Slow DNA Loss in the Gigantic Genomes of Salamanders

Cheng Sun, José R. López Arriaza, and Rachel Lockridge Mueller*

Department of Biology, Colorado State University

*Corresponding author: E-mail: Rachel.Mueller@colostate.edu.

Abstract

Evolutionary changes in genome size result from the combined effects of mutation, natural selection, and genetic drift. Insertion and deletion mutations (indels) directly impact genome size by adding or removing sequences. Most species lose more DNA through small indels (i.e., ~1–30 bp) than they gain, which can result in genome reduction over time. Because this rate of DNA loss varies across species, small indel dynamics have been suggested to contribute to genome size evolution. Species with extremely large genomes provide interesting test cases for exploring the link between small indels and genome size; however, most large genomes remain relatively unexplored. Here, we examine rates of DNA loss in the tetrapods with the largest genomes—the salamanders. We used low-coverage genomic shotgun sequence data from four salamander species to examine patterns of insertion, deletion, and substitution in neutrally evolving non-long terminal repeat (LTR) retrotransposon sequences. For comparison, we estimated genome-wide DNA loss rates in non-LTR retrotransposon sequences from five other vertebrate genomes: *Anolis carolinensis*, *Danio rerio*, *Gallus gallus*, *Homo sapiens*, and *Xenopus tropicalis*. Our results show that salamanders have significantly lower rates of DNA loss than do other vertebrates. More specifically, salamanders experience lower numbers of deletions relative to insertions, and both deletions and insertions are skewed toward smaller sizes. On the basis of these patterns, we conclude that slow DNA loss contributes to genomic gigantism in salamanders. We also identify candidate molecular mechanisms underlying these differences and suggest that natural variation in indel dynamics provides a unique opportunity to study the basis of genome stability.

Key words: indel spectrum, insertion, deletion, mutation, genome size, transposable element.

Introduction

Evolutionary changes in genome size result from the combined effects of mutation, natural selection, and genetic drift (Gregory 2005; Lynch 2007). Insertion and deletion mutations (indels) directly impact genome size by adding or removing sequences. Such sequences can range in size from a single base pair to tens of thousands of bases (Hu et al. 2011; Ju et al. 2011). Establishing the link between 1) mutational processes that add or remove sequences and 2) evolutionary patterns of genome size variation across the Tree of Life remains a central goal in genome biology (Blass et al. 2012; Nam and Ellegren 2012; Petrov 2002).

Indels are frequently binned into the categories of "small" and "large" based on sequence length. Small indels span ~1–30 bp, whereas large indels can add or remove thousands of base pairs (Bhangale et al. 2006). The molecular mechanisms underlying large indels are fairly well understood; these include transposable element proliferation, transposable element-mediated ectopic recombination, slipped-strand mispairing, and nonhomologous end joining (Petrov et al. 2003;

Bennetzen et al. 2005; Ju et al. 2011). In contrast, the molecular mechanisms underlying the formation of small indels have been less clear, although small indels have long been linked to the general processes of DNA replication and recombination (Petrov et al. 1996; Kvikstad et al. 2007).

Most species lose more DNA through small indels than they gain, which can result in genome size reduction over time (Petrov 1997). However, this rate of DNA loss varies across species (Laurie et al. 2010). Consequently, small-indel-associated DNA loss rate has been suggested to be an important determinant of genome size (Petrov et al. 2000; Bensasson et al. 2001), although the extent of its role has been debated (Gregory 2003). Recently, global analyses of small indels in several fully sequenced genomes have provided new insights into the molecular processes producing small insertions and deletions (Kvikstad et al. 2007, 2009; Nam and Ellegren 2012). These advances allow the link between DNA loss rate and genome size evolution to be re-evaluated in more detail, including the exploration of mechanistic hypotheses that unite evolutionary changes in core DNA machinery (e.g., replication,

recombination, and repair) with evolutionary expansion and contraction of genomes. Lineages with extreme genome sizes can provide critical test cases for evaluating such hypotheses. However, large genomes remain relatively unstudied because they pose analytical challenges (Ambrozova et al. 2011).

Salamanders have the largest genome sizes among tetrapods and, with the exception of lungfishes, among vertebrates as a whole. Genome sizes across the 624 species of salamanders range from ~14 Gb to ~120 Gb (Amphibiaweb 2012; Gregory 2012). Salamander genomes contain much higher levels of long terminal repeat (LTR) retrotransposons than do other vertebrate genomes, suggesting that proliferation of this transposable element type is one molecular mechanism contributing to their genomic gigantism (Sun et al. 2012). However, the dynamics of DNA loss through small indels from salamander genomes remain unexplored.

Here, we test whether genomic gigantism in salamanders is associated with slow DNA loss through small indels, as predicted if DNA loss rate is an important determinant of genome size. We focused our analysis on non-LTR retrotransposons, which proliferate via an RNA transcript of a master gene sequence (Craig et al. 2002; Han 2010). Because of errors associated with transposition, most non-LTR element copies are "Dead On Arrival" (DOA) and neutrally evolving (Deininger et al. 1992; Luan et al. 1993); thus, they have been used to examine indel mutations in a variety of taxa (Petrov and Hartl 1998; Petrov et al., 2000). Our results show that salamanders have significantly lower rates of DNA loss than do other vertebrates. More specifically, salamanders experience lower numbers of deletions relative to numbers of insertions, and both deletions and insertions are skewed toward smaller sizes. Thus, slow DNA loss contributes to genomic gigantism in salamanders. We discuss candidate molecular mechanisms underlying this difference in DNA loss rate, highlighting the unique opportunity that natural variation in indel dynamics provides to study the basis of genome stability.

