# Pervasive duplication of tumor suppressor genes preceded parallel evolution of large bodied Paenungulates

Juan Manuel Vazquez [1] [*], Vincent J Lynch

**1** Department of Human Genetics, 920 East 58th St, Chicago, IL, 60637
Department of Biological Sciences, 551 Cooke Hall, Buffalo NY, 14260

* Corresponding author: juanvazquez@uchicago.edu

## Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pellentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

## Author summary

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eget porta erat. Morbi consectetur est vel gravida pretium. Suspendisse ut dui eu ante cursus gravida non sed sem. Nullam sapien tellus, commodo id velit id, eleifend volutpat quam. Phasellus mauris velit, dapibus finibus elementum vel, pulvinar non tellus. Nunc pellentesque pretium diam, quis maximus dolor faucibus id. Nunc convallis sodales ante, ut ullamcorper est egestas vitae. Nam sit amet enim ultrices, ultrices elit pulvinar, volutpat risus.

)

## Introduction

One of the major constraints on the evolution of large body sizes in animals is an increased risk of developing cancer. If all cells in all organisms have a similar risk of malignant transformation and equivalent cancer suppression mechanisms, organism with many cells should have a higher prevalence of cancer than organisms with fewer cells. Consistent with this expectation there is a strong positive correlation between body size and cancer incidence within species, for example, human cancer incidence increases with increasing adult height [1,2] and cancer incidence is positively correlated with body size in dogs [**???**,3]. There is no correlation, however, between body size and cancer risk between species. This lack of correlation is often referred to as 'Peto's Paradox' [4–6]. While it is clear that a resolution to Peto's Paradox must involve the evolution of enhanced cancer protection alongside increases in body size and lifespan, the specific genetic, molecular, and cellular mechanisms that underlie this resistance have proven elusive. [7–11].

Among the challenges for discovering how animals evolved enhanced cancer protection mechanisms is identifying lineages in which large bodied species are nested within species with small body sizes. Afrotherian mammals are generally small-bodied, similarly to the prediced common ancestor of Eutherian mammals. For example, maximum adult weights are ˜70g in golden moles, ˜120g in tenrecs, ˜170g in elephant shrews, ˜3kg in hyraxes, and 60kg in aardvarks [12]. However, while these extant species are relatively small, the fossil evidence demonstrates that their ancestral lineages reached enormous sizes. For example, while extant hyraxes are relatively small, the extinct Titanohyrax is estimated to have weighted up to ˜1300kg [13]. The largest members of Afrotheria, too, are dwarfed by the size of their recent ancestors: extant cows manatees are large bodied (˜322-480kg) but are relatively small compared to the extinct Stellar's sea cow which is estimated to have weight 8000-10000kg [14]. Similarly African (4,800kg) and Asian elephants (3,200kg) are the largest living elephant species, but are dwarfed by the truly gigantic extinct Proboscideans such as Deinotherium (˜132,000kg), Mammut borsoni (110,000kg), and the Asian straight-tusked elephant (˜220,000kg), the largest known land mammal [15]. Remarkably these large-bodied Afrotherian lineages are nested within small bodied species (Fig. 1) [16–19], indicating that gigantism independently evolved in hyraxes, sea cows, and elephants (Paenungulates). Thus, Paenungulates are an excellent model system in which to explore the mechanisms that underlie the evolution of large body sizes and augmented cancer resistance.

Although many mechanisms can potentially resolve Peto's paradox, the most parsimonious route to enhanced cancer resistance is likely through an increased copy number of tumor suppressors. Such an example has been seen in the case of candidate genes such as *TP53* and *LIF* [11,20,21] as well as in studies involving a limited set of candidate genes [22,23]. As these studies focus on *a priori* gene sets, however, it remains unknown whether this is a general, genome-wide trend in Afrotherian genomes; and whether such a general trend is associated with the recent increases in body size – and therefore expected cancer risk – in these species.

Here, we trace the evolution of body mass and gene copy number variation in Afrotherians in order to investigate whether gene duplications are enriched in large, long-lived species for genes involved in known tumor suppression pathways. Our estimates of the evolution of body mass, similarly to previous studies [16–19], show that large body masses evolved in a step-wise manner, with major increases in body mass in the Pseudoungulata (17kg), Paenungulata (25kg), Tethytheria (296kg), and Proboscidea (4,100kg) stem-lineages. Furthermore, we see that the ancestral body size increases in Hydracoidia and Sirenia were independent events. To study the evolution of gene copy number, we used a genome-wide Reciprocal Best BLAT Hit (RBBH) method to identify gene duplications in Afrotherian genomes, and used parsimony to infer the lineages in which those duplications occurred. We found gene duplications in lineages with increased body mass were enriched in functions related to tumor suppression, including regulation of the cell cycle, DNA damage repair, and regulation of apoptosis. These data suggest that duplication of tumor suppressors played a role in the evolution of large, long-lived in Afrotherians.

# Methods

## Ancestral Body Size Reconstruction

We built a time-calibrated supertree of Eutherian mammals by combining the time-calibrated molecular phylogeny of Bininda-Emonds *et al.* [24] with the time-calibrated total evidence Afrotherian phylogeny from Puttick and Thomas [**???**].

While the Bininda-Emonds *et al.* [24] phylogeny includes 1,679 species, only 34 are Afrotherian, and no fossil data are included. The inclusion of fossil data from extinct species is essential to ensure that ancestral state reconstructions of body mass are not biased by only including extant species. This can lead to inaccurate reconstructions, for example, if lineages convergently evolved large body masses from a small bodied ancestor. In contrast, the total evidence Afrotherian phylogeny of Puttick and Thomas [19] includes 77 extant species and fossil data from 39 extinct species. Therefore we replaced the Afrotherian clade in the Bininda-Emonds *et al.* [24] phylogeny with the Afrotherian phylogeny of Puttick and Thomas [19] using Mesquite. Next, we jointly estimated rates of body mass evolution and reconstructed ancestral states using a generalization of the Brownian motion model that relaxes assumptions of neutrality and gradualism by considering increments to evolving characters to be drawn from a heavy-tailed stable distribution (the "Stable Model") [25]. The stable model allows for occasional large jumps in traits and has previously been shown to out-perform other models of body mass evolution, including standard Brownian motion models, Ornstein–Uhlenbeck models, early burst maximum likelihood models, and heterogeneous multi-rate models [25].

## Identification of Duplicate Genes

*Reciprocal Best-Hit BLAT:* We developed a reciprocal best hit BLAT (RBHB) pipeline to quickly identify homologs and estimate gene copy numbers (**Figure 1A**). The Reciprocal Best Hit (RBH) search strategy is conceptually straightforward: 1) Given a gene of interest $G_A$ in a query genome $A$, one searches a target genome $B$ for all possible matches to $G_A$; 2) For each of these hits, one then performs the reciprocal search in the original query genome to identify the highest-scoring hit; 3) A hit in genome $B$ is defined as a homolog of gene $G_A$ if and only if the original gene $G_A$ is the top reciprocal search hit in genome $A$. We selected BLAT [26] as our algorithm of choice, as this algorithm is sensitive to highly simliar (>90% identity) sequences, thus identifying the highest-confidence homologs while minimizing many-to-one mapping problems when searching for multiple genes. RBH performs similar to other more complex methods of orthology prediction, and is particularly good at identifying incomplete genes that may be fragmented in low quality/poor assembled regions of the genome [**???**,27].

*Effective Copy Number By Coverage:* In lower-quality genomes, many genes are fragmented across multiple scaffolds, which results in BLAT calling multiple hits when in reality there is only one gene. To compensate for this, we came up with a novel statistic, Estimated Copy Number by Coverage (ECNC), which averages the number of times we see each nucleotides of a query sequence in a target genome over the total number of nucleotides of the query sequence found overall in each target genome (Supplementary Figure 1). This allows us to correct for genes that have been fragmented across incomplete genomes, while also taking into account missing sequences from the human query in the target genome. Mathematically, this can be written as:

$$ECNC = \frac{\sum_{n=1}^{l} C_n}{\sum_{n=1}^{l} bool(C_n)}$$

where $n$ is a given nucleotide in the query, $l$ is the total length of the query, $C_n$ is the number of instances that $n$ is present within a reciprocal best hit, and $bool(C_n)$ is 1 if $C_n > 0$ or 0 if $C_n = 0$.

*RecSearch Pipeline:* We created a custom Python pipeline for automating RBHB searches between a single reference genome and multiple target genomes using a list of query sequences from the reference genome. For the query sequences in our search, we used the hg38 Proteome provided by UniProt [28], which is a comprehensive set of

protein sequences curated from a combination of predicted and validated protein sequences generated by the UniProt Consortium. In order to refine our search, we omitted protein sequences originating from long, noncoding RNA loci (e.g. LINC genes); poorly-studied genes from predicted open reading frames (C-ORFs); and sequences with highly repetitive sequences such as zinc fingers, protocadherins, and transposon-containing genes, as these were prone to high levels of false positive hits. After filtering out problematic protein queries, we then used our pipeline (Figure 1A) to search for all copies of our 20456 query genes in publicly available Afrotherian genomes, including African savannah elephant (*Loxodonta africana*: loxAfr3, loxAfr4, loxAfrC), African forest elephant (*Loxodonta cyclotis*: loxCycF), Asian Elephant (*Elephas maximus*: eleMaxD), Woolly Mammoth (*Mammuthus primigenius*: mamPriV), Colombian mammoth (*Mammuthus columbi*: mamColU), American mastodon (*Mammut americanum*: mamAmeI), Rock Hyrax (*Procavia capensis*: proCap1, proCap2, proCap2_HiC), West Indian Manatee (*Trichechus manatus latirostris*: triManLat1, triManLat1_HiC), Aardvark (*Orycteropus afer*: oryAfe1, oryAfe1_HiC), Lesser Hedgehog Tenrec (*Echinops telfairi*: echTel2), Nine-banded armadillo (*Dasypus novemcinctus*: dasNov3), Hoffman's two-toed sloth (*Choloepus hoffmannii*: choHof1, choHof2, choHof2_HiC), Cape golden mole (*Chrysochloris asiatica*: chrAsi1), and Cape elephant shrew (*Elephantulus edwardii*: eleEdw1). For many of these species, we covered multiple assemblies in order to test the effects of assembly size and quality on our hits.

*Duplication gene inclusion criteria:* In order to condense transcript-level hits into single gene loci, and to resolve many-to-one genome mappings, we removed exons where transcripts from different genes overlapped, and merged overlapping transcripts of the same gene into a single gene locus call. The resulting gene-level copy number table was then combined with the maximum ECNC values observed for each gene in order to call gene duplications. We called a gene duplicated if its copy number was two or more, and if the maximum ECNC value of all the gene transcripts searched was 1.5 or greater; previous studies have shown that incomplete duplications can encode functional genes, therefore partial gene duplications were included provided they passed additional inclusion criteria. The ECNC cut off of 1.5 was selected empirically, as this value minimized the number of false positives seen in a test set of genes and genomes. The results of our initial search are summarized in Figure 1B. Overall, we identified [MEDIAN] genes across all species, or [%HITS/QUERIES] of our starting query genes.

*Duplicate gene exclusion criteria:* We excluded genes from downstream analyses for which assignment of homology was uncertain, including uncharacterized ORFs (17), LOC (17), HLA genes (17), replication dependent histones (17), odorant receptors (17), ribosomal proteins (17), zinc finger transcription factors (17), viral and repetitive-element-associated proteins (17) and any protein described as either "Uncharacterized," "Putative," or "Fragment" by UniProt in UP000005640 (17).

**Orthogonal Genome Assessment using CEGMA** In order to determine the effect of genome quality on our results, we used the gVolante webserver and CEGMA to assess the quality and completeness of the genome. CEGMA was run using the default settings of [], and the mammalian-specific core gene sets.

## Evidence for Functionality of Identified Genes

To validate and filter out RBHB results, we intersected our results with either gene prediction or transcriptomic evidence as a proxy for functionality.

**Transcriptome Assembly:** For the African Savana Elephant, Asian Elephant, West Indian Manatee, and Nine-Banded Armadillo, we generated *de novo* transcriptomes using publically-available RNA-sequencing data from NCBI SRA. We mapped reads to all genomes available for each species, and assembled transcripts using HISAT2 and StringTie, respectively [**???**,**??**,**??**]. RNA-sequencing data was not

available for Cape Golden Mole, Cape Elephant Shrew, Rock Hyrax, Aardvark, or the Lesser Hedgehog Tenrec.

**Gene Prediction:** We obtained tracks for genes predicted using GenScan for all the genomes available via UCSC Genome Browser: African savannah elephant (loxAfr3), Rock Hyrax (proCap1), West Indian Manatee (triManLat1), Aardvark (oryAfe1), Lesser Hedgehog Tenrec (echTel2), Nine-banded armadillo (dasNov3), Hoffman's Two-Toed Sloth (choHof1), Cape golden mole (chrAsi1), and Cape Elephant Shrew (eleEdw1); gene prediction tracks for higher-quality assemblies were not available.

**Evidenced Duplicate Criteria:** We intersected our records of duplicate hits identified in each genome with the gene prediction tracks and/or transcriptome assemblies using `bedtools` [**???**]. When multiple lines of evidence for functionality were present for a genome, we used the union of all intersections as the final output for evidenced duplicates. When analyzing the highest-quality assemblies available for each species, if a species had neither gene prediction tracks nor RNA-seq data for the highest-quality genome available, we conservatively included all hits for the genome in the final set of evidenced duplicates.

### Reconstruction of Ancestral Copy Numbers

We implemented a maximum likelihood method for determining the ancestral copy numbers of genes in *Atlantogenata* using IQ-Tree. For this analysis, we used an unrooted subset of our prior species tree, including only the aforementioned *Atlantogenata* species. We generated PHYLIP files containing the copy number of each gene in the highest quality genome for each species, encoding genes on a scale from 1-31+ copies as 1-9, A-V; and encoding a gene's copy number as uncetain ("?") when we did not identify it in the genome. We used the included tree-searching and model-testing functionality in IQ-Tree to determine the most likely topology for the species tree, and to obtain the most likely model for copy number changes in the genome. We defined the ancestral state of a node if it had greater than an 80% posterior probability.

