Relaxation of Natural Selection in the Evolution of the Giant Lungfish Genomes

Silvia Fuselli,*,[1] Samuele Greco,[2] Roberto Biello,[1] Sergio Palmitessa,[2] Marta Lago,[1] Corrado Meneghetti,[1] Carmel McDougall,[3] Emiliano Trucchi,[4] Omar Rota Stabelli,[5,6] Assunta Maria Biscotti,[4] Daniel J. Schmidt,[3] David T. Roberts,[7] Thomas Espinoza,[7] Jane Margaret Hughes,[3] Lino Ometto (iD),[8] Marco Gerdol,[2] and Giorgio Bertorelle*,[1]

[1]Department of Life Sciences and Biotechnology, University of Ferrara, Ferrara, Italy

[2]Department of Life Sciences, University of Trieste, Trieste, Italy

[3]Australian Rivers Institute, Griffith University, Brisbane, Queensland, Australia

[4]Department of Life and Environmental Sciences, Marche Polytechnic University, Ancona, Italy

[5]Research and Innovation Centre, Fondazione Edmund Mach, 38010 San Michele all'Adige, Italy

[6]Center Agriculture Food Environment, University of Trento, 38010 San Michele all'Adige, Italy

[7]Seqwater, Ipswich, 4305 Queensland, Australia

[8]Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

*Corresponding authors: E-mails: fss@unife.it; ggb@unife.it.

Associate editor: Yuseob Kim

Abstract

Nonadaptive hypotheses on the evolution of eukaryotic genome size predict an expansion when the process of purifying selection becomes weak. Accordingly, species with huge genomes, such as lungfish, are expected to show a genome-wide relaxation signature of selection compared with other organisms. However, few studies have empirically tested this prediction using genomic data in a comparative framework. Here, we show that 1) the newly assembled transcriptome of the Australian lungfish, *Neoceratodus forsteri*, is characterized by an excess of pervasive transcription, or transcriptional leakage, possibly due to suboptimal transcriptional control, and 2) a significant relaxation signature in coding genes in lungfish species compared with other vertebrates. Based on these observations, we propose that the largest known animal genomes evolved in a nearly neutral scenario where genome expansion is less efficiently constrained.

Key words: genome size evolution, lungfish, pervasive transcription, relaxation of natural selection, Australian lungfish (*Neoceratodus forsteri*).

Introduction

Genome size varies by at least five orders of magnitude in eukaryotic organisms, displaying significant interspecific differences within phyla and even within lower taxonomic rank (e.g., vertebrates; Blommaert 2020). Although this remarkable variation is explained by the different genomic content of noncoding DNA and repeated sequences, among which transposable elements (TEs) usually play a relevant role (Hidalgo et al. 2017; Wright 2017), the evolutionary processes underlying the control (or lack thereof) of genome size in eukaryotes are still debated (Whitney and Garland 2010; Lynch 2011; Lynch and Marinov 2015; Wright 2017).

Adaptive hypotheses assume that the evolution of genome size is the result of natural selection acting on several phenotypic correlates (Cavalier-Smith 1982; Vinogradov 1995). Indeed, although genome size does not correlate with organism complexity (the so-called *C*-value paradox, Thomas 1971; Gregory 2001), it does correlate positively with cell and nucleus size and negatively with cell division rate, which, in turn, may affect the metabolic rate or life cycle complexity (Gregory 2002; Liedtke et al. 2018; Roddy et al. 2021). Alternatively, nonadaptive and (nearly) neutral hypotheses assume that most insertions and deletions are neutral or slightly deleterious, so that they cannot be effectively removed by natural selection when the effective population size (Ne) is small, and random drift is consequently large (Lynch and Conery 2003; Wright 2017). Studies explicitly testing the nearly neutral hypothesis led to contrasting results about the role of evolutionary factors, such as Ne, in the evolution of genome size (e.g., Mohlhenrich and Mueller 2016; Lefébure et al. 2017; Roddy et al. 2021), leaving the debate still open.

