nature ecology & evolution

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41559-018-0733-x

In the format provided by the authors and unedited.

Giant tortoise genomes provide insights into longevity and age-related disease

Víctor Quesada[1,19], Sandra Freitas-Rodríguez[1,19], Joshua Miller[2,19], José G. Pérez-Silva[1,19], Zi-Feng Jiang[3], Washington Tapia[4,5], Olaya Santiago-Fernández[1], Diana Campos-Iglesias[1], Lukas F. K. Kuderna[6,7], Maud Quinzin[2], Miguel G. Álvarez[1], Dido Carrero[1], Luciano B. Beheregaray[8], James P. Gibbs[9], Ylenia Chiari[10], Scott Glaberman[10], Claudio Ciofi[11], Miguel Araujo-Voces[1], Pablo Mayoral[1], Javier R. Arango[1], Isaac Tamargo-Gómez[1], David Roiz-Valle[1], María Pascual-Torner[1], Benjamin R. Evans[2], Danielle L. Edwards[12], Ryan C. Garrick[13], Michael A. Russello[14], Nikos Poulakakis[15,16], Stephen J. Gaughran[2], Danny O. Rueda[4], Gabriel Bretones[1], Tomàs Marquès-Bonet[6,7,17,18], Kevin P. White[3], Adalgisa Caccone[2*] and Carlos López-Otín[1*]

[1]Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias, CIBERONC, Universidad de Oviedo, Oviedo, Spain. [2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. [3]Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA. [4]Galapagos National Park Directorate, Galapagos Islands, Ecuador. [5]Galapagos Conservancy, Fairfax, VA, USA. [6]Institute of Evolutionary Biology (UPF-CSIC), Barcelona, Spain. [7]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology , Barcelona, Spain. [8]College of Science and Engineering, Flinders University, Adelaide, South Australia, Australia. [9]College of Environmental Science and Forestry, State University of New York, Syracuse, NY, USA. [10]Department of Biology, University of South Alabama, Mobile, AL, USA. [11]Department of Biology, University of Florence, Florence, Italy. [12]School of Natural Sciences, University of California, Merced, CA, USA. [13]Department of Biology, University of Mississippi, Oxford, MS, USA. [14]Department of Biology, The University of British Columbia, Kelowna, British Columbia, Canada. [15]Department of Biology, School of Sciences and Engineering, University of Crete, Heraklion, Greece. [16]Natural History Museum of Crete, Heraklion, Greece. [17]Catalan Institution of Research and Advanced Studies, Barcelona, Spain. [18]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. [19]These authors contributed equally: Víctor Quesada, Sandra Freitas-Rodríguez, Joshua Miller, José G. Pérez-Silva.
*e-mail: adalgisa.caccone@yale.edu; clo@uniovi.es

Supplementary Information

Giant tortoise genomes provide insights into longevity and age-related disease

Quesada et al.

Index

1. Genomic sequencing and annotation

1.1. Genome sequencing and assembly

DNA was recovered from a blood sample from Lonesome George, the last member of the species *Chelonoidis abingdonii*, one of the several giant tortoises living in the Galapagos Islands (**Fig. 1a, main text**). This DNA was sequenced using the Illumina HiSeq 2000 platform, from a 180 bp-insert paired-end library, a 5 kb-insert mate-pair library and a 20 kb-insert mate-pair library. The assembly of these libraries was performed with the AllPaths algorithm[2] to form a draft genome of 64,657 contigs with an N50 of 74 kb. Then, we scaffolded these contigs using Sspace[3] (v3.0) employing the long-insert mate-pair libraries. Finally, we filled the gaps using PBJelly[4] (v15.8.24) and the reads obtained from 18 PacBio cells. Since PacBio sequencing features frequent errors, we compiled the coordinates of these regions and added them to the header of the assembly. Point variants belonging to these regions are not reported unless validated by other means. This procedure generated 10,623 scaffolds with an N50 of 1.27 Mb (**Table S1**). The final assembly (CheloAbing 1.0) is 2.3 Gb long, a size consistent with other turtle assemblies such as *Chrysemys picta bellii* (2.59 Gb)[5], and *Chelonia mydas* and *Pelodiscus sinensis* (about 2.2 Gb)[6].

Table S1. *C. abingdonii* genome statistics.

	Scaffolds with GAPs	Scaffolds without GAPs	Contigs with Ns
Seqs	10,623	10,623	64,657
Min	886 bp	886 bp	130 bp
Median	6,499 bp	4,304 bp	16,213 bp
Mean	216,581 bp	204,233 bp	33,555 bp
Max	10,495,589 bp	10,223,643 bp	1,220,627 bp
Total	2,300,749,194 bp	2,169,570,871 bp	2,169,570,871 bp
n50	1,277,207 bp	1,227,724 bp	74,527 bp
n90	337,476 bp	331,274 bp	18,547 bp
n95	174,219 bp	172,536 bp	10,818 bp
Non-gapped Ns	23,455 bp		

Data provided by PBJelly[4], comparing the scaffolding, pre and post degapping, and the contigs assemblies.

Over this final assembly, we soft-masked repeated regions with *RepeatMasker*[7] using a database of chordate repeated elements (provided by the software) as reference. These repetitive elements represent 28.49% of the genome, including retroelements (17.86%) (**Table S2**).

Table S2. Repeated elements in *C. abingdonii* genome.

	RepeatMasker			*C. p. bellii*	
Repeat type	n	bp	%	bp	%
SINEs	215,691	33,047,802	1.4	44,277,662	1.87
PLEs	156,168	35,941,354	1.56	-	-
LINEs	572,190	231,609,979	10.07	248,848,977	10.52
LRTs	243,123	146,157,459	6.35	123,030,173	5.20
DNA Transposons	1,044,675	227,423,152	9.88	18,952,044	9.80
Unclassified	84,016	17,050,133	0.74	20,076,666	0.80

Small RNAs	42,858	8,412,400	0.37	8,527,612	0.36
Simple Repeats	8,199	1,475,310	0.06	11,395,517	0.48
Low complexity	192	64,291	~0.00	2,244,030	0.09

The type of repeat, occurrences (n), length (bp) and total percentage of the genome covered by them are indicated. The same values for *C. p. bellii*'s genome are shown for comparative purposes.

In parallel, we used the *Benchmarking Universal Single Copy Orthologs* (BUSCO v3.0.0)[8] to estimate the overall completeness of the genome. This tool tests the status of a set of 2,586 well-conserved vertebrate genes (from the comprehensive catalog of complete, fragmented or missing orthologs[9]). The results showed that a 92.4% of the genes were complete (0.6% of them with duplications), 5.7% were fragmented, and 1.9% were missing. This result is quite similar to those of other genomes we tested for comparison, such as humans or the Agassiz's desert tortoise, *Gopherus agassizii* (*G. agassizii*) (**Table S3**). The average characteristics of the predicted genes are also similar to those predicted in other genomes (**Table S16**).

Table S3. Comparative BUSCO analysis of *C. abingdonii* and *G. agassizii* tortoises.

	C. abingdonii		***G. agassizii***		***H. sapiens (CH38)***	
	Genes	*(%)*	*Genes*	*(%)*	*Genes*	*(%)*
Completed	2,391	92.4	2,387	92.4	2,355	91.1
Duplicated	16	0.6	22	0.9	72	2.8
Fragmented	147	5.7	138	5.3	112	4.3
Missing	48	1.9	61	2.3	119	4.6
Total	**2,586**	**100.0**	**3,023**	**100.0**	**2,586**	**100.0**

Data shown for *H. sapiens* are also calculated *de novo*, using vertebrata OrthDB v9.1 as database.

For the DNA sample from the Aldabra tortoise we used Illumina technology and a 180 bp paired-end library to obtain whole-genome data. We then aligned the resulting reads to CheloAbing 1.0 with BWA[10] (v0.7.5a). The resulting alignment mapped 97.8% of the 511,873,470 reads. We also calculated the callable proportion of the genome with the *A. gigantea* reads using GATKs CallableLoci walker. We found 1,885,758,801 callable bases, or 81.9% of the assembly size. We similarly aligned the genomic reads from *C. abingdonii* to CheloAbing 1.0.

With both alignments, we found 427,949 SNPs in *C. abingdonii* and 18,852,620 SNPs in *A. gigantea*. An analysis of coalescence using the Pairwise Sequentially Markovian Coalescent (PSMC) model[11] with these data showed that the effective population of *C. abingdonii* has been steadily declining for the past million years, with a slight uptick about 90,000 years ago (**Fig. 1b, main text**). Using a generation time of 25 years and a genome-wide mutation rate of 2.5×10^{-8}, we estimated that the effective population size was no more than a few thousand individuals for the past tens of thousands of years. By contrast, this analysis suggests that the effective population of *A. gigantea* has experienced greater variations in the last million years, though some have pointed out that similar patterns may represent historical population structure, leading to an overestimation of effective population size by the model[12]. It must be noted that the prediction for *C. abingdonii* loses statistical power at the million-year time frame, probably

due to complete coalescence. This suggests that overall diversity in *C. abingdonii* must have been very low throughout many generations. Consistent with this, the genome-wide median heterozygosity per bp is 0.00013 for *C. abingdonii* – one of the lowest values reported to date – and 0.00078 for *A. gigantea*.

Finally, we aligned RNA-Seq data from *C. abingdonii*'s blood and *A. gigantea*'s granuloma to the assembled genome using TopHat[13] (v2.0.14). Additionally, we used Illumina technology and a 180 bp paired-end library to obtain whole-genome data from *A. gigantea*. We then aligned the resulting reads to our *C. abingdonii*'s assembly with BWA[10] (v0.7.5a). Similarly, raw reads from *C. abingdonii* were aligned to CheloAbing 1.0 for manual curation purposes.

1.2. Automatic annotation of CheloAbing 1.0

We performed *de novo* annotation on the genome assembly of *C. abingdonii* using MAKER2, a multi-threaded, parallelized computational tool designed to produce accurate annotations for novel genomes based on a machine-learning approach[14]. We provided the algorithm with both the CheloAbing 1.0 assembly and the RNA-Seq data, as well as reference sequences from human and *P. sinensis*. It completed two runs in a Microsoft Azure virtual machine. As a result, we obtained 27,208 predicted genes, to which a putative function was assigned using MAKER2.

In a preliminary analysis of the general expansion of gene families, we performed several pairwise alignments of the predicted proteins from the automatic annotation to the Uniprot[15] databases of human and *P. sinensis* proteins using BLAST[16] (v2.6.011). Using in-house perl scripts (https://github.com/vqf/LG), we grouped these sequences into one-to-one, one-to-many and many-to-many orthologous relationships. Only alignments with a coverage of at least 80% of the longer protein and with more than 60% of identity were considered for the analysis. Finally, we searched for family expansions specifically present in *C. abingdonii*, by examining the aforementioned groups of orthologs. The results were manually curated. This way, we constructed extended orthology sets that may contain more than one sequence per species. These orthology sets capture most of the known protein families, although some of these families appear split according to sequence similarity. Almost all of these splits occur both in the human-to-*P. sinensis* and in the human-to-*C. abingdonii* comparisons. Since assembly errors may mimic gene losses, we decided to only test these sets for *C. abingdonii*-specific expansions. The interrogation of these sets suggests the existence of several high copy-number gene families in tortoises and turtles but absent in humans. Most of these families show homology to viral retrotranscriptases, consistent with the hypothesized expansion of CR1 retrotransposons in ancestral amniotes followed by dominance of L1 LINEs in therians[17]. After manual curation of the results, we found 12 examples of gene families displaying extra copies in *C. abingdonii* compared to turtles (**Table S4**). Each of those genes was also identified from the aligned reads in the Aldabra tortoise genome, and 10 of these amplifications were also identified in the recently sequenced genome of Agassiz's desert giant tortoise (*G. agassizii*)[18]. Interestingly, a functional annotation clustering analysis using DAVID[19] of these 12 genes found a significant enrichment in the "extracellular exosome" GO category (8 genes, p=0.044 after Benjamini

correction). Exosomes play an important role in cell-to-cell communication and have a strong impact on multiple biological processes and signalling pathways related to immunity, cancer and ageing[20-22]. Therefore, these data suggest that gene expansion of exosome-related genes may have provided a mechanism related to gigantism and cancer protection in giant tortoises.

Table S4. *Tortoise*-specific gene expansions.

Family	Human	*C. abingdonii*	*G. agassizii*	*P. sinensis*	Family	Human	*C. abingdonii*	*G. agassizii*	*P. sinensis*
OTX2	1	2	2	1	RPL11	1	2	2	1
ATP6V1A	1	2	2	1	EEF2	1	2	2	1
					GJD2	1	2	2	1
LAMTOR4	1	2	2	1	RPS25	1	2	1	1
STXBP1 STXBP2	2	4	6	2	POLR2L	1	2	2	1
TPT1	1	2	1	1	EEF1A1 EEF1A2	2	6	5	3
VCP	1	2	2	1					

Next, we searched for signatures of selection affecting the predicted set of genes. For this purpose, we used BLAST and in-house perl scripts to pairwise align all available protein sequences from human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis lupus familiaris*), gecko (*Gekko japonicus*), green anole lizard (*Anolis carolinensis*), python snake (*Python bivittatus*), common garter snake (*Thamnophis sirtalis*), Habu viper (*Trimeresurus mucrosquamatus*), budgerigar (*Melopsittacus undulatus*), zebra finch (*Taeniopygia guttata*), flycatcher (*Ficedula albicollis*), duck (*Anas platyrhynchos*), turkey (*Meleagris gallopavo*), chicken (*Gallus gallus*), Chinese softshell turtle (*P. sinensis*), green sea turtle (*C. mydas*) and painted turtle (*C. p. bellii*). We focused only on those genes with a one-to-one ortholog status in every species, and missing in no more than 3 species (excluding *C. abingdonii*), as described in previous studies[23]. This way, we obtained 1,592 groups of sequences. We then aligned each group separately with PRANK v.150803 using the codon model and analysed the alignments with *codeml* from the PAML package[24].

To search for genes with signatures of positive selection affecting *C. abingdonii* genes specifically, we executed two different branch models, M0, with a single ω0 value (where ω represents the ratio of non-synonymous to synonymous substitutions) for all the branches (nested), and M2a, with a foreground ω2 value exclusive for *C. abingdonii* and a background ω1 value for all the other branches. Genes with a high ω2 value (>1) and a low ω1 value (ω1<0.2 and ω1≈ω0) in *C. abingdonii*, but not in *P. sinensis* (**Table S5** and **Table S17**) were then considered to be under positive selection. As a control, M2a was repeated using *P. sinensis* as the foreground branch, and no overlapping genes were found in the result. After this, we used the M8 branch model to assess the individual importance of every site in these positively selected genes, obtaining a list of sites possibly under selection. The equivalent sites were examined in the Aldabra tortoise through alignments, to evaluate which of these important residues were altered (**Table S18**).

Table S5. Omega values of *C. abingdonii* genes under putative positive selection (ω>1).

GENE	ω2 (*C. abingdonii*)	ω1 (Background)	ω0 (Model 0)	ω2-ω1
TUBE1	153.06935	0.07402	0.04658	152.99533
FRRS1L	30.98228	0.01332	0.01626	30.96896
TUBG1	7.98622	0.00904	0.00918	7.97718
CLRN2	3.88988	0.15350	0.15807	3.73638
MVK	3.77348	0.17305	0.17735	3.60043
CMAS	3.10102	0.13879	0.15952	2.96223
CREBL2	3.09570	0.08965	0.04218	3.00605
MTFMT	2.60690	0.21941	NA	2.38749
IRAK1BP1	2.35841	0.17743	0.13865	2.18098
ALG1	2.12288	0.16317	0.17949	1.95971
COX18	1.71757	0.19049	0.21782	1.52708
NTF3	1.61603	0.08596	0.08035	1.53007
IL1R2	1.60957	0.30844	0.30517	1.30113
AHSG	1.54836	0.43872	0.36735	1.10964
PTCD2	1.42073	0.37172	0.37426	1.04901
FGF19	1.40904	0.14895	0.16238	1.26009
UBE2J1	1.39297	0.13930	0.01943	1.25367
F2R	1.33392	0.21719	0.167	1.11673
EMC2	1.27666	0.03209	0.05771	1.24457
MED19	1.26411	0.01153	0.00912	1.25258
VPS35	1.26021	0.00873	0.00869	1.25148
ATP4B	1.24188	0.17286	0.18168	1.06902
CRLS1	1.23942	0.06322	0.52669	1.1762
ITM2B	1.23539	0.07216	0.07512	1.16323
HRH1	1.20455	0.23600	0.16651	0.96855
PCP4	1.18271	0.01507	0.01507	1.16764
PIM3	1.17259	0.02294	0.02177	1.14965
GTPBP10	1.17153	0.18357	0.23034	0.98796
BAG2	1.16059	0.07724	0.07518	1.08335
GDPD1	1.13870	0.07981	0.08303	1.05889
RPS8	1.10896	0.01185	NA	1.09711
WDR27	1.07450	0.26707	0.17619	0.80743
C16orf87	1.07403	0.05566	NA	1.01837
COTL1	1.06762	0.02572	0.02609	1.0419
UQCRB	1.06365	0.13591	0.12339	0.92774
TDO2	1.06041	0.08567	0.04462	0.97474
MALSU1	1.05997	0.16755	0.18792	0.89242
SGPP2	1.05829	0.17148	0.16672	0.88681
CHIC2	1.05334	0.05987	0.00887	0.99347
DYNLRB1	1.04162	0.03182	0.03816	1.0098
TMEM38A	1.02214	0.08349	0.05464	0.93865
KCNK18	1.01755	0.24701	0.21284	0.77054
KDSR	1.00704	0.08597	0.05731	0.92107

The ω value for *C. abingdonii* (ω2) and the background (ω1), both of them under selection model 2, and the ω value under selection model 1 for all the aligned species (ω0) are indicated. Differences between *C. abingdonii*-specific and the general values (ω2-ω1) are also shown. The complete Table resulting from this analysis is presented as **Table S17**.

The results of this analysis point to multiple biochemical pathways which may have been affected by selection, although automatic functional annotation of this set did not find significant overrepresentation of any GO term (data not shown). Two of the top three genes with the strongest evidence for positive selection were tubulins (*TUBE1* and *TBG1*), suggesting that changes in cytoskeletal dynamics have been important in the evolution of giant tortoises. Consistent with the important role of this pathway in the biology of the cell, the alterations

found in the selected residues (p.E169D and p.I186V in TUBE1) are conservative and probably do not affect the main role of this protein in microtubule formation. However, the effects of regulating tubulin assembly on cell cycle progression[25] suggest that these alterations might be related to the increase in the number of cellular divisions associated to gigantism. In addition, two genes with evidence for positive selection, *BAG2* (*NEF*) and *UBE2J1* (Ubc6/7) are involved in endoplasmic-reticulum-associated protein degradation (ERAD) (**Fig. S1**). Notably, one of the genes duplicated in giant tortoises (*VCP*, also known as p97) also plays a central role in this pathway. ERAD is important in the unfolded-protein response and consequently in the correct proteostasis of the cell, whose deregulation constitutes a hallmark of ageing[26,27]. In this regard, the list of positively selected genes also features *TDO2*, whose product is involved in the regulation of tryptophan-mediated proteostasis. Interestingly, the inhibition of TDO2 has been shown to protect against age-related neurodegeneration[28]. Therefore, the selective process acting on these proteins might reflect the constraints imposed on proteostasis due to increasing lifespan.

In addition, two positively selected genes, *AHSG* and *FGF19*, are listed in a panel of four proteins whose expression levels correlate with successful ageing in humans[29]. Notably, one of the selected alterations in FGF19 (p.S116A) is expected to affect the receptor-interaction site of this protein. These factors intervene in the regulation of glucose and lipid metabolism[30,31], another hallmark of ageing, suggesting that the adaptation to the challenges that longevity poses on this system may have been important in the evolution of giant tortoises. Finally, three genes with evidence for positive selection in these organisms (*MVK*, *IRAK1BP1* and *IL1R2*) play important roles in the modulation of the immune system, which in turn participates in the phenotypes of altered intercellular communication associated with ageing[26]. In this regard, it is important to notice that, in addition to its role in proteostasis, ERAD is a target of viral infection, as multiple viruses depend on this process for successful delivery to the cytoplasm[32]. Taken together, this hypothesis-free analysis highlights proteostasis, metabolism regulation and immune response as key processes during the evolution of giant tortoises, and provide starting points for future work on this subject.

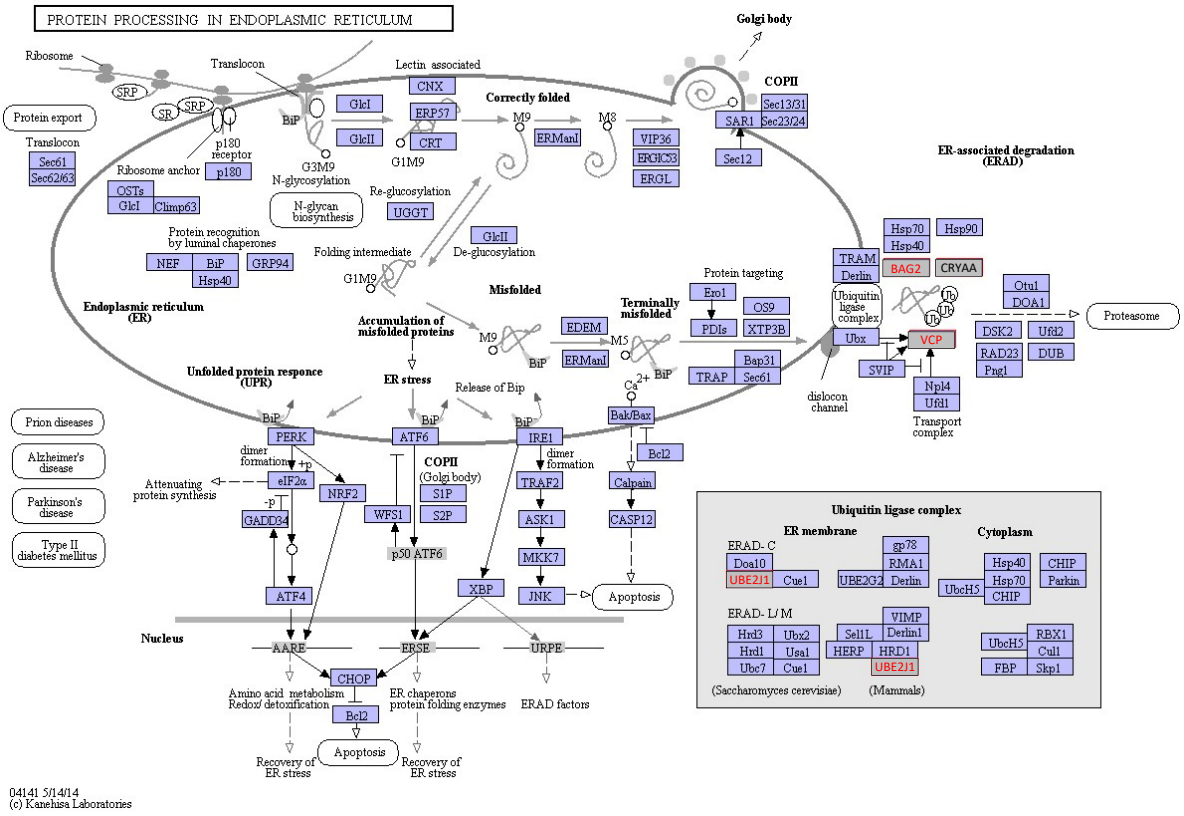


Figure S1. Positively selected and amplified genes in the genome of *C. abingdonii* implicated in Endoplasmic Reticulum Associated Degradation (ERAD). Highlighted genes are shown on top of the protein processing pathway in endoplasmic reticulum (KEGG ko04141)[33-35].

We also performed an analysis of microRNA (miRNA) sequences in *C. abingdonii*'s genome with BLASTN using as databases the high-confidence miRNA set available at miRBase v21 (http://www.mirbase.org/) and the list of miRNAs from *P. sinensis*. Due to limitations in the BLASTN algorithm, we used the longer consensus precursor sequence instead of the seed sequence. The results were manually analysed to assess their reliability. We identified a total of 316 miRNAs, 123 of those being present in both lists (**Table S19**).

1.3. Manual annotation of CheloAbing 1.0

In addition to the automatic annotation process, we manually annotated about 3,000 genes of general interest in the fields of ageing, cancer, immunology and DNA repair, and involved in physiological functions such as reproduction, development, shell formation and sensorial perception (**Fig. S2**). These genes were specifically selected *a priori* based on our experience in these fields and after a revision of the available publications in each subject.

We used a custom human protein database from Ensembl (http://www.ensembl.org/index.html) as a reference to predict each ortholog in the genome of *C. abingdonii* by using the BATI algorithm (Blast, Annotate, Tune, Iterate)[36]. Accordingly, protein residues in all alignments have been numbered using the human transcript as reference. Those cases with a low identity level between human and *C. abingdonii* were treated separately using sequences from closer relatives such as the chicken (*G. gallus*), the green anole lizard (*A. carolinensis*) and other turtles like *P. sinensis*, the western painted turtle (*C. p. bellii*) or the green turtle (*C. mydas*) as references. Once the selected genes were properly annotated, we performed several comparisons between the gene sequences of *C. abingdonii*, human (*H. sapiens*), mouse (*M. musculus*), Sauria (*A. carolinensis*), chicken (*G. gallus*), and the other turtles (*P. sinensis*, *C. p. bellii* and *C. mydas*). We also listed changes considered interesting in aligned genomic reads from *A. gigantea* as a representative of another giant tortoise species that is distantly related to Galapagos giant tortoises (**Table S6**). We first focused on collecting the putatively most relevant truncations, point mutations (according to functional information in the human ortholog), and copy-number variations. The last column of the table summarizes the results of our screen of other Galapagos tortoise species, the outgroups for both the Galapagos and Aldabra tortoises, and the other turtle species for which we have data publicly available (*G. agassizii*, *P. sinensis* and *C. p. bellii*) species, and the continental relative for giant tortoise species (continental outgroups).

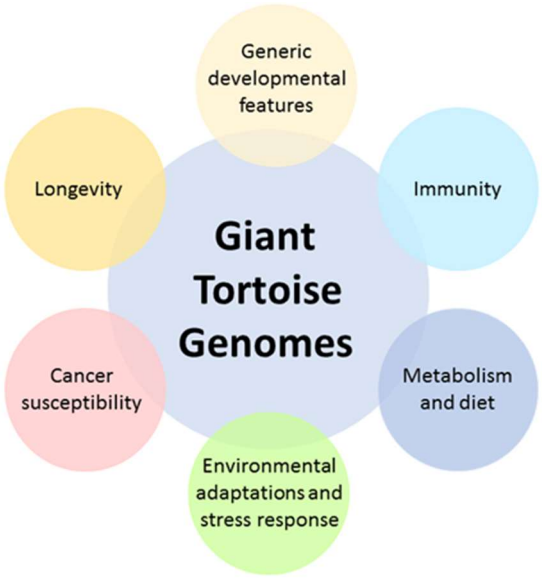


Figure S2. Features explored in the genomic analysis of giant tortoises.

Table S6. Highlighted variants and truncations found during manual analysis of *C. abingdonii*'s genome.

Gene	Variant	Affected site/Predicted effect	Present in (besides *C. abingdonii*)
ADORA2B	C72D; A150G; T153P; N163Q	Transmembrane receptor 7	G.aga
AGTR1	C355T	Lipidation residue	G.aga; A.gig
AICDA	D96N; L104M	RNF126 interaction region	G.aga (L104M)
ALDH2	Q366M	NAD binding site	Galapagos tortoises
	M487T	Homotetrameric interface	Galapagos tortoises; A.gig; Continental outgroups
ALKBH3	E248D	Region of alpha-ketoglutarate binding	
ANKRD17	Multiple	Whole protein	G.aga
ANXA1	N153C; E158D; T169V; C324F	Annexin domain	G.aga (N153C; E158D; T169V; C324F)
APEX1	K31A	TOMM20 binding	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
APLF	S116N	Diminishes phosphorylation by ATM	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
ARRB2	L244S	TRAF6 interaction domain	

ATM	R250Q	Variants related to ataxia-telangiectasia	Galapagos tortoises; A.gig; Continental outgroups
	Y316H	Variants related to ataxia-telangiectasia	Galapagos tortoises
	V2873I	Catalytic loop	Galapagos tortoises; Fishes
BCL2L1	D61A	No cleavage by caspase-1 nor caspase-3	G.gal; P.sin; C.pic; G.aga; A.gig
BIN2	N49D	Dimer interface	C.pic; G.aga; A.gig
BRCA1	S308N	Phosphoserine	P.sin; G.aga; A.gig
BRCA2	S491V; S755V; S2095Q; T3387P	Phosphoserine/Phosphothreonine	A.car; P.sin; C.pic; G.aga; A.gig
C1QA	K100R	Hydroxylysine	
C1QBP	D177N	C1Q globular heads binding domain	
C6	N167K	Calcium binding site	A.gig
	L673M	Receptor-ligand interactions	P.sin; C.pic; G.aga; A.gig
C7	N600R	Receptor-ligand interactions	G.gal; P.sin; C.pic; G.aga; A.gig
C8A	N437S	Loss of glycosylation site	C.pic; G.aga
C8B	W554S	Loss of glycosylation site	C.pic
CAMP4	Q34Lfs*9	Frameshift leading to premature stop codon	A.gig
CANT1	D169N	Activity reduced by 96%	A.car (D169E); G.gal (D169E); P.sin (D169E); C.pic (D169E); G.aga (D169E); A.gig (D169E)
CARD11	S466G	Phosphoserine	C.pic; G.aga; A.gig
CARD14	Q711*	Premature stop codon	
CAV1	K5Q	Acetyllysine	G.aga, A.gig
CDH1	Y754H	CBLL1 interaction site	G.gal (Y754N); C.pic; A.gig
CDK6	S222N	Active site and substrate binding	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
CDKN1B	Y89F	Reduced CDK4 binding and increases inhibition of CDK4	A.car; P.sin; C.pic; G.aga; A.gig
CHGA	S300N	Phosphoserine	A.car; G.aga
CHUK	D154E; I159T; I160V; H306Y	Serine-threonine kinase domain	M.mus (I159T); P.sin (D154E); C.pic (H306Y); G.aga (H306Y)
CLIP4	Q438*	Premature stop codon	A.gig
CLNK	L333M; E353R	Hydrophobic pocket	
CNDP1	A274Dfs*1; V295Sfs*7	Frameshifts leading to premature stop codons	Galapagos tortoises; A.gig; Continental outgroups
CRYAA	Q50H	Deamidated glutamine	P.sin; C.pic; G.aga; A.gig
DCLRE1B	R498C	TERF2 interaction site	G.aga; Galapagos tortoises; A.gig; Continental outgroups
DDX58	S532del; K814R; G900S	S532del Helicase domain; K814R G900S RIG domain	G.aga (S532del; G900S)
DHX58	D83N	Helicase domain	
DLL1	G417E; V444M	Calcium binding EGF-like domain	C.pic; G.aga (V444M)
DNAJA3	R293H	CARM1 methylation site of activation. Zn binding motif	
DRD2	L170I; R217Q; L236M; V255I	G-coupled receptors domain	A.car (L236M); G.gal (L170I; L236M); P.sin (L236M); C.pic (L236M); G.aga (L170I; V255I)
EBI3	E213N	Cytokine binding motif	G.aga
EIF2AK3	S1022N	Binding site	C.pic; G.aga; A.gig
EPHB6	A321T	GCC domain	
	R677L	Fibronectin type 3 domain	
	L959F	SAM domain tyrosine kinase receptor	
EPRS	V1302I; R1415K; T1429N	Prolyl-anticodon binding domain	G.aga (R1415K; T1429N)
	K844R	tRNA binding WHEP-TRS domain	G.aga
ERN1	K837R	Ligand binding site	P.sin; C.pic; A.gig
EXO1	S610G	Phosphoserine in MSH2 interaction domain	G.aga, A.gig
F10	E91K	Involved in Factor 10 deficiency	G.aga (E91D); A.gig

F11	E315K	Involved in Factor 11 deficiency	G.aga; A.gig
F7	S113D; F64L	Involved in Factor 7 deficiency	G.aga; A.gig
FADD	E37A; V39I	DED-FADD domain	
FAS	D93H; G112D; N124D	TNFRSF domain	C.pic (G112D; N124D); G.aga (D93H; G112D)
	A301F	DD death domain	G.aga
FGA	R123I	Cleavage site by plasmin	
FGF19	S116A	Receptor-interaction site	A.gig
FOXJ1L	F158Y	DNA binding domain	G.aga
GSK3A	R272Q	Activation loop	C.pic; G.aga; Galapagos tortoises; A.gig; Continental outgroups
HMGB1	S107A	DNA binding site	C.pic; G.aga; A.gig
HSPA5	T643E	Phosphothreonine	C.pic; G.aga; A.gig
HSPD1	K125R	Lys acetylation site	P.sin; C.pic; G.aga; A.gig
	K369R	Lys acetylation site	C.pic; G.aga; A.gig
IGF1R	N724D	IGF1/2 binding site	G.aga; Galapagos tortoises; A.gig; Continental outgroups
IGF2R	N1757_Y1758del	N1757 is a glycosylation site	G.aga, Galapagos tortoises; A.gig; Continental outgroups
IRAK4	V37I	IRAK4-IRAK2 interaction site	
ITGA1	R990*	Premature stop codon	
JAK2	L579V	ATP binding site	G.aga; A.gig
LY75	Q471R; K785R	Ligand binding surface	P.sin; C.pic; G.aga; A.gig
MAP3K8	E279D	Activation loop	P.sin; C.pic; G.aga; A.gig
MAP4K2	G159A; R169K	Activation loop	P.sin; C.pic; G.aga; A.gig
MARCH8	S253A	Phosphoserine	A.gig
MB	K97T	Heme interaction site	
MDM2L	E430*; P451*; N478*	Premature stop codons	G.aga; A.gig
	S235*	Premature stop codon	
MEP1A	W317Lfs*3; N529Tfs*6	Frameshifts leading to premature stop codons	Galapagos tortoises
MIF	N111C	Locked trimer	G.aga; Galapagos tortoises; A.gig; Continental outgroups
MSH2	V102I	Variant related to HNPCC1	C.pic; G.aga; A.gig
MSH6	S137P	Phosphoserine	A.gig
MYO5AB	E114K	ATP binding site	P.sin; C.pic; G.aga; A.gig
NEIL2	K154R	Acetylation site	A.gig
NF2L	S13G	Phosphoserine	P.sin; C.pic; G.aga; A.gig
	D268E	Peptide binding site	P.sin; C.pic; G.aga; A.gig
	K289Q; K291Q	PIP2 binding site	P.sin; C.pic; G.aga; A.gig
NFKBID	Y66Q; A94T; P98A; L154Q; Q194H; A217D; A218G	Interaction sites	P.sin; C.pic; G.aga; A.gig
NFKBIZ	L553F; L631F	Interaction sites	C.pic; G.aga; A.gig
NLN	W50*	Premature stop codon	G.aga (R45*, evolutive convergence); Galapagos tortoises
	L532Ffs*1	Frameshift leading to premature stop codons	Galapagos tortoises; Continental outgroups
NOTCH1	A2022G	Hydroxylated by HIFAN	A.gig
NTRK1	L333del	Receptor binding site	G.gal; P.sin; C.pic; G.aga; A.gig
NUTM1	A672*	Premature stop codon	
PDE4DIP	H1598R	Associated with risk of ischemia	C.pic; G.aga; A.gig
PIK3R1	I287V	Putative GTPase interaction site (RAS)	C.pic; G.aga; A.gig
PKN3	L623V; T639M; F641Y; C757S	Active sites	G.gal; P.sin; C.pic; G.aga
	Multiple	Rho binding domain	G.gal; P.sin; C.pic; G.aga

