

DISSERTATION

EXPANDING ON EXPANSION: GENOME GIGANTISM AND ITS EFFECTS ON DNA  
METHYLATION, RNA SPLICING AND ORGANELLAR SCALING

Submitted by:

Alexander Nichols Adams

Department of Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2023

Doctoral Committee:

Advisor: Rachel Mueller

Jeffrey Hanson

Kim Hoke

Dan Sloan

Copyright by Alexander Nichols Adams 2023

All Rights Reserved

## ABSTRACT

### EXPANDING ON EXPANSION: GENOME GIGANTISM AND ITS EFFECTS ON DNA METHYLATION, RNA SPLICING AND ORGANELLAR SCALING

Across the tree of life, the correlated traits of genome size and cell size both vary by orders of magnitude, with the increase in genome size being largely attributable to an increase in transposable elements (TEs) throughout the genome. This accumulation of TEs affects many facets of the cell including DNA regulation, organellar scaling, and RNA transcription. This dissertation will explore all 3 of these facets through the lens of genome gigantism and address how these facets differ in large cells in comparison to cells that are more typical in size.

The first chapter will discuss methylation of cytosines at genomic CpG dinucleotide sites that silence TEs. TE abundance drives differences in genome size, but TE silencing variation across genomes of different sizes remains largely unexplored. Salamanders include most of the largest C-values — 9 to 120 Gb. We measured CpG methylation levels in salamanders with genomes ranging from  $2N = \sim 58$  Gb to  $4N = \sim 116$  Gb. We compared these levels to results from endo- and ectothermic vertebrates with more typical genomes. Salamander methylation levels are  $\sim 90\%$ , higher than all endotherms. However, salamander methylation does not differ from the other ectotherms, despite a  $\sim 100$ -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 have comparable CpG levels to other ectotherms, 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 by considering endothermy and ectothermy, which may be more important drivers of methylation in vertebrates than genome size.

The next chapter will look at how an increase in cell size affects organellar structure and abundance. Depending on their shape, organelles can scale in larger cells by increasing volume, length, or number. Scaling may also reflect demands placed on organelles by increased cell size. The 8,653 species of amphibians exhibit diverse cell sizes, providing a powerful system to investigate organellar scaling. Using transmission electron microscopy and stereology, we analyzed three frog and salamander species whose enterocyte cell volumes range from 228 to 10,593  $\mu\text{m}^3$ . We show that the nucleus increases in radius (i.e. spherical volume) while the mitochondria increase in total network length; the endoplasmic reticulum and Golgi apparatus, with their complex shapes, are intermediate. Notably, all four organelles increase in volume proportionate to cell volume. This pattern suggests that protein concentrations are the same across amphibian species that differ 50-fold in cell size, and that organellar building blocks are incorporated into more or larger organelles following the same “rules” across cell sizes, despite variation in metabolic and transport demands. This conclusion contradicts results from experimental cell size increases, which produce severe proteome dilution. We hypothesize that salamanders have evolved the biosynthetic capacity to maintain a functional proteome despite a huge cell volume.

Finally, the last chapter will be discussing differences in intronic splicing, an important step that pre-mRNA transcripts undergo during processing in the nucleus to become mature mRNAs. Although long thought to occur exclusively in a single step, some introns are now also known to be removed in multiple steps through a process called recursive splicing. This non-

canonical form of splicing is hypothesized to aid with intron splicing fidelity, particularly in longer introns. Using West African lungfish (*Protopterus annectens*; genome size ~40Gb) as a model, we use total RNA-seq data to test the hypothesis that gigantic genomes, which have relatively long introns, have increased levels of recursive splicing compared to genomes of more typical size. Our results reveal levels of recursive splicing at conserved sites similar to those seen in humans, suggesting that genome-wide intronic expansion accompanying evolutionary increase in genome size is not associated with the evolution of high levels of recursive splicing. However, in addition to these results, we also observed patterns of decreasing RNA-seq read depths across entire intron lengths and note that both canonical co-transcriptional splicing and stochastic recursive splicing using many random splice sites could produce this pattern. Thus, we infer canonical co-transcriptional splicing and/or stochastic recursive splicing — but not widespread recursive splicing at conserved sites — manage the removal of long introns.

## ACKNOWLEDGMENTS

Thanks to the members of my dissertation committee K Hoke, D Sloan, and J Hansen for helpful discussion and feedback throughout the project. Additional thanks to R Denton co-author of chapter 1. B Smith and T Raad co-authors of chapter 2. Thanks to D Levy for *S. tropicalis* tissues. Thank you to M Itgen for *P. montanus* and *P. cinereus* tissues. Thanks to W Zhou, G Liang, PL Molloyd, and PA Jones for providing their full dataset for analysis. Thank you to the *Ambystoma* Genetic Stock Center, funded by NIH Grant P40-OD019794. Electron microscopy was done at the University of Colorado, Boulder EM Services Core Facility in the Molecular, Cellular and Developmental Dept, with the technical assistance of facility staff. 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. Finally, my sincerest thanks to R Mueller, co-author of all sections, advisor, mentor, confidant, and general Rupert Giles to this often stressed and tired graduate student.

## DEDICATION

To my partner in all that I do Crystal, for her unwavering support and love and for taking a chance on a delivery boy all those years ago. To my daughter Hazel, you are the best gift your mother and I could have ever asked for, may the world bring you the same feeling of wonder it has always brought me.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	v
DEDICATION .....	vi
Introduction.....	1
Chapter 1 - Gigantic Genomes of Salamanders Indicate Body Temperature, not Genome Size, is the Driver of Global Methylation and 5-Methylcytosine Deamination in Vertebrates.....	4
Materials and Methods.....	6
Results.....	12
Discussion.....	16
References.....	19
Chapter 2 - Gigantic Animal Cells Suggest Organellar Scaling Mechanisms Across a 50-fold Range in Cell Volume.....	23
Materials and Methods.....	26
Results.....	30
Discussion.....	35
References.....	42
Chapter 3 - Long Introns from the Gigantic Lungfish Genome are Recursively Spliced at Levels Comparable to the Human Genome .....	49
Materials and Methods.....	51
Results.....	54
Discussion.....	57
References.....	61
Conclusion.....	64

## INTRODUCTION

There are currently over 70,000 species of described vertebrates, with the number continuously expanding (Pough et al., 2022); across these extant species, we see genome size spanning from 0.34Gb in the puffer fish *Sphoeroides spengleri* at the low end to 132.83Gb in the lungfish *Protopterus aethiopicus* at the high end (Gregory, 2023; Noleto et al., 2009; Pedersen, 1971), an almost 400-fold difference in genome size in vertebrates alone. However, of these 70,000 species, only 813 species of salamanders and 6 species of lungfish encompass the far upper edge of this range (Amphibiaweb, 2023; Brownstein et al., 2023). These species combined span from 9.3Gb in the salamander *Thorius spilogaster* at the low end (Decena-Segarra et al., 2020) to the 132.83 Gb lungfish. These so-called “genomic giants” encompass just ~1% of known vertebrate species, but they span ~90% of the known vertebrate genome size range; the remaining 99% of species only encapsulate ~10% of the full genome size range.

A major part of this genomic expansion is from an accumulation of transposable elements (TEs) (Canapa et al., 2016; Dufresne and Jeffery, 2011; Sotero-Caio et al., 2017). These pieces of parasitic DNA proliferate within the genome and need to be silenced to prevent adulteration of the host’s genetic information (Bourque et al., 2018). This problem exists in genomic giants such as salamanders and lungfish arguably to a greater extent because of the overwhelming amount of TEs that make up the majority of their genomes (Decena-Segarra et al., 2020; Gregory, 2023; Sclavi and Herrick, 2019; Šímová and Herben, 2012; Sun et al., 2012). TE silencing mechanisms including the piRNA pathway and NuRD complex have been explored in these species; both were shown to be active, suggesting that genome expansion occurs despite intact TE silencing pathways (Carducci et al., 2021; Madison-Villar et al., 2016; Wang et al., 2023). However,

epigenetic modification of the DNA itself in gigantic genomes had yet to be investigated. Utilizing luminometric methylation assays, we test the hypothesis that levels of DNA methylation are different in gigantic genomes than they are in genomes of more typical size (Karimi et al., 2006). More specifically, this analysis tests if levels of transcriptional silencing using cytosine methylation at CpG dinucleotides are akin to other species or if they are faulty in gigantic genomes, potentially revealing a cause of genomic expansion.

As genome expansion has occurred, cell size has also increased in-kind (Gregory, 2001, 2005a; Mueller, 2015). As cell size increases, previous work focused on unicellular and mitotic systems has shown that organelles increase in size, length or number (Chan and Marshall, 2010; Marshall, 2015, 2020; Okie, 2013). We test the hypothesis that the nucleus, mitochondria, endoplasmic reticulum, and Golgi apparatus scale with cell volume in multicellular animals that have evolved huge differences in cell size. Amphibians grant us the opportunity to test this hypothesis because the clade includes species with small and gigantic genomes and cells. We use stereology and electron microscopy to assess the cellular ultrastructure of three species that differ 50-fold in cell size. More specifically, we estimate organelle volume and surface area per unit of cell volume and infer how evolutionary changes in cell size have influenced these structures and their functions (Howard and Reed, 2004; Russ and Dehoff, 2012). We interpret these results in light of differences in metabolic rate and intracellular distance (Chan and Marshall, 2010; Szarski, 1983), both of which we hypothesized would impact functional demands of the cell and organism on the organelles. We draw conclusions about evolved differences in protein synthesis that accompany evolutionary increase in cell size.

Lastly, we chose to explore the other major factor that contributes to genome gigantism, intronic expansion (Nowoshilow et al., 2018; Wang et al., 2021). Long introns present almost the

opposite problem from TE silencing; instead of asking how cells silence long stretches of DNA, we now raise the question of how cells handle exceptionally long introns as part of mRNA transcription and protein synthesis. Cells with smaller genomes have been shown to utilize a form of non-canonical splicing called recursive splicing for this exact purpose, removing long introns in multiple pieces as oppose to a single large piece (Burnette et al., 2005; Moon and Zhao, 2022; Sibley et al., 2015). Removing these introns in multiple steps has been proposed to reduce splicing noise and increase accuracy of splicing overall (Pai et al., 2018; Sibley et al., 2015). Using publicly available total RNA-seq data and the annotated genome from the lungfish *Protopterus annectens* (~40 Gb), we apply a novel pipeline to identify recursive splicing in the 200 longest genes — where most if not all introns are comparable to the longest introns in previously studied model organisms. In this way, we test the hypothesis that levels of recursive splicing in gigantic genomes have increased to handle the intronic load.

This thesis explores these 3 questions in organisms that have evolved gigantic genomes and cells. Together, these results lead to a better understanding of how transposon silencing, cellular structure, and RNA transcription occur in these unusual organisms. In line with that pursuit, I hope that it sheds light on these foundational processes and helps us to better understand how they function, not just on the far end of the genome size gradient, but in all cells in general.

# GIGANTIC GENOMES OF SALAMANDERS INDICATE BODY TEMPERATURE, NOT GENOME SIZE, IS THE DRIVER, OF GLOBAL METHYLATION AND 5-METHYLCYTOSINE DEAMINATION IN VERTEBRATES

## Introduction

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 elements (TEs)—sequences capable of proliferation and movement throughout genomes—are 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; Deniz et al., 2019; Fedoroff, 2012; Law and Jacobsen, 2010; Venkatesh and Workman, 2015). 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 (Jones and Wolffe, 1999; Wolffe, 1998).

There have been few attempts to explore global methylation at the upper limits of genome size. In *Picea abies*, the 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 frequency as TE loads and associated methylation levels increase. Whether this relationship holds for animals with gigantic 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 gigantic genomes because it includes diploid genomes that range from 9 to 120 Gb (Decena-Segarra et al., 2020; Gregory, 2023; Sclavi and Herrick, 2019; Šímová and Herben, 2012; Sun et al., 2012). 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 ~200 mya (Batistoni et al., 1995; Liedtke et al., 2018; Sun and Mueller, 2014). 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, 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 (Bernardi, 2007; Bucciarelli et al., 2009; Jabbari et al., 1997; Sved and Bird, 1990; Varriale and Bernardi, 2006). We use 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 with large genome size. The integration of the data from salamanders — which combine huge 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**

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*, widely used for studies of development and regeneration (Keinath et al., 2015; Seifert et al., 2012; Sessions and Wake, 2020; Tank et al., 1976). 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 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 gigantic genomes. For this analysis, we used muscle tissue as the single representative tissue, and we chose species that span the upper range of. 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 ~22 Gb (Gregory, 2023), although our measurements indicate a genome size of 29.3 Gb (Itgen et al., 2022). The genome size of *Necturus beyeri* is unknown, but size estimates for congeners are 120 Gb (*N. lewisi* and *N. punctatus*) and ~85 Gb (*N. maculosus*) (Gregory, 2023). *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 1.1).

**Table 1.1** Genome sizes for salamanders sampled for global methylation

Species	Genome Size (Gb)
<i>Plethodon cinereus</i>	29
<i>Ambystoma jeffersonianum</i>	29
<i>Ambystoma laterale</i>	29
<i>Ambystoma mexicanum</i>	32
<i>Necturus beyeri</i> <sup>1</sup>	85 - 120

<sup>1</sup>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*) diploid genome sizes and ploidy in salamanders

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. *Necturus beyeri* were obtained commercially through a private vendor in November of 2019. *Ambystoma mexicanum* were obtained from the Ambystoma Genetic Stock Center in July of 2019. *Ambystoma jeffersonianum* and *Ambystoma laterale* were field collected from northern Kentucky (Kenton County) and Connecticut (Litchfield County). Polyploid taxa were collected from New Jersey (Morris County), Kentucky (Kenton County), and Ohio (Crawford County). All field collections took place in the spring of 2019 and 2020.

Specimens of *A. mexicanum*, *P. cinereus*, and *N. beyeri* were euthanized in MS222 and tissues were dissected, and flash frozen in liquid nitrogen. DNA was extracted using a DNeasy Blood and Tissue extraction kit (Qiagen) following manufacturer instruction, including the optional RNase treatment step. DNA concentration and quality were measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific). Tissue samples were stored at -80°C until use. For the polyploid *Ambystoma* samples, clippings of 1-3 mm of tissue from each animal's tail were collected and stored immediately in 90% ethanol, then stored at -20°C until use. 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 DNA concentration. DNA concentration was measured using a Qubit 3 fluorometer (ThermoFisher Scientific) with the dsDNA BR (Broad Range) assay. 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 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.

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 done 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.

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  $\geq 29$  Gb vs. genome size  $\leq 6.4$  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 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.

