doi:10.1111/evo.14468

Gigantic genomes of salamanders indicate that body temperature, not genome size, is the driver of global methylation and 5-methylcytosine deamination in vertebrates

Alexander Nichols Adams,[1,2] Robert Daniel Denton,[3,4] and Rachel Lockridge Mueller[1]

[1]*Department of Biology, Colorado State University, Fort Collins, CO 80523-1878*

[2]*E-mail: admsmkay@colostate.edu*

[3]*Department of Biology, Marian University, Indianapolis, IN 46222*

[4]*Division of Science and Math, University of Minnesota Morris, Morris, MN 56267*

Received October 1, 2021

Accepted January 14, 2022

Transposable elements (TEs) are sequences that replicate and move throughout genomes, and they can be silenced through methylation of cytosines at CpG dinucleotides. TE abundance contributes to genome size, but TE silencing variation across genomes of different sizes remains underexplored. Salamanders include most of the largest C-values – 9 to 120 Gb. We measured CpG methylation levels in salamanders with genomes ranging from 2N = ∼58 Gb to 4N = ∼116 Gb. We compared these levels to results from endo- and ectothermic vertebrates with more typical genomes. Salamander methylation levels are approximately 90%, higher than all endotherms. However, salamander methylation does not differ from other ectotherms, despite an approximately 100-fold difference in nuclear DNA content. Because methylation affects the nucleotide compositional landscape through 5-methylcytosine deamination to thymine, we quantified salamander CpG dinucleotide levels and compared them to other vertebrates. Salamanders and other ectotherms have comparable CpG levels, and ectotherm levels are higher than endotherms. These data show no shift in global methylation at the base of salamanders, despite a dramatic increase in TE load and genome size. This result is reconcilable with previous studies that considered endothermy and ectothermy, which may be more important drivers of methylation in vertebrates than genome size.

KEY WORDS: CpG dinucleotides, ectothermy, endothermy, genome size, methylation, salamanders.

Genomes are composed of sequences with radically different effects on organismal survival and reproduction. Essential genes are required to sustain life, and many are constitutively transcribed in all tissues. In contrast, transposable element (TE)–sequences capable of proliferation and movement throughout genomes–are often mutagenic, and their transcriptional silencing is required to maintain germline integrity and reproduction (Bourque et al. 2018). Transcription or silencing of different DNA sequences is facilitated by conformational changes to chromatin, which is achieved through epigenetic modifications to both DNA and histones (Venkatesh and Workman 2015). The suppression of TEs relies on transcriptional silencing through methylation of cytosines (5mC) that are adjacent to guanines (i.e., CpG dinucleotide sites) (Bird 2002; Law and Jacobsen 2010; Fedoroff 2012; Venkatesh and Workman 2015; Deniz et al. 2019). TEs are a major determinant of overall genome size, which varies >65,000-fold across eukaryotes (Pellicer et al. 2010). How TE suppression via methylation at CpG sites scales with increased TE load and genome size remains incompletely understood (Wolffe 1998; Jones and Wolffe 1999).

There have been few attempts to explore global methylation at the upper limits of genome size. In *Picea abies*, Norway spruce–which has an unusually large diploid genome size of 20 Gb–the percentage of methylated CpG sites is higher than that found in other monocots and eudicots with smaller genome sizes. CHG sites (where H is any nucleotide that is not a guanine) show the same pattern, whereas methylated CHH sites do not (Ausin et al. 2016). 5mC at CpG and CHG are both associated with TE silencing in plants, while 5mC at CHH sites is not, instead forming a barrier between heterochromatin and other genes (Kenchanmane Raju et al. 2019). This single data point suggests that as genome size increases, the percentage of methylated cytosines at TE-silencing sites increases. However, to our knowledge, the relationship between genomic gigantism and global methylation in animals remains untested, as does the relationship in any taxon with a genome size above 20 Gb.

Global methylation levels can also affect the nucleotide compositional landscape because methylation causes an increase in the rate of specific mutations. Methylated cytosines at CpG sites are predisposed to undergo transition mutations from C to T via deamination (Ehrlich and Wang 1981). Recently, Zhou et al. (2020) demonstrated that the proportion of CpG sites decreases as genome size increases across animal species, with genome sizes ranging from 89 Mb to 4 Gb, which was interpreted to reflect deamination-driven transition mutations occurring at higher frequencies as TE loads and associated methylation levels increase. Whether this relationship holds for animals with large genome sizes remains unknown.

The salamander clade is an appropriate model system for studying the epigenetic factors that control–or fail to control–TE activity in large genomes because it includes diploid genomes that range from 9 to 120 Gb (Šímová and Herben 2012; Sun et al. 2012; Sclavi and Herrick 2019; Decena-Segarra et al. 2020; Gregory 2020). These gigantic and variably sized genomes reflect the accumulation of extreme levels of TEs, associated with a shift in the dynamics of genome size evolution at the base of the clade approximately 200 million years ago (Batistoni et al. 1995; Sun and Mueller 2014; Liedtke et al. 2018). Additionally, there are several polyploid salamander species, providing a different mechanistic path to high levels of nuclear DNA by combining large genomes with increased chromosome copy number (Bogart et al. 2007, Bogart et al. 2009).

Here, we quantify methylation levels across salamanders that encompass their upper range of genome sizes, and we compare these values to those of other vertebrates with more typically sized genomes. We include both endotherms (i.e., birds and mammals) and ectotherms (i.e., fish, other amphibians, reptiles) because body temperature is associated with differences in global methylation (Jabbari et al. 1997). Endotherms, as well as ectotherms inhabiting warm environments, have higher rates of deamination of 5mC nucleotides, resulting in lower global methylation levels and fewer CpG dinucleotides genome-wide (Sved and Bird 1990; Jabbari et al. 1997; Varriale and Bernardi 2006; Bernardi 2007; Bucciarelli et al. 2009). We used these data to test whether methylation levels and genome-wide CpG dinucleotide levels are driven by genome size and/or mode of body temperature regulation. In previous studies, endothermy was confounded by a large genome size. The integration of the data from salamanders–which combine large genome sizes with ectothermy–allows us to decouple genome size from metabolic heat production, revealing that genome size alone lacks explanatory power for methylation and CpG dinucleotide levels across vertebrates.