PLAG1	P346S	Repression domain	
PLG	R134L	Fibrin-binding site	G.aga; A.gig
	R136K	Fibrin-binding site	P.sin; G.aga; A.gig
PPP2R1A	T78A	Polypeptide binding site	C.pic; G.aga; A.gig
PRDM1	G74S	Irritable bowel syndrome-diarrhea	C.pic; G.aga; A.gig
PRDM1L	D69E	Autoimmune diseases	A.car; G.aga; A.gig
PRF1G	Q386*	Premature stop codon	Galapagos tortoises
PRF1I	W483*	Premature stop codon	C.dun
PRF1K	S388*; H398*; T450*; T452*; T453*; C511*	Premature stop codons	
PRG4	A1246V	Metal binding site (ion binding site)	P.sin; C.pic; G.aga; A.gig
PRKACA	V105I	Active site	G.aga; A.gig
PRKCH	V374I	Increases cerebral infarction risk in human	M.mus; C.pic; G.aga
PRKCZ	A155G; R160K	Putative DAG/PE binding sites	G.gal (R160K); P.sin (A155G; R160K); C.pic (A155G; R160K); G.aga
PRKDC	S2612T	Phosphoserine in catalytic subunit	C.pic; G.aga; A.gig
PRSS12	R691*	Premature stop codon	
PSEN1	R352E	Pathological significance	G.aga; Galapagos tortoises; A.gig; Continental outgroups
PSIP1	A153S	Nuclear localization signal	P.sin; C.pic; G.aga; A.gig
PSMD10	E54D; A72S; K116R	Oligomer interface	A.car (E54D); C.pic (K116R)
PTGS1	F209L	Heme-binding site	A.gig
	L534F	Substrate binding site	P.sin; C.pic; G.aga; A.gig
PTGS2	I331V	Substrate binding site	C.pic; G.aga; A.gig
PTK2B	R309K	Putative peptide binding site	A.car; G.gal; P.sin; C.pic; G.aga
	D578E	Activation loop	A.car; G.gal; P.sin; C.pic; G.aga
PTPRC	V1125S	Active site	G.aga
PTX3	N275S	Intermolecular salt bridges	M.mus; G.gal; P.sin; C.pic; G.aga
PVRL1	V89I; R110Q; F126Y; M143L	Ligand binding site	C.pic (V89I; R110Q; M143L); G.aga (V89I; R110Q; M143L)
RAB17	L48M; K58Q; V59L	GEF interaction sites	A.car (K58Q); G.gal (K58Q); P.sin (K58Q; L48M; V59L); C.pic (K58Q; V59L); G.aga (K58Q; V59L)
	G28T; S29A; K45R	GTP/Mg^{2+} binding sites	A.car (K45R); G.gal (G28T); P.sin (G28T; S29A; K45R); C.pic (G28T; S29A; K45R); G.aga (G28T; S29A; K45R)
	L35I; L37F; D43V; T101A; V125L	Rab subfamily motif	P.sin (L37F); C.pic (L35I; L37F; T101A; V125L); G.aga (T101A; V125L); A.gig (D43V)
RAD51B	S194T	ATP binding site	P.sin; C.pic; G.aga; A.gig
RAD51D	S197L	ATP binding site	P.sin; C.pic; G.aga; A.gig
RB1	K720R	Binding site 2 (LxCxE peptide; DNA; CDK2)	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
REL	I230S	Ankyrin protein binding site	G.gal; P.sin; C.pic; G.aga
RELA	R253K	Ankyrin protein binding site	A.car; P.sin; C.pic; G.aga
RELB	R231K; I308M	Ankyrin protein binding site; DNA binding site	A.car; G.gal; P.sin; C.pic; G.aga
RGCC	T111A	Loss phosphorylation site. Less CDK1 stimulation	M.mus; P.sin; C.pic; G.aga
RGS1	N97V; D194E; I195A	Homodimer interface	P.sin (I195A); C.pic; G.aga; A.gig (N97V; D194E; I195A)
RHOH	N62S	GAP/GEF interaction site	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
	L71I	GTP/Mg^{2+} binding site GAP/GEF/GDIa/Effector interaction site	G.gal; P.sin; C.pic; G.aga; A.gig

RIPK1	S25A; I43V	Active sites	A.car (S25A); G.gal (S25A; I43V); P.sin (S25A; I43V); C.pic (S25A; I43V); G.aga (S25A; I43V); A.gig (S25A; I43V)
RNF168	R57W	May loose ability to ubiquitinate H2A	C.pic; G.aga; A.gig
RNASE1	C68*	Premature stop codon	
ROS1	P182S	Cytokine receptor motif (FN type 3 domain)	P.sin; C.pic; G.aga; A.gig
	E274P	Cytokine receptor motif (FN type 3 domain)	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
	S641Y	Interdomain contacts (FN type 3 domain)	P.sin (S641F); C.pic; G.aga; A.gig
SDHCL	R50H	Iron-sulfur protein interface	G.aga; A.gig
	R72H	Iron-sulfur protein interface and proximal heme binding site	G.aga (R72G); A.gig
	L78T	Proximal heme binding site	G.aga (L78M in SDHC and SDHCL); A.gig (L78M); C.abi (L78M in SDHC)
SERPING1	E472D	Reactive centre loop	C.pic; G.aga; A.gig
SETBP1	TQ1462-1463FS	DNA binding	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
SIRT6	T294L; S330A	Phosphothreonine; Phosphoserine	P.sin; C.pic; G.aga; A.gig
SNAP23	M169L; Q183R	Heterotetramer interface	C.pic; G.aga; A.gig
SOX21	I171Pfs*64	Premature stop codon	A.gig
TAS1R1	T436*	Premature stop codon	G.aga (W297*); A.gig
TERF2	S161T	DCLRE1B interaction site	A.car; G.gal; P.sin; C.pic; G.aga (S161E); A.gig
	M164T		P.sin (M164K); C.pic (M164K); G.aga (M164I)
	N176G		A.car (N176H); G.gal; P.sin; C.pic; G.aga (N176V); A.gig
	M180L		A.car (M180K); G.gal; P.sin; C.pic; G.aga (M180I); A.gig
TICAM2	S16N; C117S	Phosphorylation site; dimerization site	G.gal (S16N); P.sin (S16N; C117S); C.pic (S16N; C117S); G.aga (S16N; C117S)
TIRAP	S180K	S180L protects against pneumococcal disease, malaria, bacteraemia and tuberculosis	G.gal; P.sin; C.pic; G.aga
TLR10	D508*; S512*; E530*; D540*; V542*; L567*; M572*; I602*; L610*; R620*	Premature stop codons	A.gig
TMEM173	V147I	(V147L) vascular disease in human	A.car; P.sin; C.pic; G.aga
TP53	R72P	Human polymorphism	A.car; G.gal; C.pic; G.aga; A.gig
	S106E	Adaptation to hypoxia	P.sin; C.pic; G.aga; Galapagos tortoises; A.gig; Continental outgroups; M.oec; B.bar; L.for
	R290C	DNA-binding domain. Li-Fraumeni mutation	G.aga; A.gig
TP53BP1	T543I	Thr phosphorylation site	G.aga; A.gig
TRAIP	R18K	TRAF interacting protein	G.aga; A.gig
TRIM24	S862R	Histone H3 binding site	P.sin; C.pic; G.aga; A.gig
TSHR	K621R	K621 is important for G alpha subunit activation	Galapagos tortoises; A.gig; Continental outgroups
WHSC1L1	K245L	Sumoylation site	A.car; G.gal; P.sin; C.pic; G.aga; A.gig
XPNPEP1	D16*	Premature stop codon	Galapagos tortoises; A.gig; Continental outgroups
XRCC5	K568N	Sumoylation site	P.sin; C.pic; G.aga; Galapagos tortoises; A.gig; Continental outgroups
XRCC6	K556R	Sumoylation site	Galapagos tortoises; A.gig; Continental outgroups

	N275S; R403C	DNA binding site	A.gig
ZBTB16	A469T	DNA binding site	A.gig

The variants are referenced using the Human Genome Variation Society (HGVS) nomenclature[37]. Frameshifts indicate the first stop codon of the new reading frame (if any) with a number after the asterisk. A.car, *A carolinensis*; A.gig, *A. gigantea*; B.bar, *Barbus barbus*; C.dun, *C. duncanensis*; C.pic, *C. p. bellii*; G.aga, *G. agassizii*; G.gal, *G. gallus*; M.mus, *M. musculus*; M.oec, *M. oeconomus;* L.for, *Loligo forbesi*; P.sin, *P. sinensis.*

To validate the most relevant findings, we used DNA samples of different tortoises from the Galapagos Islands, including the islands of Santa Cruz (*Chelonoidis porteri*), Española (*Chelonoidis hoodensis*), Isabela (*Chelonoidis becki* and *Chelonoidis vicina*), Pinzón (*Chelonoidis duncanensis*) and Pinta (*Chelonoidis abingdonii*). The DNA sample from the island of Pinta corresponded to Lonesome George, as it was the only available individual from the *C. abingdonii* species. We also used DNA samples of tortoises from the Aldabra atoll (*Aldabrachelys gigantea*) as well as from two continental outgroups: *Chelonoidis chilensis, Chelonoidis carbonaria* and *Chelonoidis denticulata* for Galapagos Islands tortoises, and *Stigmochelys pardalis* for Aldabra tortoise (**Fig. S3**).

To validate the existence of specific SNPs we carried out PCR experiments with locus-specific primer pairs. When studying specific variants, primers were designed targeting the region where they are located. To study copy-number variations, we performed PCR reactions with primer pairs which allow the amplification of each copy separately. When this method was not feasible, we designed specific pairs to amplify a target region with different nucleotide variants in each copy, and then examined the resulting electropherogram for evidence of both copies. A list of the primers used and other additional information about every PCR reaction is summarized in **Table S20**. In all cases, PCR reactions were carried out with 10 ng of DNA under the same specific conditions, where melting temperature (Tm) depends on the pairs used. For each PCR reaction these conditions were: 95 ^{o}C for 3 min; 35 cycles (95 ^{o}C for 60 s; Tm for 30 s; 72 ^{o}C for 30 s); 72 ^{o}C for 7 min; 20 ^{o}C for 5 min. Then, we tested the success of the PCR reactions by electrophoresis in a 2% agarose gel with the resulting products. Finally, the products were sequenced using the Sanger method with 1 µl of one of the PCR primers at 10 µM and 50 ng of amplified DNA in a total volume of 10 µl. When feasible, the results of the manual annotation were also confirmed via RNA-Seq data from *C. abingdonii* whole blood and from an *A. gigantea* granuloma (**Table S7**).

When orthology relationships were dubious, phylogenetic analyses were performed. In the cases of myosins and RAS proteins, the alignment of the sequences was performed separately for each gene with MAFFT v.7 with default parameters and auto-strategy[38]. The Bayesian Information Criteria (BIC) implemented in ProtTest v.3.4.2 was used for the selection of the best-fit model of amino acid replacement[39].

Bayesian Inference (BI) and Maximum Likelihood (ML) analyses were performed with MrBayes[40] and RAxML 8.1.21[41] in RAxMLGUI v.1.5.2, respectively. In BI, eight concurrent chains (one cold and seven heated) for $5x10^7$ generations were run and recorded samples every 5,000 generations. The first 25% of the samples were discarded as burn-in, and the remaining samples were used to summarize the posterior probability distributions for parameters (≥95% indicate

significant support)[42] . Results were analysed in Tracer v.1.6 to assess convergence and effective sample sizes (ESS) for all parameters. The best ML tree was selected from 500 iterations and the confidence of the branches of the best ML tree was assessed based on 1,000 thorough bootstrap replicates using the PROTGAMMA model. In both analyses, *Drosophila* amino acid sequences were used as outgroup. Maximum likelihood bootstrap values (bs) and Bayesian posterior probability (pp) values were joined and mapped to the Bayesian tree (i.e., the 50% majority-rule consensus tree calculated from the posterior distribution of trees).

Table S7. Gene variants validated by Sanger sequencing and supported by aligned RNA-Seq reads.

		Validations	
Gene	Observation	Sanger sequencing	RNA-Seq
ALDH2	Variant p.Q366M	Galapagos tortoises	Confirmed
	Variant p.M487T	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
APOBEC1	7 copies	C.abi; A.gig	All copies confirmed in C.abi and A.gig
ATM	Variant p.R250Q	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
	Variant p.Y316H	Galapagos tortoises	Confirmed
	Variant p.V2873I	Galapagos tortoises	Confirmed
CDH1	Variant p.Y754H	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
CHIA	Triplication	Galapagos tortoises; A.gig; Continental outgroups. One of these copies is specific of Galapagos tortoises	No reads aligned
CMA1L	Duplication + pseud.	Galapagos tortoises; A.gig; Continental outgroups. Only one copy validated by Sanger sequencing	No reads aligned
CNDP1	A274Dfs*1	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
	V295Sfs*7	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
CTSGL	8 copies + 1 pseud.	Galapagos tortoises; A.gig; Continental outgroups. Two copies were not validated by Sanger sequencing	No reads aligned
DCLRE1B	Variant p.R498C	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
EEF1A1	Duplication + 1 pseud.	Galapagos tortoises; A.gig; Continental outgroups. The pseudogene was not confirmed	Non-specific reads
GAPDH	Duplication + retrogene	Galapagos tortoises; A.gig; Continental outgroups. The retrogene was not confirmed	One copy confirmed
GSK3A	Variant p.R272Q	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
GZMB	5 copies + 1 pseud.	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
GZMH	Duplication + 3 pseud.	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
IGF1R	Variant p.N724D	Galapagos tortoises; A.gig; Continental outgroups	Reads only in A.gig granuloma
IGF2R	N1757_Y1758del	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
LRAS	New gene	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
MEP1A	W317Lfs*3	Galapagos tortoises	No reads aligned
	N529Tfs*6	Galapagos tortoises	No reads aligned
MIF	Variant p.N111C	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
MYCN	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Only one copy with reads in A.gig granuloma
NEIL1	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
NF2	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
NLN	Stop codon W50*	Galapagos tortoises	Confirmed

	L532Ffs*1	Galapagos tortoises; Continental outgroup	No reads aligned
P2RY8	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Only one copy confirmed in C.abi. Both copies confirmed in A.gig
PML	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
PRF1	9 copies + 3 pseud.	C.abi; 1 pseud. present in Galapagos tortoises but absent in A. gig and Continental outgroups; another pseud. is unique to C.abi; last pseudogene is not present in A.gig but it is not validated in other turtles	Few reads
PRSS12	Stop codon R691*	C.abi	No reads aligned
PSEN1	Variant p.R352E	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
RMI2	Duplication	C.abi; A.gig	Only one copy confirmed in C.abi. Both copies confirmed in A.gig
SMAD4	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
TLR13	Duplication	C.abi; A.gig	No reads aligned
TLR2	Duplication + 2 pseud.	C.abi; A.gig	No reads aligned
TLR5	Duplication	C.abi; A.gig	No reads aligned
TLR8	Duplication	C.abi; A.gig	No reads aligned
TP53	Variant p.S106E	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
TSHR	Variant p.K621R	Galapagos tortoises; A.gig; Continental outgroups	No reads aligned
USP12	Duplication	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
XPNPEP1	Stop codon D16*	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
XRCC5	Variant p.K568N	Galapagos tortoises; A.gig; Continental outgroups	Confirmed
XRCC6	Variant p.K556R	Galapagos tortoises; A.gig; Continental outgroups	Confirmed

A.gig, *A. gigantea*; C.abi, *C. abingdonii*. *Confirmed* by RNA-Seq means that reads were aligned unequivocally at the expected location. Lack of expression of a gene in the RNA-Seq samples may lead to *No reads aligned*.

1.4. Species names and groups

To clarify the usage of the different taxonomic terms, “giant tortoises” refers to *C. abingdonii*, other tortoises from Galapagos Islands (described below in section 1.5.1.) and *A. gigantea*. The term “tortoises” also includes *G. agassizii*. Sauropsida is a reference to turtles, birds and lizards, while “reptiles” excludes the birds. However, Archelosauria refers to birds and turtles only. Additionally, when using the terms Testudines, Chelonii or “turtles” we are referring to all turtles, but when referring to Testudinoidea, we are excluding the soft-shell turtle. Divergence times between taxa where obtained using *TimeTree*[43]. This analysis places the split between the two clades leading to the Galapagos and the Aldabra giant tortoises at about 40 million years ago, and the split between the ancestors of these tortoises and the ancestors of humans at about 312 million years ago (**Fig. S3**).

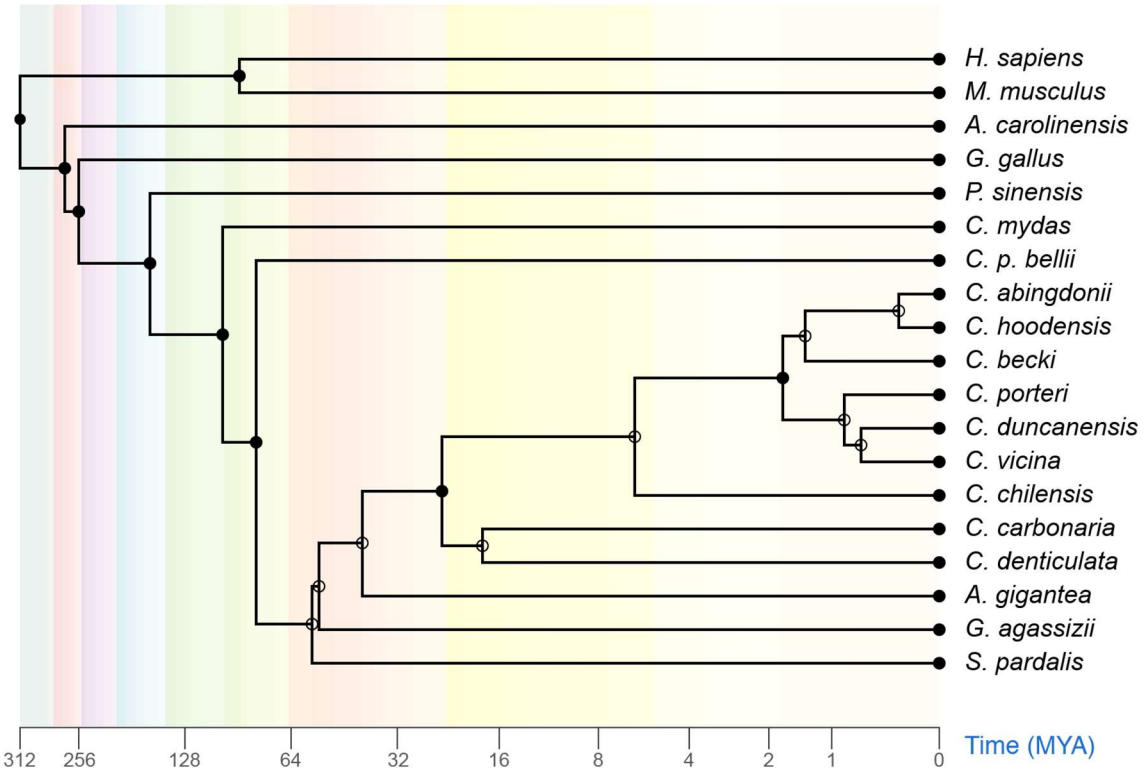


Figure S3. Time-tree showing evolutionary divergence between some of the studied species. The divergence between *C. abingdonii* and humans occurred 312 million years ago (MYA), while the divergence between *C. abingdonii* and *A. gigantea* occurred approximately 40 MYA. Data were obtained from the TimeTree public database on the tree of life and its evolutionary timescale[44].

We manually annotated about 3,000 genes in CheloAbing 1.0, including the complete degradome[45] (set of protease genes) and other genes related to the above mentioned topics of special interest. We observed hallmarks suggesting selective pressure over some of these processes, such as longevity and stress response (including its interplays with pathogens). The immune system, sometimes referred to as an 'evolutionary driver'[46], presents several variants that, taken together, suggest a stronger innate immune system in turtles and a tight modulation of the inflammatory process, which could also play a role in the control of cancer development and longevity. Another major source of variations in *C. abingdonii*'s genome was found in genes associated with development. Thus, we observed a substantial copy-number variation in the β-keratin family, which is likely to play an important role in shell development. Additionally, we present evidence for variants in the transcription factor *TP53* and in the family of globins that may be related to enhanced hypoxia tolerance in turtles. We also describe the putative role of the *IGF* system in both size and longevity of giant tortoises. Finally, other variants found in different giant tortoises suggest biochemical changes that might merit further study (**Tables S6 and S7**).

2. Generic developmental features

Turtles present a series of anatomical and physiological features that make them one of the most peculiar vertebrate groups. Besides the fact that they are the last representative member of anapsids, their back-bones grow in a unique way to give rise to their shell, a complex mosaic of keratinocytic plates. Also, turtles show a striking diversity in body size. We annotated more than 380 genes involved in development to better understand the genetic basis of the differential phenotypic traits present in giant tortoises, along with other traits associated to all Testudines (**Table S8**).

Table S8. Status of development-related genes analysed in *C. abingdonii*.

ABCA6	CLCN2	FGF3	HMGB1	HOXD12	MCOLN2	RBP1	SOX5	USP26
ACAN	CNDP1	FGF4	HMGB2	HOXD13	MCOLN3	RBP2	SOX6	VTG1
ADAM15	COL11A1	FGF5	HMGB3	HOXD3	MEOX1	RBP5	SOX7	VTG2
ADAM18	CRABP1	FGF6	HMGB4	HOXD4	MEOX2	RBP7	SOX8	VTG3
ADAM2	CRABP2	FGF7	HMGN1	HOXD8	MMP20	RH1	SOX9	VMO1
ADAM20	CRISPLD2	FGF8	HMGN2	HOXD9	MNX1	RH2	SP1	VMO1L1
ADAM7	CTGF	FGF9	HMGN3	HR	MSX1	RUNX1	SP2	VMO1L2
ADAMTS3	CYP2J19	FIBIN	HMGN4	IGF1	MSX2	RUNX2	SP3	VMO1L3
ADAMTS6	CYR61	FOXP2	HMGN5	IGF1R	MSX2P1	RUNX3	SP4	VMO1L4
AMBN	DIXDC1	FSHR	HOXA1	IGF2	MSX3	SCNN1A	SP5	WISP1
AMELX	DMP1	FZD1	HOXA10	IGFBP1	MTR	SCNN1B	SP6	WISP2
AMELY	DSPP	FZD10	HOXA11	IGFBP2	MUC5B	SCNN1G	SP7	WISP3
AMTN	DVL1	FZD2	HOXA13	IGFBP3	MYF5	SCPPPQ1	SP8	WNT1
APCDD1	DVL2	FZD3	HOXA2	IGFBP4	MYF6	SERPINA1	SP9	WNT10A
APOV1	DVL3	FZD4	HOXA3	IGFBP5	MYOD1	SETD2	SPP1	WNT10B
APP	ELN	FZD5	HOXA4	IGFBP6***	MYOG	SFRP4	STC1	WNT11
AR	ENAM	FZD6	HOXA5	IGFBP7	NOV	SLC8A1	STC2	WNT16
ATM	EPS15	FZD7	HOXA6	IGF2R	NOX4	SMAD1	SUSD5	WNT2
AVD	ERN1	FZD8	HOXA7	IHH	NPEPPS	SMAD2	SWS1	WNT2B
AVDL1	EVX1	FZD9	HOXA9	IL11	NSD1	SMAD3	SWS2	WNT3
AVDL2	EVX2	GBX1	HOXB1	IL11RA	ODAM	SMAD4	TAS1R1	WNT3A
AVDL3	FABP1	GBX2	HOXB13	INS	ORs (622)*	SMAD5	TAS1R2	WNT4
AVDL4	FABP2	GDF10	HOXB2	INSL3	OTUD4	SMAD6	TAS1R3	WNT5A
BMP1	FABP3	GDF11	HOXB3	INSL4	OXTR	SMAD7	TAS2R (8)**	WNT5B
BMP10	FABP5	GDF15	HOXB4	INSL5	PAM	SMAD9	TCF7	WNT6
BMP15	FABP6	GDF1L	HOXB5	INSL6	PAPPA	SMO	TCF7L1	WNT7A
BMP2	FABP7	GDF2	HOXB6	INSR	PDE5A	SOD1	TCF7L2	WNT7B
BMP3	FABP9	GDF5	HOXB7	INSRR	PDX1	SOD2	TESSP5	WNT8A
BMP4	FGF1	GDF6	HOXB8	ITGA8	PF4	SOD3	TIPARP	WNT8B
BMP5	FGF10	GDF7	HOXB9	KLK4	PIEZO1	SOX1	TRPA1	WNT9A
BMP6	FGF11	GDF8	HOXC10	LEF1	PKD2L1	SOX10	TRPC1	WNT9B
BMP7	FGF12	GDF9	HOXC11	LGALS3	PLCB2	SOX11	TRPC2	XPC
BMP8L	FGF13	GHR	HOXC12	LOXL4	PMP2	SOX12	TRPC3	XPNPEP1
CACNA1D	FGF14	GLT8D2	HOXC13	LRP5	POLB	SOX13	TRPC4	ZBTB7B
CACNA1E	FGF16	GNA11	HOXC3	LRP6	PPBP	SOX14	TRPC5	ZFAT
CACNA1H	FGF17	GREM1	HOXC4	LTF	PRKCB	SOX15	TRPC6	αKRTs (35)**
CALHM1	FGF18	GREM2	HOXC5	LWS	PRKCG	SOX17	TRPC7	βKRTs (26)**
CBLC	FGF19	GSK3A	HOXC6	MC1R	PRRX1	SOX18	TRPM5	
CCNB1IP1	FGF2	GSK3B	HOXC8	MC2R	PRRX2	SOX2	TRPV1	
CCND3	FGF20	GSX1	HOXC9	MC3R	PRSS12	SOX21	TRYX2	
CDX1	FGF21	GSX2	HOXD1	MC4R	PSEN1	SOX3****	TSP50	
CDX2	FGF22	HMGA1	HOXD10	MC5R	PSEN2	SOX30	TWIST1	
CDX4	FGF23	HMGA2	HOXD11	MCOLN1	PTH1R	SOX4	TWIST2	

Expansion | Absence | Without relevant variants | Loss-of-function variants | Remarkable missense variants

Tables S8, S9, S10, S11, S12, S14 and **S15** summarize the status of the analysed genes in the different fields of interest. Green boxes indicate the duplication/expansion of the mentioned gene, while blue boxes indicate its putative absence in the genomic sequences of *C. abingdonii*. Red boxes indicate premature stop codons or frameshifts in the genomic sequence. In addition, the presence of missense variants localized in functional domains of the protein was

indicated with orange boxes. *, annotated using automatic approaches. **, genes belonging to a specific gene family, but whose orthology relationship is unclear. The approximate number of genes found in these cases is indicated between parentheses. ***, genes with very low identity to their human orthologs. ****, shorter genes, possibly due to a different initial methionine.

2.1. Genetic basis of development

The determination of morphological identity and anteroposterior regionalization of structures in vertebrates has been classically considered of main interest in evolutionary and developmental biology[47]. As central players of animal development, homeobox-containing (Hox) genes are involved in controlling and determining the final shape and body-plan of an organism[48]. Thus, loss of Hox gene functions or differences in their regulation can lead to striking morphological changes[49,50]. In this context, we analysed the Hox gene set in *C. abingdonii* and identified 39 different family members in this tortoise, which are highly conserved throughout evolution. This set of 39 Hox genes was also conserved in *A. gigantea*.

As in other Testudines, placental mammals, crocodiles and birds, no *HOXC3*-related sequences could be identified in *C. abingdonii*, *A. gigantea* or *G. agassizii* (**Fig. S4**). This gene is only present in fishes, amphibians, snakes and lizards[51]. The absence of *HOXC3* in this tortoise supports its early loss in the radiation of Archelosauria[6,52]. Furthermore, we assessed the presence of paraHox genes in *C. abingdonii*, involved in the anteroposterior development of gut and nervous system of animals[53]. The annotation of this group of genes revealed the presence of a duplication affecting *CDX4* (**Table S21**) which is characteristic of Sauria.

Other genes that perform basic functions in the embryo developmental process and the maintenance of homeostasis, such as those coding for bone morphogenetic proteins (BMPs) and growth differentiation factors (GDFs)[54], are well conserved in CheloAbing 1.0. Thus, we identified orthologues for *BMP1*, *BMP2*, *BMP3*, *BMP4*, *BMP5*, *BMP6, BMP7*, *BMP10* and *BMP15*, but contrary to primates which have duplicated *BMP8* (*BMP8A* and *BMP8B*), there is only a single *BMP8*-like gene in *C. abingdonii*. We also identified the orthologs of *GDF2*, *GDF5*, *GDF6*, *GDF7, GDF8*, *GDF9*, *GDF10*, *GDF11 and GDF15* in *C. abingdonii.* By contrast, we could not distinguish *GDF1* and *GDF3* orthologs, as only one common ancestor-related gene, *GDF1L*, is present in Archelosauria. The high level of identity between these genes in the selected species suggests that their function is indispensable in Testudines.

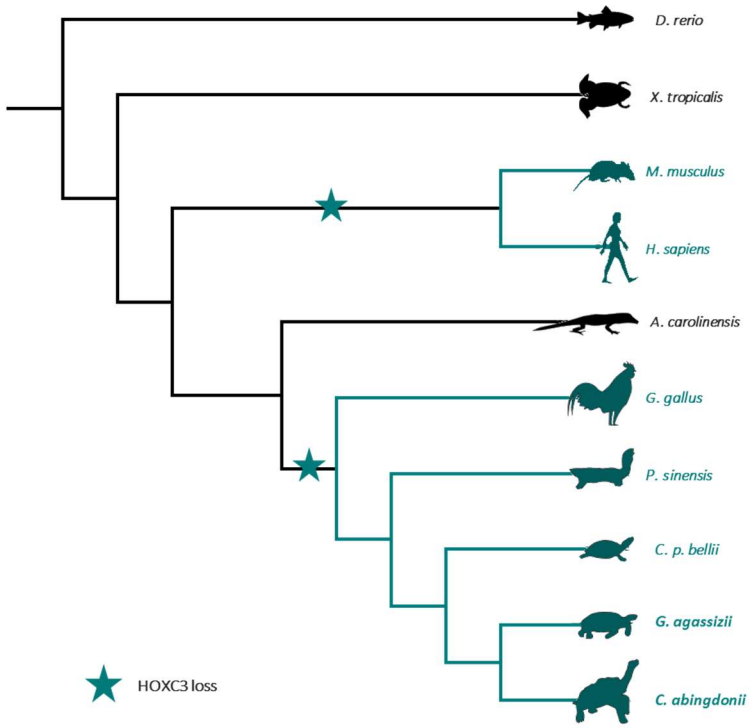


Figure S4. Evolutionary scenario of HOXC3 loss within different vertebrates. Green lines show the putative independent losses of the HOXC3 gene in mammalian and Archelosauria lineages, in contrast to fish, amphibians and the rest of reptiles. The phylogenetic tree was generated using Common Tree Taxonomy tool from NCBI[1].

Finally, we studied the WNT family, which includes a large number of cysteine-rich glycoproteins that activate signal-transduction cascades. These cascades are involved in the control of a wide range of developmental events during embryogenesis, and in adult tissue homeostasis[55]. The analysis of CheloAbing 1.0 revealed duplications in two members of this family, *WNT4* and *WNT11* (**Table S21**). Although both genes are essential during the development of the embryo, the case of *WNT11* is particularly interesting, since this gene contributes to skeletal development through osteoblast maturation[56]. This duplication was found in turtles and *G. gallus*, but not in *A. carolinensis*. On the other hand, the two copies of *WNT4* were only identified in turtles. Given the role of these genes in the development of bone structures, these duplications might be related to the growth of the shell.

2.2. Teeth

Testudines lost the ability to grow teeth approximately 150-200 million years ago, becoming the oldest extant edentulous lineage of tetrapods[57]. Previous studies in birds and edentulous Mysticeti whales proved that tooth loss is closely associated with the pseudogenization and subsequent loss of the tooth-specific genes enamelin (*ENAM*), amelogenin (*AMEL*), ameloblastin (*AMBN*), dentin sialophosphoprotein (*DSPP*), and the proteases kallikrein-4 (*KLK4*) and enamelysin (*MMP20*)[23,58]. Recently, these findings were also confirmed in the western painted turtle[5]. As expected, we did not find any of these tooth-specific genes in *C. abingdonii* or in *A.*

gigantea, indicating their possible absence or pseudogenization in Testudines. These results confirm a pattern of multiple pseudogenizations associated with tooth loss, followed consistently and independently in multiple vertebrates.

2.3. Neurological development

The development of the nervous system is a complex and intricate process in which, alongside the paraHox family, the participation of multiple genes is required. Our study revealed a series of variants in different genes involved in this mechanism and putatively associated with neurological disorders.

In this regard, the serine/threonine protein kinase ATM, involved in cell cycle arrest, apoptosis, brain development and synaptic function, showed two variants in conserved residues (p.R250Q and p.Y316H). Both variants are present in all the Galapagos tortoises, while the first variant (p.R250Q) is also present in their continental outgroups and the Aldabra giant tortoise (**Fig. S5**). These variants (confirmed by Sanger sequencing) have been linked to ataxia telangiectasia in humans (ClinVar[59]), a syndrome characterized by impairment in movement and coordination, weakened immune system and difficulty to repair DNA breaks[60]. Presumably, these variants found in tortoises are associated with other balancing changes to avoid any deleterious consequences. This mechanism of the Dobzhansky class[61] would allow the genome of tortoises to explore different solutions for the regulation of ATM that are not accessible to other organisms. Moreover, this gene also presents a variant (p.V2873I) at one of the nine residues involved in the catalytic loop formation[62]. This variant was also confirmed through Sanger sequencing and is present in all the different species of Galapagos giant tortoises tested and also in fishes, but it is absent in *A. gigantea* and in both continental outgroups.