### Pathway Enrichment Analysis

To determine which pathways were associated with duplicated genes in each species and lineage, we used WEBGESTALT to perform overrepresentation analysis (ORA) of the duplicated gene lists relative to our initial query gene list [**???**]. For the database of pathways used in the analysis, we used Reactome [**???**], Wikipathways, and Wikipathways_cancer [**???**], and KEGG [**???**]. For the ORA, we used FDR for determining significance, and ran the analysis at FDR=0.1, FDR=0.2, FDR=0.3, and FDR=0.5.

### Lifespan Phylogenetic Least-Square Regression and Calculating Estimated Cancer Risk Thoughout Atlantogenata

In order to determine the cancer risk at each node, we used a simplified multistage cancer risk model for body size and lifespan [**???**,**??**,**??**]. We defined the

## Results

### Body size frequently and independently expands and contracts throughout *Atlantogenata*

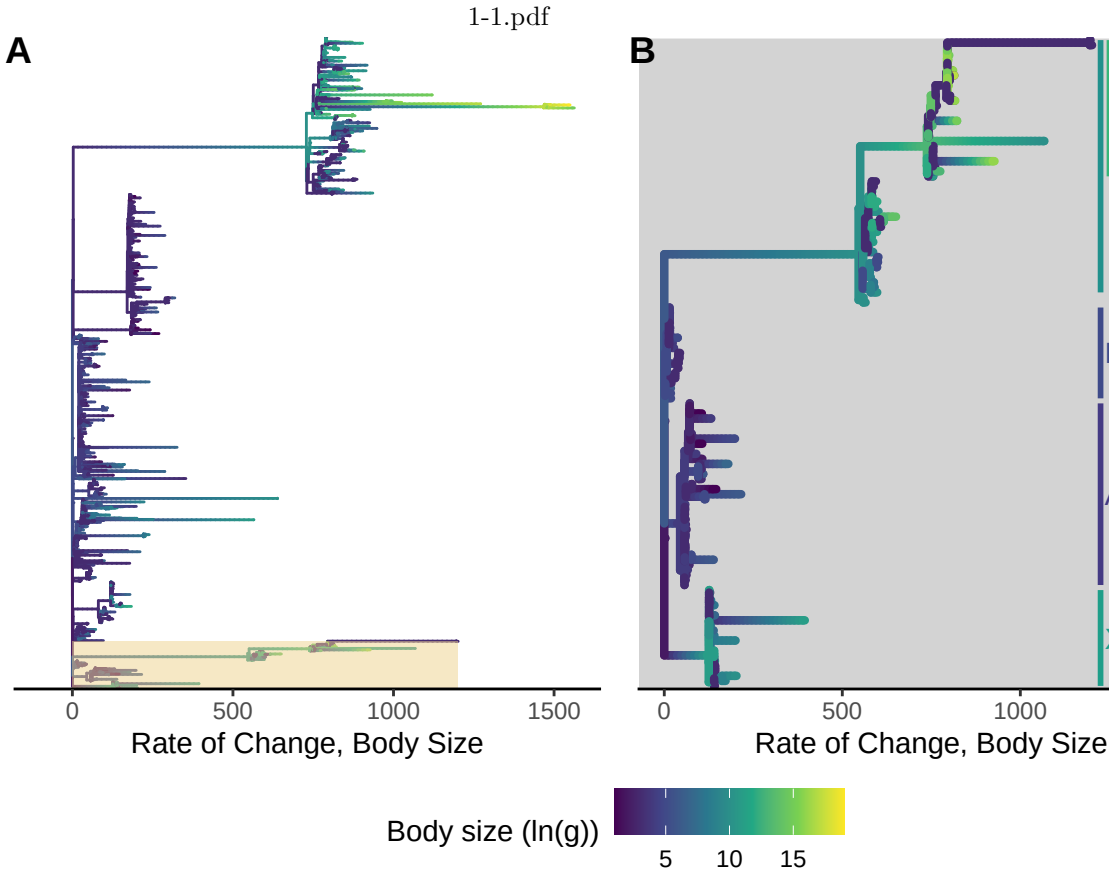



**Fig 1.** Figure 1: Body sizes rapidly and frequently expand in Eutherians, especially in Atlantogenata. **A)** Tree of Eutherian species, colored by ln(Body Size) and with branch lengths set to the rate of change in body sizes, normalized by the square root of the root branch. Atlantogenata is highlighted at the bottom. **B)** Zoom-in of (A) on Atlantogenata. Silhuetes for the African Elephant, West Indian Manatee, Cape Elephant Shrew, Lesser Hedgehog Tenrec, Cape Golden Mole, Nine-Banded Armadillo, and Hoffman's Two-Toed Sloth are colored by their extant body sizes, while clade labels are colored based on the common ancestor's estimated body size

**Table 1.** A table generated by the longtable package.

| Ancestor/Species | Estimated Body Size (log(g)) | 95% CI (Low) | 95% CI (High |
|---|---|---|---|
| Cryptochloris wintoni | 3.13 | 3.13 | 3.13 |
| Amblysomus marleyi | 3.53 | 3.53 | 3.53 |
| Elephantulus revoili | 3.48 | 3.48 | 3.48 |
| Titanohyrax andrewsi | 12.97 | 12.97 | 12.97 |
| Titanohyrax ultimus | 14.08 | 14.08 | 14.08 |
| Megalohyrax sp nov | 12.52 | 12.52 | 12.52 |
| Elephas maximus asurus | 15.66 | 15.66 | 15.66 |
| Protenrec tricuspis | 1.14 | 1.14 | 1.14 |
| Microgale parvula | 1.16 | 1.16 | 1.16 |
| Microgale pusilla | 1.25 | 1.25 | 1.25 |
| Geogale aurita | 1.90 | 1.90 | 1.90 |
| Microgale longicaudata | 2.09 | 2.09 | 2.09 |
| Microgale brevicaudata | 2.19 | 2.19 | 2.19 |
| Microgale jobihely | 2.30 | 2.30 | 2.30 |
| Microgale principula | 2.32 | 2.32 | 2.32 |
| Dilambdogale gheerbranti | 2.38 | 2.38 | 2.38 |
| Microgale taiva | 2.47 | 2.47 | 2.47 |
| Microgale cowani | 2.62 | 2.62 | 2.62 |
| Eremitalpa granti | 3.14 | 3.14 | 3.14 |
| Calcochloris obtusirostris | 3.27 | 3.27 | 3.27 |
| Neamblysomus julianae | 3.33 | 3.33 | 3.33 |
| Chlorotalpa duthieae | 3.38 | 3.38 | 3.38 |
| Chlorotalpa sclateri | 3.54 | 3.54 | 3.54 |
| Macroscelides proboscideus | 3.64 | 3.64 | 3.64 |
| Chrysochloris stuhlmanni | 3.74 | 3.74 | 3.74 |
| Oryzorictes hova | 3.79 | 3.79 | 3.79 |
| Elephantulus myurus | 3.81 | 3.81 | 3.81 |
| Elephantulus brachyrhynchus | 3.81 | 3.81 | 3.81 |
| Elephantulus rozeti | 3.81 | 3.81 | 3.81 |
| Elephantulus fuscus | 3.82 | 3.82 | 3.82 |
| Elephantulus intufi | 3.82 | 3.82 | 3.82 |
| Microgale talazaci | 3.88 | 3.88 | 3.88 |
| Chrysochloris asiatica | 3.89 | 3.89 | 3.89 |
| Elephantulus edwardii | 3.90 | 3.90 | 3.90 |
| Carpitalpa arendsi | 3.94 | 3.94 | 3.94 |
| Amblysomus corriae | 3.94 | 3.94 | 3.94 |
| Amblysomus hottentotus | 3.98 | 3.98 | 3.98 |
| Elephantulus fuscipes | 4.04 | 4.04 | 4.04 |
| Elephantulus rufescens | 4.05 | 4.05 | 4.05 |
| Neamblysomus gunningi | 4.09 | 4.09 | 4.09 |
| Elephantulus rupestris | 4.12 | 4.12 | 4.12 |
| Amblysomus septentrionalis | 4.23 | 4.23 | 4.23 |
| Chambius kasserinensis | 4.27 | 4.27 | 4.27 |
| Amblysomus robustus | 4.33 | 4.33 | 4.33 |
| Micropotamogale lamottei | 4.36 | 4.36 | 4.36 |
| Echinops telfairi | 4.47 | 4.47 | 4.47 |
| Limnogale mergulus | 4.52 | 4.52 | 4.52 |
| Hemicentetes semispinosus | 4.75 | 4.75 | 4.75 |
| Chrysospalax villosus | 4.77 | 4.77 | 4.77 |
| Petrodromus tetradactylus | 5.29 | 5.29 | 5.29 |

| Ancestor/Species | Estimated Body Size (log(g)) | 95% CI (Low) | 95% CI (High |
|---|---|---|---|
| Herodotius pattersoni | 5.50 | 5.50 | 5.50 |
| Setifer setosus | 5.61 | 5.61 | 5.61 |
| Rhynchocyon cirnei | 5.86 | 5.86 | 5.86 |
| Metoldobotes sp nov | 5.93 | 5.93 | 5.93 |
| Chrysospalax trevelyani | 6.13 | 6.13 | 6.13 |
| Rhynchocyon petersi | 6.15 | 6.15 | 6.15 |
| Rhynchocyon chrysopygus | 6.28 | 6.28 | 6.28 |
| Potamogale velox | 6.49 | 6.49 | 6.49 |
| Rhynchocyon udzungwensis | 6.57 | 6.57 | 6.57 |
| Tenrec ecaudatus | 6.75 | 6.75 | 6.75 |
| Dasypus sabanicola | 7.05 | 7.05 | 7.05 |
| Tolypeutes matacus | 7.11 | 7.11 | 7.11 |
| Dasypus septemcinctus | 7.30 | 7.30 | 7.30 |
| Zaedyus pichiy | 7.31 | 7.31 | 7.31 |
| Dasypus hybridus | 7.31 | 7.31 | 7.31 |
| Chaetophractus villosus | 7.61 | 7.61 | 7.61 |
| Chaetophractus nationi | 7.67 | 7.67 | 7.67 |
| Heterohyrax brucei | 7.78 | 7.78 | 7.78 |
| Cabassous centralis | 7.92 | 7.92 | 7.92 |
| Seggeurius amourensis | 7.98 | 7.98 | 7.98 |
| Procavia capensis | 8.01 | 8.01 | 8.01 |
| Dendrohyrax dorsalis | 8.06 | 8.06 | 8.06 |
| Microhyrax lavocati | 8.13 | 8.13 | 8.13 |
| Bradypus tridactylus | 8.23 | 8.23 | 8.23 |
| Bradypus torquatus | 8.27 | 8.27 | 8.27 |
| Dasypus novemcinctus | 8.37 | 8.37 | 8.37 |
| Euphractus sexcinctus | 8.43 | 8.43 | 8.43 |
| Choloepus hoffmanni | 8.47 | 8.47 | 8.47 |
| Bradypus variegatus | 8.49 | 8.49 | 8.49 |
| Tamandua tetradactyla | 8.52 | 8.52 | 8.52 |
| Cyclopes didactylus | 8.53 | 8.53 | 8.53 |
| Choloepus didactylus | 8.71 | 8.71 | 8.71 |
| Thyrohyrax meyeri | 8.78 | 8.78 | 8.78 |
| Saghatherium bowni | 9.13 | 9.13 | 9.13 |
| Dasypus kappleri | 9.23 | 9.23 | 9.23 |
| Thyrohyrax domorictus | 9.30 | 9.30 | 9.30 |
| Dimaitherium patnaiki | 9.57 | 9.57 | 9.57 |
| Phosphatherium escuilliei | 9.62 | 9.62 | 9.62 |
| Saghatherium antiquum | 9.73 | 9.73 | 9.73 |
| Thyrohyrax litholagus | 10.01 | 10.01 | 10.01 |
| Myrmecophaga tridactyla | 10.26 | 10.26 | 10.26 |
| Myorycteropus africanus | 10.27 | 10.27 | 10.27 |
| Selenohyrax chatrathi | 10.73 | 10.73 | 10.73 |
| Priodontes maximus | 10.82 | 10.82 | 10.82 |
| Orycteropus afer | 10.87 | 10.87 | 10.87 |
| Antilohyrax pectidens | 10.93 | 10.93 | 10.93 |
| Bunohyrax fajumensis | 11.32 | 11.32 | 11.32 |
| Afrohyrax championi | 11.32 | 11.32 | 11.32 |
| Geniohyus mirus | 11.33 | 11.33 | 11.33 |
| Prorastomus sirenoides | 11.49 | 11.49 | 11.49 |
| Elephas antiquus falconeri | 11.51 | 11.51 | 11.51 |

| Ancestor/Species | Estimated Body Size (log(g)) | 95% CI (Low) | 95% CI (High |
|---|---|---|---|
| Pachyhyrax crassidentatus | 11.81 | 11.81 | 11.8 |
| Megalohyrax eocaenus | 11.95 | 11.95 | 11.9 |
| Elephas cypriotes | 12.21 | 12.21 | 12.2 |
| Bunohyrax major | 12.36 | 12.36 | 12.3 |
| Titanohyrax angustidens | 12.48 | 12.48 | 12.4 |
| Daouitherium rebouli | 12.80 | 12.80 | 12.8 |
| Arcanotherium savagei | 12.89 | 12.89 | 12.8 |
| Dugong dugon | 12.92 | 12.92 | 12.9 |
| Trichechus senegalensis | 13.03 | 13.03 | 13.0 |
| Trichechus inunguis | 13.08 | 13.08 | 13.0 |
| Protosiren smithae | 13.20 | 13.20 | 13.2 |
| Numidotherium koholense | 13.23 | 13.23 | 13.2 |
| Omanitherium dhofarensis | 13.35 | 13.35 | 13.3 |
| Trichechus manatus | 13.44 | 13.44 | 13.4 |
| Moeritherium spp | 13.82 | 13.82 | 13.8 |
| Phiomia spp | 13.89 | 13.89 | 13.8 |
| Elephas maximus | 15.02 | 15.02 | 15.0 |
| Barytherium spp | 15.20 | 15.20 | 15.2 |
| Mammuthus primigenius | 15.27 | 15.27 | 15.2 |
| Mammut borsoni | 16.49 | 16.49 | 16.4 |
| Mammuthus trogontherii | 16.38 | 16.38 | 16.3 |
| Loxodonta africana | 15.35 | 15.35 | 15.3 |
| Loxodonta cyclotis | 15.37 | 15.37 | 15.3 |
| Palaeoloxodon antiquus | 16.14 | 16.14 | 16.1 |
| Palaeoloxodon namadicus | 16.81 | 16.81 | 16.8 |
| Mammut americanum | 15.61 | 15.61 | 15.6 |
| Mammuthus columbi | 15.71 | 15.71 | 15.7 |
| Hydrodamalis gigas | 15.72 | 15.72 | 15.7 |
| Atlantogenata | 5.55 | 4.06 | 7.9 |
| Afrotheria | 5.55 | 4.05 | 7.9 |
| Afrosorcida | 4.35 | 2.58 | 6.1 |
| Macroscelidae | 5.27 | 3.98 | 6.8 |
| Pseudoungulata | 9.76 | 5.21 | 12.7 |
| Paenungulata | 10.13 | 7.24 | 13.0 |
| Tethytheria | 12.60 | 10.25 | 13.8 |
| Proboscidae | 15.23 | 14.22 | 16.2 |
| Elephantidae | 15.49 | 14.89 | 16.1 |
| Elephantina | 15.51 | 15.08 | 15.9 |
| Mammuthus | 15.54 | 15.24 | 15.8 |
| Loxodontini | 15.55 | 15.02 | 16.1 |
| Loxodona | 15.72 | 15.16 | 16.3 |
| Xenarthra | 7.57 | 5.96 | 9.1 |