Materials and Methods

Salamander Shotgun Sequences

454 shotgun reads providing ~1% genomic coverage for four species of salamanders with genome sizes ranging from 15 to 44 Gb were obtained from our previous study (*Aneides flavipunctatus*, 44 Gb; *Batrachoseps nigriventris,* 25 Gb; *Desmognathus ochrophaeus,* 15 Gb; and *Eurycea tynerensis*, ~25 Gb) (Sun et al. 2012). The four species span the basal phylogenetic split within Plethodontidae, the largest family of salamanders (Vieites et al. 2011).

DNA Gain and Loss in Non-LTR Elements from Fully Sequenced Genomes

We first identified all the non-LTR transposable element copies from five vertebrate genomes: *Anolis carolinensis, Danio rerio,*

Gallus gallus, Homo sapiens, and *Xenopus tropicalis.* We downloaded the most recent RepeatMasker-generated pairwise alignment (.align) files from the RepeatMasker Genomic Datasets (http://www.repeatmasker.org/genomicDatasets/RMGenomicDatasets.html, last accessed 8 November 2012). These files contain pairwise alignments of a single non-LTR element sequence and its inferred master gene (i.e., ancestral) sequence; we identified the subset of each genome masked by non-LTR element master sequences.

Master sequences are estimated as the consensus sequence of all the genomic copies of a transposon family/subfamily. However, some of the non-LTR sequences identified as confamilial may have been generated by multiple active master genes that differed from one another in sequence. In such a case, a single consensus sequence would not accurately represent the ancestral state of all individual element copies; some of the differences between ancestor and descendant sequences would correspond to substitutions that occurred along the active master element lineage rather than in the neutrally evolving DOA copy. This would produce upwardly biased estimates of neutral substitution rates. To address this problem, we filtered our alignments to exclude non-LTR element sequences with substitutions that likely occurred along active lineages. We identified such sequences based on the fact that substitutions accumulating in active lineages show evidence of purifying selection acting on the proteins underlying transposition (i.e., substitutions are disproportionately at third codon positions) (Petrov et al. 1996). On the basis of each alignment, we partitioned substitutions by codon position and eliminated all non-LTR sequences with nonrandom distributions of substitutions across codon positions (χ^2 test; $P < 0.05$). Although there are other methods for identifying descendant copies from multiple master sequences (i.e., splitting subfamilies based on shared pairs of substitutions) (Price et al. 2004), the method we chose can be applied to both fully sequenced genomes and low-coverage shotgun data, enabling comparisons between salamanders and other vertebrates. In addition, as alignment accuracy decreases with sequence divergence (Lunter et al. 2008), we limited our analyses to element copies ≥80% identical to their respective master gene (ancestral) sequences, with a minimum overlap of 300 bp. The number of non-LTR elements found in each genome that meet these criteria are 64,449 (*A. carolinensis*), 24,030 (*D. rerio*), 43,723 (*G. gallus*), 255,079 (*H. sapiens*), and 16,881 (*X. tropicalis*). For each remaining non-LTR element copy, the number of insertions, deletions, and substitutions (after Jukes–Cantor correction) relative to the ancestral sequence were obtained based on the RepeatMasker-generated alignment. For each species, the sums of these values for every individual element copy were used to calculate the total amounts of DNA gained and lost through small indels.

DNA Gain and Loss in Non-LTR Elements from Salamander Shotgun Data

Consensus sequences for all repeat elements present in the 454 data for each salamander species were generated using RepeatScout (Price et al. 2005). Such consensus sequences were used as queries to BlastX against the amino acid sequences of transposable element-encoded proteins (http://www.repeatmasker.org/RepeatProteinMask.html#database, last accessed 8 November 2012), with an e-value threshold cutoff of 1e−10, to identify consensus sequences derived from non-LTR elements. The obtained consensus sequences were trimmed to the protein-coding regions. To increase accuracy, we limited our analysis to trimmed consensus sequences >330 bp in length that were estimated from at least five element copies, each of which we required to be >300 bp in length and ≥80% identical to the estimated consensus. The numbers of consensus sequences that meet these criteria in each genome are 41 (A. flavipunctatus), 57 (B. nigriventris), 19 (D. ochrophaeus), and 62 (E. tynerensis). We note that these consensus sequences are not full-length element sequences. Thus, the actual number of non-LTR families/subfamilies examined in each genome is likely lower than the number of consensus sequences because more than one consensus sequence may be derived from the same family/subfamily. All eligible consensus sequences from each species were used as a repeat library to mask the corresponding shotgun reads with RepeatMasker to generate pairwise alignment files. Finally, non-LTR sequences with nonrandom distributions of substitutions across codon positions were eliminated, as for fully sequenced genomes. The numbers of non-LTR elements analyzed in each salamander species are 1,109 (A. flavipunctatus), 1,495 (B. nigriventris), 179 (D. ochrophaeus), and 2,346 (E. tynerensis). Total amounts of DNA gained and lost through small indels, as well as total numbers of substitutions, were calculated as for fully sequenced genomes, with the exception that the minimum overlap cutoff of 100 bp was used.

