

Published in final edited form as:

Nat Ecol Evol. 2019 January ; 3(1): 87–95. doi:10.1038/s41559-018-0733-x.

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

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

<sup>1</sup>Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología del Principado de Asturias, Ciberonc, Universidad de Oviedo, Oviedo, Spain <sup>2</sup>Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA <sup>3</sup>Institute for Genomics and Systems Biology, The University of Chicago, Chicago, Illinois 60637, USA <sup>4</sup>Galapagos National Park Directorate, Galapagos Islands, Ecuador <sup>5</sup>Galapagos Conservancy, Fairfax, VA, USA <sup>6</sup>Institute of Evolutionary Biology (UPF-CSIC), PRBB, 08003 Barcelona, Spain <sup>7</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain <sup>8</sup>College of Science and Engineering, Flinders University, Adelaide SA 5001, Australia <sup>9</sup>College of Environmental Science & Forestry, State University of New York,

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

<sup>^</sup>**Corresponding authors:** Carlos López-Otín; Departamento de Bioquímica y Biología Molecular, Facultad de Medicina, Universidad de Oviedo; 33006 Oviedo-Spain. Tel. 34-985-104201; Fax: 34-985-103564; clo@uniovi.es; Adalgisa Caccone; Department of Ecology and Evolutionary Biology, YIBS-MSCG Center, Yale University, New Haven, USA; Tel. 1-203-432-5259; Fax 1-203-432-7394; adalgisa.caccone@yale.edu.

### Data availability

Data supporting the findings of this study are available within the paper and its supplementary information files. Sequencing data have been deposited at the Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>) with BioProject accession number PRJNA416050. The accession number of the assembled genomic sequence is PKMU00000000. MAKER2-predicted protein sequences can be downloaded from <https://github.com/vqf/LG>.

### Code availability

The scripts for manual annotation (BATI) can be accessed at <http://degradome.uniovi.es/downloads.html>. Custom scripts used to produce multiple alignments for positive selection and copy-number studies are freely available at <https://github.com/vqf/LG>.

**Author Contributions.** V.Q. and J.G.P.-S. performed the automatic analysis of genomes; S.F.-R. coordinated the manual genomic annotation which was performed by J.G.P.-S., O.S.-F., D.C.-I., M.G.A., M.A.-V., D.C., P.M., J.R.A., I.T.-G., D.R.-V. and M.P.-T.; S.F.-R. and D.C.-I. performed the validation of the identified genomic variants; G.B. coordinated the functional analyses of the identified genomic variants, which were carried out by O.S.-F., D.C.-I., M.G.A., M.A.-V., D.C., P.M., J.R.A., and I.T.-G.; J.M. helped in the screen of the wild samples for SNP validation and contributed to results interpretation; M.Q., L.B.B., J.P.G., Y.C., S.G., C.C., B.R.E., S.J.G., D.L.E., R.C.G., M.A.R. and N.P. contributed to early data collection and analyses; W.T., D.O.R., and J.P.G. helped to obtain material securing permits and biological samples; K.P.W. partly supported data collection and supervised initial analysis; Z.F.J. prepared DNA and RNA samples for genomic analyses and conducted raw data quality check; L.F.K.K. and T.M.-B. performed population history and diversity studies; V.Q., A.C., and C.L.-O. directed the research, analysed the data and wrote the manuscript.

**Author Information.** The authors declare no competing financial interests.

Syracuse, New York, 13210, USA <sup>10</sup>Department of Biology, University of South Alabama, Mobile, AL 36688, USA <sup>11</sup>Department of Biology, University of Florence, 50019 Sesto Fiorentino (FI), Italy <sup>12</sup>School of Natural Sciences, University of California, Merced, CA 95343, USA <sup>13</sup>Department of Biology, University of Mississippi, Oxford, MS 38677-1848, USA <sup>14</sup>Department of Biology, The University of British Columbia, Kelowna, BC, V1V 1V7 Canada <sup>15</sup>Department of Biology, School of Sciences and Engineering, University of Crete, Heraklion, Crete, GR-71409, Greece <sup>16</sup>Natural History Museum of Crete, Heraklion, Crete, GR-71409, Greece <sup>17</sup>Catalan Institution of Research and Advanced Studies (ICREA), 08010, Barcelona, Spain <sup>18</sup>Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, 08193, Barcelona, Spain

# These authors contributed equally to this work.

## Abstract

Giant tortoises are amongst the longest-lived vertebrate animals and as such provide an excellent model to study traits like longevity and age-related diseases. However, genomic and molecular evolutionary information on giant tortoises is scarce. Here, we describe a global analysis of the genomes of Lonesome George, the iconic last member of *Chelonoidis abingdonii*, and the Aldabra giant tortoise (*Aldabrachelys gigantea*). The comparison of these genomes to those of related species, using both unsupervised and supervised analyses, led us to detect lineage-specific variants affecting DNA repair genes, inflammatory mediators and genes related to cancer development. Our study also hints at specific evolutionary strategies linked to increased lifespan and expands our understanding of the genomic determinants of ageing. These new genome sequences also provide important resources to help the efforts for restoration of giant tortoise populations.

## Introduction

Comparative genomic analyses leverage the mechanisms of natural selection to find genes and biochemical pathways related to complex traits and processes. Multiple works have used these techniques with the genomes of long-lived mammals to shed light on the signalling and metabolic networks that might play a role in regulating age-related conditions<sup>1,2</sup>. Similar studies on unrelated longevous organisms might unveil novel evolutionary strategies and genetic determinants of ageing in different environments. In this regard, giant tortoises constitute one of the few groups of vertebrates with an exceptional longevity, in excess of 100 years according to some estimates.

In this manuscript, we report the genomic sequencing and comparative genomic analysis of two long-lived giant tortoises: Lonesome George – the last representative of *C. abingdonii*<sup>3</sup>, endemic to the island of Pinta (Galapagos Islands, Ecuador) – and an individual of *A. gigantea*, endemic to the Aldabra Atoll and the only extant species of giant tortoises in the Indian Ocean<sup>4</sup> (Fig. 1a). Unsupervised and supervised comparative analyses of these genomic sequences adds new genetic information on the evolution of turtles, and provides novel candidate genes that might underlie the extraordinary characteristics of giant tortoises, including their gigantism and longevity.

## Results and discussion

The genome of Lonesome George was sequenced using a combination of Illumina and PacBio platforms (Supplementary Information, Section 1.1). The assembled genome (CheloAbing 1.0) has a genomic size of 2.3 Gb, and contains 10,623 scaffolds with an N50 of 1.27 Mb (Supplementary Information, Section 1.1, Tables S1-S3). We also sequenced with the Illumina platform the closely related tortoise *A. gigantea* at an average read depth of 28x. These genomic sequences were aligned to CheloAbing 1.0.

TimeTree database estimations (<http://www.timetree.org>) indicate that Galapagos and Aldabra giant tortoises shared a last common ancestor about 40 million years ago, while both diverged from the human lineage more than 300 million years ago (Supplementary Information, Section 1.4). A preliminary analysis of demographic history using the pairwise sequentially Markovian coalescent (PSMC)<sup>5</sup> model showed that while the effective population size of *C. abingdonii* has been steadily declining for the past million years, with a slight uptick about 90,000 years ago, the population of Aldabra giant tortoises experienced substantial fluctuations over this period (Fig. 1b). Effective population size reconstructions for *C. abingdonii* lose statistical power at the million-year time frame, probably due to complete coalescence. In turn, this suggests that overall diversity in these giant tortoises must have been low throughout many generations. Altogether, these results prompt us to propose that the populations of these insular giant tortoises were vulnerable at the time of human discovery of the Galapagos Islands, likely elevating their extinction risk.