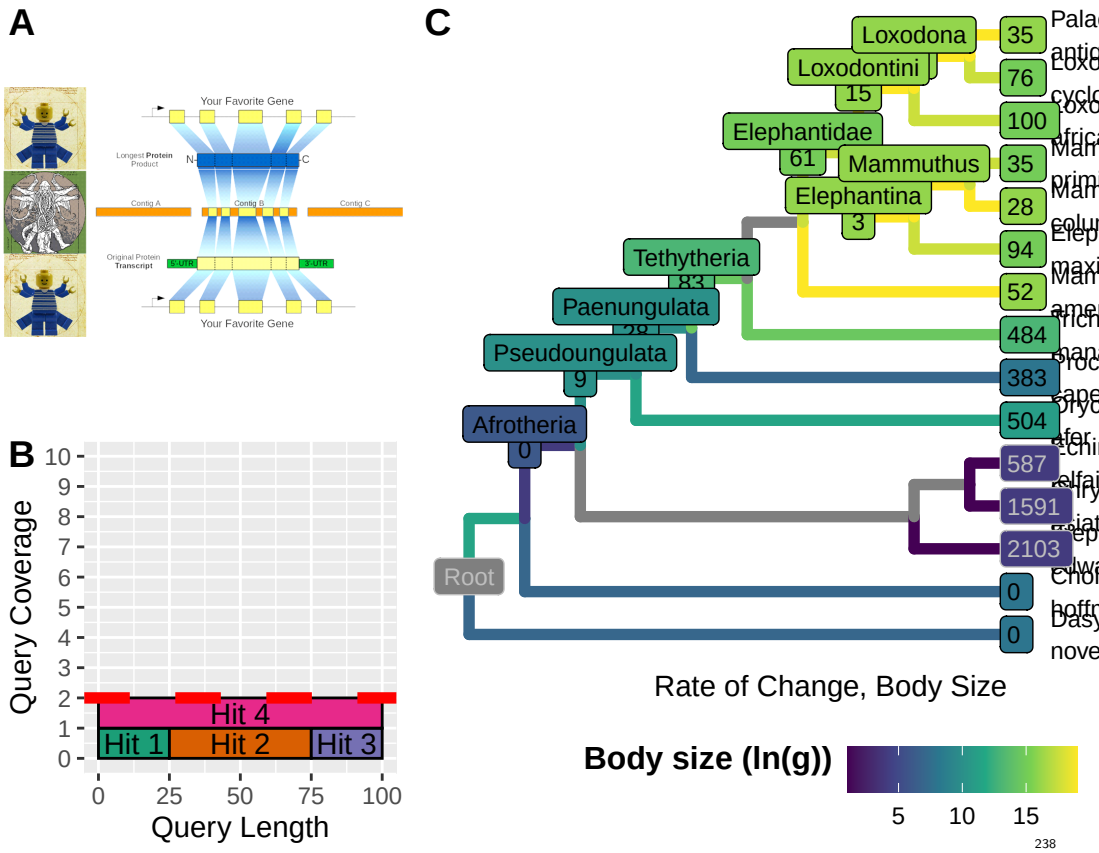
To trace the evolutionary history of body mass and lifespan in Afrotherians, we built a time-calibrated supertree of Eutherian mammals combining 1,679 species from Bininda-Emonds et al [24] with a total evidence Afrotherian phylogeny including 77 extant and fossil data from 39 extinct species [19]. Fossil data from extinct species were included to ensure that ancestral state reconstructions of body mass in Afrotherians were not biased by only including extant species, which can lead to inaccurate reconstructions, for example, if lineages multiple lineages evolved large body masses from a small bodied ancestor. We jointly estimated rates of body mass evolution and

reconstructed ancestral states using a generalization of a Brownian model of character evolution, which allows for occasional large jumps in traits (stable model) and out performs standard Brownian motion and Ornstein?Uhlenbeck models of character evolution [25].

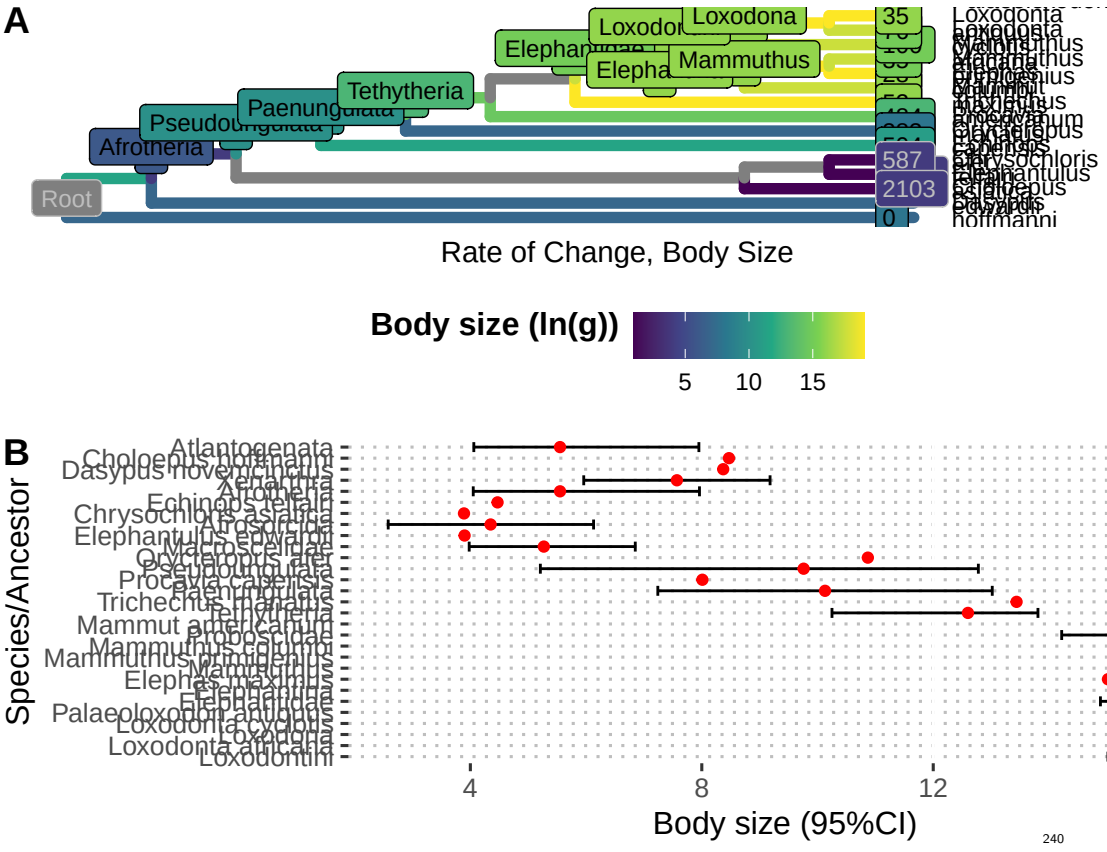
Similar to previous studies of Afrotherian body size [19,25], we found that the body mass of the Afrotherian ancestor was inferred to be small (0.26kg, 95% CI: 0.31-3.01kg) and that substantial accelerations in the rate of body mass evolution occurred coincident with a 65× increase in body mass in the stem-lineage of *Pseudoungulata* (17kg), a 1.5× increase in body mass in the stem-lineage of *Paenungulata* (25kg), a 12× increase in body mass in the stem-lineage of *Tehthytheria* (296kg), and a 14× increase in body mass in the stem-lineage of *Proboscidea* (4,100kg; Figure 1). The ancestral *Hyracoidea* was inferred to be relatively small (2.86-15.71kg), and rate accelerations were coincident with independent body mass increases in large hyraxes such as *Titanohyrax andrewsi* (67× increase in body mass). While the body mass of the ancestral *Sirenian* was inferred to be large (61-656kg), a rate acceleration occurred coincident with a 10× body mass increase in Stellar's sea cow. Rate accelerations also occurred coincident with 36× body mass reduction in the stem-lineage of the dwarf elephants *Elephas* (*Palaeoloxodon*) *falconeri* and *Palaeoloxodon cypriotes*. These data suggest that gigantism in *Afrotherians* evolved step-wise, from small to medium bodies in the *Pseudoungulata* stem-lineage, medium to large bodies in the *Tehthytherian* stem-lineage and extinct hyraxes, and from large to exceptionally large bodies independently in the *Proboscidean* stem-lineage and Stellar's sea cow (Figure 1).

## Pervasive duplication of tumor suppressors during the origins of large bodied Afrotheirans

2C1-1.pdf


A
Your Favorite Gene
Longest Protein Product
N
C
Contig A
Contig B
Contig C
Original Protein Transcript
5'-UTR
3'-UTR
Your Favorite Gene
B
Query Coverage
Query Length
Hit 1
Hit 2
Hit 3
Hit 4
C
Loxodona
Loxodontini
Elephantidae
Mammuthus
Elephantina
Tethytheria
Paenungulata
Pseudoungulata
Afrotheria
Root
Rate of Change, Body Size
Body size (ln(g))

2C2-1.pdf




```
## Warning: TODO: Table 2 is done, but Stargazer has some weird bug stopping it
## from running...
```

Enhanced cancer suppression may have evolved through many mechanisms; among the most parsimonious is an increase in the copy number of genes with tumor suppressor functions. Previous studies focusing on candidate gene studies, for example, have identified increased copy number of the tumor suppressors *TP53* and *LIF* in elephants [**???**,11,21–23]. Therefore, in order to test whether this was a pervasive phenomena genome-wide in *Afrotherians*, we used a Reciprocal Best Hit BLAT (RBHB) approach to infer gene copy number in *Afrotherian* and *Atlantogenatan* genomes (Fig. 2A). Because RBHB-like approaches can over-estimate copy number when genes are fragmented or incorrectly assembled across multiple scaffolds, we also inferred copy number using a complementary method that quantifies the ratio between observed and expected gene coverage per nucleotide (ECNC) (Supplementary Figure 1). By only including nucleotides from the query sequence that were observed in the target genome, we also correct for partial hits where some or all of the homologs of a gene have diverged from the human homolog.

Because our sequence database included various protein transcripts for each gene, in order to obtain gene-level copy number information and elimate any many-to-one mappings of hits, we labeled each exon of every reciprocal best hit (RBH) with the gene corresponding to the query transcript and merged all overlapping exons; next, we eliminated any many-to-one exons that resulted from the previous step. Finally, we reassembled the gene loci based on the original transcript starts and ends, and the collapsed exon data, obtaining the full sequence of each RBH locus. Genes were considered to be duplicated if its copy number via RBHB was greater than or equal to 2, and the maximum ECNC among all transcripts prior to filtering was greater than or

equal to 1.50. This cutoff of ECNC was selected to account for truncated gene duplications, which have been shown to be functional in various examples [**???** examples of this].

