bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genome size variation and species diversity in salamander families

Bianca Sclavi [1*] and John Herrick [2]**

Author affiliations:

1. LBPA, UMR 8113, ENS de Cachan, Cachan, France
2. Department of Physics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia VSA 1S6, Canada

* Corresponding author

** Corresponding author at 3, rue des Jeûneurs, 75002 Paris, France.

E-mail addresses: jhenryherrick@yahoo.fr (J. Herrick), sclavi@lbpa.ens-cachan.fr (B. Sclavi).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

***Abstract**:* Salamanders (urodela) have among the largest vertebrate genomes, ranging in size from 10 to 120 pg. The urodela are divided into ten extant families each with a characteristic range in genome size. Although changes in genome size often occur randomly and in the absence of selection pressure, non-random patterns of genome size variation are evident among specific vertebrate lineages. Here we report that genome size in salamander families varies inversely with species diversity and other ecological factors: clades that began radiating earlier (older crown age) tend to have smaller genomes, higher levels of diversity and larger geographical ranges. These observations support the hypothesis that urodela families with larger genomes either have a lower propensity to diversify or are more vulnerable to extinction than families with smaller genomes.

Keywords: C-value, C-value, species diversity, body size, urodela, geographical area

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

Genome size in vertebrates varies more than two hundred fold from 0,4 picograms (pg) in pufferfish to over 120 picograms (pg) in lungfish (*1*). Most of the variation in vertebrate genome size corresponds to differences in non-coding DNA such as transposable elements, microsatellites and other types of repetitive and intergenic DNA (*2*). The DNA accounting for differences in genome size between related species has previously been considered devoid of any universal function such as gene regulation, structural maintenance and protection against mutagens (*3*). Genome size, however, is known to have a direct impact on important physiological parameters such as cell size and cell cycle duration (*4, 5*). C-values, which measure genome size, are exceptionally high and vary more widely in salamanders compared to most other vertebrates (*6*), making urodelas an attractive model for studying genome size variation and its biological effects between and within taxonomically related vertebrate lineages.

In addition to its impact on cell physiology, genome size frequently correlates with a number of ecological and physiological parameters including climate, metabolic rate, extinction risk and species diversity (*7-9*). A comparative study across vertebrates, for example, has found a negative correlation between species diversity and genome size but only above the class level on the taxonomic hierarchy (*10, 11*). The strongest association between genome size on species diversity was observed at C-values greater than 5 pg. Consistent with reduced species diversity in vertebrates with large genomes, at higher taxonomic levels extinction risk was found to increase with genome size (*12*). Consequently, species number might correlate negatively with the proportion of repetitive DNA (*10*), suggesting that either evolvability (propensity to speciate) or extinction risk varies according to the amount of non-coding DNA present in the genome.

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

The large genomes of salamanders containing disproportionate amounts of non-coding DNA therefore represent a potentially important factor that is influencing species diversity across urodelas. Urodelas are comprised of ten distinct families (*13*), which is relatively low compared to anura (fifty five different families). While some urodela families, such as the Plethodontidae and Salamandridae, are exceptionally rich in species number, other families are extremely species poor. Five of the salamander families, for example, contain only three or four known species whereas the Plethodontidae contains over five hundred. Why species diversity varies so widely among the ten families remains unclear, though many physiological and environmental factors are known to contribute to differences in urodela species diversity (*14, 15*).

The large urodela genomes have been attributed to slow losses of DNA in salamanders compared to other vertebrates (*16*). The mechanisms underlying the slower loss rates remain to be elucidated, but DNA damage and repair systems likely play an important role in the evolution of large genomes and mutation rate variation (*17, 18*). Variations in the proportion of non-coding DNA in genomes (eg. indels), for example, correspond to *de facto* mutations that have occurred during the course of genome evolution (*19, 20*). Previously, we examined rates of genetic variation in salamanders and found preliminary evidence that mutation rates tend to decrease as genome size increases (*21*). Consequently, the slow loss of DNA in salamanders might reflect relatively low mutation rates in urodelas. Similar studies have shown that low mutation rates and low species diversity are also be associated with large genomes in plants (*22, 23*), while another study in plants suggests variations in species diversity depend less on absolute values of genome size than on rates of genome size evolution (*24*).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Consistent with low mutation rates in organisms with large genomes, earlier studies have revealed exceptionally low levels of genetic diversity in salamanders (*14, 25*), yet the sources of the low genetic variation remain unclear (*26, 27*). Genetic variation is a determining factor in both evolvability and extinction risk (*12, 28-30*), and is therefore potentially associated with species diversity. Hence, the low levels of genetic variation in gene expressing regions in urodela genomes (*25, 31*), suggests that genome size might also influence species diversity in salamanders. A potential relationship between genome size and species diversity, however, remains to be established across salamander families. In the following we investigate the relationship between genome size and species diversity at the family level of the urodela.

Materials and methods

Data on snout to vent length (SVL) for the species for which genome size is known were obtained from AmphibiaWeb (http://www.amphibiaweb.org) using the Rafaelli account. SVL is commonly used as a proxy for body size, and is measured as the distance from the tip of the snout to the anterior posterior to the cloaca. In some cases, only either male or female body size was found. When both male and female body size were reported, the average of the two was used. Average SVL for each family was calculated as the average value of the species in that family.

Genome sizes were obtained from the Animal Genome Size Database (*1*). C-value refers to haploid nuclear DNA content (1C). Reported polyploids, when indicated in the Animal Genome Size Database, were removed from the analyses. The value for each family represents the weighted average of all the genera available in the database for that family. The genome size (C-value) for each genus corresponds to the weighted average C-value of the species available in the database that belong to the genus. The C-value for each species is determined as the average of the available

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

values for that species when more than one C-value is recorded. The distributions in genome size for each salamander family and the genera of Plethodontidae have been published previously (*21*).