By using homology searches with known gene sets from humans and *Pelodiscus sinensis* (Chinese soft-shell turtle), along with RNA-Seq data from *C. abingdonii* blood and from an *A. gigantea* granuloma, we automatically predicted a primary set of 27,208 genes from the genome assembly using the MAKER2 algorithm<sup>6</sup>. We then performed pairwise alignments between each of the primary predicted protein sequences and the Uniprot databases for humans and *P. sinensis*, whose annotated sequences show a relatively high quality when compared to data available for other turtles<sup>7</sup>. Using alignments spanning at least 80% of the longest protein and showing more than 60% identity, we constructed sets of protein families shared among these species. This preliminary analysis singled out several protein families that seem to have undergone moderate expansion in a common ancestor of *C. abingdonii* and *A. gigantea*. Almost all of these expansions were also confirmed in the genome of the related, long-lived tortoise *Gopherus agassizii* (Supplementary Information, Section 1.2, Table S4). Most of these genes have been linked to exosome formation, suggesting that this process may have been important in tortoise evolution.

We also interrogated the predicted gene set for evidence of positive selection in giant tortoises. This analysis singled out 43 genes with evidence of giant tortoise-specific positive selection (Supplementary Information, Section 1.2, Table S5, Fig. S1). This list includes genes with known roles in the dynamics of the tubulin cytoskeleton (*TUBE1* and *TUBG1*) and in intracellular vesicle trafficking (*VPS35*). Importantly, the analysis of genes showing evidence of positive selection also includes *AHSG* and *FGF19*, whose expression levels have been linked with successful ageing in humans<sup>8</sup>. The role of both factors in metabolism regulation<sup>9,10</sup> – another hallmark of ageing<sup>11,12</sup> – suggests that the specific changes

observed in these proteins may have arisen in order to accommodate the challenges that longevity poses on this system. The list of genes with signatures of positive selection also features *TDO2*, whose inhibition has been proposed to protect against age-related diseases through regulation of tryptophan-mediated proteostasis<sup>13</sup>. In addition, we found evidence for positive selection affecting several genes involved in immune system modulation, like *MVK*, *IRAK1BP1* and *IL1R2*. Taken together, these results identify proteostasis, metabolism regulation and immune response as key processes during the evolution of giant tortoises *via* effects on longevity and resistance to infection.

Parallel to this automatic analysis, we used manually-supervised annotation on more than 3,000 genes selected *a priori* for a series of hypothesis-driven studies on development, physiology, immunity, metabolism, stress response, cancer susceptibility and longevity (Supplementary Information, Section 1.3, Fig. S2). We searched for truncating variants, variants affecting known motifs and variants whose human counterparts are related to known genetic diseases (Supplementary Information, Section 1.3, Table S6). These variants were first confirmed with the RNA-Seq data. Then, more than 100 of the most interesting variants in terms of putative functional relevance were also validated by PCR amplification followed by Sanger sequencing. To this end, we used a panel of genomic DNA samples of eleven different species of giant tortoises endemic to different islands from the Galapagos Archipelago (Supplementary Information, Section 1, Table S7, Fig. S3).

The manually-supervised annotation of development-related genes showed the complete conservation of the Hox gene set among giant tortoises, with the exception of *HOXC3*, which seems to have been lost in the radiation of Archelosauria<sup>14,15</sup> (Supplementary Information, Section 2, Table S8, Fig. S4). *BMP* and *GDF* gene families were also found to be conserved, although the duplication event that gave rise to *GDF1* and *GDF3* in mammals did not occur in turtles, birds and crocodiles. By contrast, we found a duplication of the ParaHox gene *CDX4* in giant tortoises, also present in other reptiles as well as in avian reptiles (birds). This annotation also showed the duplication of *WNT11* in turtles and chicken (but not in the lizard *Anolis carolinensis*) and the specific duplication of *WNT4* in turtles. Given the roles of these duplicated genes and their conservation in most vertebrate species, they could prove to be useful candidates to study the morphological development of turtles, particularly in relation to shell formation. Of note, *KDSR* – one of the genes possibly under positive selection in giant tortoises – has been linked to hyperkeratinisation disorders<sup>16</sup>. Also, in this regard, we annotated 30  $\beta$ -keratins in *C. abingdonii*, 26 of which seem to be functional. These numbers are lower than those previously reported for  $\beta$ -keratins in other turtles<sup>17</sup>. Finally, we did not find in *C. abingdonii* or *A. gigantea* any functional orthologs of genes specifically involved in tooth development (such as *ENAM*, *AMEL*, *AMBN*, *DSPP*, *KLK4* and *MMP20*). This finding confirms a pattern in the evolutionary molecular mechanisms for tooth loss, which seems to have been followed consistently and independently across vertebrates. Taken together, these results offer multiple candidates to study developmental traits in tortoises (Supplementary Information, Section 2, Figs. S5-S8).

In most species, the immune function is an evolutionary driver that is under strong selective pressure and has important implications in ageing and disease<sup>18</sup>. The specific components

and functionality of immune system components in Reptilia, however, has not been extensively characterized beyond the major histocompatibility complex (MHC)<sup>19,20</sup>. Our detailed analysis of 891 genes involved in immune function consistently found duplications affecting immunity genes in giant tortoises compared to mammals (Supplementary Information, Section 3, Table S9, Figs. S9-S13). We found a genomic expansion of *PRFI* (encoding perforin) in giant tortoises and other turtles, compared to chicken (one copy), *A. carolinensis* (two copies) and most mammals (one copy). Both *C. abingdonii* and *A. gigantea* possess 12 copies of this gene (validated by Sanger sequencing), although three of them have been pseudogenized in *C. abingdonii*. In addition, we detected and validated by Sanger sequencing an expansion of the chymase locus, containing granzymes, in giant tortoises (Supplementary Information, Section 3.1, Fig. S10). Both expansions are expected to affect cytotoxic T lymphocyte (CTL) and natural killer (NK) functions, which play important roles in defence against both pathogens and cancer<sup>21,22</sup>. Other concurrent expansions involve *APOBEC1*, *CAMP*, *CHIA* and *NLRP* genes, which participate in viral, microbial, fungal and parasite defence, respectively. These results suggest that the innate immune system in turtles, and especially in giant tortoises, may play a more relevant role than in mammals, consistent with the less important role that adaptive immunity seems to play<sup>19</sup>. We found that class-I and class-II MHC genes likely underwent a duplication event in a common ancestor between giant tortoises and painted turtles (*Chrysemys picta bellii*). We also annotated 40 class-III MHC genes, thus confirming the conservation of this cluster in giant tortoises. The large number of MHC genes in giant tortoises is consistent with the suggestion that ancestors of archosaurs and chelonians did not possess a minimal essential MHC as found in the chicken genome<sup>20</sup>. (Supplementary Information, Section 3.3, Table S10, Figs. S14-S16).