Materials and Methods

SAMPLING DESIGN, TISSUE DISSECTION, AND DNA EXTRACTION FOR ANALYSIS OF GLOBAL METHYLATION

Our first goal was to test for variation in methylation levels across tissues in salamanders.

For this analysis, we focused on the model salamander *Ambystoma mexicanum*, which is widely used for studies of development and regeneration (Tank et al. 1976; Seifert et al. 2012; Keinath et al. 2015; Sessions and Wake 2020). Additionally, it is one of the few species of salamanders to have its entire genome assembled and annotated (Keinath et al. 2015; Nowoshilow et al. 2018). We analyzed the brain, heart, lung, liver, intestine, muscle, skin, and testes or ovaries from three adult males and three adult females.

Our second goal was to measure methylation levels in salamanders with large genomes. For this analysis, we used muscle tissue as the single representative tissue, and we chose species that span the upper range of diploid genome sizes and ploidy in salamanders. The diploid species were *Plethodon cinereus* and *Necturus beyeri*, as well as *A. mexicanum. Plethodon cinereus* has a range of published genome sizes with an average of approximately 22 Gb (Gregory 2020), although our measurements indicate a genome size of 29.3 Gb (Itgen et al. 2021). The genome size of *Necturus beyeri* is unknown, but size estimates for congeners are 120 Gb (*N. lewisi* and *N. punctatus*) and approximately 85 Gb (*N. maculosus*) (Gregory 2020). *Ambystoma mexicanum* has a genome size of 32.15 Gb (Keinath et al. 2015). The polyploid taxa were triploid and tetraploid unisexual biotypes formed naturally via kleptogenesis (Bogart et al. 2007), including one haploid *A. laterale* genome (∼29 Gb) and either two or three *A. jeffersonianum* genomes (haploid = ∼29 Gb; ∼87–116 Gb total). We also included the diploid *A. jeffersonianum*. Three individuals were sampled from each species/biotype (Table 1).

Table 1. Genome sizes for salamanders sampled for global methylation.

Species	Genome Size (Gb)
Plethodon cinereus	29
Ambystoma jeffersonianum	29
Ambystoma laterale	29
Ambystoma mexicanum	32
Necturus beyeri[1]	85–120

[1]**The genome size for *Necturus beyeri* is unknown, but congeneric genome sizes range from 85 Gb (*N. maculosus*) to 120 Gb (*N. lewisi*, *N. punctatus*).**

Our third goal was to compare methylation levels in salamanders with those from vertebrates with more typical genome sizes. For this analysis, we combined our new data with previously published methylation levels of brain and liver tissue from fish (*Carassius auratus*, *Perca flavescens, Salvelinus namaycush*), birds (*Coturnix japonica*, *Gallus gallus*), amphibians (*Xenopus laevis*, *Xenopus tropicalis*), reptiles (*Anolis carolinensis)*, and mammals (*Delphinus delphis*, *Lagenorhynchus acutus*, *Neovison vison, Ursus maritimus*) (Basu et al. 2013; Head et al. 2014).

Plethodon cinereus were field collected between May and August of 2018 in South Cherry Valley and Oneonta, Otsego County, New York, under the New York State Department of Environmental Conservation scientific collection permit #2303. *Necturus beyeri* were obtained commercially through a private vendor in November 2019. *Ambystoma mexicanum* was obtained from the Ambystoma Genetic Stock Center in July 2019. *Ambystoma jeffersonianum* and *Ambystoma laterale* were field collected from northern Kentucky (Kenton County; Kentucky Fish and Wildlife collection permit SC20111082) and Connecticut (Litchfield County; no permit required). Polyploid taxa were collected from New Jersey (Morris County; collected by staff members of New Jersey Division of Fish and Wildlife, nongame species program, under staff jurisdiction), Kentucky (Kenton County; permit SC20111082), and Ohio (Crawford County; Department of National Resources permit 18–164). *Ambystoma* field collections took place in the spring of 2019 and 2020.

Specimens of *A. mexicanum*, *P. cinereus*, and *N. beyeri* were euthanized in MS222. Tissues were dissected, flash frozen in liquid nitrogen, and stored at −80°C until use. DNA was extracted using a DNeasy Blood and Tissue extraction kit (Qiagen) following the manufacturer's instructions, including the optional RNase treatment step. DNA concentration and quality (260/280 ratio) were measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific). For the polyploid *Ambystoma* samples and *A. jeffersonianum*, clippings of 1–3 mm of tissue from each animal's tail were collected and stored immediately in 90% ethanol and then stored at −20°C until use. Because these animals are difficult to obtain in the field, we were unable to sacrifice them to obtain internal organ tissue. DNA was extracted using DNeasy Blood and Tissue extraction kits (Qiagen) using a double final elution step with a 25% decrease in volume to increase the DNA concentration. DNA concentration was measured using a Qubit 3 fluorometer (ThermoFisher Scientific) with the dsDNA BR (Broad Range) assay. The ploidy and genomic composition of each sample were evaluated by sequencing mitochondrial loci and genotyping species-specific microsatellite markers as detailed in previous work (Denton et al. 2017). All animal work was carried out in accordance with either Colorado State University IACUC protocol number 17–7189A or University of Minnesota Morris IACUC protocol number 1901–36686A.

LUMINOMETRIC METHYLATION ASSAY AND DATA STANDARDIZATION

To measure DNA methylation levels, we used the luminometric methylation assay (LUMA), a form of pyrosequencing that targets CpG dinucleotides capable of methylation (Karimi et al. 2006). LUMA runs were carried out by the sequencing facility EpigenDx (Hopkinton, MA, USA) using a PyroMark MD system from Qiagen. Duplicate LUMA runs were performed for each DNA sample. All assays included four Lambda DNA standards with methylation percentages of 0%, 50%, 60%, and 100% as internal controls. To account for any differences across assay runs, we calibrated against the internal controls using inverse regression calibration (Ott and Longnecker 2015). Random effects were added per subject and per combination of subject and tissue type to limit batch effects and account for pseudoreplication of duplicate runs.