The data on crown age, stem age, species diversity and geographic area were obtained from Pyron and Wiens (*15*). Both the maximum likelihood (ML) and the time-calibrated trees used here were obtained from those of Pyron and Wiens (*32, 33*). They obtained the time-calibrated tree by determining divergence times from a set of fossil constraints using treePL developed by S.A. Smith (Smith S.A., O'Meara B. 2012 treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* 28, 2689-2690), and applied to the ML phylogeny determined previously for 2871 species using data from 3 mitochondrial and 9 nuclear genes (*32*). They determined species diversity from the assignment of all known amphibian species to genera and families as classified in their phylogeny (*32*).

The Pyrons and Wiens ML and time-calibrated trees were used to create a tree at the family level by first automatically assigning the species to families using the taxize package in R (Scott Chamberlain, Eduard Szocs, Carl Boettiger, Karthik Ram, Ignasi Bartomeus, and John Baumgartner (2014) taxize: Taxonomic information from around the web. R package version 0.3.0. https://github.com/ropensci/taxize). These were manually verified against the taxonomy of the Pyrons and Wiens tree. The HighLevelTree function in the EvobiR package in R by Heath Blackmon (Heath Blackmon (2013) was used to obtain the family level trees. (evobiR: evolutionary biology in R. R package version 1.0. http://CRAN.R-project.org/package=evobiR).

Regression analysis: Independent contrasts were carried out in R using the ape library based on the branch lengths of the ML tree shown in Figure 1. The regression of the independent contrasts was

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

forced through the origin. The analysis was carried out in R with the caper package using the time-calibrated salamander family tree obtained from the amphibia tree of Pyron and Wiens (*15*). The residuals of the regression for the independent contrasts (log C-value vs. log body size and log C-value vs. area, respectively) were controlled for normality using the Shapiro-Wilk test, confirming a normal distribution.

Results

Species diversity is negatively associated with C-value.

The origin of urodela dates from 155 to 170 Mya (*34, 35*). Urodela inhabit a wide variety of ecological niches and exhibit a large diversity of life history traits, including small and large body sizes, paedomorphy, neoteny, metamorphosis and direct development (*13*). In an earlier study of amphibia, Pyron and Wiens revealed a number of ecological correlates between species diversity and variables such as geographical latitude, environmental energy and climatic niche rate (*15*). Species diversity in frogs, salamanders and caecilians also varies according to abiotic factors such as humidity and temperature and biotic factors such as productivity and rates of diversification (extinction and speciation). Figure 1 shows the phylogenetic tree derived from Pyron and Wiens that was used here to investigate the relationship between salamander genome size and species

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

diversity (*32*).