Giant tortoises are at the upper end of the size scale for extant Chelonii, and have often been used as an example of gigantism<sup>23</sup>. We analysed a series of genes involved in size regulation in vertebrates, most notably dogs (Supplementary Information, Section 2, Table S8, Fig. S6). Our results on genes related to growth hormone, IGF system and stanniocalcins suggest that these genes are well conserved, and therefore additional size determinants may exist in giant tortoises. As a complex phenotype, gigantism in tortoises is expected to be caused by interactions between different genetic and environmental factors. An interesting finding in this regard is the presence of several gene variants in tortoises (including *G. agassizii*) likely affecting the activities of glucose metabolism genes, such as *MIF* (p.N111C, expected to yield a locked trimer) and *GSK3A* (p.R272Q in the activation loop). Given the roles of these positions in the mammalian orthologs of these genes, tortoise-specific changes could point to differences in the regulation of glucose intake and tolerance (Supplementary Information, Section 4, Table S11, Figs. S17 and S18). We also found expansions and inactivations in other genes involved in energy metabolism. Thus, glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), a glycolytic enzyme with a key role in energy production as well as in DNA repair and apoptosis<sup>24</sup>, is expanded in giant tortoises. Conversely, the *NLN* gene encoding neurolysin is pseudogenized in tortoises. The loss of this gene in mice has been related to improved glucose uptake and insulin sensitivity<sup>25</sup>. Taken together, these results lead us to hypothesize that genomic variants affecting glucose metabolism may have been a factor in the development of tortoises.

The analysis of genes related to stress response has also highlighted several putative variants in giant tortoises affecting globins and DNA repair factors (Supplementary Information, Section 5, Tables S12-S13, Figs. S19-S22 and S32-S33). We found that, despite living terrestrially, giant tortoises conserve the hypoxia-related globin *Gbx26*. Together with coelacanths, turtles, including giant tortoises, are the only organisms known to possess all eight different types of globins<sup>27</sup>. Consistent with this, we found in both giant tortoise genomes a variant in the transcription factor *TP53* (p.S106E) that has been linked to hypoxia resistance in some mammals and fishes<sup>28</sup>. The presence of the same residue in Testudines strongly suggests a process of convergent evolution in the adaptation to hypoxia, likely driven by an ancestral aquatic environment, which left this footprint in the genomes of terrestrial giant tortoises.

An important trait of large, long-lived vertebrates is their need for tighter cancer protection mechanisms, as illustrated by Peto's paradox<sup>29,30</sup>. In turn, this need for additional protection illustrates the deep relationship and interdependence between cancer and longevity (Fig. 2). Notably, tumours are believed to be very rare in turtles<sup>31</sup>. Therefore, we analysed more than 400 genes classified in a well-established census of cancer genes as oncogenes and tumour suppressors<sup>32</sup>. Although most presented a highly conserved amino acid sequence when compared with the sequences of other organisms, we uncovered alterations in several tumorigenesis-related genes (Fig. 2a, Supplementary Information, Section 6, Table S14, Figs. S23-S29). First, we found that several putative tumour suppressors are expanded in turtles compared to other vertebrates, including duplications in *SMAD4*, *NF2*, *PML*, *PTPN11* and *P2RY8*. In addition, the aforementioned expansion of *PRF1*, together with the tortoise-specific duplication of *PRDMI*, suggests that immunosurveillance may be enhanced in turtles. Likewise, we found giant tortoise-specific duplications affecting two putative proto-oncogenes, *MYCN* and *SET*. Notably, the SET complex mediates oxidative stress responses induced by mitochondrial damage through the action of PRF1 and GZMA in CTL and NK-mediated cytotoxicity<sup>33</sup>. Taken together, these results suggest that multiple gene copy-number alterations may have influenced the mechanisms of spontaneous tumour growth. Nevertheless, further studies are needed to evaluate the genomic determinants of putative giant-tortoise-specific cancer mechanisms.

Finally, we selected for manually-supervised annotation a set of 500 genes which may be involved in ageing modulation (Supplementary Information, Section 7, Table S15). The extreme longevity of giant tortoises is expected to involve multiple genes affecting different hallmarks of ageing<sup>11</sup>. We found several alterations in the genomes of giant tortoises that may play a direct role in six of them, and impinge on other ageing hallmarks and processes, such as cancer progression<sup>34</sup> (Fig. 2b). First, we identified changes in three candidate factors (*NEIL1*, *RMI2* and *XRCC6*) related to the maintenance of genome integrity, a primary hallmark of ageing<sup>11</sup> (Fig. 3a). Thus, we found and validated a duplication affecting *NEIL1*, a key protein involved in the base-excision repair process whose expression has been linked to extended lifespans in several species<sup>35</sup>. Likewise, *RMI2* is duplicated in tortoises, suggesting an enhanced ability to resolve homologous recombination intermediates to limit DNA crossover formation in cells<sup>36</sup>. In a preliminary exploration of this hypothesis, we overexpressed *NEIL1* and *RMI2* in HEK-293T cells and exposed the infected cells to a sub-lethal dosage of H<sub>2</sub>O<sub>2</sub> or UV light, monitoring DNA damage by

Western blot analysis at 24 and 48 h after treatment. As shown in Supplementary Information, Figs. S22, S32 and S33, the expression of both genes results in reduced levels of phosphorylated histone H2AX and cleaved poly (ADP-ribose) polymerase (PARP), suggesting reduced levels of DNA damage<sup>37</sup>. In turn, this result is consistent with the hypothesis that NEIL1 and RMI2 levels may regulate the strength of DNA repair mechanisms. Also in relation to DNA repair mechanisms, we identified and validated a variant affecting *XRCC6* – encoding a helicase involved in non-homologous end-joining of double-strand DNA breaks – which may affect a known sumoylation site (p.K556R). This lysine is conserved in diverse vertebrates but, notably, is changed in giant tortoises, and also in the naked mole rat (p.K556N), the longest-lived rodent, which suggests a putative process of convergent evolution (Fig. 3b). Since sumoylation is induced following DNA damage and plays a key role in DNA repair response and in multiple regulatory processes<sup>38</sup>, this variant may reflect selective pressures acting on the regulation of the repair of double-strand DNA breaks in long-lived organisms (Supplementary Information, Section 5.5).

Regarding telomere attrition, another primary hallmark of ageing<sup>11</sup>, we uncovered in giant tortoises one variant in *DCLRE1B* (p.R498C) potentially affecting its binding interface with telomeric repeat binding factor 2 (TERF2) (Fig. 3b, Supplementary Information, Section 7.2). This change, together with the aforementioned variants affecting DNA repair genes that may also impinge on telomere dynamics<sup>39–41</sup>, highlights the relevance of telomere maintenance as a regulatory mechanism of longevity in tortoises. Moreover, we found changes potentially affecting proteostasis (Fig. 2a). We independently found the specific expansion in *C. abingdonii*, *A. gigantea* and *G. agassizii* of the elongation factor gene *EEF1A1*, as described with the automatic annotation. Importantly, overexpression of *EEF1A1* homologs in *Drosophila melanogaster* has been linked to an increased lifespan in that species<sup>42</sup>.

Over time, nutrient-sensing deregulation – another hallmark of ageing – can result from alterations in metabolic control mechanisms and signalling pathways<sup>12</sup>. The aforementioned variant affecting the activation loop of GSK3A (Supplementary Information, Section 4.1), which is present in *C. abingdonii* and all tested tortoises from the Galapagos Islands, Aldabra Atoll, their continental outgroups, *G. agassizii* and *C. p. bellii*, may be involved in the maintenance of glucose homeostasis. Interestingly, the inhibition of GSK3 can extend lifespan in *D. melanogaster*<sup>43</sup>. Likewise, the identified alterations in other giant tortoise genes implicated in glucose metabolism, such as the aforementioned inactivation of *NLN*, may provide interesting candidates to study nutrient sensing in these long-lived species (Supplementary Information, Section 7.4).