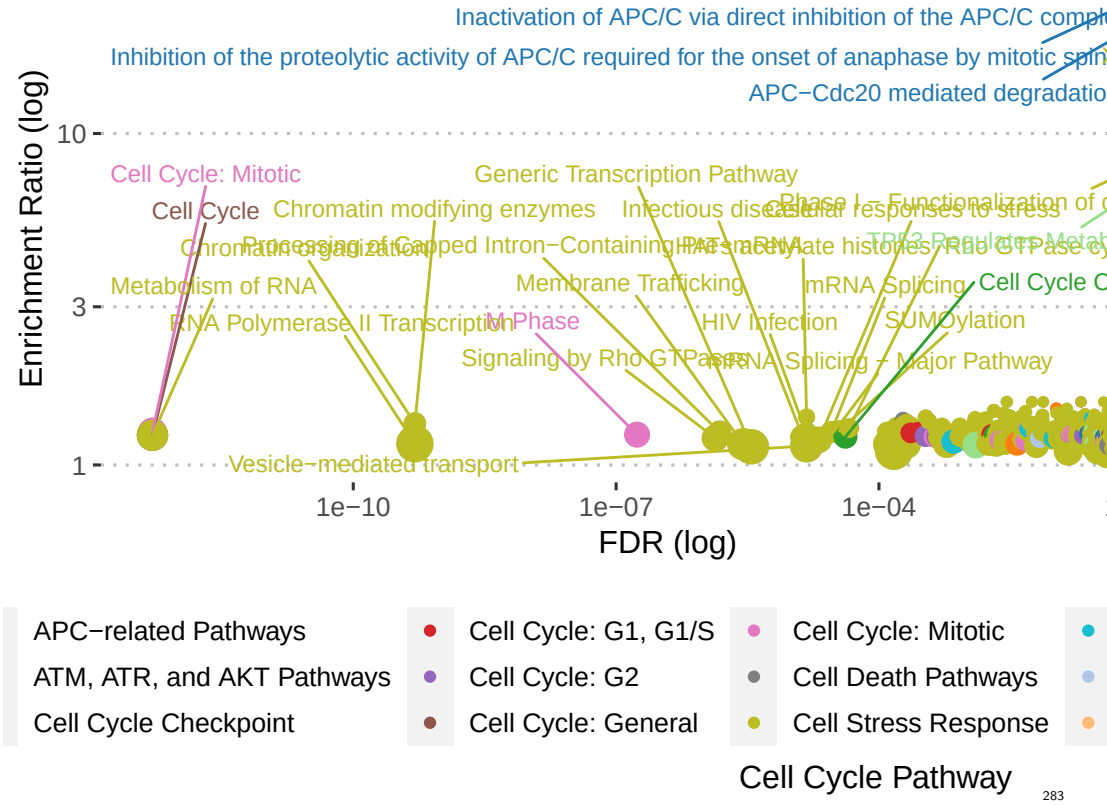
To reconcile the Atlantogenatan phylogeny with duplication events, we used maximum likelihood to reconstruct likely ancestral copy numbers for each gene at each node in the phylogeny. To define the copy number of a gene, we conservatively used the lesser value between the RBHB hit count, and the ECNC value rounded to the nearest whole number. In order to perform

Next, in order to select genes and duplicates which were likely functional, we omitted any hits that were not supported by either the gene prediction method GenScan, or by at least one transcript assembled from publically-available RNA-seq data.

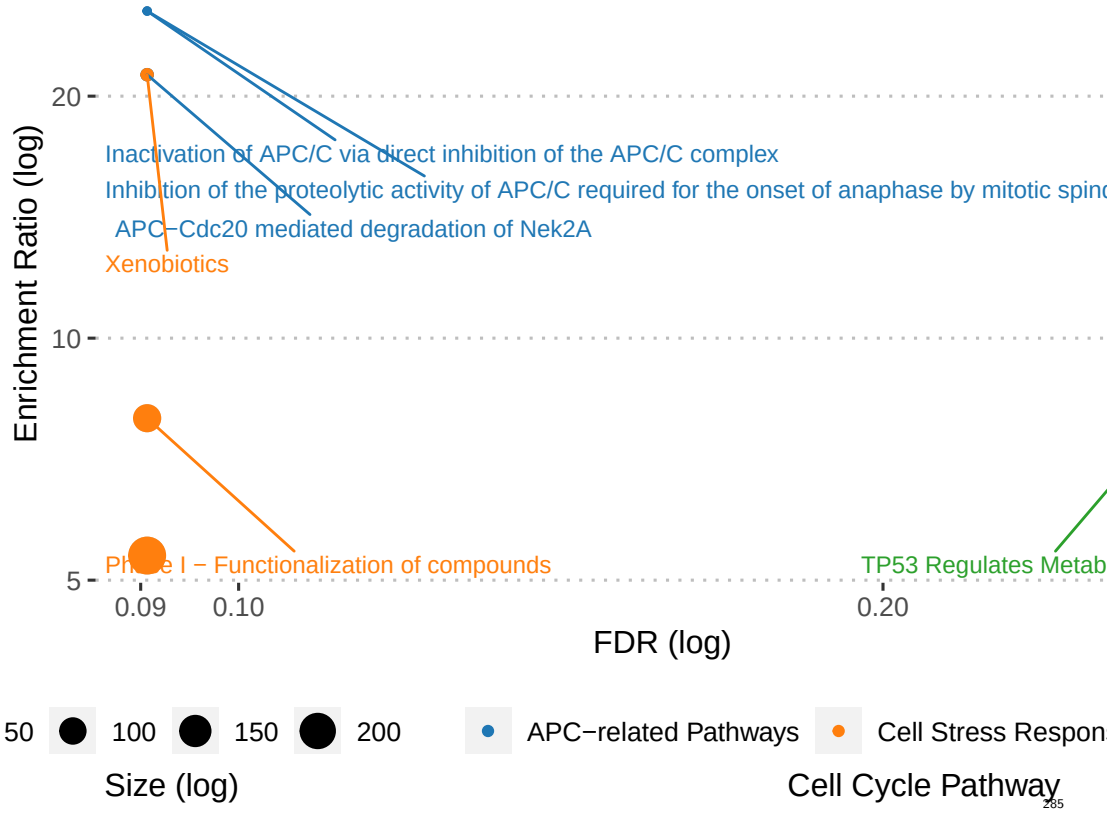
We describe the number of genes that increased in each lineage in *Atlantogenata* in Figure 2. Among the genes that increased in copy number in the elephant lineage are TP53 and LIF, as previously described. Furthermore, we identify

## Duplications that occured recently in Probodiscea are enriched for tumor suppressor pathways

3-1.pdf




3-2.pdf




In order to infer the functional consequences of these gene duplications, we tested if duplicate genes were enriched in specific pathways relative to our initial query set of genes. We used

## Concerted duplication of TP53 and TP53-related genes towards *Probodiscea*

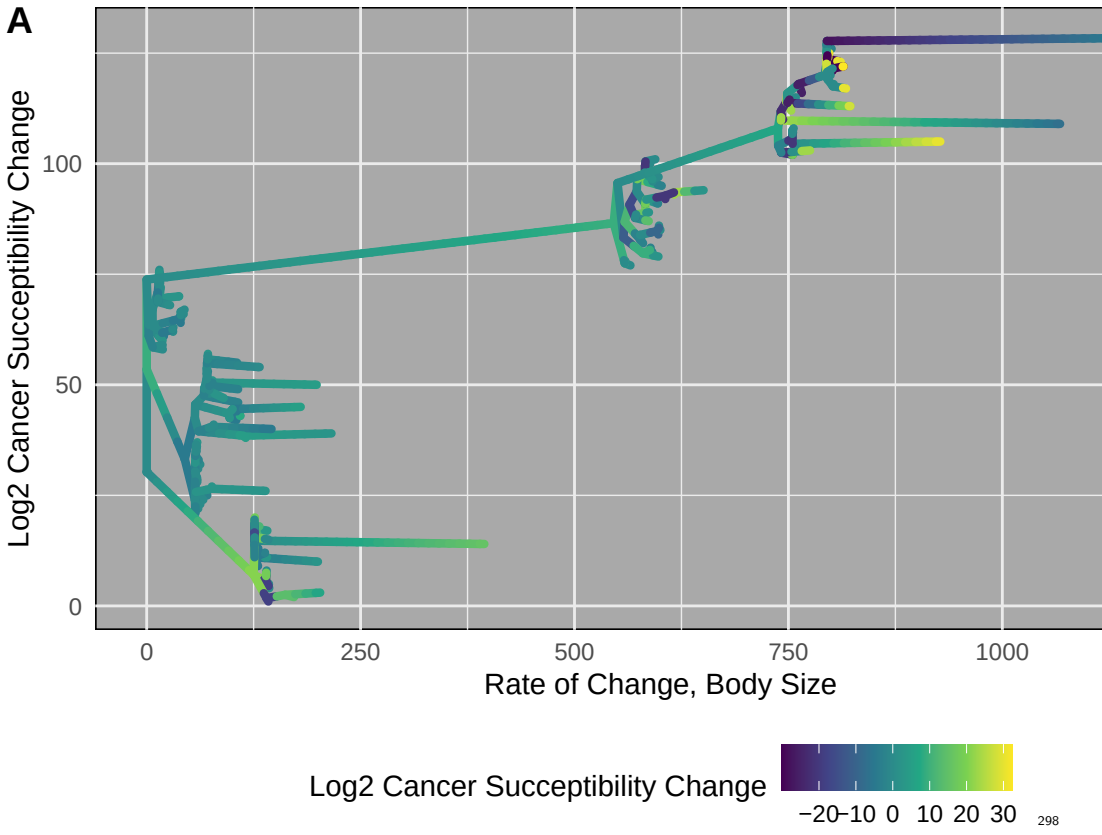
```
## Warning: TODO: Make Fig4
```

## Step-wise reduction of intrinsic cancer risk in large, long-lived Afrotherians

```
## Warning: TODO: Caption Fig 5

## Warning: TODO: de-comment geom_tiplab in f5a! Commented currently for offlin
## usage!
```

5-1.pdf




```
##
## % Table created by stargazer v.5.2.2 by Marek Hlavac, Harvard University E-
## % Date and time: Wed, May 13, 2020 - 20:08:58
## \begin{table}[!htbp] \centering
##   \caption{Phylogenetic Least Squares: ln(Lifespan) & ln(Body Size) Regressi
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{1}{c}{\textit{Dependent variable:}} \\
## \cline{2-2}
## \\[-1.8ex] & Lifespan \\
## \hline \\[-1.8ex]
##  lnSize & 0.100 \\
##   & (0.121) \\
##   & t = 0.826 \\
##   & p = 0.409 \\
##   Constant & 1.943 \\
##   & (1.385) \\
##   & t = 1.403 \\
##   & p = 0.161 \\
##  \hline \\[-1.8ex]
## Observations & 28 \\
## Log Likelihood & $-$39.636 \\
## Akaike Inf. Crit. & 85.273 \\
## Bayesian Inf. Crit. & 89.047 \\
```

```
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$
## \end{tabular}
## \end{table}
```

**Table 2.** A table generated by the longtable package.

| Node | lnSize | Est. Lifespan | Est. Cancer Susceptibility (K1) | Est. Anc |
|---|---|---|---|---|
| Loxodontini | 16 | 34.38 | 1.47e+16 | 2.97e+15 |
| Loxodonta africana | 15 | 65.00 | 2.47e+17 | 1.47e+16 |
| Loxodona | 16 | 34.38 | 1.47e+16 | 1.47e+16 |
| Loxodonta cyclotis | 15 | 31.12 | 2.97e+15 | 1.47e+16 |
| Palaeoloxodon antiquus | 16 | 34.38 | 1.47e+16 | 1.47e+16 |
| Elephantidae | 15 | 31.12 | 2.97e+15 | 1.40e+07 |
| Elephantina | 16 | 34.38 | 1.47e+16 | 2.97e+15 |
| Elephas maximus | 15 | 65.50 | 2.58e+17 | 1.47e+16 |
| Mammuthus | 16 | 34.38 | 1.47e+16 | 1.47e+16 |
| Mammuthus primigenius | 15 | 31.12 | 2.97e+15 | 1.47e+16 |
| Mammuthus columbi | 16 | 34.38 | 1.47e+16 | 1.47e+16 |
| Mammut americanum | 16 | 34.38 | 1.47e+16 | 1.40e+07 |
| Tethytheria | 13 | 25.49 | 1.21e+14 | 1.01e+12 |
| Trichechus manatus | 13 | 69.00 | 4.77e+16 | 1.21e+14 |
| Paenungulata | 10 | 18.91 | 1.01e+12 | 1.01e+12 |
| Procavia capensis | 8 | 14.80 | 3.13e+10 | 1.01e+12 |
| Pseudoungulata | 10 | 18.91 | 1.01e+12 | 1.69e+09 |
| Orycteropus afer | 11 | 29.80 | 4.19e+13 | 1.01e+12 |
| Elephantulus edwardii | 4 | 10.40 | 6.90e+07 | 1.69e+09 |
| Afrosorcida | 4 | 10.40 | 6.90e+07 | 1.69e+09 |
| Chrysochloris asiatica | 4 | 10.40 | 6.90e+07 | 6.90e+07 |
| Echinops telfairi | 4 | 19.00 | 2.57e+09 | 6.90e+07 |
| Afrotheria | 6 | 12.69 | 1.69e+09 | 2.83e+06 |
| Xenarthra | 11 | 20.89 | 4.97e+12 | 2.83e+06 |
| Dasypus novemcinctus | 8 | 22.30 | 3.67e+11 | 4.97e+12 |
| Choloepus hoffmanni | 8 | 41.00 | 1.42e+13 | 4.97e+12 |
| Atlantogenata | 2 | 8.52 | 2.83e+06 | 2.83e+06 |
| Afroinsectivora | 6 | 12.69 | 1.69e+09 | 1.69e+09 |
| NA | 3 | 9.41 | 1.40e+07 | 1.21e+14 |

The dramatic increase in body mass and lifespan in some Afrotherian lineages implies those lineages evolved reduced cancer risk. To infer the magnitude of these reductions we estimated differences in cancer risk between small bodied, short-lived species and large bodied, long-lived species as well as for reconstructed ancestral Afrotherians. Following [**???**] we estimate the intrinsic cancer risk as the product of risk associated with body mass and lifespan. Differences in cancer susceptibility $K$ due to body mass differences between species can be approximated simply as the fold difference in body mass ($D$) between species [**???**]. The risk of developing cancer also increases in proportion to the sixth power of age and is approximated by the formula $Ct^6$, in which the proportionality constant C that determines susceptibility to cancer induction is multiplied by the sixth power of the age in years, $t$ [**???**,**??**,**??**]. Thus we can estimate the intrinsic cancer risk for a species as $K \approx Dt^6$.

In order to estimate the intrinsic cancer risk of a species, we first obtained estimates

for lifespans at ancestral nodes using PGLS and the model $ln(lifespan) = \beta_1 corBrownian + \beta_2 ln(Size) + \epsilon$ (Figure ). With this information in hand, we calculated $K_1$ at all nodes, and then estimated the fold change in cancer susceptibility between an ancestral node and a given node as $\frac{K_2}{K_1}$ (**Table 4**).

