CENTRO DE INVESTIGACIÓN Y ESTUDIOS AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD IRAPUATO

Genome size evolution in neotropical salamanders

Tesis que presenta

Biól. Louis Paul Decena Segarra

Para Obtener el Grado de

Maestro en Ciencias

En la Especialidad de

Biotecnología de Plantas

Director de Tesis: Sean M. Rovito

Irapuato, Guanajuato Febrero 2018

El presente trabajo de investigación se desarrolló en el laboratorio de Genómica y Biodiversidad de Vertebrados de la Unidad de Genómica Avanzada del CINVESTAV unidad Irapuato. Bajo la dirección del Dr. Sean M. Rovito y con la asesoría de los doctores Alexander De Luna Fors y Ruairidh J. H. Sawers. Agradezco al Consejo Nacional de Ciencia y Tecnología por la beca de estudios de maestría número 590252 otorgada durante el periodo comprendido entre Septiembre de 2015 a Agosto de 2017.

CONTENTS

Figure and Table Index

Abstract

The wide distribution of genome sizes among vertebrates has intrigued biologists since they started researching the nature of nuclei and the amount of DNA contained in them. Following the first descriptions of genome size in organisms, the fact that there is a constant amount of DNA among different individuals of the same species became evident, and the term C-value was born to refer to this phenomenon. An observation that baffled scientists at the time was that some organisms considered 'lower' or 'less evolved' (from an anthropocentric point of view), such as some fishes and amphibians, had a larger amount of DNA than 'higher' species such as humans. This apparent discrepancy between genome size and organismal complexity was called the C-value paradox. With the advent of sequencing technologies, this paradox was partially resolved as we learned that the amount of DNA does not correlate with the number of genes in an organism. Even though we now understand that there is no correlation between organismal complexity and genome size, there are still unanswered questions about how the huge variation in genome size across taxa arose and what, if any, are the consequences of having a very large genome for the organism. Neotropical salamanders are a diverse group of amphibians that show wide variation in both body size, ranging from less than 20 mm to over 160 mm in maximum Snout Vent-Length, and genome size, ranging from 9.3 pg to 81.1 pg of DNA per haploid cell. Because of their large cells, miniature salamanders (<35 mm SVL) have had to undergo several extreme morphological changes. Because of the close relationship between genome size and cell size, I expected to find a strong relationship between genome size and body size in neotropical salamanders. I estimated the genome size of 54 neotropical salamander species without previously reported genome size estimates, including multiple miniature species, using Feulgen Image Analysis Densitometry. I found the smallest genome sizes ever reported for a salamander in *Thorius spilogaster* (9.3 pg), and the largest genome size for a plethodontid salamander in *Bolitoglossa macrinii* (81.1pg). Without accounting for phylogeny, there is a strong correlation between genome size and body size. When taking into account phylogeny using Phylogenetic Generalized Least Squares there is no significant correlation between these two traits although clades with miniature salamanders have smaller genomes overall. I also reconstructed ancestral states for genome size to hypothesize the evolutionary history of changes that lead to the distribution of genome sizes on extant salamander clades and found that there is strong phylogenetic signal in this trait. The results from this large survey of genome sizes make us rethink some ideas about the relationships between genome size and body size (particularly as it relates to miniaturization) and give us a framework for future studies on the dynamics of genome size evolution in neotropical salamanders and animals in general.

Resumen

La amplia distribución de tamaños del genoma en vertebrados ha sido una gran para los biólogos desde que comenzaron a estudiar la naturaleza del núcleo y la cantidad de ADN en él. A través de los años de estudio siguiendo las primeras descripciones del tamaño del genoma en distintos organismos, el hecho de que existe una cantidad constante de ADN entre individuos de una misma especie se volvió evidente, y se acuño el término “C-value” para referirse a este fenómeno. Una observación que desconcertó a los científicos de la época fue que organismos que se consideraban ‘menos evolucionados’ (desde un punto de vista antropocéntrico), como algunos peces y anfibios, contuvieran una mayor cantidad de ADN que organismos ‘superiores’, como los humanos. Esta aparente discrepancia con la ideología de la época fue llamada “la paradoja del C-value”. Con el advenimiento de las tecnologías de secuenciación, esta paradoja fue disipada al tiempo que descubríamos que la cantidad de ADN no se correlaciona con el número de genes en el organismo. Aunque ahora entendemos que no existe una correlación entre la complejidad del organismo y la cantidad de ADN en su genoma, aún existen preguntas sin responder acerca de cómo surgió esta gran variación de tamaños del genoma entre diferentes taxa, o cuales, si existen sin las consecuencias de tener genomas particularmente grandes para el organismo. Las salamandras Neotropicales son un grupo diverso de anfibios que muestran una amplia variación en tamaño del cuerpo, abarcando desde menos de 20 mm hasta más d e160 mm en longitud hocico-cloaca máxima, y en tamaño del genoma, abarcando desde 9.3 pg hasta 81.1 pg de ADN por célula haploide. Debido al tamaño pronunciado de sus células, las salamandras miniatura (<35mm LHC) han sufrido varias modificaciones morfológicas extremas. Dada la cercana relación que existe entre el tamaño del genoma y el tamaño de la célula, esperaba encontrar una fuerte relación entre el tamaño del cuerpo y el tamaño del genoma en salamandras neotropicales. Utilizando el método de Análisis de Imagen por Densitometría de Feulgen estimé el tamaño del genoma de 54 especies de salamandras neotropicales sin estimaciones previas, y encontré el tamaño del genoma más pequeño jamás reportado en una salamandra en *Thorius spilogaster* (9.3 pg), y el genoma más grande reportado en una salamandra neotropical en *Bolitoglossa macrinii* (81.1pg). Utilicé el método de Mínimos Cuadrados Generalizados Filogenéticos para probar la correlación entre el tamaño del genoma y el tamaño del cuerpo y encontré que, aunque los clados con salamandras miniatura tienen genomas más pequeños en general, no hay una correlación entre estas dos características. También reconstruí los estados ancestrales del tamaño del genoma para hipotetizar la historia evolutiva de cambios que resultaron en la distribución de tamaños del genoma en clados de salamandras extantes y encontré que hay existe una fuerte señal filogenética en este clado. Los resultados de este estudio nos hacen reflexionar acera de las ideas acerca de la relación que existe entre el tamaño del genoma y el tamaño del cuerpo (en particular en lo que respecta a la miniaturización) y nos da un marco de trabajo para futuros estudios acerca de las dinámicas de la evolución del tamaño del genoma en salamandras Neotropicales y animales en general.

Genome size evolution in neotropical salamanders.

1. Introduction

1.1. The genome and the C-value paradox

All cellular organisms store their genetic information on DNA. The total amount of DNA contained in the nucleus of an organism, including all the genes, non-coding regions is known as the genome.

We can think of the genome as a long sequence of nucleic acids that can, in turn, be divided into multiple different elements that can be classified according to their functionality and/or position relative to other elements. Genes, usually referred to as the basic unit of heredity, are but one kind of many different types of DNA sequences. Despite being formed by the same four bases, their sequences make their functions and roles inside the genome vary widely. Our understanding of the different elements in a genome has improved a lot in the past century. Nonetheless, we are still far from completely understanding the genome as a whole, and the evolutionary history and composition of the genomes of countless organisms and groups remain unexplored. There are still many unanswered questions about the relationship between genomic features and traits at the organism level.

Even before it was discovered that DNA was the material of heredity, biologists made observations about the characteristics of the cell. A few years after the discovery of DNA in 1871 (Miescher-Rüsch, 1871), Gulliver (1875) noted that in vertebrates there was a huge variation in the size of erythrocyte nuclei, and that the size of the nucleus was correlated with the size of the cell. Mammals, the only group of vertebrates in which all species have enucleated erythrocytes, have very small erythrocytes, while all the other groups have cells proportional to the size of their nuclei.

During the early 1900s, scientists became interested in the nature of DNA. To understand this biomolecule, they would extract DNA from cattle tissue by dissolving large quantities of it in sulfuric acid. They would then quantify the amount of DNA in their cells by dividing the total amount of DNA yield by the number of cells in the tissue based on cellular suspensions. These early assays gave a rough estimate of the amount of DNA per cell, and soon researchers noticed that nuclei from different tissues contained the same amount of DNA, roughly twice of that present in sperm from the same species. This lead Vendrely and Vendrely in 1949 to report “a remarkable constancy in the nuclear DNA content of all the cells in all the individuals within a given animal species” (translation by T. Ryan Gregory) (Vendrely & Vendrely, 1949). This observation, which predated the blender experiment (Hershey & Chase, 1952) and the elucidation of the double helix (Watson, Crick, & Franklin, 1953), was fundamental to cement the idea that DNA was the material of heredity.

Other methods to quantify the amount of DNA in the cells of organisms soon emerged. One such method is the Feulgen Microdensitometry, in which the Feulgen reaction is used on fixed cells in a microscope slide. In this method, the DNA contained in the nuclei is depurinated by hydrolyzing it with a strong acid and the resulting free aldehyde groups are exposed to the Schiff reagent, which turns them pink (Feulgen & Rossenbeck, 1924). After staining the nuclei, the amount of light absorbed by the nuclei is measured. Because this staining is heterogeneous and individual nucleus vary, it is not sufficient to take a single densitometric measurement, and rather the sum of several measurements is compiled in what is known as the Integrated Optical Density (IOD) (Gregory, 2011). A more recently developed method to estimate the genome size, flow cytometry, works by extracting and staining isolated nuclei and measuring the amount of fluorescence they emit (Hare & Johnston, 2011). Both methods rely on having a standard species with known genome size to derive the genome size of the sample.

In the years that followed the first studies on genome size, several studies were conducted to determine the DNA content of different organisms. Hewson Swift established the constancy of DNA content between different tissues of the same species by analyzing various tissues of mouse, frog (*Rana pipiens*), grasshopper (*Dissoteira carolina*), maize (*Zea mays*), and several species of wild flowers (*Tradeschantia spp.)*. In his research, Swift classified DNA into two classes based on its amount. “Class I” DNA was the most common, corresponding to a diploid genome, and “Class 1C value” which represented haploid genomes (Swift, 1950). To this day, we still use the term C-value to talk about “the amount of DNA (in pg) within a haploid genome” or half the mount of DNA of somatic cells. This term and the term “genome size” are interchangeable when dealing with diploid organisms.

During the 1950’s the first broad survey of genome size for animals was conducted. Under the assumption that the amount of DNA correlated directly to the number of genes in an organism, Mirsky and Ris (1951) noted that:

> “Comparing the largest and one of the smallest examples among vertebrates, one finds that a cell of amphiuma [an aquatic salamander] contains 70 times as much DNA as is found in a cell of the domestic fowl, a far more highly developed animal. It seems most unlikely that amphiuma contains 70 times as many different genes as does the fowl or that a gene of amphiuma contains 70 times as much DNA as does one in the fowl.” (Mirsky & Ris, 1951).

Repeated analyses continued to baffle scientists for this apparent decoupling between genome size and organism complexity for over two decades, so much that in 1971 C. A. Thomas described this problem as the “C-value paradox”. The problem with this paradox was the assumption that DNA content was directly correlated with gene content. Under the assumption that the genome only contains genes, the extremely wide range of genome size among eukaryotes seems indeed paradoxical. With this discovery of additional genomic elements, the

C-value paradox was partially resolved within a few years of its inception. For a time, non-functional elements in the genome were classified as “junk DNA” (Ohta, 1973), but with the advent of the genomic era, we have now started to find some evidence that this junk DNA might not only be a useless byproduct of gene duplications or the proliferation of selfish elements, but might indeed have several functions, or at the very least, some consequences at the organism level.

While we now know that genome size does not correlate with organism complexity, the questions about the huge variation of genome size between taxa, the effects genome size has at an organismal level, and the mechanisms involved in the evolution of genome size remain an interesting topic of research. As Gregory (2001) stated, these questions should be addressed to understand the still relevant and unresolved “C-value enigma”.

1.2. Genome structure.

The genome of prokaryotes (Eubacteria and Archaea domains) is located in the cytoplasm of the cell without any kind of membrane enclosing it. Prokaryotic genomes are usually a single, long circular double stranded string of DNA ranging from about 0.5MB up to 10 MB (Cole & Saint-Girons, 1999). Eukaryotic genomes are contained within a double membrane organelle called the nucleus. The DNA inside the nucleus is organized in structures called chromosomes. The DNA in chromosomes is combined with histones and several other proteins to form a condensed structure called chromatin (Alberts et al., 2002). In most vertebrates, the genome is organized in pairs of chromosomes, making them diploid organisms.

The elements in the genome can be classified into two broad groups: functional elements (those with a described function) and non-functional elements (those having no described or apparent function). Most of the functional elements are genes, sequences that code for proteins and RNAs. Other functional elements include transcribed RNA molecules (tRNA, rRNA, RNAi, miRNAs, piRNA, etc) and structural DNA such as heterochromatin. Non-functional DNA includes pseudogenes, unique sequences, and a wide variety of **Repeated Elements (RE)**. Functional DNA makes up ~30% of the human genome. Repeated elements make up about ~45% of the human genome, with other non-functional elements making up the other ~15% (E. D. Green, Watson, & Collins, 2015).

RE can be further classified in different categories. Simple Tandem Repeats (STRs), or microsatellites, are elements formed by tandem iterations of mono-, di-, tri- or tetranucleotide sequence motifs repeated 5–50 times. Minisatellites and satellite DNA have longer sequence motifs that range from 10–30bp for the first, and hundreds or thousands of bp for the latter (Ellegren, 2004). Another kind of RE are Transposable Elements (TE). TE can change their location in the genome using different mechanisms, and thus can be considered mobile elements. Because they can proliferate in the host genome without any apparent benefit to the

organism, they are considered selfish elements. TE are divided into two classes depending on the mechanism is involved in their transposition.

Class I TE are called retrotransposons. They transpose via an RNA intermediate transcribed from a genomic copy which is then reverse-transcribed by a reverse transcriptase coded in the TE (Wicker et al., 2007). Because the DNA from which the RNA is transcribed remains in the genome, while the reverse-transcribed copy inserts itself in a new site in the genome, this kind of element is called “copy-and-paste” transposons. Given their tendency to proliferate, they make up a large fraction of the repetitive regions of many large genomes.

According to Wicker’s 2007 classification, Class I transposons can be subdivided in 6 Orders depending their mechanistic features, organization, and reverse transcriptase (RT) phylogeny: *Dyctyoselium* Intermediate Repeat Sequence (DIRS); Long Interspersed Nuclear Elements (LINEs); Short Interspersed Nuclear Elements (SINEs); *Penelope*-like Elements (PLEs); Terminal Inverted Repeat (TIR); Long Terminal Repeat retrotransposons (LTRs).

Long Terminal Repeat (LTR) retrotransposons can be extremely large (up to 25kb), are flanked by repeated sequences from a few hundred to 5000 bp and start with 5’-TG-3’ and end with 5’-CA-3’. They typically contain Open Reading Frames (ORFs) for GAG, a structural protein of virus-like particles, POL, an aspartatic proteinase, reverse transcriptase, RNAse H, and DDE integrase (INT). They can also contain an ORF with unknown function (Neumann, Požárková, & Macas, 2003). LTR retrotransposons aren’t very abundant in most animals but are the predominant order in plants.