Regarding the mitochondrial function, we found two variants (p.Q366M and p.M487T) potentially affecting the function of ALDH2, a mitochondrial aldehyde dehydrogenase involved in alcohol metabolism and in lipid peroxidation, among other detoxification processes<sup>44</sup>. Notably, the p.Q366M variant, which may alter the NAD-binding site of ALDH2, is exclusively found in Galapagos giant tortoises, but not in their continental close relative *C. chilensis*, nor in the more distantly related Aldabra or Agassiz tortoises. Thus, these changes could also alter the detoxification process and contribute to pro-longevity mechanisms. Together with the above described specific alterations in other genes of giant

tortoises such as *NLN* and *GAPDH*, which encode enzymes associated with mitochondrial functions<sup>45,46</sup>, these variants may also impinge on mitochondrial dysfunction, an antagonistic hallmark of ageing<sup>11</sup> (Supplementary Information, Section 7.5).

We have also found evidence in tortoises of some variants related to altered intercellular communication (Supplementary Information, Section 7.6, Fig. S30), an integrative hallmark of ageing<sup>11</sup>. Thus, we have detected exclusively in *C. abingdonii* a premature stop codon affecting *ITGA1* (p.R990\*), an essential integrin involved in cell-matrix and cell-cell interactions. In addition, the aforementioned variant affecting *MIF* is also expected to cause the formation of inactivating interchain disulfide bonds, inhibiting intracellular signalling cascades<sup>47</sup>. Moreover, *MIF* deficiency reduces chronic inflammation in white adipose tissue and expands lifespan, especially in response to caloric restriction<sup>48,49</sup>. Finally, we have annotated a specific variant in *IGF1R* which is expected to affect the interaction between this receptor and the IGF1/2 growth factors<sup>50</sup>. Notably, a homology model of this region in *C. abingdonii*'s IGF1R suggests that position 724 is located at the surface of the protein, and the presence of an aspartic acid residue changes the local electrostatic field (Fig. 4a). The extended lifespan in different species correlates with IGF signalling decrease<sup>51,52</sup>, which suggests that this unique change in *IGF1R* may provide an attractive target to study the cellular mechanisms underlying the exceptional lifespan of these animals. To explore the functional consequences of differential IGF1 signalling caused by the N724D variant found in IGF1R, we infected HEK-293T cells with pCDH, pCDH-IGF1R<sup>WT</sup> and pCDH-IGF1R<sup>N724D</sup> plasmids. Cells expressing the mutant receptor showed an attenuation of IGF1 signalling, compared with those expressing the wild-type protein, measured as a significant reduction in the phosphorylation levels of IGF1R at 5 min. (95% C. I. of difference 0.1119-1.533,  $t=2.454$ ,  $P=0.026$ ) and 10 min. (95% C. I. of difference 0.1991-1.62,  $t=2.714$ ,  $P=0.0153$ ) after IGF1 treatment (Fig. 4b, Supplementary Information, Section 7.6.2, Fig. S31). According to a two-way ANOVA analysis, the exogenous IGF1R form accounted for 16.07% of total variation, with  $F(1,4)=20.91$ ,  $P=0.0102$  (time accounted for 44.23% of total variation, with  $F(3,12)=6.57$ ,  $P=0.0071$ ). Interestingly, we have also found in tortoises a short deletion in the coding region of *IGF2R* which results in the loss of two amino acids. The fact that *IGF2R* variants have been associated with human longevity<sup>53</sup> opens the possibility that the variant found in tortoises could also contribute to increasing the lifespan of these long-lived animals.

In summary, in this work we report the preliminary characterization of giant tortoise genomes. We complemented the automatic annotation of genomes from two giant tortoise species with a hypothesis-driven strategy using manually-supervised annotation of a large set of genes. The analysis of the resulting sequences offers candidate genes and pathways that may underlie the extraordinary characteristics of these iconic species, including their development, gigantism and longevity. A better understanding of the processes that we have studied may help to further elucidate the biology of these species and therefore aid the ongoing efforts to conserve these dwindling lineages. Lonesome George, the last representative of *C. abingdonii* and renowned emblem of the plight of endangered species, left a legacy including a story written in his genome whose unveiling has just started.



## Methods

### Genome sequencing and assembly

We obtained DNA from a blood sample from Lonesome George, the last member of *C. abingdonii*. 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. These libraries were assembled with the AllPaths algorithm<sup>54</sup> for a draft genome containing 64,657 contigs with an N50 of 74 kb. Then, we scaffolded the contigs with Sspace v3.055 using the long-insert mate-pair libraries. Finally, we filled the gaps with PBJelly v15.8.2456 using the reads obtained from 18 BioPac cells. This step yielded 10,623 scaffolds with an N50 of 1.27 Mb, for a final assembly 2.3 Gb long. Then, we soft-masked repeated regions using *RepeatMasker*<sup>57</sup> with a database containing chordate repeated elements (included in the software) as a reference. Additionally, we assessed the completeness of the assembly by their estimated gene content, using the Benchmarking Universal Single Copy Orthologs (BUSCO v3.0.0)<sup>58</sup>, which tested the status of a set of 2,586 vertebrata genes from the comprehensive catalogue of orthologs<sup>59</sup>. We also performed RNA-Seq from *C. abingdonii* blood and *A. gigantea* granuloma and aligned the resulting reads to the assembled genome using TopHat60 (v2.0.14). Finally, we obtained whole-genome data from *A. gigantea* with one Illumina lane of a 180 bp paired-end library. The resulting reads were aligned to *C. abingdonii* genome with BWA61 (v0.7.5a). Raw reads from *C. abingdonii* were also aligned to the genome for manual curation of results. All work on field samples was conducted at Yale University under IACUC permit number 2016-10825, Galapagos Park Permit PC-75-16 and CITES number 15US209142/9.

### Genome annotation

Using the genome assembly of *C. abingdonii* and the RNA-Seq reads from *C. abingdonii* and *A. gigantea*, we performed *de novo* annotation with MAKER2. The algorithm was also fed reference sequences from both human and *P. sinensis*, and performed two runs in a Microsoft Azure virtual machine (Supplementary Table 16). In parallel, we used selected genes from the human protein database in Ensembl as a reference to manually predict the corresponding homologues in the genome of *C. abingdonii* by using the BATI algorithm (Blast, Annotate, Tune, Iterate)<sup>62</sup>. Briefly, this algorithm allows a user to annotate the position and intron/exon boundaries of genes in novel genomes from tblastn results. In addition, tblastn results are integrated to search for novel homologues in the explored genome. Sequencing data have been deposited at the Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>), with comments showing which regions were filled with the BioPac reads and therefore may contain frequent errors.

### Effective population size changes and diversity

We reconstructed changes in effective population over time using the PSMC model<sup>5</sup> in the following way: the reads of both individuals were aligned to the reference assembly using bwa mem (v. 0.7.15-r1140). We then constructed pseudo-diploid sequences using variant calls generated with samtools and bcftools<sup>63</sup>, requiring minimal base and mapping qualities of 30. We additionally masked out any region with coverage below 36 or above 216 for the *C. abingdonii* sample, and below 8 or above 52 for the *A. gigantea* sample, as a function of

their respective genome wide average coverage. The resulting sequences were used to run 100 psmc bootstrap replicates per individual, using the following parameters: -N25 -t15 -r5 -p "4+25\*2+4+6". The result was averaged and scaled to real time assuming a mutation rate ( $\mu$ ) of  $2.5 \times 10^{-8}$  and a generation time (g) of 25 years.

### Expansion of gene families

To detect expansion of gene families, we aligned pairwise all the predicted proteins from the automatic annotation to the Uniprot64 database of human proteins and to the Uniprot database of *P. sinensis* proteins using BLAST65 (v2.6.011). Then, we used in-house Perl scripts to group these proteins in one-to-one, one-to-many and many-to-many orthologous relationships. Only alignments spanning at least 80% of the longer protein and with more than 60% identities were considered. Finally, we interrogated the resulting database to find families with *C. abingdonii*-specific expansions and curated the results manually. This way we constructed extended orthology sets that may contain more than one sequence per species. These sets recapitulate most of the known families, although some of these families appear split according to sequence similarity.

