genomes. Dodecaploids (*X. ruwenzoriensis* and *X. longipes*) originated twice and have different tetraploid ancestors; however, they may share recent ancestry with an octoploid species that has no known extant octoploid descendant.

Extinction of lower ploidy ancestors may have occurred on multiple occasions, although discovery of new species could prove otherwise. Depending on whether the ancestor of *Xenopus* was allo- versus autopolyploid, two or three diploid species may have gone extinct while their tetraploid descendants survived (fig. 4C). Three tetraploid ancestors of octoploids and at least one octoploid ancestor of dodecaploids are also not known and may be extinct (fig. 4C).

## The Fates of Duplicated RAG-1 Genes

Possible genetic fates of a gene duplicate are silencing, redundancy, subfunctionalization, or novel function. Alternatively, gene conversion might homogenize copies of a gene within a species but not prevent divergence of gene copies between species (Hurles 2004). Molecular evolution of RAG-1 appears to be inconsistent with gene conversion: gene paralogs have closer interspecific relationships than intraspecific relationships (fig. 3). However, in some species we did not detect all of the expected genes based on ploidy. This is probably due to insufficient sequencing and/or biased amplification of certain alleles, but it could also be caused by directional gene conversion of a subset of genes. Sammut, Marcuz, and Du Pasquier (2002) reported gene conversion in duplicated major histocompatibility complex genes of *X. ruwenzoriensis*—four expressed alleles were identified that shared the same deletion, and this deletion was not present in other closely related species, including *X. amieti*, *X. fraseri*, *X. wittei*, and *X. vestitus*. Future studies of additional nuclear genes are needed to shed light on the extent and variability in levels of recombination and gene conversion among duplicate genes at a genomic level and over time.

In most species, we identified at least two copies of RAG-1 that appear functional at the DNA level (no stop codons or frameshift mutations) in the region we sequenced. Two differently sized RAG-1 mRNA transcripts, potentially derived from different paralogs, are expressed in the thymus of *X. laevis* (Greenhalgh, Olesen, and Steiner 1993), and both may be functional at the protein level. These nondegenerate genes are thus candidates for redundancy, subfunctionalization, or novel function. A significant signal of weak positive selection ( $\omega = 1.13$ ) was detected at some sites in these genes, but these sites do not include residues critical to *V(D)J* recombination. The estimated value for  $\omega$  over sites under positive selection is near to that expected under neutral evolution, but some individual sites show a high probability of weak positive selection. However, caution should be exercised in making strong conclusions about positive selection when the estimated  $\omega$  is only marginally greater than 1 (Wong et al. 2004). If novel function has indeed evolved in duplicated

RAG-1 genes, it more likely concerns nuances of RAG-1 function, such as specificity of each copy of RAG-1 for a particular copy of RAG-2, rather than affecting the role of RAG-1 in  $V(D)J$  recombination.

Seven examples of sequences with unique (apomorphic) instances of gene degeneration (stop codons and/or frameshift mutations) were observed, and all of them occurred in the  $\beta$  genealogy of RAG-1 (fig. 3), as opposed to occurring randomly either in the  $\alpha$  or in the  $\beta$  genealogy. If expressed, these degenerate genes are almost certainly nonfunctional for  $V(D)J$  recombination because mutations have disrupted amino acids that are necessary for this process (Kim et al. 1999). If the seven examples of gene degeneration did occur independently, the probability that all seven occurred by chance in the same RAG-1 genealogy ( $\alpha$  or  $\beta$ ) is equivalent to the probability of seven coin tosses yielding all heads or all tails ( $P = 0.015625$ )—an improbable event.