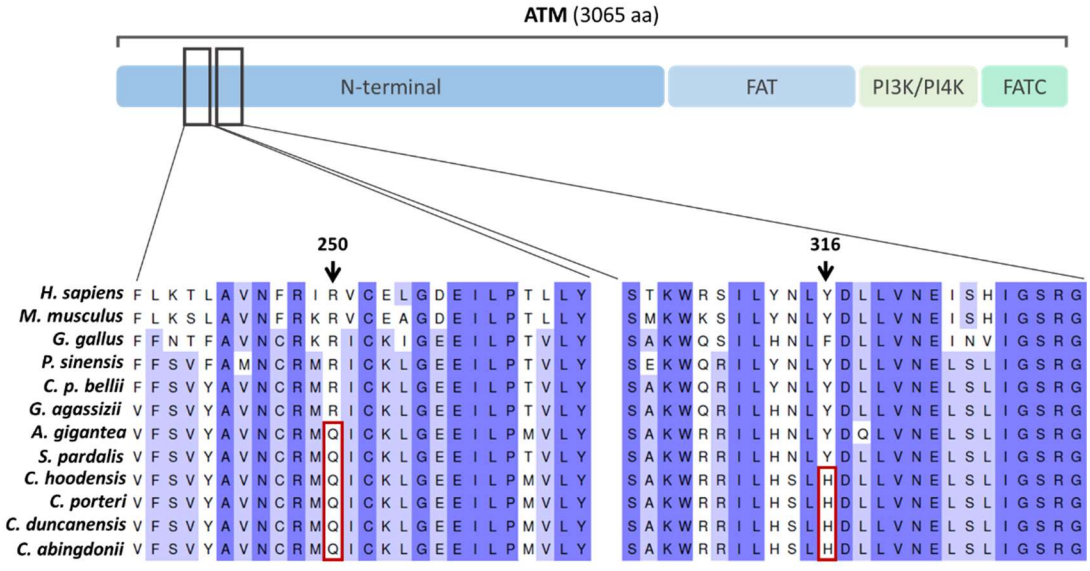


Figure S5. Partial amino acid alignment of the ATM sequence in *C. abingdonii, A. gigantea*, outgroups and other vertebrate species, including three turtles. Variants p.R250Q and p.Y316H present in giant tortoises are emphasized with a red rectangle.

Alongside these changes in *ATM*, the neurology-related protease gene *XPNPEP1* presents a premature stop codon (confirmed by RNA-Seq) at the beginning of the protein (p.D16*). This variant is present in *C. abingdonii*, other Galapagos tortoises, *A. gigantea* and their continental outgroups, as revealed through Sanger validation, and likely results in the loss of the function of this protease. The absence of *XPNPEP1* is associated with several dysfunctions, including microcephaly and neurodevelopmental retardation[63]. Similarly, *CNDP1* presents a truncating frameshift (at residue p.E274) in all Galapagos giant tortoises, the Aldabra tortoise and their continental outgroups, as validated by Sanger sequencing. In mammals, the loss of this protein causes carnosinemia, a recessive deficiency that leads to mental retardation, developmental delay and various neuropathies[64,65].

Motopsin, *PRSS12*, encodes a serine protease linked to nonsyndromic mental retardation[66]. This gene appeared to be truncated only in *C. abingdonii* (validated by Sanger sequencing) due to a premature stop codon (p.R691*). Similarly, the aspartyl protease PSEN1 was found to present a point variant (p.R352E) common to all Sauropsida, validated through Sanger sequencing in giant tortoises and their relative outgroups. In humans, a mutation at the same residue (p.R352C) has been linked to the early development of Alzheimer's disease[67,68].

Altogether, these alterations suggest that the neurological development of giant tortoises has undergone a number of molecular changes involving losses of gene functions. On the other hand, the evidence for positive selection affecting genes involved in neural development (*FGF19*, *ITM2B* and *NTF3*) suggests that this system has evolved through complex mechanisms, which may include and balance apparently deleterious variants. In this context, changes specifically associated to giant tortoises, such as those in *XPNPEP1*, could be linked to the evolutionary history of these species.

2.4. Reproduction

In contrast with placental mammals, turtles and reptiles are oviparous[69]. To explore the genomic hallmarks associated with this trait, we annotated genes related to egg and eggshell formation. Among the egg-related genes, we identified in *C. abingdonii* and *A. gigantea* the presence of vitellogenins (VTG1, 2 and 3), apovitellenin-1 (APOV1), ovomucin-α (also known as MUC5B), 4 paralogues of VMO1 (vitelline membrane outer layer protein 1), and 4 paralogues of AVD (avidin). The formation of the eggshell is crucial for the protection of the egg contents and the embryo after the lay, as well as for the prevention of contamination by microorganisms. To exert its function correctly, the eggshell has to undergo a complex process of calcification, which occurs in the uterus and is one of the most rapid biological mineralization processes known[70]. To study this process both in *C. abingdonii* and *A. gigantea*, we analysed several genes that play an important role in both Ca^{2+} transport and Ca^{2+} signalling pathways, which are essential mechanisms in eggshell formation. The presence of *CACNA1D, CACNA1E, CACNA1H, PRKCB, NCX1* (or *SLC8A1*)*, GNA11* and *OXTR* reinforces the importance of these pathways in eggshell calcification. Likewise, it is well known that calcium transport in the uterus is associated with the counter-transport of Na^+ and Cl^-, and with the secretion of shell matrix proteins[71]. In this context,

we found different genes involved in these processes, like *SCNN1A, SCNN1B, SCNN1G* (transport of Na^+) and *CLCN2* (transport of Cl^-). Finally, we also identified in CheloAbing 1.0 and confirmed in the genome of *A. gigantea* some genes, such as *SPP1, LTF, ITGA8, LGALS3* and *COL11A1*, which encode proteins of the eggshell matrix, the non-mineral component of the eggshell.

Between his discovery in 1971 and his death in 2012 there were many attempts to mate the last individual of *C. abingdonii* with females from other species of Galapagos giant tortoises[72]. Unfortunately, non-fertilized eggs were found in the pen where he lived together with two closely related females. To explore his inability to reproduce, we analysed key genes involved in this process such as *CCNB1IP1* or *FSHR*. However, we did not find any specific change to explain this observation. That being said, alterations in genes related to the thyroid function that might be also involved in reproductive difficulties were found (**see section 4.3**).

2.5. Shell formation

One of the most conspicuous characteristics of turtles is the presence of a shell that covers their entire body. In fact, the turtle shell constitutes an evolutionary novelty in which the developmental pattern of the ribs is radically modified[73]. Specifically, turtle ribs grow laterally to form the carapace, the dorsal roof of the shell, unlike the ribs of other amniotes that grow ventrally[74]. A possible candidate for these turtle-specific changes in development is the carapacial ridge (CR), a specific structure of the turtle embryo[75]. Some CR-specific genes – such as *LEF1*, *APCDD1*, *CRABP1* and *SP5* – which are effectors or transcriptional targets of the canonical WNT pathway - were identified in both the *C. abingdonii* and *A. gigantea* genomes. Then, by using manual annotation, we detected most of the remaining genes involved in this pathway. All the members of the WNT family are conserved in giant tortoises including the aforementioned duplication of *WNT4* and *WNT11*. The *DVL* and *FZD* families are also apparently complete and functional in *C. abingdonii* and *A. gigantea*. *LRP5*, *LRP6*, *GSK3A*, *GSK3B* and the orthologues of *LEF1*, *TCF7*, *TCF7L1* and *TCFL2* are also present in these tortoises. Likewise, *RUNX1*, *RUNX2*, *GREM1*, *GREM2*, *MSX1*, *TWIST1*, *SHH*, *HGF*, *MYF5*, *FGF8*, *FGF10* and genes of the BMP family (discussed in **section 2.1**), likely involved in shell formation, are also present in *C. abingdonii* and *A. gigantea*. In summary, we found a wide range of presumably functional genes that likely participate in the formation of the shell. These genes are well conserved between vertebrates. Currently, the place and time of expression is thought to be a main determinant to define the role of these genes in shell formation[75].

During evolution, the integument of amniotes has transitioned from a basic cornified epidermis, for protection against the environment and retention of water, into an elaborate covering studded with epidermal structures[76]. These structures, such as claws, beaks, scales or feathers of reptiles and birds, are used additionally for sexual display, camouflage, locomotion and thermoregulation[77]. These complex evolutionary inventions result from the products of two gene families: alpha (α) and beta (β) keratins. α-keratins are divided in two groups, type I and type II, and form heterodimers that constitute the structural basis of the cornified epidermis and epidermal appendages. β-keratins are exclusive of Sauropsida and contribute to the formation

of the hard keratinous claws and scales of reptiles, as well as the beaks and feathers in birds[78]. Our results based on the manual annotation of type I and II α-keratins in CheloAbing 1.0 suggest a reduction in the number of these fibrous proteins when compared to other organisms. Specifically, we identified a total of 35 α-keratins (19 type I and 16 type II). By contrast, mammals present more than 50 family members on average and other reptiles display 40 or more. For instance, the green anole lizard has 41 α-keratins, while the American alligator and the saltwater crocodile have 44 and 40, respectively[77]. It must be noted that the annotation of keratins is very sensitive to the quality of the assembly, and these numbers can be over- or underestimated in some cases. If confirmed, we hypothesize that this reduction in the α-keratins repertoire of *C. abingdonii* could be a consequence of the emergence of β-keratins, essential for the development of the turtle shell, an evolutionary novelty exclusive of this clade.

Several studies suggest that, concomitant with shell formation, the β-keratin gene family underwent repeated duplications in the turtle lineage from genes involved in the formation of the claws and scales in the ancestor of Sauropsida[78,79]. Using manual annotation, we found 30 β-keratins in CheloAbing 1.0, 26 of which seem to be functional. These results suggest that the number of these genes in *C. abingdonii* is comparable to that of birds[77], but differ from recent studies that have identified 89 β-keratins in *C. p. bellii* and 74 in *P. sinensis*, respectively[78]. However, using our supervised approach, we found 70 apparently functional β-keratins in *C. p. bellii* and 57 in *P. sinensis*. Finally, and also in relation to genes implicated in shell formation, it is remarkable that the list of genes under positive selection in *C. abingdonii* features *KDSR,* whose mutations cause keratinisation disorders[80].

2.6. Growth and size

Giant tortoises are classified in the upper end of the size scale for extant Chelonii[81]. However, the genetic determinants of body size and growth in these turtles remain unexplored. In this context, we analysed different genes involved in the growth process in order to better understand the mechanism underlying the gigantism of these animals and provide baseline data for future functional studies.

We first examined in CheloAbing 1.0 and in the *A. gigantea* genomes the *GHR* gene, coding for the growth-hormone receptor, as well as all the components of the insulin-like growth factor (IGF) system, including *IGF1/2*, *IGF1R*, *IGF2R*, *INS*, *INSR*, *INSRR*, *IGFBPs,* and *PAPP-A*. IGFs and insulin play important roles in the regulation of multiple processes, like metabolism and growth, and disruptions in the GH/IGF pathway cause growth disorders in humans[82,83]. Furthermore, studies carried out with domestic dogs (*C. lupus familiaris*) have suggested that genetic variants in *IGF1*, *IGF1R* and *GHR* genes are related to differences in breed size[84]. Our annotation revealed a high degree of conservation between all species considered, illustrating the basic functions of these proteins. None of the changes in IGF1R (p.R204H) or GHR (p.P177L; p.E191K) that are associated with small size[84] were found in *C. abingdonii*, *A. gigantea* or *G. agassizii*. In addition to the IGF system, other genes have been recently connected to the regulation of animal size. Therefore, we also annotated stanniocalcins (*STC1* and *STC2*), widespread glycoproteins that

seem to be involved in growth through the interaction with PAPP-A, a protease essential in IGF signalling[85]. We found a high degree of STC1 and STC2 conservation in *C. abingdonii* and *A. gigantea*, suggesting that these proteins are functional in giant tortoises.

The gigantism in tortoises is probably caused by a combination of different genetic factors, as body size is regulated at multiple levels. Our finding that most known factors involved in this trait are perfectly conserved in *C. abingdonii* and *A. gigantea* suggests that other evolutionary pathways may have contributed to this gigantism in these tortoises (**Fig. S6**), which was likely a preadaptation that contributed to the successful colonization of remote oceanic islands[86].

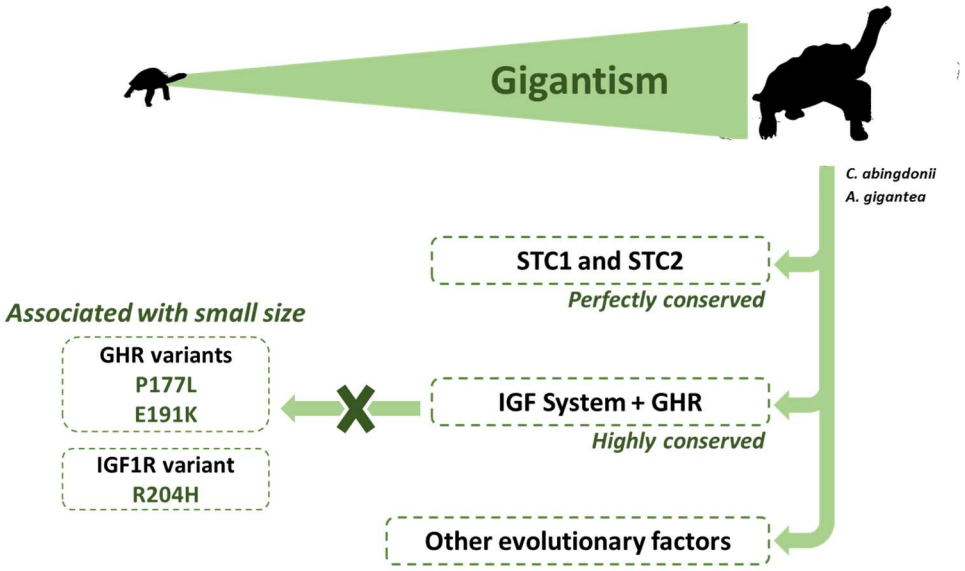


Figure S6. Status of different genes involved in growth and size in *C. abingdonii* and *A. gigantea*.

2.7. Sensorial perception

Due to the extraordinary diversity of turtles, the study of sensorial perception in these animals including giant tortoises could reveal some interesting insights about their lifestyle. To study these traits, we analysed genes associated with the senses of sight, taste and smell.

2.7.1. Sight

For most Craniata, photosensitivity is critical for the control of behavioural response. In vertebrates, vision is mediated via cones that are functional under bright-light conditions and rods that detect low levels of light[87]. The photopigments that reside in the retinal photoreceptors are responsible for the detection of light and receive the name of opsins. Five classes of these visual pigments have been identified in Craniata: rod opsin *RH1* (*RHO*), and cone opsins *RH2*, *LWS* (*OPN1LW*), *SWS1* (*OPN1SW*) and *SWS2* (*OPN1SW2*)[88]. We identified all five

opsin genes in the genomes of *G. agassizii*, *C. abingdonii* and *A. gigantea*, providing the potential for tetrachromacy in these tortoises (**Fig. S7**).

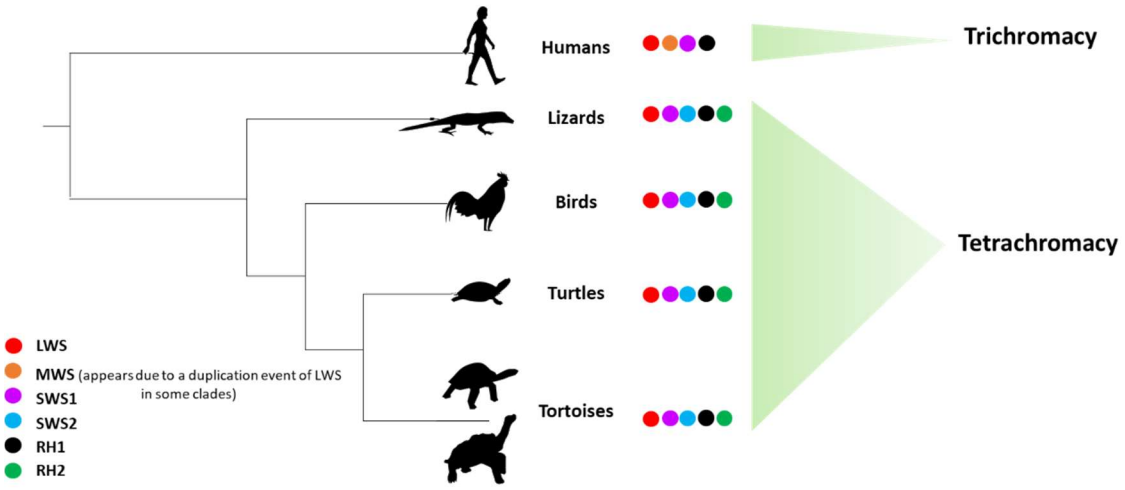


Figure S7. Divergence tree showing the repertoire of opsin genes along different clades.

2.7.2. Taste

Taste perception is a key ability for the survival of animals and is highly specialized to sense nutritious or harmful compounds in potential food sources. This sense is triggered by the physical interactions between tastant molecules from the external environment and taste receptors[89]. Vertebrates can typically detect five taste qualities: sweet, umami, bitter, sour and salty. The candidate receptors for sour and salty tastes are transient receptor potential ion channel PKD2L1 and the sodium channel ENaC (constituted by subunits SCNN1A, SCNN1B and SCNN1G), respectively. Umami and sweet tastes are sensed by the TAS1R family of G protein-coupled receptors (GPCRs), with the TAS1R1-TAS1R3 heterodimer functioning as the umami receptor and the TAS1R2-TAS1R3 heterodimer being the sweet receptor. Bitter tastants are detected by the TAS2R family of GPCRs[90]. We identified all the genes encoding the receptors responsible for sour and salty tastes (*PKD2L1, SCNN1A, SCNN1B* and *SCNN1G*) in the genomes of *C. abingdonii*, *A. gigantea* and *G. agassizii*. However, in the TAS1R family we found that only TAS1R2 and TAS1R3 are apparently functional. The other member of this gene family, TAS1R1, is truncated by a premature stop codon (p.T436*, found in *C. abingdonii* and *A. gigantea;* and p.W297* in *G. agassizii*). Finally, regarding bitterness receptors, our results showed that, as in birds, tortoises seemed to have a lower number of members of this gene family (8 in CheloAbing 1.0, 9 annotated in desert tortoises) than mammals (36 members in humans and 42 in mice)[91]. Altogether these results suggest that while sour, salty and sweet tastes could be perfectly perceived by *C. abingdonii*, *A. gigantea* and *G. agassizii*, these species might have lost the ability

of tasting umami. In addition, the reduced number of TAS2R receptors suggests a lesser perception of the bitter taste (**Fig. S8**).

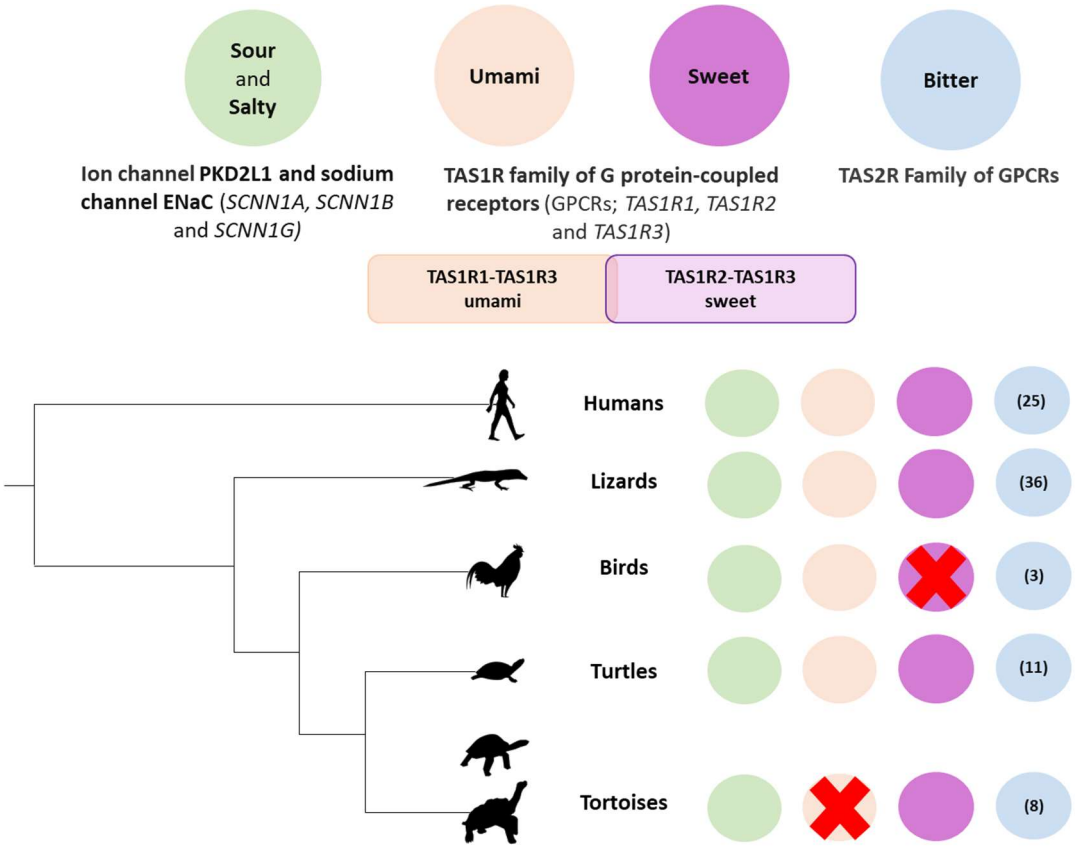


Figure S8. Divergence tree showing the status of taste receptors in different clades. As representatives for each clade we chose *A. carolinensis* (lizards), *G. gallus* (birds), *P. sinensis* (turtles) and *C. abingdonii, A. gigantea,* and *G. agassizii* (tortoises). The number of TAS2R family members that are apparently functional in these organisms is shown in parentheses[92].

2.7.3. Smell

Olfaction is also essential for survival in mammals. This important physiological function is controlled by a large multigene family of olfactory receptors (ORs) that recognize different odour molecules from the environment[93]. These receptors are known to be the largest multigene family in vertebrates. However, the number of OR genes differs largely among different species, and include several pseudogenes in addition to the functional copies[94]. In this context, we identified 622 apparently functional genes of this family in CheloAbing 1.0. This number is higher than the ORs annotated in *Chelonia mydas* (green sea turtle), in which a total of 254 intact receptors have been found[6].

3. Immunity

The identification and analysis of specific genes related to immune response could provide new insights about the development and evolution of the immune system in different organisms. The dynamics of host-pathogen interactions promotes a strong selective pressure during evolution, increasing the diversity of the immune system among different species[95]. Even though immunity has been extensively studied in mammals and birds, little is known about its specific components and functionality in non-avian reptiles, including turtles[96]. As the sister group of archosaurs[97], chelonians fill a critical gap in the evolution of immunity in amniotes[98]. Therefore, and due to the major roles that the immune system plays in multiple biological processes such as tumour protection and ageing, we exhaustively studied a group of 892 immunity-related genes in the genome of *C. abingdonii* (**Table S9**), in addition to the MHC gene family, which due to its importance and complexity was analysed separately. Finally, variants selected for their potential functional associations were subsequently investigated in the genome of *A. gigantea*.

Table S9. Status of immune system-related genes analysed in *C. abingdonii*.

ABI1	CCL1	CSF1R	ELF1	IFNA6	ITGB2	NFAT5	PUM2	TKFC
ABL1	CCL11	CSF3	ELMO1	IFNA7	ITGB7	NFATC1	PVRL1	TLR1
ABR	CCL13	CTLA4	ELMO2	IFNA8	ITK	NFATC2	PYCARD	TLR10
ADAR	CCL16	CTNNB1	EMP2	IFNAR1	JAG1	NFATC3	RAB17	TLR13
ADORA2B	CCL17	CTSGL	ENDOU	IFNAR2	JAK1	NFATC4	RAB27A	TLR14
AICDA	CCL2	CUEDC2	ENPP1	IFNB1	JAK2	NFIL3	RABGEF1	TLR15
AIRE	CCL20	CX3CR1	ENPP2	IFNE	JAK3	NFKB1	RAC2	TLR2
ANG	CCL3	CXCL1	EOMES	IFNG	JAM3KIT	NFKB2	RAF1	TLR21
ANGPT1	CCL4	CXCL10	EPHB6	IFNGR1	KIR2DS1	NFKBIB	RAG1	TLR3
ANKRD17	CCL5	CXCL11	EPRS	IFNGR2	KIR2DS2	NFKBID	RARA	TLR4
ANXA1	CCL7	CXCL12	ERBB2	IFNK	KIR2DS4	NFKBIZ	RASGRP1	TLR5
ANXA3	CCL8	CXCL14	ERBB2IP	IFNL1	KIR2DS5	NKIRAS1	RBPJ	TLR6
AP1G1	CCR2	CXCL16	ERBB4	IFNL2	KIR3DL2	NKIRAS2	RC3H1	TLR7
APBB1IP	CCR3	CXCL2	ERCC1	IFNL3	KIR3DS1	NLRC3	RC3H2	TLR8
APEX2	CCR4	CXCL5	EREG	IFNLR1	KLHL6	NLRC4	REL	TLR9
APLF	CCR6	CXCL8	ETS1	IFNW1	KLRB1	NLRPs	RELA	TMEM173
APOA1	CCR7	CXCL9	EXO1	IK	KLRC2	NLRX1	RELB	TNFAIP1
APOA2	CD14	CXCR2	EXOSC3	IKBKAP	KLRD1	NMI	RELT	TNFAIP8L2
APOA4	CD164	CXCR3	EZR	IKBKB	KLRF2	NOD1	RFX1	TNFRSF11A
APOBEC1	CD180	CXCR4	F2RL1	IKBKE	KLRG1	NOD2	RGCC	TNFRSF11B
APOBEC2	CD19	CXCR5	FADD	IKBKG	KLRG2	NOTCH1	RGS1	TNFRSF12A
APOBEC3A	CD1D	CYP27B1	FAS	IL12B	KLRK1	NOTCH2	RIPK1	TNFRSF13B
APOBEC3B	CD200	CYSLTR2	FASLG	IL12RB1	LAG3	NPY	RIPK2	TNFRSF14
APOBEC3C	CD207	DAB2IP	FAU	IL13	LAIR1	NRROS	RNF135	TNFRSF17
APOBEC3D	CD209	DAPK1	FBXO9	IL13RA1	LAMP1	OASL	RORA	TNFRSF18
APOBEC3F	CD22	DAPK3	FCER1A	IL15	LAT	OPRD1	RPS6KA3	TNFRSF1A
APOBEC3G	CD226	DDX1	FCER1G	IL16	LAT2	OPRK1	RSAD2	TNFRSF1B
APOBEC3H	CD247	DDX21	FCER2	IL17A	LBP	OPRM1	S1PR4	TNFRSF21
APOBEC4	CD27	DDX3X	FCGR1A	IL17B	LCK	OTUD7B	SAMHD1	TNFRSF4
AQP4	CD274	DDX41	FCGR1B	IL17C	LCP1	OTULIN	SAMSN1	TNFRSF6B
AQP9	CD28	DDX58	FCGR2A	IL17F	LCP2	P2RX7	SARM1	TNFRSF8
ARHGDIB	CD300A	DDX60	FCGR2B	IL17RD	LFNG	PAWR	SASH3	TNFRSF9
ARRB2	CD300E	DEFA1	FCGR3A	IL17RE	LGALS3	PAX5	SBNO2	TNFSF10
ART1	CD300LB	DEFA1B	FCGR3B	IL18R1	LGR4	PAXIP1	SEMA3C	TNFSF11
ATF1	CD300LF	DEFA3	FCN1	IL18RAP	LIG4	PDCD1	SEMA4A	TNFSF13B
ATF2	CD302	DEFA4	FCN2	IL19	LILRA2	PDE4B	SEMA4D	TNFSF14
ATF3	CD34	DEFA5	FCN3	IL1A	LILRA4	PDE4D	SEMA7A	TNFSF15
ATF4	CD36	DEFA6	FCRL5	IL1B	LILRA5	PDGFA	SERINC3	TNFSF8

ATF5	CD37	DEFB103A	FER	IL1R1	LPXN	PDGFB	SERINC5	TNFSF9
ATF6	CD38	DEFB103B	FGA	IL1R2	LSM14A	PDGFRB	SERPING1	TNIP1
ATF6B	CD3E	DEFB104A	FGB	IL1RAP	LSP1	PDPK1	SFTPD	TNIP2
ATF7	CD3G	DEFB104B	FLOT1	IL1RL1	LTB4R	PELI1	SH2B2	TNIP3
ATG12	CD4	DEFB105A	FOXA2	IL1RL2	LTBR	PGLYRP2	SH2D1A	TOLLIP
ATG16L1	CD40LG	DEFB105B	FOXJ1	IL20	LY75	PHB	SIGIRR	TRADD
ATG5	CD44	DEFB106A	FOXJ1L	IL20RA	LY9	PHPT1	SIKE1	TRAF1
ATG9A	CD46	DEFB106B	FOXP3	IL20RB	LY96	PIGR	SIN3A	TRAF2
ATP6V0A2	CD53	DEFB107A	FYB	IL21	LYST	PIK3AP1	SIRT1	TRAF2
ATP7A	CD55	DEFB107B	FYN	IL22	LYZ	PIK3R6	SIVA1	TRAF3
B2M	CD58	DEFB108B	FZD5	IL23A	MADCAM1	PKN1	SKAP1	TRAF3IP2
BATF	CD59	DEFB108P1	GAPDH	IL23R	MAP2K1	PKN2	SLAMF1	TRAF5
BCL10	CD5L	DEFB108P2	GAS6	IL24	MAP2K2	PKN3	SLC11A1	TRAF6
BCL2	CD6	DEFB109P1	GATA3	IL25	MAP2K3	PLCL2	SLC11A2	TRAFD1
BCL2A1	CD63	DEFB109P1B	GBF1	IL26	MAP2K4	PLEKHA1	SLC26A6	TRAIP
BCL2L1	CD69	DEFB110	GCH1	IL27	MAP2K6	POLR3B	SLC39A10	TRAT1
BCL3	CD74	DEFB112	GCNT3	IL27RA	MAP2K7	POLR3D	SMAD3	TREM1
BCL6	CD79B	DEFB113	GEM	IL2RB	MAP3K1	POU2AF1	SNAP23	TREM2
BCR	CD80	DEFB114	GFI1	IL2RG	MAP3K14	POU2F2	SNCA	TREML1
BDKRB2	CD81	DEFB115	GLL13L	IL31	MAP3K5	PPP2R3C	SNX4	TREML2
BECN1	CD83	DEFB116	GLL14L	IL31RA	MAP3K7	PPP3CB	SOCS1	TRIL
BID	CD86	DEFB118	GNLY	IL32	MAP3K8	PRF1	SOCS3	TRIM13
BIRC5	CD8A	DEFB119	GPLD1	IL33	MAP4K2	PRF1G	SOCS5	TRIM14
BLK	CD96	DEFB121	GPR183	IL34	MAPK1	PRF1I	SOCS6	TRIM28
BLNK	CD97	DEFB123	GPR65	IL36A	MAPK10	PRF1K	SPG21	TRIM32
BMP6	CDC37	DEFB124	GPRC5B	IL36B	MAPK11	PRG4	SPNS2	TRIM8
BMPR1A	CDK1	DEFB125	GPX1	IL36G	MAPK14	PRKCA	SPON2	TSC1
BPI	CDK9	DEFB126	GSTP1	IL36RN	MAPK8	PRKCB	SPPL2A	TSPAN6
BST1	CEBPB	DEFB127	GZMA	IL37	MAPKAPK2	PRKCD	SRC	TUBB
BTK	CEBPG	DEFB128	GZMBL	IL4	MAPKAPK3	PRKCE	ST6GAL1	TXK
BTLA	CFD	DEFB129	GZMHL	IL4R	MARCH1	PRKCG	STAP1	TYRO3
C1QA	CFH	DEFB130	GZMK	IL5	MARCH8	PRKCH	STAT1	UBASH3A
C1QB	CFHR5	DEFB131	HAMP	IL5RA	MARCO	PRKCQ	STK11	UBE2K
C1QBP	CFP	DEFB132	HAVCR2	IL6ST	MASP2	PRKCZ	STOML2	ULK1
C1QC	CHGA	DEFB133	HBEGF	IL7R	MAVS	PRKD2	STX7	UNC13D
C1R	CHIA	DEFB134	HCK	IL9	MB21D1	PRKRA	STXBP3	UNC93B1
C3	CHID1	DEFB135	HRH4	ILF2	MBL2	PRNP	SUSD4	UNG
C3AR1	CHUK	DEFB136	ICOS	ILF3	MED1	PROS1	SWAP70	VAMP7
C4BPA	CIITA	DEFB4A	IDO1	INPP5D	MEF2C	PSMC1	SYK	VAMP8
C4BPB	CITED1	DEFB4B	IFI16	IPO7	MEP1A	PSMC2	SYNCRIP	VAV1
C5	CLCF1	DHX15	IFI27	IRAK1	MFNG	PSMC3	TAB2	VAV3
C5AR1	CLEC2A	DHX29	IFI30	IRAK1BP1	MID2	PSMC4	TAC1	VCAM1
C5AR2	CLEC2B	DHX33	IFI35	IRAK2	MR1	PSMC5	TANK	VEGFA
C6	CLEC2D	DHX36	IFI6	IRAK3	MS4A1	PSMC6	TAPBP	VEGFC
C7	CLEC2L	DHX58	IFIH1	IRAK4	MSH2	PSMD10	TBK1	VIMP
C8A	CLNK	DHX9	IFIT1	IRF1	MSH6	PSMD11	TCF12	VIP
C8B	CLU	DLG1	IFIT2	IRF2	MSRB1	PSMD12	TCF7	VIPR1
C8G	CMA1L	DLL1	IFIT3	IRF2BP1	MUL1	PSMD13	TEC	VSIR
C8ORF4	CMKLR1	DNAJA3	IFIT5	IRF3	MVK	PSMD2	TEK	WASL
C9	CMTM3	DOCK1	IFITM1	IRF4	MYB	PSMD4	TESPA1	XBP1
CACNB3	CNR2	DOCK2	IFITM2	IRF5	MYD88	PSMD5	TFE3	XIAP
CACNB4	COCH	DRD2	IFITM3	IRF6	NAIP	PTAFR	TFEB	YTHDF2
CALCOCO2	COL3A1	DUSP10	IFNA1	IRF7	NCAM1	PTGER4	TGFB2	ZAP70
CAMP	COL4A3BP	DUSP3	IFNA10	IRF8	NCF1	PTK2	TGFB3	
CARD11	COLEC12	EBI3	IFNA13	IRF9	NCF2	PTK2B	THBS1	
CARD16	COX5B	EDA	IFNA14	IRG1	NCF4	PTPN1	THEMIS	
CARD18	CPA3	EDN1	IFNA16	IRGM	NCKAP1L	PTPN2	THOC1	
CARD6	CR1	EGF	IFNA17	ISG20	NCR2	PTPN6	THY1	
CARD8	CR2	EGLN2	IFNA2	ITCH	NCR3	PTPRC	TICAM1	
CARD9	CREBBP	EGR1	IFNA21	ITGA4	NDFIP1	PTPRJ	TICAM2	
CASP12	CRTAM	EIF2AK2	IFNA4	ITGAL	NEDD4	PTX3	TINAGL1	
CAMP4	CSF1	EIF4E	IFNA5	ITGB1	NFAM1	PUM1	TIRAP	

Expansion Absence Without relevant variants Loss-of-function variants Remarkable missense variants

3.1. Balance between innate and acquired immunity

Although several immune mediators can have dual functions in innate and adaptive immune responses, it is thought that the innate branch of the immune system in vertebrates evolved earlier than the adaptive route[96]. All multicellular organisms have some form of innate immune response, which acts as an initial step in the defence against pathogens[96]. Among vertebrates, Reptilia are the only ectothermic amniotes, and therefore the study of their immune system could provide new important insights into its evolution under different circumstances[96].