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* (~15 Gb; SRA046120.1), *Pleurodeles waltl* (~25 Gb; DRX131369), *Eurycea tynerensis* (~25 Gb; SRA046121.1), *Batrachoseps nigriventris* (~26 Gb; SRA046116.1), *Ambystoma mexicanum* (~30 Gb; ASM291563v2), *Bolitoglossa occidentalis* (~43.5 Gb; SRA046118.1), *Aneides flavipunctatus* (~45 Gb; SRA046114.1), *Bolitoglossa rostrata* (~48 Gb; SRA046119.1), and *Cryptobranchus alleganesis* (~55 Gb; SRA073787) (Gregory, 2023). *Ambystoma mexicanum* has a well-assembled genome

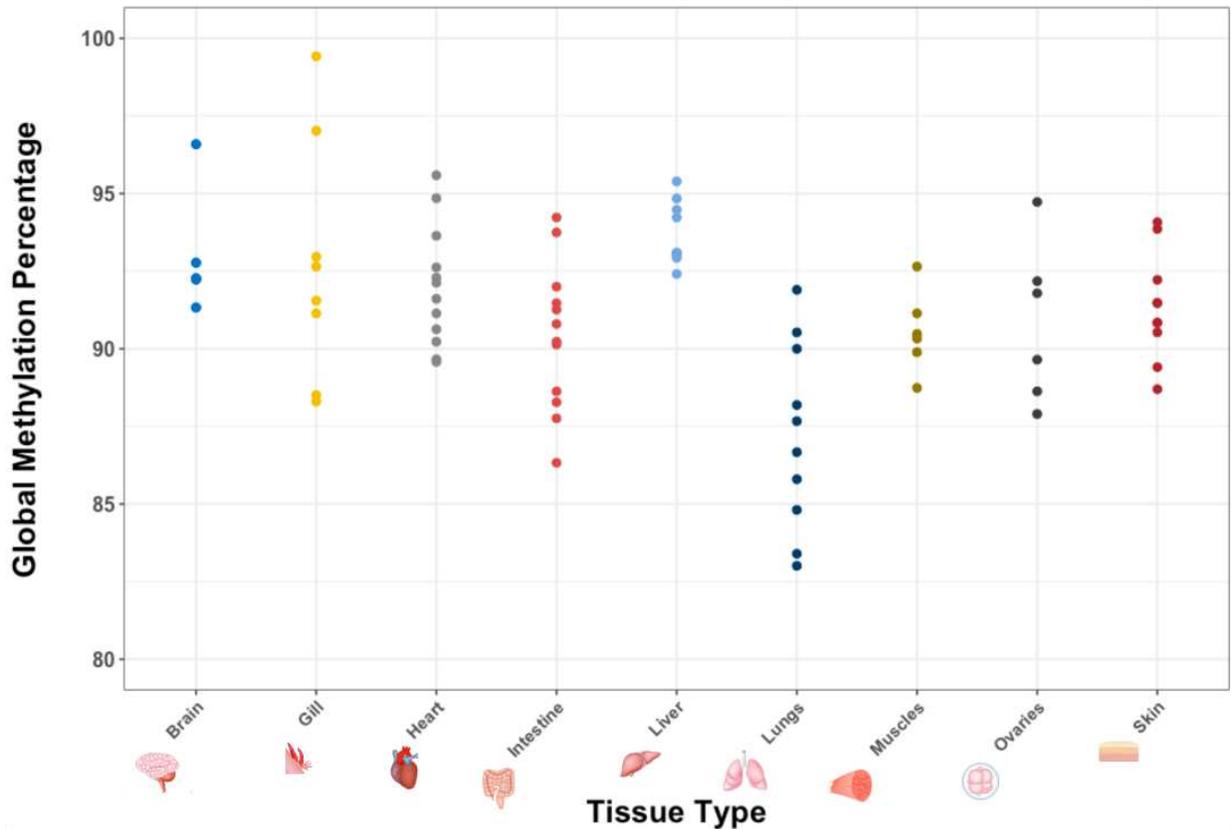
(including the repetitive regions) based on deep, short-read coverage, long reads, and optical mapping (Nowoshilow et al., 2018). *Pleurodeles waltl* has a short-read Illumina assembly, with the repeat elements constructed through a majority vote k-mer extension algorithm (Elewa et al., 2017); however, we used the unassembled Illumina HiSeq 2000 trimmed and quality filtered reads (~0.25X 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 about 0.1% - 1% of the total genome (Sun and Mueller, 2014; Sun et al., 2012). Each dataset was run through a pipeline bash script that removed tags and counted the numbers of each individual nucleotide and CpG dinucleotides 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 =  $(\text{CpG}/N)/((C \times G)/(N^2))$  (Zhou et al., 2020).

We tested for a relationship between genome size and CpG O/E dinucleotides among the nine salamander species using a 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 R packages ape (Paradis and Schliep, 2019), Geiger (Harmon et al., 2008), nlme (Pinheiro et al., 2020), and phylools (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-geome estimates), suggesting that comparisons across these different datasets are unlikely to introduce bias due to comparing whole-genome assemblies to unassembled read datasets.

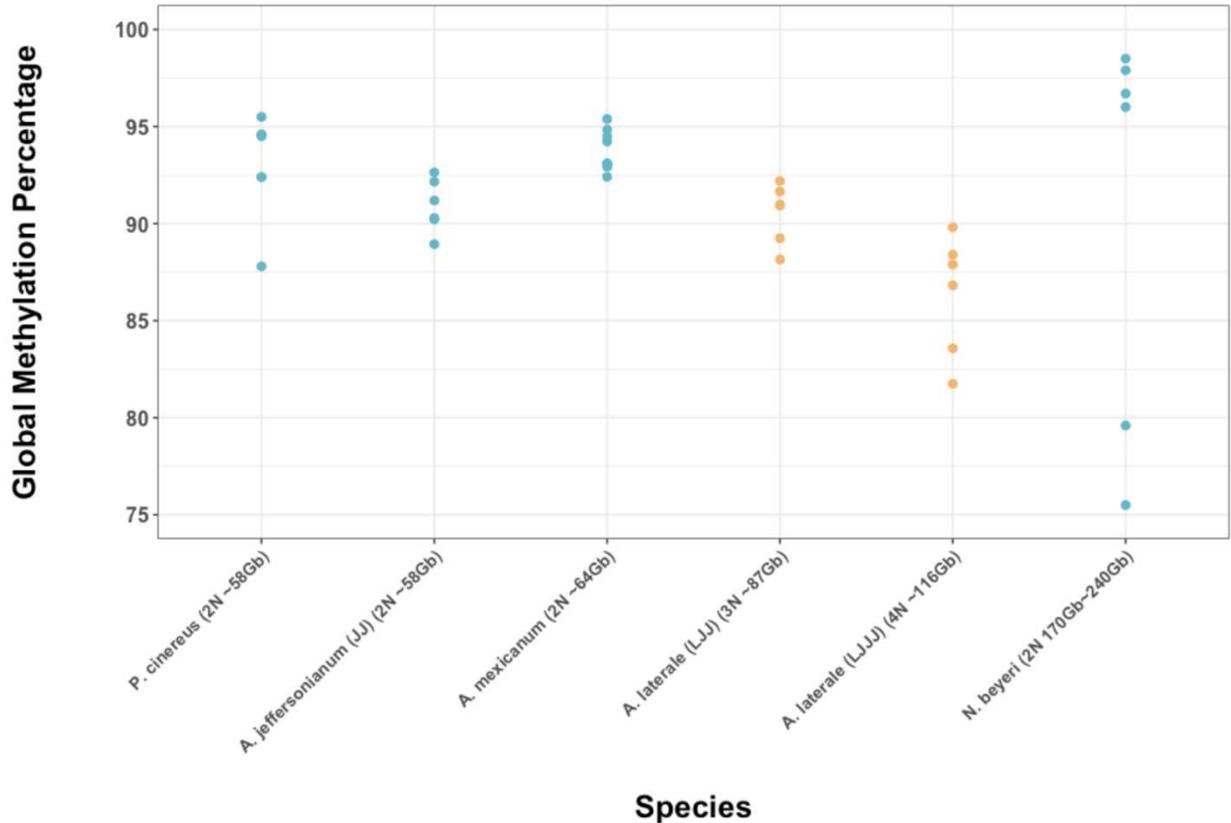
## **Results**

Methylation levels significantly differ between some 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) (Figure 1.1). Lungs were not used for downstream across-species comparisons.



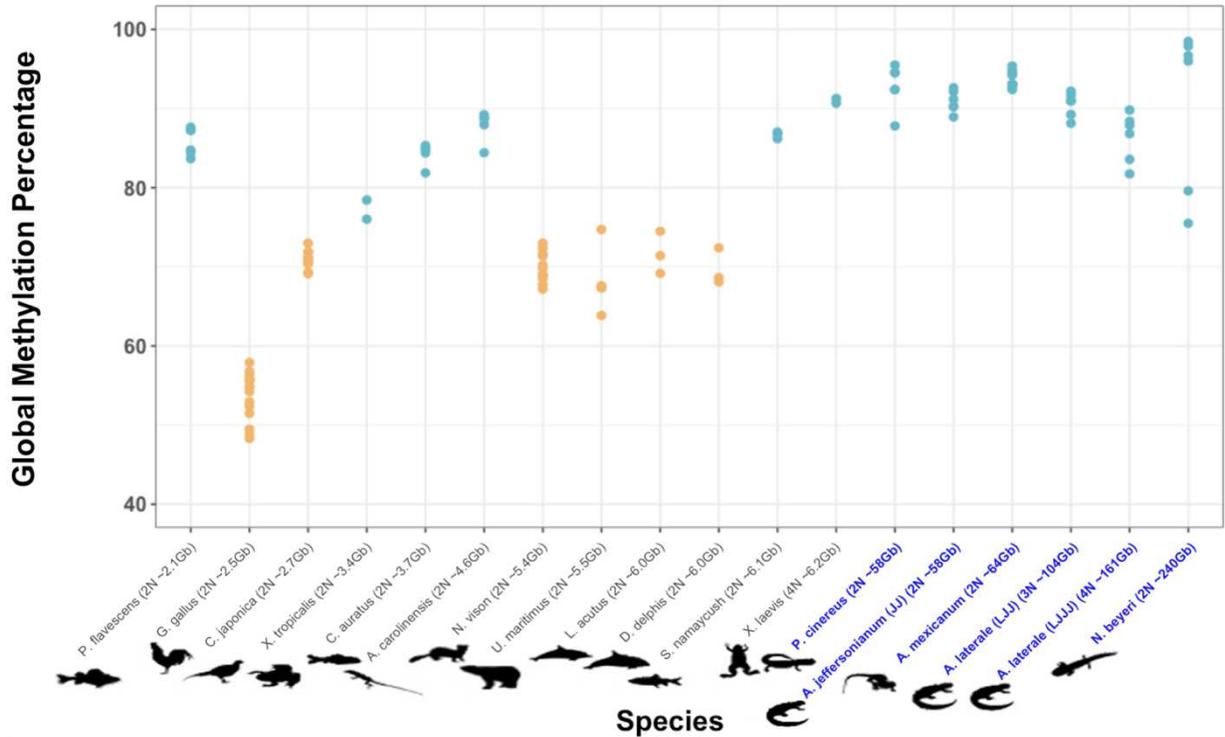
**Fig. 1.1** Methylation levels (percentage methylated cytosines at CpG dinucleotide sites) of DNA extracted from different tissues in *Ambystoma mexicanum*. Lungs have significantly lower methylation levels compared to some tissues: Brain – Lungs, Heart – Lungs, Gill – Lungs, and Liver – Lungs.

Methylation levels are not significantly different between salamander species, despite differences in diploid genome size and ploidy ( $df = 5$ ,  $F\text{-value} = 1.11$ ,  $p = 0.4$ ; Figure 1.2). 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.



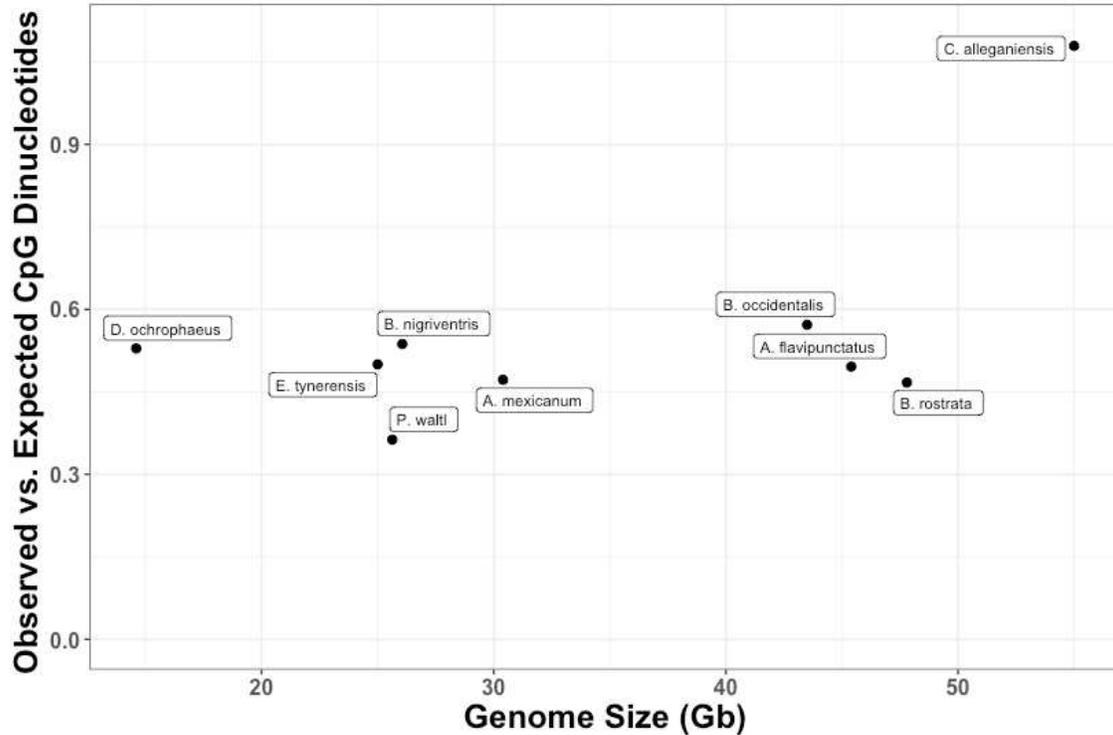
**Fig. 1.2** Methylation levels (percentage of methylated cytosines at CpG dinucleotide sites) of six different species/biotypes of salamanders with different amounts of nuclear DNA. Teal represents diploid species and orange represents polyploid species. There are no significant differences among groups.

Ectotherms as a whole (including salamanders) have significantly higher levels of methylation than endotherms ( $df = 108$ ,  $t\text{-ratio} = -15.84$ ,  $p < 0.001$ , Figure 1.3). Salamanders themselves have methylation levels that are not significantly different from other ectotherms ( $df = 108$ ,  $t\text{-ratio} = 0.88$ ,  $p = 0.38$ ), despite their enormous genome sizes. Excluding salamanders, there is still a significant difference between ectotherms and endotherms ( $df = 108$ ,  $t\text{-ratio} = -9.47$ ,  $p < 0.001$ ). Polyploidy (seen in *X. laevis* and the two uniexual *Ambystoma* biotypes) does not significantly affect methylation levels ( $df = 108$ ,  $t\text{-ratio} = 0.38$ ,  $p = 0.71$ ), 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 (Hughes and Hughes, 1993; Tymowska and Fischberg, 1973), whereas polyploidization happens anew each generation in *Ambystoma*.



**Fig. 1.3** Methylation levels (percentage of methylated cytosines at CpG dinucleotide sites) across vertebrates. Diploid, triploid, or tetraploid genome sizes are listed with species names. Orange dots indicate endothermy and teal dots indicate ectothermy. Blue font indicates salamanders. Ectotherms have higher methylation levels than endotherms ( $p < 0.001$ ).

The observed versus expected ratios of CpG dinucleotides (CpG O/E) for nine species of salamanders range from 0.47 – 1.08. There is no significant relationship between genome size and CpG O/E dinucleotides across salamanders, despite a ~3-fold difference in genome size (Figure 1.4). We note that *C. alleganiensis* is an 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 size.