### Phylogenetic, evolutionary and structural analyses

Next, we assessed evidence for signatures of positive selection affecting the predicted set of genes. For this purpose, we used databases 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 soft-shell turtle (*P. sinensis*), green sea turtle (*C. mydas*) and painted turtle (*C. p. bellii*) to generate pairwise alignments of all available genes one by one. To this end, we used BLAST and simple in-house perl scripts (<https://github.com/vqf/LG>), which allowed us to group the genes by identity (focusing only in those presenting one-to-one orthology). We then discarded those groups in which there was more than 3 species missing (always excluding those in which *C. abingdonii* was missing). This way, we obtained 1,592 groups of sequences (similar to other studies). We then aligned them with PRANK v.150803 using the codon model and analysed the alignments with *codeml* from the PAML package66. To search for genes with signatures of positive selection affecting genes specific to *C. abingdonii*, we executed two different branch models, M0, with a single  $\omega$  value (where  $\omega$  represents the ratio of non-synonymous to synonymous substitutions) for all the branches (nested), and M2a, with a foreground  $\omega_2$  value exclusive for *C. abingdonii* and a background  $\omega_1$  value for all the other branches. As a control, the second model was repeated using *P. sinensis* as the foreground branch. Genes with a high  $\omega_2$  value ( $>1$ ) and a low  $\omega_1$  value ( $\omega_1 < 0.2$  and  $\omega_1 \approx \omega_0$ ) in *C. abingdonii*, but not in *P. sinensis* (Supplementary Information, Section 1.2 and Supplementary Tables 5 and 17) were then considered to be under positive selection. After this, we used the M8 model to assess the individual importance of every site in these positively selected genes, obtaining a list of sites of special interest in this evolutionary effect. These results were compared to those of Aldabra tortoise's through alignments, to evaluate which of these important residues were altered (Supplementary Table

18). Homology models were performed with SWISS-MODEL67 from the closest template available. The results were inspected and rendered with DeepView v4.0.1. Electric potentials were calculated with DeepView using the Poisson-Boltzmann computation method. Figures were generated with PovRay (<http://povray.org>).

## Functional analyses

HEK-293T cells were infected with pCDH, pCDH-NEIL1, pCDH-RMI2 or pCDH-NEIL1 + pCDH-RMI2 in the case of repair studies, and pCDH, pCDH-IGF1R<sup>WT</sup> or pCDH-IGF1R<sup>N724D</sup> in the case of IGF1R analyses. For the repair studies, we isolated clones of infected HEK-293T with proper expression levels of NEIL1 and RMI2. Cells were exposed to UV light (20 J/m<sup>2</sup>) or H<sub>2</sub>O<sub>2</sub> (500 μM) 24 and 48 h before being lysed in NP-40 lysis buffer containing 50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM EDTA pH 8, and 1% NP-40, and supplemented with protease inhibitor cocktail (cOmplete, EDTA-free, Roche), as well as phosphatase inhibitors (PhosSTOP, Roche; NaF, Merck). For the *IGF1R* variant analyses, cells were serum-starved for 14 h and then treated with 100 nM *IGF1* for 5, 10 and 20 min before lysis in the same buffer. Equal amounts of protein were resolved by 8 to 13% SDS-PAGE and transferred to PVDF membranes (GE Healthcare Life Sciences). Membranes were blocked for 1 h at room temperature with TBS-T (0.1% Tween 20) containing 5% BSA. Immunoblotting was performed with primary antibodies diluted 1:500 to 1:1000 in TBS-T and 1% BSA and incubated overnight at 4 °C. The primary antibodies used were: anti-phospho-Histone H2AX (Ser139) (EMD Millipore, 05-636, clone JBW301, lot 2854120), anti-PARP (Cell Signalling, 9542S, rabbit polyclonal, lot 15), anti-FLAG (Cell Signalling, 2368S, rabbit polyclonal, lot 12), anti-IGF1R (Abcam, ab182408, clone EPR19322, lot GR312678-8), anti-IGF1R (p Tyr1161) (Novus Biologicals, NB100-92555, rabbit polyclonal, lot CJ36131), anti-β-actin (Sigma, A5441, clone AC-15, lot 014M4759) and anti-α-tubulin (Sigma, T6074, clone B-5-1-2, lot 075M4823V). After washing with TBS-T, membranes were incubated with secondary antibodies conjugated with IRDye 680RD (LI-COR Biosciences; 926-68071, polyclonal goat-anti-rabbit, lot C41217-03; and 926-32220, polyclonal goat-anti-mouse, lot C00727-03) or IRDye 800CW (LI-COR Biosciences; 926-32211, polyclonal goat-anti-rabbit, lot C60113-05; and 926-32210, polyclonal goat-anti-mouse, lot C50316-03) for 1 h at room temperature. Protein bands were scanned on an Odyssey infrared scanner (LI-COR). Band intensities were quantified by ImageJ and used to calculate the pIGF1R/IGF1R ratio in the case of the IGF1R assay. In each replicate, cells were infected independently. For the samples from UV treatment, 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<sub>2</sub>O<sub>2</sub> 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). Each sample contains one replicate. Statistical comparisons consisted of two-way ANOVA analyses performed using GraphPad Prism 7.0 software. Differences were considered statistically significant when P<0.05 (\*P<0.05). Effect sizes are expressed as group sum-of-squares divided by the total sum-of-squares (R<sup>2</sup>). In each time point, both groups were also compared with Fisher's LSD test (uncorrected, α=0.05).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank J.R. Obeso for support and J.M. Freije, X.S. Puente, R. Valdés-Mas, F.G. Osorio, D. López-Velasco, A. Corrales, P. Salinas, D. Rodríguez, A. López-Soto, A.R. Folgueras and M. Mittelbrunn for helpful comments and advice; M. Garaña, O. Sanz, J. Isla, and A. Marcos (Microsoft) for computing facilities; and F. Rodríguez, D.A. Puente and S.A. Miranda for excellent technical assistance. We also acknowledge the generous support by J.I. Cabrera. We thank Banco Santander for funding a short stay of S.F-R. and D.C-I. in Yale University. V.Q. is supported by grants from the Principado de Asturias and Ministerio de Economía y Competitividad, including FEDER funding. L.F.K.K. is supported by an FPI fellowship associated to BFU2014-55090-P (FEDER). T.M-B. is supported by MINECO BFU2017-86471-P (MINECO/FEDER, UE), NIH U01 MH106874 grant, Howard Hughes International Early Career, Obra Social "La Caixa" and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya. C.L-O. is supported by grants from European Research Council (DeAge, ERC Advanced Grant), Ministerio de Economía y Competitividad, Instituto de Salud Carlos III (RTICC) and Progeria Research Foundation. The Instituto Universitario de Oncología is supported by Fundación Bancaria Caja de Ahorros de Asturias. We also thank the Galapagos National Park and Galapagos Conservancy for logistic and financial support.