COMPARATIVE ANALYSES OF THE GLOBAL METHYLATION

First, we tested for differences in methylation levels among the *A. mexicanum* tissues using a mixed model ANOVA, with tissue as a fixed effect factor and individual and duplicate runs as random effect factors, followed by a Tukey HSD to test for significance between tissues. Next, we tested for differences in methylation levels among the polyploid unisexual salamander biotypes *A. laterale* (LJJ) and *A. laterale* (LJJJ) and the four diploid species *A. jeffersonianum* (JJ), *A. mexicanum*, *P. cinereus*, and *N. beyeri* using a mixed model ANOVA, with species as a fixed effect factor and individual and duplicate runs as random effect factors. We then performed a Tukey HSD to test for significance between species/biotype. Finally, we tested whether methylation levels vary across vertebrates as a function of salamander vs. non-salamander (i.e., genome size ≥29 Gb vs. genome size ≤6.4 Gb), ploidy (i.e., diploid vs. polyploid), and body temperature regulation (i.e., endothermy vs. ectothermy). We assigned species

to the appropriate subgroup(s) and tested for variation between subgroups using linear regression contrasts. We note that we are combining our data with previously published results, but are unable to incorporate a study effect because of the small sample size and dependency in the categorical variables (i.e., our study is the only one to include salamanders). We carried out all analyses in R Studio (RStudio Team 2021; R Core Team 2021) using R packages emmeans (Lenth 2021), parameters (Lüdecke et al. 2020), and lme4 (Bates et al. 2015). We used the ggpubr package (Kassambara 2020) to visualize the results.

EFFECTS OF GLOBAL METHYLATION AND GENOME SIZE ON CpG DINUCLEOTIDE LEVELS

Our final goal was to test whether enormous genome sizes translate into changes in the nucleotide compositional landscape, as predicted based on the relationships among genome size, transposable element load, methylation level, and C $\rightarrow$ T deamination mutation inferred across 53 animals (47 vertebrates, 6 invertebrates) with more typical genome sizes (89 Mb to 4 Gb) (Zhou et al. 2020). C $\rightarrow$ T deamination mutations occur preferentially at methylated CpG dinucleotides (Ehrlich and Wang 1981); thus, higher levels of methylation, associated with methylation-based silencing of a greater TE load, are predicted to correlate with fewer CpG dinucleotides genome-wide than are expected based on nucleotide frequencies (Zhou et al. 2020). We obtained publicly available genomic sequence data for nine species of salamanders with varying genome sizes: *Desmognathus ochrophaeus* ($\sim$15 Gb; SRA046120.1), *Pleurodeles waltl* ($\sim$25 Gb; DRX13131369), *Eurycea tynerensis* ($\sim$25 Gb; SRA046121.1), *Batrachoseps nigriventris* ($\sim$26 Gb; SRA046116.1), *Ambystoma mexicanum* ($\sim$30 Gb; ASM291563v2), *Bolitoglossa occidentalis* ($\sim$43.5 Gb; SRA046118.1), *Aneides flavipunctatus* ($\sim$45 Gb; SRA046114.1), *Bolitoglossa rostrata* ($\sim$48 Gb; SRA046119.1), and *Cryptobranchus alleganiensis* ($\sim$55 Gb; SRA073787) (Gregory 2020). *Ambystoma mexicanum* has a well-assembled genome (including repetitive regions) based on deep, short-read coverage, long reads, and optical mapping (Nowoshilow et al. 2018). *Pleurodeles waltl* has a short-read Illumina assembly, with the repeat elements constructed through a majority vote k-mer extension algorithm (Elewa et al. 2017); however, we used the unassembled Illumina HiSeq 2000 trimmed and quality filtered reads ($\sim$0.25$\times$ coverage) to avoid bias introduced by the relative ease of assembling genic versus repetitive sequences. Datasets for the remaining species are all low-coverage 454 shotgun reads representing approximately 0.1–1% of the total genome (Sun et al. 2012; Sun and Mueller 2014). Each dataset was run through a pipeline bash script that removed tags and counted the numbers of each individual nucleotide and CpG dinucleotide present. The observed/expected ratio (O/E) of CpG dinucleotides was calculated as the observed number of CpG dinucleotides, CpG/N, divided by the expected number of CpG dinucleotides, $(C \times G)/(N \times N)$. C is observed cytosines, G is observed guanines, CpG is the observed CpG dinucleotides, and N is the total number of base pairs. Overall CpG O/E = $(CpG/N)/((CxG)/(N^2))$ (Zhou et al. 2020).

We tested for a relationship between genome size and CpG O/E dinucleotides among the nine salamander species using linear regression analysis. We also corrected for phylogeny using phylogenetic independent contrasts (PIC) with the tree for the nine focal species subsampled from a comprehensive amphibian phylogeny (Pyron and Wiens 2011). PIC analyses were carried out using the R packages ape (Paradis and Schliep 2019), Geiger (Harmon et al. 2008), nlme (Pinheiro et al. 2020), and phylotools (Revell 2012). Next, we compared the salamander CpG O/E dinucleotide values to all species published in (Zhou et al. 2020); this study included full genome assemblies from species with genome sizes ranging from 0.36 to 4.78 Gb and showed a negative correlation between genome size and CpG O/E (Zhou et al. 2020). To ensure that the datasets were comparable, we subsampled Illumina reads of 20 species from the Zhou et al. dataset to generate datasets with coverage that approximated the coverage of the salamander datasets, and we calculated CpG O/E from these subsampled datasets. We then compared the CpG O/E values based on full genome assemblies to those we calculated from subsampled short read datasets using linear regression and found a significant positive correlation between the two sets of estimates ($p < 0.00001$, $R = 0.83$; subsampled estimates equally likely to be slightly above or below whole-genome estimates), suggesting that comparisons across these different datasets are unlikely to introduce bias due to comparing whole-genome assemblies to unassembled read datasets.