Previous genomic studies in Testudines suggested that the adaptive immune system in these animals is less robust than in mammals[5]. In this regard, we identified some features in the genomes of *C. abingdonii* and *A. gigantea* that support the enhanced role of innate immune defence in these organisms. Specifically, we found some putatively deleterious changes in genes involved in B-lymphocyte maturation. As an illustrative example, *MEP1A* – a metalloprotease involved in the inactivation of interleukin-6 (IL6) – shows several premature stop codons (p.Q320*, p.I535*, and p.I538*) due to the presence of two separate frame-shift mutations both detected only in the Galapagos giant tortoises (as assessed by Sanger sequencing validation, **Table S7**). Consistent with the role of IL6 in the differentiation of B cells[99], disruptions in *MEP1A* have been associated with altered homeostasis of monocytes and natural killer (NK) cells in mice[100].

On the other hand, the immune system seems to have been strengthened at different stages in Galapagos and Aldabra giant tortoises. Thus, we have identified an expansion of the *PRF1* gene in *C. abingdonii*, with 12 copies (**Table S21**), although three of them are pseudogenized through the generation of premature stop codons. Similarly, *A. gigantea* genome exhibits an expansion of *PRF1* and contains at least 12 functional copies of this gene. The existence of all these *PRF1* genes in both species of giant tortoises was validated by Sanger sequencing (**Table S7**). *PRF1* encodes a perforin protein, which is a component of the cytotoxic T lymphocyte (CTL) and NK cell secretory granules. This protein plays a key role in granule-dependent cell death and protection against virus-infected or neoplastic cells[101]. Other Testudines also present several copies of this gene, whereas other reptiles present two copies and birds only one (**Fig. S9**).

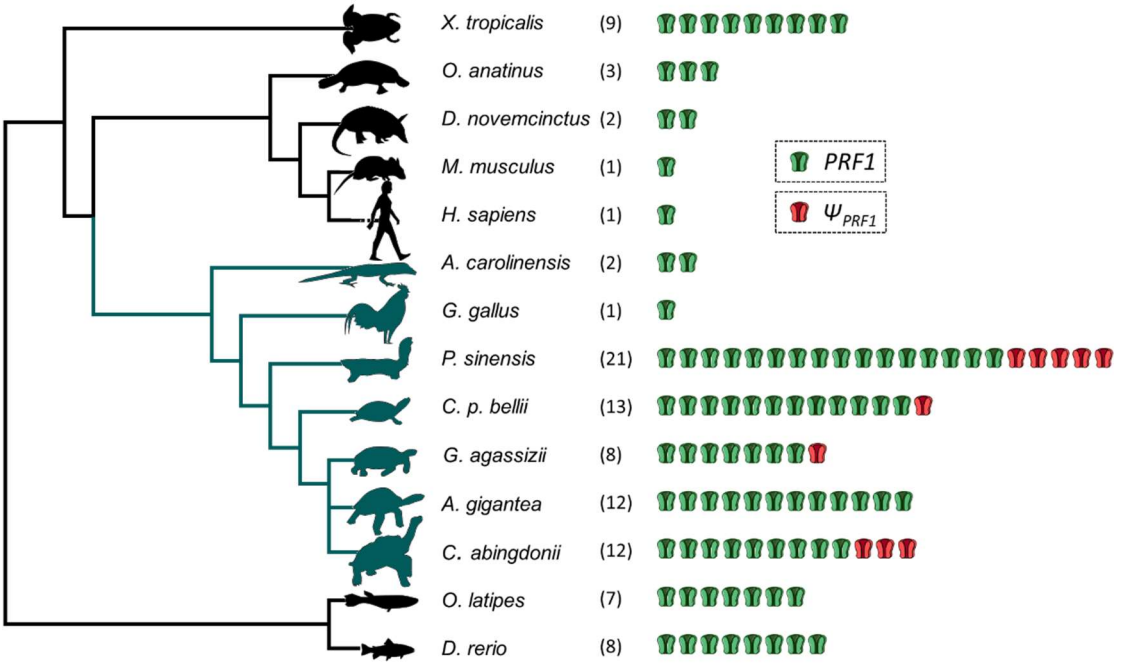


Figure S9. Divergence tree showing *PRF1* copy number in different species and the expansion of this gene in *C. abingdonii*, *A. gigantea* and other organisms. Icons represent *PRF1* copy number in each species. Ψ_{PRF1} represents a copy of *PRF1* with premature stop codons. Green lines indicate that manual annotation has been instrumental in characterizing these duplications. The phylogenetic tree was generated using the Common Tree Taxonomy tool from NCBI[1].

We also found another family expansion involving the granzyme (GZM) serine proteases (**Fig. S10**). This well-conserved family of proteolytic enzymes is expressed in 3 clusters, the chymase, the met-ase and the GZM A/K loci, the latter being conserved among Craniata[102]. The chymase locus, encoding *GZMH*, *GZMB*, *GZMG*, *CMA1* and *CTSG*, is greatly expanded, with 4 copies of *GZMH*, 5 of *GZMB*, 7 of *CTSG* and 2 of *GZMH*, but only one member of *CMA1* (**Table S21**).

These duplications detected in the genome of *C. abingdonii* are also present in all the other Galapagos giant tortoises tested and in *A. gigantea*, as assessed by Sanger sequencing of these genomic regions (**Table S7**). We detected a similar amplification in the genome of *G. agassizii*. Although some of these copies are pseudogenes, this copy-number variation evidences the importance of effector immunological pathways in tortoises, as previously reported in other organisms such as *M. musculus*, where a similar expansion was linked to an enhanced immune response[103]. Similar to PRF1, these serine proteases are key components of the CTL and NK cell secretory granules, playing important roles in defence against both pathogens and cancer[102].

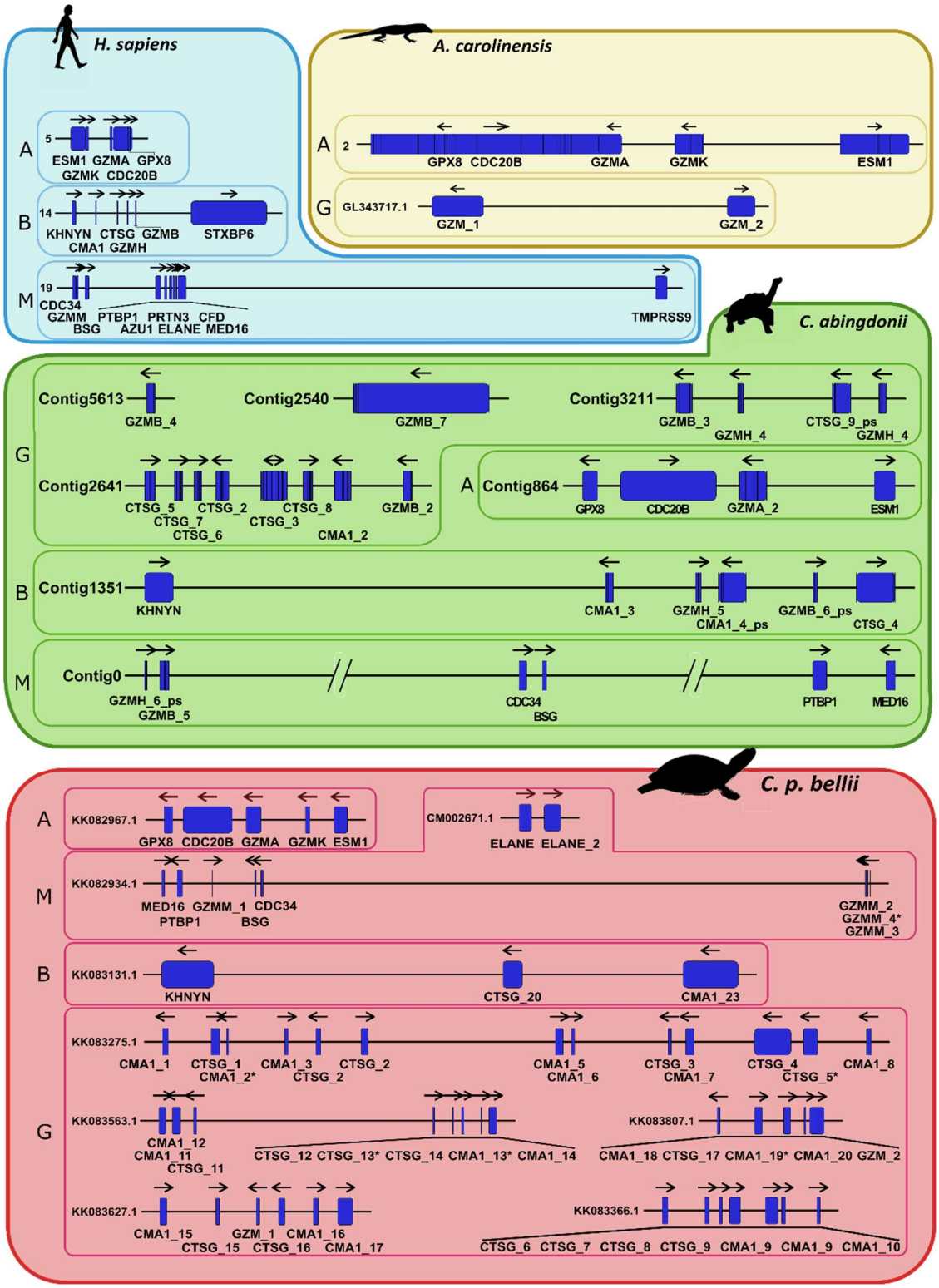


Figure S10. Granzyme clusters in different species and their expansion in *C. abingdonii* and other organisms. Blue boxes display the whole gene (both exons and introns), while the vertical black lines (when feasible) mark the exons of these genes. The cluster to which each group of contigs belongs, based on the flanking genes or the overall

presence of granzymes, is indicated by a letter preceding the box of the cluster. The expansions are showed by the extension number following the underscore. Pseudogenized copies are showed by adding 'ps' or an asterisk at the end of the name.

Similar to perforins and granzymes, the *APOBEC1* gene shows a consistent increase in copy number in different turtles, with three copies in *P. sinensis*, five in *G. agassizii*, six in *C. p. bellii* and seven in both *C. abingdonii* and *A. gigantea* (**Fig. S11; Table S21**). The antiviral function of this family in humans is reduced to the *APOBEC3* genes, with *APOBEC1* expression restricted to small intestine where it acts as an RNA editor[104]. In other organisms, such as *Rattus norvegicus* or *M. musculus*, *APOBEC1* expression is much higher and its antiviral activity has been demonstrated[105,106]. In this context, the identified amplification could participate in the enhancement of innate immunity.

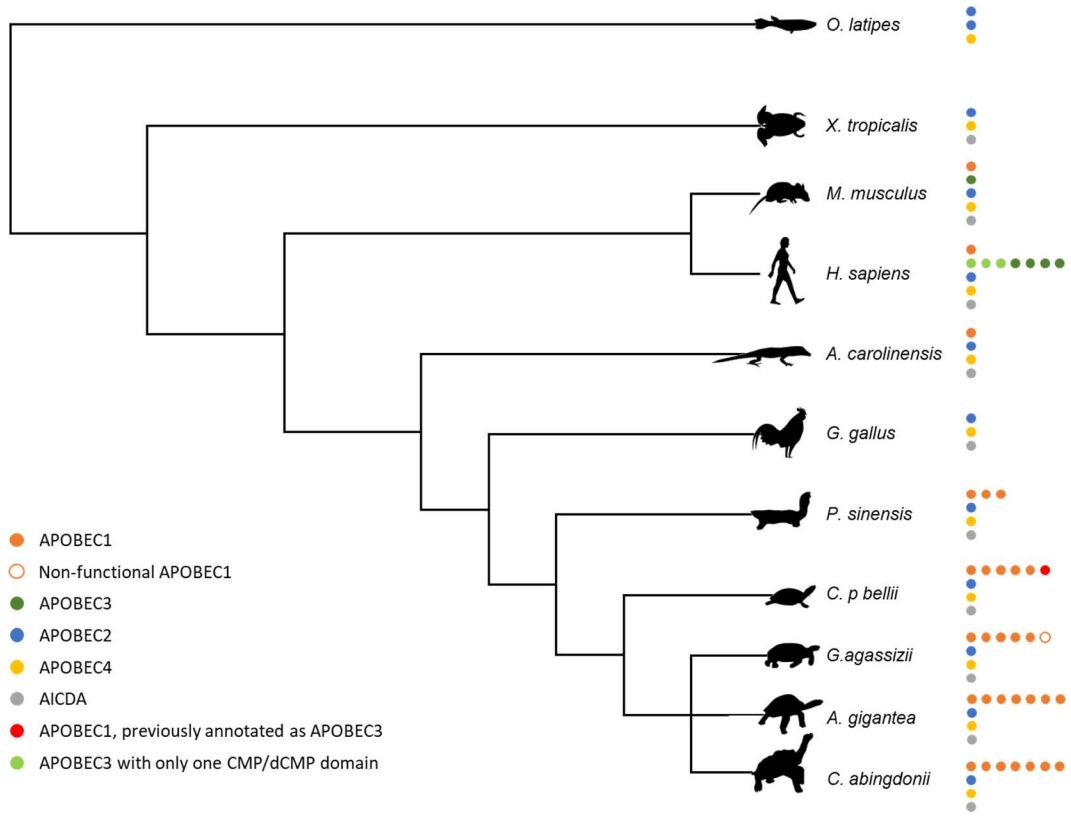


Figure S11. Divergence tree showing APOBEC gene family copy-number variations along different organisms. The phylogenetic tree was generated using the Common Tree Taxonomy tool from NCBI[1].

Another remarkable gene expansion involves the cathelicidin family of antimicrobial peptides (CAMPs). These proteins play important roles in host immune defence against microbial infections, and as anti-inflammatory and immunomodulatory factors[107]. We identified twelve putative copies in *C. abingdonii* and *A. gigantea*, and at least one pseudogene with premature stop codons (**Fig. S12**). Only two copies show aligned RNA-Seq data (**Table S7**), suggesting that the expression of the remaining copies could be conditioned to stress situations[108]. Interestingly, the production of these peptides appears to be associated with healthy ageing, with comparable

levels between healthy elderly and young individuals and lower levels in elder adults with recurrent infections[109]. The expansion of this family may therefore be related to prevention of infections and be an important factor in the longevity of giant tortoises.

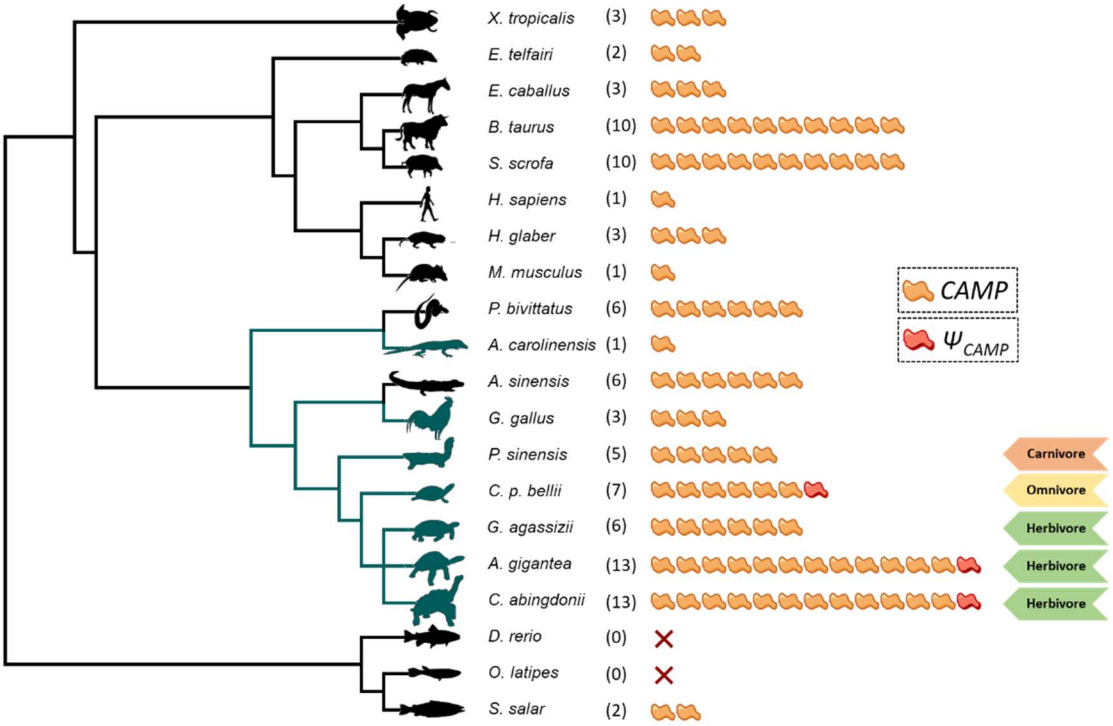


Figure S12. Divergence tree showing *CAMP* copy number in different species and the expansion of this gene in *C. abingdonii*, *A. gigantea* and other organisms. Icons represent *CAMP* copies in each species. Ψ_{CAMP} represents a copy of *CAMP* with premature stop codons. Species emphasized in green indicate that manual annotation has been instrumental in characterizing these duplications. Crosses represent the absence of these peptides in two different species of fishes and the arrows show the type of feeding of each turtle analysed. The phylogenetic tree was generated using the Common Tree Taxonomy tool from NCBI[1].

In addition, Testudines present an expansion, validated by Sanger sequencing, of the *CHIA* gene (**Table S7**; **Table S21**), which encodes a chitinase involved in the degradation of fungal cell walls, among other functions[110]. We found 3 copies of this gene in *C. abingdonii*, although one of them seems to be pseudogenized. Furthermore, we found in *C. abingdonii* and *A. gigantea* a unique set of toll-like receptors (TLRs), which are key elements of the innate immune system. Compared with *C. p. bellii*, these giant tortoises and *G. agassizii* seem to have lost several TLRs (TLR6, TLR9, TLR15 and TLR21). On the other hand, we found expansions in giant tortoises, validated by Sanger sequencing, of TLR2, TLR5, TLR8 and TLR13 (**Fig. S13; Table S21**). Out of these, only TLR2 and TLR13 are also expanded in the genome of *G. agassizii*. Each one of these receptors recognizes different specific elements of bacteria and viruses: TLR2 binds to lipoproteins and other bacterial cell wall components, TLR5 recognises bacterial flagellin, and

TLR8 and TLR13 detect different types of bacterial and viral RNAs[111-113]. These changes might be linked to adaptations to specific pathogenic challenges in the habitat of giant tortoises.

Notably, we documented the absence of most of the interferon (IFN) and interferon-related genes in Sauropsida, including *C. abingdonii* (**Table S9**). However, some genes described as conserved in reptiles, such as IFTM5 and IFTM10, are present[114,115]. Finally, we found that defensins, small proteins that participate in basic defence against bacterial, fungal or viral infections, were expectedly missing in the *C. abingdonii* genome assembly. Their avian counterparts, gallinacins[116], were annotated using the available sequences from *C. mydas, C. p. bellii* and *P. sinensis*.

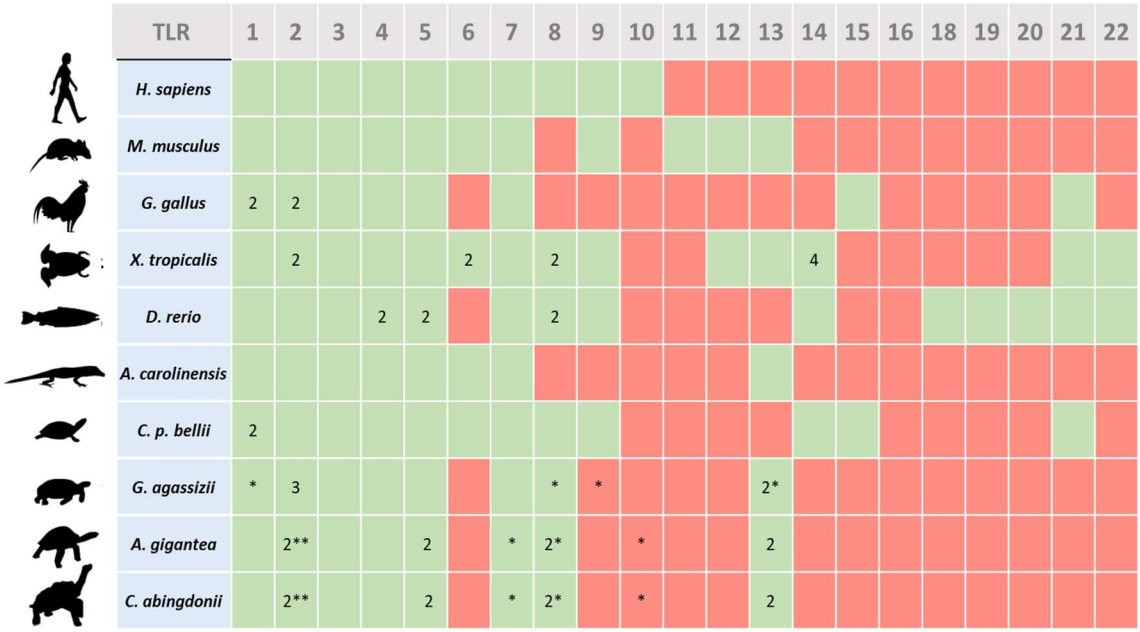


Figure S13. Comparative analysis of TLR gene set in *C. abingdonii, A. gigantea* and different vertebrates. Boxes marked with an asterisk refer to pseudogenes. Green boxes indicate the presence of at least one functional copy, while red boxes refer to absences. If there is more than one functional copy, the number is indicated. Data from other species was obtained from multiple studies[5,117-119].

3.2. Immune regulators of inflammation

Parasitic microorganisms depend on another organism for survival. Thus, they have developed strategies to invade, multiply and propagate within their host while avoiding its defences. Likewise, hosts have acquired different mechanisms to detect and respond rapidly to diverse groups of pathogens, including the activation of an innate immune response and apoptosis. However, this response must be tightly regulated, as an excessive reaction could lead to sepsis, a deleterious systemic hyperinflammation.

In this context, we analysed several members of *NLRP* family, which play a key role in the inflammasome, an innate immune system protein complex that recognizes infectious pathogens and regulates the activation of inflammatory caspases[120]. We found 26 putative *NLRP* genes in

C. abingdonii, although at least one of them seems to be pseudogenized (**Table S21**). While there is a similar number in *G. agassizii* and in *C. p. bellii* (28 genes), there is an amplification compared with *P. sinensis* (18 genes) and *H. sapiens* (14 genes). Although its importance in mammals has been widely described, its role in other vertebrates remains uncharacterized[120,121].

We also examined *CASP12*, a modulator of the activity of inflammatory caspases which is highly conserved throughout evolution and is associated with higher sensitivity to infection and sepsis in humans[122]. Notably, most human populations and *Balaena mysticetus* show different loss-of-function variants[23]. We found that *CASP12* was perfectly functional in all Testudines, without changes in essential residues.

Overall, these results suggest that the innate immune system of *C. abingdonii, A. gigantea*, and other turtles has been modulated to play a predominant role in the immune response, with genomic expansions affecting components involved mainly in this process.

3.3. Major Histocompatibility Complex

The major histocompatibility complex (MHC) is an essential component in both innate and adaptive immune systems which can be functionally divided into class I and II. MHC class I is present in almost all cells, except enucleated cells of the blood myeloid lineage, and presents an extreme variability in populations, conferring individual specificity while providing a mechanism for the immune system to recognize host cells. MHC class II genes are only expressed in antigen-presenting cells, helping them in the recognition of foreign antigens and recruiting immune response components. These protein complexes are encoded by numerous genes that are classified, according to their classical or non-classical function, in 6 MHC-I gene families (A, B, C, E, F and G) and at least 12 MHC-II genes in the human genome[123]. MHC genes are located in a well-conserved cluster, with MHC class-II-related genes grouped closer to the centromere and MHC class-I-related genes closer to the telomere. There is an additional MHC class, known as MHC-III, that is defined by its genomic location between MHC-I and MHC-II[124] . This complex comprises several genes with heterogeneous functions, some of them implicated in the complement cascade or the inflammatory response.

3.3.1. MHC class I and II genes

Due to the difficulty to find reliable sequences of MHC genes in reptiles, we decided to first manually annotate them in *P. sinensis*, *C. p. bellii* and *G. agassizii*, and then perform a comparison with the sequences present in *C. abingdonii*. The available data from Aldabra tortoise did not meet the requirements to include this organism in our comparison. Our annotation showed the presence of 10 MHC class-I-related and 8 MHC class-II-related genes apparently functional in *P. sinensis*, in contrast to the 8 and 10 respectively found in NCBI databases. In addition, we found 20 MHC class-I-related and 12 MHC class-II-related genes in *C. p. bellii* and 17 MHC class-I-related and 19 MHC class-II-related genes in *G. agassizii*, showing an expansion of the family in these turtles. *C. abingdonii* presents 20 MHC class-I-related and 15 MHC class-II-related genes, exhibiting a similar MHC expansion as *C. p. bellii* (**Fig. S14**). We also

annotated CD1D sequences, which, while phylogenetically closer to MHC class I, are similar to class II proteins in terms of protein functions and distribution[125].

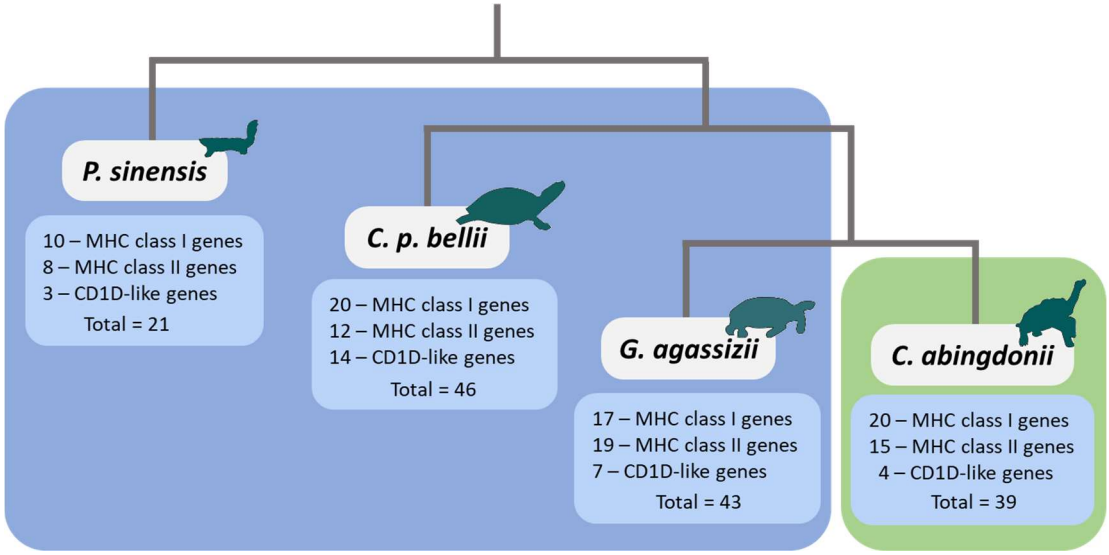
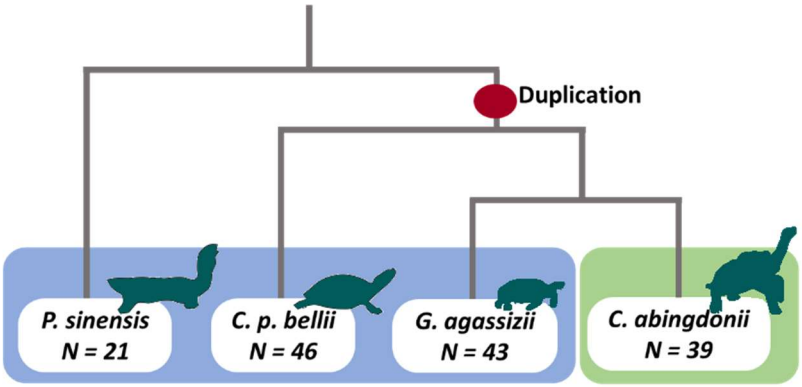


Figure S14. Results of the manual annotation of MHC class I and II genes in *P. sinensis, C. p. bellii, G. agassizii* and *C. abingdonii*. Tree view of the results of the manual annotation that shows a shared expansion of the MHC cluster in the lineage common to *C. p. bellii, G. agassizii* and *C. abingdonii. Green box*, Galapagos giant tortoises; *blue box*, other turtles.

The extreme variability between closely related species and within individuals in a species makes these results difficult to interpret. Thus, we were not able to precisely identify the human orthologs of MHC genes found in *C. abingdonii*, a difficulty experienced by MHC analysis in other reptiles. Studies about the MHC complex in other turtles are focused on allele variability in populations or individuals, without describing any defined gene[126]. This makes the study of MHC variants in *C. abingdonii* difficult. Nevertheless, our results suggest that the observed expansion of the MHC cluster in turtles could have been caused by a duplication event in a common ancestor of *C. p. belli, G. agassizii* and *C. abingdonii*, a conclusion supported by the fact that the ratios between class I and II genes have barely changed between these three lineages (**Fig. S15**). Further investigation of MHC gene patterns and levels of variability could be useful to understand the demographic history of different species of Galapagos giant tortoises and also to better inform conservation management strategies, given the known link between MHC genes with the species abilities to face off environmental changes[127]. This is particularly important as we know that different species have experienced bottlenecks of different size and duration since the colonization of the Galapagos Islands[128].



Class I (%)	47.6	43.5	39.5	51.3
Class II (%)	38.1	26.1	44.2	38.5
CD1d (%)	14.3	30.4	16.3	10.2
$^{\text{Class I}}/_{\text{Class II}}$	1.25	1.67	0.89	1.33

Figure S15. Hypothesis for MHC cluster duplication in a common ancestor of *C. p. bellii, G. agassizii* and *C. abingdonii*. Ratios between class I and II MHC genes appear to be unchanged, supporting the idea of cluster duplication.

3.3.2. MHC class III genes

As mentioned above, the MHC class III region includes a heterogeneous group of genes, some of which play a role in the inflammatory response. Due to this heterogeneity, its definition is uniquely based on their location between MHC class I and MHC class II regions in eutherian mammals[129,130]. In other species, like *G. gallus*, MHC III genes are more variable in their location[131]. To determine how these genes have evolved in turtles, we manually annotated in the genome of *C. abingdonii* the 60 MHC class III genes present in humans and found the presence of 40 of them in this giant tortoise (**Table S10**). These genes are also clustered in this species, which supports the conservation of this cluster through evolution (**Fig. S16**).

Table S10. Status of MHC class III gene set analysed in *C. abingdonii*.

ABHD16A	ATP6V1G2	C6orf48	EGFL8	HSPA1L	LY6G6C	NELFE	PRRT1	TNXB
AGER	BAG6	CFB	EHMT2	LSM2	LY6G6D	NEU1	RNF5	VARS
AGPAT1	C2	CLIC1	FKBPL	LST1	LY6G6F	NFKBIL1	SAPCD1	VWA7
AIF1	C4A	CYP21A2	GPANK1	LTA	MCCD1	NOTCH4	SKIV2L	ZBTB12
APOM	C6orf10	DDAH2	HSPA1A	LY6G5B	MPIG6B	PBX2	SLC44A4	
ATF6B	C6orf47	DXO	HSPA1B	LY6G5C	NCR3	PRRC2A	STK19	

Expansion | Absence | Without relevant variants | Loss-of-function variants | Remarkable missense variants

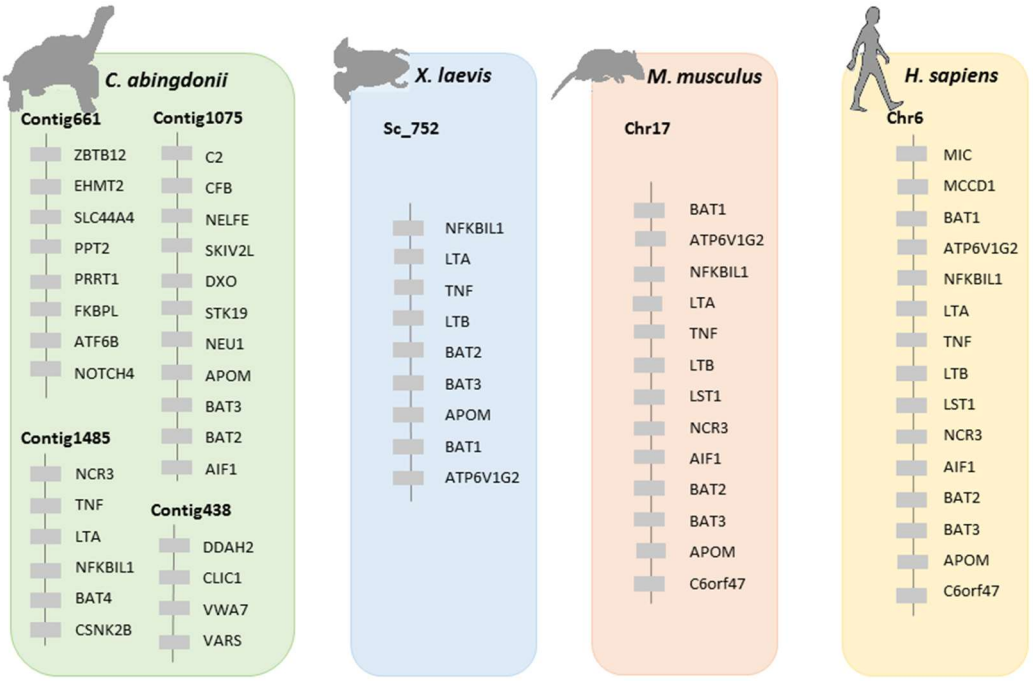


Figure S16. MHC III genes present in *C. abingdonii, X. laevis, M. musculus* and *H. sapiens.* *C. abingdonii* presents 40 MHC III genes, a reduced number compared to the 60 genes present in humans but expanded in relation to the 21 genes in chicken. In this tortoise, these genes appear to be grouped in clusters, similar to some the other species analysed -*H. sapiens, M. musculus* and *X. laevis*-.

4. Metabolism and Diet

Metabolism is essential for life and impacts every aspect of biology. In fact, we are beginning to understand the key roles played by metabolites in both physiological and pathological processes, including cancer and age-associated diseases[132,133]. In this regard, genomic studies of different organisms could broaden our knowledge about how diet and lifestyle influence metabolism. To this end, we manually annotated a set of more than 140 key components of metabolic pathways in the *C. abingdonii's* genome (**Table S11**). As before, all variants of special interest in these metabolic genes were checked in the genome of the Aldabra tortoise.