**Fig.1. 4** The observed versus expected ratios of CpG dinucleotides (O/E) for nine species of salamanders. There is no significant correlation between genome size and O/E CpG ( $p = 0.15$ )

## Discussion

Despite spanning a ~100-fold difference in genome size, the methylation levels of ectotherms were not significantly different from one another. Additionally, the methylation levels of polyploids were not significantly different from diploids. In contrast, endotherms had significantly lower levels of methylation compared to 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 proposed as the primary determinants of global methylation levels in vertebrates before (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

paralleled 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 (Bernardi, 2007; Bucciarelli et al., 2009; Jabbari et al., 1997; Varriale and Bernardi, 2006). 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 – 2.86 Gb. In contrast, the 5 bird and 88 mammal species sampled by Zhou et al., with genome sizes ranging from 1.06 – 4.78 Gb, have lower CpG O/E, as predicted by their endothermic body temperature regulation. We re-analyzed 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 also tend to have larger genomes than fish, reptiles, and frogs—but it is endothermy itself, rather than genome size, that is driving the CpG O/E pattern (Supplemental Figures 1 and 2). Within the non-salamander ectotherms alone, however, genome size is 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 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 non-salamander vertebrates.

## References

- 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.*
- 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, D. L. Mager, and C. Feschotte. 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. *Nature Reviews Genetics.* 20(7):417-31.
- 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, Y. Abbas, I. Abdullayev, R. Sandberg, M. Grabherr, B. Andersson, and A. Simon. 2017. Reading and editing the *Pleurodeles waltl* genome reveals novel features of tetrapod regeneration. *Nat. Commun.* 8(1):2286.
- Fedoroff, N. V. 2012. Transposable elements, epigenetics, and genome evolution. *Science* 338:758–767.
- Gregory, T. R. 2023. Animal Genome Size Database. <http://www.genomesize.com>
- Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2008. GEIGER: Investigating evolutionary radiations. *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. Pais De Barros, J. Desgrès, and G. Bernardi. 1997. Evolutionary changes in CpG and methylation levels in the genome 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. *Sci. Rep* 5:16413
- Law, J. A., and S. E. Jacobsen. 2010. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Essays Biochem* 63:743-755

- 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.* 5(53):2445.
- 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, F. Falcon, D. Knapp, S. Powell, A. Cruz, H. Cao, B. Habermann, M. Hiller, E. M. Tanaka, and E. W. Myers. 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 models 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. U. S. A.* 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 the 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. USA* 117:19359–19366.

# GIGANTIC ANIMAL CELLS SUGGEST ORGANELLAR SCALING MECHANISMS ACROSS A 50-FOLD RANGE IN CELL VOLUME

## **Introduction**

Cell size shows spectacular diversity across the Tree of Life, with unicellular organisms – where cell size is body size – varying up to a 1,000,000-fold in volume (Mueller 2015; Smith 2017; Malerba and Marshall 2021). In multicellular organisms – where cell size shapes the basic building blocks of tissues and organs – cells show size variation of around 1,000-fold (Gregory 2001, 2005b). As cell size changes, assuming shape remains constant, the surface area to volume ratio (SA:V) scales in predictable ways dictated by the cell's shape. Previous studies of small cells have shown how this scaling of SA:V impacts many aspects of a cell's biology including movement across the cell membrane, metabolic flux, the reach of microtubules, and nutrient exchange (Marshall et al. 2012; Marshall 2020).

As cell size increases, organelles must scale to maintain overall cell functionality (Reber and Goehring 2015). Within the natural range of cell size, scaling among smaller cells has been shown to happen in different ways: by increasing in volume, increasing of linear network length, or increasing in copy number (Chan and Marshall 2010). Specific examples include organelles like the nucleus and nucleolus increasing in volume (Noel et al. 1971; Jorgensen et al. 2007), while mitotic spindles and mitochondria increase in network length (Hara and Kimura 2009; Rafelski et al. 2012) and multiple-unit organelles like peroxisomes increase in copy number (Titorenko et al. 2000). Other organelles of more complex shape like the endoplasmic reticulum (ER) and Golgi apparatus have been the focus of less research, and their scaling patterns are less clear (Chan and Marshall 2010; Marshall 2015).

As with cells, as organellar volume increases, the surface area scales differently depending on shape; an organelle shaped like a sphere shows a SA:V ratio that scales as  $r^2 : r^3$ , where  $r$  is the radius of the sphere. In contrast, an organelle shaped like a tube maintains a SA:V ratio of 1 : 1 as it grows in length (Chan and Marshall 2010). These two basic structures account for most organelles and reflect their functions; the nucleus is a large space that houses DNA and its associated molecules, and it tends to be more spherical. The mitochondria weave throughout the cell providing energy, and they tend to be more tubular (Chan and Marshall 2010; Marshall 2015). The ER and Golgi apparatus, however, are made up of cisternae, which are flattened membrane vacuoles that share attributes of both structures (Day et al. 2013; Schwarz and Blower 2016).

In plant species and some fungal species, cell size can be heavily impacted by large fluid-filled vacuoles, which can comprise 30% to 90% of the cell volume; additionally, storage plastids also contribute to intracellular volume (Klionsky et al. 1990; O'Connor et al. 2010). The size of these vacuoles can be highly variable and can cause overall cell size to shift dramatically (Chan and Marshall 2014; Tan et al. 2019); although the cells are enclosed by a cell wall, the walls themselves have the ability to expand and contract rapidly to accommodate changes in volume (Cosgrove 1993; Marshall et al. 2012).

In contrast, animal cells do not contain these storage organelles and plastids (Wise 2007; O'Connor et al. 2010), meaning their cell volume is almost exclusively derived from cytosol and membrane-bound organelles used for intra- and intercellular functions other than storage. In addition, animal cells closely regulate their osmotic gradient and maintain tight control on their cell size (Alberts et al. 2017; Cooper and Adams 2022). Thus, animal cells do not undergo rapid size change outside of growth and size reduction associated with the cell cycle, remaining at a

more stable equilibrium size than plant cells. Two animal groups are composed of somatic cells at the farthest end of this size spectrum — salamanders and lungfishes. Salamanders are one of the three extant clades of amphibians, which include 8,653 species (Amphibiaweb 2023) with genome sizes ranging from 0.95 Gb -120 Gb; because genome size and cell size are correlated, the clade also exhibits a huge range of cell sizes (Gregory 2023).

Amphibians have been used for decades as a model taxon to examine how cell size affects an organism at the tissue and organismal levels (Hanken 1983; Roth et al. 1994; Womack et al. 2019; Decena-Segarra et al. 2020; Miller et al. 2020; Itgen et al. 2022). We continue to leverage this natural diversity of cell size across extant amphibians to quantify how organelles scale across a ~50-fold difference in cell volume using transmission electron microscopy (TEM) and stereology (Howard and Reed 2004; Russ and Dehoff 2012; Winey et al. 2014). We focus our analyses on the nucleus, mitochondria, endoplasmic reticulum (ER), and Golgi apparatus. The nucleus is the primary site for DNA and RNA synthesis, as well as the cell's central hub for detecting deformations in cell shape (Venturini et al. 2020). Mitochondria are a network of membrane-bound organelles whose functions include ATP synthesis, apoptosis, and signaling (Brand et al. 2013). The ER is a network of membranous sacs and tubules that functions in lipid synthesis and protein synthesis, modification, and sorting (Schwarz and Blower 2016). The Golgi apparatus is a collection of stacked membrane-bound organelles that functions in protein sorting and trafficking and carbohydrate and lipid synthesis (Day et al. 2013). Larger cell size increases intracellular trafficking distance, as well as increasing per-cell demand for ATP, transcript and protein production, and lipid and carbohydrate synthesis (Guo and Fang 2014); thus, functional demands on all of these organelles are likely impacted by evolutionary increases

in cell size. Taken together, our results reveal how organelles scale to maintain cell functionality at the extremes of animal cell size.

## **Materials and Methods**

Intestinal tissue was chosen for analysis as it is made up of only four cell types, and 80 percent of the total cell population is enterocytes, resulting in a relatively homogenous population of cells (De Santa Barbara et al. 2003). Three species of amphibians were chosen that span much of the range of amphibian genome and cell sizes: the western clawed frog *Silurana tropicalis* (genome size = 1.2 Gb), the northern gray-cheeked salamander *Plethodon montanus* (genome size = 35 Gb), and the western waterdog *Necturus beyeri* (genome size ~100 Gb based on congeners that range from 80.5-120.6 Gb). *Silurana tropicalis* were obtained from a lab-reared colony following standard husbandry conditions and *Necturus beyeri* were obtained commercially. *Plethodon montanus* were field collected between May and August of 2018 in Avery County, North Carolina under the wildlife collection license # 18-SC01250 issued by the North Carolina Wildlife Resources Commission. One individual was sampled per species, and all specimens were euthanized in MS222. Work was carried out in accordance with Colorado State University (*P. montanus*, *N. beyeri*) and University of Wyoming (*S. tropicalis*) IACUC protocols (17-7189A and 20200714DL00443-01, respectively).

Intestinal tissue from each individual was dissected and immersion fixed in 2.5% glutaraldehyde/2% formaldehyde. The tissues then underwent secondary fixation and staining in 1% OsO<sub>4</sub> in a 0.1 M cacodylate buffered solution followed by embedding in PELCO Eponate 12 epoxy (Cushing et al. 2014). Thin sections (60-80 nm) of resin-embedded samples were cut

using a Leica UCT ultramicrotome, collected onto Formvar-coated TEM slot grids, and poststained with 2% aqueous uranyl acetate followed by Reynold's lead citrate.

Sample preparation, fixation, and mounting were done at Colorado State University. The samples were then sectioned, stained, and imaged at the University of Colorado, Boulder Electron Microscopy Services Core Facility. Sections were imaged using a Tecnai T12 Spirit transmission electron microscope, operating at 100 kV, with an AMT CCD digital camera. *Silurana tropicalis* and *P. montanus* tissues were imaged at 9,300x direct magnification. *Necturus beyeri* tissue was imaged at 6,800x direct magnification because the larger cell sizes could not be captured at the higher magnification. All images were evaluated for quality, ensuring intact tissues undamaged by the fixation process.

Stereology uses 2-dimensional image sampling protocols that allow the estimation of surface area and volume of 3-dimensional shapes through unbiased sampling using grids (Howard and Reed 2004; Russ and Dehoff 2012). Grids are superimposed randomly onto preselected, non-overlapping TEM images. Depending on the type of probes used in the grid and how the user measures the probes' interactions with the images, the volume, surface area, length, and number of 3-dimensional structures can be estimated from the 2-D TEM images.

For the individual representing each species, 60 TEM images were randomly selected, and each image was overlaid with a grid of probes using the IMOD image processing program (Kremer et al. 1996). Volume fraction of organelles per cell (point probes), as well as surface area of organelles per unit of cell volume (line probes), were measured for nucleus, mitochondria, ER, and Golgi apparatus using the 3dmod stereology plugin in IMOD (Noske 2010).

For measuring volume fractions of the organelles (i.e. the volume of each type of organelle per total cell volume), each of the 60 images for each individual was fitted with a 7 x 7 grid using crosshair probes (points) for a total of 2,940 probes per species. Next, each crosshair was visually identified and manually assigned to one of the following categories: nucleus, mitochondria, ER, Golgi apparatus, or cytoplasm (which included cytosol and other non-focal organelles). If a crosshair fell on a part of the image that included damaged cellular material or extracellular material, it was categorized as “not in bounds” and was removed from the data set. The center of the crosshair was used to define the object category, and only one class was permitted per point (Howard and Reed 2004; Russ and Dehoff 2012). A denser grid was used for the Golgi apparatus due to the rarity of the organelle; the same 60 images per individual were overlaid with 18 x 18 grids for a total of 19,440 crosshair probes.

For measuring organelle and cell surface area densities (i.e. surface area per unit of cell volume), the same 60 images per individual were used as for volume fraction estimates. The nucleus, mitochondria, ER, and plasma membrane were measured simultaneously using alternating cycloid probes in a 2 x 2 grid per image for a total of 240 probes. The Golgi apparatus was again measured using a denser grid due to the organelle’s scarcity, using a 6 x 6 grid of alternating cycloids on each image for a total of 2,160 cycloids per species. The cycloids were visually identified and manually assigned as being “in bounds” and “not in bounds” as in the volume fraction estimation. From there, cycloids were marked with intercepts each time a cycloid interacted with the boundary of an organelle or cell; if a cycloid went entirely through an organelle, then an intercept was marked both entering and leaving the structure. Each intercept was categorized as one of the four groups of organelles from the volume fraction estimation or as the cell boundary; cytoplasm was not used in surface area density estimation. Cycloids that were

fully inside an organelle and did not contact the outside of the organelle were not marked with intercepts. The surface area density estimates were then calculated from these results with the equation  $2 * \text{Intercepts} / \text{Total Length}$  of all “In Bound” Cycloids (Howard and Reed 2004; Russ and Dehoff 2012).

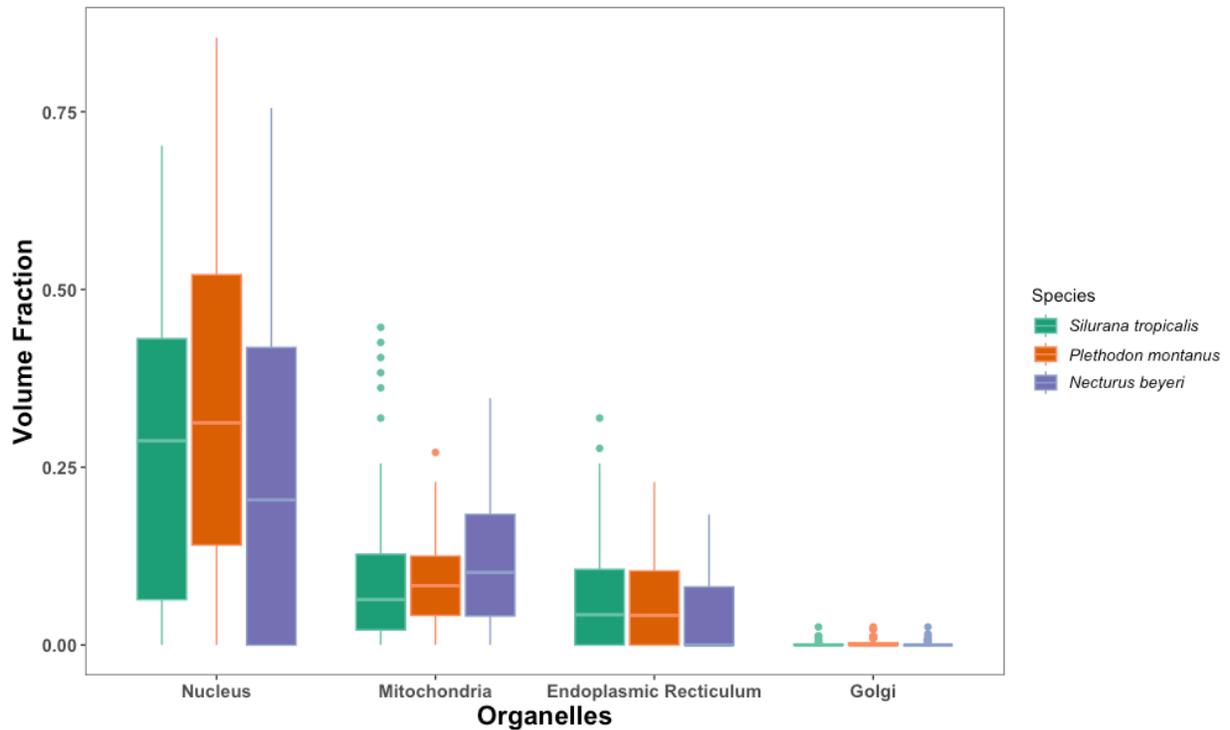
Organelle and whole cell SA:V ratios for each species were calculated by dividing the estimates for surface area density by the volume fraction estimates for each image. Because organellar structures are so variable in shape (e.g. the nucleus and mitochondria exhibiting spherical vs. tubular network shapes) (McCarron et al. 2013; Malerba and Marshall 2021), their SA:V ratios scale in different ways with increases in organellar size (Chan and Marshall 2010; Marshall 2020). Thus, SA:V estimates across species act as a proxy for organelle shape.