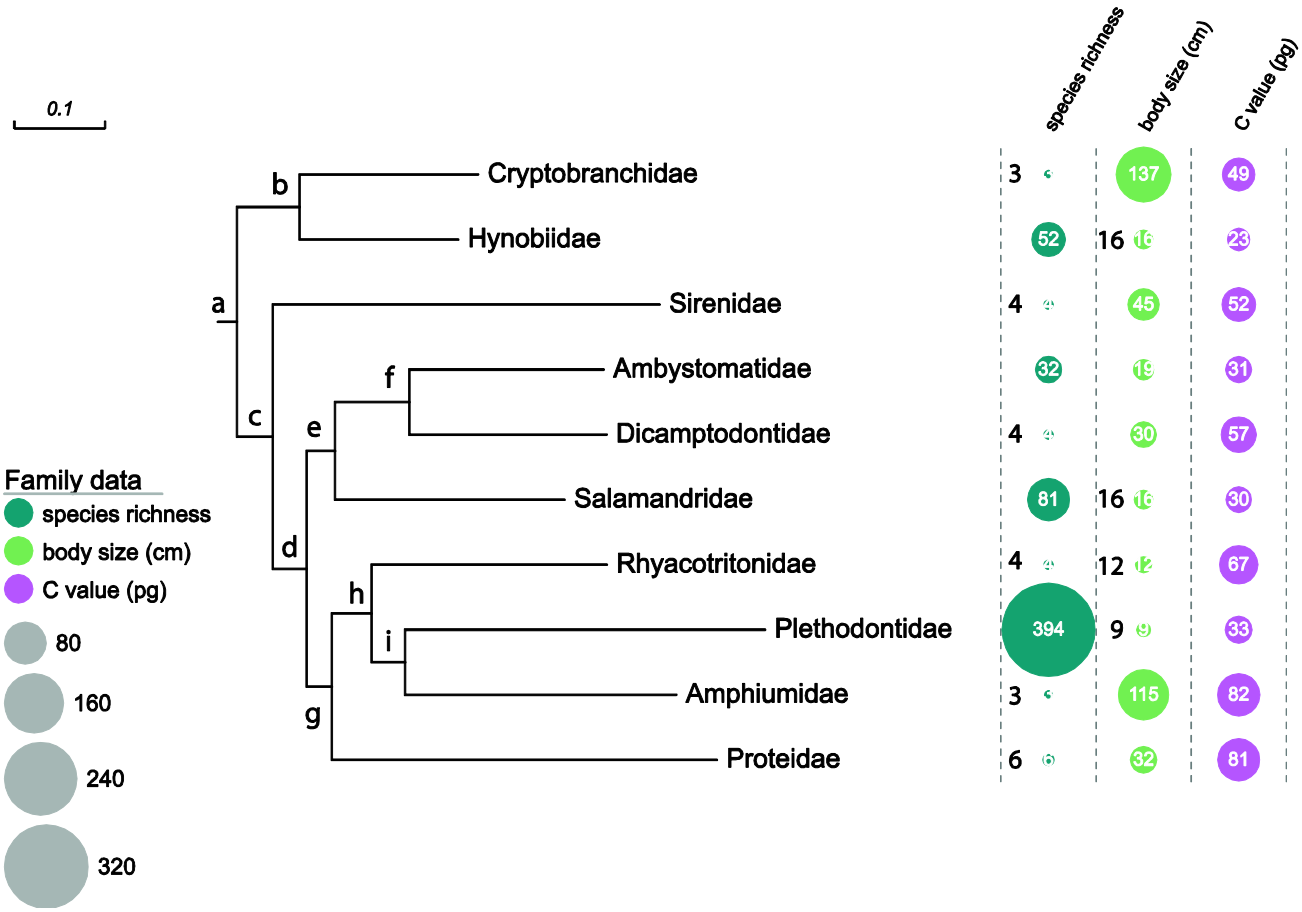

Figure 1. Maximum likelihood phylogenetic tree of caudata derived from Pyron and Wiens 2011. Branch lengths shown are the distances measured as substitutions per site up to the crown origin of each family. Species diversity, body size and average C-value are shown next to the family name in circles proportional to size. The branch length is indicated by the scale bar. The letters denote the node identities used in the analysis of phylogenetically independent contrasts (Figures 3B and 5C). Six families form three sister-pair taxa: nodes i, f and b, which correspond to the most recent common ancestor (MRCA) of the subtending families. The time calibrated tree, stem and crown ages are provided in supplementary Figure 3S. Figure created using EvolView (*52*).

Figure 1 suggests that C-value is negatively associated with species diversity. To further characterize the apparent trend between C-value and species diversity revealed in the tree, we plotted species diversity values from the Pyron and Wiens dataset against C-values obtained from the Animal Genome Size Database (*1*). The scatter plot presented in Figure 2A appears to divide salamanders into two broad classes in a genome size dependent manner: families with C-values either greater or

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

lower than 40-50 pg, suggesting that genome size and species diversity tend to vary inversely with respect to each other.

The urodela, the crown-group of the salamanders, is divided into three suborders, which span the basal phylogenetic split in Caudata: the Chriptobranchoidea, the Sirenidae and the Salamandroidea (Figure 1). The boxplot in Figure 2B confirms that species diversity in salamander families varies significantly with average genome size independently of the suborders to which they belong (p-value = 0.009; assuming equal variance: p-value = 0.00013). Families with smaller genomes on average, which are known to have correspondingly higher genome size diversity (*6*), are therefore associated with larger species diversity in urodela (see discussion).

Figure 2B indicates that variance in species richness is lower in the group with larger genomes. At the same time,these same families(C-value > 50 pg) exhibit significantly less genome size variation than the other salamander families . To assess the relationship between genome size variation and species diversity, we examined the relationship between species diversity and the coefficient of variation (CV) in genome size across the ten families. Figure 2C shows that as the CV of genome size increases, there is a consistent increase in species diversity (Table 1). Extant species diversity in salamander families therefore reflects extant genome size diversity within each family. In contrast, no relationship is found between variance in body size and species diversity (not shown).

Closely related taxa are expected to have similar phenotypes. We therefore assessed genetic distance between species in the salamander families within and between the two distinct genome-size classes in Figure 2B. Figure 2D shows the genetic distance between species within and between the two classes. If the species in the respective genome size classes in Figure 2B were more closely

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

related to each other, then the genetic distances between the species in each box plot would be smaller than the distances between species across the two classes. Figure 2D shows no significant difference in the genetic distances between salamander species in each family, whether they are measured within or between the groups. Species diversity therefore varies with genome size independently of the genetic distance between salamander species.