Table S11. Status of the metabolism gene set analysed in *C. abingdonii*.

ACAD10	CEL	DIO2	IDI1	NDUFA4	REN	SLC2A6
ACAD11	COX10	DIO3	IDI2	NDUFA5	RNASE1	SLC2A7
ACAD8	COX11	DUOX1	ITGA1	NDUFA6	SC5D	SLC2A8
ACAD9	COX15	DUOX2	ITGAV	NDUFA7	SCARB1	SLC2A9
ACADL	COX17	DUOXA1	ITGB3	NDUFA8	SDHC	SLCO1C1
ACADM	CST1	DUOXA2	IYD	NDUFA9	SDHCL	SQLE
ACADS	CST11	EBP	KNG1	NDUFB1	SLC16A10	TG

ACADSB	CST2	EMB	LDHA	NDUFB10	SLC26A4	THRB
ACADVL	CST3	FAM11A	LDHB	NDUFB4	SLC2A1	TM7SF2
ACAT1	CST4	FDFT1	LDHC	NDUFB5	SLC2A10	TNFSF4
ACAT2	CST5	FDPS	LDHD	NDUFB6	SLC2A11	TPO
ADCY1	CST6	FETUB	LRP2	NDUFB7	SLC2A11L	TSHB
AHSG	CST7	GAPDH	LSS	NDUFB8	SLC2A11L4	TSHR
ALDH2	CST8	GOT2	MIF	NDUFB9	SLC2A12	TTF1
ALDH5	CST9	GSK3A	MSMO1	NDUFC2	SLC2A13	TTF2
ASGR1	CST9L	GULO	MVD	NLN	SLC2A14	UQCR11
ASPG	CSTA	HMGCR	MVK	NSDHL	SLC2A15*	UQCRC2
ATP5H	CSTB	HMGCS1	NAPSA	PDIA4	SLC2A16**	
ATP5L	CSTL1	HMGCS2	NDUFA1	PLPP1	SLC2A2	
ATP6V1G3	CTRB1	HRG	NDUFA10	PMVK	SLC2A3	
BACE1	CYP51A1	HSD3B1	NDUFA2	PRKACA	SLC2A4	
CANX	DIO1	HSD3B2	NDUFA3	RCAN2	SLC2A5	

Expansion Absence Without relevant variants Loss-of-function variants Remarkable missense variants

* Gene sharing a common ancestor with *SLC2A3* and *SLC2A14*. ** Gene sharing a common ancestor with *SLC2A1*, *SLC2A2*, *SLC2A3*, *SLC2A4* and *SLC2A14*.

4.1. Glucose metabolism

Glucose is one of the most important carbohydrates due to its pivotal position as a source of energy, as well as a precursor of vitamins and different polymers essential for cells. Among the multiple enzymes involved in these metabolic pathways, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is a key component of the glycolysis pathway. We found three copies of this gene in CheloAbing 1.0, one of them being a retrotranscript (**Fig. S17**). The presence of the two apparently functional *GAPDH* copies – with a 95% of identity between them – was confirmed by RNA-Seq and Sanger sequencing in all analysed tortoises (including the other Galapagos tortoises, the Aldabra tortoise and their continental outgroups) (**Table S7; Table S21**). We also evaluated their selection status by comparing several selection models, as explained in **section 1.2**, with no evidence of a positive selection process in any of them, as suggested by its dN/dS value. On the other hand, the analysis of the retrotranscript suggested that it is not present in any of the other tortoise species tested. Although the putative physiological effect of *GAPDH* duplication has not been described to date, based on its metabolic role it may promote a faster consumption of glucose.

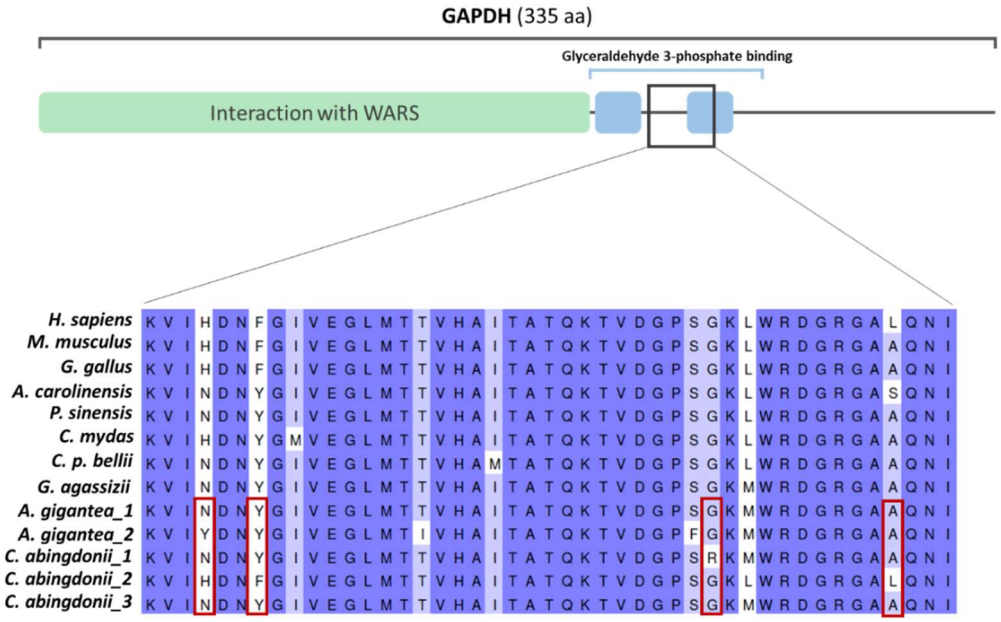


Figure S17. Partial amino acid alignment of the GAPDH sequence in *C. abingdonii, A. gigantea*, and other vertebrate species, including four turtle species. The copy-specific variants used to validate this amplification in *C. abingdonii* and *A. gigantea* are emphasized with a red rectangle.

Similarly, the mitochondrial metalloprotease neurolysin (NLN) is truncated due to a premature stop codon (W50*, present in all tested species of Galapagos giant tortoises and their continental outgroups, but not present in Aldabra tortoise). This alteration was validated through PCR amplification and Sanger sequencing (**Table S7**). The genome of *G. agassizii* shows a different premature stop codon, which suggests a case of convergent evolution. Interestingly, *NLN knock-out* mice have greater glucose tolerance, insulin sensitivity and an upregulation of several mRNAs involved in hepatic gluconeogenesis pathways[134]. *C. abingdonii* also displayed a duplication of *CTRB1*, a pancreatic serine protease involved in digestion (**Table S21**). This gene is also duplicated in *A. gigantea*, while it is not duplicated in other Sauria (**Table S7**). Along with the benefits to digestion itself, this could also acts as an extra defence against glucose intolerance, since mutations in this gene are related to this trait[135].

It is well known that glucose levels have to be carefully maintained in the organism in order to keep homeostasis, which is mainly achieved by the action of various hormones[136]. We found in the *C. abingdonii* genome that the gene encoding glycogen synthase kinase 3 alpha (*GSK3A*) presents a missense variation that affects a conserved residue (p.R272Q) in the activation loop of its protein. This variation was also validated by Sanger sequencing, and found in all Galapagos giant tortoises tested, the Aldabra tortoise, their continental outgroups, *G. agassizii* and *C. p. bellii* (**Table S7**). GSK3A acts as a negative regulator in the hormonal control of glucose homeostasis, Wnt signalling and regulation of transcription factors and microtubules[137]. Thus, the observed variation may alter its function, causing differences in the regulation of glucose levels in blood. Moreover, the loss of function of *ITGA1* (see **section 7.6**) has been correlated

with severe hepatic insulin resistance in a null mouse model when fed a high-fat diet[138]. Thus, we found evidence for selective pressure acting on the regulation of glucose metabolism in giant tortoises and related turtles. Our results point to a putative deregulation of glucose levels in the blood of these animals, which might be compensated with a greater tolerance to and faster consumption of glucose.

Furthermore, the pro-inflammatory cytokine MIF (macrophage migration inhibitory factor) presents a p.N111C variant in *C. abingdonii*, *A. gigantea*, as well as in the other species of Galapagos giant tortoises, their outgroups and *G. agassizii*, as confirmed through Sanger sequencing (**Table S7**). Surprisingly, previous functional analyses of this variant have shown that it leads to the inhibition of intracellular signalling cascades[139]. MIF deficiency has been shown to hinder the development of insulin resistance, glucose intolerance and associated atherosclerotic disease[140].

Altogether, these results suggest that tortoises present an altered glucose metabolism compared to other turtles (**Fig. S18**).

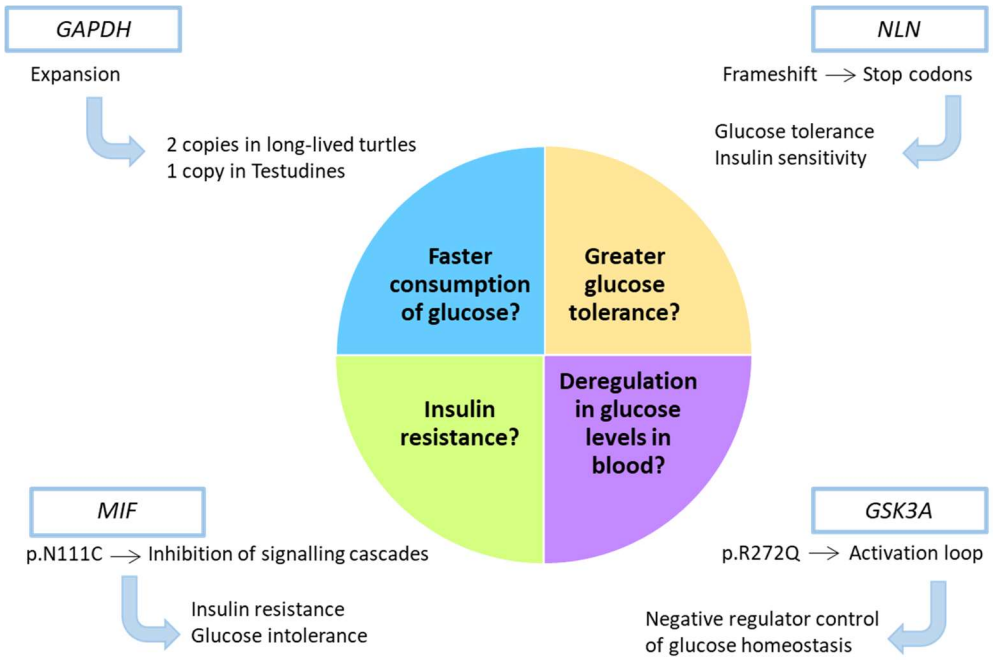


Figure S18. Potential glucose metabolism alterations in giant tortoises. The putative alterations affecting glucose metabolism may be associated with four different effects. The extra copies of *GAPDH* may be linked to a faster consumption of glucose. *NLN* loss of function could point to a greater glucose tolerance in Galapagos tortoises and their outgroups, whereas the mutation in the activation loop of *GSK3A* may contribute to a deregulation of glucose levels in blood in these organisms. Finally, the mutation in *MIF*, present in all analysed tortoises, may be related to insulin resistance.

4.2. Diet

The source of food and its relationship with the environment that a species inhabits is a driver of evolution and has been shown to modulate, among others, body size through genetic changes that include copy-number variation and point mutations[141]. In this regard, vitamins are essential in the maintenance of systemic homeostasis and normally obtained from the diet, but in some species, they can be also synthesized by the organism itself. In this context, vitamin C is unique in that some vertebrates have lost the ability to synthesize it due to loss of the *GULO* gene[142]. Gulonolactone (L-) oxidase, the enzyme encoded by *GULO*, is indispensable for the final step in vitamin C biosynthesis, and the absence of this protein alone is sufficient to disrupt this process[143]. Primates, guinea pigs, bats, some birds and teleost fishes do not have a functional *GULO* gene[142]. However, *GULO* is present in *C. abingdonii*, *A. gigantea* and in other Testudines, thereby indicating the capacity to synthesize vitamin C in this clade. This feature may be an important factor for protecting cells and tissues against oxidative damage. In addition, we identified a functional copy of embigin (*EMB*) in the genome of *C. abingdonii*. None of the variants in this gene known to be associated with carnivorism[144] were present in *C. abingdonii* or in *A. gigantea*, consistent with the well-known fact that giant tortoises are herbivores. In this regard, and as discussed above (see **section 3.1**), we have found an expansion in cathelicidin-related genes in *C. abingdonii* and *A. gigantea* (**Fig. S12**), which is also present in other species[145-147], and could be related with an herbivore diet[148].

4.3. Thyroid metabolism

Iodine is an essential nutrient due to its role as a component of thyroid hormones that must be obtained through the diet. Thyroid hormones are involved in a variety of aspects in human health, including relevant physiological processes, such as neurologic development or ageing[149]. To evaluate the genetic status of *C. abingdonii*'s thyroid hormones, we studied the main genes involved in thyroid metabolism. Our results showed that the TSH receptor (TSHR), which plays a pivotal role in regulating thyroid gland physiology, contains a variant that affects a conserved residue (p.K621R) in its G-protein recognition intracellular loop. This variant was validated by Sanger sequencing and found in all the other Galapagos giant tortoise species tested, the Aldabra tortoise and their continental outgroups (**Table S7**). However, this variant is not present in the genome of *G. agassizii*. Mutations in this and nearby residues are associated with several problems in the TSH response along with T3 and T4 hormone synthesis defects[150]. Consistently, an association between reduced thyroid signalling and extended longevity has been reported, as hypothyroidism is overrepresented among supercentenarians[151,152]. Therefore, it is tempting to speculate that this TSHR variant could contribute to the extreme longevity found among these tortoises. However, the absence of this variant in the long-lived *G. agassizii* suggests that other mechanisms are involved in this trait.

5. Environmental adaptations and stress response

The appropriate response to stress conditions is essential to maintain cellular homeostasis and ensure the survival of the organism. Reptiles, and specifically Testudines, have experienced several adaptations to survive oxygen limitation as many turtles spend their winters under anoxic and cold conditions[153]. To study these environmental interactions in *C. abingdonii*, we annotated more than 250 genes involved in oxygen homeostasis, stress response, and DNA repair (**Table S12**). Our analysis identified several variants in *C. abingdonii*, many of them shared with other Testudines, in stress-responsive and anaerobic metabolism proteins.

Table S12. Status of stress-response and DNA repair genes analysed in *C. abingdonii*.

ADGB	CYGB	FAAP24	GPX4	HSPA5	MSH5	RAD23A	TDG
ALKBH2	DCLRE1A	FABCG	GPX5	HSPA6	MUS81	RAD23B	TDP1
ALKBH3	DCLRE1B	FAN1	GPX6	HSPA8	MUTYH	RAD50	TDP2
APEX1	DCLRE1C	FANCA	GPX7	HSPA9	NEIL1	RAD51	TOPBP1
APEX2	DDB1	FANCB	GPX8	HSPB1	NEIL2	RAD51B	TP53
APLF	DDB2	FANCC	GTF2H1	HSPB2	NEIL3	RAD51C	TP53BP1
APOLD1	DMC1	FANCD2	GTF2H2	HSPB3	NGB	RAD51D	TRAP1
APTX	DUT	FANCE	GTF2H3	HSPB6	NHEJ1	RAD52	TREX1
ATM	EIF2AK1	FANCF	GTF2H4	HSPB8	NTHL1	RAD54B	TREX2
ATR	EIF2AK2	FANCI	GTF2H5	HSPB9	NUDT1	RAD54L	TTDN1
BAD	EIF2AK3	FANCL	H2AFX	HSPD1	OBFC2B	RAD9A	UBE2B
BAK1	EIF2AK4	FANCM	HELQ	HSPH1	ODF1	RBBP8	UBE2N
BCL2A1	EIF2S1	FEN1	HLTF	HUS1	OGG1	RDM1	UBE2V2
BCL2L1	EIF2S2	FOS	HP	HYOU1	PARP1	RECQL	UNG
BCL2L10**	EIF2S3	FOSB	HSBP11	JUN	PARP2	RECQL4	UVE2A
BCL2L11	EME1	FOSL1	HSF1	JUNB	PARP3	RECQL5	UVSSA
BCL2L12	EME2	FOSL2	HSF2	JUND	PCNA	REV1L	WRN
BCL2L13	ENDOV	FOXO1	HSF3	KLKB1	PER1	REV3L	XAB2
BCL2L14	ENOX1	FOXO3	HSF4	LIG1	PER2	RIF1	XPA
BCL2L15	ERCC1	FOXO4	HSF5	LIG3	PLAT	RMI2	XPC
BCL2L2	ERCC2	FOXO6	HSP90AA1	LIG4	PLAU	RNF168	XRCC1
BLM	ERCC3	GADD45A	HSP90AB1	MAD2L2	PLG	RNF4	XRCC2
BNIP2	ERCC4	GADD45B	HSP90B1	MB	PMS1	RNF8	XRCC3
BOK	ERCC5	GADD45G	HSPA12A	MBD4	PMS2	RPA1	XRCC4
BTBD12	ERCC6	GBE	HSPA12B	MDC1	PNKP	RPA2	XRCC5
CCNH	ERCC8	GBX	HSPA13	MGMT	PRKDC	RPA3	XRCC6
CETN2	ERN1	GBY	HSPA14	MLH3	PROC	RPA4	ZFAND2A
CHAF1A	ERN2	GEN1	HSPA1A	MMS19	PRPF19	RRM2B	ZFAND2B
CHEK1	EXO1	GIYD1	HSPA1B	MNAT1	PTGS1	SETMAR	αHB (3)***
CLK2	F10	GIYD2	HSPA1L	MPG	PTGS2	SHFM1	βHB (3)***
CRYAA	F11	GPX1	HSPA2	MSH2	RAD1	SHPRH	
CRYAA*	F7	GPX2	HSPA4	MSH3	RAD17	SMUG1	
CRYAB	FAAP20	GPX3	HSPA4L	MSH4	RAD18	SPO11	

Expansion | Absence | Without relevant variants | Loss-of-function variants | Remarkable missense variants

Boxes marked with * represent duplications where both copies have exactly the same exonic and intronic sequence, so the duplication is probably due to an assembly error. Boxes marked with ** represent genes with very low identity to their human orthologs. Boxes marked with *** represent genes belonging to a specific gene family, but whose orthology relationship is unclear.

5.1. Oxygen homeostasis

Globins play an essential role in oxygen binding and transport. Furthermore, these haem-containing proteins are also involved in nitric oxide metabolism, detoxification of reactive oxygen species (ROS), and intracellular signaling[154].

The family of globins has been extensively studied and constitutes a classical model to unravel the divergence and evolution of genes and proteins[155]. Five types of globins may be present in most jawed vertebrates (*Gnathostomata*): haemoglobin (Hb), myoglobin (Mb), cytoglobin (Cygb), neuroglobin (Ngb) and androglobin (Adgb)[154]. However, three additional globins with still poorly defined functions have been recently discovered. These proteins, globin E (GbE), globin Y (GbY), and globin X (GbX) have a more restricted taxonomic distribution and have been lost in many vertebrate taxa[156]. Ancestral *Gnathostomata* had the complete repertoire of globins with distinct functions, but nowadays only coelacanths and turtles show all eight different types of globins[155] (**Fig. S19**). Using human, *C. p. bellii*, and *P. sinensis* sequences as references, we annotated the full set of globins in *C. abingdonii* and *A. gigantea*, including three copies of α-Hb and β-Hb genes. Multiple α-like and β-like globin genes have been recently derived from a common ancestor gene. In both types of globins, the observed gene amplifications seem to be lineage-specific, and therefore there is not a clear orthology relationship among them.

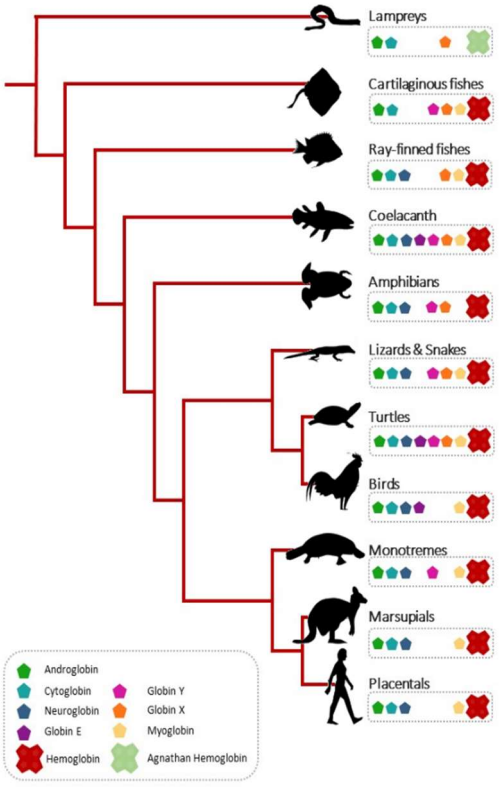


Figure S19. Set of globins present in different vertebrates. Note that only turtles and the coelacanth have the complete repertoire of globins. The phylogenetic tree was generated using Common Tree Taxonomy tool from NCBI[1].

Recently, it has been discovered that GbX, which was first designated as a neuronal globin, is actually present in fish red blood cells, suggesting a role in oxygen binding and transport similar to Hb[157]. Hypoxia-resistant fishes and Testudines exhibit Hbs with faster nitrite reductase activity[158,159], indicating that nitric oxide metabolites may contribute to the physiological response to anoxia/hypoxia periods[158]. It has been also reported that deoxygenated GbX is involved in nitrite reduction to nitric oxide, faster than human Hb. This suggests that GbX may participate in the anoxia/hypoxia response mechanism, releasing oxygen at very low oxygen pressures when Hb is already deoxygenated[157]. Accordingly, the presence of this globin in *C. abingdonii* and *A. gigantea*, as well as in other Testudines, may be one of the mechanisms underlying the exceptional tolerance to hypoxic conditions of these vertebrates[160].

5.2. Coagulation

We next analysed different genes involved in the coagulation pathway, as both coagulation and blood homeostasis can be greatly impacted by environmental changes and species adaptation to new habitats[23]. We found interesting variations in some members of the coagulation factor family, such as factor VII, factor X, and factor XI. These variants, common to all species of Galapagos giant tortoises tested, *A. gigantea* and *G. agassizii*, lead to putative F7 (p.F64L), F10 (p.E91K), and F11 (p.E315K) deficiencies (**Table S7**). This could result in altered functions of these proteases, since mutations in these residues are associated with pathologies such as factor VII, X, and XI deficiency[161,162].

Additionally, *KLKB1* – encoding a serine protease that participates in the kinin-kallikrein system[163] – was absent both in *C. abingdonii* and in *P. sinensis*. This loss could affect blood pressure and coagulation control mechanisms, through an increased serum level of vasoactive peptides[164]. Likewise, plasminogen (*PLG*) displays point variants in fibrin-binding sites (p.R134L, also present in *A. gigantea* and *P. sinensis*, and p.R136K, found in all studied turtles) that may alter the protein function, despite not being specific of these species. On the other hand, the serine protease PROC that inhibits the generation of plasmin, was absent only in *C. abingdonii*. The absence of this gene may contribute to compensate other alterations in the coagulation system in turtles.

5.3. Transcription factor TP53

The p53 transcription factor is a conserved DNA damage response protein with different roles in regulating cell cycle, apoptosis and energy metabolism, and plays a master role in tumour suppression[165]. As in all other Reptilia genomic studies so far[6,69], we found one functional copy of this transcription factor in *C. abingdonii* and *A. gigantea*, which is likely essential for proper DNA maintenance and cancer protection. However, *C. abingdonii* and *A. gigantea* share with other turtles a modification at codon 106 (p.S106E) found also in root voles (*Microtus oeconomus*) (**Fig. S20**; **Table S7**). This variant was first described as an adaptation to hypoxia and cold, given that this vole species lives on the Tibet plateau, and it is also shared by four fish species (*Barbus barbus, Tetraodon miurus, Platichthys flesus and Xiphophorus hellerii*) and a squid species (*Loligo forbesi*), which are also hypoxia-tolerant. Collectively, these results suggest a convergent evolutionary process. Specifically, p.S106E hinders p53 from activating the apoptotic genes *IGFBP3, APAF1, BAX, NOXA, PUMA* and *HDM2*, and therefore suppresses cell apoptosis by p53 under stress conditions, without affecting cell-cycle genes[166]. In addition, giant tortoises present the ancestral p.R72P variant in p53, which has been previously associated with longevity and better cancer prognosis in human populations[167].

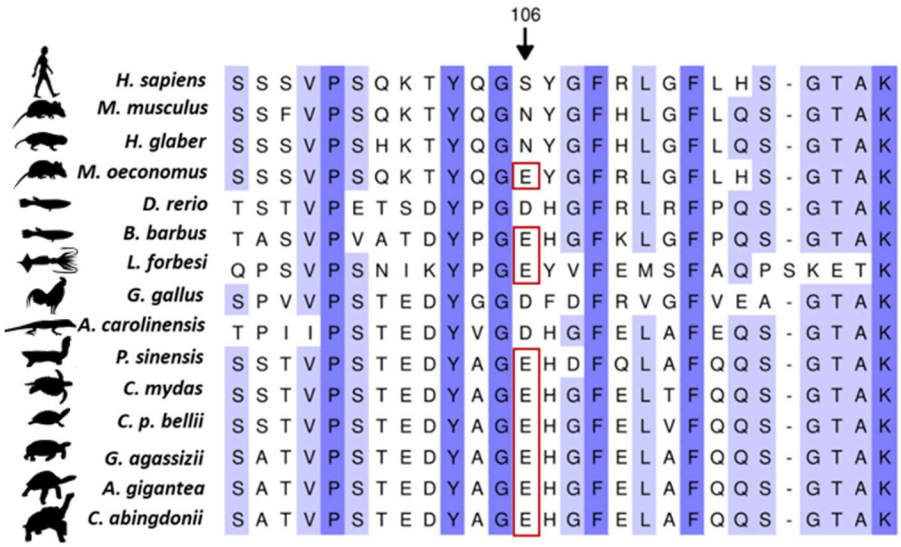


Figure S20. Partial amino acid alignment of the TP53 sequence in *C. abingdonii, A. gigantea* and other vertebrate species, including four turtles. Codon 106 changes from serine (S) to glutamic acid (E) are highlighted with red boxes.

5.4. Cell-stress response proteins

In addition to the involvement of globins and p53 in anoxia/hypoxia resistance, other cellular proteins, such as heat shock proteins (HSPs), play essential roles in maintaining the homeostasis because of their implication in cellular response processes. HSPs are a group of chaperones that act in response to different types of cellular stress, such as elevated temperature, heavy metals, proteasome inhibitors, and hypoxic/anoxic conditions[153,168]. *HSP* expression is regulated at the transcriptional level by heat shock factors (HSFs), being HSF1 the main mediator in stressful events. Previous studies carried out in Testudines showed that HSF1 and other HSPs play an important role in the anoxia response mechanism, in addition to other cellular processes such as antioxidant defences and regulation of apoptosis[153]. In our study, we identified homologs in *C. abingdonii* for different members of *HSPA* and *HSPB* families (**Table S13**). We also found the complete *HSF* repertoire, including *HSF3*, which has been pseudogenized in primates. The presence of the complete set of *HSF* genes in all Testudines studied so far suggests the existence in these animals of fine-tuning regulatory mechanisms, acting in the cellular responses to physiological and environmental insults.

Table S13. Members of HSPB, HSP60, HSPA and HSP90 families identified in the genomic sequence of *C. abingdonii*.

	Approved symbol	Approved name	Synonyms
Small heat shock proteins (HSPB)	**HSPB1**	heat shock protein family B (small) member 1	HSP27, HSP28, Hs.76067, Hsp25, CMT2F
	HSPB2	heat shock protein family B (small) member 2	Hs.78846, MKBP
	HSPB3	heat shock protein family B (small) member 3	HSPL27
	CRYAA	crystallin alpha A	CRYA1, HSPB4
	CRYAB	crystallin alpha B	CRYA2, HSPB5
	HSPB6	heat shock protein family B (small) member 6	FLJ32389, Hsp20, PPP1R91
	HSPB7	heat shock protein family B (small) member 7	cvHSP

	HSPB8	heat shock protein family B (small) member 8	H11, E2IG1, HSP22, HspB8, CMT2L
	ODF1	outer dense fiber of sperm tails 1	HSPB10, ODF27, ODFPG, CT133, RT7
	HSPB11	heat shock protein family B (small) member 11	C1orf41, HSPCO34, PP25, IFT25
Heat shock 60kDa proteins (HSP60)	**HSPD1**	heat shock protein family D (Hsp60) member 1	GroEL, HSP60
Heat shock 70kDa proteins (HSPA)	**HSPA1A**	heat shock protein family A (Hsp70) member 1A	HSP70-1, HSPA1
	HSPA1B	heat shock protein family A (Hsp70) member 1B	HSP70-2
	HSPA2	heat shock protein family A (Hsp70) member 2	
	HSPA4	heat shock protein family A (Hsp70) member 4	HS24/P52, HSPH2
	HSPA4L	heat shock protein family A (Hsp70) member 4 like	APG-1, Osp94, HSPH3
	HSPA5	heat shock protein family A (Hsp70) member 5	BiP, GRP78
	HSPA8	heat shock protein family A (Hsp70) member 6	HSC71, HSC70, HSP73, HSPA10
	HSPA9	heat shock protein family A (Hsp70) member 9	GRP75, HSPA9, BPBP74, mot-2, mthsp75
	HSPA12A	heat shock protein family A (Hsp70) member 12A	FLJ13874, KIAA0417
	HSPA12B	heat shock protein family A (Hsp70) member 12B	C20orf60, dJ1009E24.2
	HSPA13	heat shock protein family A (Hsp70) member 13	STCH
	HSPA14	heat shock protein family A (Hsp70) member 14	HSP70-4, HSP70L1
	HSPH1	heat shock protein family H (Hsp110) member 1	HSP105B, KIAA0201, HSP105A, NY-CO-25
	HYOU1	hypoxia up-regulated 1	ORP150, HSP12A, Grp170
Heat shock 90kDa proteins (HSP90)	**HSP90AA1**	heat shock protein 90 alpha family class A member 1	Hsp89, Hsp90, HSP90N, HSPC1, HSPCA, FLJ31884
	HSP90AB1	heat shock protein 90 alpha family class B member 1	HSPC2, HSPCB
	HSP90B1	heat shock protein 90 beta family member 1	GP96, GRP94, TRA1
	TRAP-1	TNF receptor associated protein 1	HSP75, HSP90L

5.5. DNA-damage response

Maintaining the integrity of the genome during environmental stress is important for survival, as the accumulation of DNA damage increases the risk of cancer and is linked to ageing[26]. A proper DNA repair response to ionizing ultraviolet A light (UVA), reactive oxygen species (ROS) or other mutagenic agents is critical for this purpose. Accordingly, we performed a deep analysis of genes involved in different steps of DNA repair, such as base excision repair (BER), mismatch excision repair (MER), nucleotide excision repair (NER) and homologous recombination (HR). Overall, these genes were highly conserved, which highlights the importance of DNA repair and genome integrity for proper organismal homeostasis. Nevertheless, potentially relevant variants and copy number variations were found in some of the analysed genes, suggesting different expression, regulation and interactions of the DNA repair factors in *C. abingdonii* and other turtles (**Fig. 3a, main text**).

We found one example of copy-number variation in *NEIL1* (**Table S21**), a DNA glycosylase involved in the removal of oxidatively damaged DNA bases[169]. Analysis of RNA-Seq data showed that both copies are expressed, and their dN/dS value suggested that there is no evidence of a positive selection process in either of them. This gene was found duplicated in *C. p. bellii, G. agassizii, C. abingdonii* and *A. gigantea*, as well as in all the other Galapagos giant tortoises

tested and in their corresponding continental outgroups analysed by Sanger sequencing (**Table S7 and Fig. S21**). Previous studies have associated an elevated expression of *NEIL1* with extended longevity in naked mole rat and humans compared to mice[170]. Furthermore, downregulation of this gene has been described in different types of cancer[171]. This result points to *NEIL1* as a candidate gene for longevity extension and cancer protection through genomic stability control.

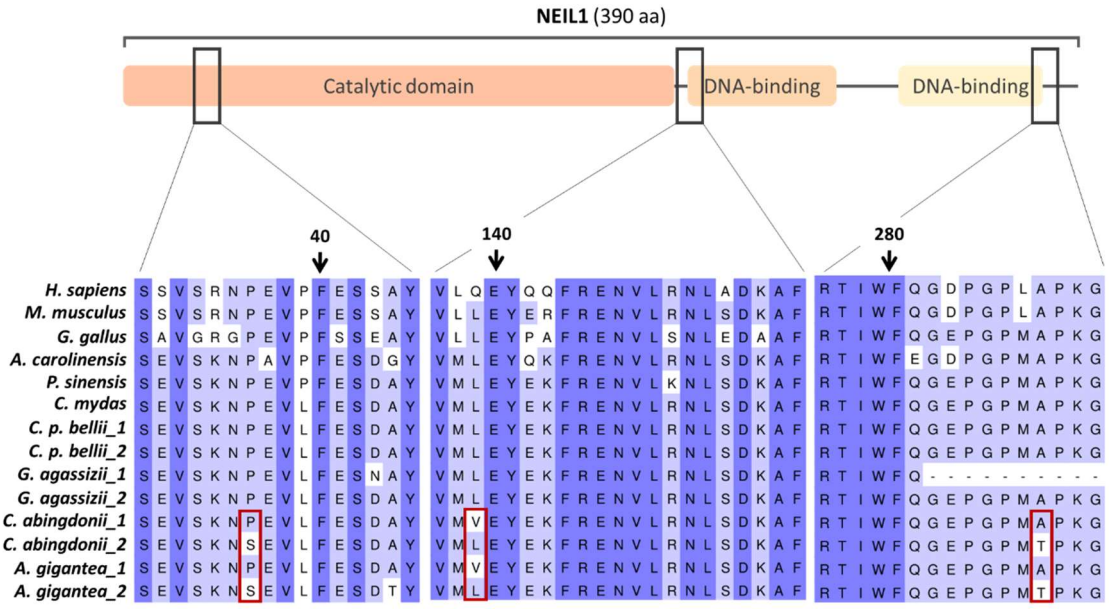


Figure S21. Partial amino acid alignment of NEIL1 sequence in *C. abingdonii, A. gigantea*, and other vertebrate species, including four turtles. Copy-specific variants present in *C. abingdonii* and *A. gigantea* are emphasized with a red rectangle.

We found another duplication affecting *RMI2* (**Table S21**), which plays an essential role in the processing of homologous recombination intermediates to limit DNA crossover formation in cells[172]. Both copies seem to be conserved and expressed in *G. agassizii*, *C. abingdonii* and *A. gigantea*, although the latter may have additional similar pseudogenes. Other Sauropsida such as *C. p. bellii*, *P. sinensis*, *A. carolinensis*, and *G. gallus*, only have one functional copy, although a second pseudogenized copy is present in some of these species.