One explanation for this pattern of gene degeneration is that gene silencing of RAG-1 actually occurred fewer times than is suggested by the number of apomorphic degenerations in the portion of the  $\beta$  paralogs that were sequenced. For example, a mutation in the promoter or in the 5' region that we did not sequence might have caused gene silencing in the  $\beta$  paralog of RAG-1 in an ancestor. Descendants of this ancestor would inherit this silenced gene, and further degeneration might then have occurred in the 3' portion of some (but not all) of the  $\beta$  paralogs of RAG-1. A lack of observable gene degeneration in some closely related paralogs (such as *X. amieti*  $\beta$ 1, *X. longipes*  $\beta$ 3, *X. laevis*  $\beta$ 1, and *X. gilli*  $\beta$ 1) is inconsistent with this explanation, but this could be because insufficient time has elapsed since ancestral gene silencing for gene degeneration to occur in all of the descendant paralogs. If the two differently sized RAG-1 transcripts that are expressed in *X. laevis* thymus (Greenhalgh, Olesen, and Steiner 1993) are derived from different paralogs, this is also potentially inconsistent with ancestral gene silencing of a RAG-1  $\beta$  paralog, depending on the relationship of *X. laevis* with respect to species with RAG-1 degeneration. However, these differently sized transcripts could also be derived from alternative splicing of an untranslated portion of the RAG-1  $\alpha$  paralog (and not from the  $\beta$  paralog).

If gene silencing did occur independently seven times in the RAG-1  $\beta$  genealogy, another explanation is that there was selection for in nonrandom gene degeneration. In allopolyploids, especially octoploids and dodecaploids, it could be advantageous if functional copies of RAG-1 are mostly  $\alpha$  copies—which are ultimately derived from the same diploid ancestor—rather than a random mixture of  $\alpha$  and  $\beta$  copies—which are derived from different diploid ancestors. This might be the case if degeneration of the  $\beta$  paralogs of RAG-2 (which are linked, coinherited, and coevolved with the  $\beta$  paralogs of RAG-1) occurred in an ancestor of these polyploid species and then was inherited by the polyploid descendants. Under this scenario, gene ancestry might influence gene fate in allopolyploid species because selection might favor exclusive interactions among genes in a heterodimer that are derived from the same ancestor. As a further test of the influence of gene ancestry on gene fate, it is necessary to determine whether degeneration of related copies of RAG-2 also occurred and, if so, whether they

are derived from the same ancestor as the degenerate RAG-1 genes.

### Speciation by Allopolyploidization

Beyond the level of individual genes lies the question of why speciation by allopolyploidization is so prevalent in clawed frogs, even though it is comparatively less common in other animals (Mable 2004). Of potential relevance is the observation that females are the heterogametic sex but hybrid males are sterile (Chang and Witschi 1956; Kobel 1996b), a violation of Haldane's (1922) rule. There is evidence that Haldane's rule for sterility in species with male heterogamy is a result of (1) dominance (genes that cause sterility are mostly recessive and thus more adversely affect the heterogametic sex) and also (2) faster evolution of genes expressed only in males that cause hybrid sterility, probably because of sexual selection (Orr 1997; Presgraves and Orr 1998). In species with female heterogamy, only the first mechanism is thought to apply. In clawed frogs, however, one could speculate that Haldane's rule is violated because fast evolution of male-expressed genes drives hybrid males to sterility to a greater degree than recessive mutations sterilize heterogametic hybrid females. Whatever the cause of this violation of Haldane's rule, fertility of hybrid females provides an evolutionary opportunity for allopolyploid speciation via backcrossing with parental males.

Moreover, allopolyploid clawed frogs probably form in nature in three steps: reproduction among species produces a hybrid female, unreduced diploid eggs from this female are crossed with haploid sperm from one parental species to make a triploid female, and then unreduced triploid eggs from this female are crossed with haploid sperm of the other parental species to produce a tetraploid (Kobel 1996a). The challenge of sex determination in the allopolyploid clawed frogs could be overcome by temperature-dependent sex determination or by variation among species in the dominance of sex-determining factors (Kobel 1996b).