Despite the astonishing diversity of the genome sizes of eukaryotes, species possessing enormous genomes (i.e., >40 Gb) are exceptionally found only in plants (Psilotales in ferns, Liliales, and Santalales in flowering plants) and in

Open Access

Article

lungfish and salamanders among vertebrates (Gregory 2001; Hidalgo et al. 2017). Lungfish genomes and cells are among the largest known in vertebrates (Pedersen 1971), a feature that negatively correlates with their metabolic rates, among the lowest measured for any fish species (Brett 1972; Seifert and Chapman 2006). To trace the rate of expansion of lungfish genome size, Thomson (1972) studied cell size in osteocyte lacunae of fossil osteoclasts, under the assumption that cell size and genome size are correlated (see also Davesne et al. 2021). The fossil cell size trend suggests that genomes were small in the Devonian, and then their size started increasing after the main diversification of the group and the significant decline in taxonomic diversity that occurred in the Carboniferous. Depending on the species, cell size and thus genome size in lungfish reached a plateau between 200 and 100 Ma (Thomson 1972). Recent genomic analyses support this view, indicating that genome size expanded at a rate of 160 Mb/My until ~200 Ma and more slowly afterwards (Meyer et al. 2021; Wang et al. 2021). The same genomic analysis showed that genomes are large mainly because they accumulated TEs and noncoding DNA. Indeed, repeated sequences account for about 90% of the *Neoceratodus forsteri* genome and at least 61.7% of the *Protopterus annectens* genome. Interestingly, the chromosomes of *N. forsteri*, although considerably expanded, retain the synteny of chordate linkage groups (Meyer et al. 2021).

In this study, we test the hypothesis of (nearly) the neutral evolution of lungfish genome size, whereby a relaxation of natural selection would have allowed the accumulation of noncoding and repetitive genomic elements. Using the newly assembled *N. forsteri* transcriptome in a comparative molecular evolutionary framework, our analyses support a model where a less efficient purifying selection method might have significantly contributed to genomic gigantism in lungfish.

Results

Transcriptome Assembly, Annotation, Assessment of Transposon Activity, and Pervasive Transcription

Two adult male specimens of *N. forsteri*, with an estimated age of 45 and 39 years (Fallon et al. 2019), respectively, were sampled in the Brisbane River, southeast Queensland, Australia. We de novo assembled a transcriptome, evaluated its completeness, and performed a functional annotation from the paired end reads obtained from five different tissues of the first individual (Supplementary Material online). Assembly metrics (a high fraction of transcripts with full-length coding sequences, 93.76% of vertebrate BUSCOs detected as complete, 5.41% fragmented, and only 0.81% missing) indicated the presence of several protein-coding transcripts that could not be accurately annotated in the reference genome of the same species because of technical constraints (Meyer et al. 2021; supplementary table S3, Supplementary Material online). Although significant TE activity was detected (supplementary table S4 and fig. S1, Supplementary Material online), our analyses also revealed

that 189,822 out of 289,539 contigs (65.56% of the total) were categorized as noncoding mRNAs associated with nonrepetitive regions. These results suggest that nearly two-thirds of all assembled expressed sequences in *N. forsteri* may be linked to pervasive noncoding transcription (supplementary fig. S2, Supplementary Material online). A comparative analysis with other bony fish species and with the axolotl, an amphibian with a giant genome, revealed that the total amount of pervasively transcribed genomic sequence was positively correlated with the size of the assembled genomes (Spearman $r_s = 0.95$, $P < 10^{-4}$, fig. 1a, supplementary tables S1 and S5, Supplementary Material online). Nevertheless, in the Australian lungfish, such regions spanned a much larger fraction of the genome than in other bony fish (i.e., 8×) or axolotl (2.5×; fig. 1b).