In a preliminary attempt to explore the way these duplications in *NEIL1* and *RMI2* could modulate the repair response, we cloned the corresponding cDNAs on a pCDH vector and infected HEK-293T cells with them. Clones from each condition -pCDH, pCDH-NEIL1, pCDH-RMI2 and pCDH-NEIL1+pCDH-RMI2- were isolated and cells were exposed to sub-lethal doses of UV light stress (20 J/m^2) or H_2O_2 treatment (500 µM). After 24 or 48 h, we studied the activation of DNA damage response pathways by Western blot analysis, targeting the phosphorylation of the histone H2AX and the cleavage of poly (ADP-ribose) polymerase (PARP). We observed a reduction of both damage-associated forms in all different conditions (**Fig. S22 and Figs. S32-S33**). This suggests that overexpression of these factors may lead to an improved response to stress and a quicker repairing capacity. Therefore, duplications in *NEIL1* and *RMI2* could provide

giant tortoises with an enhanced DNA-repair response, contributing to their expanded lifespan and protection against cancer.

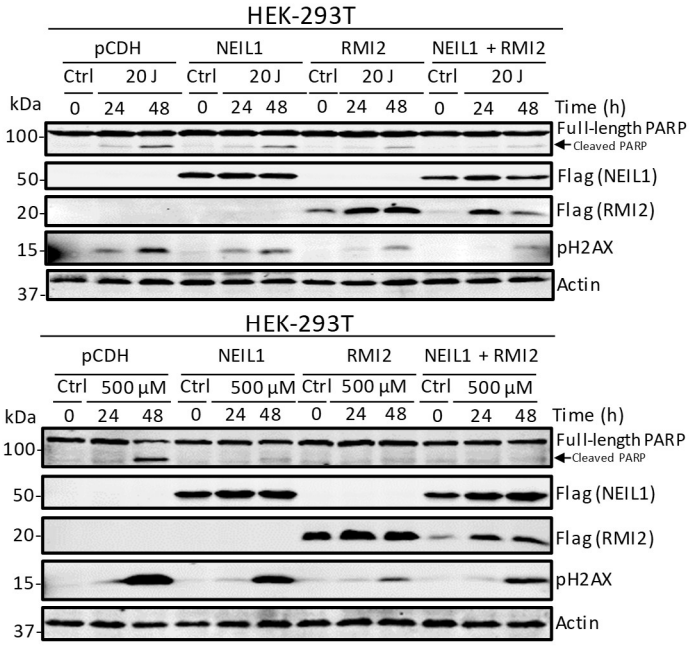


Figure S22. Effect of *NEIL1* and *RMI2* duplication in HEK-293T cells. Western blot analysis of the phosphorylation of histone H2AX at serine 139, and cleaved PARP at 24 and 48 hours after UV light exposure (20 J/m^2) (upper panel) or H_2O_2 treatment (500 µM) (lower panel). For the samples from UV treatment, note that Flag (RMI2) was detected on the same samples used for the remaining Western-blots shown in this panel, run in parallel on an identical blot. Similarly, for the samples from H_2O_2 treatment, the Western-blots shown were carried out with the same samples run in parallel in three identical blots (one for PARP and actin; a second blot for Flag (NEIL1 and RMI2); and a third blot for pH2AX).

In addition to these duplications, some potentially relevant point mutations were found in genes involved in DNA repair. For example, we found one variant affecting protein kinase ATM, which activates checkpoint signalling upon double strand breaks, apoptosis and genotoxic stresses, thereby acting as a DNA damage sensor. As discussed above (see **section 2.3**), one of the nine residues involved in ATM catalytic loop formation[62,173] is mutated in *C. abingdonii* (p.V2873I). This mutation was also found in all the analysed Galapagos giant tortoises and in fishes but was absent in Aldabra tortoise and in their continental relatives (as confirmed through Sanger sequencing; **Table S7**).

Other alterations were detected in genes involved in non-homologous end-joining (NHEJ), an important pathway that repairs double-strand breaks in DNA. In this context, XRCC6 (Ku70) and XRCC5 (Ku80), two ssDNA-dependent ATP-dependent helicases involved in this process[174,175], present changes affecting sumoylation sites in *C. abingdonii* (p.K556R and p.K568N, respectively)[176]. The variant affecting *XRCC6* is present in *C. abingdonii, A. gigantea* and other

Galapagos giant tortoises used for Sanger sequencing validation. The variant affecting *XRCC5* was also found in all the Galapagos giant tortoises species tested, Aldabra tortoise, and in other turtles, such as *G. agassizii, C. p. belli*, and *P. sinensis*. Interestingly, the naked mole rat, the longest-lived rodent, also presented a similar point mutation in XRCC6 (p.K556N)[177] (**Fig. 3b, main text**). Sumoylation is generally induced after DNA damage and plays a key role in DNA repair response and in multiple regulatory processes by enabling specific protein interactions[178]. Therefore, changes in this process may have an important effect in the repair system.

Collectively, the results described in this section point to a proper or even enhanced stress response in *C. abingdonii* and *A. gigantea* with expansion of DNA damage response genes. Minor changes in some of them could suggest small differences in the regulation, interaction and activation of repair proteins, especially those involved in non-homologous end-joining. In addition, duplications in some DNA repair genes could enhance DNA integrity maintenance, contributing to the extended longevity and cancer resistance of *C. abingdonii* and related turtles. The analysis of globins and p53 showed specific adaptations to hypoxia shared with other Testudines. As *C. abingdonii* and *A. gigantea* did not experience the pressure of living in an anoxic or cold environment, these variants are probably relics of the environment where their ancestors thrived.

6. Cancer susceptibility

Cancer and ageing can be considered as two different manifestations of the same underlying process[26], since extended lifespan means that cells have more time to accumulate mutations, increasing the probability of carcinogenesis[179]. Peto's paradox suggests that large, long-lived animals, such as giant tortoises, have evolved mechanisms to suppress cancer more efficiently than other animals. Trying to explore these questions, we selected genes involved in cancer whose alterations had been associated with oncogenesis[180], either as oncogenes or as tumour suppressors. Approximately 420 genes were annotated and analysed (**Table S14**). Overall, compared to other organisms, most of these genes had a well-conserved amino acid sequence. Nevertheless, various changes in several genes could point out mechanisms protecting giant tortoises from cell transformation and cancer (**Fig. S23**).

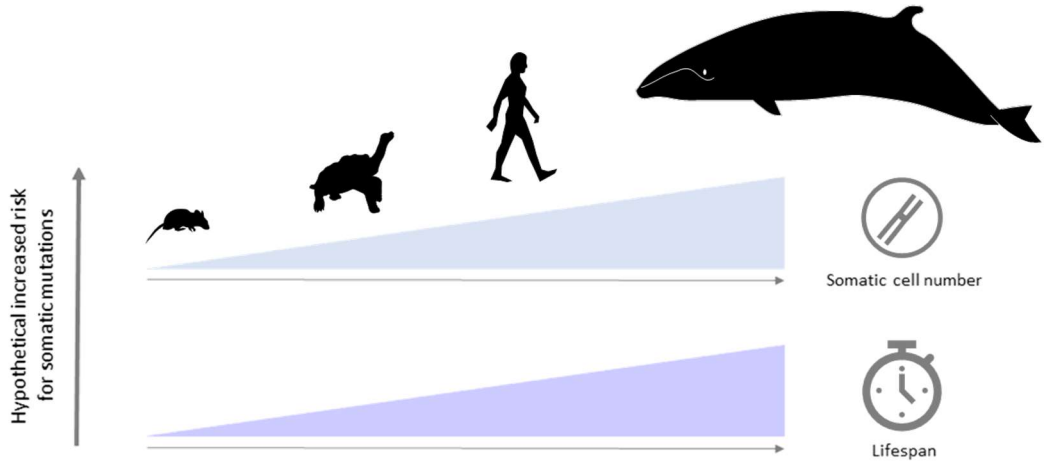


Figure S23. Diagram summarizing Peto's paradox. This theory suggests that large, long-lived animals have developed several mechanisms to suppress cancer better than other animals.

Table S14. Status of cancer-gene set analysed in *C. abingdonii*.

ABL2	CDH11	ETNK1	IL2	MYH11	PLAG1	SRGAP3
ACKR3	CDH2	ETV1	IL21R	MYH9	PML	SRSF2
ACSL3	CDH3	ETV4	ITGA1	MYO5A	POU5F1	SRSF3
ACSL6	CDK12	ETV5	JAZF1	MYO5AB	PPFIBP1	SS18
ACVR1	CDK4	ETV6	KAT6A	MYO5B	PPP2R1A	SS18L1
AFF1	CDKN2A	EWSR1	KAT6B	NAB1	PPP2R1B	SSX1
AFF3	CDKN2B	EXT1	KCNJ5	NAB2	PPP6C	SSX2
AFF4	CDKN2C	EXT2	KDM5A	NACA	PRCC	SSX4
AKAP9	CDKN2D	EZH2	KDM5C	NCKIPSD	PRDM1	STAG2
AKT1	CEP89	FAM131B	KDM6A	NCOA1	PRDM16	STAT3
AKT2	CHCH7	FAM46C	KDR	NCOA2	PRDM1L	STAT5B
AKT3	CHIC2	FBXO11	KDSR	NCOA4	PRF1	STAT6
ALK	CHN1	FBXW7	KIAA1549	NDRG1	PRF1G	STK3
AMER1	CIC	FCRL4	KIF5B	NDRG2	PRF1I	STK4
APC	CLIP1	FEV	KLF4	NDRG3	PRF1K	STRN
APC2	CLIP2	FGCR2B	KLF6	NDRG4	PSIP1	SUFU
ARHGAP26	CLIP3	FGFR1OP	KMT2A	NF1	PTCH1	SUZ12
ARHGEF12	CLIP4	FGFR2	KMT2C	NF2	PTPN11	TAF15
ARID1A	CLP1	FGFR3	KMT2D	NF2L	PTPRB	TAL1
ARID1B	CLTC	FGFR4	KRAS	NFIB	PTPRK	TAZ
ARID2	CLTCL1	FH	KTN1	NFKBIE	PWWP2A	TCEA1
ARNT	CNBP	FIP1L1	LASP1	NIN	RABEP1	TCF3
ASPSCR1	CNOT3	FLCN	LATS1	NKX2-1	RAC1	TCF7L2
ASXL1	CNTRL	FLI1	LATS2	NONO	RAD21	TET1
ATIC	COL1A1	FLT3	LHFP	NOTCH1	RALGS	TET2
ATP1A1	COL2A1	FLT4	LIFR	NOTCH2	RALGSL	TFG
ATP2B3	COX6C	FOXA1	LMO1	NPM1	RANBP17	TFPT
ATRX	CREB1	FOXL2	LMO2	NRAS	RB1	THRAP3
AXIN1	CREB3L1	FOXP1	LPP	NRG1	RBM15	TLX1
AXIN2	CREB3L2	FSTL3	LRAS	NSD1	RET	TLX3
BCL11A	CRLF2	FUBP1	LRIG3	NT5C2	RHOH	TOP1
BCL11B	CRTC1	FUS	LYL1	NT5DC4	RMI2	TP53
BCL5	CRTC3	GAS7	MAF	NTRK1	RNF213	TPM3
BCL7A	CSF3R	GATA1	MAFB	NTRK3	RNF43	TPM4
BCL9	CUX1	GATA2	MAML2	NUMA1	ROS1	TRAF7
BCOR	DAXX	GMPS	MAP3K13	NUP214	RPL10	TRIM24
BIRC2	DCTN1	GNA11	MAX	NUTM1	RPL22	TRIM33
BIRC3	DDIT3	GNAQ	MDM2	NUTM2A	RPL5	TRIP11
BRAF	DDX10	GNAS	MDM2L	NUTM2B	RPN1	TRRAP
BRCA1	DDX5	GOLGA5	MDM4	NUTM2D	RSPO2	TSC2
BRCA2	DDX6	GOPC	MDS2	NUTM2E	RSPO3	TTL
BRD3	DEK	GPC3	MED12	NUTM2F	SBDS	U2AF1
BRD4	DICER1	GPHN	MEN1	NUTM2G	SDHAF2	UBR5
CACNA1D	DNM2	GRIN2A	MET	OLIG2	SDHB	USP12
CALR	DNMT3A	H3F3A	MLLT1	OMD	SDHD	VHL

CANT1	EBF1	H3F3B	MLLT10	P2RY8	SET	VTI1A
CASC5	EGFR	HAS2	MLLT11	PAFAH1B2	SETBP1	WAS
CBFB	EIF3E	HERPUD1	MLLT3	PAX3	SETD2	WHSC1L1
CBL	EIF4A2	HEY1	MLLT4	PAX7	SF3B1	WIF1
CBLB	ELF4	HIST1H3B	MLLT6	PAX8	SH3GL1	WT1
CBLC	ELK4	HIST1H4I	MN1	PBRM1	SLC34A2	WWTR1
CCDC6	ELL	HLA-A	MPL	PBX1	SLC45A3	XPO1
CCNB1IP1	EML4	HLF	MRAS	PCM1	SMAD4	YAP1
CCND1	EP300	HNF1A	MSI2	PDCD1LG2	SMARCA4	YWHAE
CCND2	EPS15	HNRNPA2B1	MSN	PDE4DIP	SMARCD1	ZBTB16
CCND3	ERBB2	HOOK3	MTCP1	PDGFRA	SMARCE1	ZMYM2
CCNE1	ERBB3	HRAS	MUSK	PHF6	SMO	ZNF278
CD79A	ERBB4	IDH1	MYC	PHOX2B	SPECC1	ZNF331
CDC73	ERC1	IDH2	MYCL	PIK3R1	SPEN	ZNF384
CDH1	ERG	IKZF1	MYCN	PIM1	SPOP	

Expansion Absence Without relevant variants Loss-of-function variants Remarkable missense variants

6.1. Expansion of tumour suppressor genes

We found in the genomes of both *C. abingdonii* and *A. gigantea* several extra copies of tumour suppressor genes, such as *SMAD4* and *NF2*, which are implicated in different physiological and pathological processes[181,182]. SMADs behave as transcription factors involved in the transduction of TGF-β signalling, cooperating with other proteins to carry out their functions[183]. We identified orthologs of *SMAD* genes in *C. abingdonii* and in other turtles, including in all cases two different copies of *SMAD4* (**Table S21**) (**Fig. S24**). Both resulting proteins contain highly conserved N-terminal MH1 and C-terminal MH2 domains and are expressed in *C. abingdonii* as assessed by RNA-Seq data, suggesting that they are both functional. *NF2* encodes a large FERM-domain-containing protein that plays an essential role in controlling several signalling pathways, including PI3K-AKT, RAC-PAK, FAK-SRC and EGFR-RAS-ERK[184]. This genomic amplification is also present in *P. sinensis*, *C. p. bellii* and *G. agassizii*. The presence of both duplications in Testudines may reflect a potential role in the protection against cancer.

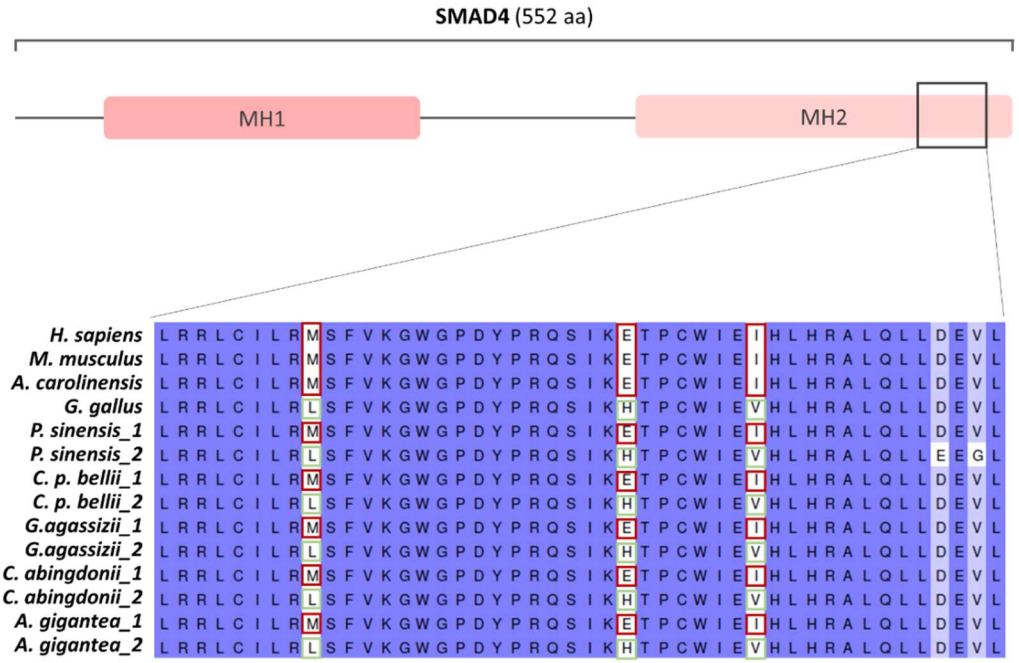


Figure S24. Partial amino acid alignment of the SMAD4 sequence in *C. abingdonii, A. gigantea*, and other vertebrate species, including three turtles.

Other tumour suppressors duplicated in *C. abingdonii* and *A. gigantea* among other analysed Sauropsida species include *PML*, *PTPN11* and *P2RY8* (**Table S21**). Manual annotation was instrumental in characterizing these duplications. In some species, such as *G. gallus* and *P. sinensis,* the specific expansion of *P2RY8* was not clearly detected, although there is an expansion in some of the known purinergic receptors. Similarly, the aforementioned premature stop codon found in *ITGA1* (see **section 7.6**) removes the GFFKR motif which has been proposed to interact with RASSF5[185], a recognized tumour suppressor involved in the RAS pathway. The putative loss of this interplay might impact on the activity of RASSF5 leading to a tumour-protecting effect.

Some oncogenic processes can be suppressed by a mechanism known as immunosurveillance[186]. The copy-number expansion of *PRF1* in Testudines could participate in this process. The absence of perforin-dependent cytotoxicity seems to play a critical role during the earliest stages of epithelial carcinogenesis[187], and these proteins act as tumour suppressors in B-cell malignancies[188]. Moreover, perforin-deficient mice are more susceptible to the development of spontaneous lymphomas[189]. Therefore, these reports suggest that the expansion of perforins could have a protective effect against cancer and ageing in Testudines.

We also found in *C. abingdonii* and *A. gigantea* two copies of *PRDM1* (**Table S21**), a transcriptional repressor that binds specifically to the PRDI element in the promoter of the *β-IFN* gene, driving the maturation of B-lymphocytes into Ig-secreting isotypes[190]. This copy-number gain was also found in *C. p. bellii* and *G. agassizii*, but not in other Testudines. In addition, *PRDM1* is a tumour suppressor gene in NK cell malignancies[191], suggesting a protective

role against cancer and a robust innate immune system in these animals carrying an extra copy of *PRDM1*[5].

Altogether, the redundancy of tumour suppressor genes observed in Testudines, and particularly in giant tortoises, could contribute to explain their apparent natural resistance to cancer[192] (**Fig. S25**). This might provide an interesting example of convergent evolution, as other large and long-lived vertebrates also show genomic expansions affecting tumour suppressors[193].

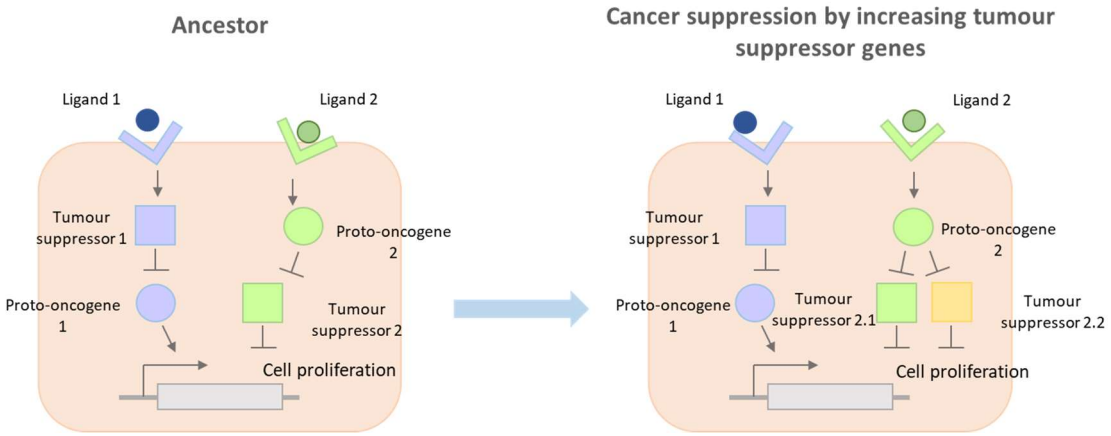


Figure S25. Putative mechanism of cancer control by expansion of tumour suppressor genes.

6.2. Proto-oncogene evolution

Proto-oncogenes play important roles in multiple biological pathways. However, when activated, the resulting proteins may favour unregulated cell growth and cancer. Therefore, oncogene activation is a tightly regulated process resulting from the interplay of multiple factors. In this regard, we detected the duplication of several proto-oncogenes in *C. abingdonii* genome. Namely, we found two conserved copies of *MYCN* and *SET* (**Table S21**), also present in *A. gigantea* but not in other Testudines, which suggests that these changes are exclusive of the group including giant tortoises. Of note, the sequences of both *SET* copies are almost identical, and therefore we have not been able to independently validate this duplication. *MYCN* encodes a member of the MYC family, highly expressed in the fetal brain and critical for its normal development[194]. The newly identified MYCN-like gene (*MYCNL*) is present in all the species of Galapagos giant tortoises tested and in *A. gigantea*, but it is not present in other turtles such as *P. sinensis*, *C. p. bellii* or *G. agassizii*. *SET* encodes a PP2A inhibitor which regulates a variety of processes such as apoptosis, cell cycle and cell motility[195], and is overexpressed in many human cancers. In addition, this protein belongs to the SET complex, which mediates oxidative stress responses induced by mitochondrial damage through the action of PRF1 and GZMA in CTL and NK-mediated cytotoxicity[196]. This extra copy of *SET* is also present in *A. gigantea* but not in *G. agassizii*, *P. sinensis* or *C. p. bellii*. Related to this, *USP12* – encoding a cysteine protease involved in the activation of c-MYC[197] – was found to be duplicated in *G. agassizii*, *C. abingdonii*, other

Galapagos giant tortoises, *A. gigantea* and their continental outgroups, as assessed through Sanger sequencing validation (**Table S7; Table S21**).

Another remarkable duplication in *C. abingdonii* affects unconventional myosins[198]. Thus, we have annotated a new myosin copy that resembles *MYO5A* and *MYO5B*. This additional copy is also present in *A. gigantea* and has been annotated in *P. sinensis*, *C. p. bellii, C. mydas* and *G. agassizii*. Based on sequence identity of this novel gene with both *MYO5A* and *MYO5B* (**Table S21**), we have called it *MYO5AB*, even though the phylogenetic analysis does not support a direct relationship with those genes (**Fig. S26**). This new gene contains a missense mutation in its ATP-binding site (p.E114K), which is expected to ablate its ATPase activity[199]. This variant was also present in non-muscle myosins type II as MYH9, which is involved in diverse functions such as cytokinesis, cell motility, maintenance of cell shape and tumour suppression[200]. Thus, the presence of a non-functional copy of this protein might modulate the activity of the functional copy by sequestering substrates, suggesting a new regulatory mechanism affecting this family.

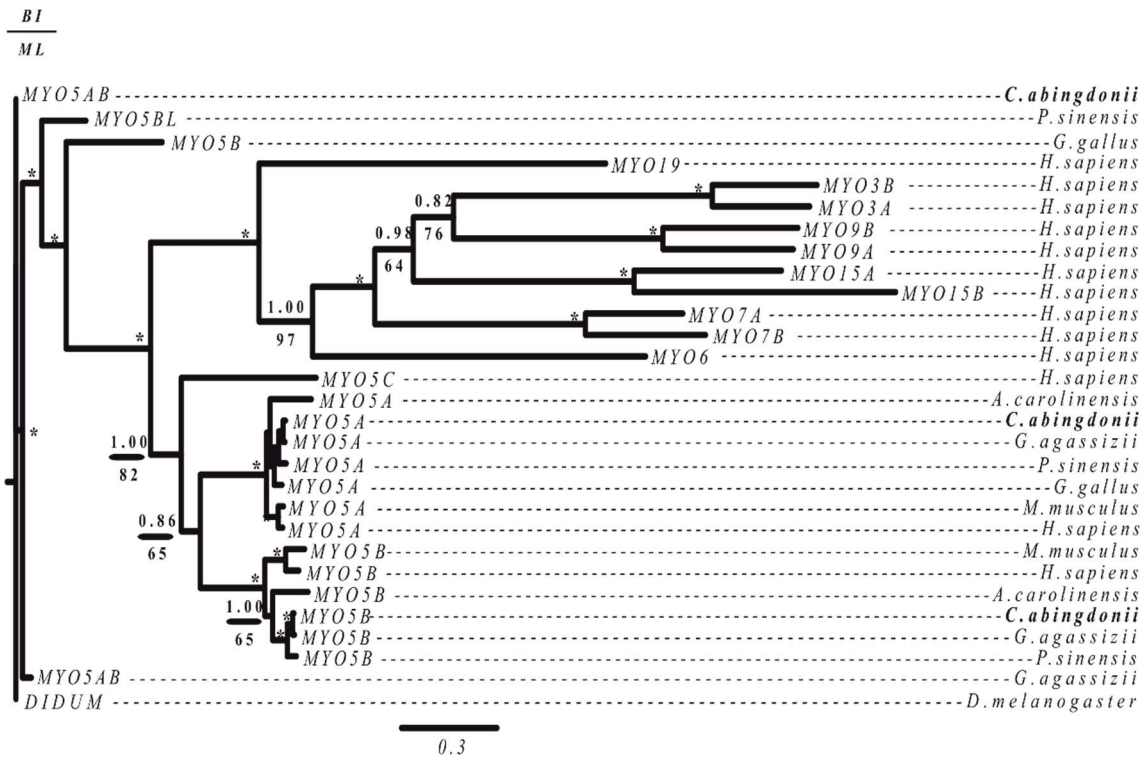


Figure S26. Topology of the Bayesian inference (BI) and maximum likelihood (ML) trees showing the relationship of unconventional myosin genes in *C. abingdonii, A. gigantea* (in bold) and other vertebrate species. Numbers represent Bayesian posterior probabilities and bootstrap scores to estimate the robustness of the phylogenetic tree. Asterisks (*) indicate 1.00 and 100 pp and bs, respectively.

Another gene duplication of potential interest due to its role in cancer involves *RAS* genes. RAS is a family of related proteins belonging to the small GTPase class, which are implicated in transmitting signals within cells[201]. We annotated a new gene, which we have called *LRAS*, in *C.*

abingdonii, *A. gigantea, P. sinensis*, *C. p. bellii, G. agassizii* and *A. carolinensis*. This gene shares a common ancestor with *HRAS*, *NRAS* and *KRAS* (**Fig. S27**).

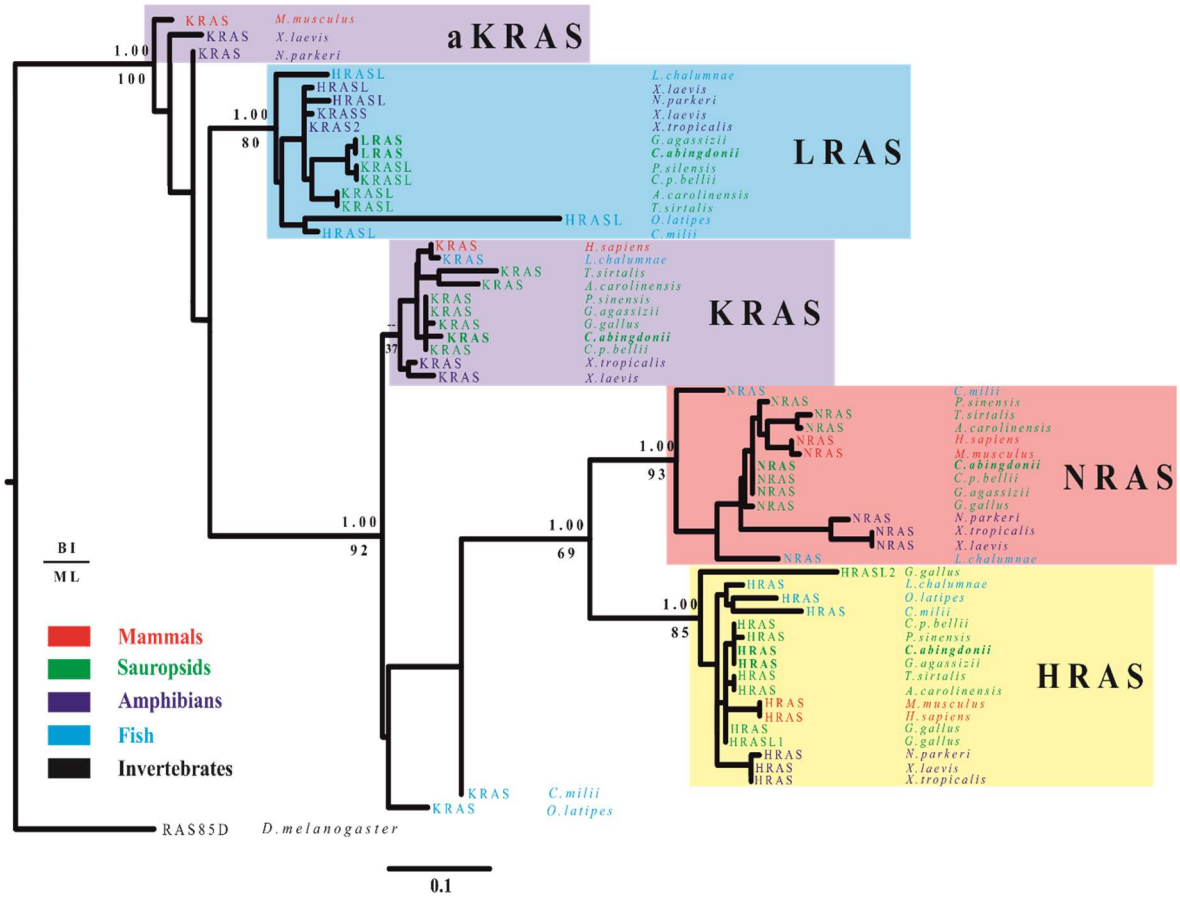


Figure S27. Topology of the BI and ML trees showing the phylogenetic relationships of the RAS family proteins in *C. abingdonii, A. gigantea,* and other vertebrate species. Numbers represent Bayesian posterior probabilities and bootstrap scores (above and below each branch, respectively). Text colour represents the type of animal to which each species belongs. Dash line (-) indicates no support in this analysis.

We also identified an additional copy of *NRAS* in *C. abingdonii* and *A. gigantea*, probably a pseudogene due to the presence of several premature stop codons in both species. Interestingly, *LRAS* displays a divergent C-terminus with an intact farnesylation signal. The accumulation of amino acid substitutions at this site could provide a modulatory effect of signal transduction mediated by RAS (**Fig. S28**).

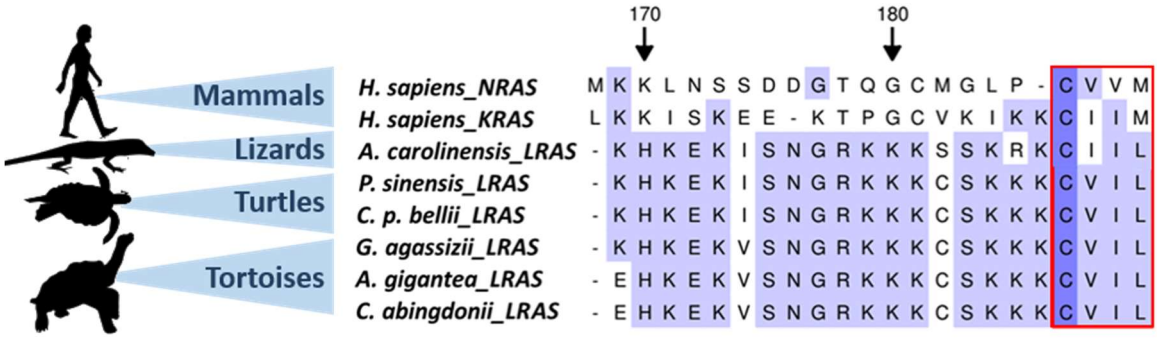


Figure S28. Dominant-negative mutations in proto-oncogenes evolution in *C. abingdonii, A. gigantea* and other vertebrate species. NRAS, KRAS and LRAS amino acid alignment including different representative species. The CAAX-farnesylation signal sequence, essential for the functional role of Ras and Ras-like proteins[202] is emphasized with a red rectangle.

Dominant-negative mutations were described as mutations that act in opposition to the normal activity of the gene product when overexpressed[203]. The presence of a second copy of a gene, as described herein for genes of oncological relevance present in the genomes of giant tortoises, would open up additional opportunities in evolution to develop new functional properties[204].

6.3. Modifications in the Hippo pathway

The Hippo pathway consists of a large network of proteins that play a pivotal role in organ size control during development and regeneration by promoting cell-cycle arrest and apoptosis, as well as in pathological states like cancer. The core of this pathway comprises a kinase cascade in which STK3/MST2 and STK4/MST1, in association with its regulatory protein SAV1, phosphorylate and activate LATS1/2, which in turn phosphorylates and inactivates YAP oncoprotein and WWTR1/TAZ. Phosphorylation of YAP by LATS2 inhibits its translocation into the nucleus, where it regulates important genes for cell proliferation, migration and death[205] (**Fig. S29**).

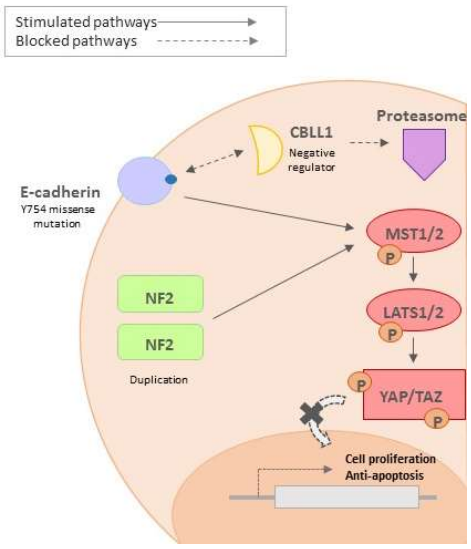


Figure S29. Modifications in Hippo pathway in *C. abingdonii* and other analysed tortoises. Results suggest a differential modulation of Hippo pathway, promoting cell cycle arrest and inhibiting anti-apoptotic signals.

