A genomic perspective on the origins, evolution and adaptation of Galápagos iguanas

Julia López-Delgado [1*†], Ian M. Carr [2], Cecilia Paradiso [3,4], Paolo Gratton [5], Giuliano Colosimo [5,6], Matthijs P. van den Burg [6,7], Adolphe O. Debrot [8], Christian Sevilla [9], Mohd Noor Mat-Isa [10], Richard Bayliss [11,12], Glenn P. Gerber [6,13], Mohd Firdaus-Raih [14], Mary J. O'Connell [15,16], Gabriele Gentile [5,6*], Simon J. Goodman [1*]

1 School of Biology, University of Leeds, Leeds, United Kingdom

2 Leeds Institute for Molecular Medicine, University of Leeds, Leeds, United Kingdom

3 PhD Program in Evolutionary Biology and Ecology, Department of Biology, University of Rome Tor Vergata, Rome, Italy

4 Mississippi State University, Mississippi, United States of America

5 Department of Biology, University of Rome Tor Vergata, Rome, Italy

6 IUCN SSC Iguana Specialist Group, Gland, Switzerland

7 Burg Biologica, The Hague, The Netherlands

8 Wageningen Marine Research, Wageningen Research, Den Helder, The Netherlands

9 Galápagos National Park Directorate, Ecuador

10 Malaysia Genome and Vaccine Institute, National Institutes of Biotechnology Malaysia, Malaysia

11 School of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom

12 Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom

13 Beckman Center for Conservation Research, San Diego Zoo Wildlife Alliance, Escondido, California, USA

14 Department of Applied Physics, Faculty of Science and Technology, Bioinformatics and Molecular Simulations Research Group, Universiti Kebangsaan Malaysia, Bangi, Malaysia

15 Computational and Molecular Evolutionary Biology Group, School of Life Sciences, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom

16 Natural History Museum, London, United Kingdom

† Present address: Gestión y Planeamiento Territorial y Medioambiental, Las Palmas de Gran Canaria, Spain.

Gabriele Gentile and Simon J. Goodman contributed equally and are joint senior authors.

* Correspondence:

JLD – fbsjlop@leeds.ac.uk

GG – gabriele.gentile@uniroma2.it

SJG – s.j.goodman@leeds.ac.uk

Abstract

Island systems provide unique opportunities to explore the genomic basis of species diversification and novel trait evolution driven by adaptation to environmental challenges and exploitation of new ecological niches. Here we generate the first genomic sequences for the four Galápagos iguana species to show that they evolved *in situ* into the marine *Amblyrhynchus* and land *Conolophus* sister lineages on now submerged islands, following a single colonisation event approximately 13 million years ago, with evidence for recent inbreeding coincident with the timing of human settlement. Selection scans identified genes linked to traits facilitating adaption to the Galápagos environment, including specialised pigmentation phenotypes, as well as DNA damage and UV irradiation inflammation responses. These may contribute to adaptive enhanced cancer resistance, in relation to elevated UV exposure at equatorial latitudes. In the marine iguana, genes under positive selection are associated with their transition to underwater foraging, as well as their unique ability to repeatedly shrink and grow in body size. Our study provides insights into the origins and diversification of the iguanas and the molecular basis of adaptation to life in the Galápagos, facilitating future conservation genomic management of threatened iguanid populations.

Main

Remote oceanic islands have long been recognised as natural laboratories for studying evolutionary processes owing to their geographical isolation and habitat diversity, promoting the development of unique evolutionary trajectories. The Galápagos, in particular, have helped shape evolutionary theory [1]. The extant islands emerged from 30,000 to 4.0 million years ago as the Nazca tectonic plate moved over a stationary mantle hotspot, creating an eastward cycle of island formation and subsidence as the plate drifts toward South America [2]. The equatorial archipelago is situated *ca*. 930 km west of the continent and hosts a unique assemblage of species.

Among the islands' iconic species, Galápagos iguanas consist of two monophyletic sister lineages (Fig. 1a, b), the marine *Amblyrhynchus* and land *Conolophus*, comprising the yellow land iguana (*C. subcristatus*), the Santa Fe land iguana (*C. pallidus*) and the pink land iguana (*C. marthae*). These endemic species presumably diverged within the archipelago from a common Central American ancestor that rafted from the continent [3]. The timing of colonisation and lineage divergence remain uncertain, as alternate studies suggest that differentiation occurred on the extant islands or point towards a much older split on now submerged islands [3,4].

The diverse habitats, environmental gradients and climatic fluctuations (such as El Niño events) of the Galápagos have driven the evolution of remarkable adaptations in the iguanas. Marine iguanas transitioned from a terrestrial to a marine ecology by evolving a specialised foraging strategy feeding on benthic algae, an exclusive trait amongst lizards. This species also has the unique adaptation of repeatedly shrinking and regrowing in body length, an adaptive strategy to cope with resource scarcity [5]. The pigmentation of Galápagos iguanas ranges from amelanism in the pink iguana to black colouration in the marine species, having profound effects on photoprotection, thermoregulation, sexual display and camouflage. Notably, to our knowledge, the Galápagos iguanas show no recorded incidences of cancer, despite prolonged exposure to extremely high levels of ultraviolet radiation [6].

Despite the outstanding biological features and ecological significance of the Galápagos iguanas, our understanding of their natural history and phenotypic diversity remains poor. Here, we generated the first complete genome sequences for the Galápagos iguanas as well

as for the Turks and Caicos (*Cyclura carinata*), Lesser Antillean (*Iguana delicatissima*) and Útila spiny-tailed (*Ctenosaura bakeri*) iguanas. We then performed evolutionary and demographic reconstructions to gain insight into the origins and evolution of the species. Finally, we analysed signatures of positive selection to unravel the molecular underpinnings of adaptations to life in the Galápagos.