There are five LTR retrotransposons superfamilies; *Copia*, *Gypsy*, *BEL-Pao*, *Retrovirus* and *ERV*. Retroviruses are closely related to LTR retrotransposons, and it has been proposed that retroviruses evolved from a LTR retrotransposon of the *Gypsy* superfamily after it acquired an envelope protein and other proteins and regulatory sequences (Frankel & Young, 1998; Seelamgari et al., 2004). A retrovirus can be transformed into an LTR by inactivating or deleting the domains that enable extracellular mobility (Capy, 2005). No longer able to infect a new host, inactivated retroviruses rely on vertical transmission as means of propagation. Thus, Wicker placed these so-called endogenous retroviruses (*ERV*s) (Bannert & Kurth, 2006) in his system as a super family within LTR. The *Copia, Gypsy* and *BEL-Pao* super families are structurally very similar, although the *BEL-Pao* elements have been only detected in metazoans.

Elements of the *DIRS* order contain a tyrosine recombinase gene instead of an INT and their features indicate that they have a different integration mechanism than LTRs and LINEs. They can be found in several groups including green algae, animals, and fungi. PLE encode an RT that is more closely related to telomerase than to the RT of LTR. These elements were first detected in *Drosophila virilis* and although they have been detected in animals, fungi and plants they are not very widely distributed in these groups. For example, they have been found in conifers but no other gymnosperms, and it has been shown that the main lineage of the conifer

PLEs is closely related to that of arthropods, suggesting a transkingdom horizontal transposon transfer in eukaryotes (Lin, Faridi, & Casola, 2016). LINEs are several kb in length and lack LTRs. They can be found in most eukaryotic kingdoms. They vary in abundance in animal genomes, but usually predominate over LTRs. SINEs are non-autonomous and have a different origin to the other elements in Class I. They are small elements (80–500 bp) and rely on LINEs for trans-acting transposition functions, such as RT.

Class II TE are called DNA transposons. They are found in most eukaryotes in low to moderate numbers. Unlike Class I elements, they do not rely on a RNA intermediate to transpose. There are two subclasses in this class. Subclass 1 includes Terminal Inverted Repeats (TIRs), the traditional Class II elements. Their transposition mechanism relies on a transposase enzyme that recognizes the TIR and cuts both strands at each end. The gap is then closed by the DNA repair mechanisms and the double stranded transposon is inserted in a new position in the genome. These TE are usually called “cut-and-paste” transposons. Besides TIRs, Wicker includes the obscure *Crypton* TE as a second order in subclass 1. These TEs have only been found in fungi to date and contain a tyrosine recombinase and lack a RT, similar to *DIRS* retrotransposons. Subclass 2 includes several “copy-and-paste” TE that transpose by replication involving the displacement of only one DNA strand.

Given the relatively high percentage of TE in the average vertebrate genome, one might assume that C-value of a species would tend to increase over time. That assumption would be correct if these selfish elements were given free reign over their proliferation, but several pathways that suppress novel TE insertions involving small RNAs have been described among different organisms (Siomi, Sato, Pezic, & Aravin, 2011). On the other hand, DNA loss rate has also been proposed as a mechanism through which genome size can change over time (Petrov, 2002). Both factors must be considered to research the evolution of genome size.

1.3. The C-value enigma in vertebrates

All of the elements in the genome are subject to change. Indel mutations, gene duplications, chromosome inversions, and transpositions can be fixed in a species, either by natural selection or genetic drift and, in the long run, change its genome size. Different lineages of organisms follow different trends in their genome size distributions; some groups show little variation in genome size whereas others vary widely. In eukaryotes alone, genome size varies 200,000 fold (Gregory & Hebert, 1999).

Among vertebrates, genome size varies widely (Fig. 1), from 0.34 pg in the Bandtail puffer fish (*Sphoeroides spengleri*, family Tetradontidae) to an astonishing 132 pg in the Marbled lungfish (*Protopterus aethiopicus*, family Protopteridae). Genome sizes of vertebrates vary 350-fold, but not all taxonomic classes have this extreme range. Birds, for example, have very consistent and relatively small genome sizes, which range from 0.9 pg in the Black-chinned hummingbird (*Achilochus alexandri*) to 2.16 pg in the Ostrich (*Struthio camelus*). Mammals have a slightly

larger variation in genome size, with the Carriker's round-eared bat (*Lophostoma carrikeri*) having the smallest estimate at 1.6 pg and the red viscacha rat having the largest genome size reported for this group at 8.4 pg. Humans (*Homo sapiens*) have an slightly above average genome size for this group at 3.5 pg. The same pattern of variation can be found in crocodilians, lizards and turtles. Fishes in general, including jawless fishes, sharks and rays, have a wider distribution of genome size. Osteichthyes (not including lungfushes) have a genome size ranging from 0.34 pg in the bandtail puffer (*Spheroides spengleri*) to 7.25 in the Nile bichir (*Polypterus bichir bichir*). Lungfishes have the largest genome size among fishes, ranging from 40 pg to 132 pg. Amphibians also show wide variation in genome size, with the ornate burrowing frog (*Limnodynastes ornatus*) having the smallest genome size of the group with 0.95 pg and the gulf coast waterdog salamander (*Necturus lewisi*) having the largest genome at 120 pg.

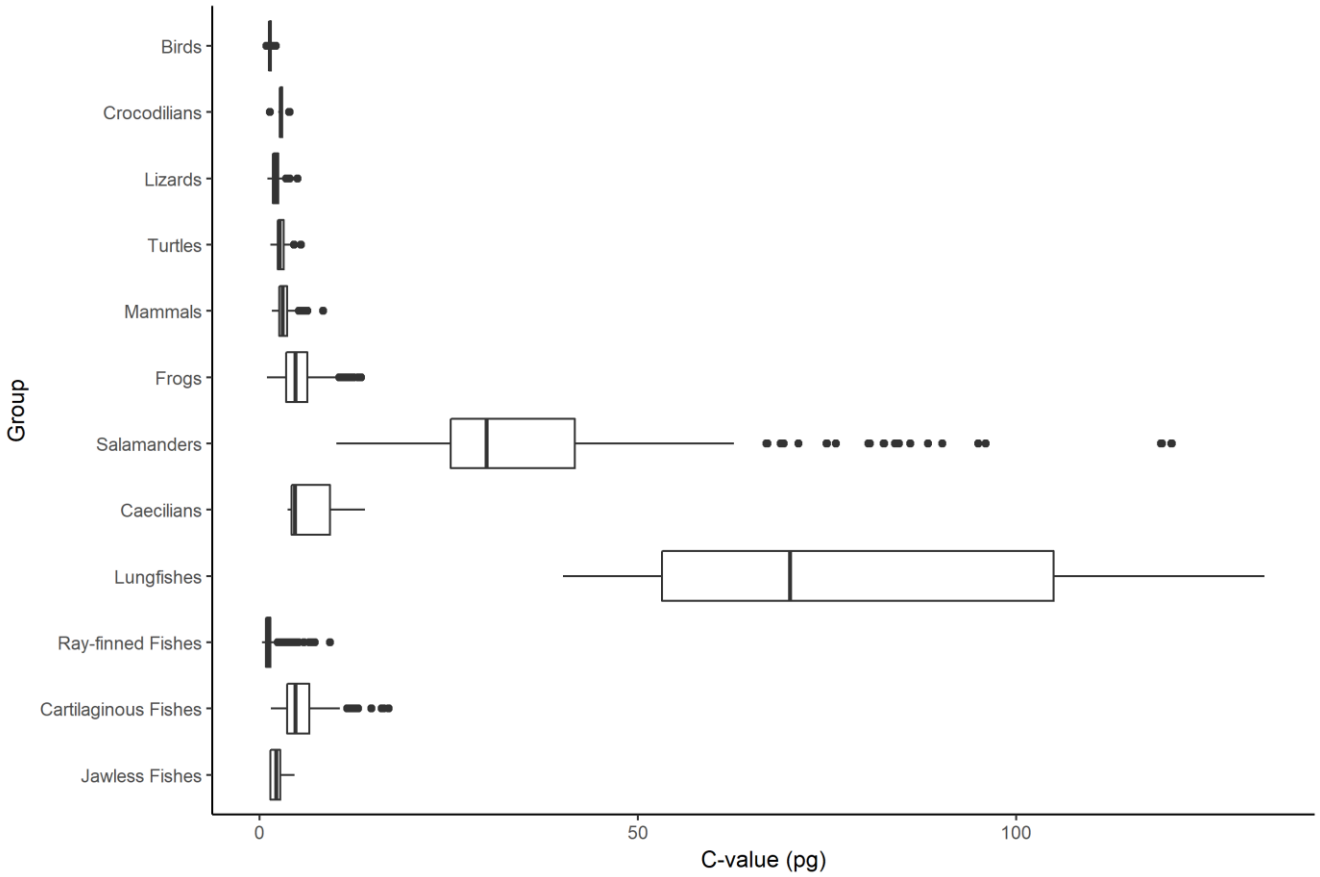

Figure 1: Genome size in picograms of DNA per haploid cell in different groups of vertebrates

The variation of genome size in vertebrates is not a trivial matter of biology, especially because the distribution of this trait appears to be non-random. Because we now know that genome size differs largely because of noncoding DNA, it might be assumed that differences in genome size are unimportant for the organisms. Besides its importance in sequencing projects (the larger the genome, the harder and costlier it is to sequence), genome size also has some interesting relationships with some organism traits.

As previously mentioned, early observations of the genome (in the form of the nucleus; Gulliver, 1875) of different organisms revealed that genome size has a strong positive correlation with cell size. The simplest example for this relationship can be found in species with polyploids. As the ploidy level increases, so does cell volume (Bogart, 1982). It is easy to measure the difference between the genome size of a diploid cell and a tetraploid cell in an organism, because one has half as much DNA as the other, but the relationship between genome size and cell area/volume has also been studied among different species. This relationship between genome size and cell area was observed more than 140 years ago in the red blood cells (RBC) of vertebrates. A positive, significant correlation ($R^2 > 0.82$, $p = 0.0001$) between erythrocyte size (measured as dry RBC area) and genome size and nuclei size was found among jawless fishes, cartilaginous fishes, bony fishes and lungfishes. In amphibians, this correlation persists, being stronger in salamanders ($R^2 = 0.58$, $p < 0.0001$) than in frogs ($R^2 = 0.51$, $p < 0.0001$). The same trend can be seen in reptiles, birds, and even mammals, despite having enucleated RBCs (Gregory, 2001b).

Genome size has also been found to be correlated with metabolic rate. Even though this characteristic is more closely related to RBC size, the fact that genome size is so significantly correlated with erythrocyte size, and the non-random distribution of genome sizes among vertebrate clades, leads to the assumption that genome size is inversely proportional to metabolic rate. In birds (class Aves), C-value and mass specific resting metabolic rate (RMR) have a significant negative relationship ($R^2 = -0.39$, $p < 0.005$) at the species level, although the relationship between active metabolic rate wasn't significant ($p > 0.18$) (Gregory, 2002).

Even though flying is a highly demanding metabolic activity for vertebrates, metabolic rate and genome size did not correlate at any levels among bats. Interestingly enough, like birds, bats also have small genome sizes, which have been shown to be constrained for mammals (Smith, Bickham, Gregory, & Bainard, 2013). On the other side of the spectrum, amphibians, which have significantly larger genome sizes, display a highly significant negative correlation in their C-values and the RMR ($R^2 = -0.75$, $p < 0.0001$) measured as mass-specific oxygen consumption at 15°C (Gregory, 2003) . This correlation does not hold up within frogs or salamanders taken separately, but when metabolic rate was measured at 25°C in salamanders, a robust (albeit small) negative correlation was found (Licht & Lowcock, 1991).

Cell division rate is another trait that has been long hypothesized to be correlated with genome size. Not only will cells with larger C-values contain more DNA to replicate, but this excess DNA also has to be folded and unfolded to complete the cell cycle. In plants facing certain ecological circumstances under which a fast mitotic division rate and a small cell area are favorable, selection leads to reductions in genome size and nuclear area, either by DNA loss or by an increase on the general heterochromatinization of the genome (Nagl, 1974). Under other, less favorable conditions like extreme cold or hot, selection has led to increased cell size and a reduced mitotic rate (Price & Bachmann, 1976). The mechanisms observed in plants regarding cell division rate provide circumstantial evidence for a negative correlation between cell division

rate and genome size. It is still not clear whether the effect of genome size in cell division rate is due to the amount of DNA or if it depends only on the level of compaction in the nucleus (Gregory, 2001a). The Hayflick limit is the idea that there is a limit to the number of times a cell can divide before its telomeres get to a critical short length and enters a senescence phase (Shay & Wright, 2000). Because of this, it is thought that species with larger genomes have both longer telomeres, allowing for more divisions, and a slower division rate that would also increase the life span of a cell lineage Thus, a genome size correlation with division rates would be concordant with the empirical observations of an increased longevity in species with larger genomes, such as salamanders and lungfishes.

The influence of genome size on organismal traits does not stop at the cellular and molecular level. Genome size and cell size have been shown to be strongly correlated in plants, one of the best examples being the strawberry fruit, which is an octaploid or a decaploid. The wild variety of the strawberry fruit (not an polyploid) is significantly smaller. This trend is much less obvious among animals, although it has been shown that there are positive significant correlations between genome size and body size in several groups (Jeffery, Ellis, Oakley, & Gregory, 2017). In mammals this relationship can be found among rodents, and in birds the relationship can be seen in hummingbirds. In bats, however, genome size and body size aren't correlated (Smith et al., 2013). In salamanders of the genus *Eurycea*, there is a clear relationship between genome size and body size in artificially induced polyploidy larvae (Levy & Heald, 2016). In treefrogs of the *Hyla versicolor* complex, the tetraploid species possesses larger adhesive digital pads than its diploid sibling species, although there are no differences in body overall body size (D. M. Green, 1980)

Lungfishes and salamanders have the widest variation range in genome size among vertebrates and studying the underlying relationships behind genome size and other traits in these two groups could help us understand better the mechanisms and processes involved in the evolution of genome size in vertebrates. Given the enormous size of the genome in these two clades, it is sensible to assume that any phenotypic effects that genome size could have on an organism would be exacerbated in lungfishes and salamanders. Indeed, these two taxa exhibit most of the traits one would expect of species with large genomes, such as a reduced metabolic rate and relatively long lifespans. Lungfishes have the largest genome size among vertebrates, but there are only six extant species distributed in South America, Africa and Australia. Salamanders, on the other hand, have a greater relative variation in genome size, and there are over 700 species distributed mostly in the northern temperate zone, as well as in Central and South America. Given their diversity, salamanders are better suited as a model to study the evolution of genome size.

1.4. Plethodontid salamanders

Salamanders are amphibians of the order Caudata, and with 707 described species in the world, they represent 9% of all amphibian species. They are distributed through Eurasia and

America, where many of the salamander species of the world can be found. There are ten families in this order, the largest of which is the lungless salamanders (family Plethodontidae). Salamanders typically have a slender lizard-like body, with a tail present through all the stages of their lives and four limbs in all species except those of the family Sirenidae. Salamanders of the family Plethodontidae lack lungs and conduct respiration entirely through the skin (AmphibiaWeb, 2017).

There are 468 described species of plethodontid salamanders in the world, and with the exception of six species that occur in Europe and one species in South Korea, they exclusively inhabit the new world (AmphibiaWeb, 2017). They can be found from southern Canada, through the Eastern and Western US, Mesoamerica, and south to Bolivia and Brazil. The plethodontid salamanders that inhabit Central and South America represent nearly all the tropical salamanders in the world, and they have diversified to an impressive degree. Tropical plethodontids are classified in a single inclusive taxon, the tribe Bolitoglossini tribe (Wake, 2012), which includes 294 species, roughly 40% of all the salamanders in the world (Rovito, Parra-Olea, Recuero, & Wake, 2015).