Results

METHYLATION LEVELS ARE LOWER IN THE LUNGS

DNA concentrations and 260/280 ratios are summarized in Table S1. Methylation levels significantly differed between the lungs and some of the other tissues: brain–lungs ($p = 0.003$), gill–lungs ($p = 0.012$), heart–lungs ($p = 0.012$), and liver–lungs ($p = 0.001$; $df = 8$ and an overall F-value of 4.5; Fig. 1). Lungs were not used for downstream across-species comparisons.

METHYLATION LEVELS DO NOT VARY ACROSS SALAMANDER SPECIES

DNA concentrations and 260/280 ratios, when available, are summarized in Table S1. Methylation levels were not significantly different between salamander species, despite differences in diploid genome size and ploidy ($df = 5$, F-value = 1.11, $p = 0.4$; Fig. 2).

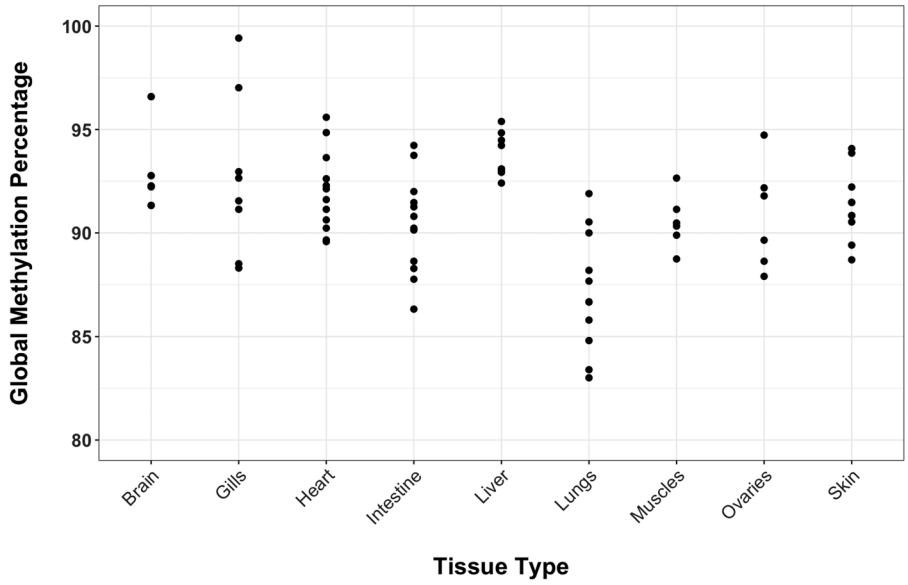

Figure 1. Methylation levels (percentage of methylated cytosines at CpG dinucleotide sites) of DNA extracted from different tissues in *Ambystoma mexicanum*. Lungs have significantly lower methylation levels than some tissues: Brain–Lungs, Heart–Lungs, Gill–Lungs, and Liver–Lungs.

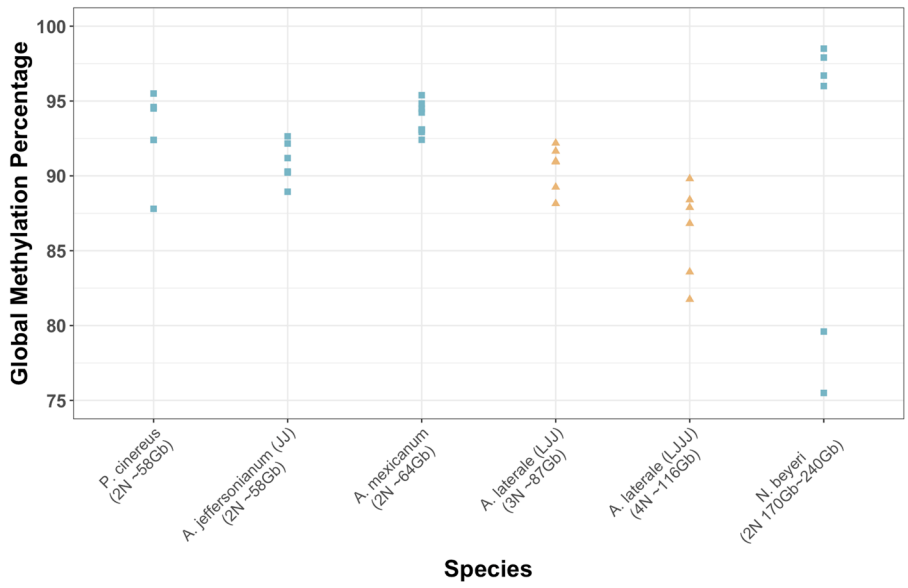

Figure 2. Methylation levels (percentage of methylated cytosines at CpG dinucleotide sites) of six different species/biotypes of salamanders with different amounts of nuclear DNA. Diploid species are represented by teal squares, and polyploid species are represented by orange triangles. There were no significant differences among the groups.

ACROSS VERTEBRATES, METHYLATION LEVELS REFLECT BODY TEMPERATURE REGULATION, NOT NUCLEAR DNA CONTENT

Ectotherms as a whole (including salamanders) have significantly higher levels of methylation than endotherms ($df = 108$, t-ratio $= -15.84$, $p < 0.001$, Fig. 3). Salamanders themselves have methylation levels that are not significantly different from those of other ectotherms ($df = 108$, t-ratio $= 0.88$, $p = 0.38$), despite their enormous genome sizes. Excluding salamanders, there was still a significant difference between ectotherms and endotherms ($df = 108$, t-ratio $= -9.47$, $p < 0.001$). Because ectotherms and endotherms were assayed in the same studies (Basu et al. 2013; Head et al. 2014), a study effect is unlikely to have produced this pattern. Because our study includes only salamanders, we cannot formally exclude the possibility that a study effect produced the overlap in methylation levels between