Nuclear volume was estimated using the Nucleator probe (Gundersen et al. 1988) in the Visiopharm VIS stereology software (version 2017.7). The Nucleator randomly assigns two perpendicular rays that radiate outward from a fixed point in the nucleus, defined to be the nucleolus, and uses these rays and their intersection points with the nuclear membrane to estimate mean nuclear volume. 100 nuclei were analyzed from each species. Then, using the proportional estimates of nuclear volume obtained from IMOD (Kremer et al. 1996), cell volume and surface area, and all other organelle volumes and surface areas, were extrapolated from nuclear volume.

We tested for associations between cell size (i.e. species) and 1) organelle volume fraction, 2) organelle surface area density, and 3) organelle surface area to volume ratio using ANOVA, followed by a pairwise Tukey post-hoc method of comparison. We carried out all analyses in R Studio (RStudio Team 2021; R Core Team 2021) using R packages emmeans (Lenth 2021), car (Fox and Weisberg 2011) and lme4 (Bates et al. 2015).

## Results

Across the three species, the cytoplasm made up the largest fraction of cell volume (0.48 – 0.57), followed by the nucleus (0.27 – 0.36), mitochondria (0.09 – 0.12), ER (0.04 – 0.07), and Golgi apparatus (0.002 – 0.003) (Figure 2.1). None of the four organelles showed any statistically significant differences in volume fraction across species based on the ANOVA ( $p > 0.05$ ). The cytoplasm volume fraction was statistically significantly different between *S. tropicalis* and *P. montanus* (ANOVA  $p = 0.002$ ; Tukey  $p = 0.002$ ). All mean volume fractions and standard errors are listed in Table 2.1. All  $p$ -values are listed in Table 2.2.



**Fig. 2.1:** Volume fraction of each organelle. Species are shown left-to-right from smallest to largest cell sizes.

**Table 2.1:** Measurements of the cell and organelles (\* Indicates significant difference  $p < 0.05$ )

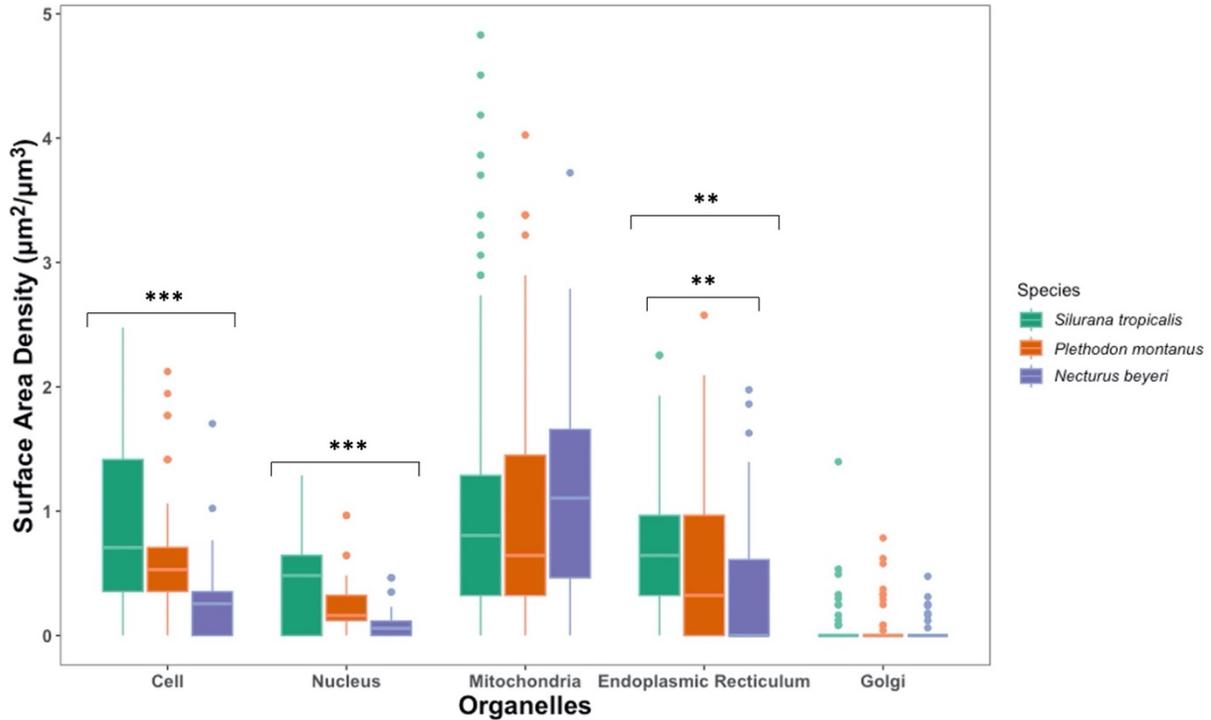
Measurement	Component	<i>Silurana tropicalis</i> Mean (SE)	<i>Plethodon montanus</i> Mean (SE)	<i>Necturus beyeri</i> Mean (SE)
<b>Volume Fraction</b>	Cytoplasm	<b>0.57*</b> (0.02)	<b>0.49*</b> (0.02)	0.56 (0.02)
	Nucleus	0.27 (0.03)	0.36 (0.03)	0.27 (0.03)
	Mitochondria	0.10 (0.01)	0.09 (0.01)	0.12 (0.01)
	Endoplasmic Reticulum	0.06 (0.008)	0.07 (0.008)	0.04 (0.008)
	Golgi Apparatus	0.002 (0.0001)	0.003 (0.0001)	0.002 (0.0001)
<b>Surface Area Density (<math>\mu\text{m}^2/\mu\text{m}^3</math>)</b>	Cell	<b>0.93*</b> (0.06)	<b>0.61*</b> (0.06)	<b>0.27*</b> (0.06)
	Nucleus	<b>0.46*</b> (0.04)	<b>0.25*</b> (0.04)	<b>0.10*</b> (0.04)
	Mitochondria	1.19 (0.05)	1.01 (0.05)	1.15 (0.05)
	Endoplasmic Reticulum	<b>0.71*</b> (0.08)	0.57 (0.08)	<b>0.34*</b> (0.08)
	Golgi Apparatus	0.07 (0.02)	0.07 (0.02)	0.03 (0.02)
<b>Surface Area to Volume Ratio (<math>\mu\text{m}^2/\mu\text{m}^3</math>)</b>	Cell	<b>0.97*</b> (0.07)	<b>0.68*</b> (0.07)	<b>0.31*</b> (0.07)
	Nucleus	<b>1.52*</b> (0.20)	<b>0.83*</b> (0.19)	<b>0.26*</b> (0.19)
	Mitochondria	12.31 (1.8)	11.54 (1.8)	12.57 (1.8)
	Endoplasmic Reticulum	10.05 (1.9)	7.50 (1.9)	4.83 (1.9)

**Table 2.2:** ANOVA p-values for volume fraction, surface area density, and SA:V ratio for the cell, cytoplasm, and organelles across species. All Tukey post-hoc p-values for between-species comparisons. (\* Indicates significant difference)

	<b>Across All Species</b> (p-value)	<i>S. tropicalis</i> – <i>P. montanus</i> (p-value)	<i>P. montanus</i> – <i>N. beyeri</i> (p-value)	<i>S. tropicalis</i> – <i>N. beyeri</i> (p-value)
<b>Cytoplasm Volume Fraction</b>	<b>0.002*</b>	<b>0.002*</b>	0.052	0.478
<b>Cell SA Density</b>	<b>8.23E-11*</b>	<b>0.002*</b>	<b>0.001*</b>	<b>&lt;0.001*</b>
<b>Cell SA:V Ratio</b>	<b>1.04E-08*</b>	<b>0.018*</b>	<b>0.001*</b>	<b>&lt;0.001*</b>
<b>Nucleus Volume Fraction</b>	0.102	0.233	0.110	0.925
<b>Nucleus SA Density</b>	<b>5.14E-10*</b>	<b>0.0002*</b>	<b>0.014*</b>	<b>&lt;0.001*</b>
<b>Nucleus SA:V Ratio</b>	<b>4.34E-05*</b>	<b>0.034*</b>	0.093	<b>&lt;0.001*</b>
<b>Mitochondria Volume Fraction</b>	0.218	0.656	0.188	0.660
<b>Mitochondria SA Density</b>	0.585	0.606	0.688	0.990
<b>Mitochondria SA:V Ratio</b>	0.899	0.921	0.911	0.999
<b>ER Volume Fraction</b>	0.078	0.999	0.129	0.118
<b>ER SA Density</b>	<b>0.007*</b>	0.446	0.124	<b>0.005*</b>
<b>ER SA:V Ratio</b>	0.568	0.580	0.113	0.136
<b>Golgi Volume Fraction</b>	0.481	0.739	0.454	0.893
<b>Golgi SA Density</b>	0.320	1.000	0.388	0.396

All mean surface area densities and standard errors are listed in Table 2.1. Across the three species, the largest organellar surface area density is seen in the mitochondria (1.0 – 1.2  $\mu\text{m}^2/\mu\text{m}^3$ ), followed by the ER (0.3 – 0.7  $\mu\text{m}^2/\mu\text{m}^3$ ), nucleus (0.1 – 0.5  $\mu\text{m}^2/\mu\text{m}^3$ ), and Golgi apparatus (0.03 – 0.07  $\mu\text{m}^2/\mu\text{m}^3$ ) (Figure 2.2). All 3 species show statistically significant differences in nuclear surface area density ( $p < 0.02$ ; Table 2.2), with nuclear surface area density decreasing as cell size increases. The ER also shows a decrease in surface area density as cell

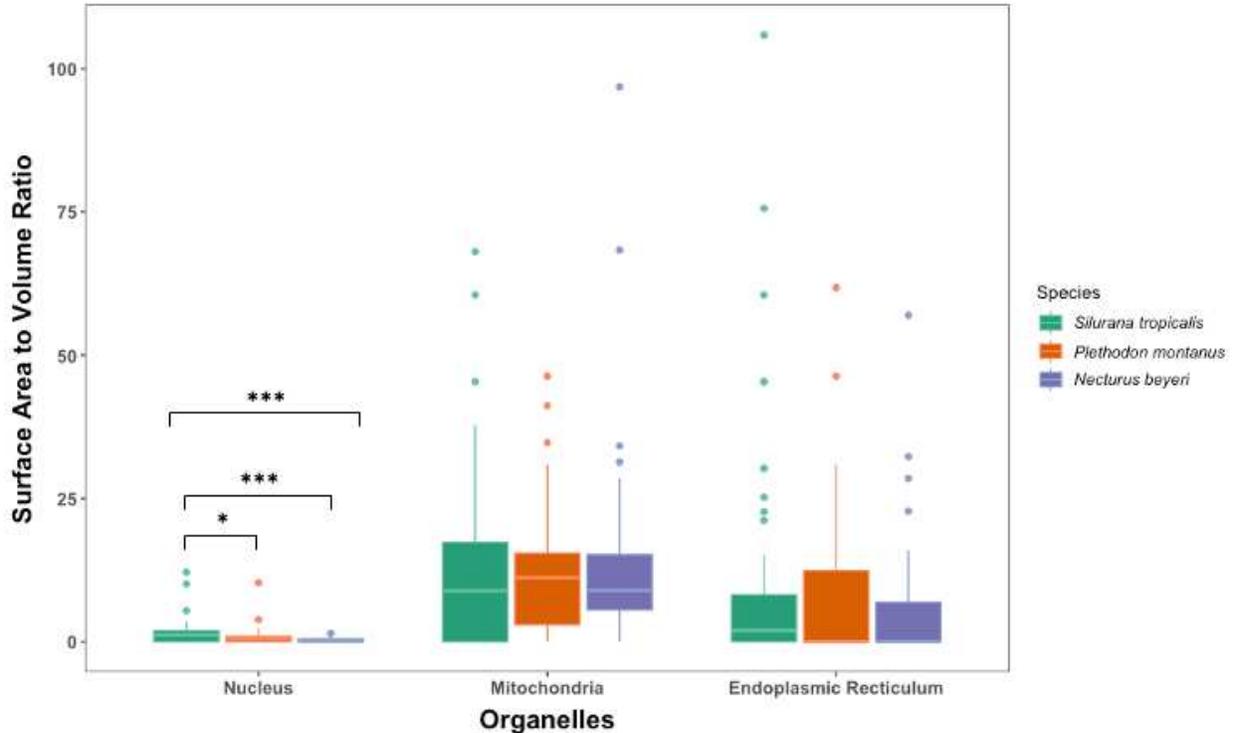
size increases, with a statistically significant difference between *S. tropicalis* and *N. beyeri*, the smallest and largest cells ( $p < 0.007$ ; Table 2.2).



**Fig. 2.2:** Surface area density for each organelle and the overall cell. Species are shown left-to-right from smallest to largest cell sizes. Groups with significant differences are marked with asterisks; the one pairwise comparison with the ER that shows a significant difference in the post-hoc test is indicated further (\*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$ ).

All mean SA:V ratios for the cell and organelles and standard errors are listed in Table 2.2. Across the three species, SA:V ratio of the cell is statistically significantly different ( $p < 0.02$ ; Table 2.1), decreasing from  $0.97 \mu\text{m}^2/\mu\text{m}^3$  in *S. tropicalis* to  $0.31 \mu\text{m}^2/\mu\text{m}^3$  in *N. beyeri*. As expected, the SA:V ratio of the nucleus also decreases from  $1.52 \mu\text{m}^2/\mu\text{m}^3$  to  $0.26 \mu\text{m}^2/\mu\text{m}^3$  as cell size increases and is statistically significantly different between *S. tropicalis* and *P. montanus* as well as between *S. tropicalis* and *N. beyeri* ( $p < 0.034$ ; Table 2.1). In contrast, the mitochondria show no differences in SA:V ratio across species, consistent with the expectation of increasing the length of a tube-like network as cell size increases. The ER also shows no

differences in SA:V ratio across species. Because of its scarcity, the Golgi was absent from many grids, so per-grid SA:V ratios were not calculated (Figure 2.3).



**Fig. 2.3:** Surface Area to Volume ratio of the cell and organelles. Species are shown left-to-right from smallest to largest cell sizes. Groups with significant differences are marked with asterisks, subgroupings that show significant differences in post-hoc tests are indicated further (\* =  $p < 0.05$ , \*\*\* =  $p < 0.001$ )

All estimates of absolute volume and surface area for cells and organelles are listed in Table 2.3. Cell volume in *N. beyeri* is 46 times larger than *S. tropicalis*. The majority of cell volume estimates from diverse taxa in the literature are for erythrocytes; comparison of our enterocyte volumes with erythrocyte volumes for the same or congeneric species reveals that enterocytes are larger (*S. tropicalis*  $228 \mu\text{m}^3$  vs.  $122 \mu\text{m}^3$ ; *P. montanus*  $3,191 \mu\text{m}^3$  vs  $1,476 - 3,204 \mu\text{m}^3$  for congeners *P. cinereus* (20 Gb) and *P. dunni* (47.5Gb) (Gregory 2005a, 2023)), but that our estimates are in keeping with other empirical estimates.

**Table 2.3:** Estimated volumes and surface areas extrapolated from Nucleator estimates of nuclear volume and point count estimates of volume fractions and surface area densities.