Besides having a huge range of genome sizes, varying almost ten-fold, plethodontid salamanders also exhibit great interspecific variation in body size. Salamanders of the genus *Isthmura* can be as large as 160 mm in Snout-Vent Length (SVL) while some species of the genus *Thorius* can be as small as 18 mm. The variation in body size of plethodontid salamander species represents an interesting model to study different evolutionary processes involving this trait, among the most interesting of which is miniaturization. The evolution of an extremely small body size is widespread in animals, and in salamanders there have been several independent origins for the miniature phenotype. There are at least 10 genera of plethodontids containing species below a critical size of 35 mm of SVL, set up as the standard length for miniature salamanders (Hanken & Wake, 1993a).

There are several ecological and evolutionary hypotheses for the occurrence of this phenotype. One of the most obvious advantages of reduced body size is a better ability to hide from predators. Being able to hide in small spaces improves the chances of survival against predation. The smaller body size of miniature salamanders also allows them to occupy otherwise empty ecological niches that bigger species cannot, such as small epiphytes, small cracks and the space between the bark and wood of logs (Hanken & Wake, 1993b). Finally, another hypothesis proposed by Feder (1982) is that miniature salamanders can regulate their temperature by selecting environments with an optimal temperature in a process called behavioral thermoregulation. Feder's hypothesis derived from his observations of miniature salamanders (genus *Thorius*) living between the bark and the wood of fallen logs. These miniature salamanders moved through the log looking for the place with optimal temperatures, which larger salamanders are unable to do because of a lack of thermally diverse, moist environments (Feder, 1982).

Besides these ecological advantages, miniaturization in salamanders also has several consequences. The reduction of body size can have ramifications on practically all aspects of an organism's biology. One of the most common effects of miniaturization is the reduction and structural simplification of morphology (Hanken & Wake, 1993b). In salamanders of the genus *Thorius*, this is exemplified by the morphological reduction and loss of several head bones. The otic capsules, eyes, and brain also occupy a bigger portion of the skull than in non-miniature salamanders (although they have not been reduced in proportion to the rest of the body). There have also been several rearrangements inside the skull of miniature salamanders, with one of the most dramatic being the displacement of the brain further back in the skull. Several of the sensory organs, including the eyes, are much closer to the brain as well. The nasal capsule occupies most of the snout, and most species have lost their maxillary teeth (Hanken, 1983b).

1.5. Salamanders' genome size and its consequences

Plethodontid salamanders have a highly conserved chromosome number, with either 2n = 26 or 28, while some other salamanders, like those of the family Hynobiidae, can have from 38 to 78 chromosomes of diverse sizes. Although chromosome number varies, there is no direct correlation between chromosome number and genome size; rather, species within a family with larger genomes tend to have longer chromosomes (Sessions, 2008). All species of neotropical salamanders have 13 pairs of chromosomes that vary in arm length, although their genome size is extremely varied. This observation is suggestive of massive gains and/or losses in DNA content throughout the chromosomes over the course of the evolution and diversification of this group, rather than through gains or losses of chromosomes through genome duplication or other mechanisms.

As with the examples given previously, this massive variation in genome size in salamanders appear to have several phenotypic consequences in salamanders beyond their chromosome size. Observing the different developmental strategies in different salamander families (Sessions, 2008), a trend seems to emerge. Families with larger genomes tend to be strongly paedomorphic (larvae become sexually mature) or direct developing (with no larval stage), while the metamorphic species appear to have smaller genomes overall. In plethodontids this is even more evident, with some species (those with smaller genomes) having aquatic larvae and complete metamorphosis, and the rest being direct developers. One possible explanation for this phenomenon is that species with larger genomes would have to spend much more time (because cell division rate is affected by genome size) and resources on an already complicated and metabolically costly process, while species with smaller genome sizes have had the chance to retain the ancestral biphasic life history characteristic of amphibians. Although there are some strongly paedomorphic species with medium genome sizes (for salamanders) like the axolotl (*Ambystoma mexicanum,* C-val = 32), they usually can undergo metamorphosis without any problem after hormone treatment, while this is not true for species with larger genomes.

The impact of genome size on development extends beyond life history of salamanders. All salamanders can regenerate their limbs, and an experiment conducted by Sessions and Larson (1987) estimated the growth rate and differentiation rate by the amputation of the right hind limb of several species of plethodontid salamanders with different genome sizes. The absolute growth rate was estimated by dividing the measured area of regrowth by the number of days since amputation, while the relative growth rate was calculated by dividing the absolute growth rate by limb diameter. Both of these measurements had a negative correlation with genome size. Differentiation rate, determined by the morphology of the regenerating limb, was also determined and had a negative relationship with genome size (Sessions & Larson, 1987). These observations are consistent with the hypothesis that the larger genome cause slower cell replication times. Since lungfishes also present limb (fin) regeneration abilities, equivalent to those seen in salamanders (Nogueira et al., 2016), it has been hypothesized that genome size could be correlated with this trait.

It has been long known that nucleus size is highly correlated with cell size (Gulliver, 1875), and we now know that genome size is tightly correlated with nucleus size. The size of the cells can have major implications for the development of certain structures and systems within the organism. One of the most important systems in vertebrates is the nervous system. Compared to amniotes, amphibians and lungfishes (which have large genome sizes) do not differ greatly in their spinal cord, medulla oblongata, midbrain, and the preoptic-hypothalamic diencephalic region. But they show major differences in the thalamopallial system, visual system, pallial regions, striatopallidum and amygdaloid complexs. Even compared to fishes, amphibians are thought to have relatively simple brain morphology (Dicke & Roth, 2010). Large genome size has a strong positive correlation with nerve cell size, both in frogs and salamanders. The complexity of the tectum mesencephali—a part of the nervous system responsible for auditory and visual reflexes—judged by the number of alternating cell and fiber layers, has a significantly negative correlation with genome size in both frogs and salamanders. The trend is also persistent in the torus semicircularis, another multi-tissue structure in the midbrain involved in major audiomotor interfaces and ascending sensory input. The cerebellum, formed by the corpus cerebelli and two lateral auricular lobes, is very small in amphibians and also has a negative correlation with genome size. In small-sized species, the corpus cerebelli remains paired with the auricles, and in the smallest salamanders, found in the tribe Bolitoglossini, it has partly retreated under the tectum (Roth & Walkowiak, 2015).

The secondary simplifications in the cranium and nervous system of miniature salamanders, coupled with the large nervous cell size that genome size causes, poses an interesting challenge for miniature species. Because plethodontid salamanders are active foragers that use their projectile tongue to catch food, the eyes are an extremely important organ. Because space is reduced, and cells are large, the number of cells involved in vision is severely restricted. The number of photoreceptors and retinal ganglion cells in the eye and of visual neurons in the brain is believed to be decisive in the correct functioning of eyes (Roth, Rottluff, & Linke, 1988). To avoid loss of function in the eyes, miniature salamanders can do one of two

things: they can maintain eyes that are proportionally larger to the skull than other salamanders, to be able to sustain the number of cells needed for a functioning, accurate vision, or they can reduce their genome size in order to have smaller cells.

The difficulties for miniature salamanders with big genome sizes extend beyond the nervous system and the eyes. Blood vessels in minute species can be extremely narrow, restricting the passage of RBCs potentially leading to the formation of blood clots. For this reason, several species have evolved a high percentage of enucleated RBCs, a characteristic usually restricted to mammalians among vertebrates. Although not all the cells are enucleated in miniature salamanders, the percentage of enucleation is significantly higher among miniature salamanders compared to their non-miniature relatives. Among small salamanders of the genus *Batrachoseps*, enuclation is more severe among those species with larger genomes (Mueller, Ryan Gregory, Gregory, Hsieh, & Boore, 2008).

Although the consequences of genome size on salamander physiology, development, and life history are evident and have been extensively studied, questions of how these amphibians ended up with such a large and varied genome size and how selection acts on this trait remain unanswered. Given the difficulties of sequencing and completely assembling genomes of such large size, the molecular basis for the increases and reductions on this trait have proven to be difficult. Still, using low coverage genome sequencing, Sun et al. (2012), explored the amount of RE present in salamander genomes. They compared the genomic makeup of six species of plethodontid salamanders and found that their genomes have a significantly higher percentage of transposable elements compared to other vertebrates. The main family of TE present in the species analyzed are those of the LTR type (Sun, Shepard, et al., 2012). The proliferation of this particular TE class on salamanders' genomes begin to explain their genomic gigantism, but it still does not explain the huge interspecific variation featured on these amphibians.

The runaway expansion of TE in salamander genomes and the apparent disadvantages of having a huge genome open the question of why selection did not act against the expansion of genome size. One hypothesis proposes that plethodontid salamanders lack TE silencing pathways conserved among vertebrates, which in turn would allow the proliferation of these selfish elements. Although this would be a compelling explanation, a recent study shows that the piRNA pathway, responsible for silencing novel TE insertions, is present in *Desmognathus fuscus* and is likely conserved and expressed in other salamanders (Madison-Villar, Sun, Lau, Settles, & Mueller, 2017). There might be other molecular processes that could inactivate piRNA's, salamanders' TE might have evolved an independent mechanism to circumvent this silencing mechanism, or there might be a strong, yet unseen, positive selection for the huge genome phenotype. Increasing the amount of sequence data available could help us postulate novel hypothesis for the high load of TE in salamanders. Along with possible TE expansions, salamanders seem to possess particularly efficient DNA repair mechanisms, which could be a factor for the large genomes they possess. DNA loss rate (bp lost per substitution) in non-LTR

elements can be half as much as that of humans and can be as small as 6 times less than in species like *Xenopus tropicalis* and *Danio rerio* (Sun, López Arriaza, & Mueller, 2012).

There is no doubt that the extreme genome size present in salamanders has major consequences at multiple organismal and systemic levels. These consequences are particularly evident in species with reduced body sizes because of constraints on the minimum size of functional sensory and information processing organs. The salamanders of the tribe Bolitoglossini, with their huge interspecific variation in both genome and body size, provide an excellent model to examine the impact of miniaturization and body size evolution on the evolution of genome size, with its myriad consequences at the organismal level.

2. Hypothesis

Given the morphological simplifications on the nervous system of neotropical salamanders and the relationship between genome size and cell size, as well as the independent origins for a miniature phenotype, I expect miniature salamanders to have a relatively small genome size.

There should be strong selection for small genome size in miniature species because of the genome-cell size correlation. There must exist an upper limit for how big the genome of a miniature salamander can get before the organism can no longer achieve miniaturization (i.e. have functional sensory systems). On the other hand, non-miniature salamanders need not have a lower limit on their genome size, so I expect a wider distribution of genome size in larger species than in miniature ones.

3. Goals

3.1. Main goal

Analyze the relationship between genome size and body size and its evolutionary implications on neotropical salamanders in a phylogenetic context.

3.2. Specific goals

a) Collect a representative sample of neotropical salamanders, including miniature species.

b) Estimate genome size using Feulgen Image Analysis Densitometry.

c) Test for a correlation between body size and genome size in a phylogenetic context using phylogenetic comparative methods.

4. Materials and methods

4.1. Sampling

All salamander specimens used in this project were collected by Dr. Sean M. Rovito or members of the Vertebrates Genomics and Biodiversity Lab between 2013 and 2017. These salamanders were all collected in Mexican territory (figure 2) with a collecting permit (Number 06720 and 06329) issued by the Secretariat of Environment and Natural Resources (SEMARNAT). All animals were treated under international ethics conventions.

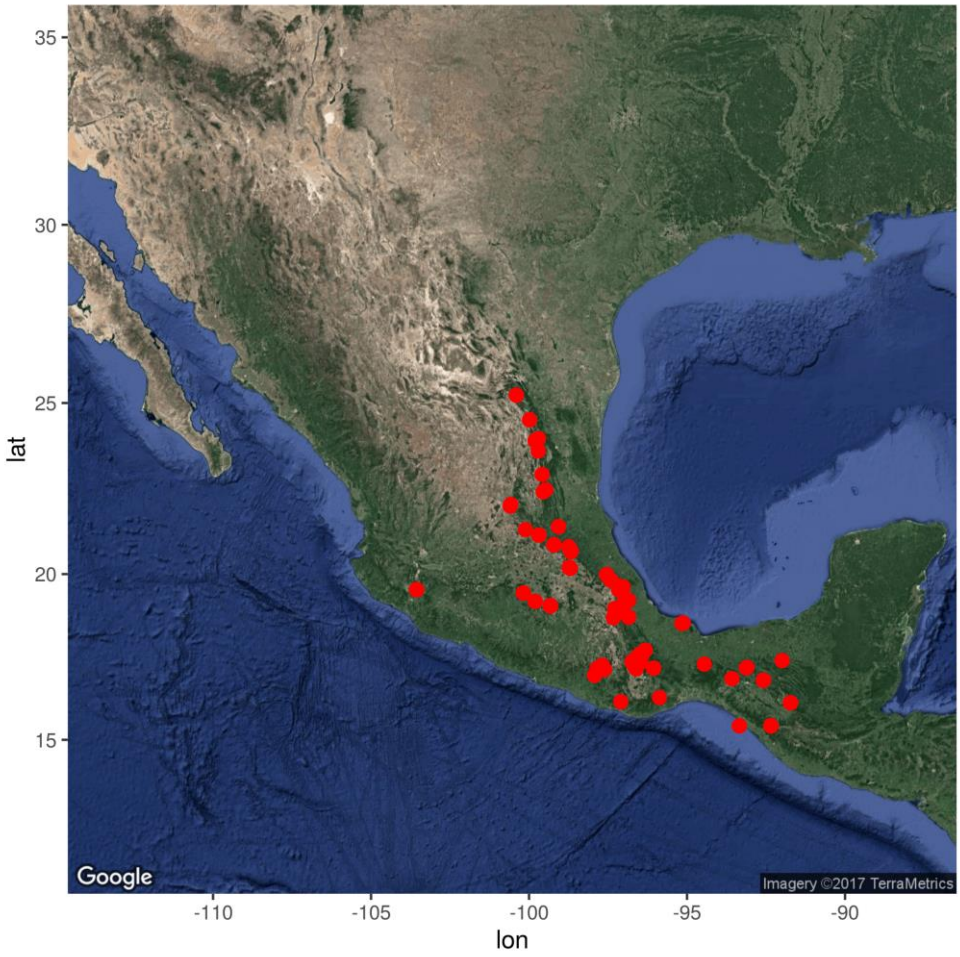

Figure 2: Specimen collection sites

The sampling effort for this project was aimed at salamander species without genome size estimates, especially miniature salamanders of the genus *Thorius*. Because many of these salamanders are endangered, and some of their populations have been extirpated, there are some species that are more abundant than others. Furthermore, some species are more widely distributed (such as *A. cephalica* and *P. leprosa*) than others, and thus are overrepresented in the dataset compared to rare or narrowly endemic species.

Each salamander was collected, anesthetized using a chlorotone solution until it no longer responded to stimuli, and dissected. Internal organs, heart, liver, intestine, blood vessel, and testes and mental gland (in the case of males) were removed preserved in a RNAlater solution using liquid nitrogen. These samples are stored at -80 °C. During the dissection of each salamander, a blood smear was prepared on a glass slide as indicated in Hardie et al. (2002). In some cases, when the salamander used was too small or had too little blood to place a drop on the slide, a small circular smear would be done with the heart. The specimen would then be fixed using formalin and after two days would be stored in 70% ethanol. In the case of the premature death of a salamander, dissection and slide preparation was conducted on site.