Comparison of DNA Gain and Loss Rates among Vertebrates

Because the absolute dates of transposition events are unknown, calculation/comparison of absolute rates of small insertions and deletions are not possible. Thus, we compared relative rates of DNA loss across taxa (Petrov 2002). We calculated the ratio of net DNA loss (i.e., total bp deleted – total bp inserted) to number of substitutions for every taxon using all the non-LTR copies that we confirmed to be evolving neutrally.

Analysis of the Indel Spectrum

To gain insight into the molecular mechanisms underlying DNA gain and loss through small indels, we characterized individual insertion and deletion events in each taxon. Custom Perl scripts were used to calculate the number of deletion events, the size of each individual deletion, the number of insertion events, and the size of each individual insertion based on these sets of alignments.

Comparison of the Indel Spectra among Vertebrates

We binned individual insertion and deletion events by size (1–5 bp, 6–10 bp, 11–15 bp, and >15 bp) in each taxon and calculated the number of events in each bin relative to the taxon-specific total number of insertions or deletions. Next, we calculated the relative rate of DNA loss ([bp deleted – bp inserted]/substitution) for each bin in each taxon. Finally, we plotted these results to allow comparisons 1) across taxa for a given bin size and 2) across bin size for a given taxon.

Results

Rates of DNA Loss per Substitution

We calculated the relative rates of DNA loss ([bp deleted – bp inserted]/substitution) based on the sum totals of deletions, insertions, and substitutions that occurred in all the non-LTR copies identified in our nine focal taxa; these results are shown in figure 1 and included in table 1. The four salamander species have at least 2-fold lower rates of DNA loss than do the five remaining vertebrate species. Rates of DNA loss range from 0.019 bp/substitution (B. nigriventris) to 0.039 bp/substitution (E. tynerensis) across the four salamander species, which encompass a ~3-fold difference in genome size (Gregory 2012). In contrast, rates of DNA loss range from 0.080 bp/substitution (H. sapiens) to 0.315 bp/substitution (D. rerio) across the five remaining vertebrate species, which also encompass a ~3-fold difference in genome size (Gregory 2012). Table 1 lists the mean ratios of bp deleted/substitution and bp inserted/substitution for each of our nine focal taxa. Salamanders have 1.6- to 5.4-fold lower values of bp deleted/substitution than do the other vertebrates. In contrast, bp inserted/substitution values for salamanders are closer to and/or overlap with those of the other vertebrates.

For all nine of our focal taxa, DNA loss rate is highly variable among individual non-LTR element sequences (fig. 2; outliers excluded from boxplot). For the fully sequenced genomes, H. sapiens shows the most consistent rates across elements, whereas D. rerio shows the greatest variance. Variance across the four salamander species is more equivalent, although this may reflect the smaller shotgun data sets. Taken together, these results demonstrate highly variable rates of DNA loss among different non-LTR sequences within single genomes, consistent with the pattern reported previously for some primate and bird genomes based on slightly different methods (Kvikstad et al. 2007; Nam and Ellegren 2012).

Comparison of the Indel Spectra among Vertebrates

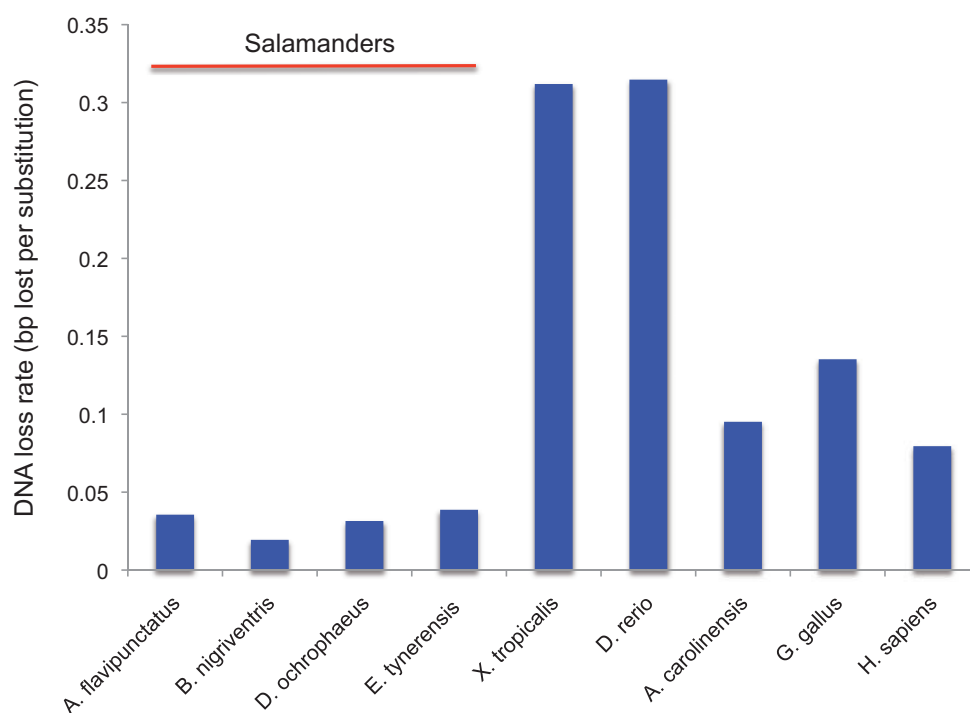The proportions of the total numbers of insertion and deletion events adding/removing ≤5 bp, 6–10 bp, 11–15 bp, and

Fig. 1.—Overall DNA loss rates from nine vertebrate genomes. Salamanders lose DNA more slowly than do other vertebrates.