The success of allopolyploid species could be attributable in part to the resistance to the parasites of both of their ancestors (Jackson and Tinsley 2003) or because the rate of adaptation is faster at higher ploidy levels if beneficial alleles are partially dominant (Otto and Whitton 2000). Disomic polyploid genomes might also have the advantage of avoiding complications associated with multivalents during cell division and allopolyploids may become disomic quicker than autopolyploids. The possibility that polyploids have a wider ecological tolerance is suggested in *Xenopus* because tetraploids have completely replaced diploids but is not consistent with the current distributions of octoploids and dodecaploids, which have much smaller ranges than most tetraploids (Tinsley, Loumont, and Kobel 1996). Of course, we do not know how the distributions of various ploidy levels varied in the past, and on an evolutionary timescale, the success of polyploids relative to their lower ploidy ancestors is difficult to assess empirically because the rates of allopolyploid speciation and extinction are unknown.

## Conclusions

In conclusion, recombination between alleles of different duplicated RAG-1 genes appears to be infrequent in clawed frogs, weak positive selection was identified at some amino acid positions, and gene degeneration occurred independently in some closely related paralogs in the  $\beta$  lineage of RAG-1. Comparison of RAG-1 and mtDNA genealogies suggests that extant tetraploid species evolved twice (once in *Xenopus* and at least once in *Silurana*) but that extant octoploids are each of independent allopolyploid origin. Dodecaploids evolved twice but may share the same octoploid ancestor. Additional fieldwork and data from other nuclear loci undoubtedly will provide further phylogenetic resolution and information on the ancestry of polyploids, and will shed light on the possibility of extinction of lower ploidy ancestors. Future studies will also offer insight into variation in recombination and gene conversion among paralogous genes, restructuring of these polyploid genomes (D. E. Soltis and P. S. Soltis 1999; Ozkan, Levy, and Feldman 2001), how and whether the extent of recombination has changed over time, and the mechanisms that drive speciation of this fascinating group.

## Acknowledgments

We are grateful to D. Rungger, J. Montoya-Burgos, C. Thiebaud, A. Solaro, M. Picker, and M. Lebreton, and the Station de Zoologie Expérimentale, Université de Genève, for their hospitality. We also thank P. Chippindale, B. Colombelli, L. Du Pasquier, L. Du Preez, M. Fischberg, D. Foguekem, D. Hillis, M. Klemens, H. Kobel, J. Le Doux Diffo, G. Legrand, C. Lieb, C. Loumont, J. Perret, D. Rungger, M. Ruggles, E. Rungger-Brandle, C. Thiebaud, R. Tinsley, T. Titus, and J. Tymowska for their fieldwork, without which this study would have been impossible. For loans of genetic samples, we thank the following individuals and their respective institutions (museum abbreviations follow Leviton et al. 1985): J. Montoya-Burgos, S. Fisch-Muller, and A. Schmitz (MHNG); J. Vindum and A. Leviton (CAS); L. Ford (AMNH), T. LaDuc (TNHC), L. Trueb, and J. Simmons (KU); and R. Murphy (ROM). We also thank ministries of the environment, science, and forestry of the many countries that granted research permission and export permits for these genetic samples. We thank T. Townsend for providing RAG-1 sequences for *S. hurterii* and *R. dorsalis* and for providing primers that were used to gather preliminary information for primer design. We thank D. Martin, M. Suchard, and M. Worobey for their advice on the analysis of recombination. We also thank B. Golding, A. Holloway, M. Tobias, and two anonymous reviewers for comments.