These observations were further supported by the contribution provided by pervasive transcription to the total transcriptional effort in multiple tissues. In fact, the proportion of reads mapped to noncoding transcripts unrelated to TE activity, compared with protein-coding transcripts and TEs, was highly prevalent in all lungfish tissues. Little variation was found between the two analyzed individuals, ranging from 24.30% in the liver of the first individual to 38.68% in the testis of the second individual (fig. 1c, supplementary table S6, Supplementary Material online). Overall, the testis was the tissue where pervasive transcription was more significant, consistent with the expected relaxed state of chromatin during spermatogenesis, which may increase its accessibility by the RNA polymerase II complex and accessory factors regulating transcription (Villa and Porrua 2023). Widespread intergenic transcription was less prominent in the liver, whereas the lung and brain showed an intermediate level of activity. Based on these data, the Australian lungfish clearly emerged as an outlier compared with all other species, including axolotl, where no more than 6.39% of all transcriptional effort was linked to this class of mRNAs (fig. 1c). On average, despite the slightly higher expression observed in *Thymallus thymallus*, mean pervasive transcription accounted for only 3.06% (with a standard deviation of 2.34%) in bony fishes.

Relaxation of Selection in Lungfish

We analyzed a set of 1,790 conserved single copy genes (BUSCO) in 15 different species (supplementary table S2, Supplementary Material online), including 5 of the 6 (or 7 according to Carneiro et al. 2021) recognized lungfish species (fig. 2a), while for *Protopterus amphibious*, genomic data are not available. We followed two approaches to detect positive and relaxed selection (CODEML, Yang 1998 and RELAX, Wertheim et al. 2015). Only those genes that showed a significant and consistent signature in both approaches were considered for further analyses.

The intersection of CODEML and RELAX results showed that 139 and 6 lungfish genes were more relaxed and more conserved, respectively, compared with the rest of the tree (4 fish and 6 tetrapods, SR1; fig. 2b, supplementary fig. S3, Supplementary Material online).