Table 1

Deletion and Insertion Rates Relative to Substitutions

Species	Mean bp Deleted per Substitution	Mean bp Inserted per Substitution	DNA Loss Rate (bp/Substitution)
Anolis carolinensis	0.25	0.15	0.095
Danio rerio	0.49	0.18	0.315
Gallus gallus	0.19	0.06	0.135
Homo sapiens	0.21	0.13	0.080
Xenopus tropicalis	0.44	0.13	0.312
Aneides flavipunctatus	0.10	0.07	0.036
Batrachoseps nigriventris	0.10	0.08	0.019
Desmognathus ochrophaeus	0.12	0.09	0.031
Eurycea tynerensis	0.09	0.06	0.039

>15 bp are summarized for each taxon in figure 3. For all nine of our focal taxa, indels ≤ 15 bp in length make up >90% of the total indel events, and the majority of indel events add or remove <5 bp, consistent with previous studies from diverse taxa (Petrov et al. 2000; Laurie et al. 2010; Hu et al. 2011). This result demonstrates that 1) our shotgun data set for salamanders, despite yielding shorter non-LTR consensus sequences, is sufficient to characterize the small indel spectrum and 2) salamander genomes are qualitatively similar to those of other taxa in being highly skewed toward the smallest insertion and deletion sizes.

Additional information on indel size and number of indel events is summarized in table 2. For all nine of our focal taxa,

more base pairs are deleted than inserted. The mean sizes of both deletions and insertions are smaller in salamanders than in the other vertebrate taxa. Except for *D. ochrophaeus*, which has the smallest data set, there are more deletion events than insertion events in all taxa. The four salamander species have the lowest # deletions/# insertions (0.87–1.19), whereas *G. gallus* has the highest (2.42). DNA loss rates ([bp deleted − bp inserted]/substitution) for each indel size category across taxa are summarized in figure 4. In general, salamanders have lower rates of DNA loss across all indel sizes than do the other vertebrate taxa. There are two exceptions to this pattern: 1) rates of DNA loss in humans for indels > 6 bp in length are comparable to the low loss rates in salamanders
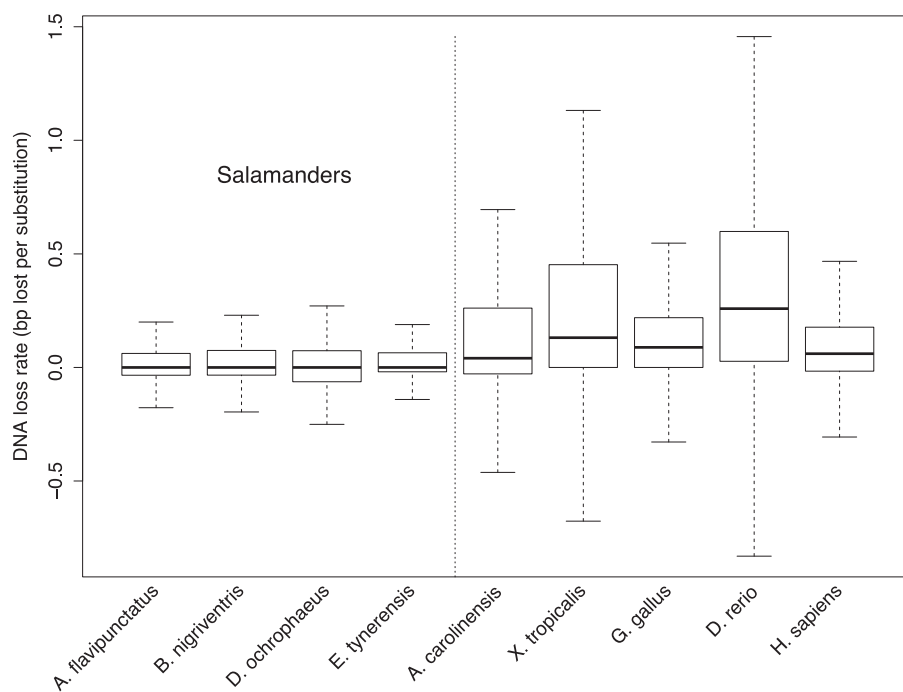
FIG. 2.—DNA loss rate estimates for individual non-LTR element copies. All species show heterogeneous rates of DNA loss across the genome. Outliers are excluded from plot to allow visualization of data. Element sequences are different lengths and include variable numbers of indels. Lower variance in salamanders may reflect smaller data sets for these taxa.
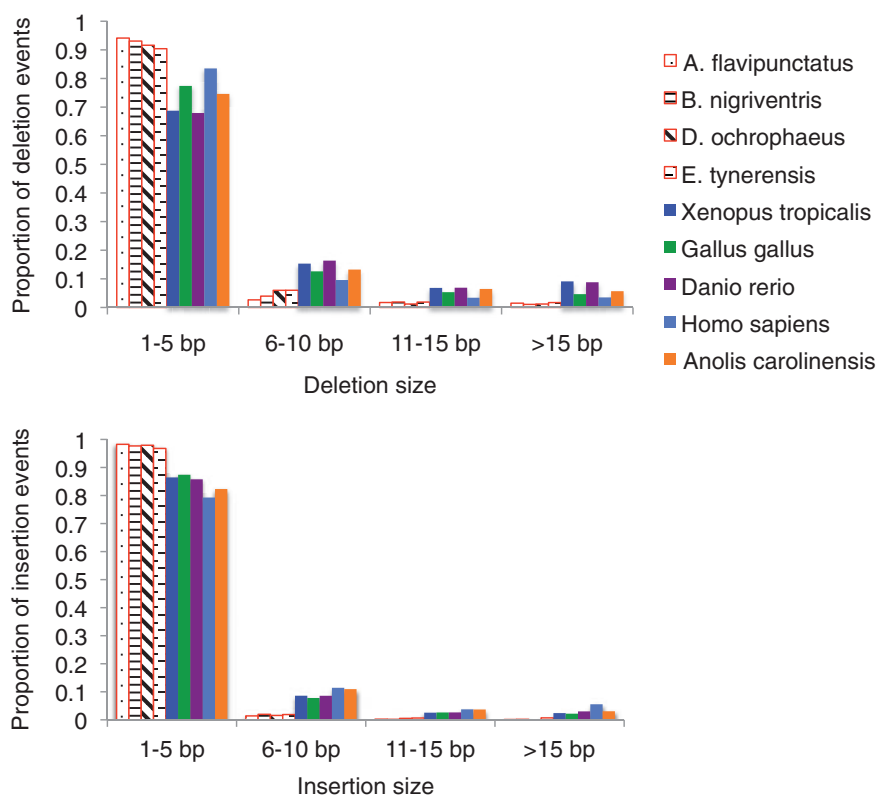