As shown in **Table 4**, cancer susceptibility skyrocketed at the initial divergence of Atlantogenata, followed by a generally upwards trend. At the common ancestor of Afrotheria there is an inital 9.22-fold increase in cancer risk. In parallel to Afrotheria, cancer succeptibility increases 20.75-fold in Xenarthra. However, cancer risk slowly deflates as size decreases as one moves along the tree towards extant species, such as in Hoffman's Two Toed Sloth (-fold change) and in the Nine-banded Armadillo (-fold change).

Within Afrotheria, cancer succeptibility drops in Afrosorcida as species shrink (-4.61-fold, then stagnates for the Cape Golden Mole) - but then rises -fold towards the Lesser Hedgehog Tenrec. In parallel, Afroinsectivora does not increase in cancer succeptibility, and decreases once more at the Cape Elephant Shrew (-fold). The emergence of Pseudoungulata sees the next big leap in cancer succeptibility with a 9.22-fold increase. The Aardvark further increases -fold, while we don't observe an increase at the common ancestor of Paenungulates. While the Rock Hyrax decreases in cancer succeptibility as expected (-fold), Tethytheria sees a sharp increase in cancer risk (6.92-fold). Within Tethytheria, the Manatee's cancer risk increases once more -fold, while Proboscidae's cancer risk drops precipitously along with its body size (-23.05-fold). Yet, within Proboscidae we see the biggest increases: right off the bat, we see that the cancer succeptibility of Elephantidae and the American Mastodon skyrocket by 27.66-fold and -fold, respectively. Both Elephantina and Loxodontini in Elephantidae have a 2.31-fold increase in cancer succeptibility. Within Elephantina, cancer susceptibility stays stable at Mammuthus and in the Colombian Mammoth, and slightly decreases in the Wooly Mammoth (-fold). The three extant elephants - Asian Elephant in Elephantina, the African Savana Elephant in Loxodontini, and the African Forest Elephant in Loxodona, meanwhile, have parallel and similar decreases in both size and cancer succeptibility (-, -, and -fold, respectively). Neither the common ancestor of Loxodonta, nor the Straight-Tusked Mammoth see any further changes in cancer succeptibility.

# Discussion

What biological mechanisms underlie the evolution of thousand- to hundred million-increases in cancer susceptibility during the origins of Afrotherians, which are essential for large body size and long lifespan to evolve

Candidate gene studies, for example, have identified functional duplicates of the tumor suppressors TP53 and LIF in elephants. In a larger candidate gene study, Caulin et al. characterized the copy number of 830 known tumor-suppressor genes across 36 mammals and identified 382 putative duplicates, including duplicates in species with large body sizes and long life-spans. However, the probability of developing cancer is similar for small, short-lived mammals such as mice and for large, long-lived mammals such as elephants.

In stark contrast, genome-wide studies of unusually large or long-lived species such as the bowhead whale (Keane et al., 2015), Myotid bats (Seim et al., 2013; Zhang et al., 2013), naked mole rat (Kim et al., 2011), and blind mole rat (Fang et al., 2014) did not find an over representation of tumor suppressors among duplicate genes.

A genomic analysis of genetic changes associated with the evolution of enhanced cancer resistance in the elephant lineage has yet to be performed. Thus it is not clear if the duplication of TP53 and LIF reflects a general pattern of tumor suppressor

duplication in the elephant lineage, unlike other lineages that resolved Peto?s paradox, or from the kinds of ascertainment biases common in candidate gene studies.


## Acknowledgements

I would like to thank Olga Duchenko at the Aiden Lab, D.H. Vazquez for his indispensible support.


## Conflicts of Interest

The Authors have no conflicts of interest to report


## Funding Source

We would like to thank the Department of Human Genetics at the University of Chicago for supporting this project.


# SUPPLEMENTARY FIGURES

```
## Warning: TODO: Make HQ version of this figure
## Warning: Removed 1 rows containing missing values (geom_text).
## $RBX1
##
## $LIN9
##
## $MND1
##
## $E2F2
##
## $MAX
##
## $SYCP3
##
## $CDK1
##
## $RNF168
##
## $RBBP8
##
## $MAD2L1
## Warning: TODO: this causes a segfault, find out why!
```

Figure 1-1.pdf

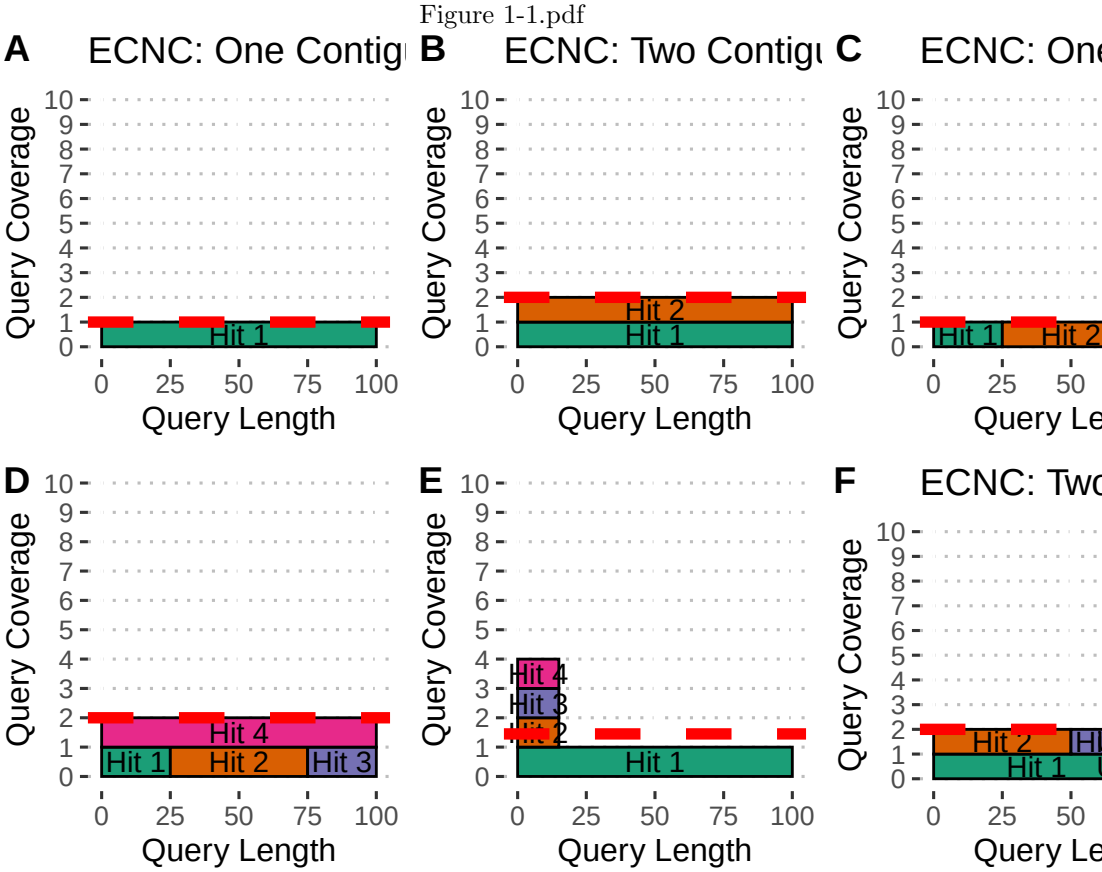



**Fig 2.** Supplementary Figure 1: Estimated Copy Number by Coverage (ECNC) consolidates fragmented genes while accounting for missing domains in homologs. **A)** A single, contiguous gene homolog in a target genome with 100% query length coverage has an ECNC of 1.0. **B)** Two contiguous gene homologs, each with 100% query length coverage have an ECNC of 2.0. **C)** A single gene homolog, split across multiple scaffolds and contigs in a fragmented target genome; BLAT identifies each fragment as a single hit. Per nucleotide of query sequence, there is only one corresponding nucleotide over all the hits, thus the ECNC is 1.0. **D)** Two gene homologs, one fragmented and one contiguous. 100% of nucleotides in the query sequence are represented between all hits; however, every nucleotide in the query has two matching nucleotides in the target genome, thus the ECNC is 2.0. **E)** One true gene homolog in the target genome, plus multiple hits of a conserved domain that span 20% of the query sequence. While 100% of the query sequence is represented in total, 20% of the nucleotides have 4 hits. Thus, the ECNC for this gene is 1.45. **F)** Two real gene homologs; one hit is contiguous, one hit is fragmented in two, and the tail end of both sequences was not identified by BLAT due to sequence divergence. Only 75% of the query sequence was covered in total between the hits, but for that 75%, each nucleotide has two hits. As such, ECNC is equal to 2.0 for this gene.

Figure 2-1.pdf

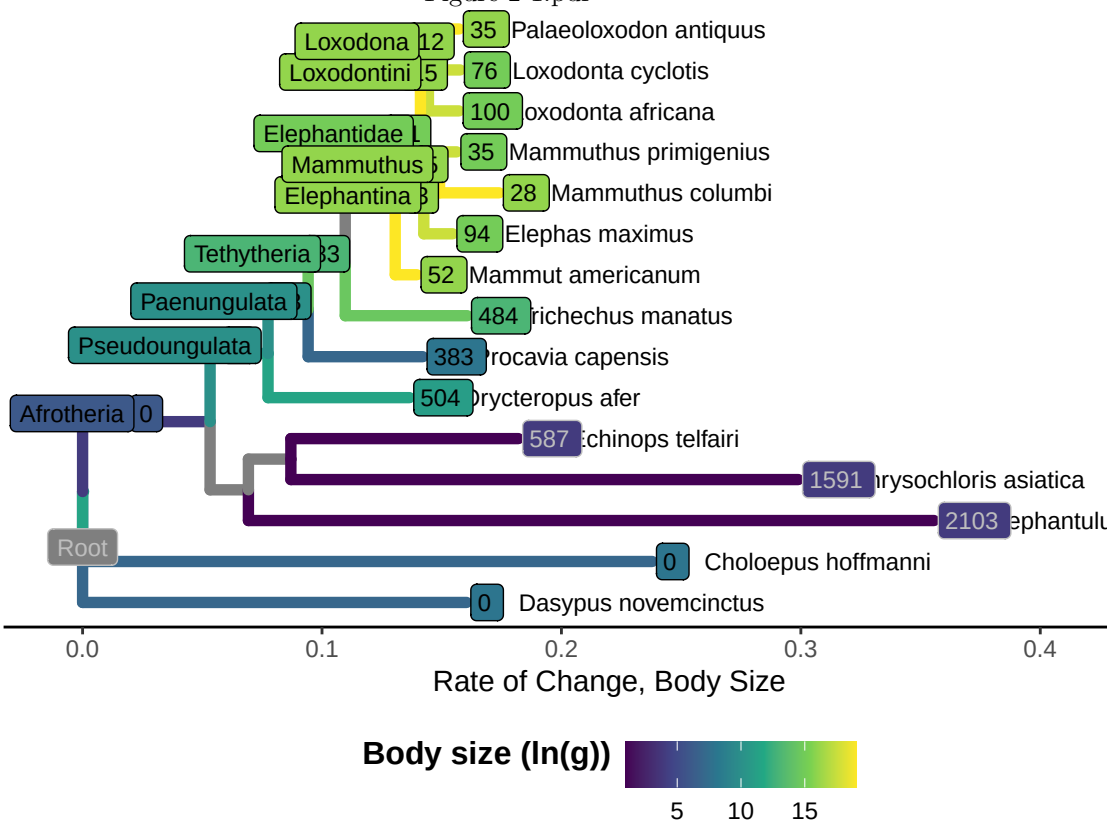



**Fig 3.** Supplementary Figure 2: Gene copy increases polarized along Atlantogenata, colored by ln(Body Size), with branch lengths equal to the change in gene copy number