## Literature Cited

- Agrawal, A., Q. M. Eastman, and D. G. Schatz. 1997. Implications of transposition mediated by V(D)J recombination proteins RAG1 and RAG2 for origins of antigen-specific immunity. *Nature* **394**:744–751.
- Brandt, V. L., and D. B. Roth. 2002. A recombinase diversified: new functions of the RAG proteins. *Curr. Opin. Immunol.* **14**:224–229.
- Bürki, E., and M. Fischberg. 1985. Evolution of globin expression in the genus *Xenopus* (Anura: Pipidae). *Mol. Biol. Evol.* **2**:270–277.
- Cannatella, D. C., and L. Trueb. 1988a. Evolution of pipoid frogs: intergeneric relationships of the aquatic frog family Pipidae (Anura). *Zool. J. Linn. Soc.* **94**:1–38.
- . 1988b. Evolution of pipoid frogs: morphology and phylogenetic relationships of *Pseudhymenochirus*. *J. Herpetol.* **22**:439–456.
- Chang, C. Y., and E. Witschi. 1956. Genic control and hormonal reversal of sex differentiation in *Xenopus*. *Proc. Soc. Exp. Biol. Med.* **93**:140–144.
- de Sá, R. O., and D. M. Hillis. 1990. Phylogenetic relationships of the pipid frogs *Xenopus* and *Silurana*: an integration of ribosomal DNA and morphology. *Mol. Biol. Evol.* **7**:365–376.
- Doolittle, W. F., and J. R. Brown. 1994. Tempo, mode, the progenote, and the universal root. *Proc. Natl. Acad. Sci.* **91**:6721–6728.
- DuPasquier, L., I. Zucchetti, and D. Santis. 2004. Immunoglobulin superfamily receptors in protochordates: before RAG time. *Immunol. Rev.* **198**:233–248.
- Evans, B. J., D. B. Kelley, R. C. Tinsley, D. J. Melnick, and D. C. Cannatella. 2004. A mitochondrial DNA phylogeny of clawed frogs: phylogeography on sub-Saharan Africa and implications for polyploid evolution. *Mol. Phylogenet. Evol.* **33**:197–213.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- Ford, L. S., and D. C. Cannatella. 1993. The major clades of frogs. *Herpetol. Monogr.* **7**:94–117.
- Gibbs, M. J., J. S. Armstrong, and A. J. Gibbs. 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**:573–582.
- Goffeau, A. 2004. Seeing double. *Nature* **430**:25–26.
- Graf, J. D. 1996. Molecular approaches to the phylogeny of *Xenopus*. Pp. 379–389 in R. C. Tinsley and H. R. Kobel, eds. *The biology of Xenopus*. Clarendon Press, Oxford.
- Graf, J. D., and M. Fischberg. 1986. Albumin evolution of polyploid species of the genus *Xenopus*. *Biochem. Genet.* **24**:821–837.
- Greenhalgh, P., C. E. M. Olesen, and L. A. Steiner. 1993. Characterization and expression of recombination activating genes (RAG-1 and RAG-2) in *Xenopus laevis*. *J. Immunol.* **151**:3100–3110.
- Groth, J. G., and G. F. Barrowclough. 1999. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Mol. Phylogenet. Evol.* **12**:115–123.
- Haake, D. A., M. A. Suchard, M. M. Kelley, M. Dundoo, D. P. Alt, and R. L. Zuerner. 2004. Molecular evolution and mosaicism of leptospiral outer membrane protein involves horizontal DNA transfer. *J. Bacteriol.* **186**:2818–2819.
- Haldane, J. B. S. 1922. Sex-ratio and unisexual sterility in hybrid animals. *J. Genet.* **12**:101–109.
- Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**:160–174.
- Hoegg, S., M. Vences, H. Brinkmann, and A. Meyer. 2004. Phylogeny and comparative substitution rates of frogs inferred from sequences of three nuclear genes. *Mol. Biol. Evol.* **21**:1188–1200.
- Holland, P. W. H. 1997. Vertebrate evolution: something fishy about *Hox* genes. *Curr. Biol.* **7**:R570–R572.
- Holland, P. W. H., and J. Garcia-Fernández. 1996. *Hox* genes and chordate evolution. *Dev. Biol.* **173**:382–395.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Hurles, M. 2004. Gene duplication: the genomic trade in spare parts. *PLoS Biol.* **2**:900–904.