Multiple studies have shown that key junctional proteins can regulate YAP and TAZ by sequestering them at cell junctions, thus preventing nuclear entry and contact with their specific phosphatases[206]. E-cadherin, the best characterized and prototype member of the classical cadherins in epithelial cells, is a key component of adherens junctions which acts as a potent tumour suppressor. We found a missense variant (p.Y754H) in all analysed

tortoise species and also in *C. p. bellii,* which was confirmed by Sanger sequencing (**Table S7**). This is an essential phosphorylated residue for the interaction of E-cadherin with its negative regulator CBLL1[207], an E3 ligase that mediates ubiquitination of the E-cadherin complex. While the resulting histidine residue may also be phosphorylated, this reaction must be mediated by different enzymes. Therefore, we hypothesize that this missense variation could modify the regulation of the interaction of E-cadherin with CBLL1, perhaps promoting the stabilization of E-cadherin in the surface of cytoplasmic membrane. Due to the role of E-cadherin in preventing epithelial-mesenchymal transition and downregulating cell proliferation, the existence of this variant may suggest a new protective mechanism against tumorigenesis. Of note, the aforementioned functional duplication of *NF2* (**Table S21**) may also affect the regulation of the Hippo pathway by binding and recruiting LATS to the plasma membrane, where it is activated by MST kinases[208]. Taken together, these results suggest a tendency favouring activation of the Hippo pathway, which in turn would lead to cell proliferation arrest, and anti-apoptotic pathway inhibition (**Fig. S29**).

In summary, our analysis shows a trend towards a lower tumorigenic potential in giant tortoises. Namely, several tumour suppressors have been genomically amplified, which might prevent cells from undergoing transformation. In addition, we have found several putative modulators of proto-oncogenes, suggesting novel regulatory mechanisms specific of this group of genes. While these results may be important in understanding the development of these giant organisms, they may also provide clues about the enhanced protection against tumour development suggested by Peto's paradox. Given the limitations of this analysis, further works are needed to confirm these hypotheses.

7. Genomic basis of longevity

The unprecedented advances over the last years in ageing research have led to the postulation of nine hallmarks that define ageing and therefore affect longevity[26]. Moreover, the development of next-generation sequencing techniques has led to the comprehensive study of the genome of different long-lived organisms such as the naked mole rat and the bowhead whale[23,177]. Some members of Testudines, such as *C. abingdonii,* all other Galapagos giant tortoises and *A. gigantea* are also long-lived. Estimates have placed Lonesome George at over 100 years. The maximum lifespan of the Aldabra giant tortoise ranges from estimates of 70 years[209] to 65-90 years[210], with unverified higher estimates reported. Moreover, the Mojave Desert tortoise is estimated to live up to 56 years in the wild[211]. Accordingly, we exhaustively analysed a set of almost 495 genes involved in the molecular control of longevity and found relevant genetic modifications that could affect some of the well-established hallmarks of ageing (**Fig. 2a, main text**) (**Table S15**).

Table S15. Status of ageing-related genes analysed in *C. abingdonii*.

A2M	APP	CCDC88C	ECHS1	GSTAL2	IRS2	MYT1L	PKNOX2	RIMBP2	SYK
ABCC4	ARHGAP1	CCL5	EEF1A1	GSTAL3	ITGA1	NAV2	PLAU	RMI2	SYNE1
ABCC8	ARHGAP8	CDC42	EFEMP1	GSTAL4	ITGB3	NBEA	PLXNA4	RNF145	SYT13

ABO	ARNTL	CDH4	EGFR	GSTT1	JAK2	NBN	POLA1	RPA1	TBL1XR1
ACD	ASHG	CDK14	EIF4E3	GSTZ1	KCNMA1	NCAM2	POLA2	RPTOR	TBXAS1
ACE	ASIC2	CDK6	EIF4EBP1	GTF2H5	KCNQ4	NCOR1	POLB	RTEL1	TENM4
ACTN3	ATCAY	CDK7	EIF5A2	H2AFX	KCNT2	NCOR2	POLD1	RYR3	TERF2
ADA	ATG4C	CDKN1A	ELL2	H6PD	KCTD1	NDUFA11	POLD2	SDC4	TERT
ADAMTSL1	ATM	CDKN1B	ELN	HDAC1	KIF13B	NDUFS1	POLD3	SDHC	TGFA
ADARB1	ATP5O	CDKN2A	ELOVL2	HDAC10	KL	NEIL1	POLD4	SDHCL	TGFB1
ADARB2	ATRIP	CDKN2B	ELTD1	HDAC11	KLRF1	NETO1	POLE	SEMA3A	TH
ADCY5	ATXN2	CEBPA	ENOX1	HDAC2	LAMTOR2	NFE2L1	POLE2	SEMA6A	THSD7B
ADIPOQ	ATXN7L3B	CEBPB	EPB41L4B	HDAC3	LEP	NFE2L2	POLE3	SERPINE1	TLK2
AGER	BANF1	CETP	EPS8	HDAC4	LEPR	NFKB1	POLE4	SERPINH1	TLR4
AGPAT2	BAG6	CGNL1	EQTN	HDAC5	LINGO2	NFKBIA	POLG	SGK1	TMEM151B
AGT	BDNF	CHEK2	ERGIC1	HDAC6	LMNA	NLN	POLG2	SGK2	TMEM2
AGTR1	BHLHE41	CHIT1	ESR1	HDAC7	LMNB1	NLRC5	POLH	SGSH	TMPRSS6
AGTR2	BIN2	CHST11	ESRRG	HDAC8	LMNB2	NLRPs	POLI	SH2B3	TNF
AHSG	BLM	CISD2	EXO1	HDAC9	LMNTD1	NNMT	POLK	SH2D4A	TOMM40
AIF1	BNIP2	CLOCK	F7	HECW1	LMO4	NOS1	POLL	SIRT1	TOP3B
AIFM1	BNIPL	CLSTN2	FADS1	HECW2	LMX1B	NOS2	POLM	SIRT2	TOX
AKAP10	BRCA1	CLU	FAM13A	HELLS	LPA	NPAT	POLN	SIRT3	TP63
AKT1	BRCA2	CLYBL	FAM19A5	HFE	LRP1B	NR3C1	POLQ	SIRT4	TP73
AKT1S1	BRE	COL5A1	FGF19	HIF1A	LY6G6F	NRDE2	POMT2	SIRT5	TPP2
AKT3	BRIP1	COL6A3	FGFR1	HIP1	LY86	NRXN3	PON1	SIRT6	TRIM25
ALDH2	BSCL2	COQ7	FHIT	HLA-DQA1	LYN	NT5DC1	POT1	SIRT7	TSFM
ANGEL1	BTBD19	CPB2	FSHR	HLA-DQB1	LYST	NUDT1	POU1F1	SLC13A1	TSHR
APBB2	BTBD9	CSF1R	GABRR3	HLA-DRB1	MAPK3	OGDH	PPARG	SLC13A2	TTC6
APEX1	BTG3	CSMD3	GAPDH	HMOX1	MAPK9	OR56A1	PPARGC1A	SLC13A3	TTR
APOA1	BTNL2	CSNK1E	GATA4	HNRNPD	MECOM	ORC5	PPM1D	SLC13A4	TUSC3
APOA2	BUB1B	CTC1	GBA3	HP	MEFV	OTOL1	PPP1R1A	SLC13A5	TXN
APOA4	BUB3	CTF1	GCLC	HPCAL1	METTL1	PALB2	PPP1R26	SLC25A21	UCP1
APOA5	C1QTNF5	CTNNA2	GCLM	HRAS	METTL21B	PAPD5	PPP2R1A	SLC9A3R2	UCP3
APOB	C2	CTNNA3	GH1	HS3ST3B1	MGP	PARK2	PPP2R2B	SLX4	VASH1
APOC1	C20orf194	CTNND2	GHRH	HSPA9	MIF	PARVG	PPP2R2C	SMARCA2	VPS13D
APOC2	C20orf78	CYC1	GHRHR	HTRA2	MINPP1	PAX4	PREX1	SNCA	VWA5A
APOC3	C7orf50	CYP19A1	GHRL	HTT	MKL1	PCK1	PRKCA	SOCS2	WDR72
APOC4	C9ORF3	CYP1B1	GHSR	ICAM1	MLH1	PCK2	PROM1	SORCS1	WRAP53
APOD	CACNA1C	DCLRE1B	GLIS1	IFNG	MRE11A	PCMT1	PROP1	SORCS2	WWOX
APOE	CADM1	DCPS	GPC6	IGF1R	MSH5	PCNX2	PRR5-ARHGAP8	SPRTN	XDH
APOF	CAMK4	DDB2	GPX4	IGF2R	MSRA	PCSK1	PRR5L	SQSTM1	XKR6
APOH	CAMTA1	DEFB1	GRAMD1B	IGFBP3	MSTN	PEX5	PRUNE2	SSPN	XRCC5
APOL1	CARD14	DGAT1	GRIA1	IGFBP4	MT1A	PGPEP1	PSEN1	ST3GAL3	XRCC6
APOL2	CARS	DKC1	GRIN2B	IKBKB	MT1E	PHYHIP	PTEN	STK24	YBX2
APOL3	CASP5	DLGAP5	GRN	IL10	MT2A	PICALM	PTH	STUB1	YTHDF2
APOL4	CASP8	DOCK8	GSG1L	IL12A	MTOR	PICK3CA	RAE1	SUCLA2	YWHAG
APOL5	CAT	DOT1L	GSK3A	IL12RB2	MTR	PIGC	RBPMS	SULT1A1	ZBTB20
APOL6	CAV1	DRD4	GSR	IL18	MTTP	PIK3C2G	RECQL4	SUMF1	ZFYVE28
APOM	CCDC170	DZIP3	GSTA4	IL6	MYLK4	PIK3CA	RHOA	SUV39H1	ZMPSTE24
APOO	CCDC85A	EBF3	GSTAL1	IRS1	MYO9B	PITPNM3	RICTOR	SVEP1	ZPLD1

Expansion Absence Without relevant variants Loss-of-function variants Remarkable missense variants

7.1. Genomic instability

One characteristic feature associated with ageing is the accumulation of genetic damage as a result of continuous environmental and cellular stresses, along with the disruption of the function of the DNA repair machinery. In the context of proper DNA maintenance, and as mentioned above (**section 5.5**), we found in the genomes of *C. abingdonii*, *A. gigantea*, their continental outgroups, *C. p. bellii* and *G. agassizii* a duplication affecting *NEIL1*, a key protein involved in the base excision repair (BER) process (**Table S7**). A higher expression of *NEIL1* was previously correlated with extended lifespan in several species[170], pointing to this gene as one of the candidate putative factors involved in the exceptional longevity of giant tortoises.

Additionally, the aforementioned variants affecting *ATM*, *XRCC5* and *XRCC6,* as well as the duplications of *RMI2* and *GAPDH* (**Table S7**) could also contribute to improve the efficiency of DNA repair and therefore, the stability of the genome.

7.2. Telomere attrition

As mentioned above, cell repair mechanisms play an essential role in multiple cellular processes, including transcription, apoptosis and telomere maintenance. The accumulation of mutations due to the saturation of the DNA repair machinery with age is accompanied by telomere shortening[26]. Thus, cells need numerous strategies to counteract the natural tendency to telomere attrition. In this regard, the exonuclease DCLRE1B participates in the protection of telomeres against NHEJ-mediated repair and plays a key role in telomeric loop (T-loop) formation after being recruited by TERF2[212]. In *C. abingdonii*, *A. gigantea*, *G. agassizii* and all the other species of Galapagos giant tortoises and continental outgroups used in the Sanger sequencing validation (**Table S7**), this protein presented a modification in the heterodimer interface binding TERF2 (p.R498C)(**Fig. 3b, main text**). Furthermore, TERF2 had several alterations in the domain responsible for the binding to DCLRE1B (p.S161T, p.M164T, p.N176G and p.M180L), also shared by other turtles and birds. This variation in key binding residues could affect the interactions between these two proteins and therefore vary the T-loop formation. Moreover, the aforementioned variants present in DNA repair genes, such as *ATM*, *XRCC5* and *XRCC6*, may also impinge on telomere dynamics in tortoises[213-215].

7.3. Loss of proteostasis

Protein homeostasis (proteostasis) in cells usually deteriorates during ageing and correlates with the downregulation of genes responsible for the synthesis and recycling of proteins[26,216]. In this context, *EEF1A1* – which encodes an isoform of the alpha subunit of the elongation factor-1 complex – is responsible for the delivery of tRNAs to the ribosome. In *C. abingdonii*'s genome, we identified three copies of this gene, one of which is a retrogene. Comparative analysis with data from *A. gigantea* and *G. agassizii* revealed in both of them the existence of two or more copies and no retrotranscript, which suggests a duplication mechanism initiated in a common ancestor, and the later incorporation of the reverse-transcribed copy to *C. abingdonii*'s genome. To confirm these results, we validated all copies by Sanger sequencing (**Table S7; Table S21**), which led us to identify the two functional copies of *EEF1A1* in all the studied species (including Galapagos giant tortoises, *A. gigantea* and their continental outgroups), whereas the retrotransposed gene seems to be exclusive to *C. abingdonii*. Interestingly, overexpression of *EEF1A1* increases lifespan in *Drosophila melanogaster*[217]. These findings suggest that proteostasis maintenance may represent a pro-longevity mechanism in these animals.

7.4. Deregulated nutrient sensing

Different signalling pathways that recognize and respond to fluctuations in nutrient levels are commonly deregulated during ageing[132]. Over time, glucose dysregulation can result from alterations throughout metabolism control pathways, such as glycolysis or gluconeogenesis. In this regard, the aforementioned variant in giant tortoises affecting the activation loop of GSK3A (**section 4.1**) may be involved in the maintenance of glucose homeostasis in these organisms (*C. abingdonii*, the other Galapagos giant tortoises, *A. gigantea*, their continental outgroups, *G. agassizii* and *C. p. bellii*) (**Table S7**). Interestingly, the inhibition of *GSK3* can extend lifespan in *D. melanogaster*[218], which highlights its critical role as a regulator of longevity across the evolution and, particularly, in tortoises. Likewise, the identified alterations in giant tortoises affecting other genes implicated in glucose metabolism may also represent pro-longevity determinants in these animals. In fact, the *NLN* inactivation in giant tortoises and their continental outgroups due to the presence of premature stop codons could provide a metabolic advantage for these animals, as mutant mice deficient in this enzyme present increased glucose tolerance and insulin sensitivity[134].

7.5. Mitochondrial dysfunction

The correct function of mitochondria is critical for homeostasis maintenance, whilst these organelles are crucial in energy production through beta oxidation and Krebs cycle, and their malfunction or dysregulation leads to severe disorders[132,133,219]. In this context, two variants affecting *ALDH2*, which encodes a mitochondrial aldehyde dehydrogenase, were found in several analysed tortoises. Specifically, a variant affecting an NAD-binging site (p.Q366M)[220,221] is common to all the Galapagos giant tortoises, but not to their continental outgroup *C. chilensis*, nor the Aldabra tortoise or its outgroups. Another variant located in an homotetrameric interface site[220] (p.M487T) was found in all the tortoises studied (including both Galapagos and Aldabra tortoises, along with their outgroups) (**Table S7**). This protein is implicated in alcohol metabolism and lipid peroxidation, among other detoxification processes[222]. The metabolic machinery responsible for alcoholic degradation arose in the evolution as an adaptation to the consumption of fermented fruits[223], but it also participates in cellular homeostasis through the elimination of potential damaging molecules such as reactive aldehydes[222]. Therefore, these mutations could also alter the detoxification process and contribute to pro-longevity mechanisms. Likewise, the above described alterations in other genes such as *NLN* and *GAPDH*, which encode enzymes associated with mitochondrial functions[224,225], may also impinge on this hallmark of ageing and contribute to extend the lifespan and healthspan of these animals.

7.6. Altered intercellular communication

Several changes in intercellular signalling take place during ageing, such as those involved in immunosurveillance, neurohormonal signalling and inflammatory reactions[26]. Accordingly, we

specifically examined the genomes of *C. abingdonii* and *A. gigantea* for variants occurring in genes affecting intercellular communication of putative relevance for ageing. In this context, the integrin alpha-1 gene (*ITGA1*), a heterodimeric cell-surface protein involved in cell-matrix and cell-cell interactions, presents a premature stop codon in position p.R990* only in *C. abingdonii*. This premature stop codon is expected to remove the C-terminus of the protein, previously described as a critical structural element for the activation of alpha integrins[226] (**Fig. S30**).

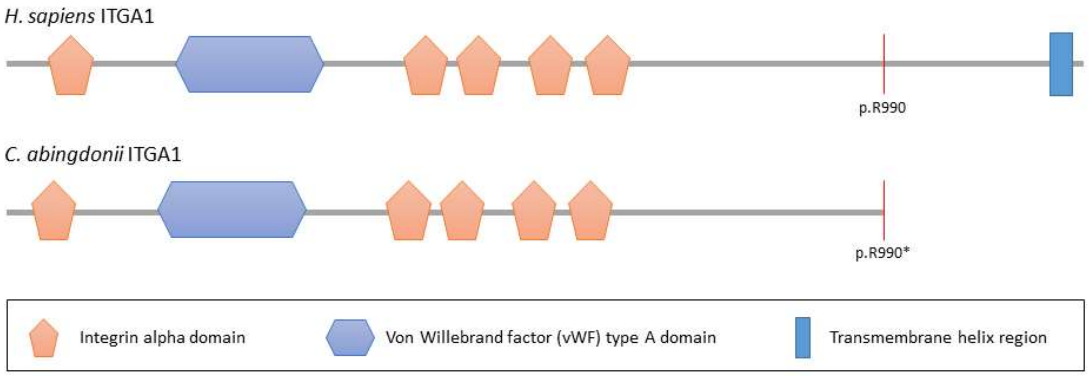


Figure S30. Domain structure of human and *C. abingdonii* ITGA1. The identified mutation affecting ITGA1 generates a stop codon (red bar) at the residue 990 in the protein, removing the C-terminal and a transmembrane region that participates in the dimerization process.

7.6.1. Inflammation and ageing

Among the multiple age-related alterations that affect intercellular communication, inflammation takes a prominent role resulting in age-associated dysfunctions. This situation is known as "inflammaging", and is caused by many factors, such as the accumulation of pro-inflammatory tissue damage, dysfunction of the immune system, or production of pro-inflammatory cytokines by senescent cells, among others[227]. In this context, the discussed variant in the pro-inflammatory cytokine MIF in giant tortoises an *G. agassizii* (**see section 4.1**) has been shown by mutagenesis studies to cause the formation of interchain disulfide bonds between the altered p.N111C and the p.C81 residue of an adjacent subunit, inhibiting intracellular signalling cascades[139]. Moreover, *MIF* deficiency tends to reduce chronic inflammation in white adipose tissue[140] and to expand lifespan, especially in response to caloric restriction[228]. In general, this result reinforces the hypothesis that giant tortoises could have decreased chronic inflammatory response, and reduced production of pro-inflammatory interleukins, thereby expanding their longevity.

7.6.2. IGF system and ageing

Insulin and insulin-like growth factors (IGFs) play important roles in the regulation of multiple processes, including metabolism and growth[229]. The IGF system is also involved in the ageing process, as suggested by the extended lifespan in different species correlated to an IGF signalling

decrease[230,231]. Thus, the characterization of the IGF system is important to further understand its contribution to growth in giant tortoises and its possible implication in the extremely long lifespan of these animals. The interaction between IGF1/2 and their receptors (IGF1R and IGF2R) is a key step in this signalling mechanism[231]. We found that all of the key residues for this interaction are conserved between the selected species, with the exception of Asn-724 in IGF1R, which corresponds to an aspartic acid in *C. abingdonii*. Further comparative analysis revealed the presence of the same specific amino acid substitution in *G. agassizii*, Galapagos and Aldabra tortoises and also in the continental outgroups, which are also long-lived (**Fig. 4a, main text**) (**Table S7**). The presence of this specific change suggests that the correct interaction between IGF1/2 and IGF1R may have changed in these tortoises, probably regulating the downstream signalling mechanism. To explore this possibility, we treated with IGF1 different cells expressing pCDH, pCDH-IGF1R^{WT} and pCDH-IGF1R^{N724D} plasmids, and observed a significant reduction in the phosphorylation levels of the receptor carrying the mutation (**Fig. 4b, main text, Fig. S31**). Consequently, this unique change in IGF1R provides an attractive target to study the cellular mechanisms underlying the exceptional long lifespan of these animals. Also in this regard, it is remarkable our finding in tortoises of a short deletion in the coding region of *IGF2R* which results in the loss of two amino acids in this growth factor receptor. Interestingly, an IGF2R SNP has been associated with human longevity[232], which opens the possibility that the variant found in the studied tortoises in this additional component of the IGF system could somewhat contribute to increase the lifespan of these long-lived animals.

8. Conclusions

Following a hypothesis-driven strategy, we have carefully annotated a large set of genes related to multiple physiological processes, focusing specially on response to stress, ageing, and DNA repair. By using this manual approximation, we could identify several changes affecting relevant residues of pivotal proteins, which may had been overlooked by more automatic methods.

By comparing our annotations to those of other turtles, our results may shed light on several aspects of the evolution of these vertebrates. Thus, we have found genomic duplications affecting WNT4 and WNT11 in turtles which, given their known roles in development, might be associated with the growth of their shell. We have also pointed out several genomic amplifications in Testudines that suggest an enhanced innate immune system, in addition to a much richer perforin and granzyme system than mammals. We have also found several examples of adaptations to hypoxia, including variants in TP53 and expression of specific globins, shared by all Testudines analysed so far.

In addition, the comparison of these data with the genomic sequence of the Aldabra tortoise may provide important clues about the specific evolution of giant tortoises. In this regard, *C. abingdonii* and the Aldabra tortoise share several variants predicted to affect their neurological development. Surprisingly, most of these alterations are expected to lead to lower cognitive

capabilities and even neurological diseases. However, these alterations may simply reflect a different regulation of these processes in these tortoises compared to other organisms. Also, the phylogenetic group containing giant tortoises seem to have undergone multiple changes affecting glucose metabolism. Several specific alterations suggest that both glucose consumption and glucose tolerance were increased in their lineage. In addition to the different caloric requirements of giant tortoises, these characteristics may also be related with changes affecting the insulin system, which is known to play a role in size regulation. This hypothesis is reinforced by our finding of a specific variant affecting an important residue in IGF1R in tortoises.

An important consequence of gigantism is a potential increase in the risk of cancer development. We have found a number of gene changes specific of giant tortoises which might contribute to explain the low prevalence of cancer in Chelonians, which is remarkably lower than in the other reptilian counterparts (2.7% in Chelonians *vs.* 15% in snakes)[192]. Indeed, when Chelonians were split into turtles and tortoises, the prevalence of neoplasia in the last group was 1.4%. Thus, changes affecting genes involved in DNA repair, like *RMI2*, *XRCC5* and *XRCC6*, suggest a differential regulation of DNA integrity in this lineage. Moreover, several known tumour suppressor genes, like *PRDM1* and *PRF1*, were expanded in a common ancestor to both giant tortoises. Notably, we have also found duplications affecting proto-oncogenes, but some of the novel copies show evidence for loss of function or dominant negative functions, which might actually point to a lower tumorigenic potential.

Finally, giant tortoises are also examples of extreme longevity. In this regard, we have found multiple specific variants that might be related to this trait, affecting ageing hallmarks like genomic instability (*NEIL1*), telomere attrition (*DCLRE1B*, *TERF2*), loss of proteostasis (*EEF1A1*), mitochondrial dysfunction (*ALDH2*), and intercellular signalling (*IGF1R*). The interplay of these factors with the immune and insulin systems may hold important clues about the ageing process in these organisms, improving our understanding of longevity by opening new lines of investigation.

9. Raw Data

Figure 4b

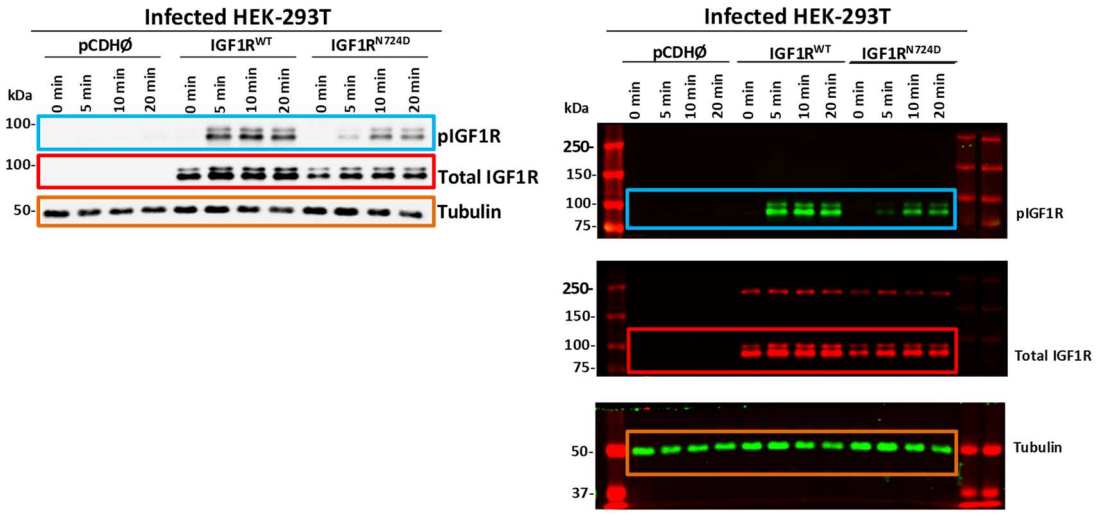


Figure S31. Raw data of Western-blots from Figure 4b, main text.

Figure S22, upper panel

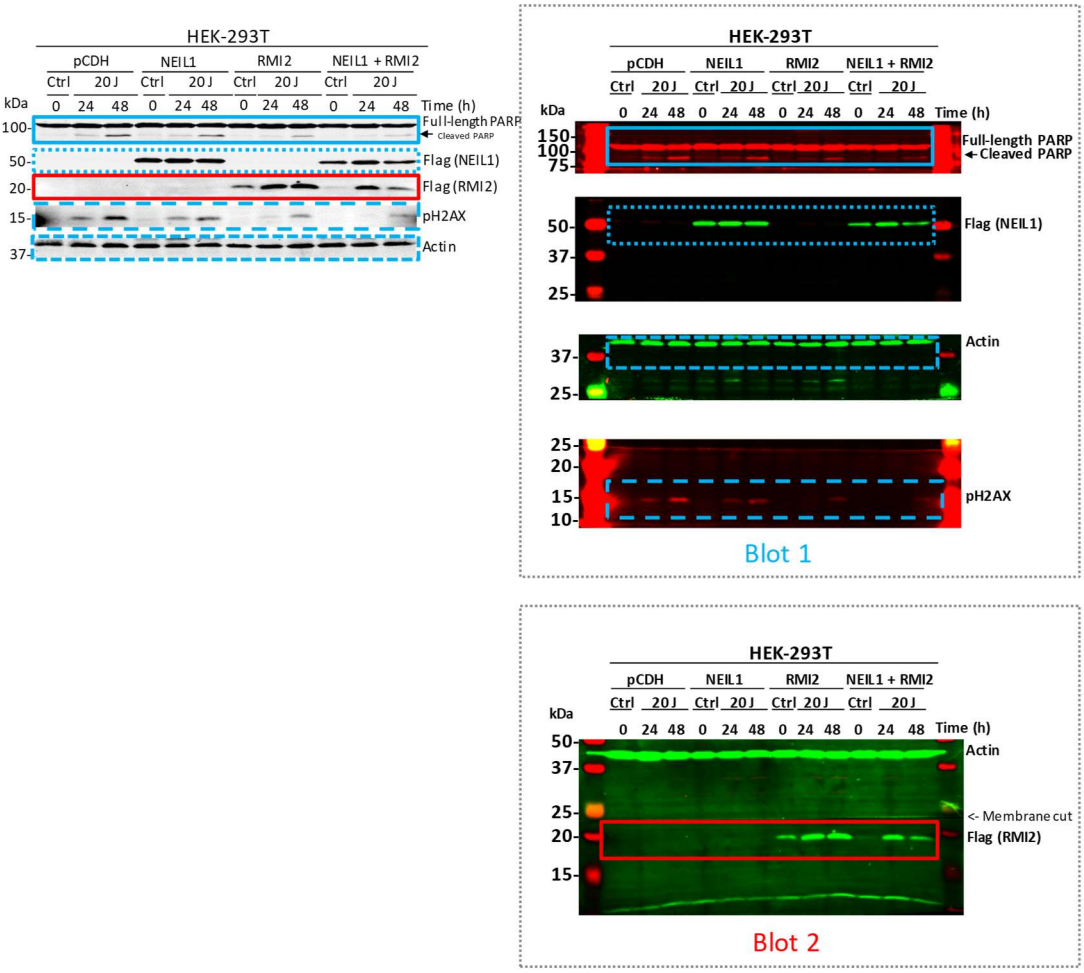


Figure S32. Raw data of Western-blots from Figure S22, upper panel.

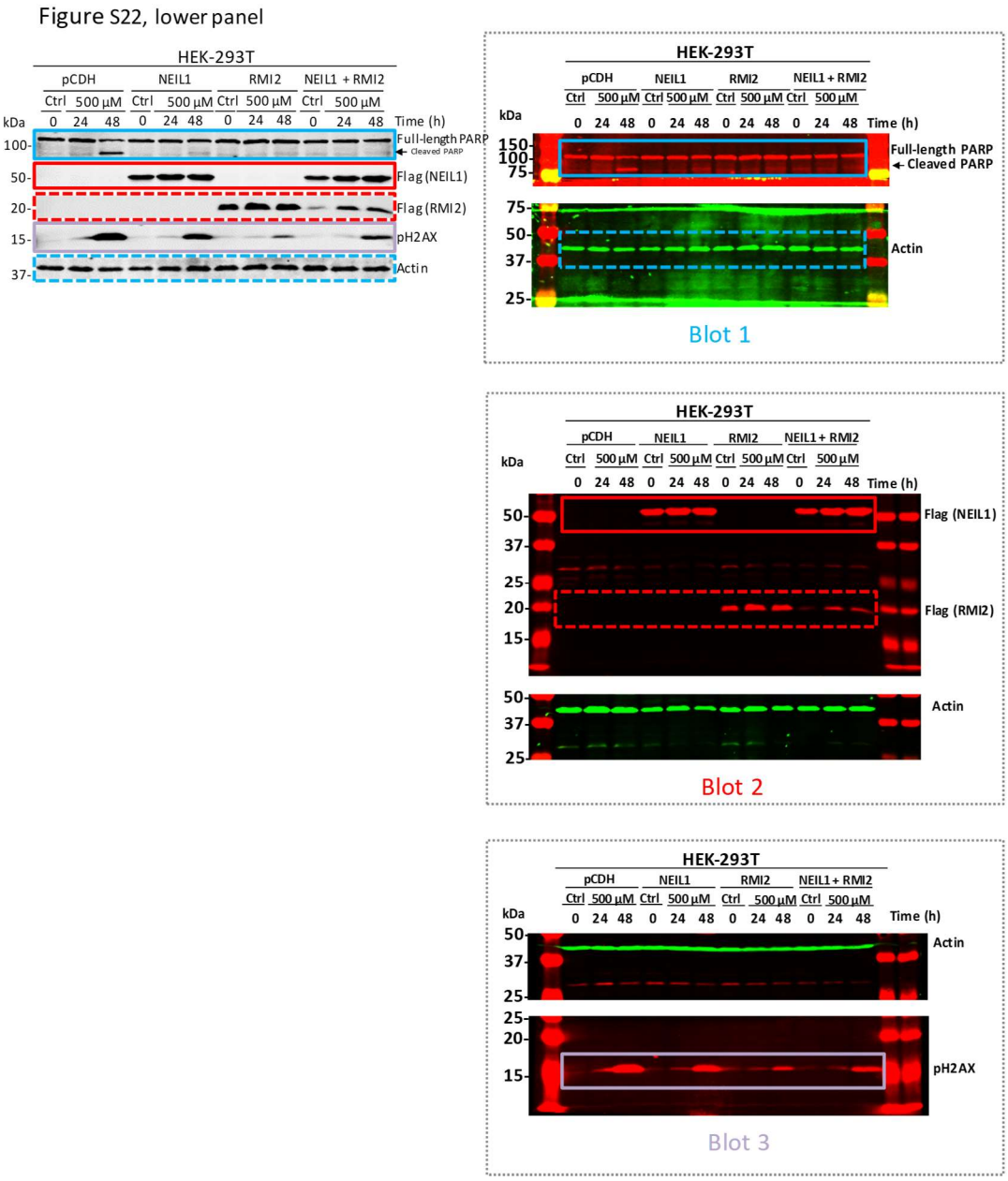


Figure S33. Raw data of Western-blots from Figure S22, lower panel.

10. References

1 Coordinators, N. R. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* **46**, D8-D13 (2018).

2 Gnerre, S. *et al.* High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* **108**, 1513-1518 (2011).

3 Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578-579 (2011).

4 English, A. C. *et al.* Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One* **7**, e47768 (2012).

5 Shaffer, H. B. *et al.* The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol* **14**, R28 (2013).

6 Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet* **45**, 701-706 (2013).

7 Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. 2013–2015. *URL http://www. repeatmasker. org* (2014).

8 Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).

9 Zdobnov, E. M. *et al.* OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res* **45**, D744-D749 (2017).

10 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).

11 Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496 (2011).

12 Mazet, O., Rodriguez, W. & Chikhi, L. Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theor Popul Biol* **104**, 46-58 (2015).

13 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111 (2009).

14 Campbell, M. S., Holt, C., Moore, B. & Yandell, M. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics* **48**, 1-39 (2014).

15 UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169 (2017).

16 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

17 Suh, A. *et al.* Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol Evol* **7**, 205-217 (2014).

18 Tollis, M. *et al.* The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PLoS One* **12**, e0177708 (2017).

19 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).

20 Prattichizzo, F. *et al.* Exosome-based immunomodulation during aging: A nano-perspective on inflamm-aging. *Mech Ageing Dev* **168**, 44-53(2017).

21 Becker, A. *et al.* Extracellular vesicles in vancer: cell-to-cell mediators of metastasis. *Cancer Cell* **30**, 836-848 (2016).
22 Baixauli, F., López-Otín, C. & Mittelbrunn, M. Exosomes and autophagy: coordinated mechanisms for the maintenance of cellular fitness. *Front Immunol* **5**, 403 (2014).
23 Keane, M. *et al.* Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep* **10**, 112-122 (2015).
24 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007).
25 Chinen, T. *et al.* The gamma-tubulin-specific inhibitor gatastatin reveals temporal requirements of microtubule nucleation during the cell cycle. *Nat Commun* **6**, 8722 (2015).
26 López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging. *Cell* **153**, 1194-1217 (2013).
27 Scheper, W. & Hoozemans, J. J. The unfolded protein response in neurodegenerative diseases: a neuropathological perspective. *Acta Neuropathol* **130**, 315-331 (2015).
28 Breda, C. *et al.* Tryptophan-2,3-dioxygenase (TDO) inhibition ameliorates neurodegeneration by modulation of kynurenine pathway metabolites. *Proc Natl Acad Sci USA* **113**, 5435-5440 (2016).
29 Sanchis-Gomar, F. *et al.* A preliminary candidate approach identifies the combination of chemerin, fetuin-A, and fibroblast growth factors 19 and 21 as a potential biomarker panel of successful aging. *Age (Dordr)* **37**, 42 (2015).
30 Pal, D. *et al.* Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *Nat Med* **18**, 1279-1285 (2012).
31 Kir, S. *et al.* FGF19 as a postprandial, insulin-independent activator of hepatic protein and glycogen synthesis. *Science* **331**, 1621-1624 (2011).
32 Morito, D. & Nagata, K. Pathogenic hijacking of ER-associated degradation: Is ERAD flexible? *Mol Cell* **59**, 335-344 (2015).
33 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361 (2017).
34 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462 (2016).
35 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
36 Quesada, V., Velasco, G., Puente, X. S., Warren, W. C. & López-Otín, C. Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals. *BMC Genomics* **11**, 220 (2010).
37 den Dunnen, J. T. *et al.* HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat* **37**, 564-569 (2016).
38 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**, 772-780 (2013).
39 Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164-1165 (2011).
40 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61**, 539-542 (2012).