FIG. 1. (a) The total size of the genomic regions subjected to intergenic pervasive transcription across different species. Please note the logarithmic scale. (b) The total fraction of the genome subjected to pervasive transcription, based on detected primary transcripts (i.e., by including introns). (c) Total contribution of pervasively transcribed intergenic regions to the global transcriptional effort in different tissues from multiple species. Gonads may refer either to the testis or to the ovary, depending on available data. Gm, *Gadus morhua*; Ol, *Oryzias latipes*; Ph, *Pangasianodon hypophthalmus*; El, *Esox lucius*; Lo, *Lepisosteus oculatus*; Pf, *Perca fluviatilis*; Aa, *Anguilla anguilla*; Tt, *Thymallus thymallus*; St, *Salmo trutta*; Am, *Ambystoma mexicanum*; Nf, *Neoceratodus forsteri*. Species are ordered based on the size of the assembled genome considered in the analyses, from the smallest (left) to the largest (right). Data are available in supplementary tables S1, S5, and S6, Supplementary Material online.
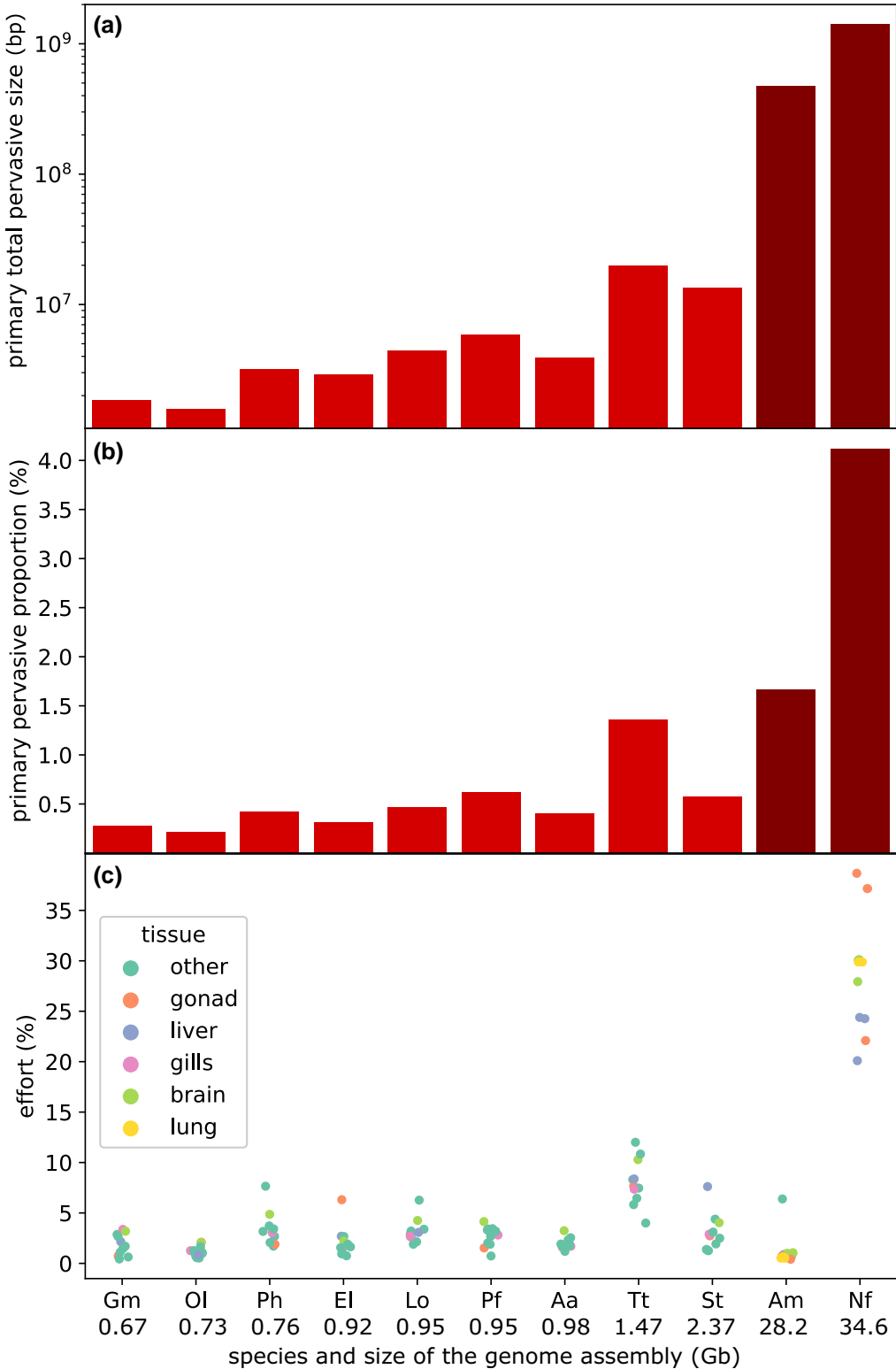
The high number of relaxed genes is surprising, considering that BUSCO genes are evolutionarily conserved. Alternative partitions (SR2-6, fig. 2b) were tested to confirm that the lungfish clade shows evidence of selective relaxation that other groups do not show. A comparison of the proportion of significantly more relaxed and significantly more constrained genes in the different partitions showed that the purifying selection process appeared to