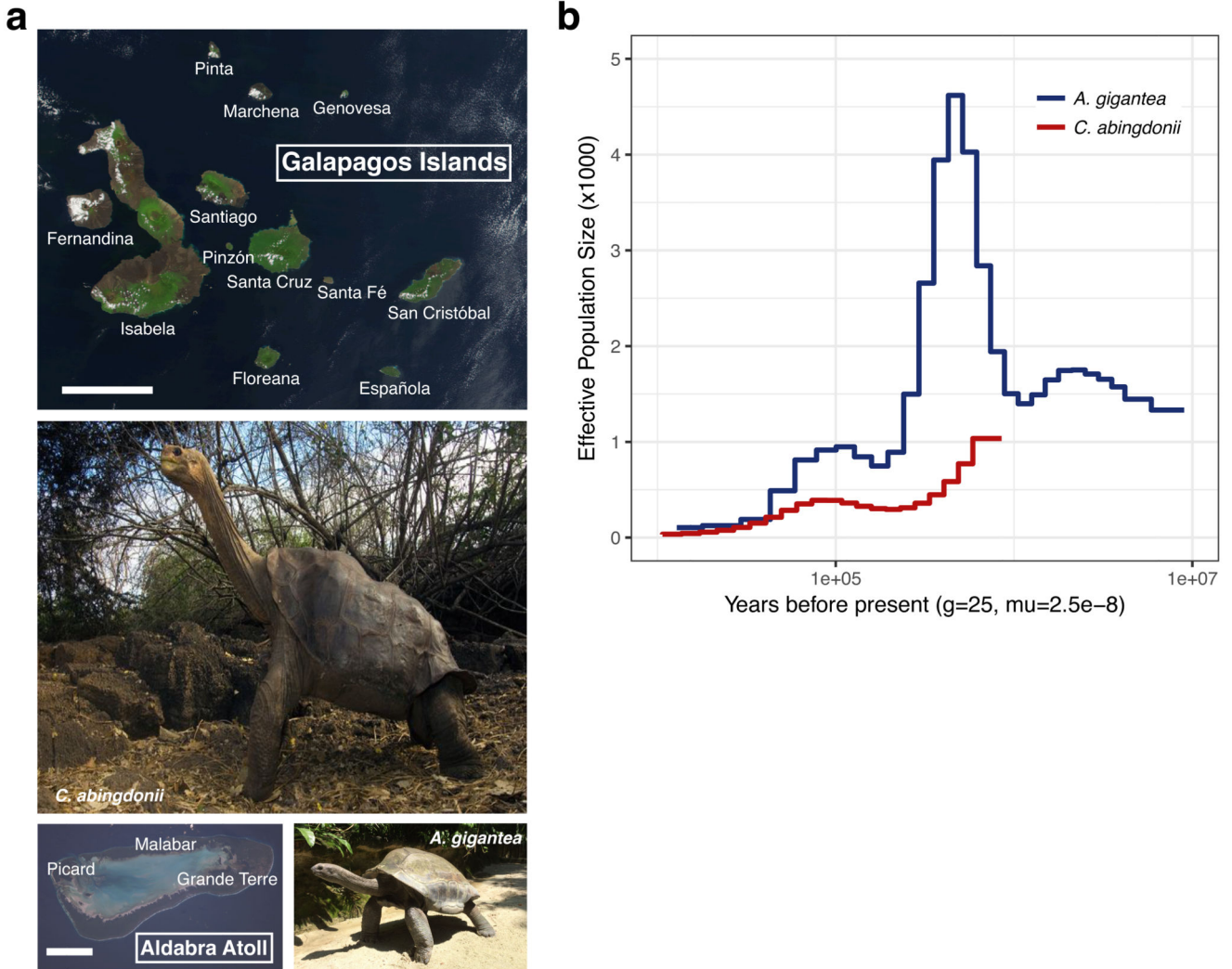
## References

- Kim EB, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature*. 2011; 479:223–227. [PubMed: 21993625]
- Keane M, et al. Insights into the evolution of longevity from the bowhead whale genome. *Cell Rep*. 2015; 10:112–122. [PubMed: 25565328]
- Nicholls H. The legacy of Lonesome George. *Nature*. 2012; 487:279–280. [PubMed: 22810667]
- Kehlmaier C, et al. Tropical ancient DNA reveals relationships of the extinct Bahamian giant tortoise *Chelonoidis alburyorum*. *Proc Biol Sci*. 2017; 284:20162235. [PubMed: 28077774]
- Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011; 475:493–496. [PubMed: 21753753]
- Campbell MS, Holt C, Moore B, Yandell M. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics*. 2014; 48:1–39. [PubMed: 25501939]
- 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*. 2013; 45:701–706. [PubMed: 23624526]
- 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)*. 2015; 37:9776. [PubMed: 25911468]
- Pal D, et al. Fetuin-A acts as an endogenous ligand of TLR4 to promote lipid-induced insulin resistance. *Nat Med*. 2012; 18:1279–1285. [PubMed: 22842477]
- Kir S, et al. FGF19 as a postprandial, insulin-independent activator of hepatic protein and glycogen synthesis. *Science*. 2011; 331:1621–1624. [PubMed: 21436455]
- López-Otín C, Blasco MA, Partridge L, Serrano M, Kroemer G. The hallmarks of aging. *Cell*. 2013; 153:1194–1217. [PubMed: 23746838]
- López-Otín C, Galluzzi L, Freije JM, Madeo F, Kroemer G. Metabolic control of longevity. *Cell*. 2016; 166:802–821. [PubMed: 27518560]
- van der Goot AT, et al. Delaying aging and the aging-associated decline in protein homeostasis by inhibition of tryptophan degradation. *Proc Natl Acad Sci USA*. 2012; 109:14912–14917. [PubMed: 22927396]
- Crawford NG, et al. A phylogenomic analysis of turtles. *Mol Phylogen Evol*. 2015; 83:250–257.
- 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*. 2012; 10:65. [PubMed: 22839781]