	<i>Silurana tropicalis</i>		<i>Plethodon montanus</i>		<i>Necturus beyeri</i>	
	Volume ( $\mu\text{m}^3$ )	Surface Area ( $\mu\text{m}^2$ )	Volume ( $\mu\text{m}^3$ )	Surface Area ( $\mu\text{m}^2$ )	Volume ( $\mu\text{m}^3$ )	Surface Area ( $\mu\text{m}^2$ )
Entire Cell	228.2	212.3	3,190.9	1926.7	10,592.8	2,777.6
Cytoplasm	128.6	N/A	1,525.6	N/A	5,983.9	N/A
Nucleus	62.2	105.0	1,154.4	797.7	2,843.1	1,059.3
Mitochondria	22.6	271.6	292.6	3,222.8	1,301.9	12,181.7
Endoplasmic Reticulum	14.3	162.0	210.6	1,818.8	448.1	3,601.5
Golgi Apparatus	0.52	16.0	7.9	223.4	15.9	317.8

## Discussion

Marshall (2020) laid out five mechanistic models to explain the scaling of organelles with cell size. In order of simplest to most complex, these are 1) *the limiting precursor model*, which posits that large cells have higher biosynthetic capacity and, thus, make more organellar building blocks, which are incorporated into more/larger organelles; 2) *the relative growth model*, which posits that cells and organelles follow their own independent growth trajectories; 3) *the demand-driven model*, which posits that organelles grow or multiply to meet the demands of their cell's size; 4) *the size measurement model*, which posits that the size of the cell is measured, and this measurement is used to set organelle size/number; and 5) *the programmed scaling model*, which posits that cells adjust biosynthesis of organellar building blocks to achieve a match between organelle size/number and target cell size (Marshall 2020). The simplest model that can explain observed patterns in the data is to be favored. Here, we discuss scaling of each organelle and what it suggests about the mechanisms of organellar scaling with cell size.

Our results on scaling of the nucleus across this ~50-fold increase in cell size are consistent with earlier results in two ways: 1) The nucleus scales allometrically with respect to its

shape, showing the decreased SA:V ratio predicted by a sphere increasing in size. 2) The volume of the nucleus scales proportionately with cell size, i.e., the nucleocytoplasmic ratio is maintained, consistent with its apparent functional significance (Balachandra et al. 2022). Mechanistically, the size of the nucleus is set by several interacting forces: DNA content, the state (i.e. compaction) of the chromatin, the total cellular amount of cytoplasmic factors involved in nuclear growth, and overall nucleocytoplasmic transport (Cantwell and Nurse 2019b,a; Heijo et al. 2022; Chen et al. 2023). Across species, the varying contributions of these forces, as well as the differences in the relevant cytoplasmic factors, can be understood in light of a general model for maintenance of the nucleocytoplasmic ratio in which the balance of mechanical forces generated by osmotic pressure within the nucleus and within the cytoplasm sets nuclear volume (Deviri and Safran 2022).

Because of salamanders' enormous cell volumes, the conserved nucleocytoplasmic ratio yields a disproportionately low nuclear surface area relative to both nuclear and cell volumes. In addition, nuclear pore complexes, which mediate the import and export of macromolecules into and out of the nucleus, are sparse across the salamander nuclear envelope relative to their density in other vertebrates (Maul and Deaven 1977). Nonetheless, the conserved nucleocytoplasmic ratio in salamander cells — as well as overall cell functionality — indicates the existence of effective nucleocytoplasmic transport, demonstrating that movement through the nuclear pore complexes is not a rate-limiting factor for salamander cell physiology, despite proportionally low nuclear surface area and low nuclear pore complex density. Overall, the proportionate increase in nuclear volume in these huge cells is consistent with nuclear scaling following a limiting precursor model.

Across the tree of life, across cell sizes that span four orders of magnitude, mitochondrial volume increases proportionately with cell size, comprising roughly 10% of overall cell volume (Okie et al. 2016). Our results are largely consistent with these broad patterns; mitochondrial volume in our representative amphibians ranges from ~9 – 12%. Mitochondrial surface area has also been shown to scale roughly proportionately with cell surface area across a dataset of eukaryotes that includes unicellular organisms, green algae and land plants, and mammals (Lynch and Marinov 2017); our results show somewhat higher mitochondrial surface area per cell surface area than predicted from this broad relationship, but given the sparse sampling and noise inherent in this growing but limited empirical dataset, we do not attach any functional interpretation to this pattern.

In heterotrophic organisms, the increase in overall mitochondrial volume accompanying increased cell volume is primarily accomplished through an increase in mitochondrial number rather than size (Okie et al. 2016). In addition, experimental manipulations in yeast demonstrate that mitochondrial shape stays the same across increasing cell and overall mitochondrial volumes (Seel et al. 2022). Our results showing unchanged mitochondrial SA:V ratio across our model amphibians also suggest isometric growth of the total mitochondrial network with respect to its shape, consistent with these previous results. Overall, our mitochondrial results are broadly consistent with global patterns emerging from limited empirical data across the eukaryotic tree, suggesting that mitochondrial scaling follows general rules that transcend species and cell type.

Despite the presence of general scaling rules suggested by these broad-scale phylogenetic patterns, mitochondrial number is also known to vary at a finer scale among cells, individuals, and species in association with metabolic demand (Schwerzmann et al. 1989; Scott et al. 2018). Salamanders have lower metabolic rates than frogs, and in fact show the lowest metabolic rates

among tetrapod vertebrates (Pough 1980; Gatten et al. 1992; Chong and Mueller 2013). Accordingly, they have long been studied with the goal of connecting metabolic rate to cell physiology (Szarski 1970; Goniakowska 1973). The association between low metabolic rate and huge cells in salamanders prompted the “frugal metabolic strategy” hypothesis that natural selection favors large cell size because it relaxes energetic requirements (Szarski 1983) — in contrast to selection at the other physiological extreme favoring small cell size to facilitate high metabolic output in bats, birds, and pterosaurs (Hughes and Hughes 1995; Organ and Shedlock 2009; Wright et al. 2014; Kapusta et al. 2017). Since then, mechanistic hypotheses connecting large cell size to low metabolic rate have been proposed based on both energy supply (e.g. constraints on intracellular resource transport) and demand (e.g. decreased relative cost of  $\text{Na}^+ - \text{K}^+$  gradient maintenance across the plasma membrane) (Glazier 2022).

Despite solid theoretical predictions for lowered metabolic rates in large cells (Kozłowski et al. 2003), empirical data have both supported and rejected a causal relationship between cell size and metabolic rate in amphibians. For example, studies of *Xenopus* frog embryos that differ in cell size and ploidy show lowered mass-specific metabolic rates because of decreased relative  $\text{Na}^+ - \text{K}^+$  ATPase activity costs in larger cells (Cadart et al. 2023). In contrast, studies of salamanders fail to show a consistent relationship between mass-specific metabolic rate and cell size’s most commonly-used proxy, genome size (Licht and Lowcock 1991; Gregory 2003, 2005a; Uyeda et al. 2017; Gardner et al. 2020; Johnson et al. 2021). Our results reveal no effect of lower metabolic rate — whatever its underlying mechanistic cause — on mitochondrial proportionate volume, surface area, or shape in amphibians, suggesting that differences in metabolic demand across species do not drive cellular mitochondrial content in the

clade. Our results are more consistent with mitochondrial volume increasing proportionately with cell volume, independent of any specific features of cell physiology.

Mitochondrial content and function can be decoupled (Miettinen and Björklund 2016), and salamander mitochondrial oxidative phosphorylation (OXPHOS) genes show evidence of relaxed purifying selection relative to frogs, consistent with lower functional demand on ATP synthesis (Chong and Mueller 2013). Thus, uniform mitochondrial abundance should not be interpreted as uniform capacity for ATP synthesis across amphibians. More generally, mitochondrial fission and fusion events maintain connectivity among the mitochondria to maximize overall function (Miettinen and Björklund 2016), possibly through the transmission of the proton motive force itself from one region of the mitochondrial network to another faster than the corresponding metabolites and oxygen could reach distant mitochondria by diffusion (Glancy et al. 2015; Miettinen and Björklund 2017). Our results are consistent with mitochondrial content scaling proportionate to cell volume following a limiting precursor model, which yields a mitochondrial network that comprises ~10% of the cell's volume. This percentage volume, in turn, produces a spatial distribution of organelles that can undergo fission and fusion events to maintain an effective functional network throughout the cell — even at the extremes of animal cell size and metabolic rate.

The largest cell (*N. beyeri*) has significantly less ER surface area density than the smallest cell (*S. tropicalis*). The ER is a complex organelle consisting of different interconnected structures that perform different functions: the nuclear envelope, peripheral cisternae around the nucleus, and a tubular network that extends throughout the cytoplasm (English and Voeltz 2013; Gubas and Dikic 2022). Our analyses were not able to distinguish among these different components of the overall ER, limiting our ability to functionally interpret differences in the

amount of ER surface area per cell volume. However, decreased ER surface area density in the largest cells suggests the possibility that ER functions that take place in the membrane (e.g. membrane lipid synthesis) (English and Voeltz 2013) are operating at lesser capacity per unit of cytoplasm in larger cells relative to smaller cells. This pattern is consistent with the existence of relatively less plasma and nuclear membrane in larger cells (Table 2.2). In contrast, ER volume scales proportionately with cell volume, suggesting that ER functions that take place in the lumen (e.g. protein folding, processing, and assembly) are operating at the same capacity per unit of cytoplasm in larger and smaller cells. Taken together, these surface area and volume patterns suggest that the overall ER network shape is different in larger cells, trending towards less tubular morphology. Overall, our results are consistent with endoplasmic reticulum content scaling proportionate to cell volume following a limiting precursor model, which yields an ER network that comprises ~6% of the cell's volume, albeit with altered SA:V ratio in the largest cells that may track membrane biosynthesis functional demands. This percentage volume, in turn, produces a spatial distribution of tubules and cisternae that can maintain the physical connections between the ER and the mitochondria, plasma membrane, nucleus, and Golgi apparatus that are necessary for the functionality of all of these organelles (Heald and Cohen-Fix 2014).

The Golgi apparatus receives processed proteins from the ER via membrane-bound vesicles and completes the final stages of the secretory pathway: additional protein processing, sorting, and export in vesicles to the proteins' final destinations in the cell (e.g. the plasma membrane for secretion or the lysosomes for degradation) (Sengupta and Linstedt 2011). Thus, Golgi size is proximately held at a dynamic equilibrium by balanced influx and outflux of cargo in membrane-bound vesicles, and ER and Golgi size are functionally interconnected (Sengupta

and Linstedt 2011; Reber and Goehring 2015). Ultimately, though, Golgi size does scale with cell volume during the growth phase of the mammalian cell cycle (Sin and Harrison 2016), and we show here that Golgi content also scales proportionately with cell volume across evolutionary increases in cell size. Intracellular transport distance is among the most obvious parameters affected by increased cell size; thus, the result that the organelle responsible for transportation out of the cell via the secretory pathway is proportionately the same size across a 50-fold difference in cell volume is noteworthy, suggesting scaling based on a limited precursor model rather than a demand-drive model.

Our overall results are broadly consistent with the limited precursor model; the cytoplasm of larger cells contains more organellar building blocks, which are simply incorporated into larger organelles following the same “rules” across cell sizes, despite differences in metabolic demand and intracellular transport distances. The fact that all organelles scale proportionately with cell volume indicates that all the proteins required for organelle biosynthesis are maintained at the same concentrations across this evolved ~50-fold difference in cell volume. This finding is significant because it directly contradicts results from experimentally manipulated increases in cell volume, which reveal proteome dilution and perturbation so severe the cells appear senescent (Neurohr et al. 2019; Cheng et al. 2021; Lanz et al. 2022; Xie et al. 2022). Our results suggest that salamanders have evolved the biosynthetic capacity to maintain a functional proteome despite a huge cell volume. Further work will reveal the dynamics of RNA and protein synthesis and stability that underlie this hypothesized alteration to biosynthetic output.

## References

- Alberts, B., A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter. 2017. How cells read the genome: from DNA to protein. *Molecular Biology of the Cell*, pp.299-368.
- Amphibiaweb. 2023. AmphibiaWeb.org.
- Balachandra, S., S. Sarkar, and A. A. Amodeo. 2022. The Nuclear-to-Cytoplasmic Ratio: Coupling DNA Content to Cell Size, Cell Cycle, and Biosynthetic Capacity. *Annu Rev Genet.* 56:165-85.
- Bates, D., M. Mächler, B. M. Bolker, and S. C. Walker. 2015. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.*
- Brand, M. D., A. L. Orr, I. V. Perevoshchikova, and C. L. Quinlan. 2013. The role of mitochondrial function and cellular bioenergetics in ageing and disease. *Br. J. Dermatol.* 169:1–8.
- Cadart, C., J. Bartz, G. Oaks, M. Z. Liu, and R. Heald. 2023. Polyploidy in *Xenopus* lowers metabolic rate by decreasing total cell surface area. *Curr. Biol.* 33:1744-1752.e7.
- Cantwell, H., and P. Nurse. 2019a. A systematic genetic screen identifies essential factors involved in nuclear size control. *PLoS Genet.* 15(2):e1007929.
- Cantwell, H., and P. Nurse. 2019b. Unravelling nuclear size control. Springer Verlag. *Current Genet.* 65(6):1281-5.
- Chan, Y. H. M., and W. F. Marshall. 2014. Organelle size scaling of the budding yeast vacuole is tuned by membrane trafficking rates. *Biophys. J.* 106:1986–1996.
- Chan, Y. H. M., and W. F. Marshall. 2010. Scaling properties of cell and organelle size. *Organogenesis.* 6(2):88-96.
- Chen, P., S. Mishra, and D. L. Levy. 2023. Nuclear growth and import can be uncoupled. *bioRxiv* 2023.04.19.537556.
- Cheng, L., J. Chen, Y. Kong, C. Tan, R. Kafri, and M. Björklund. 2021. Size-scaling promotes senescence-like changes in proteome and organelle content. *bioRxiv* 2021.08.05.455193.
- Chong, R. A., and R. L. Mueller. 2013. Low metabolic rates in salamanders are correlated with weak selective constraints on mitochondrial genes. *Evolution (N. Y).* 67:894–899.
- Cooper, G., and K. Adams. 2022. *The Cell: A Molecular Approach*. Oxford University Press.
- Cosgrove, D. J. 1993. How do plant cell walls extend? *Plant physiology.* 02(1):1.
- Cushing, P. E., P. Casto, E. D. Knowlton, S. Royer, D. Laudier, D. D. Gaffin, L. Prendini, and J. O. Brookhart. 2014. Comparative morphology and functional significance of setae called

- papillae on the pedipalps of male camel spiders (Arachnida: Solifugae). *Ann. Entomol. Soc. Am.* 107:510–520.
- Day, K. J., L. A. Staehelin, and B. S. Glick. 2013. A three-stage model of Golgi structure and function. *Histochemistry and cell biology.* 140(3):239-49.
- De Santa Barbara, P., G. R. Van Den Brink, and D. J. Roberts. 2003. Development and differentiation of the intestinal epithelium. *Cellular and Molecular Life Sciences CMLS.* 60:1322-32
- 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.
- Deviri, D., and S. A. Safran. 2022. Balance of osmotic pressures determines the nuclear-to-cytoplasmic volume ratio of the cell. *Proc. Natl. Acad. Sci. USA* 119:e2118301119.
- English, A. R., and G. K. Voeltz. 2013. Endoplasmic reticulum structure and interconnections with other organelles. *Cold Spring Harbor perspectives in biology.* 5(4):a013227.
- Fox, J., and S. Weisberg. 2011. *An R companion to applied regression.* Sage publications.
- Gardner, J. D., M. Laurin, and C. L. Organ. 2020. The relationship between genome size and metabolic rate in extant vertebrates. *Philos. Trans. R. Soc. B Biol. Sci.* 375(1793):20190146.
- Gatten, R. E., K. J. Miller, and R. J. Full. 1992. Energetics at rest and during locomotion. Pp. 314–374 in *Environmental Physiology of the Amphibians.* 314-377.
- Glancy, B., L. M. Hartnell, D. Malide, Z. X. Yu, C. A. Combs, P. S. Connelly, S. Subramaniam, and R. S. Balaban. 2015. Mitochondrial reticulum for cellular energy distribution in muscle. *Nature* 523:617–620.
- Glazier, D. S. 2022. Variable metabolic scaling breaks the law: from “Newtonian” to “Darwinian” approaches. *Proc. R. Soc. B Biol. Sci.* 289(1985):20221605.
- Goniakowska, L. 1973. Metabolism, resistance to hypotonic solutions, and ultrastructure of erythrocytes of five amphibian species. *Acta Biol. Cracoviensia* 16:113–137.
- Gregory, T. R. 2023. *Animal Genome Size Database.* <http://www.genomesize.com>
- Gregory, T. R. 2005a. *Cell Size Database.* <https://www.genomesize.com/cellsize/>
- Gregory, T. R. 2005b. Synergy between sequence and size in large-scale genomics. *Nature Reviews Genetics.* 6(9):699-708.
- Gregory, T. R. 2001. The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates. *Blood Cells, Mol. Dis.* 27:830–843.