- Jackson, J. A., and R. C. Tinsley. 2003. Parasite infectivity to hybridising host species: a link between hybrid resistance and allopolyploid speciation? *Int. J. Parasitol.* **33**:137–144.
- Kass, R. E., and A. E. Raftery. 1995. Bayes factors. *J. Am. Stat. Assoc.* **90**:773–795.
- Kim, D. R., Y. Dai, C. L. Mundy, W. Yang, and M. A. Oettinger. 1999. Mutations of acidic residues in RAG1 define the active site of the *V(D)J* recombinase. *Genes Dev.* **13**:3070–3080.
- Kobel, H. R. 1996a. Reproductive capacity of experimental *Xenopus gilli* × *X. l. laevis* hybrids. Pp. 73–80 in R. C. Tinsley and H. R. Kobel, eds. *The biology of Xenopus*. Oxford University Press, Oxford.
- . 1996b. Allopolyploid speciation. Pp. 391–401 in R. C. Tinsley and H. R. Kobel, eds. *The biology of Xenopus*. Clarendon Press, Oxford.
- Kobel, H., B. Barandun, and C. H. Thiebaud. 1998. Mitochondrial rDNA phylogeny in *Xenopus*. *Herpetol. J.* **8**:13–17.
- Kobel, H. R., C. Loumont, and R. C. Tinsley. 1996. The extant species. Pp. 9–33 in R. C. Tinsley and H. R. Kobel, eds. *The biology of Xenopus*. Clarendon Press, Oxford.
- Kraysberg, Y., M. Schwartz, T. A. Brown, K. Ebralidse, W. S. Kunz, D. A. Clayton, J. Vissing, and K. Khrapko. 2004. Recombination of human mitochondrial DNA. *Science* **304**:981.
- Ladoukakis, E. D., and E. Zouros. 2001. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Mol. Biol. Evol.* **18**:1168–1175.
- Leviton, A., R. Gibbs, E. Heal, and C. Dawson. 1985. Standards in herpetology and ichthyology: Part 1. Standard symbolic codes for institutional resource collections in herpetology and ichthyology. *Copeia* **1985**:802–832.
- Litman, G. W. 1993. Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. *Mol. Biol. Evol.* **10**:60–72.
- Mable, B. K. 2004. 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. *Biol. J. Linn. Soc.* **82**:453–466.
- Maddison, D. R., and W. P. Maddison. 2000. *MacClade*. Sinauer Associates, Sunderland, Mass.
- Mann, M., M. S. Risley, R. A. Eckhardt, and H. E. Kasinsky. 1982. Characterization of spermatid/sperm basic chromosomal proteins in the genus *Xenopus* (Anura, Pipidae). *J. Exp. Zool.* **222**:173–186.
- Martin, D., and E. Rybicki. 2000. RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**:562–563.
- Masterson, J. 1994. Stomatal size in fossil plants: evidence for polyploid in majority of angiosperms. *Science* **264**:421–423.
- Maynard Smith, J. 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* **34**:126–129.
- McMahan, C. J., M. J. Sadofsky, and D. G. Schatz. 1997. Definition of a large region of RAG1 that is important for coimmunoprecipitation of RAG2. *J. Immunol.* **158**:2202–2210.
- Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves-Aldrey. 2004. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**:47–67.
- Orr, H. A. 1997. Haldane's rule. *Annu. Rev. Ecol. Syst.* **28**:195–218.
- Osborn, T. C., J. C. Pires, J. A. Birchler et al. (11 co-authors). 2003. Understanding mechanisms of novel gene expression in polyploids. *Trends Genet.* **19**:141–147.
- Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**:401–437.
- Ozkan, H., A. A. Levy, and M. Feldman. 2001. Allopolyploidy-induced rapid genomic evolution of the wheat (*Aegilops-Triticum*) group. *Plant Cell* **13**:1735–1747.
- Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734–740.
- Padidam, M., S. Sawyer, and C. M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**:218–225.
- Posada, D. 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* **19**:708–717.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- . 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci.* **98**:13757–13762.
- Presgraves, D. C., and H. A. Orr. 1998. Haldane's rule in taxa lacking a hemizygous X. *Science* **282**:952–954.
- Rambaut, A., and N. C. Grassly. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput. Appl. Biosci.* **13**:235–238.
- Rambaut, A., and M. Worobey. 2001. PIST: a program for the informative-sites test. University of Oxford, Oxford.
- Rast, J. P. 1997.  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$  T cell antigen receptor genes arose early in vertebrate phylogeny. *Immunity* **6**:1–11.
- Rice, W. R. 1989. Analyzing tables of statistical tests. *Evolution* **43**:223–225.
- Rodríguez, F., J. L. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitution. *J. Theor. Biol.* **142**:485–268.
- Roth, D. B., and N. L. Craig. 1998. VDJ recombination: a transposase goes to work. *Cell* **94**:411–414.
- Sadofsky, M. J., J. E. Hesse, J. F. McBlane, and M. Gellert. 1993. Expression and V(D)J recombination activity of mutated RAG-1 proteins. *Nucleic Acids Res.* **21**:5644–5650.
- Salminen, M. O., J. K. Carr, D. S. Burke, and F. E. McCutchan. 1995. Identification of breakpoints in intergenotypic recombinants of HIV-1 by bootscanning. *AIDS Res. Hum. Retroviruses* **11**:1423–1425.
- Sammur, B., A. Marcuz, and L. Du Pasquier. 2002. The fate of duplicated major histocompatibility complex class Ia genes in a dodecaploid amphibian, *Xenopus ruwenzoriensis*. *Eur. J. Immunol.* **32**:1593–1604.
- Schierup, M. H., and J. Hein. 2000. Recombination and the molecular clock. *Mol. Biol. Evol.* **17**:1578–1579.
- Soltis, D. E., and P. S. Soltis. 1999. Polyploidy: recurrent formation and genome evolution. *Trends Ecol. Evol.* **14**:348–352.
- Suchard, M. A., R. E. Weiss, K. S. Dorman, and J. S. Sinsheimer. 2002. Oh Brother, where art thou? A Bayes factor test for recombination with uncertain heritage. *Syst. Biol.* **51**:715–728.
- Swofford, D. L. 2002. *Phylogenetic analysis using parsimony* (\*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.
- Swofford, D., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. *Phylogenetic inference*. Pp. 407–514 in D. M. Hillis, C. Moritz, and B. K. Mable, eds. *Molecular systematics*. 2nd edition. Sinauer Associates, Sunderland, Mass.
- Tavaré, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. *Lect. Math. Life Sci.* **17**:277–290.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**:4876–4882.
- Thompson, J. D., and R. Lumaret. 1992. The evolutionary dynamics of polyploid plants: origins, establishment and persistence. *Trends Ecol. Evol.* **7**:302–307.
- Tinsley, R. C., C. Loumont, and H. R. Kobel. 1996. Geographical distribution and ecology. Pp. 35–59 in R. C. Tinsley and