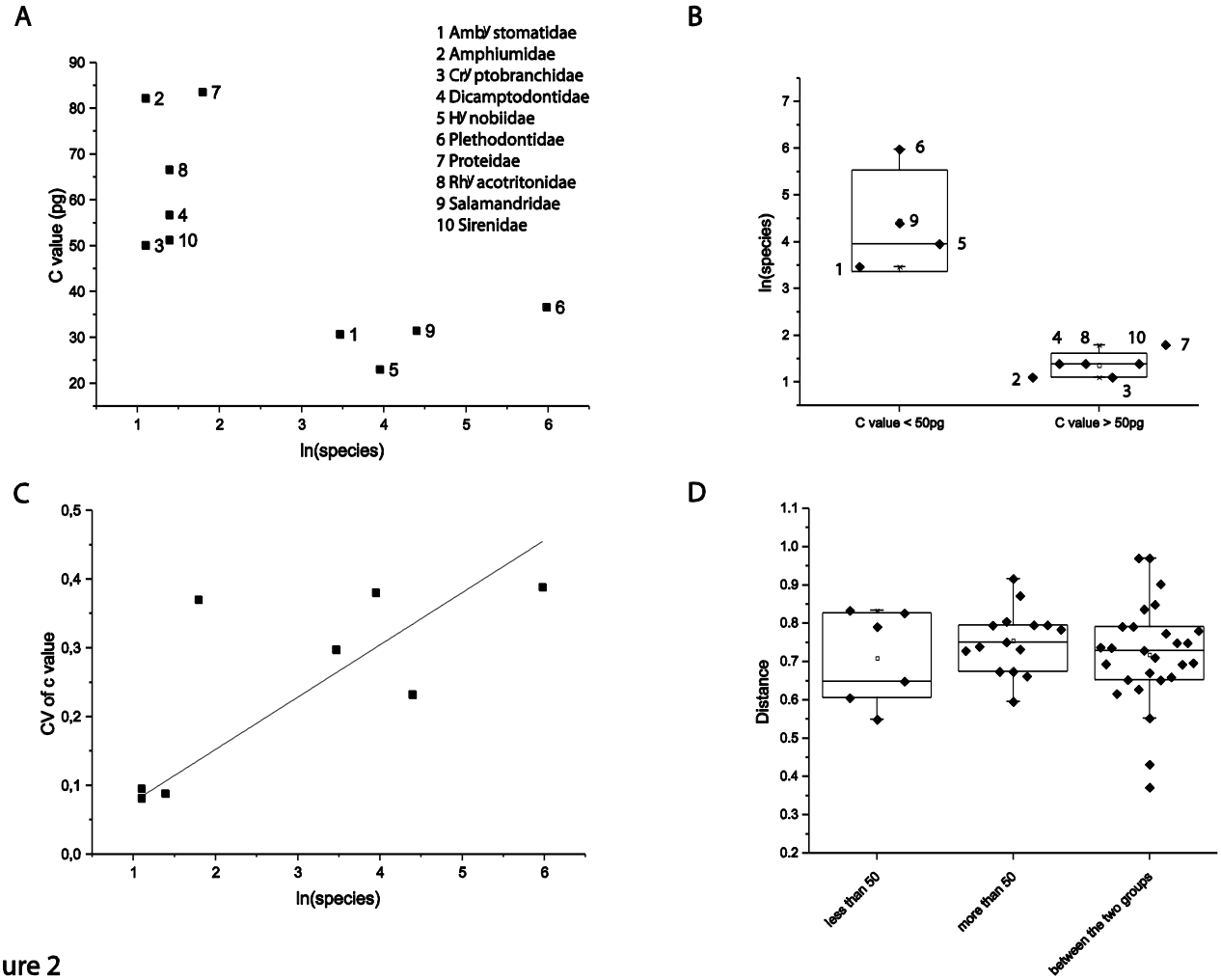

Figure 2

Figure 2: Species diversity is negatively associated with C-value.
A) Scatter plot analysis of species diversity versus C-value. Numbers refer to respective family names. B) Boxplot of species diversity for families with genome sizes <50 pg (ln(species) median = 4.17; IQR = 1.5) or >50 pg (ln(species) median = 1.4; IQR = 0.28). Open squares: mean; horizontal lines: median; whiskers: minimum and maximum values. Size of the box is +/- 1 SD from the mean. P-value = 0.009; assuming equal variance p-value = 0.00013. C) Coefficient of variation versus species richness. D) Genetic distance assessed within (left two boxes) and among (right box) salamander species used in the regression analysis.

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Changes in C-value are associated with changes in body size

Figure 1 also suggests a negative relationship between body size and species diversity. Body size in salamanders is assessed according to snout to vent lengths (SVL). Regression analysis in Figure 3A shows the association between C-value and body size, indicating that average C-value and SVL measurements are not significantly correlated (adjusted R^2 = 0.19, p-value=0.1; Table1). An assessment of independent contrasts of C-value and body size, however, did reveal a marginally significant relationship between these two variables (adjusted R^2 = 0.35, p-value = 0.04; Figure 3B; Table 1): as genome sizes increase (or decrease), body size tends to increase (or decrease) as well. If standardized contrasts represent evolutionary rates (*36, 37*), the evolution of genome size and body size tends to be positively associated, even though body size and genome size are not themselves significantly related. Together these results suggest that C-value is more significantly associated with species diversity than is body size in salamander families.

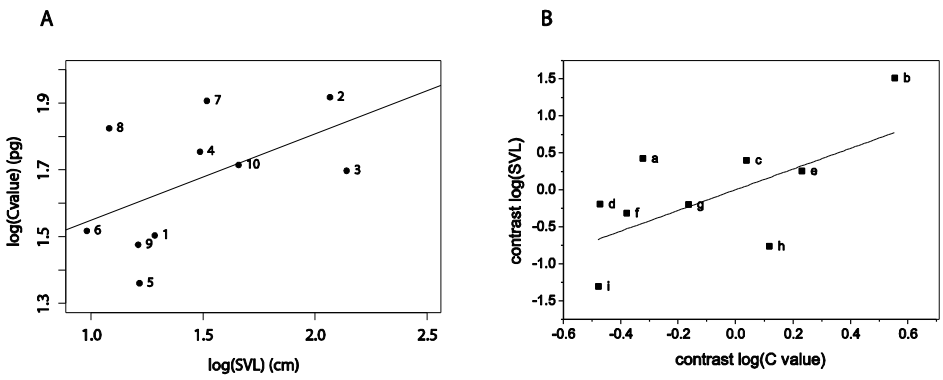

Figure 3: Changes in C-value are associated with changes in body size.
A) Standard regression analysis of log body size versus C-value (adjusted R^2 = 0.44 p=0.02). Numbers correspond to the families indicated in Figure 2. B) Assessment of independent contrasts of C-value and body size (adjusted R^2 = 0.47, p-value = 0.017).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