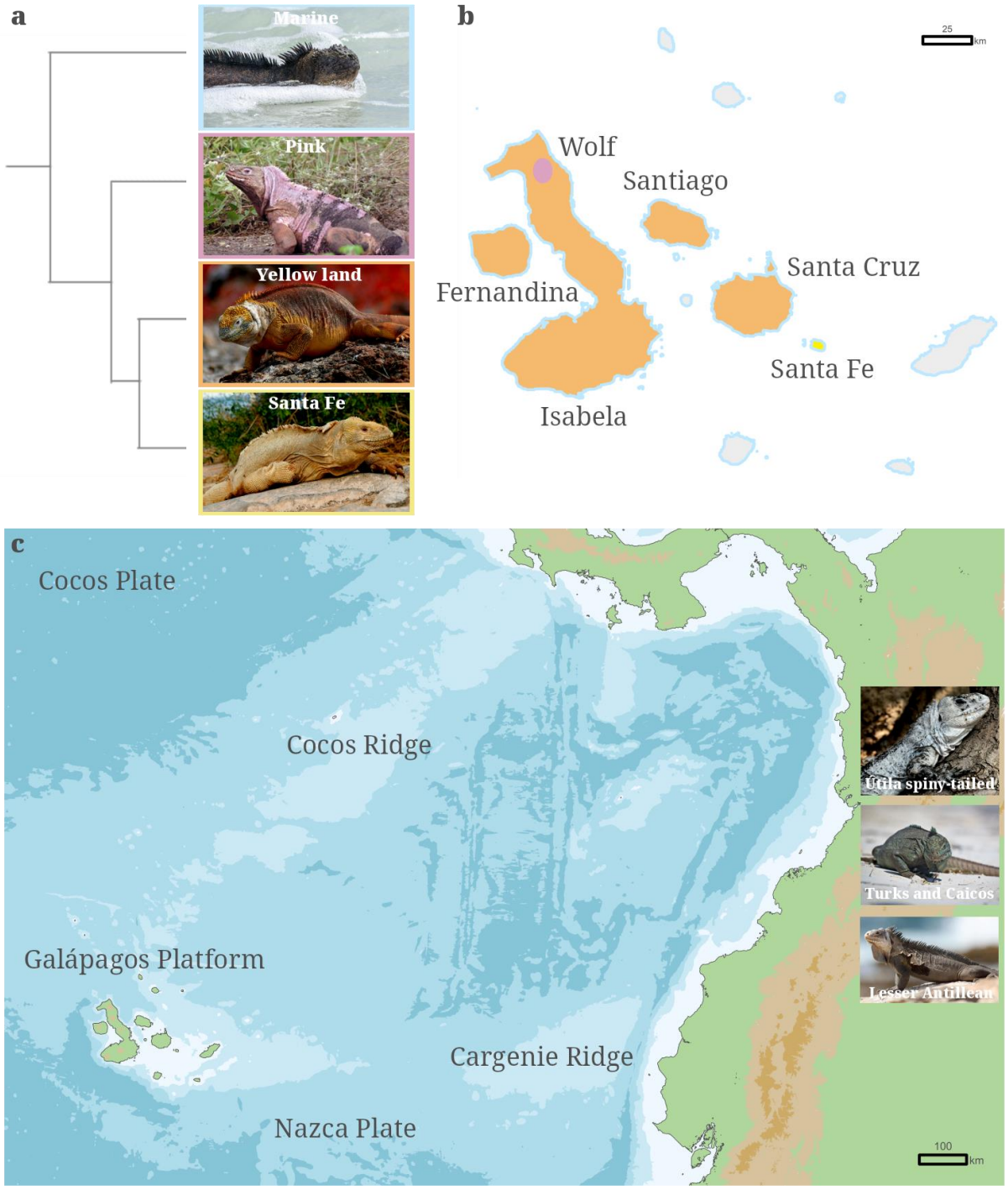


Fig. 1 Species relationships and geographic distribution of the Galápagos iguanas. (a) Cladogram of the Galápagos iguanas depicting relationships between the four species: marine iguana (*Amblyrhynchus cristatus*; blue), pink iguana (*Conolophus marthae*; pink), Galápagos yellow land iguana (*C. subcristatus;* orange) and Santa Fe land iguana (*C. pallidus*; yellow). **(b)** Distribution of the iguanas on the Galápagos islands, indicated using the respective colours from the cladogram for island shading and outline. **(c)** The geographical setting of the archipelago with shaded isobaths for 500, 1,000, 1,500 and > 1,500 m, ranging from white to dark blue. Position of images do not reflect Caribbean species distributions. Photographs by M. Arechavaleta, W. Tapia, P. Wilton, P. McFarling, J. Müller, T. Sackton.