FIG. 3.—Proportion of total numbers of deletion events (top) and insertion events (bottom) that create indels of different size categories. Salamanders are outlined in red; other vertebrates are solid colors. Salamanders have higher relative numbers of the smallest (i.e., 1–5 bp) insertions and deletions.

Table 2

Insertion and Deletion Profiles

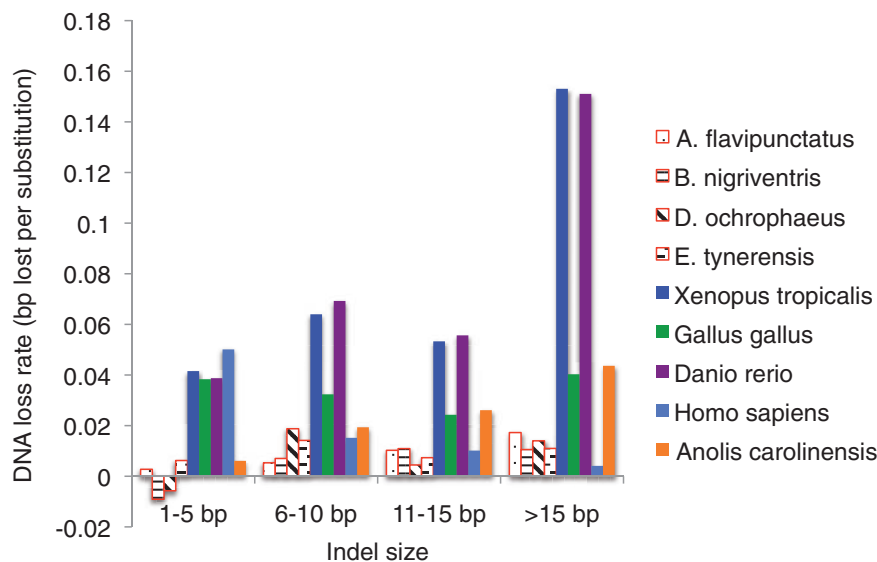| Species	Mean Deletion Size (bp)	Mean Insertion Size (bp)	Number of Deletions	Number of Insertions	No. of Deletions/ No. of Insertions	Total bp Deleted	Total bp Inserted
Anolis carolinensis	4.60	3.53	188,055	152,059	1.24	865,823	536,342
Danio rerio	5.92	3.43	104,554	64,982	1.61	619,220	223,111
Gallus gallus	4.24	3.00	158,889	65,533	2.42	674,429	196,377
Homo sapiens	3.89	4.71	1,503,779	774,286	1.94	5,851,399	3,652,366
Xenopus tropicalis	5.83	3.38	50,038	25,256	1.98	291,973	85,242
Aneides flavipunctatus	1.98	1.32	868	857	1.01	1,715	1,132
Batrachoseps nigriventris	2.04	1.79	1,213	1,118	1.08	2,476	1,998
Desmognathus ochrophaeus	2.25	1.45	167	192	0.87	377	278
Eurycea tynerensis	2.20	1.54	2,397	2,021	1.19	5,262	3,103

Fɪɢ. 4.—DNA loss rates calculated for indels of different size categories. Salamanders are outlined in red; other vertebrates are solid colors. Salamanders lose less DNA/substitution than *Xenopus tropicalis*, *Gallus gallus*, and *Danio rerio* across all indel size categories. *Homo sapiens* and *Anolis carolinensis* each show DNA loss rate overlap with salamanders in at least one indel size category. *Batrachoseps nigriventris* and *Desmognathus ochrophaeus* gain more DNA than they lose for indels 1–5 bp in length.

and 2) rates of DNA loss in *Ano. carolinensis* for indels ≤ 10 bp are comparable to the low loss rates in salamanders. Across taxa, rates of DNA loss vary among indel bin sizes in different ways; for example, *Ano. carolinensis* has increasing rates of DNA loss with increasing indel sizes, whereas *H. sapiens* shows the opposite pattern of decreasing rates of DNA loss with increasing indel size. Two salamander species (*B. nigriventris* and *D. ochrophaeus*) gain more DNA than they lose for indels 1–5 bp in length.