be significantly weaker (χ^2 test $P < 0.001$) in the lungfish clade. Furthermore, single-branch analyses within the lungfish group showed the same trend and revealed a linage-specific rather than gene-specific relaxation signal, as expected for a random, nondirectional effect (supplementary table S7 and fig. S4, Supplementary Material online). In contrast, other species with giant genomes, the urodeles *Ambystoma mexicanum* and *Cynops orientalis*, behave similarly
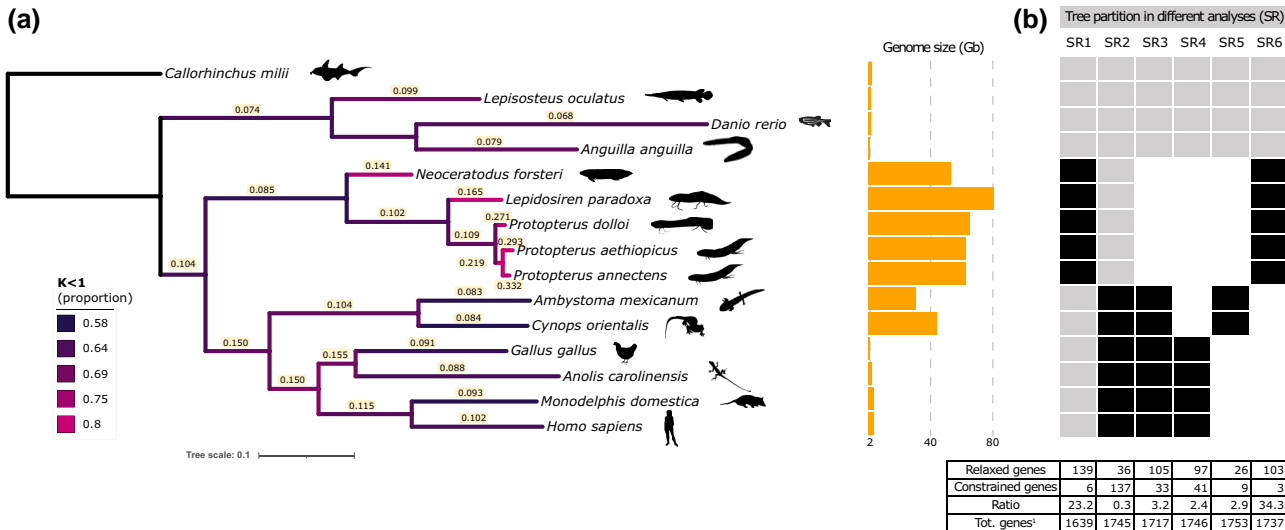
3

	SR1	SR2	SR3	SR4	SR5	SR6
Relaxed genes	139	36	105	97	26	103
Constrained genes	6	137	33	41	9	3
Ratio	23.2	0.3	3.2	2.4	2.9	34.3
Tot. genes[1]	1639	1745	1717	1746	1753	1737

FIG. 2. Branch-specific selective pressure (ω and k) and selection relaxation signature in different tree partitions estimated from 1,790 (or subsets) orthologous genes in 15 vertebrate species, including 5 lungfish. (*a*) Maximum-likelihood phylogeny was inferred using IQ-TREE v1.6.10 (Nguyen et al. 2015) and model selection via ModelFinder (Kalyaanamoorthy et al. 2017; all branches were supported by 100% bootstrap values). The scale bar represents in expected number of nucleotide replacements per site. The numbers shown along each branch are the maximum-likelihood estimates of ω obtained with CODEML free-ratio. Branches are colored according to the proportion of potentially relaxed genes ($k < 1$) estimated with RELAX (general descriptive model), increasing from blue to magenta. Orange bars show the genome size for each species (Gb) according to the Animal Genome Size Database (Gregory 2023. http://www.genomesize.com). (*b*) Six different branch partitions (SR1–SR6) defined to infer the number of genes out of the initial 1,790 orthologues showing intensified signals of relaxation (CODEML $\omega fg > \omega bg \cap$ RELAX $k < 1$; "Relaxed genes" in the table) or conservation (CODEML $\omega fg < \omega bg \cap$ RELAX $k > 1$, "Constrained genes" in the table). Black boxes: foreground (CODEML) or test (RELAX) branches, gray boxes: background (CODEML) or reference (RELAX) branches. [1]Total number of genes that produced an output in both RELAX and CODEML out of 1,790.

to other tetrapods (fig. 2b). When single branches were considered in the RELAX analysis (general descriptive model), lungfish showed a higher proportion of loci with $k < 1$ (i.e., relaxation of selection) than the rest of the tree (fig. 2a). Finally, branch-specific ω values obtained on the concatenated 1,790 genes (free-ratio model, CODEML, fig. 2a) further supported a significant trend toward higher ω values (indicative of relaxation of selection) in lungfish (Mann–Whitney U test, $P < 0.01$).

Under strong relaxation conditions, even generally conserved genes are expected to accumulate deleterious mutations, possibly leading to amino acid replacements at key protein sites. To test whether this was the case in the putatively relaxed lungfish lineages, we identified and counted, in each of the 15 terminal branches of the tree in figure 2a, the private amino acidic substitutions for each of the 15 lineages. These changes were then classified as either radical or conservative according to the conservative-radical index (Sharbrough et al. 2018; supplementary fig. S5, Supplementary Material online). With the exception of *P. annectens*, which is also a lungfish species with a larger ω, lungfish species showed trends that are similar to the rest of the tree, suggesting that the core gene set analyzed, although relaxed, is not drifting toward pseudogenization.