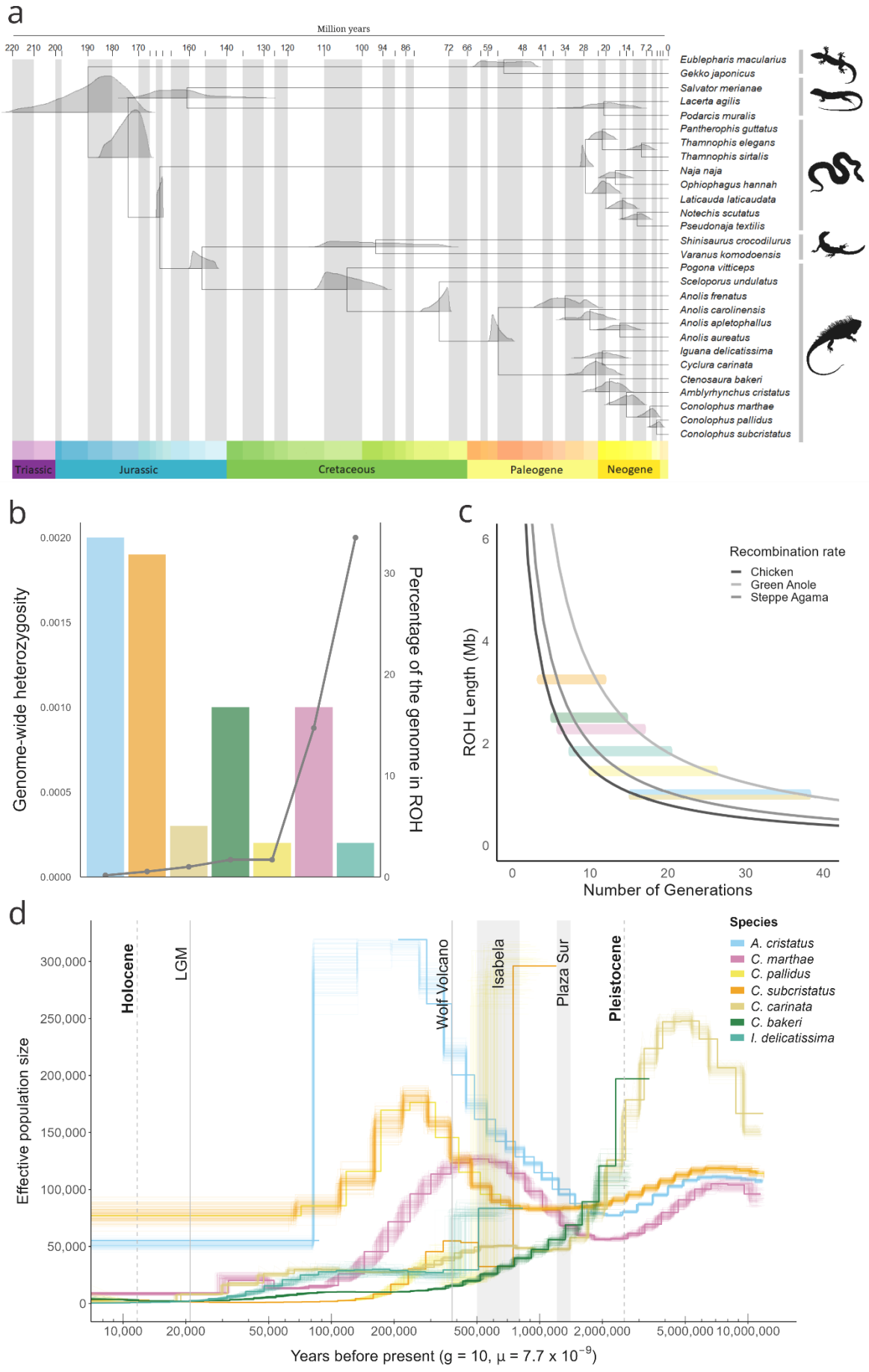
Results

Origins and evolution of the Galápagos iguanas

To uncover the origins and evolutionary history of the Galápagos iguanas, we assembled and annotated the reference genomes of the four species, as well as the Turks and Caicos, Lesser Antillean and Útila spiny-tailed iguanas. These latter three Caribbean species are hypothesised to be among the closest sister taxa to the Galápagos iguanas [3] and their inclusion enhances power to test hypotheses about the origins of, and signatures of selection in the Galápagos lineage, as well as being of conservation genomic interest in their own right. The reference genomes of the Galápagos iguanas have high gene completeness (BUSCO mean score of 93.5%) and contiguity (mean contig N50 4.7 Mb; Supplementary Table 1). Across the seven iguanas, we found an average genome size of 2.00 Gb (sd=0.03), repetitive element content of 45.9% (sd=0.8, Supplementary Table 2) and gene count of 22,002 (sd=1,067). We also assembled mitochondrial genomes for the species, which averaged 16,724 bp in length (sd=94) and had the expected 37 genes for vertebrates. In addition, transcriptome assembly and annotation of the marine iguana resulted in 27,990 functionally annotated transcripts (Supplementary Data).

Phylogenomic reconstructions, using 229 single-copy orthologs across 28 squamates, confirmed the basal position of geckos, followed by lacertids and a clade formed by snakes, iguanids and anguimorphs. Within the Galápagos, the phylogenies supported the initial land-marine split, with the pink iguana forming the most basal lineage within the terrestrial clade.

Divergence time estimation places the origin of squamates in the Jurassic period (172-215 Million years ago (Mya) 95% HPD CI), the Anguimorpha - Iguania split occurred ~153 Mya (149-157 Mya 95% HPD CI), and the split between pleurodont and acrodont iguanas at ~105 Mya (93-115 Mya 95% HPD CI). The split between the Galápagos land and marine iguanas, which illustrates the minimum amount of time that the species have inhabited the archipelago, occurred ~13 Mya (8-20 Mya 95% HPD CI). The analyses suggest the pink iguana diverged ~6 Mya (3-9 Mya 95% HPD CI) and the two other land iguanas split ~3.8 Mya (1.6-6.7 Mya 95% HPD CI). Calibration densities modelled with uniform and skewed-normal calibration distributions produced very similar divergence times for every node, whereas the Cauchy distribution resulted in much older estimates (Supplementary Table 3).



154

Fig. 2 Evolutionary and demographic history of iguanids. **(a)** Timescale of squamate evolution estimated with MCMCtree under a uniform prior probability distribution displaying the posterior distributions for each node. The geologic periods are labelled on the x-axis, with their corresponding subdivisions coloured as alternating white and grey bars. Taxonomic subdivisions are represented by vertical bars and correspond to the squamate suborders: Gekkota, Laterata, Serpentes, Anguimorpha and Iguania. **(b)** Patterns of heterozygosity reveal a recent history of inbreeding and ancient demography. Genome-wide heterozygosity (bars), calculated as Watterson's theta (θ), on the left y-axis and percentage of the genome in Runs of Homozygosity (ROH; points) on the right y-axis and species colour as per key in panel (d). **(c)** Number of generations since inbreeding estimated based on mean ROH length using the recombination rates of the chicken, the steppe agama and the green anole. **(d)** PSMC inference of the effective population size. The x-axis shows time in years before present, calibrated using a generation time (g) of 10 years and a per site per generation mutation rate (μ) of 7.7 x 10^{-9}. The y-axis represents the effective population size. The coloured lines show 100 bootstrap replicates for each species. The vertical grey bars represent the Last Glacial Maximum (LGM), the onset of the Pleistocene and Holocene, the emergence of Wolf Volcano, and the minimum and maximum emergence times of Isabela and Plaza Sur.