Discussion

Patterns of DNA Loss in Salamanders

Our results demonstrate that relative rates of DNA loss through small insertions and deletions are lower in salamanders than

in other vertebrates with more "typically" sized genomes. This pattern suggests that slow DNA loss rate contributes to genomic gigantism in salamanders. However, the 3-fold difference in genome size among our focal salamander species is not explained by variation in DNA loss rate; two species with the same genome size (*E. tynerensis* and *B. nigriventris*) have the most divergent DNA loss rates. This result suggests that among-salamander genome size differences reflect other factors (e.g., transposable element proliferation, selection on life history traits, and/or neutral drift processes) (Wake and Marks 1993; Lynch 2007; Sun et al. 2012). Among the five nonsalamander vertebrates, the ~3-fold difference in genome size tracks variation in DNA loss rate with the exception of *G. gallus*, which has the smallest genome and an intermediate rate of DNA loss. Taken together, our results suggest that DNA

loss rate contributes to genome size differences among vertebrates (e.g., salamanders vs. other lineages) but that it is not the sole determinant of genome size.

DNA loss rate reflects the overall number and size of individual insertion and deletion events. Thus, lower DNA loss rates can result from 1) smaller deletion sizes, 2) larger insertion sizes, 3) fewer deletion events, and/or 4) more insertion events. We show that salamanders lose at least 1.6-fold fewer base pairs of DNA/substitution than do other vertebrates; in contrast, base pairs inserted/substitution are more comparable between salamander and nonsalamander taxa (table 1). We show that both insertions and deletions are smaller, on average, in salamanders than in the other vertebrates we examined. We also show that salamanders have the lowest ratios of the numbers of deletion to insertion events (# deletions/# insertions = 0.87–1.19; table 2); in contrast, deletions outnumber insertions more dramatically in four of the five other vertebrates (# deletions/# insertions = 1.61–2.42). *Anolis carolinensis* is a notable exception; it has almost equivalent numbers of insertions and deletions (1.24). Thus, the difference in rates of DNA loss between salamanders and other vertebrates primarily reflects a lower number of base pairs deleted per substitution, which in turn reflects 1) fewer deletion events relative to substitutions in salamanders and 2) smaller deletion sizes (as well as insertion sizes) in salamanders relative to other vertebrates.

Candidate Molecular Mechanisms of Indel Formation in Salamanders

Small indels have long been attributed to uncharacterized errors in DNA replication and/or recombination (Petrov et al. 1996; Kvikstad et al. 2007). Recently, comparative genomic analyses have begun to leverage natural variation in indel dynamics, both across genomes and across lineages, to reveal the specific mechanisms of indel formation (Kvikstad et al. 2007, 2009; Hu et al. 2011; Nam and Ellegren 2012). Such analyses suggest that the molecular mechanisms producing insertions and deletions are partially distinct from one another, despite some overlap (Ball et al. 2005; Kvikstad et al. 2007, 2009; Messer and Arndt 2007; Tanay and Siggia 2008; Nam and Ellegren 2012). Salamanders experience fewer deletions relative to insertions than do other vertebrates; thus, deletion-specific processes are candidate molecular mechanisms to explain salamanders' slow rates of DNA loss. Additionally, both insertions and deletions are smaller in salamanders, suggesting a mechanism that reduces overall indel size. Here, we discuss such candidate mechanisms and their implications for genome size evolution.

Recombination and Meiotic Crossing-Over

Recombination rates, reflecting levels of meiotic crossing over, are negatively correlated with various measures of DNA loss (e.g., number of deletion events) within the chicken, zebra finch, and human genomes. In contrast, no relationship exists between recombination rate and insertion rate (Nam and Ellegren 2012). These results suggest a candidate mechanism for low rates of DNA loss in salamanders. Despite their enormous genome sizes, most salamanders (including those examined here) have relatively few diploid chromosomes (2N = 22–28) (Sessions 2007). Because the number of meiotic crossing-over events reflects the number of chromosome arms (Pardo-Manuel De Villena and Sapienza 2001; Lynch 2007), salamanders experience substantially fewer recombination events per base pair than do vertebrates with more typically sized genomes. This may contribute to the low number of deletion events in salamanders.

Recombination and DNA Repair

In addition to facilitating crossing-over during meiosis, recombination plays a central role in the repair of deleterious DNA double-strand breaks in germline and somatic cells. Two pathways exist to repair such breaks: homologous recombination and nonhomologous end joining (Haber 2000; Pardo et al. 2009). In homologous-recombination-mediated repair, broken DNA molecules are repaired by invasion of, and synthesis from, a sister chromatid or homologous chromosome (Srivastava and Raman 2007). In nonhomologous end joining, DNA resection and synthesis repair break site damage, and broken ends are ligated together (Lehman et al. 1994; Raghavan and Raman 2004; Forand et al. 2009; Ahmed et al. 2010).