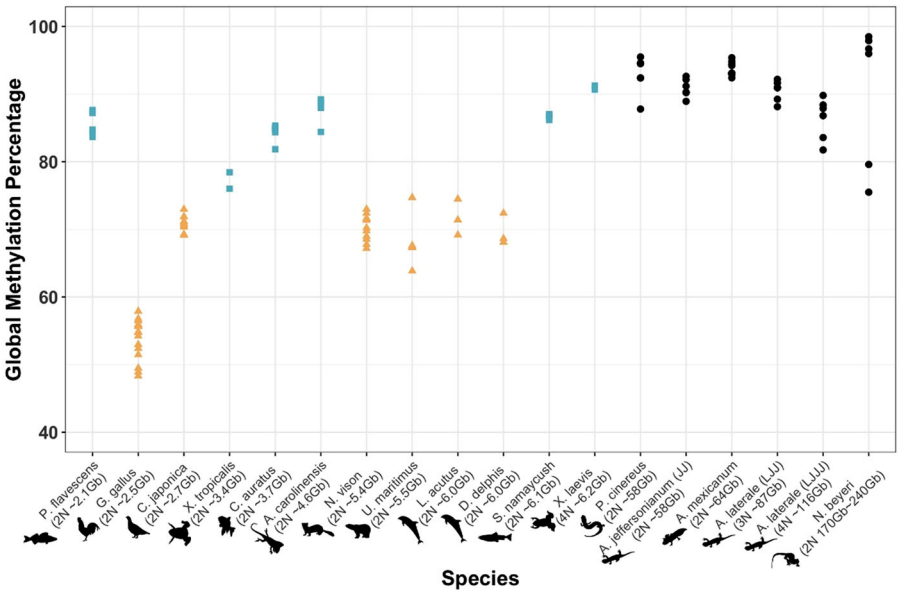

Figure 3. Methylation levels (percentage of methylated cytosines at CpG dinucleotide sites) across vertebrates. Diploid, triploid, or tetraploid genome sizes are listed with species names. Orange triangles are endotherms, teal squares are ectotherms, and black dots are salamanders. Ectotherms had higher methylation levels than endotherms ($p < 0.001$). All pictures are under public domain license, courtesy of Phylopics.org.

salamanders and the other ectotherms. However, we consider this unlikely because of the inclusion of internal controls in LUMA analyses. Polyploidy (seen in *X. laevis* and the two uniexual *Ambystoma* biotypes) does not significantly affect methylation levels ($df = 108$, t-ratio $= 0.38$, $p = 0.71$).

Because LUMA measures cytosine methylation at CpG dinucleotides within a specific genomic context (i.e., CCGG tetranucleotides), we tested whether endotherms and ectotherms differ in the percentage of CpG dinucleotides that are found within CCGG tetranucleotides in a way that would confound our interpretation of the difference we report between endotherm and ectotherm methylation. Of our 18 focal species, genome sequence data are publicly available for 10 species (4 ectotherms and 6 endotherms). The percentage of CpG dinucleotides localized within CCGG tetranucleotides ranged from 12% to 21% across these 10 species, and ANOVA showed no significant difference in the percentage between ectotherms and endotherms ($df = 1$, F-value $= 0.399$, $p = 0.545$; Table S2).

LUMA is sensitive to DNA quality, which we report in Table S1. We do not have 260/280 ratios for the *Ambystoma* biotypes; thus, we do not have the same measures of DNA quality for all samples. However, degraded DNA samples produce an underestimate of DNA methylation levels (Head et al. 2014); thus, the higher methylation levels we report in ectotherms (including all salamanders) are unlikely to reflect any undetected differences in sample quality.

LARGE, HIGHLY METHYLATED GENOMES SHOW NO RELATIONSHIP BETWEEN DEAMINATION-DRIVEN TRANSITION MUTATIONS AND GENOME SIZE

The observed versus expected ratios of CpG dinucleotides (CpG O/E) for nine species of salamanders ranged from 0.47 to 1.08. There is no significant relationship between genome size and CpG O/E dinucleotides across salamanders, despite an approximately threefold difference in genome size (Fig. 4). We note that *C. alleganiensis* is a statistical outlier, with an O/E CpG value of 1.08 that invites further study; however, our conclusions are not affected by this data point. Overall, the range of CpG O/E values in salamanders is higher than those seen in the largest genomes sampled in previous analyses of tetrapods (2 Gb–4 Gb, CpG O/E range 0.13–0.54) (Zhou et al. 2020), suggesting that the negative correlation between genome size and CpG O/E demonstrated across a smaller range of animal genome sizes does not hold across the full range of animal genome sizes.

Discussion

BODY TEMPERATURE, NOT NUCLEAR DNA CONTENT, IS THE MAIN PREDICTOR OF METHYLATION LEVELS AND CpG DINUCLEOTIDE LEVELS ACROSS VERTEBRATE GENOMES

Despite spanning an approximately 100-fold difference in genome size, our results show no significant differences in