Genome-wide diversity and inbreeding

We calculated genome-wide heterozygosity and identified runs of homozygosity (ROH) for the iguanid species. *A. cristatus* and *C. subcristatus* had the highest genome-wide heterozygosities, whereas *C. pallidus*, *C. bakeri* and *I. delicatissima* had the lowest values (Fig. 2b, Supplementary Table 4). *C. marthae* and *I. delicatissima* displayed the highest proportions of the genome in ROH. The other five species had less than 2% of their genomes in ROH, with the lowest values corresponding to *A. cristatus* and *C. subcristatus*. The minimum ROH length was ≥ 1 Mb across all species, which is a common threshold that signals recent inbreeding [7]. The number of ROH did not increase when specifying increasingly smaller window sizes (Supplementary Table 4). This suggests recent inbreeding and is supported by the low estimate of the number of generations since inbreeding occurred (Fig. 2c). Despite having reduced proportions of ROH, *C. subcristatus* and *C. bakeri* had the longest ROH, and

inbreeding occurred as recently as five generations ago. *C. marthae* and *I. delicatissima* display relatively long ROH and inbreeding is predicted to have occurred 7.3 – 16.8 and 8.8 – 20.2 generations ago, respectively.

Intraspecific demographic histories

The demographic history of the seven iguanids was investigated by Pairwise Sequentially Markovian Coalescence (PSMC) analysis of their genomes (Fig. 2d). The Galápagos species show an initial decline in effective population size (N_e) from 6 Mya until ~2 Mya, after which *C. subcristatus* remained at constant population size and *C. marthae* and *A. cristatus* increased. The N_e of *A. cristatus* increased in the period between 2 Mya to 200 Kya and then dropped rapidly. The estimated age of emergence of Wolf Volcano of 380 Kya coincides with the decrease in *C. marthae* and increase in *C. subcristatus* N_e[2]. The analysis was unable to reconstruct the N_e of *C. pallidus* and *I. delicatissima* past 700 Kya.

Signatures of positive selection

To identify signatures of positive selection in the Galápagos iguanas, we performed positive selection scans at species and branch levels (Supplementary Table 5). The marine iguana showed the highest number of candidate genes under positive selection. In addition, the stem *Conolophus* lineage had more positively selected genes than the Galápagos branch. Gene Ontology (GO) slims were enriched in all lineages, with key terms including DNA repair and replication, inflammatory response and anatomical structure development (Supplementary Fig. 2). Gene family evolution analysis revealed the Galápagos iguanas have a high number of gene gains and losses (Supplementary Table 6), with the marine and pink iguana displaying the highest number of both gene family expansions and contractions.

Marine iguanas are anatomically and physiologically specialised to thrive in the ocean [5]. Positively selected genes in the marine iguana have functions related to oxygen management, cardiovascular control, cell protection and anatomical modifications that appear to contribute towards hypoxia tolerance in this species (Fig. 3). Overrepresented GO terms across the

positively selected genes are associated with swimming, adaptive thermogenesis, respiratory burst and haemoglobin biosynthetic process. Both paralogs encoding the alpha subunit of the Hypoxia-Inducible Factor are under positive selection. Genes linked to haemoglobin affinity and oxygen binding and transport are positively selected in this species, including *HBA*, *HMOX2* and *BPGM*. We detected positive selection in genes related to osmoregulation and antioxidation, including aquaporins and solute-carriers (Fig. 3). The marine iguana also shows positive selection signals in genes related to cardiovascular control, including genes involved in vasoregulation, coagulation and the renin-angiotensin system.

OXYGEN STORES				
Blood haemoglobin	●	●		*BPGM CA2 NCOA4*
Mitochondrial function	●	●	●	*ATP1B3 COQ10B ME2 MICU3 NDUFB8 NDUFA10 PARL PPA1*
Energy metabolism	●	●	●	*ALDH1A2 CPT1A G6PC1 GLDC PAH PGM3 PPARG TLE3*
Erythropoiesis		●		*IRAK1 PPARG CISH KLF1*
CELL PROTECTION				
Antioxidants	●	●	●	*ENPEP GPX7 GPX8 IGF1 PTGS2 STC1*
Osmoregulation	●	●	●	*AQP2 AQP6 KCNJ11 SLCs*
VASCULAR RESISTANCE				
Vasoregulation	●	●	●	*ADAMTS3 ADRB1 APLNR IL6ST PDE1A PTGS2*
Renin-angiotensin	●	●	●	*AGT AGTR2 ENPEP ANPEP PPARG PRCP*
Coagulation	●		●	*ADTRP AGT TFPI2*
ANATOMY				
Heart development	●	●	●	*ADAM22 APLNR ASB2 ELK3 MYH7 SERPINF1*
Lung development		●		*BMP2 BMPR1B HOXB5 WLS*
Tail development	●			*BMP2 PTCH2 HOXB5 HOXB9*
Motor function	●			*DAGLA SFRP1 UBA5*

Fig. 3 Genomic signatures of the transition to underwater foraging in the marine iguana. Key candidate genes under positive selection exclusively in the marine iguana with known adaptive functions in marine mammals (whale icon), high-altitude species (yak silhouette) and aquatic reptiles (snake figure). Underlined genes have undergone gene family expansion or contraction.

The marine iguana exhibits high plasticity in body size driven by food availability [6]. Positively selected genes in the marine iguana have been associated with morphological plasticity in other vertebrates, contributing to bone homeostasis, muscle development and growth regulation (Fig. 4). These phenotypic responses are mediated by molecular signalling

pathways regulated by nutrient sensing and stress hormone release [8]. The marine iguana displays genes under positive selection involved in the WNT, Hedgehog, TGF-β and Notch signalling pathways and regulatory networks, as well as in bone resorption and formation, glucose homeostasis and glucocorticoid metabolism. For instance, *AREG* promotes bone homeostasis [9] and *CPNE7* induces osteogenic differentiation to promote bone tissue regeneration [10]. Notably, *PRKAA2*, the gene encoding the master regulator of energy homeostasis AMPK, is positively selected in the marine iguana.