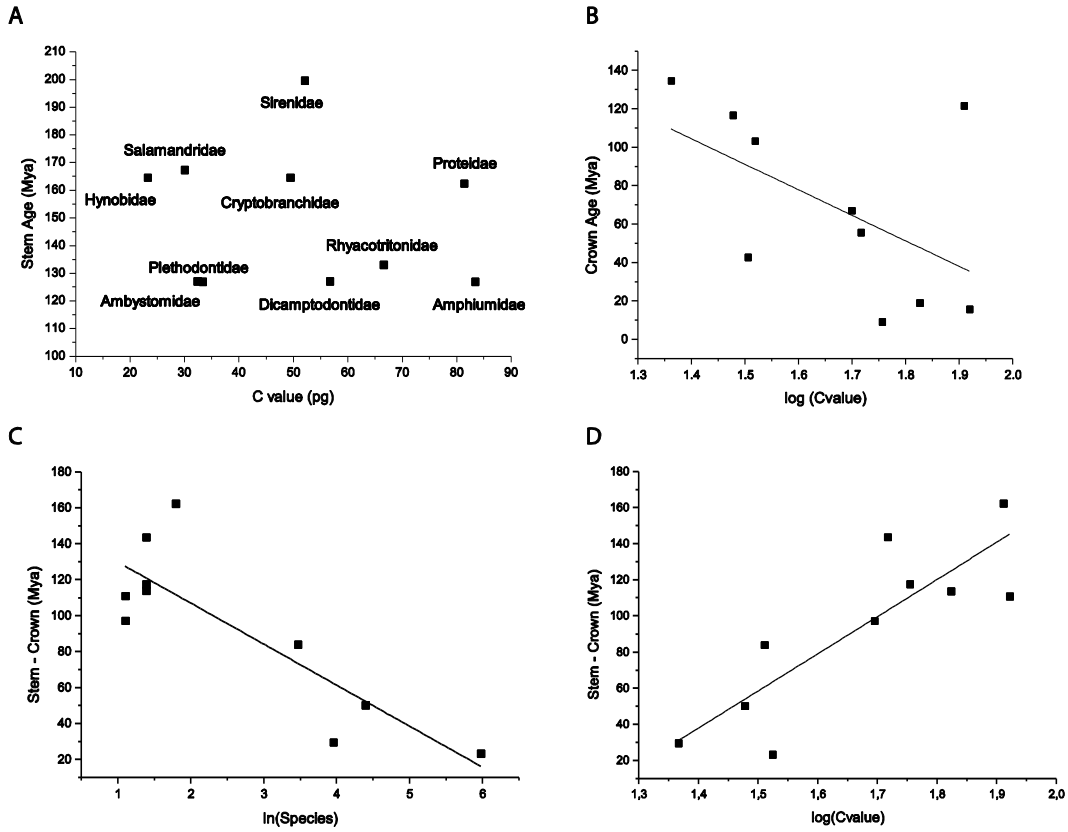

Figure 4

Figure 4: Ecological correlates with genome size (time).
A) Relationship between phylogenetic stem age and genome size. B) Standard regression analysis of crown age versus C-value (adjusted R^2 = 0.20, p-value = 0.1). C) Relationship between stem-to-crown age and species diversity (adjusted R^2 = 0.68, p-value = 0.002). D) Relationship between stem-to-crown age and C-value (adjusted R^2 = 0.70, p-value = 0.002).

Ecological correlates with genome size: time

The genomes of urodela have experienced massive amplification during evolution (*38, 39*). An earlier study reported that genome size in salamanders has increased at a rate of about 0.6 pg Mya, indicating that salamander families that have diverged earlier in time have correspondingly larger genomes (*40*). We therefore examined the relationship between phylogenetic stem age and genome size, but found no significant relationship (Figure 4A).

We next undertook an examination of the relationship between crown age, which corresponds to the origin of a clade (*41*), and average genome size in the different urodela families. We found that

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

C-value and crown age are not significantly correlated across the Caudata (adjusted R^2 = 0.20, p-value = 0.1) (Figure 4B, Table 1). We determined, however, that the family Proteidae represents an outlier in the regression analysis (studentized residual: 2.35; studentized deleted residual: 3.94). In the absence of Proteidae, the relationship becomes significant (R^2 = 0.70, p-value = 0.003; Table 1), as expected if that family constitutes an outlier in the regression analysis. Crown age rather than stem age therefore tends to be associated with C-value when the Proteidae are excluded from the regression analysis.

Relationships of stem-to-crown age with species diversity and C-value

Older clades are expected to have more species under the assumption of a molecular clock (*42*). Crown-to-stem age indicates the time elapsed before a clade began radiating relative to stem age. The scatter plot in Figure 4C reveals a clear negative trend between stem-to-crown age and species diversity (see Table 1). The most species rich clades, the Hynobidae, Ambystomatidae, Salamandridae and Plethodontidae, for example, all began radiating earlier with respect to stem age than did the less speciose clades (Figure 4C). Clades that formed and diversified earlier with respect to stem age likewise have consistently smaller genomes on average than clades corresponding to longer stem-to-crown ages (adjusted R^2 = 0.70, p-value = 0.002; Table 1; Figure 1; Figure 4D).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