Although both insertions and deletions (large and small) can result from double-strand break repair (Li et al. 2009; Lieber 2010; Ju et al. 2011; Mladenov et al. 2011), global analyses of motifs in the human genome suggest that some repair processes are more closely associated with small deletion than small insertion (Kvikstad et al. 2009). Specifically, deletion sites exhibit enrichment of translin target motifs and colocalization of translin targets with DNA polymerase pause/frameshift hotspots and topoisomerase cleavage sites (Kvikstad et al. 2009). Translin has diverse biological functions but likely plays a role in DNA damage recovery via repair/recombination pathways (Jaendling and Mcfarlane 2010). DNA polymerase pauses at both natural pause sites and damaged sites, and topoisomerase cleavage is associated with homologous recombination-mediated DNA repair (Kvikstad et al. 2009). Thus, double-strand break pathways may be particularly relevant for small deletion formation and are therefore candidate mechanisms for salamanders' slow DNA loss. Such mechanisms could act through either 1) decreased numbers of double-strand breaks during germline mitosis and meiosis, which would lower the number of deletion events and/or 2) higher-fidelity double-strand break repair, which would both lower the number of deletion events (i.e., if more breaks are repaired perfectly) and decrease overall indel sizes (i.e., if breaks are repaired with fewer changes to the break

site). We note that a possible connection to double-strand break pathways is particularly interesting, given salamanders' high levels of transposable element activity; transposition itself is associated with the introduction of double-strand breaks (Gasior et al. 2006; Suzuki et al. 2009).

DNA Replication

Finally, the small sizes of both insertions and deletions in salamanders suggest that the subset of molecular mechanisms that generate both types of mutations may produce smaller indels in salamanders than in other vertebrates. Motifs associated with DNA replication, as well as with repair of paused replication forks, are associated with both types of indels in the human genome (Kvikstad et al. 2009), suggesting that increased fidelity of replication may contribute to slow DNA loss in salamanders.

Indel Formation and Genome Size Evolution

Our previous work has shown that salamander genomes contain very high levels of transposable elements, particularly LTR retrotransposons (Sun et al. 2012). Here, we show that salamanders also lose DNA via small indels more slowly than other vertebrates. Taken together, these results suggest that salamanders' large genomes reflect both high rates of DNA "gain" (i.e., transposable element proliferation) and low rates of DNA loss. Different dynamic interactions between loss and gain, with different long-term effects on genome size evolution, are implied by the candidate molecular mechanisms for DNA loss presented here. For example, if salamanders' low levels of deletion events reflect fewer meiotic crossing-over events per bp, then genome expansion and DNA loss dynamics would mutually reinforce one another to bias genome sizes upward through the following steps: 1) transposable element proliferation increases genome size, whereas chromosome arm number remains constant. 2) Consequently, recombination rates per bp decrease, which lowers rates of DNA loss by decreasing the number of small deletions. 3) These lower DNA loss rates further facilitate genome expansion, as does the presence of more noncoding DNA; all insertions and deletions (regardless of size) become less likely to be deleterious as coding density decreases (Petrov 2002; Pettersson et al. 2009). This reinforcement toward genome expansion is analogous to the genomic contraction "vortex" described by Nam and Ellegren (2012), whereby recombination rate and deletion bias reinforce one another to drive genomic contraction. In contrast, no such predictable feedback interactions exist between transposable element-mediated genome expansion and DNA loss mediated through replication fidelity or double-strand break repair pathways.

Regardless of the underlying mechanism(s), the slow DNA loss rates exhibited by salamanders are part of the complex interactions among mutation, selection, and drift that have shaped the enormous genomes in this clade for the past >250 My. More detailed analyses are required to understand the relative contributions of these different evolutionary forces to genome size evolution. Comparative studies of natural variation in indel dynamics, combined with studies of the molecular mechanisms generating indels, are a powerful tool for connecting mutational processes that add or remove sequences with patterns of genome size variation across the Tree of Life. More generally, natural variation in indel dynamics across taxa suggests that comparative analyses can contribute substantially to studies of the core processes that maintain genomic stability over evolutionary time.

Literature Cited

Ahmed EA, De Boer P, Philippens MEP, Kal HB, De Rooij DG. 2010. Parp1-XRCC1 and the repair of DNA double strand breaks in mouse round spermatids. Mutat Res. 683:84–90.

Ambrozova K, et al. 2011. Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*. Ann Bot. 107:255.

Amphibiaweb [Internet]. 2012. AmphibiaWeb: information on amphibian biology and conservation. [cited 2012 May 16]. Available from: www.amphibiaweb.org.

Ball EV, et al. 2005. Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat. 26:205–213.

Bennetzen JL, Ma J, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 95:127.

Bensasson D, Petrov DA, Zhang D-X, Hartl DL, Hewitt GM. 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol Biol Evol. 18:246–253.

Bhangale TR, Stephens M, Nickerson DA. 2006. Automating resequencing-based detection of insertion-deletion polymorphisms. Nat Genet. 38:1457–1462.

Blass E, Bell M, Boissinot S. 2012. Accumulation and rapid decay of non-LTR retrotransposons in the genome of the threespine stickleback. Gen Biol Evol. 4:687–702.

Craig NL, Craigie R, Gellert M, Lambowitz AM. 2002. Mobile DNA II. Herndon (VA): American Society for Microbiology Press.

Deininger PL, Batzer MA, Hutchison CA III, Edgell MH. 1992. Master genes in mammalian repetitive DNA amplification. Trends Genet. 8:307–311.

Forand A, et al. 2009. Similarities and differences in the in vivo response of mouse neonatal gonocytes and spermatogonia to genotoxic stress. Biol Reprod. 80:860–873.