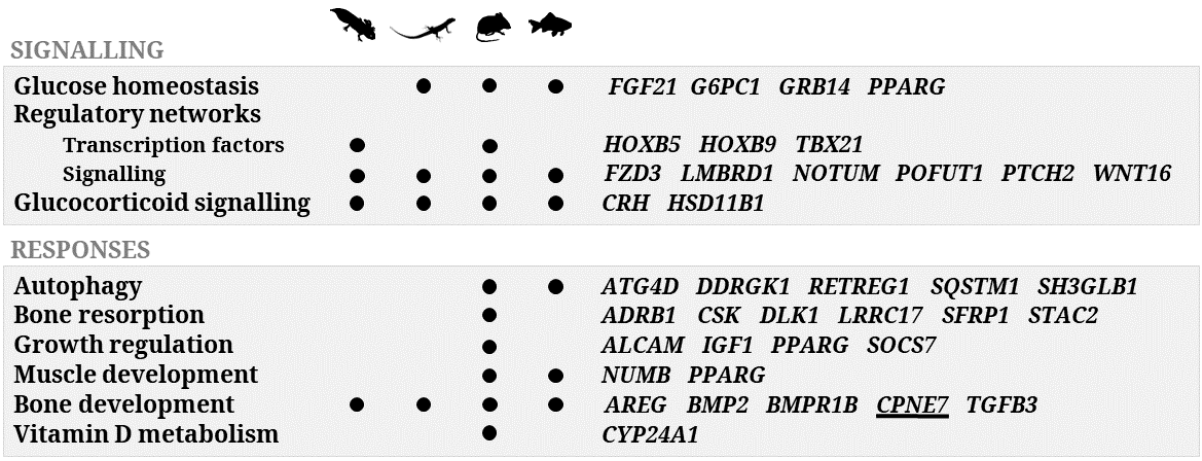


Fig. 4 Genomic signatures of morphological plasticity in the marine iguana. Key candidate genes under positive selection exclusively in the marine iguana linked to morphological plasticity with known adaptive functions in amphibians (axolotl icon), reptiles (lizard icon silhouette), mammals (mouse icon figure) and fish. The underlined gene has undergone gene family contraction.

The Galápagos iguanas show signatures of positive selection in genes involved in DNA damage sensing, repair pathways and tumour suppression (Fig. 5a). The DNA damage sensors *ATR* and DNA-PK (encoded by *PRKDC*), which transduce DNA damage signals from UV radiation to checkpoint control proteins such as *RAD17* [11], are under positive selection in most Galápagos species (Fig. 5a). We identified 17 positively selected Mitogen-Activated Protein Kinases (MAPKs) and MAPK kinase kinases in Galápagos iguanas, including the oncogenes *ARAF* and *BRAF* as well as *MAPK14*, which encodes for the critical effector in cellular stress responses p38-α [12]. p38 can be activated by the DNA damage sensors ATM and ATR or directly by UV

256 radiation, regulating downstream targets including several kinases, transcription factors and
257 cytosolic proteins (Fig. 5b). We found that the MAPK14 protein in the Galápagos iguanas has
258 an extended N-terminal region and a unique pattern of sequence variation across its MAPK
259 insert, which is a critical structural motif in the interaction with regulatory partners. Protein
260 modelling implies that the variation is likely to drive differences in the interaction of *MAPK14*
261 with its upstream kinases *MAP2K3* and *MAP2K6* (Fig. 5c). In response to DNA damage, p38
262 phosphorylates and activates the tumour suppressor protein p53, a key regulator of genomic
263 integrity [13]. Several genes from the p53 family were also under positive selection in the
264 Galápagos iguanas. Additionally, selected genes are involved in major DNA repair systems,
265 oncogenesis and tumour suppression (Fig. 5a).