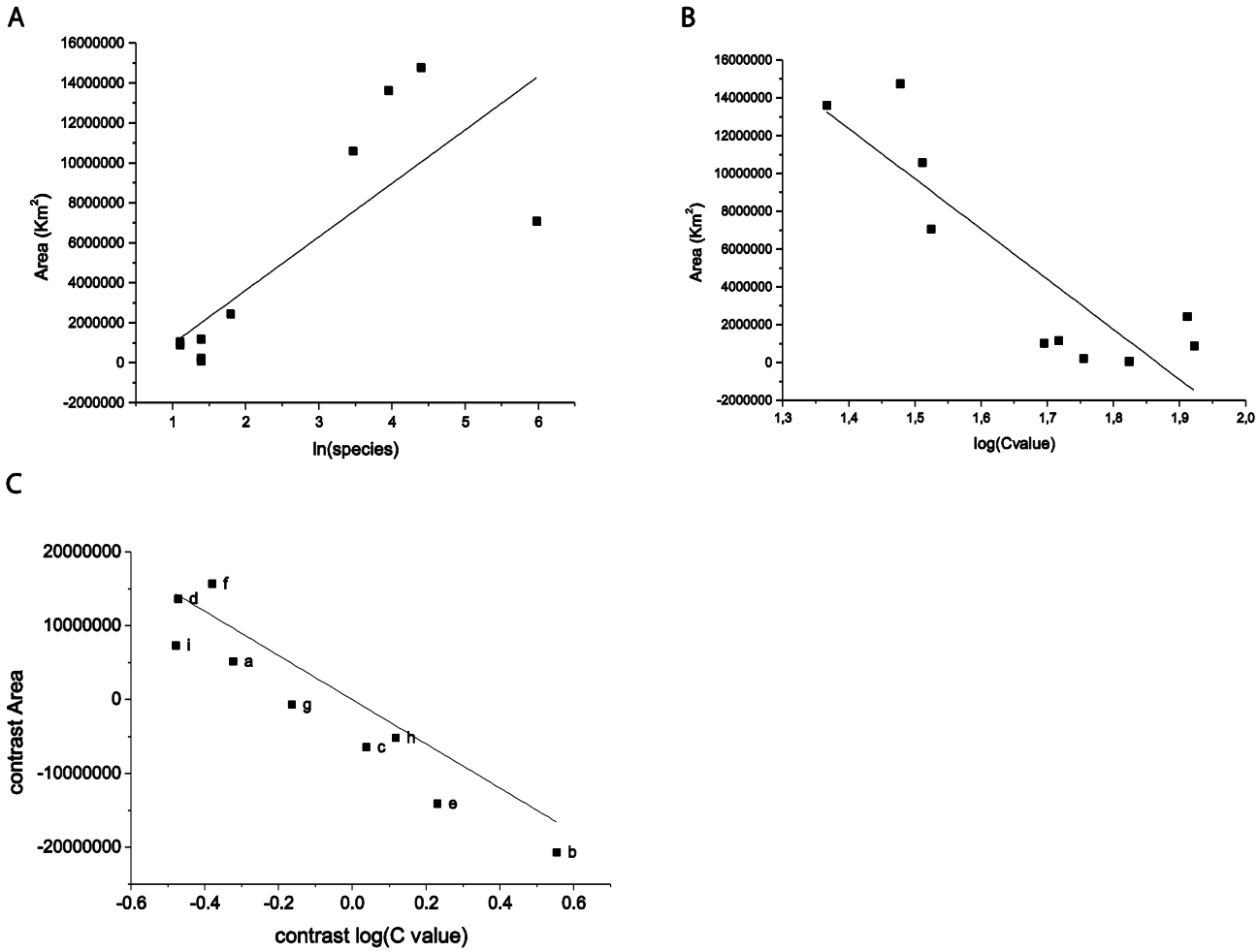

Figure 5

Figure 5: Ecological correlates with genome size (area).
A) Standard regression analysis of geographical area versus C-value (R^2 = 0.74, p-value = 0.0009). B) Regression analysis of area versus C-value using independent contrasts (R^2 = 0.78, p-value = 0.0004). C) Scatter plot of area versus species diversity.

Ecological correlates with genome size: area

Older lineages have had more time to disperse over larger areas, suggesting that geographical area might correlate with species diversity. As populations disperse they will encounter barriers that result in geographic isolation, and the number of barriers and isolated populations are expected to increase as the geographic area increases. We therefore examined the relationship between species diversity and area, assuming that larger areas will tend to accommodate greater species diversity across the salamander families. The scatter plot in Figure 5A reveals that area divides

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

species diversity into two groups (above and below a range of 45,000,000 km^2) in a manner similar to the two classes of genome size and species diversity observed in Figure 2B. Standard regression analysis suggests that area and species diversity are potentially correlated (R^2 = 0.58, p-value = 0.006; Table 1).

Given the findings described above concerning the relationship between species diversity and C-value, we next examined the relationship between genome size and geographic area for species in each of the ten salamander families. Figure 5B shows a strong correlation between C-value and geographical area: lineages with smaller genome sizes occupy larger geographical areas (adjusted R^2 = 0.74, p-value = 0.0009, Table 1) Regression analysis shown in Figure 5C using independent contrasts confirms the strong association between genome size and area (R^2 = 0.80, p-value = 0.0003), indicating that changes in area occupied correlate negatively with changes in genome size. Families with genome sizes greater than or equal to 50 pg appear, on the other hand, to be more restricted in terms of area and tend to occupy more similarly sized ranges (around 2,000,000 Km^2, Figure 5A).

	Adj. R^2	P value	Model
Log(SVL) vs ln(species)	0.44	**0.02**	linear
CV of C value vs ln(species)	0.85	**0.0002**	linear
Log(C value) vs log(SVL)	0.19	0.1	linear
Stem-Crown vs ln(species)	0.68	**0.002**	linear
Stem-Crown vs log(C value)	0.70	**0.002**	linear
Ln(species) vs Area	0.58	**0.006**	linear
Log (C value) vs Area	0.74	**0.0009**	linear
Log(C value) vs Crown age	0.20	0.1	linear
No Proteidae: Log(C value) vs Crown age	0.70	**0.003**	linear
Log(C value) vs Area	0.80	**0.0003**	Linear IC
Log(C value) vs log(SVL)	0.35	**0.04**	Linear IC

Table 1: Regression analysis and independent contrasts

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Discussion

We report here a significant relationship between genome size and species diversity across the respective salamander families. Examining other factors including changes in body size, clade age and geographical area suggests that the effect of genome size on species diversity operates directly or indirectly through a number of different variables. Although body size itself is not significantly associated with genome size, a significant relationship exists between changes in body size and changes in average genome size as measured by independent contrasts. Clade age measured in terms of crown age appears to correlate significantly with genome size (negatively) but not with species diversity when the proteidae are included in the regression analysis (Table 1). In contrast, stem-to-crown age is significantly associated with both genome size and species diversity, suggesting that species diversity and genome size evolve in parallel. Consistent with the assumption that geographical area tends to reflect a greater number and variety of habitats, we also found evidence that genome size decreases while species diversity increases with geographical area.

Species diversity in salamander families and genera has been shown here to increase as the variance in genome size increases (Figure 2C) (*6, 21*). CV in genome size, however, does not significantly correlate with genome size itself or with body size (not shown), indicating that species diversity is associated with genome size diversity in salamander families. Variance in genome size within taxa therefore appears to have important implications for species diversity. Our observations confirm and extend that view to include an additional effect due to average genome size: salamander families that tend to have overall smaller genomes comprise greater numbers of species. Recently, an early ancestor of salamanders has been estimated to have a large genome size of approximately 36.7 pg, which is typical of neotenic species (*35*). The analysis of the ten families presented here indicates that average genome size has either expanded or contracted since sister

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

pairs split from the ancestral node (Figure 1). Deciding between these alternatives at the family-level of sister-pairs nevertheless leads to the same conclusion: reductions in average genome size either promote species diversity, or amplifications in genome size diminish species diversity.