- Gregory, T. R. 2003. Variation across amphibian species in the size of the nuclear genome supports a pluralistic, hierarchical approach to the C-value enigma. *Biol. J. Linn. Soc.* 79:329–339.
- Gubas, A., and I. Dikic. 2022. ER remodeling via ER-phagy. *Mol. Cell* 82:1492–1500.
- Gundersen, H. J. G., P. Bagger, T. F. Bendtsen, S. M. Evans, L. Korbo, N. Marcussen, A. Moller, K. Nielsen, J. R. Nyengaard, B. Pakkenberg, F. B. Sorensen, A. Vesterby, and M. J. West. 1988. The new stereological tools: Disector, fractionator, nucleator and point sampled intercepts and their use in pathological research and diagnosis. *Apmis* 96:857–881.
- Guo, T., and Y. Fang. 2014. Functional organization and dynamics of the cell nucleus. *Frontiers in Plant Science.* 5:378.
- Hanken, J. 1983. Miniaturization and its effects on cranial morphology in plethodontid salamanders, genus *Thorius* (Amphibia, Plethodontidae): II. The fate of the brain and sense organs and their role in skull morphogenesis and evolution. *J. Morphol.* 177:255–268.
- Hara, Y., and A. Kimura. 2009. Cell-Size-Dependent Spindle Elongation in the *Caenorhabditis elegans* Early Embryo. *Curr. Biol.* 19:1549–1554.
- Heald, R., and O. Cohen-Fix. 2014. Morphology and function of membrane-bound organelles. *Curr. Opin. Cell Biol.* 26:79–86.
- Heijo, H., C. A. Merten, and Y. Hara. 2022. Differential contribution of nuclear size scaling mechanisms between *Xenopus* species. *Dev. Growth Differ.* 64:501–507.
- Howard, V., and M. Reed. 2004. *Unbiased Stereology: Three-Dimensional Measurement in microscopy.* Garland Science.
- Hughes, A. L., and M. K. Hughes. 1995. Small genomes for better flyers. *Nature.* 377(6548):391
- Itgen, M. W., G. R. Natalie, D. S. Siegel, S. K. Sessions, and R. L. Mueller. 2022. Genome size drives morphological evolution in organ-specific ways. *Evolution (N. Y.).* 76:1453–1468.
- Johnson, B. B., J. B. Searle, and J. P. Sparks. 2021. Genome size influences adaptive plasticity of water loss, but not metabolic rates, in lungless salamanders. *J. Exp. Biol.* 224.
- Jorgensen, P., N. P. Edgington, B. L. Schneider, I. Rupeš, M. Tyers, and B. Futcher. 2007. The size of the nucleus increases as yeast cells grow. *Mol. Biol. Cell* 18:3523–3532.
- Kapusta, A., A. Suh, and C. Feschotte. 2017. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. USA* 114:E1460–E1469.
- Klionsky, D. J., P. K. Herman, and S. D. Emr. 1990. The fungal vacuole: composition, function, and biogenesis. *Microbiol. Rev.* 54:266–292.
- Kozłowski, J., M. Konarzewski, and A. T. Gawelczyk. 2003. Cell size as a link between

- noncoding DNA and metabolic rate scaling. *Proc. Natl. Acad. Sci. U. S. A.* 100:14080–14085.
- Kremer, J. R., D. N. Mastrorade, and J. R. McIntosh. 1996. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* 116:71–76.
- Lanz, M. C., E. Zatulovskiy, M. P. Swaffer, L. Zhang, I. Ilert, S. Zhang, D. S. You, G. Marinov, P. McAlpine, J. E. Elias, and J. M. Skotheim. 2022. Increasing cell size remodels the proteome and promotes senescence. *Mol. Cell* 82:3255-3269.e8.
- Lenth, R. V. 2021. Package: “emmeans”: emmeans: Estimated Marginal Means, aka Least-Squares Means.
- Licht, L. E., and L. A. Lowcock. 1991. Genome size and metabolic rate in salamanders. *Comp. Biochem. Physiol. -- Part B Biochem.* 100:83–92.
- Lynch, M., and G. K. Marinov. 2017. Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *elife.* 6:e20437.
- Malerba, M. E., and D. J. Marshall. 2021. Larger cells have relatively smaller nuclei across the Tree of Life. *Evolution Letters.* 5(4):306-14.
- Marshall, W. F. 2020. Scaling of Subcellular Structures. *Annual Review of Cell and Developmental Biology.* 36:219-36.
- Marshall, W. F. 2015. Subcellular size. *Cold Spring Harb. Perspect. Biol.* 7:1–16.
- Marshall, W. F., K. D. Young, M. Swaffer, E. Wood, P. Nurse, A. Kimura, J. Frankel, J. Wallingford, V. Walbot, X. Qu, and A. H. K. Roeder. 2012. What determines cell size? *BMC Biol.* 10:1–22.
- Maul, G. G., and L. Deaven. 1977. Quantitative determination of nuclear pore complexes in cycling cells with differing DNA content. *J. Cell Biol.* 73:748–760.
- McCarron, J. G., C. Wilson, M. E. Sandison, M. L. Olson, J. M. Girkin, C. Saunter, and S. Chalmers. 2013. From structure to function: Mitochondrial morphology, motion and shaping in vascular smooth muscle. *Journal of Vascular Research.* 50(5):357-71.
- Miettinen, T. P., and M. Björklund. 2016. Cellular Allometry of Mitochondrial Functionality Establishes the Optimal Cell Size. *Dev. Cell* 39:370–382.
- Miettinen, T. P., and M. Björklund. 2017. Mitochondrial Function and Cell Size: An Allometric Relationship. *Trends Cell Biol.* 27(6):393-402.
- Miller, K. E., C. Brownlee, and R. Heald. 2020. The power of amphibians to elucidate mechanisms of size control and scaling. *Experimental Cell Research.* 392(1):112036.
- Mueller, R. L. 2015. Genome biology and the evolution of cell-size diversity. *Cold Spring Harb.*

- Perspect. Biol. 7:19125–19126.
- Neurohr, G. E., R. L. Terry, J. Lengefeld, M. Bonney, G. P. Brittingham, F. Moretto, T. P. Miettinen, L. P. Vaites, L. M. Soares, J. A. Paulo, J. W. Harper, S. Buratowski, S. Manalis, F. J. van Werven, L. J. Holt, and A. Amon. 2019. Excessive cell growth causes cytoplasm dilution and contributes to senescence. *Cell* 176:1083-1097.e18.
- Noel, J. S., W. C. Dewey, J. H. Abel, and R. P. Thompson. 1971. Ultrastructure of the nucleolus during the Chinese hamster cell cycle. *J. Cell Biol.* 49:830–847.
- Noske, A. B. 2010. Multi-scale, spatio-temporal analysis of mammalian cell tomograms.
- O’Connor, C. M., J. U. Adams, and J. Fairman. 2010. *Essentials of Cell Biology*. Cambridge, MA NPG Educ. 1:54.
- Okie, J. G., V. H. Smith, and M. Martin-Cereceda. 2016. Major evolutionary transitions of life, metabolic scaling and the number and size of mitochondria and chloroplasts. *Proc. R. Soc. B Biol. Sci.* 283.
- Organ, C. L., and A. M. Shedlock. 2009. Palaeogenomics of pterosaurs and the evolution of small genome size in flying vertebrates. *Biol. Lett.* 5:47–50.
- Pough, F. H. 1980. The Advantages of Ectothermy for Tetrapods. *Am. Nat.* 115:92–112.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Rafelski, S. M., M. P. Viana, Y. Zhang, Y. H. M. Chan, K. S. Thorn, P. Yam, J. C. Fung, H. Li, L. D. F. Costa, and W. F. Marshall. 2012. Mitochondrial network size scaling in budding yeast. *Science* (80-. ). 338:822–824.
- Reber, S., and N. W. Goehring. 2015. Intracellular scaling mechanisms. *Cold Spring Harb. Perspect. Biol.* 7(12):a019067.
- Roth, G., J. Blanke, and D. B. Wake. 1994. Cell size predicts morphological complexity in the brains of frogs and salamanders. *Proc. Natl. Acad. Sci. U. S. A.* 91:4796–4800.
- RStudio Team. 2021. RStudio: Integrated Development for R. RStudio. PBC, Boston, MA. <http://www.rstudio.com/>.
- Russ, J. C., and R. T. Dehoff. 2012. *Practical stereology*. Springer Science & Business Media.
- Schwarz, D. S., and M. D. Blower. 2016. The endoplasmic reticulum: Structure, function and response to cellular signaling. *Cellular and Molecular Life Sciences.* 73:79-94.
- Schwerzmann, K., H. Hoppeler, S. R. Kayar, and E. R. Weibel. 1989. Oxidative capacity of muscle and mitochondria: Correlation of physiological, biochemical, and morphometric characteristics. *Proc. Natl. Acad. Sci. U. S. A.* 86:1583–1587.

- Scott, G. R., K. H. Guo, and N. J. Dawson. 2018. The mitochondrial basis for adaptive variation in aerobic performance in high-altitude deer mice. *Integr. Comp. Biol.* 58:506–518.
- Seel, A., F. Padovani, A. Finster, M. Mayer, D. Bureik, C. Osman, T. Klecker, and K. M. Schmoller. 2022. Regulation with cell size ensures mitochondrial DNA homeostasis during cell growth. *bioRxiv* 2021.12.03.471050.
- Sengupta, D., and A. D. Linstedt. 2011. Control of organelle size: The Golgi complex. *Annu. Rev. Cell Dev. Biol.* 27:57–77.
- Sin, A. T.-W., and R. E. Harrison. 2016. Growth of the Mammalian Golgi Apparatus during Interphase. *Mol. Cell. Biol.* 36:2344–2359.
- Smith, D. R. 2017. Does cell size impact chloroplast genome size? *Front. Plant Sci.* 8:2116.
- Szarski, H. 1983. Cell size and the concept of wasteful and frugal evolutionary strategies. *J. Theor. Biol.* 105:201–209.
- Szarski, H. 1970. Changes in the amount of DNA in cell nuclei during vertebrate evolution. *Nature* 226:651–652.
- Tan, X., K. Li, Z. Wang, K. Zhu, X. Tan, and J. Cao. 2019. A review of plant vacuoles: Formation, located proteins, and functions. *Plants.* 8(9):327.
- Titorenko, V. I., H. Chan, and R. A. Rachubinski. 2000. Fusion of small peroxisomal vesicles in vitro reconstructs an early step in the in vivo multistep peroxisome assembly pathway of *Yarrowia lipolytica*. *J. Cell Biol.* 148:29–43.
- Uyeda, J. C., M. W. Pennell, E. T. Miller, R. Maia, and C. R. McClain. 2017. The evolution of energetic scaling across the vertebrate tree of life. *Am. Nat.* 190:185–199.
- Venturini, V., F. Pezzano, F. C. Castro, H. M. Häkkinen, S. Jiménez-Delgado, M. Colomer-Rosell, M. Marro, Q. Tolosa-Ramon, S. Paz-López, M. A. Valverde, J. Weghuber, P. Loza-Alvarez, M. Krieg, S. Wieser, and V. Ruprecht. 2020. The nucleus measures shape changes for cellular proprioception to control dynamic cell behavior. *Science.* 370(6514):eaba2644.
- Winey, M., J. B. Meehl, E. T. O’Toole, and T. H. Giddings. 2014. Conventional transmission electron microscopy. *Mol. Biol. Cell* 25:319–323.
- Wise, R. R. 2007. *The Diversity of Plastid Form and Function.* (pp. 3–26). Springer, Dordrecht.
- Womack, M. C., M. J. Metz, and K. L. Hoke. 2019. Larger genomes linked to slower development and loss of late-developing traits. *Am. Nat.* 194:854–864.
- Wright, N. A., T. R. Gregory, and C. C. Witt. 2014. Metabolic “engines” of flight drive genome size reduction in birds. *Proc. R. Soc. B Biol. Sci.* 281(1779):20132780.

Xie, S., M. Swaffer, and J. M. Skotheim. 2022. Eukaryotic cell size control and its relation to biosynthesis and senescence. *Annual Review of Cell and Developmental Biology*. 38:291-319.

# LONG INTRONS FROM THE GIGANTIC LUNGFISH GENOME ARE RECURSIVELY SPLICED AT LEVELS COMPARABLE TO THE HUMAN GENOME

## **Introduction**

As genomes increase in size, the amount of protein-coding DNA does not increase proportionately (Hedges and Kumar 2002; Smith et al. 2009; Wang et al. 2021). Instead, genomic expansion reflects an increase in transposable elements and intronic sequence (Sun et al. 2012; Wang et al. 2021). This increase in genomic content has been connected with an increase in nucleus size, cell size, and changes at the tissue and organ levels (Gregory 2001; Marguerat and Bähler 2012; Itgen et al. 2022). However, less research has focused on how this genomic gigantism alters the transcriptional process of the cell (Sessions and Wake 2020; Taylor et al. 2023)

In a typical eukaryotic cell, DNA is transcribed into a pre-mRNA transcript, introns are spliced out leaving only the exonic sequences, and a poly-adenylated tail is attached to make a mature mRNA transcript (Hocine et al. 2010). In the canonical form of splicing, introns are removed by spliceosomes, facilitated through small nuclear ribonucleoproteins (snRNPs) (Berget et al. 1977; Darnell 1978). The 5' snRNP attaches to the complementary sequence of the transcript downstream of the intron; this forms a lariat structure that leads to the activation of the spliceosome, which then removes the intron all at once (Konarska et al. 1985; Sharp 1985; Gehring and Roignant 2021).

An alternative to this canonical splicing mechanism was later discovered (Hatton et al. 1998; Burnette et al. 2005). This novel form of splicing — called recursive splicing (RS) — works by removing introns in multiple sequential pieces instead of all at once (Hatton et al.

1998; Burnette et al. 2005). The mechanism works by splicing out a section of the intron flanked by discreet splicing sites, referred to as recursive splice sites (RSS) (Duff et al. 2015; Sibley et al. 2015; Kelly et al. 2015). These subsections are still removed using lariat structures and spliceosomes (Hoppe et al. 2023). As each is removed sequentially, the remaining downstream intronic sequence is brought into contact with the upstream exon at the RSS (Sibley et al. 2015; Georgomanolis et al. 2016). These ephemeral partially spliced introns can be identified through deep sequencing of total RNA (Burnette et al. 2005; Sibley et al. 2015). Their existence results in uneven RNA-seq read depths across the entire length of the intron that produce a characteristic sawtooth pattern of depth of coverage, with the RSS defining the edges of the sawtooth. This phenomenon has been identified in several model species such as fruit flies, mice, and humans (Hatton et al. 1998; Sibley et al. 2015; Joseph et al. 2018; Moon and Zhao 2022).