I estimated genome sizes for 176 samples from 66 different species of 8 different genera; *Aquiloeurycea*, *Bolitoglossa*, *Chiropterotriton*, *Cryptotriton, Isthmura, Parvimolge, Pseudoeurycea,* and *Thorius*. Of the 66 species for which genome size was estimated, 12 had a previously reported genome size in the animal genome size database.

4.2. Body size measurements

Maximum snout-vent length data for 291 species was provided by Dr. Sean M. Rovito. These data were used as a measure of body size for salamanders. I took three additional morphological measurements for 587 specimens of 72 different species. Head width, head length (snout-gular length) and head depth at angle of jaw were measured using a digital caliper. These measurements were multiplied to obtain a rough head volume estimate to use as a different body size proxy given the importance of the cranial size in miniaturization processes (Hanken, 1983a). The mean for each measurement by species was calculated and used for PCMs.

4.3. Feulgen Image Analysis Densitometry

This methodology is modified from Harden (2002). This part of the project was conducted under the supervision of Dr. Liljana Bizmjak Mali and Dr. Ales Kladnik from the University of Ljubljana, Slovenia, and Dr. Stanley K. Sessions from Hartwick College in New York.

Erythrocyte nuclei in blood smear slides were stained the following way:

- **Fixation:** Slides were air dried for several hours. They were then fixed in a 100% methanol solution for 3 minutes.
- **Hydration**: Methanol residues were removed with distilled water for 3 min.
- **Hydrolysis**: Slides were placed in 5N HCl for 20 min at room temperature (~20–25°C).
- **Rinse**: Slides were washed with running distilled water for 1 min three times.
- **Staining**: Samples were stained using Schiff's Reagent for 90 min at room temperature and protected from light.
- **Rinse**. Slides rinsed as above.

- **Mounting**: Slides were dehydrated doing 70%, 90%, 100% ethanol changes (1 min each), followed by two 5 min changes in Xylene. Samples were mounted with permount and a cover slip.

Each batch of 18 slides was stained simultaneously with either a *Bolitoglossa platydactyla* or an *Ambystoma mexicanum* slide to use as a reference, and an internal control of *Ambystoma velascii* for some batches.

Stained slides were checked under a light microscope (Fig. 3) to avoid problems with poorly or incorrectly stained slides. Using the KS10D 3.0 microscope software and the macros developed for it by Dr. Barbara Vilhar (University of Ljubljana), I obtained IOD (Integrated Optical Density) measurements derived from the number of pixels and the intensity of the staining of 40–300 (mean 121.2) nuclei per sample. I used several criteria to select which nuclei to measure. These characteristics included the wholeness of the nucleus and the cell, its proximity to other nuclei and cells (to avoiding overlapping), the morphology of the nucleus. Salamander erythrocytes generally have oval shaped nuclei that can be damaged during the slide preparation, fixing or staining, although some salamanders have slightly “wrinkled” nuclei. Because salamander nuclei are very large, there can be variations in IOD measurement at even slightly different levels of focusing. A subjectively good and consistent level of focus was desired. The illumination source for the microscope, a halogen lamp, was tested prior to each use of the equipment to check that there was no variation in light intensity, which could affect the IOD measurement. Calibration was performed using the reference slide and a series of gray filters and glare correction were used.

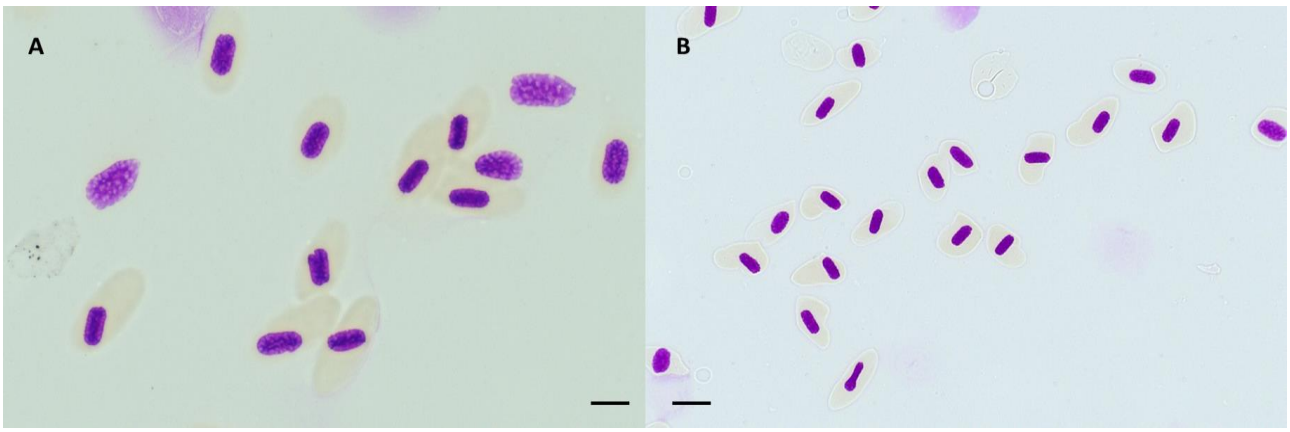

Figure 3: Stained nuclei of A) Aguiloeurycea galeanae and B) Thorius lunaris. Scale bar 20 µm

4.4. Genome size estimation

All statistical analyses were performed using the R software (R Development Core Team, 2017).

Given the large size of erythrocyte nuclei in salamanders and previously mentioned problem with microscope focusing, the distribution of IOD data was sometimes very wide with some extreme outliers. To account for this source of variability, I eliminated the measurements on the upper and lower 5% of the distribution. These outliers sometimes had over twice the mean IOD of other nuclei, and likely were in the process or meiosis, while others had substantially lower IOD and may have been poorly stained.

After narrowing the distribution of IOD data and testing for normality, we calculated the mean, median, standard deviation and coefficient of variation (CV) for each sample. Samples with a CV larger than 10% were eliminated from genome size estimation because this variation likely reflected problems with the staining of the sample. After this data pruning we were left with 176 samples of 66 different species.

Using the following formula I estimated genome size (Hardie, Gregory, & Hebert, 2002):

$$C_u = C_s \times \left(\frac{IOD_u}{IOD_s}\right)$$

Where *Cu* is the C-value of the unknown sample, *Cs* is the C-value of the standard sample (in pg), *IODu* and *IODs* are the integrated optical densities of the unknown sample and standard sample respectively. Once all C-value estimates were calculated, I estimated mean and standard deviation of C-value by species.

Mean, median, standard deviation, maximum and minimum measures were obtained for body size measurements and genome size estimates by species, genera, and for all the specimens. A Pearson's product-moment correlation coefficient was computed to assess the relationship between the log-transformed mean genome sizes and the log transformed mean maximum SVL and head volumes as proxies of body size of all the species for which both morphological estimates were available.

Using an ultrametric Maximum Likelihood phylogeny estimated using the 16S and cytochrome b mitochondrial genes, with the topology constrained to match that of the species tree from Rovito et al. (2015), I tested four different models of evolution for genome size and selected the best one using the Akaike Information Criterion. The four models tested were: Brownian Motion (BM), which assumes a "random walk" model in which a trait value can change randomly in both directions; the Ornstein-Uhlenbeck (OU) model, also a "random walk" model that assumes the evolution of a trait is bounded by an optimal value to which the character will revert over time; the Early Burst model (EB), which assumes there is an early radiation in character traits that gives rise to all the diversity within a group; and the White Noise (non-phylogenetic) model, in which it's assumed that character traits evolve in a completely stochastic process.

When testing correlations between phenotypic characters among organisms, it is important to account for the fact that shared ancestry can be potentially misleading in statistical analyses that treat data points as independent. Particularly in biology, the result of a simple correlation analysis on data from related species could produce spurious results, with correlations due to shared ancestry alone. Such is the importance of using phylogenetic comparative methods. Unlike other kinds of correlation analysis in which points are considered independent, closely related species can't be treated as such. The shared ancestry and evolutionary history of biological organisms can shroud the true relationship (or lack thereof) between phenotypical traits.

For this reason, following the development of methods to estimate phylogenies from molecular data as a means to represent evolutionary relationships, a new set of statistical methods was born. Phylogenetic comparative methods (PCMs) consider the shared history of organisms when calculating correlations. One of the most commonly used PCMs, Phylogenetic Generalized Least Squares (PGLS), is described by Grafen (1989) as a generalization of Felsenstein's (1985) independent contrasts. Both analyses recognize the problem of non-independence of species, and the results from both methods, in their raw form, are fundamentally the same (Symonds & Blomberg, 2014). This is the main reason for only reporting the results of PGLS and not those of phylogenetic independent contrasts (PIC).

The logic behind both approaches is that although species do not represent independent data points, differences ('contrasts') between closely related species are independent as they represent independent portions evolutionary of evolutionary history on the phylogeny. One advantage PGLS has over the PIC is that the first method doesn't rely on the assumption that closely related species will necessarily be similar. Under a Brownian motion model of evolution this is expected; species with a more recent common ancestor are expected to be more similar than more distant species because their traits will have less time to diverge. Since there are some situations in which traits can be evolutionary liable, having a method that can incorporate different evolutionary models such as PGLS is desired. In my analysis I found that a Brownian motion model is the best fit for genome size in Neotropical salamanders (see Results), but without this information *a priori*, it makes sense to use PGLS instead of PIC.

Because the Pearson's correlation treats each data point as independent, and all neotropical salamanders share a common evolutionary history, I used PGLS (Symonds & Blomberg, 2014) to account for the shared ancestry of the species analyzed. PGLS was estimated using the R packages 'phytools' (Revell, 2012), 'ape', 'nlme' and 'geiger'. Finally, I estimated the ancestral state of the C-value at the internal nodes of the phylogeny under the best evolutionary model, (selected using the AIC) using a maximum likelihood approach and 100 replicas using the R package 'phylotools' in addition to the packages used in previous steps.

5. Results

5.1. Genome size estimates

I estimated genome size for 176 specimens of 66 different species (Table 1), 54 of which did not have previous genome size estimates reported on the animal genome size database (www.genomesize.com). I analyzed an average of 2.7 specimens per species. The C-values estimated range from 9.3 pg in *Thorius spilogaster* to 81.1 pg in *Bolitoglossa macrinii*. These estimates represent the smallest and largest genome sizes estimated for any plethodontid salamander to date; *T. spilogaster* has the smallest genome size estimated for any salamander species. Previously, *Desmognathus wrighti* had the smallest genome size with an average C-value of 13.8 and *Hydromantes italicus* had the largest, with an average C-value of 71.6. The genome size estimates obtained in this study are comparable to previous estimates, varying by less than 10 pg in 12 of 13 species in common between my study and previous estimates. The only exception was *Bolitoglossa hartwegi*, which has a reported genome size of 42.0 pg and had an average genome size almost twice as big of 79.61 pg (n = 2, sd = 2.48) in this study.

Table 1: Genome size estimates of neotropical salamanders where: n = number of specimens, C-value in picograms and sd = standard deviation.

Species (OTU)	**n**	C-value	**sd**
Thorius spilogaster	1	**9.3**	NA
Thorius aureus	2	**9.6**	0.26
Thorius sp. San Juan del Estado	2	**12.4**	0.56
Thorius sp. 7	2	**13.6**	1.61
Thorius sp. Ixtlan	4	**15.0**	3.73
Bolitoglossa veracrucis	1	**16.0**	NA
Chiropterotriton orculus	2	**16.7**	7.65
Thorius troglodytes	1	**17.3**	NA
Thorius sp.	1	**19.2**	NA
Thorius pulmonaris	1	**19.3**	NA
Parvimolge townsendi	4	**20.0**	1.56
Thorius boreas	2	**20.2**	1.89
Thorius macdougalli	10	**21.4**	4.74
Chiropterotriton cf arboreus	3	**21.7**	3.68
Chiropterotriton terrestris	5	**21.8**	5.26
Chiropterotriton sp. "Pinguica"	1	**22.2**	NA
Chiropterotriton sp. H	2	**22.7**	4.02
Thorius narisovalis	4	**22.8**	1.93
Chiropterotriton multidentatus	5	**23.2**	3.82

Pseudoeurycea juarezi	2	**23.5**	8.61
Chiropterotriton arboreus	1	**24.7**	NA
Thorius lunaris	1	**24.9**	NA
Chiropterotriton sp. "pluvialis"	3	**25.9**	1.93
Pseudoeurycea saltator	3	**26.4**	8.6
Chiropterotriton chondrostega	1	**26.6**	NA
Chiropterotriton lavae	2	**27.7**	2.8
Pseudoeurycea leprosa	5	**27.9**	3.35
Pseudoeurycea lineola	4	**28.0**	1.35
Thorius tlaxiacus	4	**28.9**	1.47
Chiropterotriton dimidiatus	10	**30.3**	4.71
Cryptotriton alvarezdeltoroi	1	**30.6**	NA
Pseudoeurycea orchimelas	3	**32.2**	1.98
Pseudoeurycea smithi	1	**32.2**	NA
Isthmura gigantea	3	**32.2**	1.55
Chiropterotriton sp. nov.	2	**32.8**	2.84
Chiropterotriton infernalis	1	**33.2**	NA
Chiropterotriton miquihuanus	2	**33.2**	6.49
Pseudoeurycea mixteca	1	**33.2**	NA
Aquiloeurycea quetzalanensis	3	**36.7**	4.07
Chiropterotriton chico	1	**37.7**	NA
Aquiloeurycea sp. "wakei"	2	**38.2**	0.88
Pseudoeurycea conanti	2	**38.4**	6.34
Pseudoeurycea werleri	1	**39.9**	NA
Aquiloeurycea cephalica	10	**41.8**	4.16
Pseudoeurycea gadovii	1	**41.8**	NA
Pseudoeurycea altamontana	1	**42.0**	NA
Aquiloeurycea sp. "potosina"	2	**42.0**	0.05
Bolitoglossa zapoteca	1	**42.2**	NA
Bolitoglossa rufescens	4	**43.5**	4.39
Pseudoeurycea sp.	7	**43.7**	7.59
Pseudoeurycea longicauda	2	**44.1**	0.8
Pseudoeurycea melanomolga	1	**44.3**	NA
Aquiloeurycea sp.	5	**44.6**	1.68
Bolitoglossa sp. nov	1	**44.7**	NA
Aquiloeurycea galeanae	4	**44.9**	4.06
Aquiloeurycea scandens	2	**45.0**	0.36
Isthmura bellii	9	**47.8**	7.08

Pseudoeurycea robertsi	2	**48.8**	1.37
Bolitoglossa rostrata	1	**50.4**	NA
Bolitoglossa riletti	1	**57.9**	NA
Bolitoglossa occidentalis	1	**59.9**	NA
Bolitoglossa mexicana	1	**68.8**	NA
Bolitoglossa stuarti	3	**76.6**	3.49
Bolitoglossa franklini	2	**77.7**	2.59
Bolitoglossa hartwegi	2	**79.6**	2.48
Bolitoglossa macrinii	1	**81.1**	NA

In addition to the genomes sizes estimated in this project, I consulted the genome sizes of the rest of the family Plethodontidae (Figure 4) from the Animal Genome Size database (data provided in supplementary table SR1). The average C-value for 156 species of plethodontid salamanders (n = 360) is 35.2 pg, with the smallest genome size at 9.3 and the largest at 81.15 (previously mentioned). The genus *Bolitoglossa* (n = 27 spp.) has the largest genome sizes of the plethodontid family, with an average of 55.1 pg. It also has the biggest variation in genome size. *Bolitoglossa veracrucis* has the smallest genome size of this genus at 16.0 pg, 2.6 times smaller than *Bolitoglossa zapoteca*, which has the second smallest genome at 42.2 pg, and 5 times smaller than *Bolitoglossa macrinii* at 81.1 pg. The genus *Desmognathus* (a temperate genus; n = 6 spp.) has the smallest average genome size at 15.9 pg (min *Desmognathus wrighti* 13.8 pg (Sessions & Larson, 1987), max *Desmognathus ochrophaeus* 15.5 pg (Licht & Lowcock, 1991)). Even though *Desmognathus* has the smallest average genome size, the genus *Thorius* (n = 14 spp.) has the species with the smallest genome size (*Thorius spilogaster* 9.3). With an average genome size of 18.57, all the salamanders from the genus *Thorius* have relatively small genomes, and it has the smallest genomes for all the order Caudata. *Thorius aureus* has the second smallest genome at 9.6 pg (n =2), followed by *Thorius* sp. San Juan del Estado at 12.4 pg (n = 2) and *Thorius* sp. 7 at 13.65. *Thorius tlaxiacus* has the largest genome size in this genus at 28.9 pg (n = 4). Within this genus, the largest genome is 3.1 times bigger than the smallest genome.