Our findings also support earlier proposals that junk DNA negatively impacts species diversity as nearly neutral DNA accumulates within and across lineages (*10, 12, 43*). Junk DNA, although nearly neutral in terms of sequence, is increasingly recognized as a major chromatin modulator that mediates both facultative and constitutive chromatin modifications (*44, 45*). These modifications have a genome wide impact on the transcriptional and DNA replication programs during development (*46, 47*); and, as we would like to suggest here, during speciation. Genome size-associated differences in chromatin re-modeling and re-organization, for example, have the potential to affect critical evolutionary parameters including DNA replication, mutation rates and gene expression (*2, 48-51*), which can contribute to the allelic incompatibilities that are associated with reproductive isolation and speciation.

Standardized contrasts indicate evolutionary rates (*36, 37*). Our findings therefore suggest that the rate at which body size changes is significantly related to the rate at which genome size changes, and hence the morphological evolution and evolution of corresponding life-history traits in salamanders is, in part, contingent upon changes in genome size. The two distinct classes of salamander family genome sizes evident in Figure 2B, for example, correspond to the obligate paedomorphs including the Sirenidae, Cryptobranchidae, Proteidae and Amphiumidae, which have genome sizes larger than 50 pg on average, and families comprising metamorphic, facultative paedomorphs and direct developing species, which have average genome sizes that are less than 50 pg. Neoteny and related life-history traits in salamanders thus appear to be closely associated

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

with the evolution of exceptionally large genomes, which supports the proposal that non-coding DNA participates in organizing the transcriptional program that determines higher physiological traits such as paedomorphy.

Finally, we found that genome size varies with two different measurements of time: crown age and stem-to-crown age, which measures approximately the phylogenetic time elapsed before the formation of a clade (*42*). The association between older crown age and smaller genome size when the Proteidae are excluded is difficult to interpret (Figure 4B); but it suggests that families with smaller average extant genome sizes have a higher propensity to speciate, resulting in earlier adaptive radiations compared to families with larger average genome sizes. The association between higher species diversity and shorter stem-to-crown age is consistent with a greater tendency of those families to begin radiating earlier with respect to stem age (Figure 4C), and hence to accumulate higher biodiversity than the class with larger genomes.

If habitat and niche availability both increase with geographic area, then changes in genome size might have coincided with adaptations that made available habitats and niches more accessible to dispersing ancestral urodela populations. The strong correlation and negative slope between independent contrasts of geographic area and C-value, for example, indicate that the rate at which C-value changes coincides with the rate at which geographic range changes (adjusted R^2 = 0.787). Families with lower average C-value, however, tend to have higher genome size variance (*6, 21*). Whether larger geographical areas reflect an influence on species diversity due either to an overall change in average C-value per family or to an increase in genome size variance (or both) remains an open question. In conclusion, our observations indicate that genome size variation and changes in genome size are important factors contributing to salamander species diversity and evolution.

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgements

BS is supported by a grant from Human Frontier Science Program (RGY0079). JH benefited from support from John Bechhoefer's lab, Physics Department, Simon Fraser University.

References

1. T. R. Gregory. (http://www.genomesize.com, 2015).
2. C. J. Metcalfe, D. Casane, Accommodating the load: The transposable element content of very large genomes. *Mob Genet Elements* **3**, e24775 (2013).
3. A. F. Palazzo, T. R. Gregory, The case for junk DNA. *PLoS Genet* **10**, e1004351 (2014).
4. T. Cavalier-Smith, Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J Cell Sci* **34**, 247-278 (1978).
5. D. Francis, M. S. Davies, P. W. Barlow, A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Ann Bot* **101**, 747-757 (2008).
6. S. Sessions, Evolutionary cytogenetics in salamanders. *Chromosome Res.* **16**, 183-201 (2008).
7. A. Larson, Developmental Correlates of Genome Size in Plethodontid Salamanders and Their Implications for Genome Evolution Author (s): Stanley K . Sessions and Allan Larson Reviewed work (s): Published by : Society for the Study of Evolution Stable URL : http://. **41**, 1239--1251 (2011).
8. T. Gregory, Macroevolution, Hierarchy Theory, and the C-value Enigma. *Paleobiology* **30**, 179-202 (2004).
9. J. E. Mank, J. C. Avise, Cladogenetic correlates of genomic expansions in the recent evolution of actinopterygiian fishes. *Proc Biol Sci* **273**, 33-38 (2006).
10. E. Olmo, Genome size and evolutionary diversification in vertebrates. *Italian Journal of Zoology* **73**, 167 - 171 (2006).
11. K. Kraaijeveld, Genome Size and Species Diversification. *Evolutionary Biology* **37**, 227-233 (2010).
12. A. E. Vinogradov, Genome size and extinction risk in vertebrates. *Proc Biol Sci* **271**, 1701-1705 (2004).
13. D. Wake, What Salamanders have Taught Us about Evolution. *Ann. Rev. Ecol. Evol. Syst.*, 333-352 (2009).
14. E. Nevo, A. Beiles, Genetic diversity and ecological heterogeneity in amphibian evolution. *Copeia* **19**, 565-592 (1991).
15. R. A. Pyron, J. J. Wiens, Large-scale phylogenetic analyses reveal the causes of high tropical amphibian diversity. *Proc Biol Sci* **280**, 20131622 (2013).
16. C. Sun, J. R. López Arriaza, R. L. Mueller, Slow DNA loss in the gigantic genomes of salamanders. *Genome Biol Evol* **4**, 1340-1348 (2012).
17. C. Sun, R. L. Mueller, Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. *Genome Biol Evol* **6**, 1818-1829 (2014).
18. J. Herrick, Genetic variation and replication timing, or why is there late replicating DNA? *Evolution*, (2011).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