The presence of recursive splicing correlates with intron length; most RSS in *Drosophila* occur in introns longer than 40kb, and most RSS in mice occur in introns longer than 51kb (Pai et al. 2018; Moon and Zhao 2022). However, all this work has examined long introns in typically sized genomes. Natural genome size diversity extends across a much greater range than is seen across mammals and *Drosophila*; salamanders and lungfishes, for example, demonstrate up to a 40-fold increase in genome size relative to humans (Gregory 2023). On average, intron length in the model salamander *Ambystoma mexicanum* (genome size = 32 Gb) is 13 times longer than in humans, and similar intronic expansion is seen in the Australian lungfish *Neoceratodus forsteri* (genome size = ~50 Gb) (Nowoshilow et al. 2018; Meyer et al. 2021). Considering this intronic expansion in species with large genomes, and the fact that RS occurs in the longest introns in previous studies of smaller genomes, we hypothesized that there would be a higher incidence of recursive splicing to manage the job of removing introns in gigantic genomes.

To test this hypothesis, we looked for evidence of recursive splicing in a species that demonstrates genomic gigantism. Recent advances in sequencing and assembly have begun to make gigantic genomes amenable to genomic and transcriptomic analyses (Warren et al. 2015; Stevens et al. 2016; Nowoshilow et al. 2018; Wang et al. 2021). The west African lungfish *Protopterus annectens* (40 Gb) is among the largest well-annotated genome assemblies on NCBI. Using publicly available deep sequencing total RNA datasets, we tested the 200 longest genes of *P. annectens* for evidence of recursive splicing, and we compared recursive splicing levels to those found in the 200 longest genes in the human genome. Our analyses suggest that even for the longest introns in the largest vertebrate genomes, high levels of recursive splicing at conserved sites are not necessary to manage their increased length.

## **Materials and Methods**

There are currently three animal species that demonstrate genomic gigantism with genome assemblies on NCBI: the Mexican axolotl salamander *Ambystoma mexicanum*, the Australian lungfish *Neoceratodus forsteri*, and the West African lungfish *Protopterus annectens*. Of these species, only *P. annectens* had available the total RNA-seq read depths and extensive genome annotation that was needed for recursive splicing analysis (Wang et al. 2021; Sayers et al. 2022). We used the *P. annectens* genome assembly (NCBI reference number GCA\_019279795.1). In addition, we used the largest dataset for the species based on rRNA depleted RNA-seq libraries that did not use polyadenylation tail selection, as poly-A selection only selects for mature transcripts; the libraries were constructed from liver, gill, gut, and lung tissues (SRA entries SRX3392415, SRX3392416, SRX3392417, SRX3392418; ~25 billion bp total) (Chen et al. 2021).

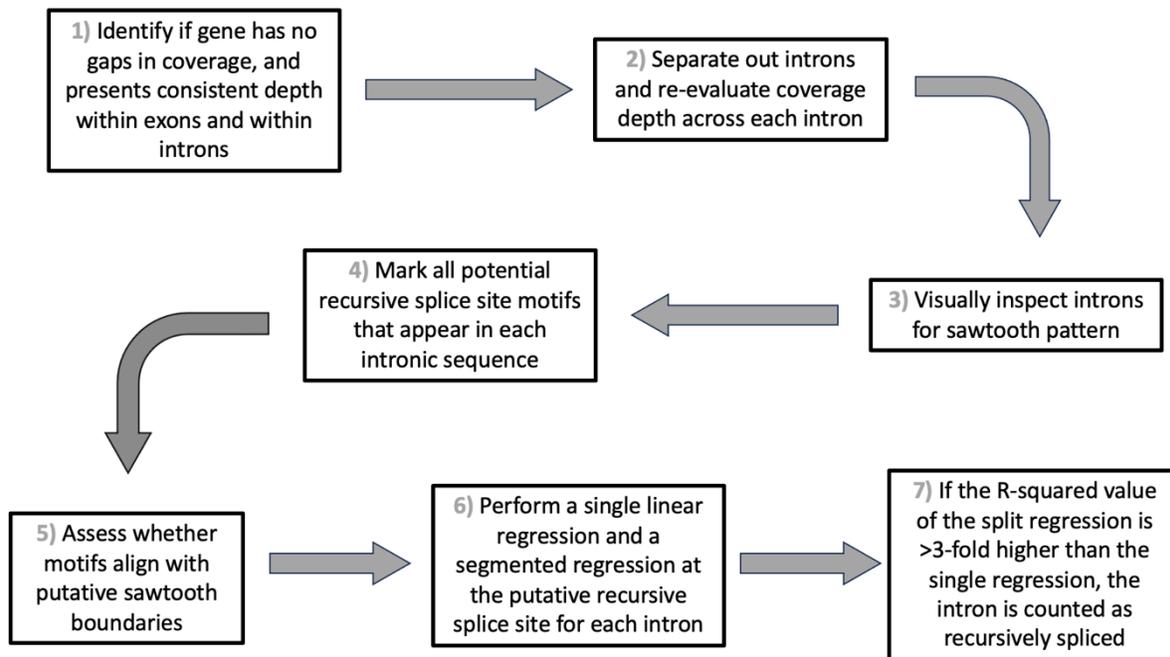
The human genome was chosen both as the positive control for validating our RSS identification pipeline, and as a point of comparison representing a typically sized vertebrate genome, as the human genome is one of the most extensively studied for recursive splicing (Sibley et al. 2015; Wan et al. 2021; Hoppe et al. 2023). We used the genome assembly used in previous human RSS analyses (GRCh37.p13, NCBI reference GCF\_000001405.25) (Sibley et al. 2015). We used rRNA depleted RNA-seq libraries that did not use poly-A selection; the libraries were constructed from liver tissue, one of the tissues represented in the lungfish dataset (SRX20105218; ~25.9 billion bp total). We subsampled the human liver RNA-seq dataset to ~25.9 billion bp total to match the size of the lungfish RNA-seq datasets.

Our overall pipeline is summarized in Figure 3.1. First, we assessed depth of RNA-seq coverage across the entire gene length. To this end, RNA reads were aligned to the genomic sequence using Burrows Wheeler Aligner (Li and Durbin 2009), then sorted and indexed using SAMtools (Danecek et al. 2021). Bam files were organized using the BEDTools suite (Quinlan and Hall 2010) to allow for extraction of read depths for specific genes. Second, for genes that showed sufficient coverage (i.e.,  $\geq 5$  reads at roughly 75% of sites), we separated out the individual introns and assessed them for depth of coverage. Third, we assessed the introns with sufficient coverage for a putative sawtooth pattern consisting of spikes in read depth within the intron with a sloping decrease in read depth on the side consistent with whether the transcript was derived from the sense or anti-sense strand (Sibley et al. 2015; Georgomanolis et al. 2016; Joseph et al. 2018). To this end, introns were visualized in R Studio (Allaire 2011; RStudio Team 2021; R Core Team 2021) using R packages dplyr (Wickman et al. 2021) and ggplot2 (Kassambara, 2020). Fourth, we mined the transcript sequences of introns showing this sawtooth pattern for the presence and location of motifs that have previously been associated with RSS:

AG/GTAAG, AG/GTGAG, AG/GTAGG, AG/GTATG, AG/GTAAA, AG/GTAAT, AG/GTGGG, AG/GTAAC, AG/GTCAG, or AG/GTACG (Sibley et al. 2015). Fifth, we visually assessed whether any of these sites lined up with the putative sawtooth boundaries. Finally, we used linear regression models to test whether the introduction of the putative RSS led to an improved fit of the relationship between intronic position and read depth. To this end, we first tested for a relationship between intronic position and RNA-seq read depth using linear regression across the entire intron. We then tested if introducing the putative RSS improves model fit by performing a segmented regression with two subsections of each intron using the RSS as the breakpoint. Based on our analyses following Sibley et al. (2015), we considered a  $\geq 1.1$  slope increase in R-squared value as evidence for recursive splicing.

Human candidate genes for pipeline validation were taken from Sibley et al. (2015). We selected the 20 genes that had the strongest evidence for recursive splicing based on the comparison of single to segmented regression models: OPCML, CADM2, HS6ST3, LRRC4C, CADM1, SPATA5, PDE4D, ANK3, NTM, SYN3, MAGI1, GPM6A, RBFOX1, PCDH15, CDH12, PCDH9, ROBO2, FRMD4A, NCAM1 and NSMCE2. We then tested whether our pipeline identified the same RSS.

We ordered all annotated genes in the lungfish genome from longest to shortest and selected the 200 longest genes. Next, we obtained all of the RNA-seq dataset fastq files using the SRA Toolkit provided by NCBI (SRA Toolkit Development Team 2019). We applied our pipeline (Figure 3.1) to these 200 genes to identify cases of recursive splicing in the lungfish. We then identified the 200 longest genes in the human genome and obtained recursive splicing levels from Sibley et al. (2015) as a point of comparison.

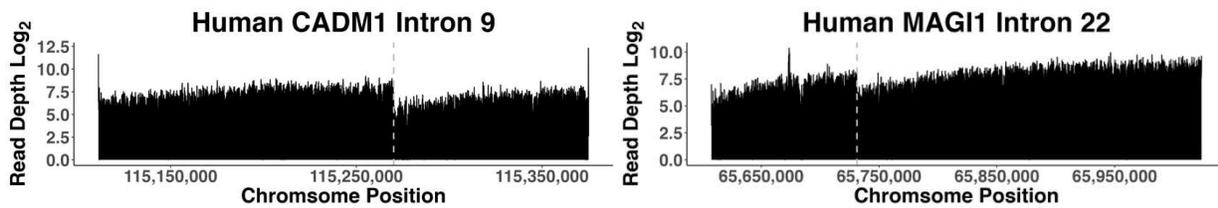


**Fig. 3.1:** Workflow to identify recursive splicing.

## Results

Of the 20 human genes selected as positive RSS controls, three had sufficient RNA-seq coverage to move past the first step of the pipeline. Our pipeline then identified the RSS annotated in Sibley et al. 2015 (Figure 3.2), validating our pipeline’s efficacy and suggesting that these RSS sites are conserved across at least brain tissue (used in Sibley et al. 2015) and liver tissue (used here)(Table 3.1). The remaining 17 genes for which we had insufficient coverage likely had lower relative expression in liver (i.e. LRRC4C, NTM, SYN3, RBFOX1 and PCDH15 are not expressed in the liver) (Yu et al. 2010). In order to establish a baseline in breadth and depth of coverage that can still identify known cases of recursive splicing, we randomly subsampled the three genes to achieve lower read depths by parsing the fastq files into smaller and smaller subsets. We determined that a minimum of 40% of all sites must have  $> 0$  reads, with

no large gaps, and the remaining sites must have an average read depth  $\geq 2.1$ . These specific metrics were then used as a final filter for lungfish introns.

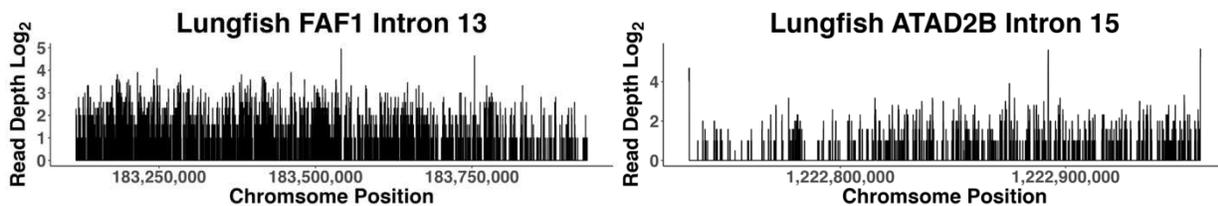


**Fig. 3.2:** Depth of coverage of RNAseq reads across two human introns with known RSS; identification of sawtooth pattern with RSS sites indicated with dotted lines (Pipeline Steps 2, 3, and 4)

**Table 3.1:** Assessing model fits with single and segmented regressions.

Gene	Gene Length	Intron #	Intron Length	R-Squared Segmented	R-Squared Whole	Fold Difference
PDE4D	1,527,249	14	1,019,451	0.205	0.039	5.3
MAGI1	685,393	22	684,471	0.485	0.453	1.1
CADM1	335,178	9	335,178	0.176	1.19E-04	1,474

The longest 200 lungfish genes ranged in size from 4,097,758 – 17,083,450 bp in length. Twenty-five of these 200 genes were identified as having sufficient coverage; the other 175 lacked consistent depth of  $\geq 5$  reads. These 25 genes were subdivided into their 856 separate introns, which ranged in size from 85 – 1,181,814 bp in length. Of these 856 introns, 79 showed a spike in read depth and were further investigated (Fig. 3.3).



**Fig. 3.3:** Two representative lungfish introns with sufficient coverage and a read-depth peak in the middle of the intron suggestive of an RSS.

Of the 79 introns with a spike in read depth, 29 introns across 16 genes contained an RSS motif that coincided with a putative sawtooth boundary. The other 50 introns contained motifs

that did not coincide with any putative sawtooth boundaries. In total, there were 21,057 RSS sites across the 30,701,965 bp of intronic sequence, corresponding to one RSS per 1,460 bp. Of these 29 introns across 16 genes, 19 introns across 12 different genes showed a  $\geq 1.1$  increase of fit in their R-squared values under the segmented regression model with the segment defined at the putative RSS site compared to the single regression model. Further analysis of the “spike” sequence was done to see if the pattern was an artifact caused by a repetitive genomic sequence (i.e. a TE) that would have high corresponding RNA-seq depth independent of recursive splicing. Ten of the 19 sequences were identified as overrepresented in the genome, mapping to  $> 10$  different unplaced genome scaffolds, and were eliminated out of an abundance of caution. Only four of the remaining introns met the minimum read-depth qualification of  $> 40\%$  coverage with an average read depth of  $\geq 2.1$ ; we considered these to show evidence of RSS.

Overall, we found evidence for recursive splicing in 8% (4/25) of the longest lungfish genes for which we had sufficient initial coverage to continue analysis. These 25 genes contain 856 introns; thus, 0.47% (4/856) of the total introns had strong evidence for recursive splicing. No genes contained strong evidence for more than one recursively spliced intron (Table 3.2).

In the human genome, of the 200 longest genes, Sibley et al. (2015) explored 170 of them for the presence of recursive splicing, and 10% (17/170) showed evidence of an intron that undergoes recursive splicing (Sibley et al. 2015). These 170 genes contain roughly 3,794 introns — depending on which annotation is selected — thus, only 0.45% (17/3,794) of those introns show evidence of recursive splicing. No genes showed evidence for more than one recursively spliced intron, although some introns had more than one RSS, indicating removal in more than two subsections.

**Table 3.2:** The 4 introns that show the strongest evidence for potential recursive splicing in lungfish.

Gene	Gene Length	Intron #	Intron Length	R-Squared Segmented	R-Squared Whole	Fold Difference
ATAD2B	4,508,040	15	227,086	0.023	0.018	1.28
CLIP1	4,646,655	2	303,186	0.028	0.021	1.32
CREBBP	4,149,154	16	147,732	0.045	0.038	1.17
FAF1	5,915,520	13	817,342	0.057	0.051	1.12

## Discussion

Sibley et al. (2015) found that 0.45% of introns in the 200 longest genes in humans are removed by recursive splicing (Sibley et al. 2015). Using a similar approach, we found that 0.47% of the introns in the longest genes in lungfish are removed by recursive splicing, although we note that this may be an underestimate based on low sequencing coverage. There are 31 genes that are among the 200 longest genes in both humans and lungfish, six of which show evidence of recursive splicing in humans (Sibley et al. 2015). Of these six, only FOXP1 had enough RNA-seq coverage for analysis in lungfish, and the 3<sup>rd</sup> intron — recursively spliced in humans — did not show alignment between RSS motifs and any sawtooth boundaries.