266

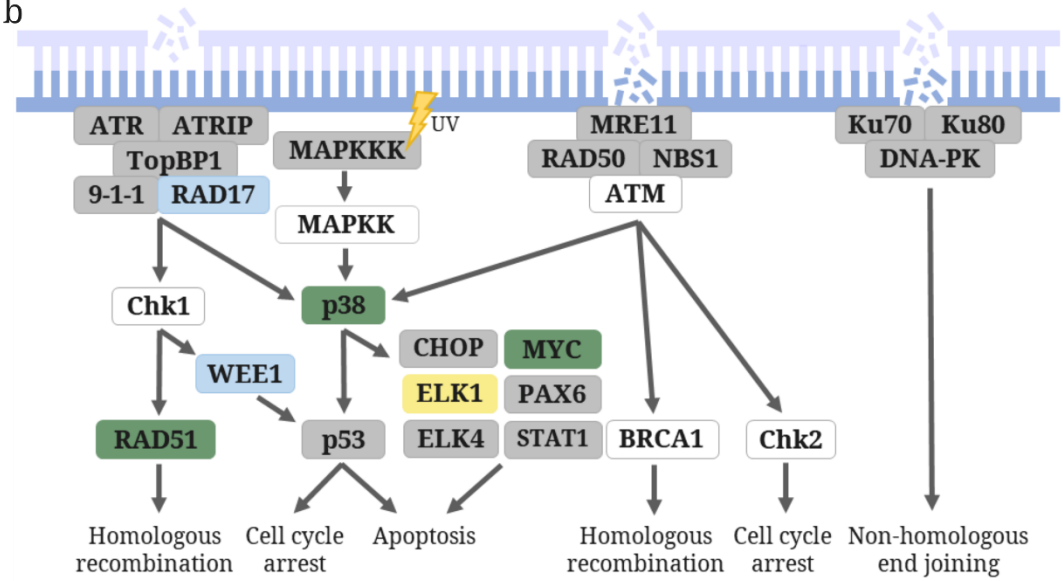
267

a

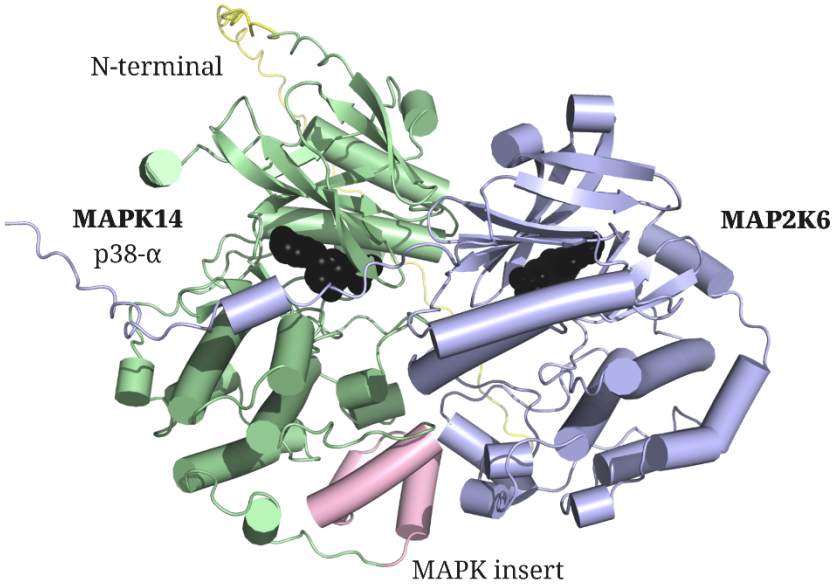
DNA REPAIR	
DNA damage sensors	*RAD17*
Base excision repair	*ADPRS APEX1 HMGB2 POLE2 POLL MUTYH RNASEH2A*
Mismatch repair	*TDG*
Nucleotide excision repair	*ACTR5 POLE2 RFC3 USP44*
Alternative end-joining	*HMCES*
Non-homologous end joining	*FEN1 PAXX POLL RNASEH2A TPK1*
Homologous recombination	*BAP1 GEN1 FEN1 IP6K1 NUDT2 NUDT5 NUDT7 PDS5A RAD51 RADX*
CELL CYCLE	
Cell signalling	*CDC5L CDK18 IGF1 IGFBP1 KIF22 MAP3K1 MAPK14 SOS1 TRIP13 MAP3K4 RNASEH2A RRM1 SOCS1*
Cell cycle regulation	*CCNA1 CDK18 RRM2B SESN1 WEE1*
Apoptosis	*CASP6 CASP7 DFFA SIVA1 USP53*
RESPONSES	
Tumorigenesis	*ALKBH2 CDKN1C MACC1 MYC PLCD1 PLCG1 RASSF3 DPH1 LATS2 MYCN RBMS1 SERPINF1 TANK KLHL4 WBP2*
Melanogenesis	*ARAF ENC1 KITLG IGF1 PAK1 MYOF1 NGFR RET SIRT7*
Longevity	*FGF21 SOCS1*

A. cristatus *C. marthae* *C. subcristatus* *C. pallidus* Galápagos Land *C. pallidus - C. subcristatus*

b



c



268

Fig. 5 Signatures of positive selection in the DNA damage response in the Galápagos iguanas. (a) Key candidate genes linked to the DNA damage response involved in damage sensing, DNA repair, cellular and general responses under positive selection in each Galápagos species or branch tested. DNA repair strategies include base excision repair, nucleotide excision repair, mismatch repair, non-homologous-end-joining and homologous recombination. Underlined genes have undergone gene family expansion or contraction. **(b)** The networks of pathways activated during the DNA damage response. Positively selected genes are coloured by lineage using the respective colours from (a), with grey indicating a gene was under positive selection in more than one lineage. Single strand breaks result in the recruitment of ATR/ATRIP, whereas double strand breaks are detected by the Ku70/Ku80 or the MRN complexes, all of which recruit and activate signal transducers. **(c)** 3D protein model of the MAPK14 (p38-α, green) and MAP2K6 (purple) complex for the pink iguana, showing the N-terminal region (yellow) and MAPK insert (pink) of MAPK14.

Signatures of positive selection were also found in key genes involved in pigment cell differentiation and pigment synthesis (Fig. 6). In the marine iguana, many genes related to melanogenesis and melanophore differentiation showed evidence of positive selection. These include the key pigmentation gene *ASIP* and transcription factors that modulate the master regulator of melanocyte development *MITF* such as *TFEC*, which also plays a crucial role in chromatophore development in reptiles [14]. Essential genes for carotene uptake and metabolism, such as beta-carotene oxygenase genes, also showed signatures of positive selection in the marine species [15]. In land iguanas, genes involved in xanthophore differentiation, pteridine synthesis, purine binding and carotenoid metabolism were under positive selection. In the pink iguana, GO terms associated to developmental pigmentation were strongly overrepresented. Evidence of positive selection were detected in transcription factors and signalling genes linked to *MITF* regulation in this species. Notably, two genes from the myosin superfamily were under positive selection in the pink iguana, with *MYO1F* also having undergone gene family expansion.