Figure 3-1.pdf

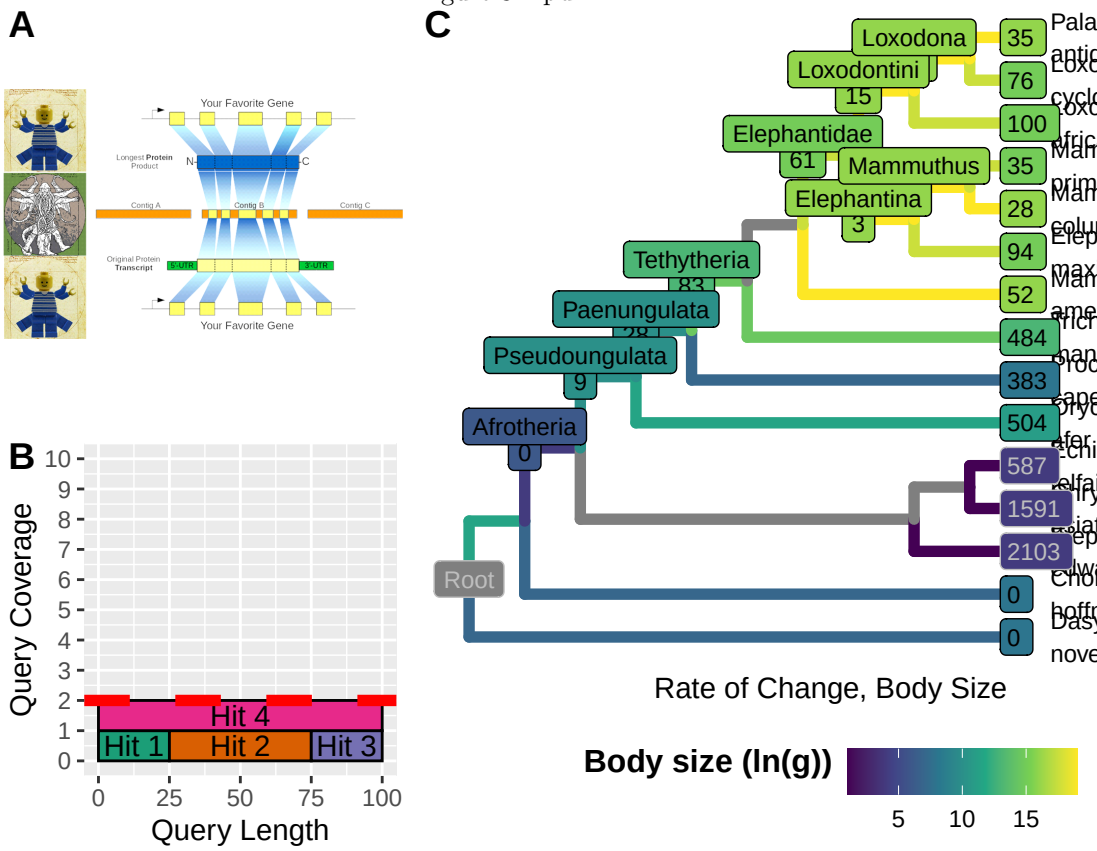



**Fig 4.** Supplementary Figure 3: Full version of RecBlat strategy, but low quality

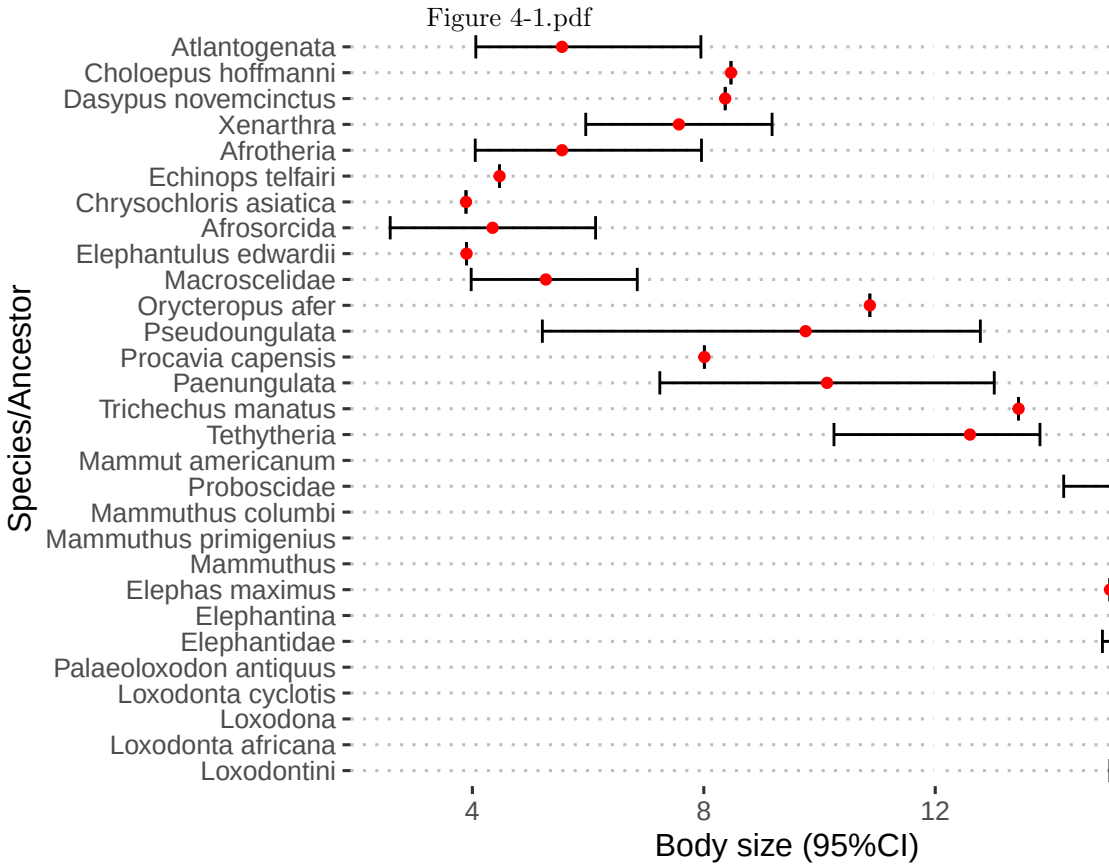



**Fig 5.** Supplementary Figure 4: Dot and Line plot for Body Sizes and such

Figure 5-1.pdf

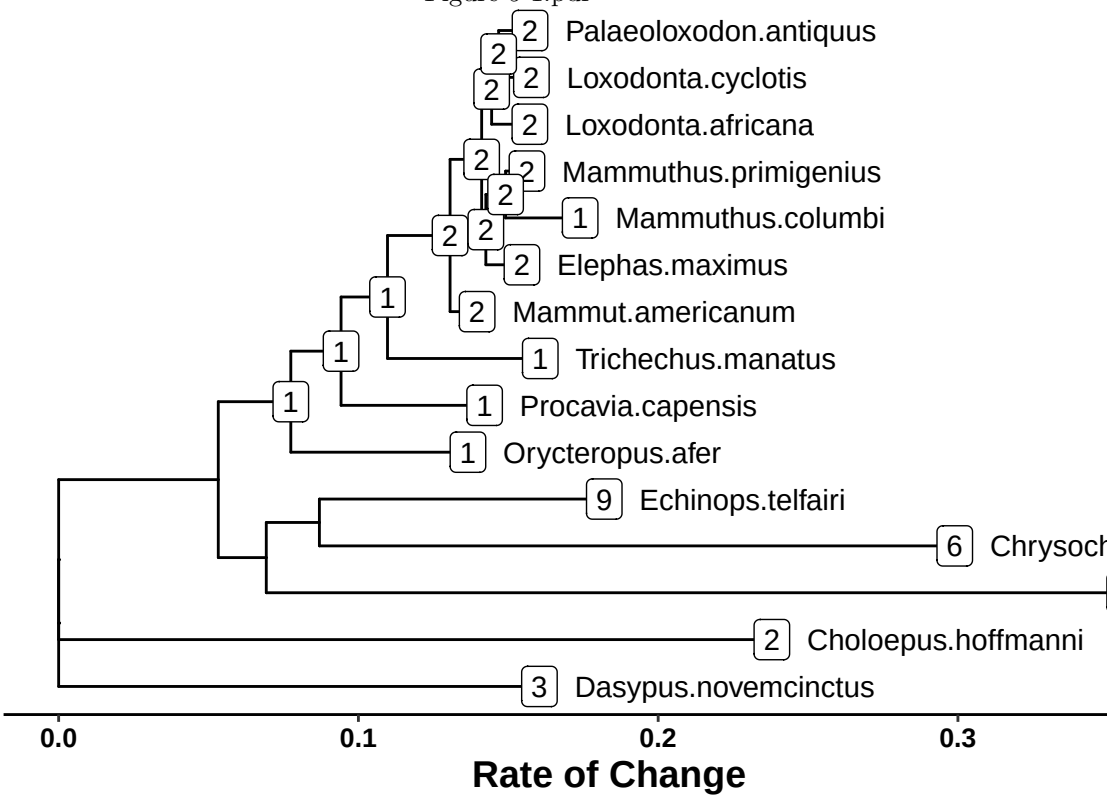



**Fig 6.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-2.pdf

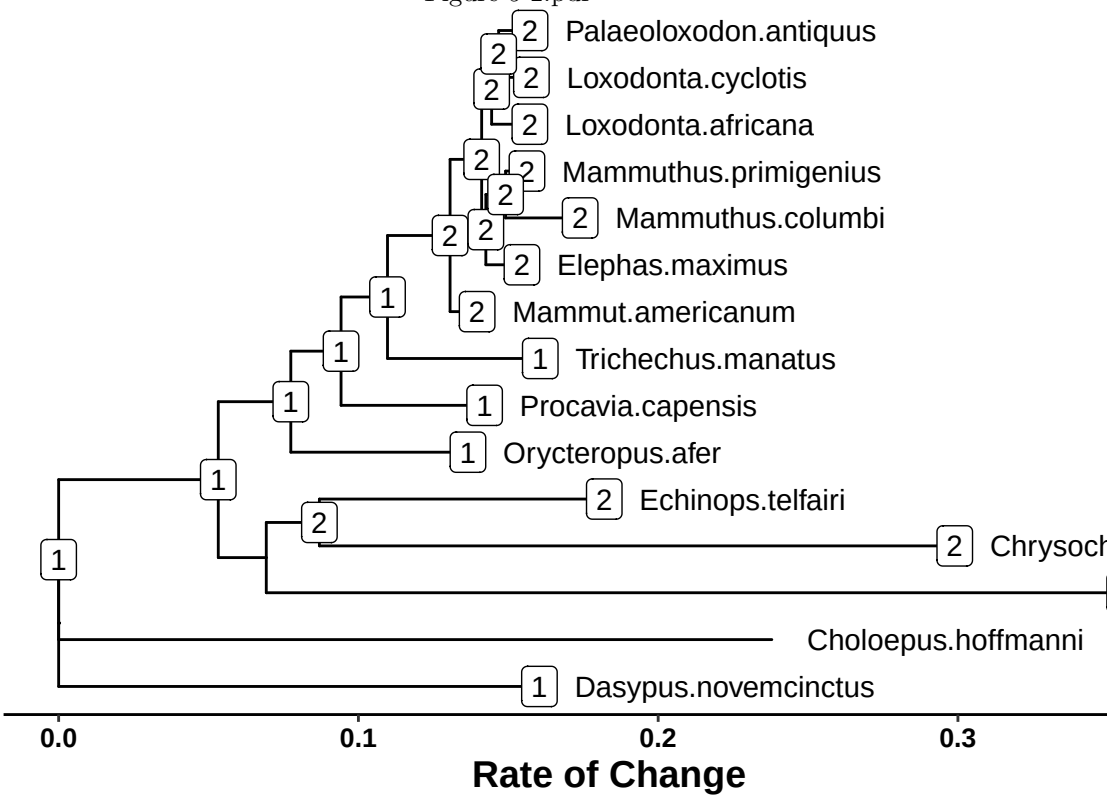



**Fig 7.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-3.pdf

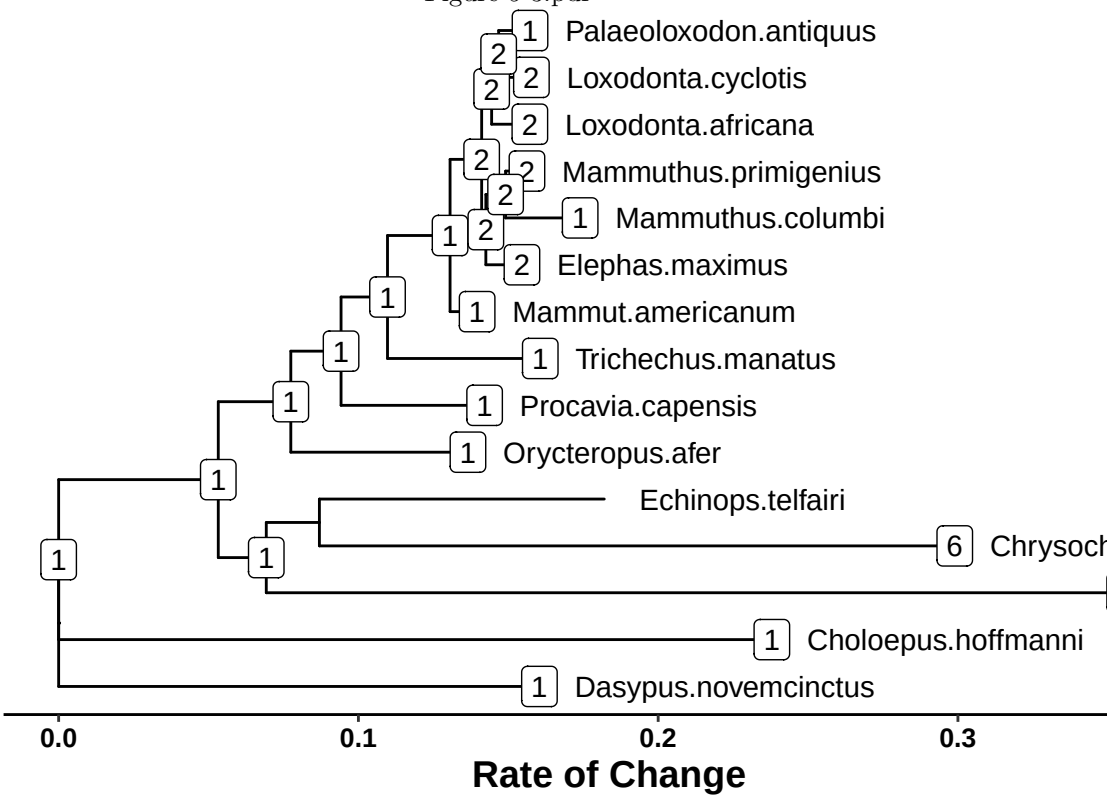



**Fig 8.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-4.pdf

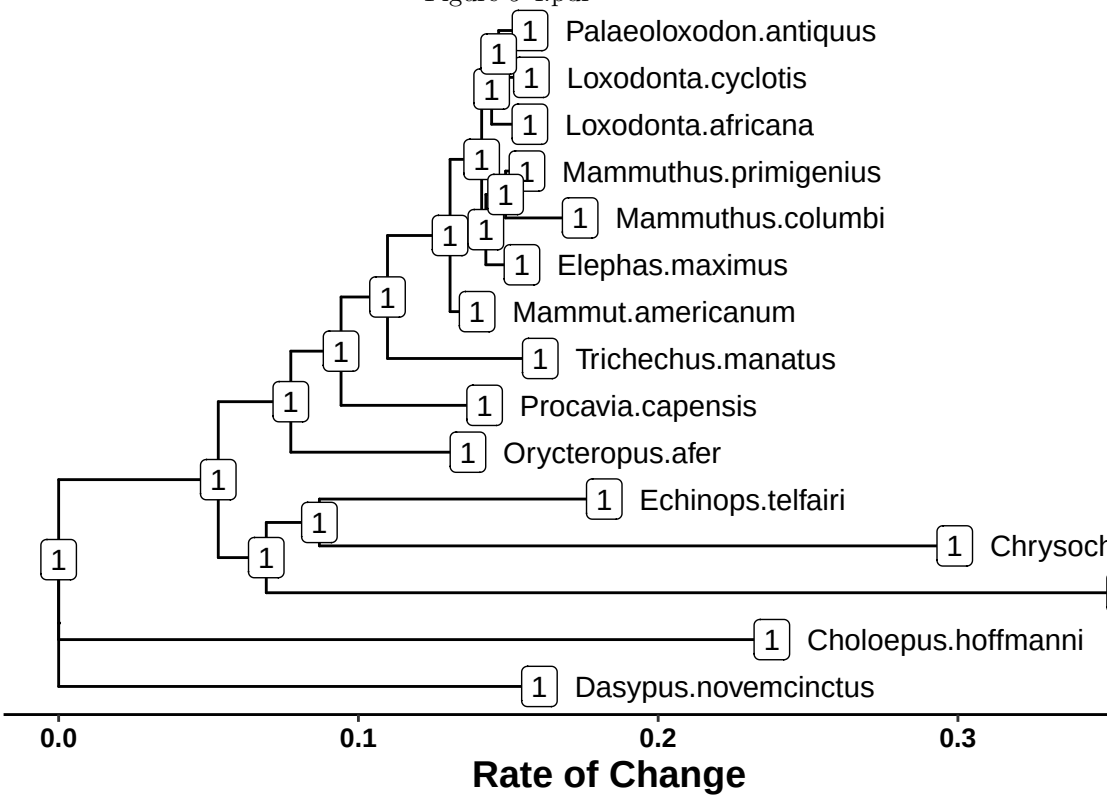



**Fig 9.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-5.pdf

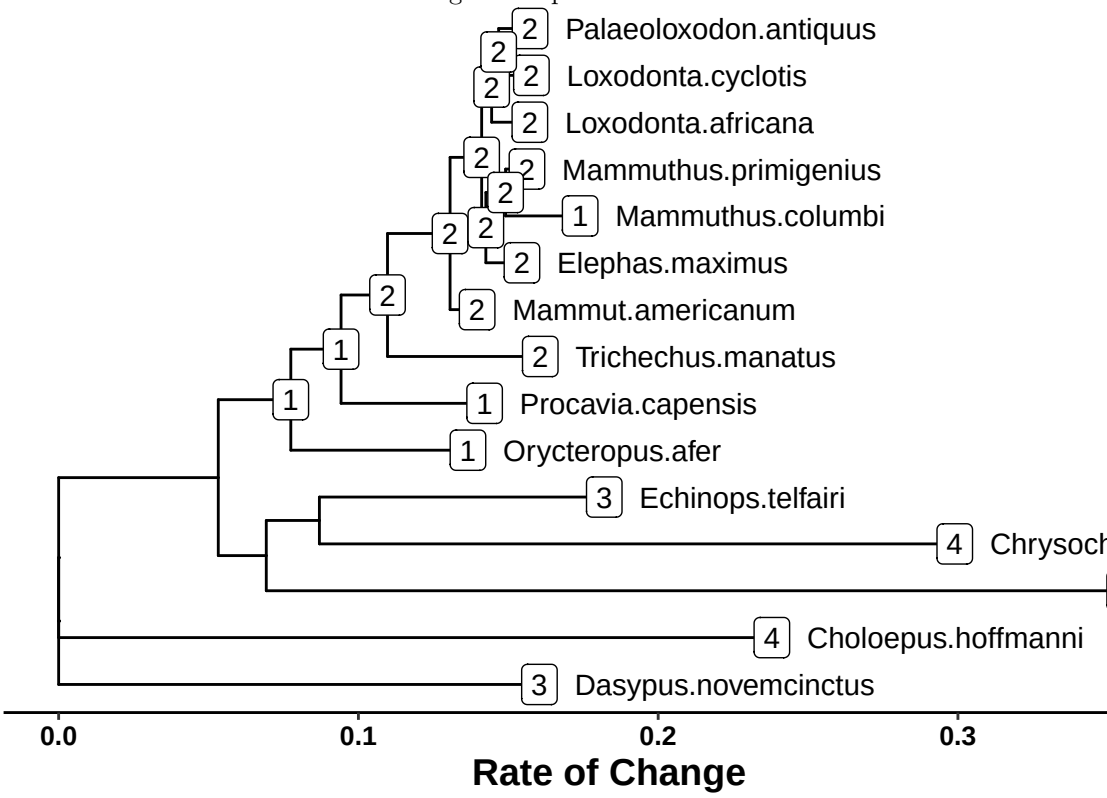



**Fig 10.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-6.pdf

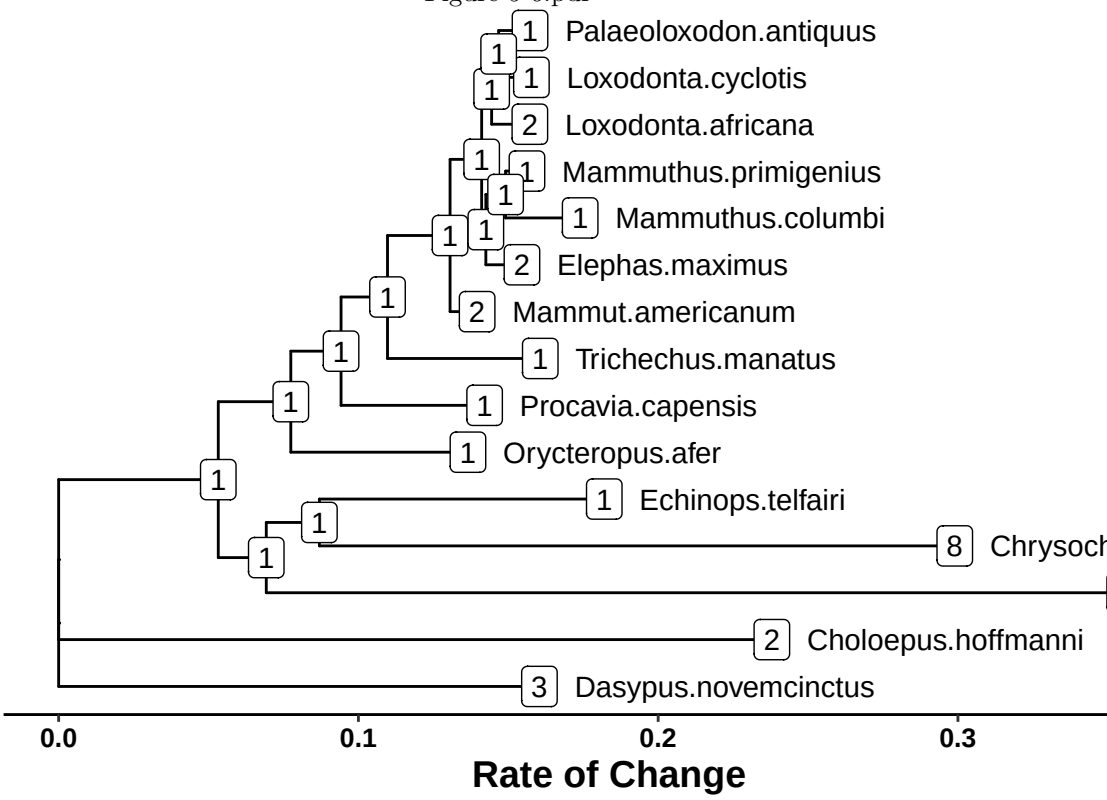



**Fig 11.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-7.pdf

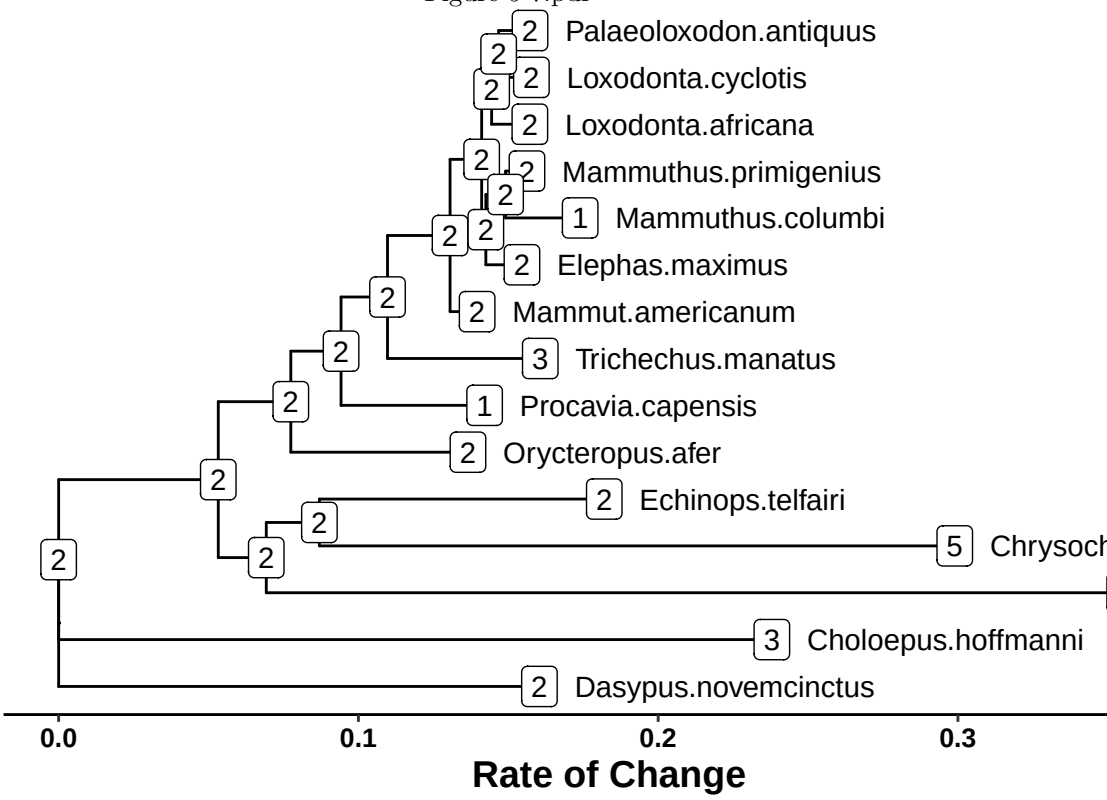



**Fig 12.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-8.pdf

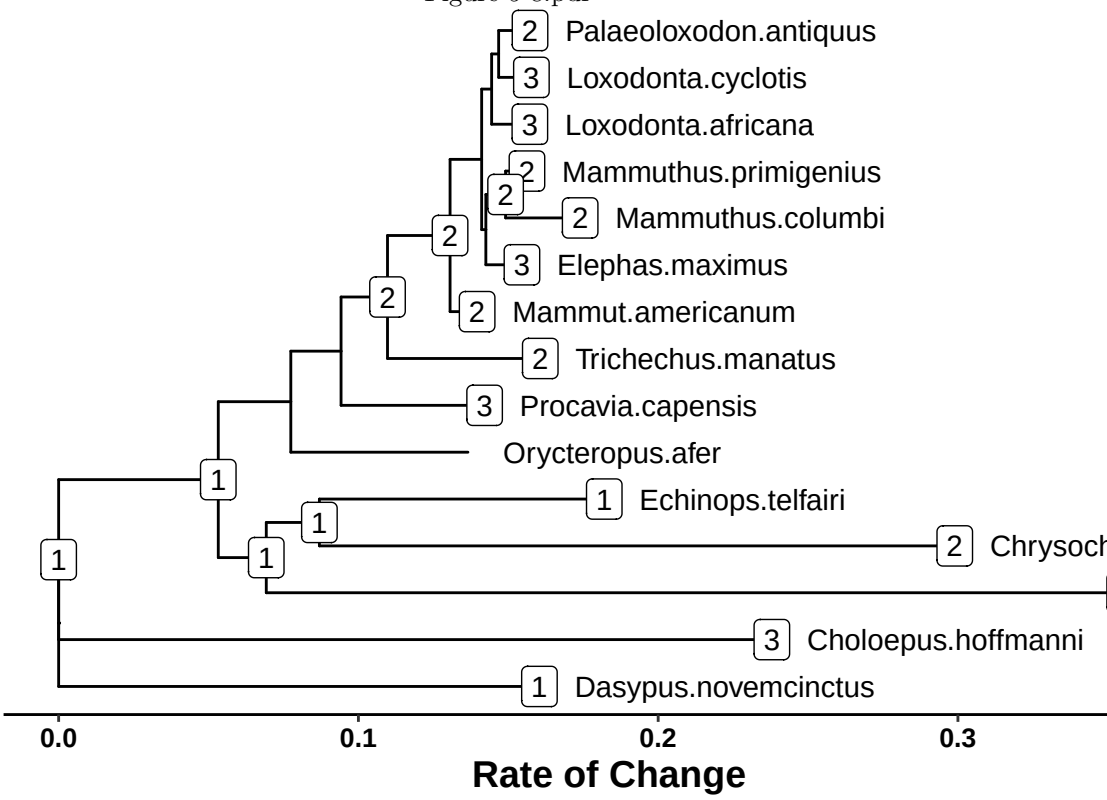



**Fig 13.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-9.pdf

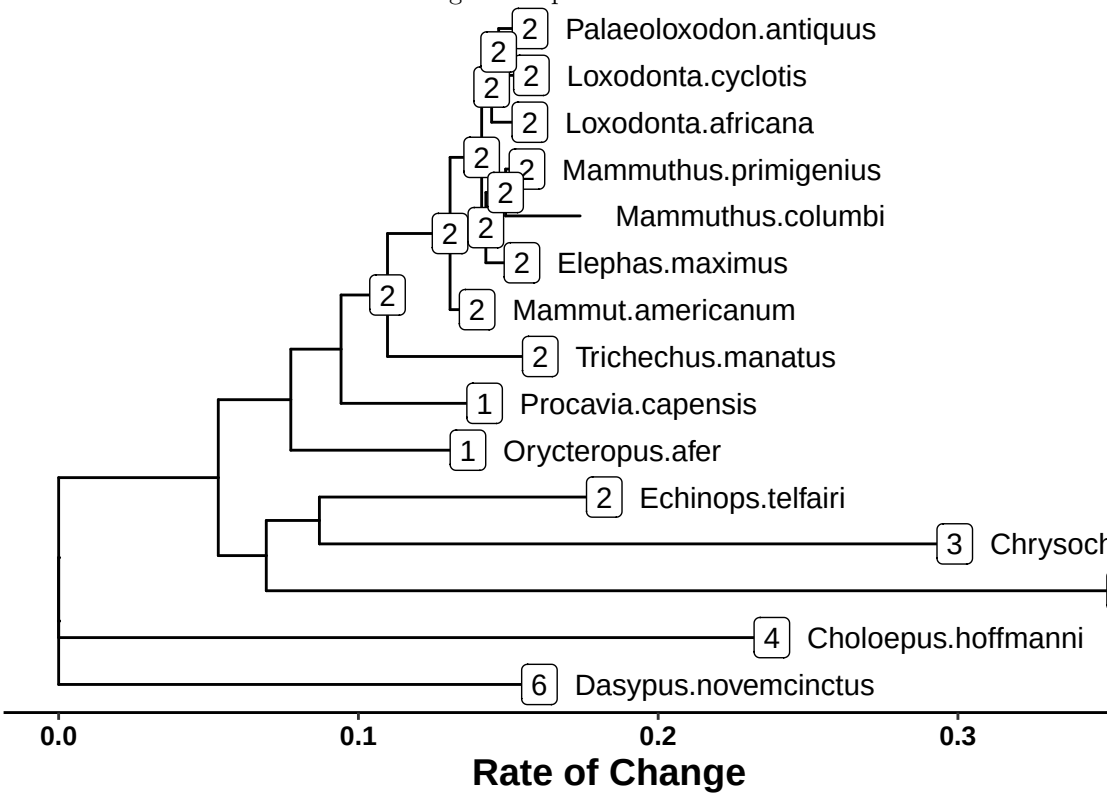



**Fig 14.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 5-10.pdf

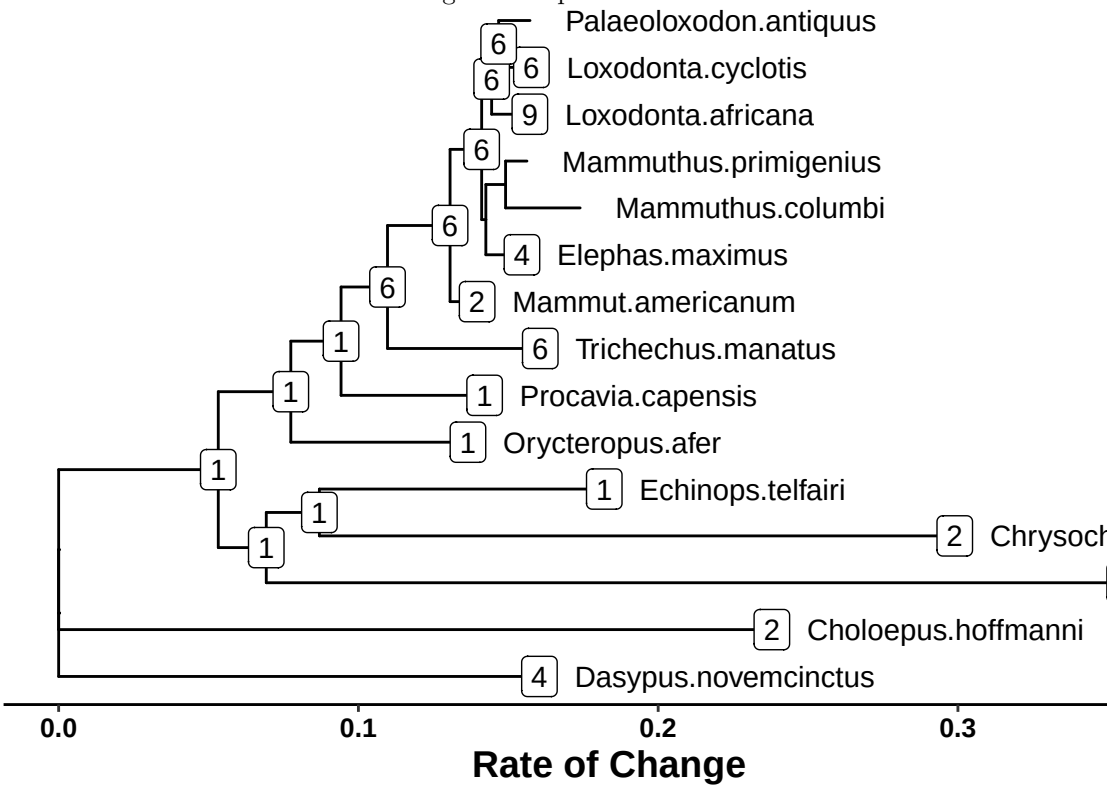



**Fig 15.** Supplementary Figure 5: All the Gene Copy Trees for interesting genes duplicated in LoxAfr4 (RIP any trees if this gets printed)

Figure 7-1.pdf

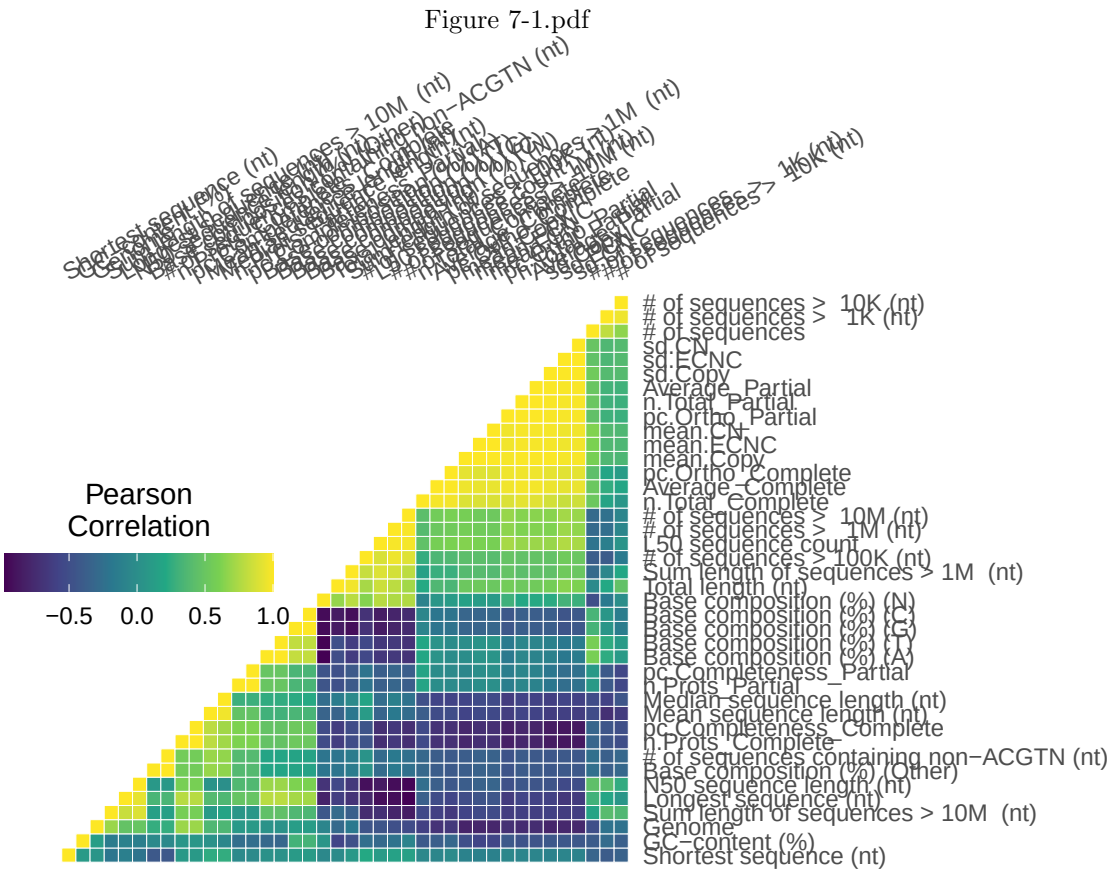



**Fig 16.** Supplementary Figure 7: Correlation matrix heat map for genome quality metrics, ECNC, RBHB Copy, and Estimated Copy Number (lesser between ECNC and RBHB)

## Supplementary data

```
## Warning: TODO: Read in Eutheria.progress and present it to the bold.
```

## References


1. Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V, et al. Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. The Lancet Oncology. 2011;12: 785–794. doi:10.1016/s1470-2045(11)70154-1

2. Nunney L. Size matters: height, cell number and a person's risk of cancer. Proc R Soc B. 2018;285: 20181743. doi:10.1098/rspb.2018.1743

3. Dobson JM. Breed-predispositions to cancer in pedigree dogs. ISRN veterinary science. 2013;2013: 941275. doi:10.1155/2013/941275