16. Boyden LM, et al. Mutations in KDSR cause recessive progressive symmetric erythrokeratoderma. *Am J Hum Genet.* 2017; 100:978–984. [PubMed: 28575652]
17. Li YI, Kong L, Ponting CP, Haerty W. Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biol Evol.* 2013; 5:923–933. [PubMed: 23576313]
18. Barreiro LB, Quintana-Murci L. From evolutionary genetics to human immunology: how selection shapes host defence genes. *Nat Rev Genet.* 2010; 11:17–30. [PubMed: 19953080]
19. Zimmerman LM, Vogel LA, Bowden RM. Understanding the vertebrate immune system: insights from the reptilian perspective. *J Exp Biol.* 2010; 213:661–671. [PubMed: 20154181]
20. Balakrishnan CN. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* 2010; 8:29. [PubMed: 20359332]
21. Dotiwala F, et al. Killer lymphocytes use granulysin, perforin and granzymes to kill intracellular parasites. *Nat Med.* 2016; 22:210–216. [PubMed: 26752517]
22. Voskoboinik I, Whisstock JC, Trapani JA. Perforin and granzymes: function, dysfunction and human pathology. *Nat Rev Immunol.* 2015; 15:388–400. [PubMed: 25998963]
23. Jaffe AL, Slater GJ, Alfaro ME. The evolution of island gigantism and body size variation in tortoises and turtles. *Biol Lett.* 2011; 7:558–561. [PubMed: 21270022]
24. Chuang DM, Hough C, Senatorov VV. Glyceraldehyde-3-phosphate dehydrogenase, apoptosis, and neurodegenerative diseases. *Annu Rev Pharmacol Toxicol.* 2005; 45:269–290. [PubMed: 15822178]
25. Cavalcanti DM, et al. Neurolysin knockout mice generation and initial phenotype characterization. *J Biol Chem.* 2014; 289:15426–15440. [PubMed: 24719317]
26. 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.* 2016; 113:8538–8543. [PubMed: 27407144]
27. Schwarze K, Singh A, Burmester T. The full globin repertoire of turtles provides insights into vertebrate globin evolution and functions. *Genome Biol Evol.* 2015; 7:1896–1913. [PubMed: 26078264]
28. 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.* 2013; 110:20639–20644. [PubMed: 24297887]
29. Caulin AF, Maley CC. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol Evol.* 2011; 26:175–182. [PubMed: 21296451]
30. Chiari Y, Glaberman S, Lynch VJ. Insights on cancer resistance in vertebrates: reptiles as a parallel system to mammals. *Nat Rev Cancer.* 2018
31. Garner MM, Hernandez-Divers SM, Raymond JT. Reptile neoplasia: a retrospective study of case submissions to a specialty diagnostic service. *Vet Clin North Am Exot Anim Pract.* 2004; 7:653–671. [PubMed: 15296868]
32. Futreal PA, et al. A census of human cancer genes. *Nat Rev Cancer.* 2004; 4:177–183. [PubMed: 14993899]
33. Martinvalet D, Zhu P, Lieberman J. Granzyme A induces caspase-independent mitochondrial damage, a required first step for apoptosis. *Immunity.* 2005; 22:355–370. [PubMed: 15780992]
34. Gorbunova V, Seluanov A, Zhang Z, Gladyshev VN, Vijg J. Comparative genetics of longevity and cancer: insights from long-lived rodents. *Nat Rev Genet.* 2014; 15:531–540. [PubMed: 24981598]
35. MacRae SL, et al. DNA repair in species with extreme lifespan differences. *Aging (Albany NY).* 2015; 7:1171–1184. [PubMed: 26729707]
36. Daley JM, Chiba T, Xue X, Niu H, Sung P. Multifaceted role of the Topo IIIalpha-RMI1-RMI2 complex and DNA2 in the BLM-dependent pathway of DNA break end resection. *Nucleic Acids Res.* 2014; 42:11083–11091. [PubMed: 25200081]
37. Ivashkevich A, Redon CE, Nakamura AJ, Martin RF, Martin OA. Use of the gamma-H2AX assay to monitor DNA damage and repair in translational cancer research. *Cancer Lett.* 2012; 327:123–133. [PubMed: 22198208]

38. Cremona CA, et al. Extensive DNA damage-induced sumoylation contributes to replication and repair and acts in addition to the mec1 checkpoint. *Mol Cell*. 2012; 45:422–432. [PubMed: 22285753]
39. Wang Y, Ghosh G, Hendrickson EA. Ku86 represses lethal telomere deletion events in human somatic cells. *Proc Natl Acad Sci USA*. 2009; 106:12430–12435. [PubMed: 19581589]
40. Tong AS, et al. ATM and ATR signaling regulate the recruitment of human telomerase to telomeres. *Cell Rep*. 2015; 13:1633–1646. [PubMed: 26586433]
41. Ribes-Zamora A, Indiviglio SM, Mihalek I, Williams CL, Bertuch AA. TRF2 interaction with Ku heterotetramerization interface gives insight into c-NHEJ prevention at human telomeres. *Cell Rep*. 2013; 5:194–206. [PubMed: 24095731]
42. Shikama N, Ackermann R, Brack C. Protein synthesis elongation factor EF-1 alpha expression and longevity in *Drosophila melanogaster*. *Proc Natl Acad Sci USA*. 1994; 91:4199–4203. [PubMed: 8183891]
43. Castillo-Quan JI, et al. Lithium promotes longevity through GSK3/NRF2-dependent hormesis. *Cell Rep*. 2016; 15:638–650. [PubMed: 27068460]
44. Ohta S, Ohsawa I, Kamino K, Ando F, Shimokata H. Mitochondrial ALDH2 deficiency as an oxidative stress. *Ann N Y Acad Sci*. 2004; 1011:36–44. [PubMed: 15126281]
45. Serizawa A, Dando PM, Barrett AJ. Characterization of a mitochondrial metallopeptidase reveals neurolysin as a homologue of thimet oligopeptidase. *J Biol Chem*. 1995; 270:2092–2098. [PubMed: 7836437]
46. Tristan C, Shahani N, Sedlak TW, Sawa A. The diverse functions of GAPDH: views from different subcellular compartments. *Cell Signal*. 2011; 23:317–323. [PubMed: 20727968]
47. 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*. 2013; 110:10994–10999. [PubMed: 23776208]
48. 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*. 2009; 105:99–107.
49. Harper JM, Wilkinson JE, Miller RA. Macrophage migration inhibitory factor-knockout mice are long lived and respond to caloric restriction. *FASEB J*. 2010; 24:2436–2442. [PubMed: 20219983]
50. Whittaker J, et al. Alanine scanning mutagenesis of a type 1 insulin-like growth factor receptor ligand binding site. *J Biol Chem*. 2001; 276:43980–43986. [PubMed: 11500492]
51. Kenyon CJ. The genetics of ageing. *Nature*. 2010; 464:504–512. [PubMed: 20336132]
52. Brohus M, Gorbunova V, Faulkes CG, Overgaard MT, Conover CA. The insulin-like growth factor system in the long-lived naked mole-rat. *PLoS One*. 2015; 10:e0145587. [PubMed: 26694858]
53. 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*. 2012; 47:379–387. [PubMed: 22406557]
54. Gnerre S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA*. 2011; 108:1513–1518. [PubMed: 21187386]
55. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011; 27:578–579. [PubMed: 21149342]
56. English AC, et al. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PLoS One*. 2012; 7:e47768. [PubMed: 23185243]
57. Smit, A; Hubley, R; Green, P. RepeatMasker Open-4.0. 2013–2015. 2014. URL <http://www.repeatmasker.org>
58. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31:3210–3212. [PubMed: 26059717]
59. Zdobnov EM, et al. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res*. 2017; 45:D744–D749. [PubMed: 27899580]