In humans, the introns that are removed by recursive splicing are always the longest intron in the gene, typically the first intron (Bradnam and Korf 2008). In contrast, in the lungfish, the introns that show evidence for removal by recursive splicing are not the longest intron in the gene, and the longest intron is not typically the first intron, potentially changing the relationship between recursive splicing and intron identity. Thus, although overall recursive splicing levels are similar between the two datasets, the identity of the RS intron with respect to gene identity, intron length, or intron position is not the same.

In *Drosophila*, several lines of evidence suggest that recursive splicing is functionally important for intron splicing. RSS are distributed non-randomly across the longest introns, subdividing the longest ones into equal subsections (Joseph et al. 2018; Pai et al. 2018). Additionally, splicing becomes increasingly noisy and error-prone as intron lengths increase, and recursive splicing leads to more accurate, albeit slower, splicing of the longest introns (Pai et al. 2018).

In humans, splicing also becomes noisier and more error-prone as intron length increases (Pickrell et al. 2010). However, initial estimates of recursive splicing in humans and mice — which targeted nervous system tissue, as it is characterized by longer transcript lengths — revealed lower levels than those seen in flies, despite overall longer intron lengths (Sibley et al. 2015; Joseph et al. 2018; Moon and Zhao 2022). Thus, the functional significance of recursive splicing remained obscure.

As analytical methods developed, higher levels of recursive splicing were identified in humans, revealing the phenomenon in introns of all lengths and multiple cell types beyond the nervous system (Wan et al. 2021; Hoppe et al. 2023). Importantly, live imaging of single-cell transcriptional and splicing dynamics revealed that recursive splicing is widespread, occurring in ~30% of human genes; however, RSS selection can be stochastic, removing introns in a variety of different segments defined not at conserved sites, but rather by random selection by the spliceosome of one of many possible RSS (Wan et al. 2021). This stochastic process produces diverse, transient intermediate RNA molecules that do not appear as a sawtooth pattern in RNA-seq datasets. Taken together, this work revealed that many human genes experience recursive splicing, that it is more prevalent in longer introns, and that it can be stochastic (particularly in shorter genes) or occur constitutively at conserved RSS (Hoppe et al. 2023).

Based on these earlier studies, we hypothesized that the long introns in the gigantic lungfish genome would experience noisy, error-prone splicing if excised as single lariats, and that recursive splicing might have evolved to be more prevalent in response to this noise. In contrast, our results do not support high levels of recursive splicing occurring at conserved RSS in the lungfish; <1% of introns show a sawtooth pattern of RNA-seq read depths. Although this is likely an underestimate, and lariat sequencing or imaging analyses would likely reveal additional recursively spliced lungfish introns (Hoppe et al. 2023), the relevant comparison is with the 13-fold smaller human genome analyzed with the same approach, which also showed <1% recursively spliced introns (Sibley et al. 2015). We note that the lungfish sample was derived from mixed tissues, so we cannot formally exclude the possibility that higher levels of recursive splicing are occurring, but at conserved sites that differ across tissue types, thus obscuring the pattern in our mixed sample.

We emphasize that our results do not preclude the possibility of high levels of stochastic recursive splicing occurring in lungfish. In all 25 lungfish genes with sufficient coverage, we observed a pattern of decreasing RNA-seq read-depth across the length of the entire intron. Although this pattern is predicted under canonical cotranscriptional splicing, we note that it could also be produced by stochastic recursive splicing using a large number of potential RSS sites. Additional approaches to test for recursive splicing such as intron lariat sequencing and single-molecule imaging would provide additional insight into stochastic splicing (Wan et al. 2021; Hoppe et al. 2023).

In summary, although the longest introns within a species' genome may be more likely to undergo recursive splicing, we do not find support for the hypothesis that species with longer introns overall show higher overall levels of recursive splicing. This suggests that, even when

genomes evolve to be huge, canonical splicing is able to excise the long introns with error levels that can be accommodated by the cell. Alternatively, it leaves open the possibility that stochastic recursive splicing does remove these long introns in pieces, but that the pieces are defined randomly from a large variety of potential RS sites. Additional approaches will be helpful in continuing to reveal how genomic gigantism affects transcriptional and splicing dynamics of RNA.

## References

- Allaire, J. 2011. RStudio: Integrated development environment for R. *J. Wildl. Manage.* 75:1753–1766.
- Berget, S. M., C. Moore, and P. A. Sharp. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc. Natl. Acad. Sci. USA* 74:3171–3175.
- Bradnam, K. R., and I. Korf. 2008. Longer first introns are a general property of eukaryotic gene structure. *PLoS One* 3:e3093.
- Burnette, J. M., E. Miyamoto-Sato, M. A. Schaub, J. Conklin, and A. J. Lopez. 2005. Subdivision of large introns in *Drosophila* by recursive splicing at nonexonic elements. *Genetics* 170:661–674.
- Chen, X. X., W. C. Wu, and M. Shi. 2021. Discovery and characterization of actively replicating dna and retro-transcribing viruses in lower vertebrate hosts based on rna sequencing. *Viruses*. 13(6):1042.
- Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, and R. M. Davies. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:1–4.
- Darnell, J. E. 1978. Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science*. 202:1257–1260.
- Duff, M. O., S. Olson, X. Wei, S. C. Garrett, A. Osman, M. Bolisetty, A. Plocik, S. E. Celniker, and B. R. Graveley. 2015. Genome-wide identification of zero nucleotide recursive splicing in *Drosophila*. *Nature* 521:376–379.
- Gehring, N. H., and J. Y. Roignant. 2021. Anything but ordinary – emerging splicing in mechanisms in eukaryotic gene regulation. *Trends in Genetics*. 37(4):355-72.
- Georgomanolis, T., K. Sofiadis, and A. Papantonis. 2016. Cutting a long intron short: Recursive splicing and its implications. *Front. Physiol.* 7:598.
- Gregory, T. R. 2023. Animal Genome Size Database. <http://www.genomesize.com>
- Gregory, T. R. 2001. The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates. *Blood Cells, Mol. Dis.* 27:830–843.
- Hatton, A. R., V. Subramaniam, and A. J. Lopez. 1998. Generation of alternative ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol. Cell* 2:787–796.
- Hedges, S. B., and S. Kumar. 2002. Vertebrate genomes compared. *Science*. 297:1283–1285.
- Hocine, S., R. H. Singer, and D. Grünwald. 2010. RNA processing and export. *Cold Spring*

- Harbor Laboratory Press. 2(12):a000752.
- Hoppe, E. R., D. B. Udy, and R. K. Bradley. 2023. Recursive splicing discovery using lariats in total RNA sequencing. *Life Sci. Alliance* 6. Life Science Alliance. 6(7).
- Itgen, M. W., G. R. Natalie, D. S. Siegel, S. K. Sessions, and R. L. Mueller. 2022. Genome size drives morphological evolution in organ-specific ways. *Evolution (N. Y.)*. 76:1453–1468.
- Joseph, B., S. Kondo, and E. C. Lai. 2018. Short cryptic exons mediate recursive splicing in *Drosophila*. *Nat. Struct. Mol. Biol.* 25:365–371.
- Kassambara, A. 2020. Package ‘ggpubr’: “ggplot2” Based publication ready plots. R Packag. version 0.4.0.
- Kelly, S., T. Georgomanolis, A. Zirkel, S. Diermeier, D. O’Reilly, S. Murphy, G. Längst, P. R. Cook, and A. Papantonis. 2015. Splicing of many human genes involves sites embedded within introns. *Nucleic Acids Res.* 43:4721–4732.
- Konarska, M. M., P. J. Grabowski, R. A. Padgett, and P. A. Sharp. 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. *Nature* 313:552–557.
- Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760.
- Marguerat, S., and J. Bähler. 2012. Coordinating genome expression with cell size. *Trends in Genetics*. 28(11):560-5.
- Meyer, A., S. Schloissnig, P. Franchini, K. Du, J. M. Woltering, I. Irisarri, W. Y. Wong, S. Nowoshilow, S. Kneitz, A. Kawaguchi, A. Fabrizius, P. Xiong, C. Dechaud, H. P. Spaink, J. N. Volff, O. Simakov, T. Burmester, E. M. Tanaka, and M. Scharl. 2021. Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* 590:284–289.
- Moon, S., and Y. T. Zhao. 2022. Recursive splicing is a rare event in the mouse brain. *PLoS One* 17:e0263082.
- 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, F. Falcon, D. Knapp, S. Powell, A. Cruz, H. Cao, B. Habermann, M. Hiller, E. M. Tanaka, and E. W. Myers. 2018. The axolotl genome and the evolution of key tissue formation regulators. *Nature* 554:50–55.
- Pai, A. A., J. M. Paggi, P. Yan, K. Adelman, and C. B. Burge. 2018. Numerous recursive sites contribute to accuracy of splicing in long introns in flies. *PLoS Genet.* 14:e1007588.
- Pickrell, J. K., A. A. Pai, Y. Gilad, and J. K. Pritchard. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* 6:1–11.

- Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- RStudio Team. 2021. RStudio: Integrated Development for R. RStudio. PBC, Boston, MA. <http://www.rstudio.com/>.
- Sayers, E. W., E. E. Bolton, J. R. Brister, K. Canese, J. Chan, D. C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B. W. Trawick, K. D. Pruitt, and S. T. Sherry. 2022. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 50:D20–D26.
- Sessions, S. K., and D. B. Wake. 2020. Forever young: Linking regeneration and genome size in salamanders. *Dev. Dyn.* 250(6):768-78.
- Sharp, P. A. 1985. On the origin of RNA splicing and introns. *Cell.* 42(2):397-400.
- Sibley, C. R., W. Emmett, L. Blazquez, A. Faro, N. Haberman, M. Briese, D. Trabzuni, M. Ryten, M. E. Weale, J. Hardy, M. Modic, T. Curk, S. W. Wilson, V. Plagnol, and J. Ule. 2015. Recursive splicing in long vertebrate genes. *Nature* 521:371–375.
- Smith, J. J., S. Putta, W. Zhu, G. M. Pao, I. M. Verma, T. Hunter, S. V Bryant, D. M. Gardiner, T. T. Harkins, and S. R. Randal. 2009. Genic regions of a large salamander genome contain long introns and novel genes. *BMC genomics.* 10:1-1.
- SRA Toolkit Development Team. 2019. SRA Toolkit.
- Stevens, K. A., J. L. Wegrzyn, A. Zimin, D. Puiu, M. Crepeau, C. Cardeno, R. Paul, D. Gonzalez-Ibeas, M. Koriabine, A. E. Holtz-Morris, P. J. Martínez-García, U. U. Sezen, G. Marçais, K. Jermstad, P. E. McGuire, C. A. Loopstra, J. M. Davis, A. Eckert, P. de Jong, J. A. Yorke, S. L. Salzberg, D. B. Neale, and C. H. Langley. 2016. Sequence of the sugar pine megagenome. *Genetics* 204:1613–1626.
- 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.
- Taylor, A. N., R. L. Mueller, and A. Prasad. 2023. Amphibian segmentation clock models suggest mechanisms of slowed development across increasing genome size and nuclear volume. *bioRxiv* 2023.07.16.549220.
- Wan, Y., D. G. Anastasakis, J. Rodriguez, M. Palangat, P. Gudla, G. Zaki, M. Tandon, G. Pegoraro, C. C. Chow, M. Hafner, and D. R. Larson. 2021. Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection. *Cell* 184:2878-2895.e20.

- Wang, K., J. Wang, C. Zhu, L. Yang, Y. Ren, J. Ruan, G. Fan, J. Hu, W. Xu, X. Bi, Y. Zhu, Y. Song, H. Chen, T. Ma, R. Zhao, H. Jiang, B. Zhang, C. Feng, Y. Yuan, X. Gan, Y. Li, H. Zeng, Q. Liu, Y. Zhang, F. Shao, S. Hao, H. Zhang, X. Xu, X. Liu, D. Wang, M. Zhu, G. Zhang, W. Zhao, Q. Qiu, S. He, and W. Wang. 2021. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* 184:1362-1376.e18.
- Warren, R. L., C. Yang, B. P. Vandervalk, B. Behsaz, A. Lagman, S. J. M. Jones, and I. Birol. 2015. LINKS: Scalable, alignment-free scaffolding of draft genomes with long reads. *Gigascience*. 4(1):s13742-015.
- Wickman, H., R. François, L. Henry, and K. Muller. 2021. dplyr: A grammar of data manipulation (version 1.0. 0). R package version. 2021.
- Yu, Y., J. Ping, H. Chen, L. Jiao, S. Zheng, Z. G. Han, P. Hao, and J. Huang. 2010. A comparative analysis of liver transcriptome suggests divergent liver function among human, mouse and rat. *Genomics* 96:281–289.

## CONCLUSION

My research investigating the impacts of genome gigantism on both genome and cellular biology reveals that, that time and time again, these outliers navigate the world without the need for wholly unique adaptations. More specifically, the aspects explored in this dissertation — epigenetic silencing, cellular ultrastructure, and recursive splicing — all share features with genomes of more typical (not gigantic) sizes. The results of this work reveal that these incredibly large cells exist in the world seemingly unperturbed by their extreme size and considerable DNA content.

Our analysis of methylation of cytosines at CpG dinucleotide sites in different salamander species helped to provide insights into epigenetic silencing. Looking across different haploid genome sizes and ploidy levels, we found no significant difference in percentage of methylated cytosines between the species. Additionally, though the overall percentage of global methylation in salamanders was high, further comparison of salamanders to both endothermic and ectothermic species showed that similarly high levels of global methylation exist across ectothermic species. All this information taken together is the first indicator that large cells and large genomes continue along, operating as normal, typically sized cells do.

Zooming out, we then looked at enterocytes of different sizes, which reflected different genome sizes across species. Across cell sizes, we observed isometric scaling of the nucleus, mitochondria, ER and Golgi, all of which maintain their shape and percentage of cell volume over a ~50-fold change in size. This seems to be true even between species with different metabolic needs, where we see mitochondria scaling with cell volume rather than the demands of the species' metabolism. Again we see these large cells bucking expectations; where we expect

mitochondria to match energetic need, we see instead isometry in their distribution. When experimental manipulations of cell size suggested that we would find proteome dilution and senescence, we see these huge cells functioning completely fine.

Finally, we looked at intronic splicing, hypothesizing that genes becoming much longer due to intronic expansion may utilize different forms of splicing. Recursive splicing has been previously reported in longer introns and shown to aid in splicing fidelity in long introns. We hypothesized that as we investigated the 200 longest genes in lungfish, the incidence of recursive splicing would be high relative to humans. Our results, however, suggest that conserved recursive splice sites in lungfish appeared at comparable levels to humans, despite a 13-fold difference in genome size and average intron length. More recent data has indicated that recursive splicing can happen stochastically and in any length of intron. The patterns we report in lungfish could also result from stochastic splicing such as this and, again, demonstrate that species with genome gigantism do not significantly differ from species with smaller genomes.

Despite the extreme differences that are seen in species with genome gigantism, both at the genome and cell levels, our results point to the conclusion that these species keep trudging along; we show no differences in DNA transcriptional silencing, RNA intronic splicing, or organellar scaling. Granted their cells are larger, but size seems to be the only real difference that separates them from other organisms because proportionally they remain the same. However, when a cell increases in size experimentally, it causes a breakdown of the cell's overall functionality; the proteome begins to unravel leading the cell into senescence. With changes in cell size that occur over evolutionary time we do not see the same occurrence; there is more silencing that must be done, a larger presence of organelles, and a greater number of long introns

that need to be splice, but they manage as well as any other vertebrate species. I guess you could say these organisms abide.