19. J. Jurka, W. Bao, K. K. Kojima, O. Kohany, M. G. Yurka, Distinct groups of repetitive families preserved in mammals correspond to different periods of regulatory innovations in vertebrates. *Biol Direct* **7**, 36 (2012).
20. A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res* **16**, 203-215 (2008).
21. J. Herrick, B. Sclavi, A new look at genome size, evolutionary duration and genetic variation in salamanders. *Comptes Rendus Palevol* **13**, 611-615 (2014).
22. L. Bromham, X. Hua, R. Lanfear, P. F. Cowman, Exploring the Relationships between Mutation Rates, Life History, Genome Size, Environment, and Species Richness in Flowering Plants. *Am Nat* **185**, 507-524 (2015).
23. H. Schneider *et al.*, Are the genomes of royal ferns really frozen in time? Evidence for coinciding genome stability and limited evolvability in the royal ferns. *New Phytol* **207**, 10-13 (2015).
24. M. N. Puttick, J. Clark, P. C. Donoghue, Size is not everything: rates of genome size evolution, not C-value, correlate with speciation in angiosperms. *Proc Biol Sci* **282**, 20152289 (2015).
25. A. Pierce, B. Mitton, The relationship between genome size and genetic variation. *The American Naturalist* **116**, 850-861 (1980).
26. E. Parker, M. Kreitman, On the relationship between heterozygosity and DNA content. *The American Naturalist* **119**, 749-752 (1982).
27. A. Larson, A reevaluation of the relationship between genome size and genetic variation. *The American Naturalist* **118**, 119-125 (1981).
28. F. Frentiu, J. Ovenden, R. Street, Australian lungfish (Neoceratodus forsteri:Dipnoi) have low genetic variation at allozyme and mitochondrial DNA loci: A conservation alert? *Conservation Genetics* **2**, 63-67 (2001).
29. M. Matsui, A. Tominaga, W. Z. Liu, T. Tanaka-Ueno, Reduced genetic variation in the Japanese giant salamander, Andrias japonicus (Amphibia: Caudata). *Mol Phylogenet Evol* **49**, 318-326 (2008).
30. J. M. Hughes *et al.*, Extremely low microsatellite diversity but distinct population structure in a long-lived threatened species, the Australian lungfish Neoceratodus forsteri (Dipnoi). *PLoS One* **10**, e0121858 (2015).
31. A. A. Karlin, D. B. Means, Genetic variation in the aquatic salamander genus Amphiuma. *American Midland Naturalist* **132**, 1-9 (1994).
32. R. A. Pyron, J. J. Wiens, A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Mol Phylogenet Evol* **61**, 543-583 (2011).
33. R. A. a. W. J. J. Pyron, A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular phylogenetics and evolution* **61**, 543--583 (2011).
34. D. Marjanovic, M. Laurin, An updated paleontological timetree of lissamphibians with comments on the anatomy of Jurassic crown-group salamanders (Urodela). *Historical Biology: An International Journal of Paleobiology*, (2013).
35. M. Laurin, A. Canoville, M. Struble, V. Buffrénil, Early genome size increase in urodeles. *Comptes Rendus Palevol*, (2015).
36. T. Garland, Rate tests for phenotypic evolution using phylogenetically independent contrasts. *Am Nat* **140**, 509-519 (1992).
37. M. J. Oliver, D. Petrov, D. Ackerly, P. Falkowski, O. M. Schofield, The mode and tempo of genome size evolution in eukaryotes. *Genome Res* **17**, 594-601 (2007).

bioRxiv preprint doi: https://doi.org/10.1101/065425; this version posted July 22, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

38. C. L. Organ, A. Canoville, R. R. Reisz, M. Laurin, Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. *J Evol Biol* **24**, 372-380 (2011).
39. C. Organ, M. Struble, A. Canoville, V. Bruffénil, M. Laurin, Macroevolution of genome size in sarcopterygians during the water–land transition. *Comptes rendus palevol* **15**, 65-73 (2016).
40. C. C. Martin, R. Gordon, Differentiation trees, a junk DNA molecular clock, and the evolution of neoteny in salamanders. *J. Evol. Biol*. **8**, 339-354 (1995).
41. T. Stadler, D. L. Rabosky, R. E. Ricklefs, F. Bokma, On age and species richness of higher taxa. *Am Nat* **184**, 447-455 (2014).
42. S. Blair Hedges, J. Marin, M. Suleski, M. Paymer, S. Kumar, Tree of life reveals clock-like speciation and diversification. 2014.
43. A. E. Vinogradov, Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet* **19**, 609-614 (2003).
44. H. C. Macgregor, S. K. Sessions, The biological significance of variation in satellite DNA and heterochromatin in newts of the genus Triturus: an evolutionary perspective. *Philos Trans R Soc Lond B Biol Sci* **312**, 243-259 (1986).
45. K. H. Wei, J. K. Grenier, D. A. Barbash, A. G. Clark, Correlated variation and population differentiation in satellite DNA abundance among lines of Drosophila melanogaster. *Proc Natl Acad Sci U S A* **111**, 18793-18798 (2014).
46. T. Chen, S. Y. Dent, Chromatin modifiers and remodellers: regulators of cellular differentiation. *Nat Rev Genet* **15**, 93-106 (2014).
47. Q. Ding, D. M. MacAlpine, Defining the replication program through the chromatin landscape. *Crit Rev Biochem Mol Biol* **46**, 165-179 (2011).
48. F. Supek, B. Lehner, Differential DNA mismatch repair underlies mutation rate variation across the human genome. *Nature* **521**, 81-84 (2015).
49. T. C. Voss, G. L. Hager, Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat Rev Genet* **15**, 69-81 (2014).
50. S. Whalen, R. M. Truty, K. S. Pollard, Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* **48**, 488-496 (2016).
51. A. W. Shermoen, M. L. McCleland, P. H. O'Farrell, Developmental control of late replication and S phase length. *Curr Biol* **20**, 2067-2077 (2010).
52. H. Zhang, S. Gao, M. J. Lercher, S. Hu, W. H. Chen, EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **40**, W569-572 (2012).