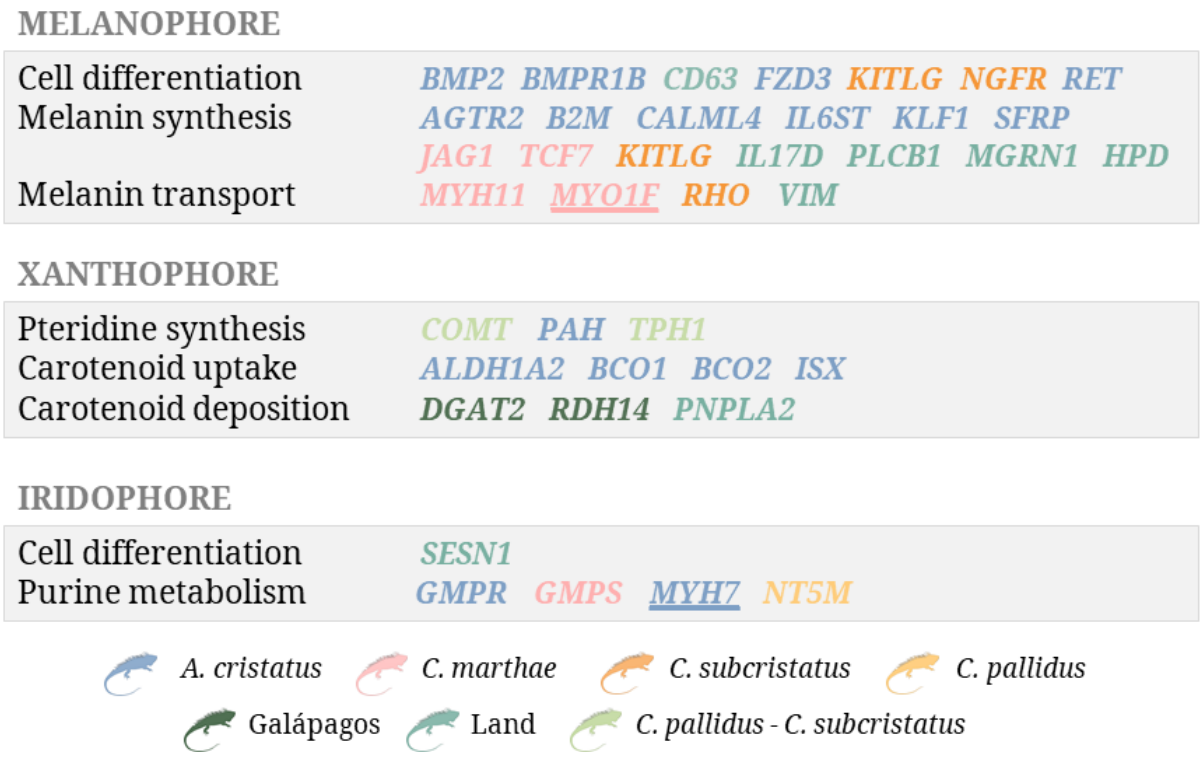


Fig. 6 Pigmentation synthesis and transport genes show lineage-specific signatures of positive selection in the Galápagos iguanas. Key candidate genes linked to pigment synthesis and differentiation of pigment cells under positive selection in each Galápagos species or branch tested. [Note: purine metabolism and binding are also involved in pteridine synthesis.] Underlined genes have also undergone gene family expansion or contraction.

Discussion

Here, we generated the first genome assemblies for the Galápagos, Turks and Caicos, Lesser Antillean and Útila spiny-tailed iguanas. These contribute towards squamate genome sequencing efforts, which remain under-sampled in genomic repositories, and support our phylogenomic and molecular evolution analyses. Our reconstructions for the squamate tree of life confirmed the basal position of geckos, followed by lacertids and the Toxicofera clade, in line with recent phylogenetic analyses [16,17]. Divergence time estimation suggests that squamates originated in the Jurassic, concordant with previous studies [18,19]. Our estimates on the origins of squamate groups largely agree with previous dating from molecular phylogenies, at the rank of suborder to genus. However, our results suggest an older split for snakes and anguimorphs [18], which likely results from our calibration having included the

316 oldest known fossil snake, which was dated 70 My earlier than previous estimates for the
317 origin of snakes [20].

318 Our results show the initial divergence among Galápagos iguanas occurred ~13 Mya. This is
319 concordant with previous findings from RADseq data, mitochondrial rRNA sequences,
320 immunological comparisons of serum albumins and electrophoresis of plasma proteins [4,21,22].
321 Divergence times estimates for the Galápagos iguanas do not overlap with extant island ages
322 [2] (< 4 Mya), implying that the species originated and diversified into the marine and land
323 lineages on now submerged islands. This idea is supported by palaeogeographical data on the
324 ontogeny of proto-Galápagos islands. Direct evidence of subaerial erosion of sunken
325 seamounts to the southeast of the extant islands ranges from 9.1 - 16 Mya [23,24]. In addition,
326 plate motion reconstructions inferred that significant subaerial landscapes existed for at least
327 the past 20 My along the Carnegie-Cocos hotspot tracks leading from America to the
328 archipelago [24] (Fig. 1c). The Galápagos also host other endemic taxa with reported divergence
329 times older than the oceanic islands, including the *Galapaganus* weevils [25], *Phyllodactylus*
330 geckos [26] and, possibly, *Pseudalsophis* racer snakes [27].

331 The monophyly of the lineage and divergence time estimates imply that the iguanas evolved
332 from a single colonisation event soon after land became available in the Galápagos platform.
333 The alternative hypothesis, whereby the marine and land iguanas evolved elsewhere, is highly
334 unlikely as it would require independent colonisation events and the extinction of both
335 ancestral stocks. Following the initial colonisation, the iguanas likely spread onto the new
336 islands in a northwest direction parallel to island formation, as reflected in the population
337 structure patterns of marine and land iguanas [28,29]. Notably, the most ancient species in the
338 land lineage is restricted to the youngest island [2] (0.5 - 0.8 Kya). In fact, the estimated origin
339 of the pink iguana at 6.1 Mya, suggests the pink and yellow land iguanas that inhabit Isabela
340 must have colonised the island separately. However, data from MacLeod et al. [30] and Paradiso
341 et al. [4] support much younger estimates and a single colonisation event of Isabela followed
342 by speciation *in situ*, with the split of *C. marthae* from the *C. subcristatus* - *C. pallidus* lineage
343 concordant with island age.

344 Our estimate for the marine - land split is significantly older than those predicted by MacLeod
345 et al. [30] (*ca*. 4.5 Mya) and Malone et al. [3] (*ca.* 5.5 Mya), although these studies also support an
346 initial divergence on now submerged islands. Their estimates were based on a reduced set of

loci, which can decrease the accuracy and robustness of divergence time estimation [31]. Moreover, both studies calibrated their molecular clock analyses by placing most of the prior probability mass near the estimated age of fossils descending from each node and with fossil age as a hard boundary for the minimum node age, which could have favoured an underestimation of divergence times [32]. Furthermore, we recover older splits within the land lineage than Paradiso et al. [4]. While Paradiso et al. [4] applied an average mutation rate for lizards to scale the species tree, we used a fossil-calibrated relaxed molecular clock, which can yield older divergence estimates by incorporating lineage-specific rate variation and paleontological constraints [33]. The youngest calibration point available for this study is 23-28.4 Mya within the snake clade and an absence of more recent points may lead to an overestimation of node ages close to the tree tips [34]. Additionally, differences in assumptions around N_e and gene flow between phylogenomic and coalescent-based approaches [4] may also affect estimates of divergence times. Further whole genome studies at the population level and a wider exploration of the influence of calibration point availability will help to clarify divergence times within the land lineage.