Biological Functions in Lungfish Relaxed and Constrained Genes

The set of relaxed genes was not associated with a statistically significant enrichment of specific biological functions, with the sole exception of the ten GO terms listed in supplementary table S8, Supplementary Material online, which mostly related to protein turnover. Among the six genes more constrained in lungfish, *SMC5* (structural maintenance of chromosomes 5) is particularly interesting, given that its protein product is involved in the repair of DNA double-strand breaks by homologous recombination, telomere maintenance, and sister chromatid cohesion during prometaphase and mitotic progression, functions that could be crucial in the presence of giant genomes.

Discussion

The transcriptome of *N. forsteri* is characterized by a strong background noise caused by the abundance of short, pervasively transcribed RNAs, the expression of which is not associated with coding genes, or caused by TE activity. Pervasive transcription is a common feature of eukaryotic genomes, occasionally providing the raw material for new genes and long noncoding RNAs (Oss and Carvunis 2019; Palazzo and Koonin 2020) but generally representing a by-product of transcription events whose excess is contained through a specific cellular "toolbox" (Jensen et al. 2013; Koonin 2016). In this framework, pervasive transcription is expected to increase when selection is less efficient in controlling this costly, inefficient, and possibly harmful molecular mechanism. The relevant pervasive transcription observed in lungfish tissues, together with the relaxation signal shown by conserved coding genes analyzed by different comparative approaches, suggests that

lungfish genomes have been under reduced selective constraints compared with those of other vertebrate species.

Although multiple lines of evidence support this interpretation, we describe some limitations of our analysis and explain why we believe that their impact on the general conclusion is negligible.

First, the large evolutionary distance between lungfish species and between lungfish and other vertebrates, and the huge size of lungfish genomes, may have produced genome annotation errors. In particular, our pipeline could not discriminate between long intergenic noncoding RNAs (lincRNAs), which are still scarcely known and rarely annotated in nonmodel organisms, and proper transcriptional leakage, intended as a phenomenon linked to a scarce control of the transcription molecular machinery (Villa and Porrua 2023). However, it is reasonable to assume that a similar number of lincRNA genes (close to zero) has been properly annotated in the genomes of all the species taken into account, and therefore, it is unlikely that it could skew our estimates. Second, when looking for a signature of relaxed selection in lungfish using the ratio dN/dS, the deep divergence times (e.g., *N. forsteri*/other lungfish about 230 My, lungfish/tetrapods: about 400 My, Kumar et al. 2022) may have introduced bias because of the saturation at third codon positions and the consequent underestimation of the synonymous substitution rate. The extent to which saturation affects the robustness of codon-based models is controversial (Seo and Kishino 2008; Gharib and Robinson-Rechavi 2013; Weber et al. 2014), but high-sequence divergence alone does not appear to be a serious problem if the alignment is reliable (Yang and dos Reis 2011). The strength of the relaxation signal in lungfish is reinforced by the conservative nature of branch models toward the selective constraint hypothesis. In fact, dN/dS was estimated by averaging nucleotide substitution rates over sites in the protein (Yang and dos Reis 2011) and by including in the data set only conserved gene sequences that were further trimmed by removing poorly aligned regions. Finally, we ruled out that the lungfish signal was random by analyzing different partitions of the tree (fig. 2b, supplementary methods, Supplementary Material online), and we have been extremely conservative in describing as *relaxed* only those genes detected by both approaches, CODEML and RELAX.

In conclusion, our results on pervasive transcription and the accumulation of nonsynonymous variants are robust and consistent with the prediction from the nearly neutral theory of the inverse relationship between the strength of purifying selection and genome size (Lynch and Conery 2003), at least when fish species are compared. As previously observed (Mohlhenrich and Mueller 2016), the two amphibians with giant genomes included in our study, *A. mexicanum* and *C. orientalis*, did not show the same signals, thus arguing against a simple and universal correlation. Consistent with a *nearly* neutral scenario, in which purifying selection is still acting on variants with a large selection coefficient, the genes analyzed were not enriched for radical amino acid changes in the five lungfish species.