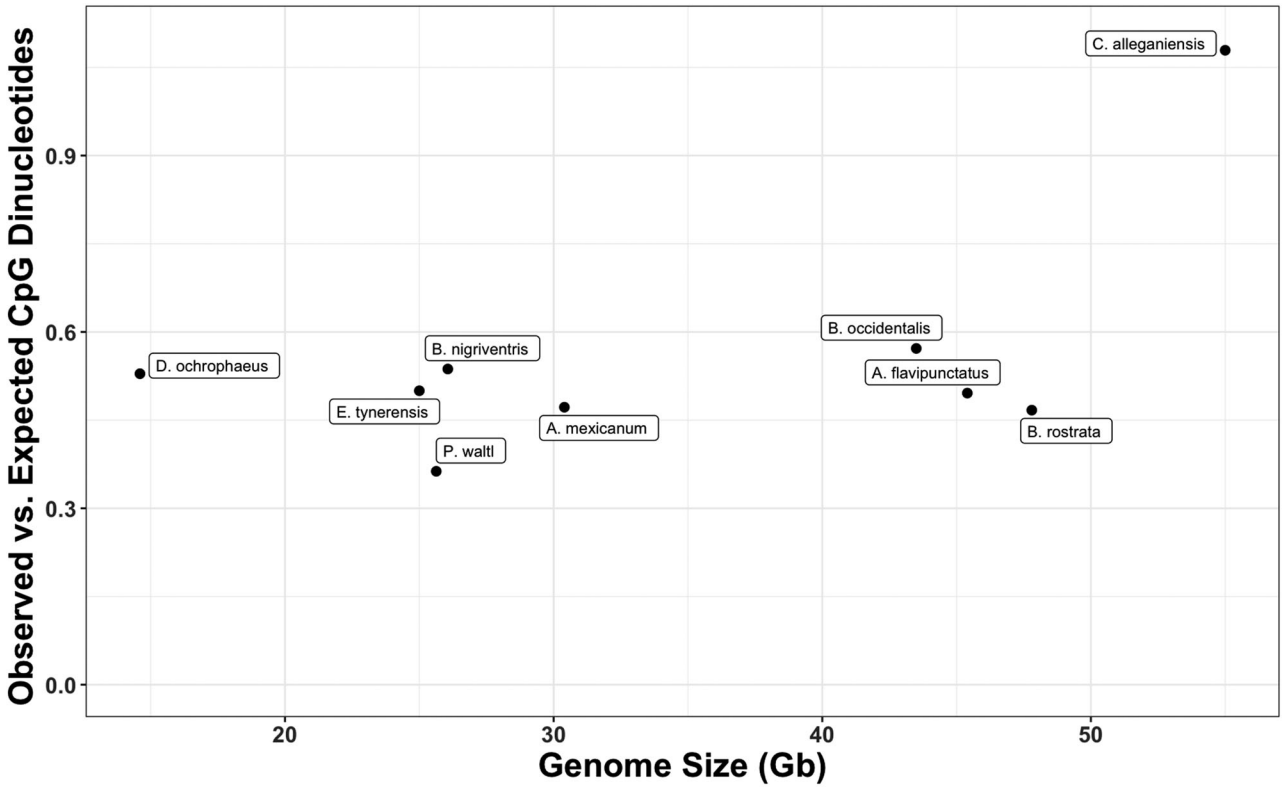

Figure 4. The observed versus expected ratios of CpG dinucleotides (O/E) for nine species of salamanders. There was no significant correlation between genome size and O/E CpG ($p = 0.15$).

methylation levels across ectotherms. Although we are comparing methylation across different tissue types (i.e., muscle for the salamanders, liver and brain for the other ectotherms), our across-tissue analysis within *Ambystoma mexicanum* shows no significant differences in methylation levels among these three tissues; using muscle allowed us to incorporate results from polyploid *Ambystoma* biotypes where we were limited to tail tips. Overall, our across-tissue analysis supports comparing different tissues from different taxa when necessary, provided the lungs are excluded.

Our results also show no significant difference in the methylation levels of polyploids versus diploids, although tetraploid *X. laevis* has higher methylation than diploid *X. tropicalis*. This may reflect the fact that *X. laevis* is a 30-million-year-old polyploidization event (Tymowska and Fischberg 1973; Hughes and Hughes 1993), whereas polyploidization happens anew each generation in *Ambystoma*. We note that the two species with the most nuclear DNA–the tetraploid *Ambystoma* biotype LJJJ and the diploid *Necturus beyeri*–have the largest intraspecific variances in methylation, encompassing the lowest values. We are unable to distinguish between biological and methodological reasons for this variation, but further sampling will be important to resolve this issue. If the variation is biological, it would suggest the interesting possibility that larger genomes with proportionally more noncoding sequences are tolerant of a broader range of epigenetic states.

Endotherms had significantly lower levels of methylation than ectotherms, with or without the inclusion of salamanders. These results suggest that the largest predictor of DNA methylation levels in vertebrates is the maintenance of a high body temperature by the production of metabolic heat, and not the amount of DNA in the nuclear genome. Endothermy and ectothermy have been previously proposed as the primary determinants of global methylation levels in vertebrates (Jabbari et al. 1997). Mammals and birds, which have high, endothermically maintained body temperatures, have lower methylation levels than amphibians and reptiles, which have lower body temperatures. This difference in methylation is parallel in congeneric species of gobies that live in vastly different temperatures in the Gulf of California–the species that lives in warmer temperatures has lower levels of DNA methylation (Bucciarelli et al. 2009). Similarly, temperate and tropical fish have lower methylation levels than polar fish (Varriale and Bernardi 2006). In all of these cases, the lower methylation levels found in animals with higher body temperatures likely reflect the higher rates of 5mC deamination of CpG dinucleotides at warmer temperatures, leading to faster loss of methylated cytosines from the genome (Jabbari et al. 1997; Varriale and Bernardi 2006; Bernardi 2007; Bucciarelli et al. 2009).

High methylation levels likely evolved at the base of vertebrates and remain high in ectotherms, irrespective of genome size; the independent evolutionary acquisitions of endothermy resulted in lower methylation levels.

5mC deamination rates should also impact the global nucleotide composition landscape because deamination causes transition mutations from C to T. Thus, endothermic vertebrates, as well as ectotherms living at high temperatures, should have fewer CpG dinucleotide sites than expected based on their nucleotide frequencies (lower O/E CpG values). Our results, in combination with those of Zhou et al. (2020), suggest that endothermy and ectothermy in vertebrates are also the primary predictors of CpG O/E values. Our salamander CpG O/E values broadly overlap with those of the 14 fish, 3 reptile, and 1 frog species sampled by Zhou et al., with genome sizes ranging from 0.36 to 2.86 Gb. In contrast, the five bird and 88 mammal species sampled by Zhou et al., with genome sizes ranging from 1.06 to 4.78 Gb, have lower CpG O/E, as predicted by their endothermic body temperature regulation. We reanalyzed the data from Zhou et al., running a linear regression with first genome size as the predictor, followed by ectothermy/endothermy. The R^2 value increased from 0.26 (when using genome size as the predictor) to 0.48 (when using ectothermy vs. endothermy). This suggests that much of the signal within the Zhou et al. dataset comes from the fact that mammals are endothermic and tend to have larger genomes than fish, reptiles, and frogs—but it is endothermy itself, rather than genome size, that drives the CpG O/E pattern (Figures S1 and S2). Within the nonsalamander ectotherms alone, however, genome size was negatively correlated with CpG O/E based on linear regression analysis ($p = 0.009$), suggesting a possible relationship between TE load, methylation, and CpG loss in ectotherms with relatively small genomes that warrants further investigation.