- H. R. Kobel, eds. The biology of *Xenopus*. Clarendon Press, Oxford.
- Trueb, L. 1996. Historical constraints and morphological novelties in the evolution of the skeletal system of pipid frogs (Anura: Pipidae). Pp. 349–377 in R. C. Tinsley and H. R. Kobel, eds. The biology of *Xenopus*. Clarendon Press, Oxford.
- Trueb, L., and A. M. Báez. 1997. Redescription of the Paleogene *Shelania pascuali* from Patagonia and its bearing on the relationships of fossil and recent pipoid frogs. *Sci. Pap. Nat. Hist. Mus. Univ. Kan.* **4**:1–41.
- Tymowska, J. 1991. Polyploidy and cytogenetic variation in frogs of the genus *Xenopus*. Pp. 259–297 in D. S. Green and S. K. Sessions, eds. Amphibian cytogenetics and evolution. Academic Press, San Diego, Calif.
- Vandepoele, K., W. De Vos, J. S. Taylor, A. Meyer, and Y. Van de Peer. 2002. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc. Natl. Acad. Sci. USA* **101**:1638–1643.
- Wong, W. S. W., Z. Yang, N. Goldman, and R. Nielsen. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* **168**:1041–1051.
- Worobey, M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Mol. Biol. Evol.* **18**:1425–1434.
- Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**:555–556.
- Yang, Z., and R. Nielsen. 2002. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol. Biol. Evol.* **19**:908–917.

Dan Graur, Associate Editor

Accepted January 25, 2005