Why do lungfish show reduced purifying selection that possibly allowed genome expansion? A reasonable potential explanation is a small, long-term population size. The decline of the lungfish group started in the Carboniferous (Jorgensen and Joss 2016) when, according to fossil bones, genome size started to increase (Thomson 1972) and the number of species started to decrease. The reasons for such a decline in species diversity are not known, but if this process also reduced the population sizes of the surviving species, enhanced drift effects and reduced purifying selection may have allowed genome size expansion through reduced control of TE proliferation (Lockton et al. 2008; Oggenfuss et al. 2021). Genome expansion may trigger evolutionary radiation and/or the evolution of new gene functions (Davesne et al. 2021), but the negative effects of TE proliferation and genomic expansion (Boissinot et al. 2006; Hollister and Gaut 2009; Wei et al. 2022) probably prevailed in lungfish history. Our data suggest a causal link between reduced population size and increased genome size through relaxed selection and TE proliferations.

Materials and Methods

A detailed account of the methods can be found in the supplementary text, Supplementary Material online.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Data Availability

The data presented in this study are openly available in NCBI BioProject; Accession: PRJNA605733. A novel custom bioinformatic script used in this work to identify and count private nonsynonymous changes in phylogeny is available at https://github.com/rsbiello/Lungfish_relax.

References

Blommaert J. 2020. Genome size evolution: towards new model systems for old questions. *Proc Biol Sci.* **287**:20201441.

Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. 2006. Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci U S A.* **103**:9590–9594.

Brett JR. 1972. The metabolic demand for oxygen in fish, particularly salmonids, and a comparison with other vertebrates. *Respir Physiol.* **14**:151–170.

Carneiro J, Dutra GM, Nobre RM, de Pinheiro LML, Oliva PAC, Sampaio I, Schneider H, Schneider I. 2021. Evidence of cryptic speciation in South American lungfish. *J Zool Syst Evol Res*. **59**: 760–771.

Cavalier-Smith T. 1982. Skeletal DNA and the evolution of genome size. *Annu Rev Biophys Bioeng*. **11**:273–302.

Davesne D, Friedman M, Schmitt AD, Fernandez V, Carnevale G, Ahlberg PE, Sanchez S, Benson RBJ. 2021. Fossilized cell structures identify an ancient origin for the teleost whole-genome duplication. *Proc Natl Acad Sci U S A*. **118**:e2101780118.

Fallon SJ, McDougall AJ, Espinoza T, Roberts DT, Brooks S, Kind PK, Kennard MJ, Bond N, Marshall SM, Schmidt D, *et al.* 2019. Age structure of the Australian lungfish (*Neoceratodus forsteri*). *PLoS One* **14**:e0210168.

Gharib WH, Robinson-Rechavi M. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol*. **30**:1675–1686.

Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev Camb Philos Soc*. **76**:65–101.

Gregory TR. 2023. Genome size and developmental complexity. *Genetica* **115**:131–146.

Hidalgo O, Pellicer J, Christenhusz M, Schneider H, Leitch AR, Leitch IJ. 2017. Is there an upper limit to genome size? *Trends Plant Sci*. **22**:567–573.

Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res*. **19**: 1419–1428.

Jensen TH, Jacquier A, Libri D. 2013. Dealing with pervasive transcription. *Mol Cell*. **52**:473–484.

Jorgensen JM, Joss J. 2016. *The biology of lungfishes*. Enfield, New Hampshire: CRC Press.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. **14**:587–589.

Koonin EV. 2016. Splendor and misery of adaptation, or the importance of neutral null for understanding evolution. *BMC Biol*. **14**: 114.

Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. 2022. TimeTree 5: an expanded resource for species divergence times. *Mol Biol Evol*. **39**:msac174.