Given these results, is there a potential mechanistic relationship between endothermy and the slightly increased TE load of mammals? Higher deamination rates would decrease the methylation of TE sequences in mammals, which might negatively impact their silencing and allow for greater activity. In addition, higher deamination rates would yield higher transition mutation rates of TE sequences, which in turn could have two outcomes: (1) the production of divergent TE sequences that escape sequence-specific TE silencing, leading to greater TE activity, and/or (2) the production of TE sequences that lose their functional ORFs, rendering them incapable of autonomous transposition and leading to decreased TE activity. There are also groups that do not show an association between endothermy and increased TE load and genome size. Birds–also endothermic–have lower TE loads and smaller genome sizes, and salamanders–ectotherms–have the highest TE loads and largest genomes among tetrapods. In total, TE activity and abundance reflect the interaction of diverse forces, which may or may not include a relationship between methylation, mutation, and TE silencing that contributes to mammals' slightly larger genome sizes relative to nonsalamander vertebrates. This is a worthy target of future research.

AUTHOR CONTRIBUTIONS

A.N.A. was associated with study conception, experimental design, data collection, data analysis, initial draft of manuscript, and revision of manuscript. R.D.D. was associated with experimental design, data collection, revision of manuscript, and funding. R.L.M. was associated with study conception, experimental design, initial draft of manuscript, revision of manuscript, and funding.

ACKNOWLEDGMENTS

The authors thank M. Itgen for *P. cinereus* tissues. The authors thank W. Zhou, G. Liang, P.L. Molloy, and P.A. Jones for providing their full dataset for analysis. The authors thank the members of A. Adams' dissertation committee K. Hoke, D. Sloan, and J. Hansen for helpful discussion and feedback throughout the project. The authors thank the *Ambystoma* Genetic Stock Center, funded by NIH Grant P40-OD019794. Funding for this project was provided by NSF grant 1911585 to R. Mueller, a UMN Grant-in-Aid award and NSF grant 2045704 to R. Denton, and by Colorado State University.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA ARCHIVING

The doi for our data is https://doi.org/10.5061/dryad.fqz612jtz

LITERATURE CITED

Ausin, I., S. Feng, C. Yu, W. Liu, H. Y. Kuo, E. L. Jacobsen, J. Zhai, J. Gallego-Bartolome, L. Wang, and U. Egertsdotter. 2016. DNA methylome of the 20-gigabase Norway spruce genome. Proc. Natl. Acad. Sci. USA 113:E8106–E8113.

Basu, N., J. Head, D.-H. Nam, J. R. Pilsner, M. J. Carvan, H. M. Chan, F. W. Goetz, C. A. Murphy, K. Rouvinen-Watt, and A. M. Scheuhammer. 2013. Effects of methylmercury on epigenetic markers in three model species: mink, chicken and yellow perch. Comp. Biochem. Physiol. Part C Toxicol. Pharmacol 157:322–327.

Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw https://doi.org/10.18637/jss.v067.i01.

Batistoni, R., G. Pesole, S. Marracci, and I. Nardi. 1995. A tandemly repeated DNA family originated from SINE-related elements in the European plethodontid salamanders (Amphibia, Urodela). J. Mol. Evol 40:608–615.

Bernardi, G. 2007. The neoselectionist theory of genome evolution. Proc. Natl. Acad. Sci. USA 104:8385–8390.

Bird, A. 2002. DNA methylation patterns and epigenetic memory. Genes Dev 16:6-21.

Bogart, J. P., J. Bartoszek, D. W. A. Noble, and K. Bi. 2009. Sex in unisexual salamanders: discovery of a new sperm donor with ancient affinities. Heredity 103:483–493.

Bogart, J. P., K. Bi, J. Fu, D. W. A. Noble, and J. Niedzwiecki. 2007. Unisexual salamanders (genus Ambystoma) present a new reproductive mode for eukaryotes. Genome 50:119–136.

Bourque, G., K. H. Burns, M. Gehring, V. Gorbunova, A. Seluanov, M. Hammell, M. Imbeault, Z. Izsvák, H. L. Levin, T. S. Macfarlan, et al. 2018. Ten things you should know about transposable elements. Genome Biol 19:1–12.

Bucciarelli, G., M. Di Filippo, D. Costagliola, F. Alvarez-Valin, G. Bernardi, and G. Bernardi. 2009. Environmental genomics: a tale of two fishes. Mol. Biol. Evol 26:1235–1243.

Decena-Segarra, L. P., L. Bizjak-Mali, A. Kladnik, S. K. Sessions, and S. M. Rovito. 2020. Miniaturization, genome size, and biological size in a diverse clade of salamanders. Am. Nat 196:634–648.

Deniz, Ö., J. M. Frost, and M. R. Branco. 2019. Regulation of transposable elements by DNA modifications. Nat. Rev. Gen 20:417-431.

Denton, R. D., K. R. Greenwald, and H. L. Gibbs. 2017. Locomotor endurance predicts differences in realized dispersal between sympatric sexual and unisexual salamanders. Funct. Ecol 31:915–926.

Ehrlich, M., and R. Y. H. Wang. 1981. 5-Methylcytosine in eukaryotic DNA. Science 212:1350-1357.

Elewa, A., H. Wang, C. Talavera-López, A. Joven, G. Brito, A. Kumar, L. S. Hameed, M. Penrad-Mobayed, Z. Yao, N. Zamani, et al. 2017. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. Nat. Commun 8.