Gasior SL, Wakeman TP, Xu B, Deininger PL. 2006. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol. 357:1383–1393.

Gregory TR. 2003. Is small indel bias a determinant of genome size? Trends Genet. 19:485–488.

Gregory TR. 2005. The evolution of the genome. San Diego (CA): Elsevier.

Gregory TR. 2012. Animal genome size database. [Internet]. [cited 2012 may 16]. Available from: http://www.genomesize.com.

Haber JE. 2000. Partners and pathways: repairing a double-strand break. Trends Genet. 16:259–264.

Han JS. 2010. Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions. Mobile DNA 1:15.

Hu TT, et al. 2011. The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change. Nat Genet. 43:476–481.

Jaendling A, Mcfarlane R. 2010. Biological roles of translin and translin-associated factor-X: RNA metabolism comes to the fore. Biochem J. 429:225–234.

Ju YS, et al. 2011. Extensive genomic and transcriptional diversity identified through massively parallel DNA and RNA sequencing of eighteen Korean individuals. Nat Genet. 43:745–752.

Kvikstad EM, Chiaromonte F, Makova KD. 2009. Ride the wavelet: a multiscale analysis of genomic contexts flanking small insertions and deletions. Genome Res. 19:1153–1164.

Kvikstad EM, Tyekucheva S, Chiaromonte F, Makova KD. 2007. A macaque's-eye view of human insertions and deletions: differences in mechanisms. PLoS Comp Biol. 3:e176.

Laurie S, Toll-Riera M, Rado-Trilla N, Alba MM. 2010. Sequence shortening in the rodent ancestor. Genome Res. 22:478–485.

Lehman CW, Trautman JK, Carroll D. 1994. Illegitimate recombination in *Xenopus*: characterization of end-joined junctions. Nucleic Acids Res. 22:434–442.

Li W, Tucker AE, Sung W, Thomas W Kelley, Lynch M. 2009. Extensive, recent intron gains in *Daphnia* populations. Science 326:1260–1262.

Lieber MR. 2010. The mechanism of double-strand DNA break repair by the nonhomologous DNA end joining pathway. Annu Rev Biochem. 79:181.

Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605.

Lunter G, et al. 2008. Uncertainty in homology inferences: assessing and improving genomic sequence alignment. Genome Res. 18:298–309.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates, Inc.

Messer PW, Arndt PF. 2007. The majority of recent short DNA insertions in the human genome are tandem duplications. Mol Biol Evol. 24: 1190–1197.

Mladenov E, Iliakis G. 2011. Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. Mutat Res. 711:61–72.

Nam K, Ellegren H. 2012. Recombination drives vertebrate genome contraction. PLoS Genet. 8:e1002680.

Pardo B, Gomez-Gonzalez B, Aguilera A. 2009. DNA double-strand break repair: how to fix a broken relationship. Cell Mol Life Sci. 66: 1039–1056.

Pardo-Manuel De Villena F, Sapienza C. 2001. Recombination is proportional to the number of chromosome arms in mammals. Mamm Genome. 12:318–322.

Petrov DA. 1997. Slow but steady: reduction of genome size through biased mutation. Plant Cell 9:1900–1901.

Petrov DA. 2002. Mutational equilibrium model of genome size evolution. Theor Popul Biol. 61:533–546.

Petrov DA, Aminetzach YT, Davis J, Bensasson D, Hirsch AE. 2003. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. Mol Biol Evol. 20:880–892.

Petrov DA, Hartl DL. 1998. High rate of DNA loss in the *Drosophila melanogaster* and *Drosophila virilis* species groups. Mol Biol Evol. 15: 293–302.

Petrov DA, Lozovskaya ER, Hartl DL. 1996. High intrinsic rate of DNA loss in *Drosophila*. Nature 384:356–349.

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL. 2000. Evidence for DNA loss as a determinant of genome size. Science 287: 1060–1062.

Pettersson ME, Kurland CG, Berg OG. 2009. Deletion rate evolution and its effect on genome size and coding density. Mol Biol Evol. 26: 1421–1430.

Price AL, Eskin E, Pevzner PA. 2004. Whole-genome analysis of Alu repeat elements reveals complex evolutionary history. Genome Res. 14: 2245–2252.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. Bioinformatics 21(Suppl 1): i351–i358.

Raghavan SC, Raman MJ. 2004. Nonhomologous end joining of complementary and noncomplementary DNA termini in mouse testicular extracts. DNA Repair 3:1297–1310.

Sessions SK. 2007. Evolutionary cytogenetics in salamanders. Chromosome Res. 16:183–201.

Srivastava N, Raman MJ. 2007. Homologous recombination-mediated double-strand break repair in mouse testicular extracts and comparison with different germ cell stages. Cell Biochem Funct. 25:75–86.

Sun C, et al. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Gen Biol Evol. 4:168–183.

Suzuki J, et al. 2009. Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. PLoS Genet. 5:e1000461.

Tanay A, Siggia ED. 2008. Sequence context affects the rate of short insertions and deletions in flies and primates. Gen Biol. 9:R37.

Vieites DR, Rom SN, Wake MH, Wake DB. 2011. A multigenic perspective on phylogenetic relationships in the largest family of salamanders, the Plethodontidae. Mol Phylogenet Evol. 59:623–635.

Wake DB, Marks SB. 1993. Development and evolution of plethodontid salamanders: a review of prior studies and a prospectus for future research. Herpetologica 49:194–203.

Associate editor: Bill Martin