Lefébure T, Morvan C, Malard F, François C, Konecny-Dupré L, Guéguen L, Weiss-Gayet M, Seguin-Orlando A, Ermini L, Sarkissian CD, *et al.* 2017. Less effective selection leads to larger genomes. *Genome Res*. **27**:1016–1028.

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

Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A*. **105**:13965–13970.

Lynch M. 2011. Statistical inference on the mechanisms of genome evolution. *PLoS Genet*. **7**:e1001389.

Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* **302**:1401–1404.

Lynch M, Marinov GK. 2015. The bioenergetic costs of a gene. *Proc Natl Acad Sci U S A*. **112**:15690–15695.

Meyer A, Schloissnig S, Franchini P, Du K, Woltering JM, Irisarri I, Wong WY, Nowoshilow S, Kneitz S, Kawaguchi A, *et al.* 2021.

Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature* **590**:284–289.

Mohlhenrich ER, Mueller RL. 2016. Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution* **70**:2865–2878.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. **32**:268–274.

Oggenfuss U, Badet T, Wicker T, Hartmann FE, Singh NK, Abraham L, Karisto P, Vonlanthen T, Mundt C, McDonald BA, *et al.* 2021. A population-level invasion by transposable elements triggers genome expansion in a fungal pathogen. *eLife* **10**:e69249.

Oss SBV, Carvunis A-R. 2019. De novo gene birth. *PLoS Genet*. **15**: e1008160.

Palazzo AF, Koonin EV. 2020. Functional long non-coding RNAs evolve from junk transcripts. *Cell* **183**:1151–1161.

Pedersen RA. 1971. DNA content, ribosomal gene multiplicity, and cell size in fish. *J Exp Zool*. **177**:65–78.

Roddy AB, Alvarez-Ponce D, Roy SW. 2021. Mammals with small populations do not exhibit larger genomes. *Mol Biol Evol*. **38**: 3737–3741.

Seifert AW, Chapman LJ. 2006. Respiratory allocation and standard rate of metabolism in the African lungfish, *Protopterus aethiopicus*. *Comp Biochem Physiol A Mol Integr Physiol*. **143**:142–148.

Seo T-K, Kishino H. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. *Syst Biol*. **57**:367–377.

Sharbrough J, Luse M, Boore JL, Logsdon JM, Neiman M. 2018. Radical amino acid mutations persist longer in the absence of sex. *Evolution* **72**:808–824.

Thomas CA. 1971. The genetic organization of chromosomes. *Annu Rev Genet*. **5**:237–256.

Thomson KS. 1972. An attempt to reconstruct evolutionary changes in the cellular DNA content of lungfish. *J Exp Zool*. **180**:363–371.

Villa T, Porrua O. 2023. Pervasive transcription: a controlled risk. *FEBS J*. **290**:3723–3736.

Vinogradov AE. 1995. Nucleotypic effect in homeotherms: body-mass-corrected basal metabolic rate of mammals is related to genome size. *Evolution* **49**:1249–1259.

Wang K, Wang J, Zhu C, Yang L, Ren Y, Ruan J, Fan G, Hu J, Xu W, Bi X, *et al.* 2021. African lungfish genome sheds light on the vertebrate water-to-land transition. *Cell* **184**:1362–1376.e18.

Weber CC, Nabholz B, Romiguier J, Ellegren H. 2014. K_r/K_c but not d_N/d_S correlates positively with body mass in birds, raising implications for inferring lineage-specific selection. *Genome Biol*. **15**:542.

Wei KHC, Mai D, Chatla K, Bachtrog D. 2022. Dynamics and impacts of transposable element proliferation in the Drosophila nasuta species group radiation. *Mol Biol Evol*. **39**(5):msac080.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol*. **32**:820–832.

Whitney KD, Garland TJ. 2010. Did genetic drift drive increases in genome complexity? *PLoS Genet*. **6**:e1001080.

Wright SI. Evolution of genome size. In: *eLS*. 1st ed. Chichester, UK: John Wiley & Sons, Ltd; 2017. p. 1–6. Available from: https://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0023983

Yang Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*. **15**:568–573.

Yang Z, dos Reis M. 2011. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol*. **28**:1217–1228.