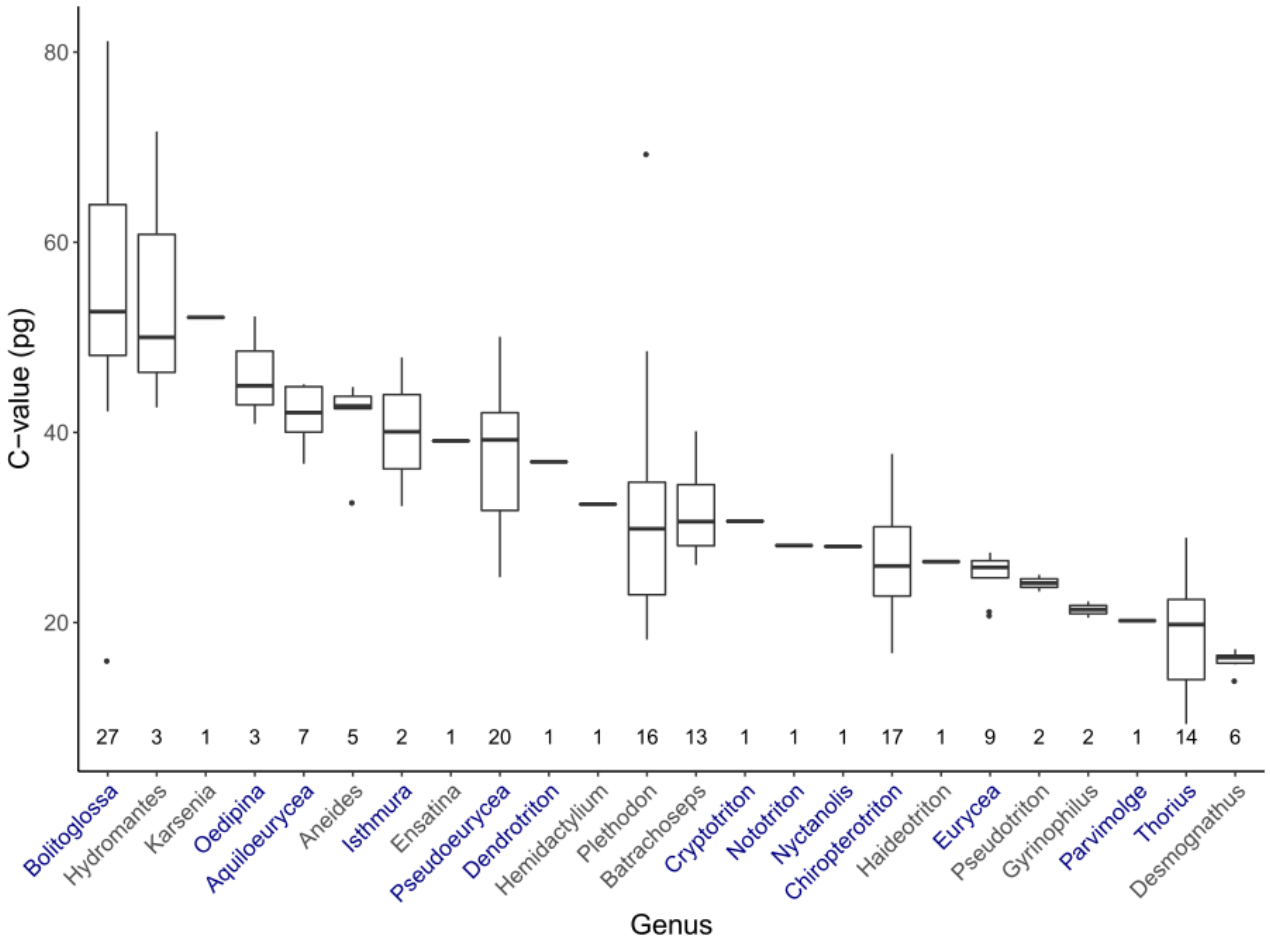

Figure 4: Genome size of plethodontid salamanders in picogams by genus. Number of species stated above the genus name. Tropical salamander genera in blue

To determine the quality of the staining we calculated the coefficient of variation (cv) of the integrated optical density of the nuclei analyzed per sample. In most cases, the cv was smaller than 10%, but in some cases, it would be much larger (up to 35%). This was most likely caused by staining problems, causing some nuclei to be more stained than others in one slide. In total, 31 samples were eliminated from this study for having a cv >10% in IOD. To make sure that the genome size estimates were consistent, an internal control sample of *Ambystoma velasci* (C-value ~40 pg) was added to all the samples in which *A. mexicanum* was used as a standard. The results obtained using *B. platydactyla* and *A. mexicanum* are comparable among them.

Maximum Snout Vent-Length

The average maximum SVL for plethodontid salamanders (n = 286 spp.) is 53.0 mm. *Thorius infernalis* is the smallest salamander, with a maximum SVL of 18.8 mm and *Isthmura gigantea* is the largest species with a maximum SVL of 161 mm. The genus *Bolitoglossa* has the largest variation in SVL. *Bolitoglossa jugivagans* is the smallest species of this genus (maximum SVL 31.0 mm), followed by *Bolitoglossa kamuk* at 34.8 mm, making them miniature salamanders by

the criteria of SVL alone. The biggest salamander of this genus is *Bolitoglossa magnifica* (maximum SVL 133.5 mm), making it 3.8 times larger than *B. jugivagans*. The genus *Thorius*, whose species are known for their minute size, has an average maximum SVL of 25.9 mm (n = 26 spp). The smallest salamander of this genus is *Thorius infernalis* with a max SVL of 18.8 mm and the largest species of the genus are *Thorius aureus* and *T. boreas*, both with a maximum SVL of 34.9 mm, 1.8 times larger than *T. infernalis* (Fig. R2). The genera *Parvimolge* (n = 1 spp.), *Thorius* (n = 26 spp.), and *Cryptotriton* (n = 6 spp) are composed only of miniature species under the criteria of max SVL ≤ 35 mm. The genera *Nototriton, Dendrotriton, Oedipina, Aquiloeurycea,* and *Bolitoglossa* also have some species that are classified as miniature based on body size (figure 5).

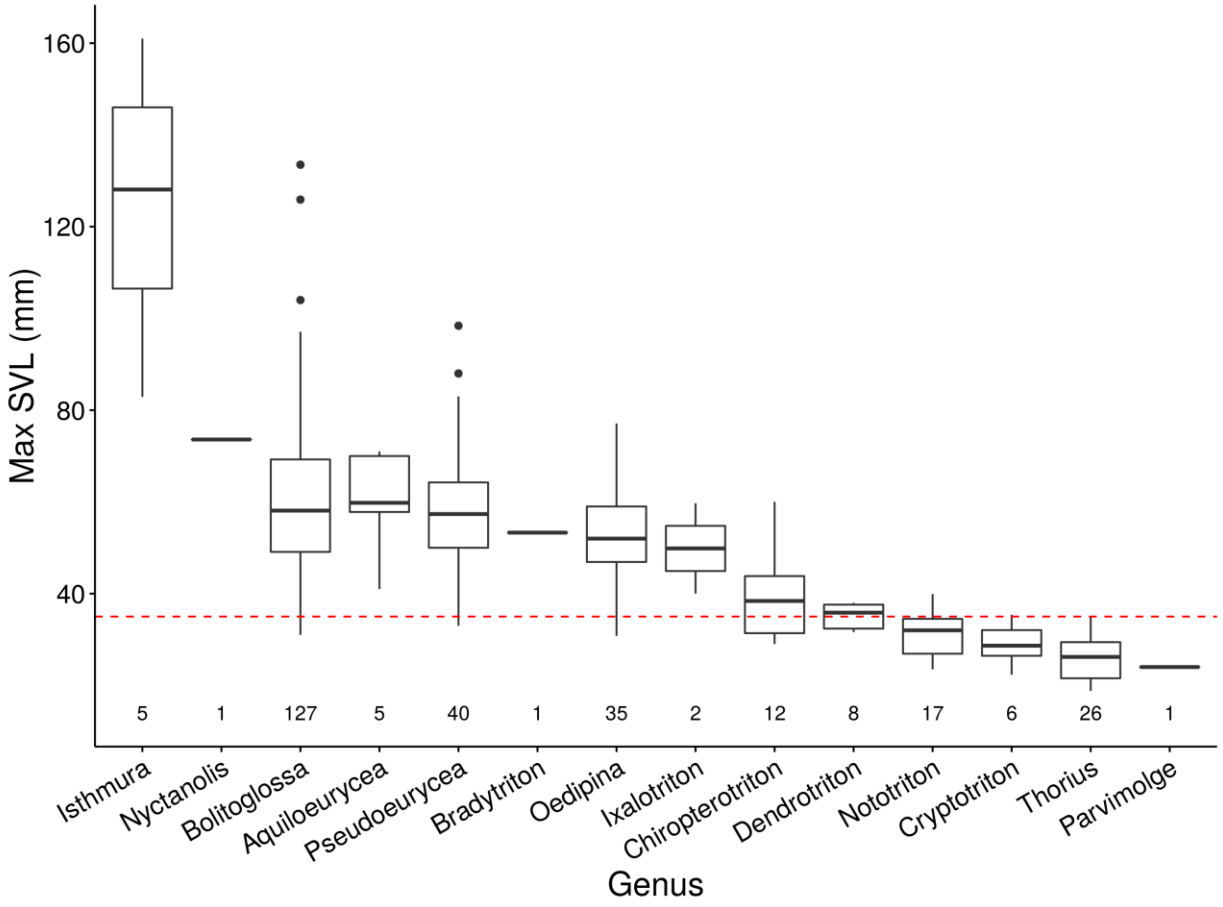

Figure 5: Maximum Snout-Vent Length of plethodontid salamanders by genus. Number of specimens measured is given above genus name. The red line indicates the miniature threshold of 35 mm.

5.2. Head measurements

The mean head length (snout-gular length) for neotropical salamanders (n = 581) is 8.3 mm, with a minimum of 2.9 mm and a maximum of 29.1 mm. Mean head width is 5.4 mm, 2.1 mm min and 22.0 max. Mean head depth is 2.7 mm, 0.9 mm min and 14.3 max.

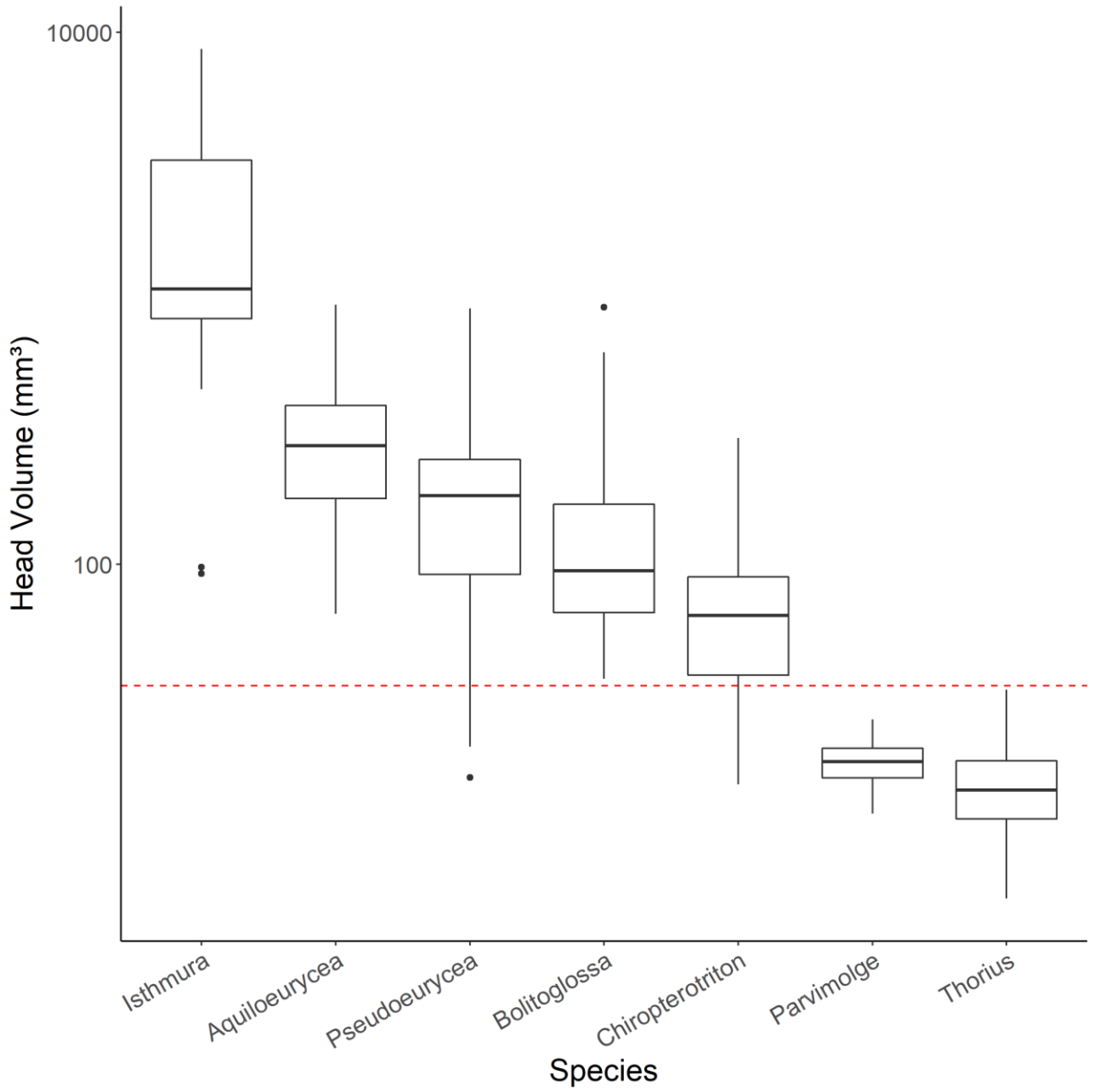

Figure 6: Average head volume of plethodontid salamanders by genus. Number of specimens measured given above genus name. The red line indicates the miniature threshold for salamanders with head volume smaller than the largest Thorius head volume.