4. Caulin AF, Maley CC. Peto's Paradox: evolution's prescription for cancer prevention. Trends in ecology & evolution. 2011;26: 175–82. doi:10.1016/j.tree.2011.01.002

5. Leroi AM, Koufopanou V, Burt A. Cancer selection. Nature Reviews Cancer. 2003;3: 226–231. doi:10.1038/nrc1016

6. Peto R, Roe F, Lee P, Levy L, Clack J. Cancer and ageing in mice and men. British Journal of Cancer. 1975;32: 411–426. doi:10.1038/bjc.1975.242

7. Ashur-Fabian O, Avivi A, Trakhtenbrot L, Adamsky K, Cohen M, Kajakaro G, et al. Evolution of p53 in hypoxia-stressed Spalax mimics human tumor mutation. Proceedings of the National Academy of Sciences. 2004;101: 12236–12241. doi:10.1073/pnas.0404998101

8. Seluanov A, Hine C, Bozzella M, Hall A, Sasahara THC, Ribeiro AACM, et al. Distinct tumor suppressor mechanisms evolve in rodent species that differ in size and lifespan. Aging cell. 2008;7: 813–23. doi:10.1111/j.1474-9726.2008.00431.x

9. Gorbunova V, Hine C, Tian X, Ablaeva J, Gudkov AV, Nevo E, et al. Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. Proceedings of the National Academy of Sciences of the United States of America. 2012;109: 19392–6. doi:10.1073/pnas.1217211109

10. Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, Ablaeva J, et al. High molecular weight hyaluronan mediates the cancer resistance of the naked mole-rat. 2013;499. doi:10.1038/nature12234

11. Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, et al. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. eLife. 2016;5: e11994. doi:10.7554/elife.11994

12. Tacutu R, Craig T, Budovsky A, Wuttke D, Lehmann G, Taranukha D, et al. Human Ageing Genomic Resources: Integrated databases and tools for the biology and genetics of ageing. Nucleic Acids Research. 2013;41: D1027–D1033. doi:10.1093/nar/gks1155

13. Schwartz GT, Rasmussen DT, Smith RJ. Body-Size Diversity and Community Structure of Fossil Hyracoids. Journal of Mammalogy. 1995;76: 1088–1099. doi:10.2307/1382601

14. Scheffer VB. The Weight of the Steller Sea Cow. Journal of Mammalogy. 1972;53: 912–914. doi:10.2307/1379236

15. Larramendi A. Shoulder Height, Body Mass, and Shape of Proboscideans. Acta Palaeontologica Polonica. 2015;61. doi:10.4202/app.00136.2014

16. O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, et al. The placental mammal ancestor and the post-K-Pg radiation of placentals. Science