41 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
42 Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
43 Hedges, S. B., Dudley, J. & Kumar, S. TimeTree: a public knowledge-base of divergence times among organisms. *Bioinformatics* **22**, 2971-2972 (2006).
44 Poulakakis, N., Russello, M., Geist, D. & Caccone, A. Unravelling the peculiarities of island life: vicariance, dispersal and the diversification of the extinct and extant giant Galapagos tortoises. *Mol Ecol* **21**, 160-173 (2012).
45 López-Otín, C. & Overall, C. M. Protease degradomics: a new challenge for proteomics. *Nat Rev Mol Cell Biol* **3**, 509-519 (2002).
46 Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* **11**, 17-30 (2010).
47 Bohmer, C., Rauhut, O. W. & Worheide, G. New insights into the vertebral Hox code of archosaurs. *Evol Dev* **17**, 258-269 (2015).
48 Di-Poi, N. *et al.* Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99-103 (2010).
49 Pearson, J. C., Lemons, D. & McGinnis, W. Modulating Hox gene functions during animal body patterning. *Nat Rev Genet* **6**, 893-904 (2005).
50 Liang, L. *et al.* Adaptive evolution of the Hox gene family for development in bats and dolphins. *PLoS One* **8**, e65944 (2013).
51 Liang, D., Wu, R., Geng, J., Wang, C. & Zhang, P. A general scenario of Hox gene inventory variation among major sarcopterygian lineages. *BMC Evol Biol* **11**, 25 (2011).
52 Crawford, N. G. *et al.* A phylogenomic analysis of turtles. *Mol Phylogen Evol* **83**, 250-257 (2015).
53 Garstang, M. & Ferrier, D. E. Time is of the essence for ParaHox homeobox gene clustering. *BMC Biol* **11**, 72 (2013).
54 Katagiri, T. & Watabe, T. Bone morphogenetic proteins. *Cold Spring Harb Perspect Biol* **8**, a021899 (2016).
55 Niehrs, C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol* **13**, 767-779 (2012).
56 Friedman, M. S., Oyserman, S. M. & Hankenson, K. D. Wnt11 promotes osteoblast maturation and mineralization through R-spondin 2. *J Biol Chem* **284**, 14117-14125 (2009).
57 Davit-Beal, T., Tucker, A. S. & Sire, J. Y. Loss of teeth and enamel in tetrapods: fossil record, genetic data and morphological adaptations. *J Anat* **214**, 477-501 (2009).
58 Meredith, R. W., Zhang, G., Gilbert, M. T. P., Jarvis, E. D. & Springer, M. S. Evidence for a single loss of mineralized teeth in the common avian ancestor. *Science* **346**, 1254390 (2014).
59 Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* **42**, D980-985 (2014).
60 Sahin, E. & Depinho, R. A. Linking functional decline of telomeres, mitochondria and stem cells during ageing. *Nature* **464**, 520-528 (2010).
61 Orr, H. A. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* **144**, 1331-1335 (1996).

62 Muftuoglu, Y., Xue, Y., Gao, X., Wu, D. & Ha, Y. Mechanism of substrate specificity of phosphatidylinositol phosphate kinases. *Proc Natl Acad Sci USA* **113**, 8711-8716 (2016).
63 Yoon, S. H. *et al.* Developmental retardation, microcephaly, and peptiduria in mice without aminopeptidase P1. *Biochem Biophys Res Commun* **429**, 204-209 (2012).
64 Hu, Y. *et al.* Identification and validation of novel CSF biomarkers for early stages of Alzheimer's disease. *Proteomics Clin Appl* **1**, 1373-1384 (2007).
65 Bellia, F., Vecchio, G. & Rizzarelli, E. Carnosinases, their substrates and diseases. *Molecules* **19**, 2299-2329 (2014).
66 Mitsui, S., Hidaka, C., Furihata, M., Osako, Y. & Yuri, K. A mental retardation gene, motopsin/prss12, modulates cell morphology by interaction with seizure-related gene 6. *Biochem Biophys Res Commun* **436**, 638-644 (2013).
67 Jiang, H.-Y. *et al.* Identification of PSEN1 mutations p. M233L and p. R352C in Han Chinese families with early-onset familial Alzheimer's disease. *Neurobiol Aging* **36**, 1602.e03-1602.e06 (2015).
68 Ryazantseva, M., Skobeleva, K., Glushankova, L. & Kaznacheyeva, E. Attenuated presenilin-1 endoproteolysis enhances store-operated calcium currents in neuronal cells. *J Neurochem* **136**, 1085-1095 (2016).
69 Alföldi, J. *et al.* The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587-591 (2011).
70 Zhang, Q. *et al.* Integrating transcriptome and genome re-sequencing data to identify key genes and mutations affecting chicken eggshell qualities. *PLoS One* **10**, e0125890 (2015).
71 Vetter, A. E. & O'Grady, S. M. Sodium and anion transport across the avian uterine (shell gland) epithelium. *J Exp Biol* **208**, 479-486 (2005).
72 Nicholls, H. Tortoise conservation: one of a kind. *Nature* **429**, 498-500 (2004).
73 Kuraku, S., Usuda, R. & Kuratani, S. Comprehensive survey of carapacial ridge-specific genes in turtle implies co-option of some regulatory genes in carapace evolution. *Evol Dev* **7**, 3-17 (2005).
74 Rieppel, O. Turtles as hopeful monsters. *Bioessays* **23**, 987-991 (2001).
75 Nagashima, H. *et al.* On the carapacial ridge in turtle embryos: its developmental origin, function and the chelonian body plan. *Development* **134**, 2219-2226 (2007).
76 Chuong, C. M. & Homberger, D. G. Development and evolution of the amniote integument: current landscape and future horizon. *J Exp Zool B Mol Dev Evol* **298**, 1-11 (2003).
77 Greenwold, M. J. *et al.* Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. *BMC Evol Biol* **14**, 249 (2014).
78 Li, Y. I., Kong, L., Ponting, C. P. & Haerty, W. Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biol Evol* **5**, 923-933 (2013).
79 Holthaus, K. B. *et al.* Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell. *Mol Biol Evol* **33**, 726-737 (2015).
80 Boyden, L. M. *et al.* Mutations in KDSR cause recessive progressive symmetric erythrokeratoderma. *Am J Hum Genet* **100**, 978-984 (2017).
81 Jaffe, A. L., Slater, G. J. & Alfaro, M. E. The evolution of island gigantism and body size variation in tortoises and turtles. *Biol Lett* **7**, 558-561 (2011).

82 Walenkamp, M. *et al.* Homozygous and heterozygous expression of a novel insulin-like growth factor-I mutation. *J Clin Endocrinol Metab* **90**, 2855-2864 (2005).
83 Brooks, A. J. & Waters, M. J. The growth hormone receptor: mechanism of activation and clinical implications. *Nat Rev Endocrinol* **6**, 515-525 (2010).
84 Rimbault, M. *et al.* Derived variants at six genes explain nearly half of size reduction in dog breeds. *Genome Res* **23**, 1985-1995 (2013).
85 Jepsen, M. R. *et al.* Stanniocalcin-2 inhibits mammalian growth by proteolytic inhibition of the insulin-like growth factor axis. *J Biol Chem* **290**, 3430-3439 (2015).
86 Caccone, A., Gibbs, J. P., Ketmaier, V., Suatoni, E. & Powell, J. R. Origin and evolutionary relationships of giant Galapagos tortoises. *Proc Natl Acad Sci USA* **96**, 13223-13228 (1999).
87 Fain, G. L., Hardie, R. & Laughlin, S. B. Phototransduction and the evolution of photoreceptors. *Curr Biol* **20**, R114-R124 (2010).
88 Davies, W. I., Collin, S. P. & Hunt, D. M. Molecular ecology and adaptation of visual photopigments in craniates. *Mol Ecol* **21**, 3121-3158 (2012).
89 Feng, P., Zheng, J., Rossiter, S. J., Wang, D. & Zhao, H. Massive losses of taste receptor genes in toothed and baleen whales. *Genome Biol Evol* **6**, 1254-1265 (2014).
90 Zhao, H., Li, J. & Zhang, J. Molecular evidence for the loss of three basic tastes in penguins. *Curr Biol* **25**, R141-R142 (2015).
91 Go, Y. Lineage-specific expansions and contractions of the bitter taste receptor gene repertoire in vertebrates. *Mol Biol Evol* **23**, 964-972 (2006).
92 Li, D. & Zhang, J. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol Biol Evol* **31**, 303-309 (2014).
93 Niimura, Y. & Nei, M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proc Natl Acad Sci USA* **102**, 6039-6044 (2005).
94 Niimura, Y. & Nei, M. Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLoS One* **2**, e708 (2007).
95 Barreiro, L. B. & Quintana-Murci, L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet* **11**, 17-30 (2010).
96 Zimmerman, L., Vogel, L. & Bowden, R. Understanding the vertebrate immune system: insights from the reptilian perspective. *J Exp Biol* **213**, 661-671 (2010).
97 Chiari, Y., Cahais, V., Galtier, N. & Delsuc, F. Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol* **10**, 65 (2012).
98 Miller, H. C. *et al.* Major Histocompatibility Complex genes map to two chromosomes in an evolutionarily ancient reptile, the tuatara Sphenodon punctatus. *G3 (Bethesda)* **5**, 1439-1451 (2015).
99 Hirano, T. *et al.* Complementary DNA for a novel human interleukin (BSF-2) that induces B lymphocytes to produce immunoglobulin. *Nature* **324**, 73-76 (1986).
100 Sun, Q., Jin, H. J. & Bond, J. S. Disruption of the meprin α and β genes in mice alters homeostasis of monocytes and natural killer cells. *Exp Hematol* **37**, 346-356 (2009).
101 Dotiwala, F. *et al.* Killer lymphocytes use granulysin, perforin and granzymes to kill intracellular parasites. *Nat Med* **22**, 210-216 (2016).
102 Akula, S., Thorpe, M., Boinapally, V. & Hellman, L. Granule associated serine proteases of hematopoietic cells – An analysis of their appearance and diversification during vertebrate evolution. *PLoS One* **10**, e0143091 (2015).

103 Puente, X. S., Sanchez, L. M., Overall, C. M. & López-Otín, C. Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet* **4**, 544-558 (2003).
104 Knisbacher, B. A., Gerber, D. & Levanon, E. Y. DNA editing by APOBECs: A genomic preserver and transformer. *Trends Genet* **32**, 16-28 (2016).
105 Petit, V. *et al.* Murine APOBEC1 is a powerful mutator of retroviral and cellular RNA in vitro and in vivo. *J Mol Biol* **385**, 65-78 (2009).
106 Bishop, K. N., Holmes, R. K., Sheehy, A. M. & Malim, M. H. APOBEC-mediated editing of viral RNA. *Science* **305**, 645 (2004).
107 Hancock, R. E., Haney, E. F. & Gill, E. E. The immunology of host defence peptides: beyond antimicrobial activity. *Nat Rev Immunol* **16**, 321-334 (2016).
108 Peric, M. *et al.* IL-17A enhances vitamin D3-induced expression of cathelicidin antimicrobial peptide in human keratinocytes. *J Immunol* **181**, 8504-8512 (2008).
109 Castaneda-Delgado, J. E. *et al.* Differences in cytokine production during aging and its relationship with antimicrobial peptides production. *Immunol Invest* **46**, 48-58 (2017).
110 Di Rosa, M., Distefano, G., Zorena, K. & Malaguarnera, L. Chitinases and immunity: ancestral molecules with new functions. *Immunobiology* **221**, 399-411 (2016).
111 Drage, M. G. *et al.* TLR2 and its co-receptors determine responses of macrophages and dendritic cells to lipoproteins of Mycobacterium tuberculosis. *Cell Immunol* **258**, 29-37 (2009).
112 Oldenburg, M. *et al.* TLR13 recognizes bacterial 23S rRNA devoid of erythromycin resistance-forming modification. *Science* **337**, 1111-1115 (2012).
113 Roach, J. C. *et al.* The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* **102**, 9577-9582 (2005).
114 Schultz, U., Kaspers, B. & Staeheli, P. The interferon system of non-mammalian vertebrates. *Dev Comp Immunol* **28**, 499-508 (2004).
115 Fu, J. P. *et al.* IFN-γ in turtle: Conservation in sequence and signalling and role in inhibiting iridovirus replication in Chinese soft-shelled turtle Pelodiscus sinensis. *Dev Comp Immunol* **43**, 87-95 (2014).
116 Harwig, S. S. *et al.* Gallinacins: cysteine-rich antimicrobial peptides of chicken leukocytes. *FEBS Lett* **342**, 281-285 (1994).
117 Ishii, A., Kawasaki, M., Matsumoto, M., Tochinai, S. & Seya, T. Phylogenetic and expression analysis of amphibian Xenopus Toll-like receptors. *Immunogenetics* **59**, 281-293 (2007).
118 Jault, C., Pichon, L. & Chluba, J. Toll-like receptor gene family and TIR-domain adapters in Danio rerio. *Mol Immunol* **40**, 759-771 (2004).
119 Keestra, A. M., de Zoete, M. R., Bouwman, L. I., Vaezirad, M. M. & van Putten, J. P. Unique features of chicken Toll-like receptors. *Dev Comp Immunol* **41**, 316-323 (2013).
120 Guo, H., Callaway, J. B. & Ting, J. P. Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nat Med* **21**, 677-687 (2015).
121 Tian, X., Pascal, G. & Monget, P. Evolution and functional divergence of NLRP genes in mammalian reproductive systems. *BMC Evol Biol* **9**, 202 (2009).
122 Hervella, M. *et al.* The loss of functional caspase-12 in Europe is a pre-neolithic event. *PLoS One* **7**, e37022 (2012).
123 Jaratlerdsiri, W., Isberg, S. R., Higgins, D. P. & Gongora, J. MHC class I of saltwater crocodiles (Crocodylus porosus): polymorphism and balancing selection. *Immunogenetics* **64**, 825-838 (2012).

124 Acevedo-Whitehouse, K. & Cunningham, A. A. Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol Evol* **21**, 433-438 (2006).
125 Yang, Z. *et al.* Analysis of the reptile CD1 genes: evolutionary implications. *Immunogenetics* **67**, 337-346 (2015).
126 Elbers, J. P. & Taylor, S. S. Major histocompatibility complex polymorphism in reptile conservation. *Herpetol Conserv Biol* **11**, 1-12 (2015).
127 Monzon-Arguello, C., Garcia de Leaniz, C., Gajardo, G. & Consuegra, S. Less can be more: loss of MHC functional diversity can reflect adaptation to novel conditions during fish invasions. *Ecol Evol* **3**, 3359-3368 (2013).
128 Garrick, R. C. *et al.* Naturally rare versus newly rare: demographic inferences on two timescales inform conservation of Galapagos giant tortoises. *Ecol Evol* **5**, 676-694 (2015).
129 Deakin, J. E. *et al.* Evolution and comparative analysis of the MHC Class III inflammatory region. *BMC Genomics* **7**, 281 (2006).
130 Xie, T. *et al.* Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse. *Genome Res* **13**, 2621-2636 (2003).
131 Shiina, T. *et al.* Comparative genomic analysis of two avian (quail and chicken) MHC regions. *J Immunol* **172**, 6751-6763 (2004).
132 López-Otín, C., Galluzzi, L., Freije, J. M., Madeo, F. & Kroemer, G. Metabolic control of longevity. *Cell* **166**, 802-821 (2016).
133 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
134 Cavalcanti, D. M. *et al.* Neurolysin knockout mice generation and initial phenotype characterization. *J Biol Chem* **289**, 15426-15440 (2014).
135 M't Hart, L. *et al.* The CTRB1/2 locus affects diabetes susceptibility and treatment via the incretin pathway. *Diabetes* **62**, 3275-3281 (2013).
136 Rosen, E. D. & Spiegelman, B. M. Adipocytes as regulators of energy balance and glucose homeostasis. *Nature* **444**, 847-853 (2006).
137 Harwood, A. J. Regulation of GSK-3: a cellular multiprocessor. *Cell* **105**, 821-824 (2001).
138 Williams, A. S. *et al.* Integrin α1-null mice exhibit improved fatty liver when fed a high fat diet despite severe hepatic insulin resistance. *J Biol Chem* **290**, 6546-6557 (2015).
139 Fan, C. *et al.* MIF intersubunit disulfide mutant antagonist supports activation of CD74 by endogenous MIF trimer at physiologic concentrations. *Proc Natl Acad Sci USA* **110**, 10994-10999 (2013).
140 Verschuren, L. *et al.* MIF deficiency reduces chronic inflammation in white adipose tissue and impairs the development of insulin resistance, glucose intolerance, and associated atherosclerotic disease. *Circul Res* **105**, 99-107 (2009).
141 Tornberg, R., Liuska, L., Rytkönen, S., Mutanen, M. & Välimäki, P. Diet shift induced rapid evolution of size and function in a predatory bird. *Oecologia* **176**, 781-788 (2014).
142 Drouin, G., Godin, J.-R. & Pagé, B. The genetics of vitamin C loss in vertebrates. *Curr Genomics* **12**, 371-378 (2011).
143 Yang, H. Conserved or lost: molecular evolution of the key gene GULO in vertebrate vitamin C biosynthesis. *Biochem Genet* **51**, 413-425 (2013).
144 Kim, S. *et al.* Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome Biol* **17**, 211 (2016).

145 Wang, Y. *et al.* Snake cathelicidin from Bungarus fasciatus is a potent peptide antibiotics. *PLoS One* **3**, e3217 (2008).
146 Chen, Y. *et al.* As-CATH1-6, novel cathelicidins with potent antimicrobial and immunomodulatory properties from Alligator sinensis, play pivotal roles in host antimicrobial immune responses. *Biochem J* **474**, 2861-2885 (2017).
147 Peel, E. *et al.* Cathelicidins in the Tasmanian devil (Sarcophilus harrisii). *Sci Rep* **6**, 35019 (2016).
148 Elsik, C. G. *et al.* The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**, 522-528 (2009).
149 Brown-Borg, H. M. Hormonal regulation of longevity in mammals. *Ageing Res Rev* **6**, 28-45 (2007).
150 Claus, M., Neumann, S., Kleinau, G., Krause, G. & Paschke, R. Structural determinants for G-protein activation and specificity in the third intracellular loop of the thyroid-stimulating hormone receptor. *J Mol Med (Berl)* **84**, 943-954 (2006).
151 Cheng, S. *et al.* What makes a 97-year-old man cycle 5,000 km a year? *Gerontology* **62**, 508-512 (2016).
152 Atzmon, G., Barzilai, N., Hollowell, J. G., Surks, M. I. & Gabriely, I. Extreme longevity is associated with increased serum thyrotropin. *J Clin Endocrinol Metab* **94**, 1251-1254 (2009).
153 Krivoruchko, A. & Storey, K. B. Turtle anoxia tolerance: biochemistry and gene regulation. *Biochim Biophys Acta* **1850**, 1188-1196 (2015).
154 Burmester, T. & Hankeln, T. Function and evolution of vertebrate globins. *Acta Physiol* **211**, 501-514 (2014).
155 Schwarze, K., Singh, A. & Burmester, T. The full globin repertoire of turtles provides insights into vertebrate globin evolution and functions. *Genome Biol Evol* **7**, 1896-1913 (2015).
156 Storz, J. F., Opazo, J. C. & Hoffmann, F. G. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol Phylogen Evol* **66**, 469-478 (2013).
157 Corti, P. *et al.* Globin X is a six-coordinate globin that reduces nitrite to nitric oxide in fish red blood cells. *Proc Natl Acad Sci USA* **113**, 8538-8543 (2016).
158 Jacobsen, S. B. *et al.* Circulating nitric oxide metabolites and cardiovascular changes in the turtle Trachemys scripta during normoxia, anoxia and reoxygenation. *J Exp Biol* **215**, 2560-2566 (2012).
159 Fago, A. & Jensen, F. B. Hypoxia tolerance, nitric oxide, and nitrite: lessons from extreme animals. *Physiology* **30**, 116-126 (2015).
160 Lutz, P. L., Prentice, H. M. & Milton, S. L. Is turtle longevity linked to enhanced mechanisms for surviving brain anoxia and reoxygenation? *Exp Gerontol* **38**, 797-800 (2003).
161 Al-Hilali, A. *et al.* Analysis of the novel factor X gene mutation Glu51Lys in two families with factor X-Riyadh anomaly. *Thromb Haemost* **97**, 542-545 (2007).
162 Quélin, F. *et al.* Identification of five novel mutations in the factor XI gene (F11) of patients with factor XI deficiency. *Blood Coagul Fibrinolysis* **17**, 69-73 (2006).
163 Wong, M. K. & Takei, Y. Lack of plasma kallikrein-kinin system cascade in teleosts. *PLoS One* **8**, e81057 (2013).

164 Khan, S. Q. *et al.* C-terminal pro–endothelin-1 offers additional prognostic information in patients after acute myocardial infarction: Leicester Acute Myocardial Infarction Peptide (LAMP) Study. *Am Heart J* **154**, 736-742 (2007).
165 Muller, P. A. & Vousden, K. H. p53 mutations in cancer. *Nat Cell Biol* **15**, 2-8 (2013).
166 Zhao, Y. *et al.* Codon 104 variation of p53 gene provides adaptive apoptotic responses to extreme environments in mammals of the Tibet plateau. *Proc Natl Acad Sci USA* **110**, 20639-20644 (2013).
167 Orsted, D. D., Bojesen, S. E., Tybjaerg-Hansen, A. & Nordestgaard, B. G. Tumor suppressor p53 Arg72Pro polymorphism and longevity, cancer survival, and risk of cancer in the general population. *J Exp Med* **204**, 1295-1301 (2007).
168 Heikkila, J. J. The expression and function of hsp30-like small heat shock protein genes in amphibians, birds, fish, and reptiles. *Comp Biochem Physiol A Mol Integr Physiol* **203**, 179-192 (2017).
169 Hegde, M. L. *et al.* Prereplicative repair of oxidized bases in the human genome is mediated by NEIL1 DNA glycosylase together with replication proteins. *Proc Natl Acad Sci USA* **110**, e3090-e3099 (2013).
170 MacRae, S. L. *et al.* DNA repair in species with extreme lifespan differences. *Aging (Albany NY)* **7**, 1171-1184 (2015).
171 Shinmura, K. *et al.* Abnormal expressions of DNA glycosylase genes NEIL1, NEIL2, and NEIL3 are associated with somatic mutation loads in human cancer. *Oxid Med Cell Longev* **2016**, 1546392 (2016).
172 Daley, J. M., Chiba, T., Xue, X., Niu, H. & Sung, P. Multifaceted role of the Topo IIIα–RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection. *Nucleic Acids Res* **42**, 11083-11091 (2014).
173 Walker, E. H., Perisic, O., Ried, C., Stephens, L. & Williams, R. L. Structural insights into phosphoinositide 3-kinase catalysis and signalling. *Nature* **402**, 313-320 (1999).
174 Kragelund, B. B., Weterings, E., Hartmann-Petersen, R. & Keijzers, G. The Ku70/80 ring in Non-Homologous End-Joining: easy to slip on, hard to remove. *Front Biosci (Landmark Ed)* **21**, 514-527 (2016).
175 Pace, P. *et al.* Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. *Science* **329**, 219-223 (2010).
176 Hendriks, I. A. *et al.* Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**, 927-936 (2014).
177 Kim, E. B. *et al.* Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* **479**, 223-227 (2011).
178 Cremona, C. A. *et al.* Extensive DNA damage-induced sumoylation contributes to replication and repair and acts in addition to the mec1 checkpoint. *Mol Cell* **45**, 422-432 (2012).
179 Caulin, A. F. & Maley, C. C. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol Evol* **26**, 175-182 (2011).
180 Futreal, P. A. *et al.* A census of human cancer genes. *Nat Rev Cancer* **4**, 177-183 (2004).
181 Moz, S. *et al.* SMAD4 loss enables EGF, TGFβ1 and S100A8/A9 induced activation of critical pathways to invasion in human pancreatic adenocarcinoma cells. *Oncotarget* **7**, 69927-69944 (2016).
182 Petrilli, A. & Fernandez-Valle, C. Role of Merlin/NF2 inactivation in tumor biology. *Oncogene* **35**, 537-548 (2016).

183 Hill, C. S. Transcriptional control by the SMADs. *Cold Spring Harb Perspect Biol* **8**, a022079 (2016).
184 Cooper, J. & Giancotti, F. G. Molecular insights into NF2/Merlin tumor suppressor function. *FEBS Lett* **588**, 2743-2752 (2014).
185 Katagiri, K., Maeda, A., Shimonaka, M. & Kinashi, T. RAPL, a Rap1-binding molecule that mediates Rap1-induced adhesion through spatial regulation of LFA-1. *Nat Immunol* **4**, 741-748 (2003).
186 Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883-899 (2010).
187 Street, S. E. *et al.* Host perforin reduces tumor number but does not increase survival in oncogene-driven mammary adenocarcinoma. *Cancer Res* **67**, 5454-5460 (2007).
188 Bolitho, P. *et al.* Perforin-mediated suppression of B-cell lymphoma. *Proc Natl Acad Sci USA* **106**, 2723-2728 (2009).
189 Smyth, M. J. *et al.* Perforin-mediated cytotoxicity is critical for surveillance of spontaneous lymphoma. *J Exp Med* **192**, 755-760 (2000).
190 Lin, Y., Wong, K.-k. & Calame, K. Repression of c-myc transcription by Blimp-1, an inducer of terminal B cell differentiation. *Science* **276**, 596-599 (1997).
191 Küçük, C. *et al.* PRDM1 is a tumor suppressor gene in natural killer cell malignancies. *Proc Natl Acad Sci USA* **108**, 20119-20124 (2011).
192 Garner, M. M., Hernandez-Divers, S. M. & Raymond, J. T. Reptile neoplasia: a retrospective study of case submissions to a specialty diagnostic service. *Vet Clin North Am Exot Anim Pract* **7**, 653-671 (2004).
193 Abegglen, L. M. *et al.* Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA* **314**, 1850-1860 (2015).
194 Knoepfler, P. S., Cheng, P. F. & Eisenman, R. N. N-myc is essential during neurogenesis for the rapid expansion of progenitor cell populations and the inhibition of neuronal differentiation. *Genes Dev* **16**, 2699-2712 (2002).
195 Li, J. *et al.* Stable SET knockdown in breast cell carcinoma inhibits cell migration and invasion. *Biochem Biophys Res Commun* **453**, 7-12 (2014).
196 Martinvalet, D., Zhu, P. & Lieberman, J. Granzyme A induces caspase-independent mitochondrial damage, a required first step for apoptosis. *Immunity* **22**, 355-370 (2005).
197 Tang, L.-J. *et al.* USP12 regulates cell cycle progression by involving c-Myc, cyclin D2 and BMI-1. *Gene* **578**, 92-99 (2016).
198 Li, J., Lu, Q. & Zhang, M. Structural basis of cargo recognition by unconventional myosins in cellular trafficking. *Traffic* **17**, 822-838 (2016).
199 Espreafico, E. M. *et al.* Primary structure and cellular localization of chicken brain myosin-V (p190), an unconventional myosin with calmodulin light chains. *J Cell Biol* **119**, 1541-1557 (1992).
200 Schramek, D. *et al.* Direct in vivo RNAi screen unveils myosin IIa as a tumor suppressor of squamous cell carcinomas. *Science* **343**, 309-313 (2014).
201 Pylayeva-Gupta, Y., Grabocka, E. & Bar-Sagi, D. RAS oncogenes: weaving a tumorigenic web. *Nat Rev Cancer* **11**, 761-774 (2011).
202 Baker, N. M. & Der, C. J. Cancer: Drug for an 'undruggable' protein. *Nature* **497**, 577-578 (2013).

203 Veitia, R. A. Dominant negative factors in health and disease. *J Pathol* **218**, 409-418 (2009).
204 Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. *Genome Res* **20**, 1313-1326 (2010).
205 Harvey, K. F., Zhang, X. & Thomas, D. M. The Hippo pathway and human cancer. *Nat Rev Cancer* **13**, 246-257 (2013).
206 Hong, J.-H. *et al.* TAZ, a transcriptional modulator of mesenchymal stem cell differentiation. *Science* **309**, 1074-1078 (2005).
207 Mukherjee, M. *et al.* Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. *EMBO J* **31**, 1308-1319 (2012).
208 Yin, F. *et al.* Spatial organization of Hippo signaling at the plasma membrane mediated by the tumor suppressor Merlin/NF2. *Cell* **154**, 1342-1355 (2013).
209 Bourn, D. & Coe, M. J. Features of tortoise mortality and decomposition on Aldabra. *Philos Trans R Soc Lond B Biol Sci* **286**, 189-193 (1979).
210 Grubb, P. The growth, ecology and population structure of giant tortoises on Aldabra. *Philos Trans R Soc Lond B Biol Sci* **260**, 327 (1971).
211 Curtin, A. J., Zug, G. R. & Spotila, J. R. Longevity and growth strategies of the desert tortoise (Gopherus agassizii) in two American deserts. *J Arid Environ* **73**, 463-471 (2009).
212 Ye, J. *et al.* TRF2 and Apollo cooperate with topoisomerase 2α to protect human telomeres from replicative damage. *Cell* **142**, 230-242 (2010).
213 Tong, A. S. *et al.* ATM and ATR signaling regulate the recruitment of human telomerase to telomeres. *Cell Rep* **13**, 1633-1646 (2015).
214 Wang, Y., Ghosh, G. & Hendrickson, E. A. Ku86 represses lethal telomere deletion events in human somatic cells. *Proc Natl Acad Sci USA* **106**, 12430-12435 (2009).
215 Ribes-Zamora, A., Indiviglio, S. M., Mihalek, I., Williams, C. L. & Bertuch, A. A. TRF2 interaction with Ku heterotetramerization interface gives insight into c-NHEJ prevention at human telomeres. *Cell Rep* **5**, 194-206 (2013).
216 Kaushik, S. & Cuervo, A. M. Proteostasis and aging. *Nat Med* **21**, 1406-1415 (2015).
217 Shikama, N., Ackermann, R. & Brack, C. Protein synthesis elongation factor EF-1 alpha expression and longevity in Drosophila melanogaster. *Proc Natl Acad Sci USA* **91**, 4199-4203 (1994).
218 Castillo-Quan, J. I. *et al.* Lithium promotes longevity through GSK3/NRF2-dependent hormesis. *Cell Rep* **15**, 638-650 (2016).
219 Waterham, H. R., Ferdinandusse, S. & Wanders, R. J. Human disorders of peroxisome metabolism and biogenesis. *Biochim Biophys Acta* **1863**, 922-933 (2016).
220 Steinmetz, C. G., Xie, P., Weiner, H. & Hurley, T. D. Structure of mitochondrial aldehyde dehydrogenase: the genetic component of ethanol aversion. *Structure* **5**, 701-711 (1997).
221 Moore, S. A. *et al.* Sheep liver cytosolic aldehyde dehydrogenase: the structure reveals the basis for the retinal specificity of class 1 aldehyde dehydrogenases. *Structure* **6**, 1541-1551 (1998).
222 Ohta, S., Ohsawa, I., Kamino, K., Ando, F. & Shimokata, H. Mitochondrial ALDH2 deficiency as an oxidative stress. *Ann N Y Acad Sci* **1011**, 36-44 (2004).
223 Carrigan, M. A. *et al.* Hominids adapted to metabolize ethanol long before human-directed fermentation. *Proc Natl Acad Sci USA* **112**, 458-463 (2015).

224 Serizawa, A., Dando, P. M. & Barrett, A. J. Characterization of a mitochondrial metallopeptidase reveals neurolysin as a homologue of thimet oligopeptidase. *J Biol Chem* **270**, 2092-2098 (1995).
225 Tristan, C., Shahani, N., Sedlak, T. W. & Sawa, A. The diverse functions of GAPDH: views from different subcellular compartments. *Cell Signal* **23**, 317-323 (2011).
226 Weitz-Schmidt, G., Schürpf, T. & Springer, T. A. The C-terminal αI domain linker as a critical structural element in the conformational activation of αI integrins. *J Biol Chem* **286**, 42115-42122 (2011).
227 Salminen, A., Kaarniranta, K. & Kauppinen, A. Inflammaging: disturbed interplay between autophagy and inflammasomes. *Aging (Albany NY)* **4**, 166-175 (2012).
228 Harper, J. M., Wilkinson, J. E. & Miller, R. A. Macrophage migration inhibitory factor-knockout mice are long lived and respond to caloric restriction. *FASEB J* **24**, 2436-2442 (2010).
229 Clemmons, D. R. Role of IGF binding proteins in regulating metabolism. *Trends Endocrinol Metab* **27**, 375-391 (2016).
230 Kenyon, C. J. The genetics of ageing. *Nature* **464**, 504-512 (2010).
231 Brohus, M., Gorbunova, V., Faulkes, C. G., Overgaard, M. T. & Conover, C. A. The insulin-like growth factor system in the long-lived naked mole-rat. *PLoS One* **10**, e0145587 (2015).
232 Soerensen, M. *et al.* Human longevity and variation in GH/IGF-1/insulin signaling, DNA damage signaling and repair and pro/antioxidant pathway genes: cross sectional and longitudinal studies. *Exp Gerontol* **47**, 379-387 (2012).

11. Annexes

Table S16. Summary of average characteristics of the automatic annotation set of *C. abingdonii* compared to sets of other species. This table is attached as *SuppTables.xlsx* (sheet A1).

Table S17. Genes with signatures of positive selection in *C. abingdonii*. This table is attached as *SuppTables.xlsx* (sheet A2).

Table S18. Positions of interest within the genes with signatures of positive selection. This table is attached as *SuppTables.xlsx* (sheet A3).

Table S19. microRNAs identified in *C. abingdonii*'s genome. This table is attached as *SuppTables.xlsx* (sheet A4).

Table S20. Primers used for validation of gene alterations identified in *C. abingdonii*'s genome. This table is attached as *SuppTables.xlsx* (sheet A5).

Table S21. Expansion events with coverage and identity for nucleotide and amino acid sequences. Each row shows the comparison between the firstly identified copy and the rest of them, consecutively. This table is attached as *SuppTables.xlsx* (sheet A6).