Average head volume is 223.7 mm^3 with a minimum of 5.5 mm^3 in *Thorius pennatulus* and a maximum of 8673.3 mm^3 in *Isthmura bellii.* The genus *Thorius* had the smallest mean head volume at 15.7 mm^3 (n = 91), *Thorius boreas* had the largest head volume of the genus at 31.2 mm^3. The genus *Isthmura* had the largest specimens with an average head volume of 2574.2 mm^3. It is hard to establish a cut off head volume size for miniature salamanders. For illustrative purposes, I define miniature salamanders by head volume as all the salamanders with a head volume smaller than the largest head volume of all *Thorius* (Figure 6). This definition will be

revised in the future when head volume data for additional genera of miniature salamanders (*Cryptotriton*) become available.

Average head volume and max SVL are strongly correlated in neotropical salamanders (R^2 = 0.84 p-value = 8.092e-14, n = 46 spp.) (Fig. 7). Even though longer salamanders tend to have bigger heads, there are some species with relatively small heads but long bodies, such as *Oedipina elongata* and *Pseudeurycea lineola* (max SVL = 44.00, head volume = 31.29 mm^3), that wouldn't be considered miniature by one criterion but would by the other.

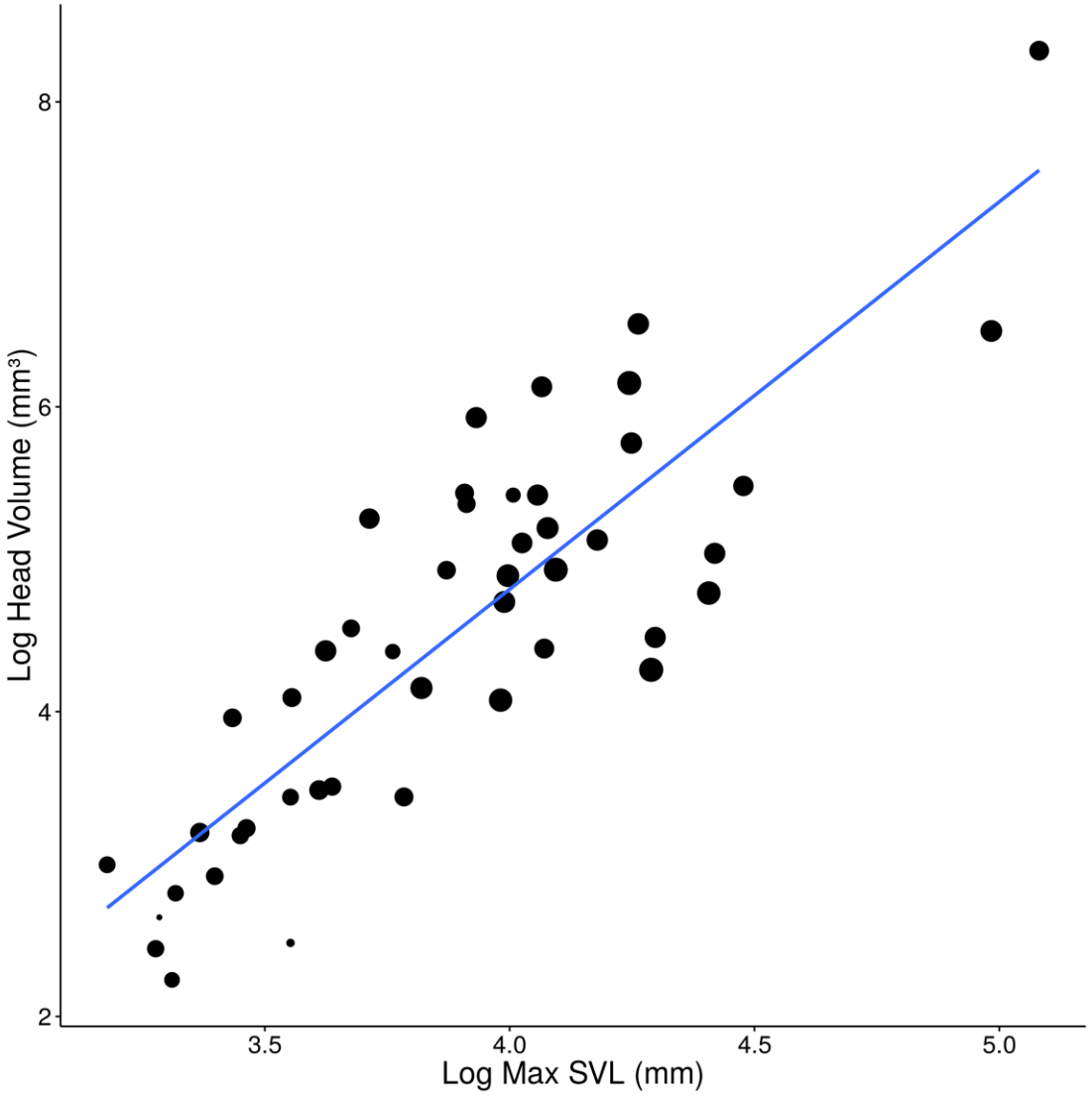

Figure 7: Log Head volume vs Log Max SVL for 46 species. The blue line indicates the simple regression line (R^2 = 0.84 p-value = 8.092e-14). Dot diameter is proportional to genome size.

5.3. Correlation between genome size and body size

There is a positive, significant correlation between maximum SVL and C-value in neotropical salamanders (R^2 = 0.611, p-value = 9.162e-09, n = 73 spp). After controlling for phylogeny

using PGLS, however, the association of max SVL and C-value is no longer significant (t = 1.253, p = 0.214). Despite the lack of a significant correlation between body size and genome size after accounting for phylogeny, there are no miniature salamanders with big genomes and larger salamanders do not have especially small genomes (figure 8).

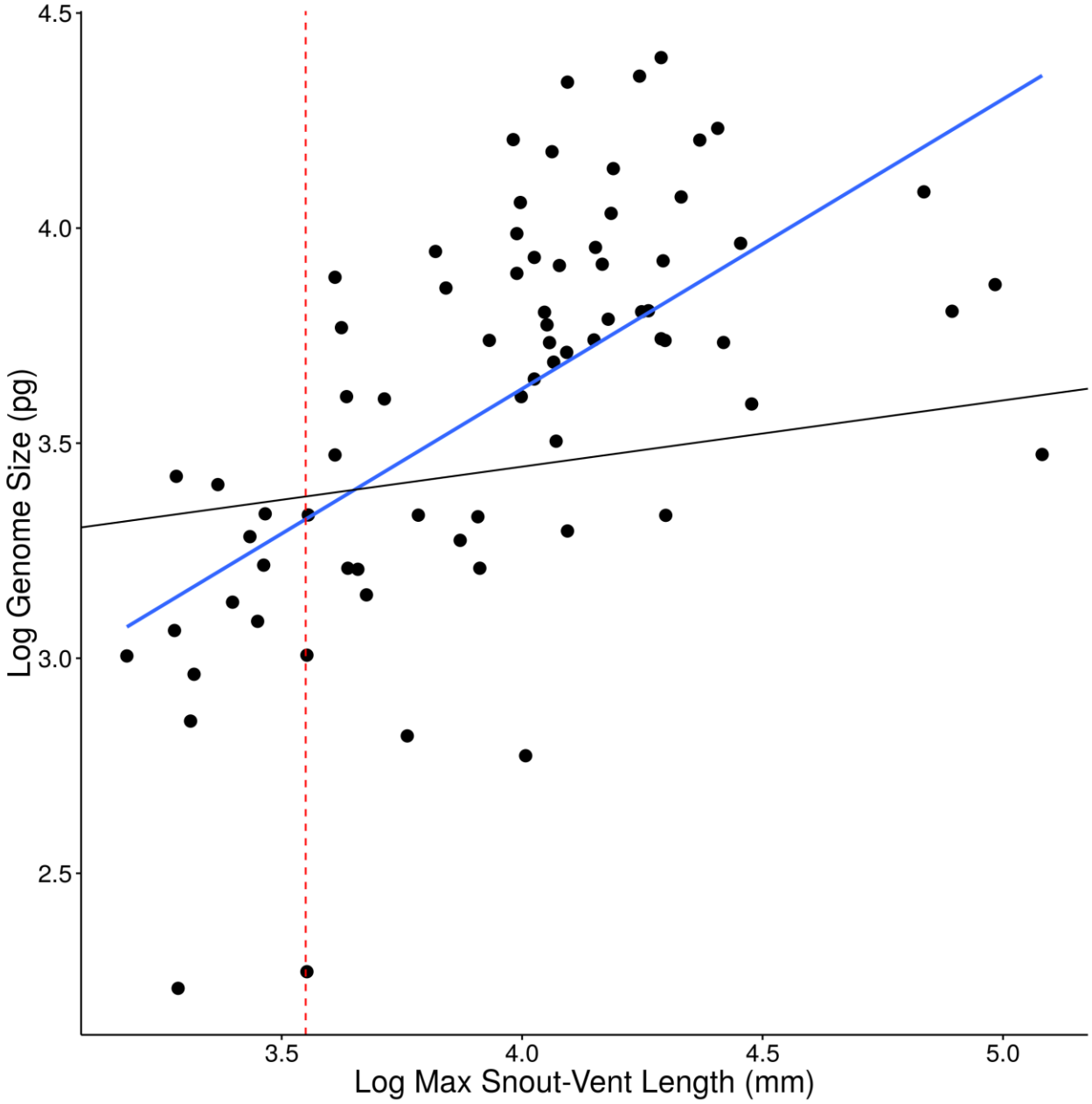

Figure 8: Log-Log genome size vs Max SVL. Miniature salamanders (>35mm SVL) are below the red dashed line.

The correlation between head volume and genome size is slightly lower but still significant (R^2 = 0.591, p-value < 2.65e-07, n = 64 spp). Accounting for phylogeny, the correlation became much lower and non-significant (t = 0.896, p-value = 0.374). Once again, we can observe that there are no miniature salamanders (based on head volume) with big genomes and there are not many salamanders with large heads and small genomes (figure 9).

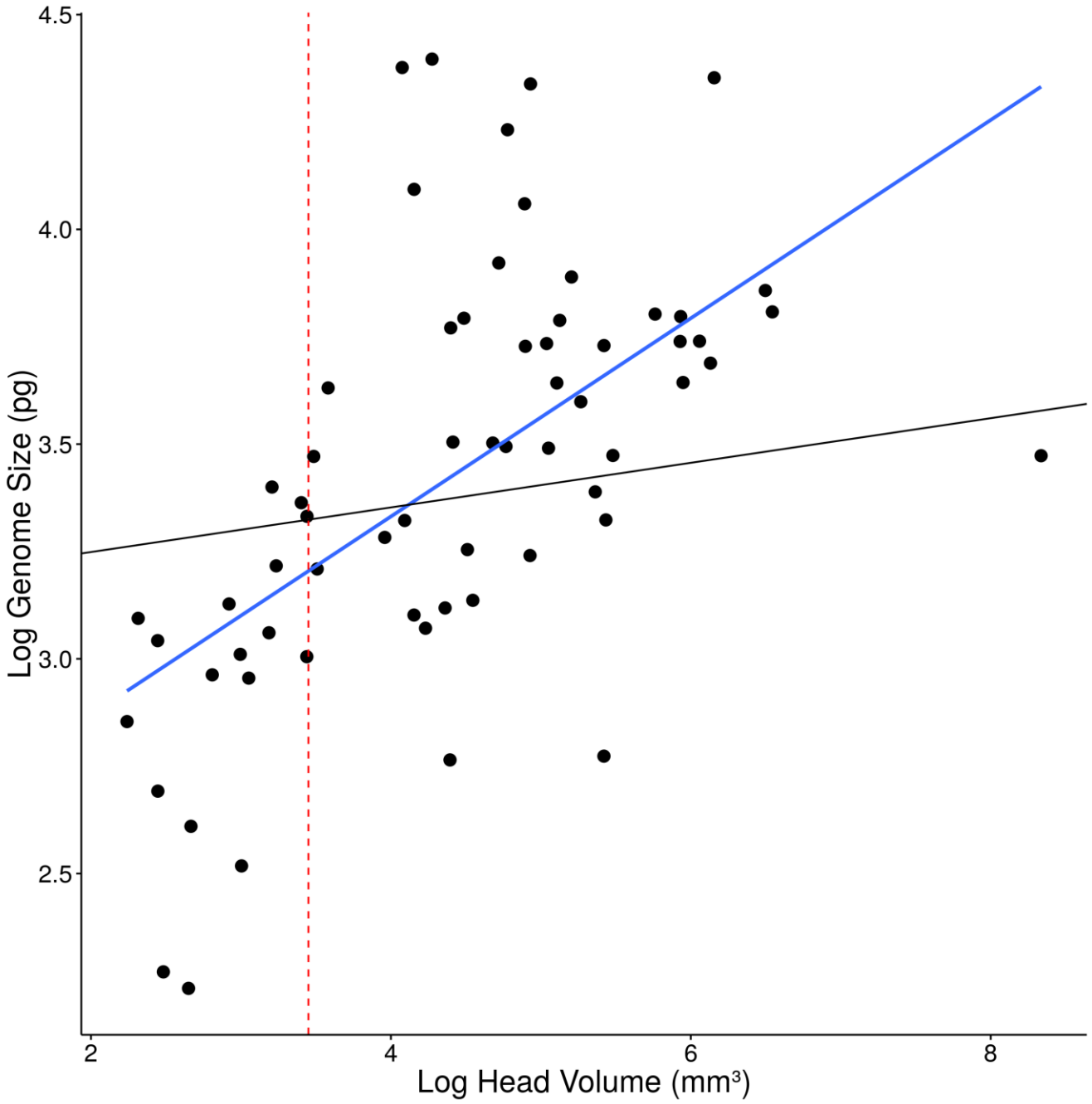

Figure 9: Log-Log genome size vs head volume. Miniature salamanders (>35mm SVL) are below the red dashed line.

5.4. Ancestral state reconstruction

Using the Akaike Information Criterion, Brownian motion (BM) was selected as the most likely model of evolution for this character. AIC scores are summarized in table 2:

Table 2 Akaike Information Criterion corrected for small sample sizes (AICc) for the four models of character evolution tested for the ancestral state reconstruction. SE models consider the standard error. Delta AICc (dAICc) represents the difference between AICc of the current model and the best model.

	BM	OU	EB	white
AICc	643.16	643.69	645.31	718.21
AICc_SE	645.27	645.89	647.47	720.36
dAICc	0	0.53	2.15	75.06
dAICc_SE	0	0.63	2.2	75.1

Ancestral states of the 169 nodes in the phylogeny were estimated under this model with a 95% confidence estimated using 100 replicates. The ancestral genome size of the neotropical salamanders was estimated to be 34.5 pg (95% CI 22.5–46.6). There appear to be some independent expansions of genome size within *Bolitoglossa* and at least one important reduction in the ancestor of all extant species of *Thorius*. There are some other independent genome size reductions in other clades of the phylogeny, such as in the genus *Chiropterotriton*. In the genus *Pseudoeurycea* we can see another independent reduction in the clade formed by sister species *Pseudeurycea saltator* and *Pseudeurycea juarezi*. *Parvimolge townsendi* is another species for which we can assume an independent reduction in genome size relative to the ancestor of the *Aquiloeurycea*, *Bolitoglossa, Isthmura,* and *Pseudoeurycea*. The *Bolitoglossa* clade is interesting, because there are signs of both independent and dramatic genome size expansions and reductions. Under the assumption of Brownian motion, there is a strong phylogenetic signal for genome size, with Blomberg's K having a value of K = 1.0043 (p-value = 0.01) and Pagel's λ = 0.9055 (p-value = 9.74 e-18).