(New York, NY). 2013;339: 662–7. doi:10.1126/science.1229237

17. Springer MS, Meredith RW, Teeling EC, Murphy WJ. Technical comment on "The placental mammal ancestor and the post-K-Pg radiation of placentals". Science (New York, NY). 2013;341: 613. doi:10.1126/science.1238025

18. O'Leary MA, Bloch JI, Flynn JJ, Gaudin TJ, Giallombardo A, Giannini NP, et al. Response to comment on "The placental mammal ancestor and the post-K-Pg radiation of placentals". Science (New York, NY). 2013;341: 613. doi:10.1126/science.1238162

19. Puttick MN, Thomas GH. Fossils and living taxa agree on patterns of body mass evolution: a case study with Afrotheria. Proceedings Biological sciences / The Royal Society. 2015;282: 20152023. doi:10.1098/rspb.2015.2023

20. Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, et al. Potential Mechanisms for Cancer Resistance in Elephants and Comparative Cellular Response to DNA Damage in Humans. JAMA. 2015;314: 1850–1860. doi:10.1001/jama.2015.13134

21. Vazquez JM, Sulak M, Chigurupati S, Lynch VJ. A Zombie LIF Gene in Elephants Is Upregulated by TP53 to Induce Apoptosis in Response to DNA Damage. Cell Reports. 2018;24: 1765–1776. doi:10.1016/j.celrep.2018.07.042

22. Caulin AF, Graham TA, Wang L-S, Maley CC. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2015;370: 20140222. doi:10.1098/rstb.2014.0222

23. Doherty A, Magalhães J de. Has gene duplication impacted the evolution of Eutherian longevity? Aging Cell. 2016;15: 978–980. doi:10.1111/acel.12503

24. Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, et al. Erratum: The delayed rise of present-day mammals. Nature. 2008;456: 274–274. doi:10.1038/nature07347

25. Elliot MG, Mooers AØ. Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. BMC evolutionary biology. 2014;14: 226. doi:10.1186/s12862-014-0226-8

26. Kent JW. BLAT—The BLAST-Like Alignment Tool. Genome Research. 2002;12: 656–664. doi:10.1101/gr.229202

27. Altenhoff AM, Dessimoz C. Phylogenetic and functional assessment of orthologs inference projects and methods. PLoS computational biology. 2009;5: e1000262. doi:10.1371/journal.pcbi.1000262

28. Consortium TU. UniProt: the universal protein knowledgebase. Nucleic Acids Research. 2017;45: D158–D169. doi:10.1093/nar/gkw1099