Fedoroff, N. V. 2012. Transposable elements, epigenetics, and genome evolution. Science 338:758–767.

Gregory, T. R. 2020. Animal genome size database. https://www.genomesize.com.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: Investigating evolutionary radiation. Bioinformatics 24:129–131.

Head, J. A., K. Mittal, and N. Basu. 2014. Application of the LUminometric Methylation Assay to ecological species: tissue quality requirements and a survey of DNA methylation levels in animals. Mol. Ecol. Resour 14:943–952.

Hughes, M. K., and A. L. Hughes. 1993. Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. Mol. Biol. Evol 10:1360–1369.

Itgen, M. W., D. S. Siegel, S. K. Sessions, and R. L. Mueller. 2021. Genome size drives morphological evolution in organ-specific ways. bioRxiv 2021.08.17.456171.

Jabbari, K., S. Cacciò, J. P. Païs De Barros, J. Desgrès, and G. Bernardi. 1997. Evolutionary changes in CpG and methylation levels in the genomes of vertebrates. Gene 205:109-118.

Jones, P. L., and A. P. Wolffe. 1999. Relationships between chromatin organization and DNA methylation in determining gene expression. Semin. Cancer Biol 9:339–347.

Karimi, M., S. Johansson, D. Stach, M. Corcoran, D. Grandér, M. Schalling, G. Bakalkin, F. Lyko, C. Larsson, and T. J. Ekström. 2006. LUMA (LUminometric Methylation Assay) A high throughput method to the analysis of genomic DNA methylation. Exp. Cell Res 312:1989–1995.

Kassambara, A. 2020. Package 'ggpubr': "ggplot2" Based Publication Ready Plots. R Packag. version 0.4.0.

Keinath, M. C., V. A. Timoshevskiy, N. Y. Timoshevskaya, P. A. Tsonis, S. R. Voss, and J. J. Smith. 2015. Initial characterization of the large genome of the salamander *Ambystoma mexicanum* using shotgun and laser capture chromosome sequencing. Sci. Rep 5:16413.

Kenchanmane Raju, S. K., E. J. Ritter, and C. E. Niederhuth. 2019. Establishment, maintenance, and biological roles of non-CG methylation in plants. Essays Biochem 63:743-755.

Law, J. A., and S. E. Jacobsen. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet 11:204-220.

Lenth, R. V. 2021. Package: "emmeans": emmeans: Estimated marginal means, aka least-squares means.

Liedtke, H. C., D. J. Gower, M. Wilkinson, and I. Gomez-Mestre. 2018. Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. Nat. Ecol. Evol 2:1792–1799.

Lüdecke, D., M. Ben-Shachar, I. Patil, and D. Makowski. 2020. Extracting, computing and exploring the parameters of statistical models using R. J. Open Source Softw https://doi.org/10.21105/joss.02445.

Nowoshilow, S., S. Schloissnig, J. F. Fei, A. Dahl, A. W. C. Pang, M. Pippel, S. Winkler, A. R. Hastie, G. Young, J. G. Roscito, et al. 2018. The axolotl genome and the evolution of key tissue formation regulators. Nature 554:50–55.

Ott, R. L., and M. T. Longnecker. 2015. An introduction to statistical methods and data analysis. Cengage Learning.

Paradis, E., and K. Schliep. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Pellicer, J., M. F. Fay, and I. J. Leitch. 2010. The largest eukaryotic genome of them all? Bot. J. Linn. Soc 164:10–15.

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R. C. Team. 2020. Nlme: linear and nonlinear mixed effects model description. R Packag. version 3. 1–149.

Pyron, A. R., and J. J. Wiens. 2011. 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.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

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

RStudio Team. 2021. RStudio: Integrated Development for R. RStudio. PBC, Boston, MA. http://www.rstudio.com/.

Sclavi, B., and J. Herrick. 2019. Genome size variation and species diversity in salamanders. J. Evol. Biol 32:278–286.

Seifert, A., J. Monaghan, R. Voss, and M. Maden. 2012. Skin regeneration in adult axolotls: a blueprint for scar-free healing in vertebrates. PLoS One 7:32875.

Sessions, S. K., and D. B. Wake. 2020. Forever young: Linking regeneration and genome size in salamanders. Dev. Dyn 250:768-778.

Šímová, I., and T. Herben. 2012. Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. Proc. R. Soc. B Biol. Sci 279:867–875.

Sun, C., and R. L. Mueller. 2014. Hellbender genome sequences shed light on genomic expansion at the base of crown salamanders. Genome Biol. Evol 6:1818–1829.

Sun, C., D. B. Shepard, R. A. Chong, J. L. Arriaza, K. Hall, T. A. Castoe, C. Feschotte, D. D. Pollock, and R. L. Mueller. 2012. LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. Genome Biol. Evol 4:168–183.

Sved, J., and A. Bird. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proc. Natl. Acad. Sci. USA 87:4692–4696.

Tank, P. W., B. M. Carlson, and T. G. Connelly. 1976. A staging system for forelimb regeneration in the axolotl, *Ambystoma mexicanum*. J. Morphol 150:117–128.

Tymowska, J., and M. Fischberg. 1973. Chromosome complements of the genus *Xenopus*. Chromosoma 44:335–342.

Varriale, A., and G. Bernardi. 2006. DNA methylation and body temperature in fishes. Gene 385:111–121.

Venkatesh, S., and J. L. Workman. 2015. Histone exchange, chromatin structure and regulation of transcription. Nat. Rev. Mol. Cell Biol 16:178-189.

Wolffe, A. P. 1998. Packaging principle: how DNA methylation and histone acetylation control the transcriptional activity of chromatin. J. Exp. Zool 282:239–244.

Zhou, W., G. Liang, P. L. Molloy, and P. A. Jones. 2020. DNA methylation enables transposable element-driven genome expansion. Proc. Natl. Acad. Sci. U. S. A. 117:19359–19366.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1
Table S2
Figure S1
Figure S2