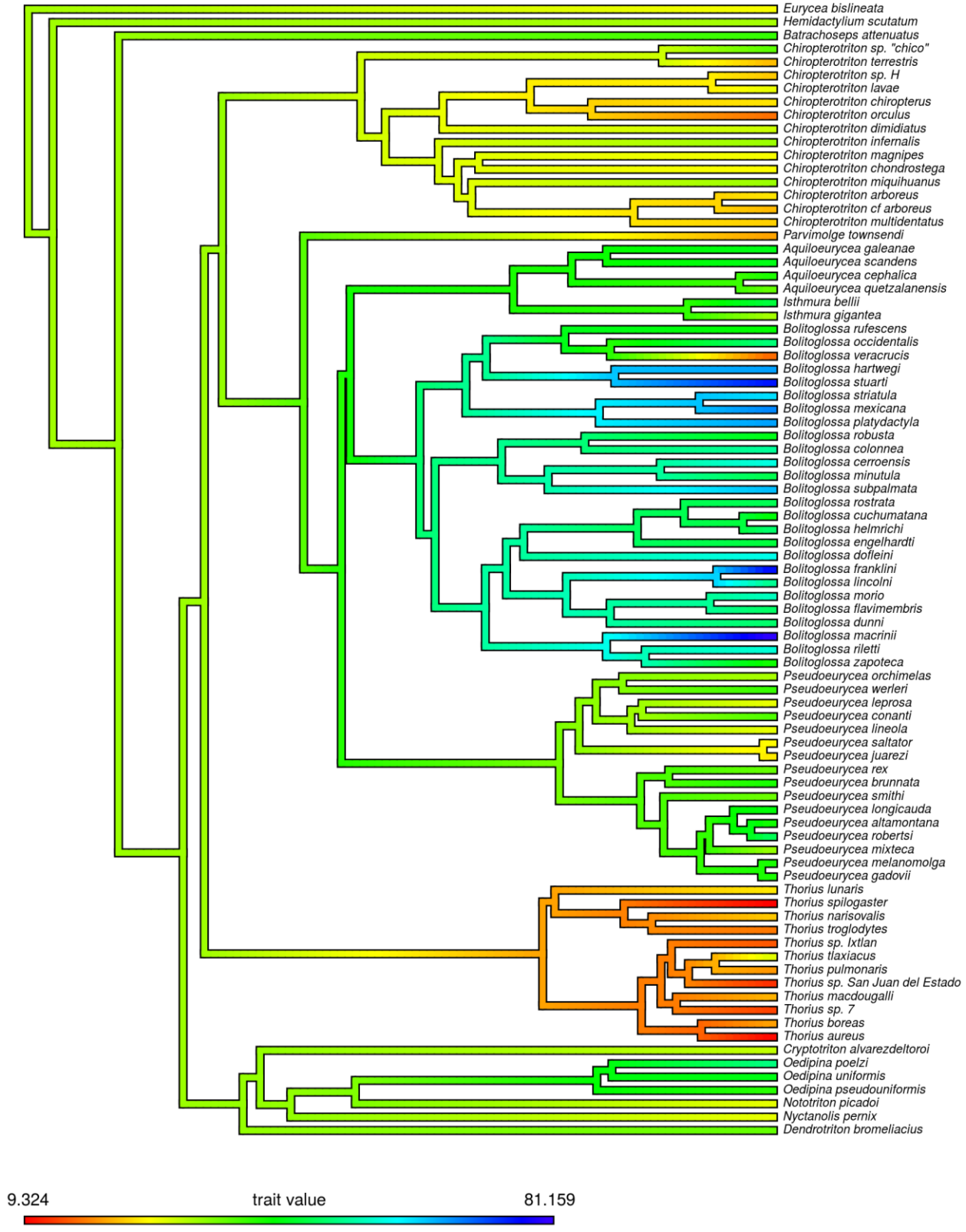

Figure 10: Ancestral state reconstruction of neotropical Plethodontid Salamanders' C-value on a phylogeny. The Thorius clade has the smallest genome sizes of all.

6. Discussion

The very large genome size of salamanders, as well as the wide interspecific variation in their DNA content, has intrigued scientists for more than a century. Studying the dynamics of genome size evolution of this group of amphibians will help us understand more about the evolution of the genome and how it might be related with other mechanistic and morphological characteristics. Their morphology, while more thoroughly studied, also has sparked interest in naturalists and biologists. There are miniature salamanders that are among the smallest extant vertebrates, but there are also giant salamanders in Asia that are the largest amphibians in the world.

The main aim of this project was to understand how genome size has evolved within the neotropical plethodontid salamanders and to test for a relationship between genome size and body size. In particular, I sought to test the hypothesis that miniature salamanders have undergone independent reductions in genome size. I originally thought that selection for a miniature phenotype in salamanders would impose a strong selection for reduced genome size mediated by cell size. This idea was mainly supported by the observations that miniature salamanders had undergone secondary simplifications in their cranial structures as well as in their nervous system, especially in organs whose function depends on cell number, like the eye.

Expanding the amount of information available on genome size for neotropical salamanders was the first step towards my objective. I measured the genome size of 65 different species, 53 of which did not have previous estimates. More importantly, I increased the number of genome size estimates of miniature salamanders. Before the conclusion of this project there were genome size estimates for only five miniature species, and now we have estimates for 29 miniature species. With the previously available information of 103 species, plus the genome size of 53 previously unreported species, we now have a more representative sample of neotropical salamanders of varying body size, allowing me to test hypotheses of how genome size might have evolved and what relationship, if any, it has with miniaturization.

In agreement with my hypothesis, I found that smaller salamanders indeed have smaller genome sizes. As well as being smaller in body size, miniature neotropical salamanders include two species with the smallest genome sizes ever found in a salamander: *Thorius spilogster* (9.3 pg) and *Thorius aureus* (9.7 pg). These two species break a record as being the only species of salamanders with a genome size under 10 pg, although their genome is still almost three times larger than the human genome. I also found several non-miniature species with small genome sizes, such as *Bolitoglossa veracrucis.* Although I expected miniature salamanders to have smaller genome sizes, there is no reason for larger species to necessarily have larger genomes. A large species could potentially accommodate as many small cells as it needed to form all necessary structures, but assuming there is a runaway genome expansion mediated

by retrotransposable elements, larger species would continue to increase their genome and cell size until some constraint stopped genome expansion.

Another interesting and somewhat puzzling result was the finding that some of the smallest salamanders in the genus *Thorius* have bigger genomes that some larger or similarly sized species of the same genus. For example, *Thorius spilogaster* and *Thorius aureus*, the species with the smallest genome sizes, aren't the smallest in the genus While the first has a maximum SVL of 26.7 and a genome size of 9.3 pg, a similarly sized salamander, *Thorius macdougalli,* has a maximum SVL of 26.5 mm and a genome size of 21.4 pg. Similarly, *Thorius aureus* has a larger body size at a maximum SVL of 34.9 mm but a smaller genome size. I would have expected to find the opposite result, because selection for small genome size, mediated by cell size, should be strongest on smaller species. The clade shows a small genome size overall.

In the present study, phylogenetic comparative methods were used alongside traditional statistical analyses to assess the correlation between body size and genome size. I hypothesized that there would be a strong correlation between these two traits, especially in miniature salamanders. My reasoning followed the idea that the constraints that miniaturization imposes on the physiology and morphology of the salamander, especially on their nervous system, would not allow for further expansions on the genome size of these salamanders. Similarly, non-miniature salamanders wouldn't have these kinds of constraints on their genome size. As shown by Hanken (1983), miniature salamanders have a larger percentage of their head volume occupied by their brain—up to 40% in several *Thorius* species compared to 25% in *Pseudoeurycea goebeli*—which indicates an allometric reduction in head size compared to brain size.

In accordance with my hypothesis, a strong and highly significant positive correlation was found between genome size and SVL ($R^2 = 0.61$) and between genome size and head volume ($R^2 = 0.59$). Without making any kind of correction for phylogeny, I could draw the conclusion that the body size of salamanders imposes a constraint on the genome size, or vice versa. But once I control for nonindependence of the data points using phylogenetic generalized least squares (PGLS) to analyze the data, the correlation between the two traits is no longer significant and I can conclude that the phylogenetic relationships and shared ancestry of the group explains the variance of the trait.

The results from the statistical analyses indicate that although small salamanders do have smaller genome sizes, the variance of the trait is better explained through the phylogeny and shared ancestry of the group. The fact that most miniature salamanders included in the PGLS analysis belong to one genus (*Thorius*) also raises concerns about whether the amount of information we can infer from the traits of extant salamander species is a limiting factor when trying to infer the relationships between genome size and mechanistic processes in the evolutionary history of the group. Having most of the miniature salamanders in one clade

becomes then an issue of statistical power. Having few independent data points could shroud a relationship if it truly exists.

The fact that miniature salamanders across all genera have small genome sizes leads me to believe that, regardless of the PGLS results, there is a relationship between morphology and genome size (albeit perhaps nonlinear). I have come up with two possible scenarios on how this relationship might have arisen:

1. The last common ancestor of all salamanders had a large genome (supported by the observations that most amphibians have large genomes as well as paleogenomic data (Organ, Canoville, Reisz, & Laurin, 2011)) and a relatively large body size. After several independent genome size reductions, the morphological constraints on the reduction of body size were lifted and gave rise to several clades of miniature salamanders.
2. The last common ancestor of all salamanders had a small genome and some species expanded their genomes while some others retained their small genome size. The species that retained the small genome size were then able to undergo morphological miniaturization.

The results of the PGLS analysis suggest that there isn't a causal relationship between genome size and body size, and as such, we cannot state that there is selection for reduced genome size in small salamanders. This result also does not support the opposite direction of causality, with genome size exerting an influence on body size, the opposite of my hypothesis. If genome size did influence body size, salamanders would need to undergo deletion of a large amount of genetic material from their genomes to reduce their genome and cell size before being able to reach a miniature size. I believe that the first scenario is the most likely, but to test this new hypothesis a broader analysis of the genomic elements present in different species of salamanders is needed.

Another possible explanation for the observed results is that there is a genome size threshold above which miniaturization in salamanders becomes impossible. An observation that supports this idea is that there aren't any miniature salamanders with average genomes sizes above 45 pg. The fact that, as stated before, there is not a straightforward relationship between body size and genome size within miniature salamanders (shown by the fact that the very smallest salamanders have larger genomes than some of the largest miniature species) could also support a genome size threshold for miniaturization hypothesis. Other compelling observations that could support this hypothesis are the severe morphological characteristics displayed on miniature salamanders, such as the significantly reduced space between the eyes and the brain or the relative position of the brain in the cranium (Hanken, 1983b). If genome size wasn't a factor to consider in miniaturization, one could argue that reducing all the organs in a proportional way would be a simpler way to achieve a miniature phenotype.

The results of the ancestral genome size reconstruction for neotropical salamanders indicate that the ancestral neotropical salamander had a moderately sized genome of about 35 pg, which is close to the average genome size of all salamanders. Some clades, such as the *Pseudoeurycea* clade, seem to have undergone an initial reduction with subsequent expansions, while others, such as the *Bolitoglossa* clade, seem to have undergone an initial expansion in genome size with several posterior reductions, one of which appears to have been rather dramatic and would be interesting to study. The results from an ancestral state reconstruction, while statistically robust, have to be viewed in light of all the assumptions implicit in the analysis. It is very important to remember that, under a Brownian model of evolution, the method will tend towards a reconstruction of an ancestral state that is very close to the mean of the trait being studied. Furthermore, directional tendencies (such as a reduction in all lineages over time) cannot be accurately reconstructed using this method.

The method I used to estimate the genome size of neotropical salamanders, FIAD, works by staining DNA in the nuclei of RBC. This method uses the same basic principle as traditional densitometric methods, in which the amount of DNA is determined by how much light is blocked by the stained nuclei. The main advantage of this method over traditional densitometry is the use of image analysis software, which allows the user to convert the amount and shading of pixels in a specific area to a measure of Integrated Optical Density. Having the ability to measure several nuclei by microscope field, instead of one nucleus at a time is a great improvement over traditional densitometric methods. One caveat of this method, especially when working with species with large genomes, is the amount of artificial variability that can be introduced to the measurement by external factors. The image analysis software can correct for several of these factors, such as glare when making a measurement. Some other sources of noise among measurements are harder to correct because they are intrinsic to the sample. One of these is the three-dimensional position the nuclei take on the slide once it is fixed. Because salamander nuclei are not totally symmetrical, slight variation in nucleus position can change the way light hits it and is scattered. This is likely one of the reasons salamanders with small genomes show smaller variability in their measurements than salamanders with larger genomes.

Other methods for measuring genome size, such as flow cytometry are extremely useful and precise to measure genome sizes by using isolated, stained nuclei. But once again, larger genomes represent a challenge for this method. The filters used for the nuclei need to be very large to have a useful sample with sufficient numbers of nuclei, but having such large filters makes it hard to have isolated nuclei. Another potential problem with this approach is the possibility of clogging the flow cytometer's sheath. In general, FIAD is a fairly cheap and precise method to attain useful genome size estimates and may be more useful than flow cytometry when working with organisms that have very large genomes.

In summary, I found that salamanders of the *Thorius* genus are the smallest in body size and have the smallest genomes sizes among all of the estimated C-values in salamanders, both

previously reported and those measured on this study. I even found that *Thorius spilogaster* (9.3 pg) and *Thorius aureus* (9.7) have the smallest genome sizes among all the salamanders for which there is data on genome size, a record previously held by the genus *Desmognathus*. Other miniature salamanders, such as those of the genera *Chiropterotriton* and *Parvimolge* also have small genome sizes. There appears to be a threshold on genome size over which miniaturization is not possible despite the results from the statistical analyses, which show that the correlation between genome size and body size is not significant once I accounted for phylogenetic relationships. There are still several questions about the direction of causality over these two traits, but given the number of extant plethodontid salamanders, we might never be able to have enough phylogenetically independent data points to answer them.

The data collected for this project will help us to understand of the broader implications that genomes size has on organism level traits in neotropical salamanders. With these new data, researching the molecular mechanisms by which the variation of genome size came to be, as well as formulating new and interesting hypotheses on the evolution of this characteristic and its relationship with other traits becomes possible. Hopefully, this project and the ones that follow will help us come closer to answer the C-value enigma and understand the huge variation in genome size across the tree of life.

7. Perspectives

To test new hypotheses, I propose to conduct low coverage sequencing experiments on different clades of salamanders, including miniature and non-miniature salamanders. Depending the genomic elements that make up salamanders' genomes, the huge variation in genome size in the group could be caused by differences in the relative percentage of repeated elements, or by differences in the elements making up the genome.

Transposable elements in salamanders seem to be one of the main drivers of genome gigantism (Sun, Shepard, et al., 2012). Exploring the diversity of TEs in salamanders will help us answer several questions related to the evolution of the genome in neotropical salamanders. Using NGS data coupled with phylogenetic analysis, I hope to identify TE bursts that might have occurred in different salamander clades. I can then test if these bursts can help explain the variation of genome sizes among clades. I can also study if a particularly abundant and active TE family in salamanders with large genomes is missing, or has less observed activity over time, in species with smaller genome sizes. I propose to use a sequence-based statistical approach to look for lone LTRs as indicators of large deletions caused by ectopic recombination between different copies of a retrotransposons (Sun, Shepard, et al., 2012) and analyze if this is the reason why some clades have smaller genome sizes.