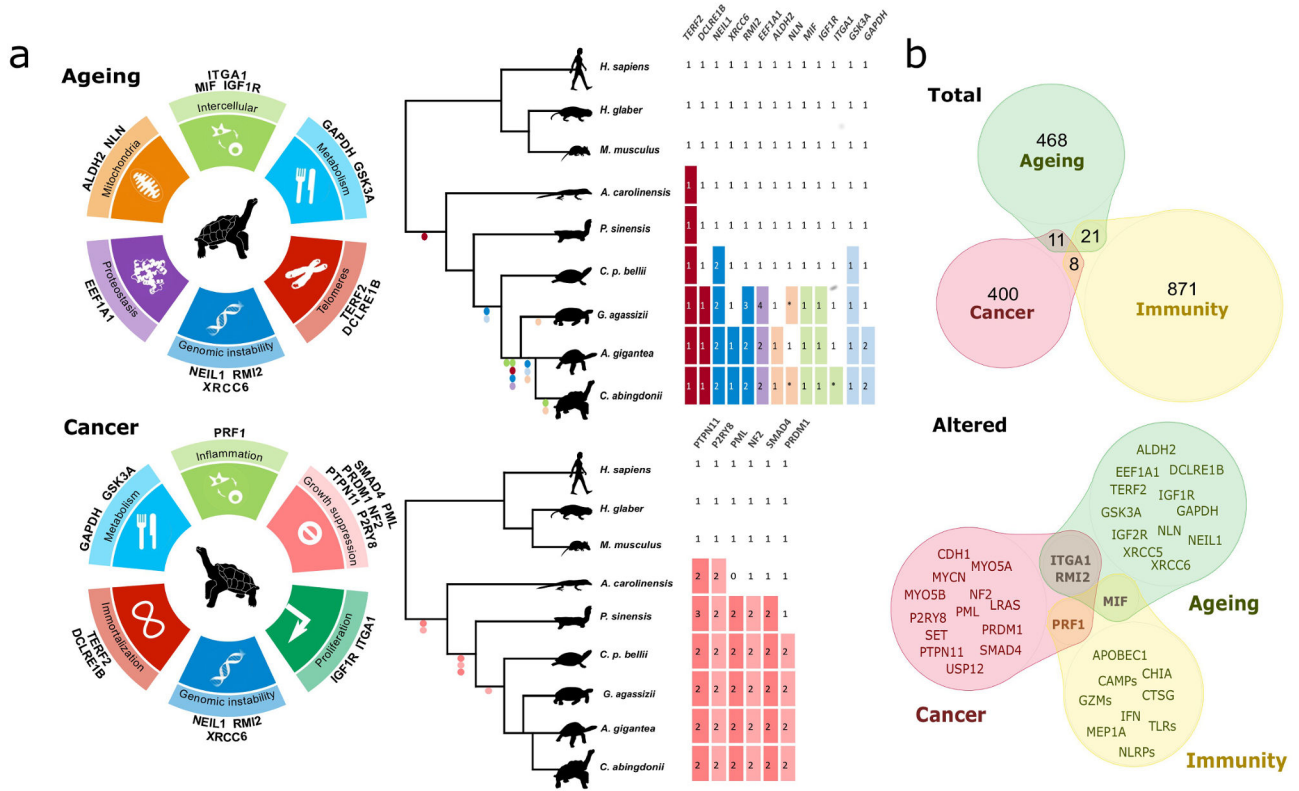
60. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009; 25:1105–1111. [PubMed: 19289445]
61. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010; 26:589–595. [PubMed: 20080505]
62. Quesada V, Velasco G, Puente XS, Warren WC, López-Otín C. Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals. *BMC Genomics*. 2010; 11:220. [PubMed: 20359326]
63. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25:2078–2079. [PubMed: 19505943]
64. UniProt: the universal protein knowledgebase. *Nucleic Acids Res*. 2017; 45:D158–D169. [PubMed: 27899622]
65. Camacho C, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009; 10:421. [PubMed: 20003500]
66. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007; 24:1586–1591. [PubMed: 17483113]
67. Biasini M, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014; 42:W252–258. [PubMed: 24782522]



**Figure 1. Geographical and temporal distribution of giant tortoises.**

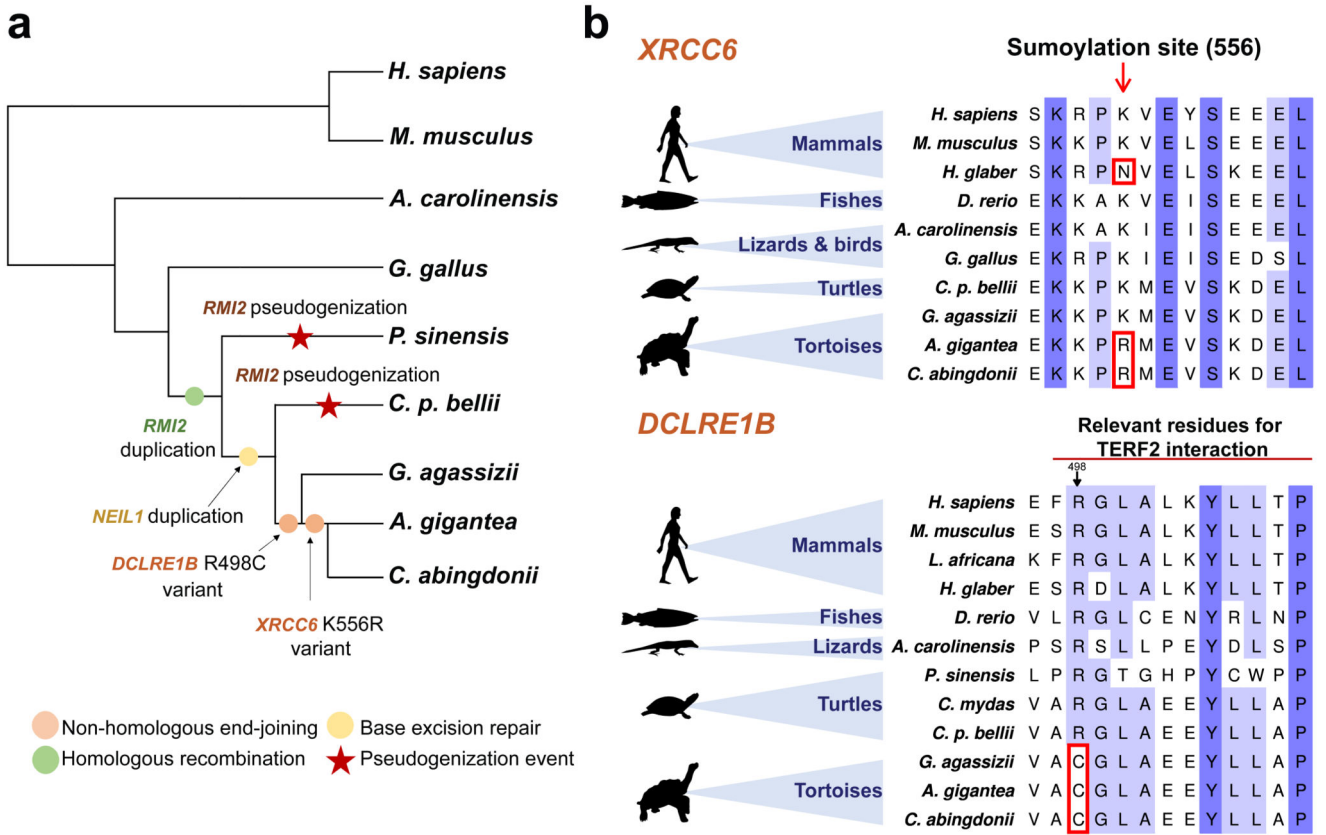
**a**, Satellite view of the Galapagos Islands (top; scale bar: 50 km) and Aldabra Atoll (bottom left; scale bar: 10 km), and pictures of *C. abingdonii* (middle), and *A. gigantea* (bottom right). Both satellite pictures are from <http://eol.jsc.nasa.gov>. **b**, Demographic history of giant tortoises, inferred using a hidden Markov model (HMM) approach as implemented in the pairwise sequentially Markovian coalescence (PSMC) model. The default mutation rate ( $\mu$ ) for humans of  $2.5 \times 10^{-8}$  and an average generation time ( $g$ ) of 25 years were used in the calculations.





**Figure 2. Genomic basis of longevity and cancer in giant tortoises.**

**a**, Genes potentially implicated in *C. abingdonii* and *A. gigantea* longevity extension and cancer resistance classified according to their putative role in the different hallmarks. Tables indicate copy-number variations and relevant variants of age-related genes and tumour suppressors found in *C. abingdonii*, *A. gigantea* and other species. In the tables, numbers indicate gene copy numbers, asterisks pseudogenization events and dots the presence/absence of the variant (in different colours, related to each hallmark). **b**, Venn diagrams showing the relationships between cancer-, ageing- and immunity-related genes, as classified before annotation. The top diagram represents all the genes related to each category that have been manually annotated, with the number of genes in each group. The bottom diagram shows those genes showing potentially interesting variations after annotation.



**Figure 3. DNA repair response in giant tortoises.**

**a**, Copy-number variations and putative function-altering point variants found in *C. abingdonii*, *A. gigantea*, and closely related species. **b**, Alignments showing the variants highlighted in *XRCC6* and *DCLRE1B*.