NGS data can also provide some information about the population dynamics that might have driven the evolution of genome gigantism in amphibians, and whether this characteristic was the result of drift or selection. Coupled with information from other studies of salamander demography and population genetics, I could try to estimate the effective population size (Ne) of different species and try to correlate it with the genome size of species in a phylogenetic context to test if Ne could be responsible for the fixation of large genome sizes by drift, or if this characteristic was driven by a selection for one of the consequences of having a large genome size, such as a slow metabolic rate, increased longevity or others. One recent study (Madison-Villar et al., 2017) shows that genetic drift alone is not responsible for genome size gigantism in salamanders, but it is important to test this theory in the group of salamanders showing the largest variation in genome size.

Depending on the results of these experiments, I expect to be able to delve deeper into the intricacies of the molecular mechanisms of genome size evolution in vertebrates. Regardless of the answers to my questions, having sequence information for the repeat elements in salamanders' genomes will be extremely valuable, and it will help improve the development of methods to study what has been sometimes dubbed the dark matter of the genome (Blaxter, 2010).

8. Conclusion

I estimated the genome size of 66 different species of plethodontid salamanders, 54 of which did not have previously reported C-value estimates in the Animal Genome Size Database. This represents an increase of 34.6% over the previous genome size estimates available for this family of salamanders. With the data available of genome size estimates for 156 plethodontid salamanders, from a total of 360 specimens, I was able to determine that the average genome size for this group of salamanders is 35.2 pg per haploid cell. The plethodontid salamander with the smallest genome reported to date is *Thorius spilogaster*, with a C-value of 9.3 pg, while the largest is *Bolitoglossa macrinii* with a C-value 81.1 pg. Both of these record genome sizes were estimated as part of this study.

There is a strong and significant correlation between maximum Snout-Vent Length and head volume, which makes both an effective proxy for body size and appropriate parameters to test correlations between this and other traits. When testing for correlations between genome size and body size (either maximum Snout-Vent Length or head volume) I found that there were no significant correlations once I adjusted for phylogenetic relationships. The very strong phylogenetic signal found shows that related species tend to resemble each other in genome size. Although there has been compelling morphological evidence to assume that there is a strong relationship between genome size and body size for several organisms, this relationship might not be as straightforward as originally thought, at least in neotropical salamanders. An argument could be made for the idea of genome size-body size thresholds. A clade with a particular upper and lower genome sizes might be able to have a particular range of body sizes, and the lower limits for this body size might be controlled by genome size, rather than the opposite.

Brownian motion is the most likely model of evolution for genome size in neotropical salamanders, implying that small, random changes in genome size over time account for the diversity of genome sizes that we see today. Under this model, the ancestral state of the genome size for neotropical plethodontid salamanders is 35.4 pg (95% CI 22.5–46.6) meaning that the ancestor of this group had a moderately sized genome compared to extant species. Reconstruction of genome size evolution on the phylogeny shows several apparent independent reductions occurring mainly in clades with miniature salamanders, while other clades tend to expand their genome sizes.

9. Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walters, P. (2002). Chromosomal DNA and Its Packaging in the Chromatin Fiber. In *Molecular Biology of the Cell*. https://doi.org/10.1091/mbc.E14-10-1437

AmphibiaWeb. (2017). AmphibiaWeb: Information on amphibian biology and conservation. Retrieved from https://amphibiaweb.org

Bannert, N., & Kurth, R. (2006). The Evolutionary Dynamics of Human Endogenous Retroviral Families. *Annual Review of Genomics and Human Genetics*, *7*(1), 149–173. https://doi.org/10.1146/annurev.genom.7.080505.115700

Blaxter, M. (2010). Revealing the dark matter of the genome. *Science*. https://doi.org/10.1126/science.1200700

Bogart, J. P. (1982). Ploidy and genetic diversity in Ontario salamanders of the Ambystoma jeffersonianum complex revealed through an electrophoretic examination. *Canadian J Zoo*. https://doi.org/10.1139/z82-117

Capy, P. (2005). Classification and nomenclature of retrotransposable elements. *Cytogenetic and Genome Research*. https://doi.org/10.1159/000084978

Cole, S. T., & Saint-Girons, I. (1999). Bacterial genomes—all shapes and sizes. In *Organization of the prokaryotic genome* (pp. 35–62). American Society of Microbiology.

Dicke, U., & Roth, G. (2010). Evolution of the amphibian nervous system. In *Evolution of Nervous Systems* (pp. 61–124). https://doi.org/10.1016/B0-12-370878-8/00129-4

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews*, *5*(June), 435–445. https://doi.org/10.1038/nrg1348

Feder, M. E. (1982). Thermal Ecology of Neotropical Lungless Salamanders (Amphibia: Plethodontidae): Environmental Temperatures and Behavioral Responses. *Ecology*, *663*(6), 1665–1674.

Feulgen, R., & Rossenbeck, H. (1924). Mikroskopisch-chemischer Nachweis einer Nucleins??ure vom Typus der Thymonucleins??ure und die-darauf beruhende elektive F??rbung von Zellkernen in mikroskopischen Pr??paraten. *Hoppe-Seyler's Zeitschrift Fur Physiologische Chemie*. https://doi.org/10.1515/bchm2.1924.135.5-6.203

Frankel, A. D., & Young, J. A. T. (1998). HIV-1: Fifteen Proteins and an RNA. *Annual Review of Biochemistry*. https://doi.org/10.1146/annurev.biochem.67.1.1

Green, D. M. (1980). Size Differences in Adhesive Toe-Pad Cells of Treefrogs of the Diploid-Polyploid Hyla versicolor complex. *Journal of Herpetology*, *141*(14). https://doi.org/10.2307/1563870

Green, E. D., Watson, J. D., & Collins, F. S. (2015). Human Genome Project: Twenty-five years of big biology. *Nature*. https://doi.org/10.1038/526029a

Gregory, T. R. (2001a). Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews of the Cambridge Philosophical Society*, *76*(1), 65–101. https://doi.org/10.1111/j.1469-185X.2000.tb00059.x

Gregory, T. R. (2001b). The Bigger the C-Value, the Larger the Cell: Genome Size and Red Blood Cell Size in Vertebrates. *Blood Cells, Molecules, and Diseases*, *27*(5), 830–843. https://doi.org/10.1006/bcmd.2001.0457

Gregory, T. R. (2002). A Bird'S-Eye View of the C-Value Enigma: Genome Size, Cell Size, and Metabolic Rate in the Class Aves. *Evolution*, *56*(1), 121. https://doi.org/10.1554/0014-3820(2002)056[0121:ABSEVO]2.0.CO;2

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. *Biological Journal of the Linnean Society*, *79*(2), 329–339. https://doi.org/10.1046/j.1095-8312.2003.00191.x

Gregory, T. R. (2011). *The Evolution of the Genome*.

Gregory, T. R., & Hebert, P. D. N. (1999). The modulation of DNA content: Proximate causes and ultimate consequences. *Genome Research*. https://doi.org/10.1101/gr.9.4.317

Gulliver, G. (1875). Observations on the Sizes and Shapes of the Red Corpuscles of the Blood of Vertebrates.

Hanken, J. (1983a). Miniaturization and its Effects on Cranial Morphology in Plethodontid Salamanders, Genus Thorius (Ambphibia: Plethodontidae). I. Osteological variation. *Journal of Morphology*, *268*, 55–75.

Hanken, J. (1983b). 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. *Journal of Morphology*, *177*(3), 255–268. https://doi.org/10.1002/jmor.1051770304

Hanken, J., & Wake, D. B. (1993a). Miniaturization of Body Size: Organismal Consequences and Evolutionary Significance. *Annual Review of Ecology and Systematics*, *24*(1), 501–519. https://doi.org/10.1146/annurev.es.24.110193.002441

Hanken, J., & Wake, D. B. (1993b). Miniaturization of Body Size: Organismal Consequences and Evolutionary Significance. *Annual Review of Ecology and Systematics*, *24*(1), 501–519. https://doi.org/10.1146/annurev.es.24.110193.002441

Hardie, D. C., Gregory, T. R., & Hebert, P. D. N. (2002). From Pixels to Picograms: A Beginners' Guide to Genome Quantification by Feulgen Image Analysis Densitometry. *Journal of Histochemistry & Cytochemistry*, *50*(6), 735–749. https://doi.org/10.1177/002215540205000601

Hare, E. E., & Johnston, J. S. (2011). Genome Size Determination Using Flow Cytometry of Propidium Iodide-Stained Nuclei. *Molecular Methods for Evolutionary Genetics*, 3–12. https://doi.org/10.1007/978-1-61779-228-1

Hershey, A. D., & Chase, M. (1952). Independent functions of viral protein and nucleic acid in growth of bacteriophage. *Journal of General Physiology*. https://doi.org/Doi 10.1085/Jgp.36.1.39

Jeffery, N. W., Ellis, E. A., Oakley, T. H., & Gregory, T. R. (2017). The genome sizes of ostracod crustaceans correlate with body size and evolutionary history, but not environment. *Journal of Heredity*. https://doi.org/10.1093/jhered/esx055

Levy, D. L., & Heald, R. (2016). Biological scaling problems and solutions in amphibians. *Cold Spring Harbor Perspectives in Biology*, *8*(1). https://doi.org/10.1101/cshperspect.a019166

Licht, L. E., & Lowcock, L. A. (1991). Genome size and metabolic rate in salamanders. *Comparative Biochemistry and Physiology*, *100*(1), 83–92. https://doi.org/10.1016/0305-0491(91)90089-V

Lin, X., Faridi, N., & Casola, C. (2016). An ancient transkingdom horizontal transfer of penelopelike retroelements from arthropods to conifers. *Genome Biology and Evolution*. https://doi.org/10.1093/gbe/evw076

Madison-Villar, M. J., Sun, C., Lau, N. C., Settles, M. L., & Mueller, R. L. (2017). Small RNAs from a Big Genome: The piRNA Pathway and Transposable Elements in the Salamander Species Desmognathus fuscus. *Journal of Molecular Evolution*, *84*(1), 67. https://doi.org/10.1007/s00239-016-9769-1

Miescher-Rüsch, F. (1871). Ueber die chemische Zusammensetzung der Eiterzellen.

Mirsky, a E., & Ris, H. (1951). The desoxyribonucleic acid content of animal cells and its evolutionary significance. *The Journal of General Physiology*. https://doi.org/10.1085/jgp.34.4.451

Mueller, R. L., Ryan Gregory, T., Gregory, S. M., Hsieh, A., & Boore, J. L. (2008). Genome size, cell size, and the evolution of enucleated erythrocytes in attenuate salamanders. *Zoology*, *111*(3), 218–230. https://doi.org/10.1016/j.zool.2007.07.010

Nagl, W. (1974). Mitotic cycle time in perennial and annual plants with various amounts of DNA and heterochromatin. *Developmental Biology*. https://doi.org/10.1016/S0012-1606(74)80038-2

Neumann, P., Požárková, D., & Macas, J. (2003). Highly abundant pea LTR retrotransposon Ogre is constitutively transcribed and partially spliced. *Plant Molecular Biology*. https://doi.org/10.1023/B:PLAN.0000006945.77043.ce

Nogueira, A. F., Costa, C. M., Lorena, J., Moreira, R. N., Frota-Lima, G. N., Furtado, C., … Schneider, I. (2016). Tetrapod limb and sarcopterygian fin regeneration share a core genetic programme. *Nature Communications*, *7*, 13364. https://doi.org/10.1038/ncomms13364

Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*. https://doi.org/10.1038/246096a0

Organ, C. L., Canoville, A., Reisz, R. R., & Laurin, M. (2011). Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. *Journal of Evolutionary Biology*. https://doi.org/10.1111/j.1420-9101.2010.02176.x

Petrov, D. A. (2002). Mutational Equilibrium Model of Genome Size Evolution. *Theoretical Population Biology*. https://doi.org/10.1006/tpbi.2002.1605

Price, H. J., & Bachmann, K. (1976). Mitotic cycle time and DNA content in annual and perennial Microseridinae (Compositae, Cichoriaceae). *Plant Systematics and Evolution*. https://doi.org/10.1007/BF00983368

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Roth, G., Rottluff, B., & Linke, R. (1988). Miniaturization, genome size and the origin of functional constraints in the visual system of salamanders. *Naturwissenschaften*, *75*(6), 297–304. https://doi.org/10.1007/BF00367320

Roth, G., & Walkowiak, W. (2015). The influence of genome and cell size on brain morphology in amphibians. *Cold Spring Harbor Perspectives in Biology*, *7*(9). https://doi.org/10.1101/cshperspect.a019075

Rovito, S. M., Parra-Olea, G., Recuero, E., & Wake, D. B. (2015). Diversification and biogeographical history of Neotropical plethodontid salamanders. *Zoological Journal of the Linnean Society*, *175*(1), 167–188. https://doi.org/10.1111/zoj.12271

Seelamgari, A., Maddukuri, A., Berro, R., de la Fuente, C., Kehn, K., Deng, L., … Kashanchi, F. (2004). Role of viral regulatory and accessory proteins in HIV-1 replication. *Frontiers in Bioscience*. https://doi.org/10.1097/ALN.0b013e31825dd7ac

Sessions, S. K. (2008). Evolutionary cytogenetics in salamanders. *Chromosome Research*, *16*(1), 183–201. https://doi.org/10.1007/s10577-007-1205-3

Sessions, S. K., & Larson, A. (1987). Developmental Correlates of Genome Size in Plethodontid Salamanders and Their Implications for Genome Evolution. *Source: Evolution Evolution*, *41*(416), 1239–1251. https://doi.org/10.2307/2409090

Shay, J. W., & Wright, W. E. (2000). Hayflick, His Limit, and Cellular Ageing. *Nature Reviews. Molecular Cell Biology*. https://doi.org/10.1038/35036093

Siomi, M. C., Sato, K., Pezic, D., & Aravin, A. A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/nrm3089

Smith, J. D. L., Bickham, J. W., Gregory, T. R., & Bainard, J. (2013). Patterns of genome size diversity in bats (order Chiroptera) [1]. *Genome*, *56*(8), 457–472. https://doi.org/10.1139/gen-2013-0046

Sun, C., López Arriaza, J. R., & Mueller, R. L. (2012). Slow DNA loss in the gigantic genomes of salamanders. *Genome Biology and Evolution*, *4*(12), 1340–1348. https://doi.org/10.1093/gbe/evs103

Sun, C., Shepard, D. B., Chong, R. A., Arriaza, J. L., Hall, K., Castoe, T. A., … Mueller, R. L. (2012). LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders. *Genome Biology and Evolution*, *4*(2), 168–183. https://doi.org/10.1093/gbe/evr139

Swift, H. H. (1950). The Desoxyribose nucleic acid content of animal nuclei. *Physiological Zoology*, *23*(3), 169–198.

Symonds, M. R. E., & Blomberg, S. P. (2014). A primer on phylogenetic generalised least squares. In *Modern Phylogenetic Comparative Methods and their Application in Evolutionary Biology*. https://doi.org/10.1007/978-3-662-43550-2_5

Vendrely, R., & Vendrely, C. (1949). La teneur du noyau cellulaire en acide désoxyribonucléique á travers les organes, les individus et les espéces animales. *Experientia*. https://doi.org/10.1007/BF02172642

Wake, D. B. (2012). Taxonomy of Salamanders of the Family Plethodontidae (Amphibia: Caudata). *Zootaxa*, *3484*, 75–82.

Watson, J. D., Crick, F. H. C., & Franklin, R. E. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*. https://doi.org/10.1038/171964b0

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., … Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*(4), 276–276. https://doi.org/10.1038/nrg2165-c4