We report reduced genome-wide heterozygosity in *Conolophus marthae* and low levels in *C. pallidus*, comparable with the values of other critically endangered reptiles [35,36]. *A. cristatus* and *C. subcristatus* harboured moderate to high levels of heterozygosity, similar to non-endangered vertebrates. The Galápagos iguanas do not seem to display signs of severe deleterious mutations or an increased prevalence of genetic diseases, which are commonly associated with low heterozygosity and high levels of inbreeding. However, it is possible that genomic diversity is at minimum levels for *C. marthae* and, especially, *C. pallidus.* These species exist as single populations restricted to very small geographic areas, with the pink iguana consisting of only 150-270 adults [37], putting them at risk from stochastic catastrophes and further erosion of genetic diversity and adaptive potential. This may increase genomic vulnerability and inbreeding risks [36], with potential fitness reductions, although the latter may depend on the extent of past purging of deleterious mutations.

In the Galápagos iguanas, the timing of ROH coalescence coincides with the arrival of sailors in the 18th century and colonisation in the 19th century. Humans introduced invasive species including cattle, cats, dogs and rats. These have had a strong impact on iguana population sizes via resource competition, habitat alteration and predation, especially since iguanas have

negligible anti-predation behaviour due to having evolved in the absence of terrestrial predators [38]. Despite harbouring high heterozygosity, *C. subcristatus* showed the highest average ROH length (3.25 Mb), which is potentially indicative of very recent inbreeding. This pattern has also been reported in the blue-tailed skink [39] and the California condor [40]. Notably, *C. subcristatus* was sampled from Wolf Volcano, where predation pressure from non-native animals is particularly high [38] and could have led to a reduction in population size and gene flow that resulted in inbreeding within the last century. Future population genomic investigations will help refine understanding of demographic histories.

PSMC analysis revealed an increase in N_e of the Galápagos iguanas throughout the early Pleistocene, possibly highlighting favourable climatic conditions in the equatorial islands during the glacial period [2]. This rise also coincides with the emergence of new islands, which increased habitat diversity and resource availability. The asynchrony between the N_e of the pink and yellow land iguanas coincides with the emergence of Wolf Volcano [2], with the pink population declining as the yellow increases. This could have been due to interspecific competition, as suggested by stable isotope profiling, as well as the lower genomic diversity, higher level of inbreeding and smaller population size of pink iguanas [36,41]. Both iguana species then declined and remained at constant levels until present day, likely resulting from environmental and climatic conditions in the Galápagos ecosystem. The N_e of *Conolophus pallidus* and *Iguana delicatissima* could not be fully reconstructed as local heterozygosity densities were insufficient to infer ancient recombination events and coalescent times. Notably, caution must be taken when interpreting the timing and magnitude of predicted changes and absolute values of N_e, especially since PSMC estimates become less reliable in the very ancient and recent past [42].

Positive selection signatures in the Galápagos iguanas are consistent with lineage-specific adaptations. We identified a larger proportion of genes under positive selection in the marine iguana, likely prompted by the transition from a land-dwelling ancestor to the marine environment. Other aquatic species experience similar challenges associated with underwater foraging. These include apnea, hypoxia tolerance, biomechanical and energetic adaptations, elaborated cardiovascular control and maintenance of body water balance [43]. Selection scans of air-breathing marine vertebrates with highly developed dive responses, or those adapted to hypoxic conditions at high altitude, have identified similar sets of positively

409 selected genes. The master regulator of oxygen homeostasis, the Hypoxia-Inducible Factor
410 (HIF), is a notable target of hypoxia adaptation in highland species such as Andean humans
411 and Tibetan horses, sheep and wolves [44,45,46]. We found both *HIF-1α* and *HIF-2α* to be under
412 positive selection in the marine iguana. In addition, we identified positively selected genes
413 linked to osmotic and cardiovascular regulation, haemoglobin metabolism and oxygen
414 binding and transport. Interestingly, molecular dynamics simulations suggest that oxygen
415 affinity of the haemoglobin tetramer $Hb(\alpha A)_2(\beta II)_2$ is higher in the marine iguana than in the
416 land species [47]. Such genes are also under positive selection or differentially expressed in
417 Tibetan human populations, sea snakes, alpine sheep, highland geladas and northern
418 elephant seals [43,48,49].

419 The marine iguana is the only vertebrate known to repeatedly shrink and grow its body size,
420 as a strategy to manage energetic overheads during resource limiting conditions [5]. This
421 morphological plasticity starts with body size reductions during El Niño events, caused by
422 bone mass reduction via resorption and loss of cartilage and connective tissue, and is followed
423 by regrowth once favourable conditions are restored [6]. Similar reductions in bone density and
424 strength are reported in astronauts and osteoporotic mammals [50], whereas bone and
425 cartilage growth have been widely described in amphibian and reptilian limb regeneration [45].
426 Genes under positive selective pressure in the marine iguana have known functions in these
427 vertebrates related to growth regulation, bone homeostasis and muscle development.
428 Furthermore, we identified genes under positive selection that are involved in nutrient
429 sensing and stress hormone release. Glucocorticoids are the main stress hormones secreted
430 as a response to starvation and have marked effects on bone metabolism by enhancing bone
431 resorption and decreasing bone formation and differentiation of osteoblastic cells [51]. Severe
432 environmental conditions during El Niño have been documented to increase glucocorticoid
433 concentrations in marine iguanas due to food shortages [52]. We uncover key genes potentially
434 involved in this remarkable adaptation, which is driven by nutrient availability and induced by
435 the dysregulation of energy metabolism and signalling cascades.

436 Our analyses reveal signatures of positive selection in genes involved in DNA damage sensing
437 and signal transduction. DNA damage sensors, such as the positively selected genes encoding
438 ATR and DNA-PKs, detect genotoxic stress and coordinate downstream responses involved in
439 DNA repair and cell cycle regulation [11]. The positively selected *MAPK14* plays a central role in

the MAP kinase signal transduction pathway, which mediates signal transduction cascades and modulates cellular responses to extracellular stimuli, including UV radiation [53]. We identified unique sequence variation in the MAPK insert and an extended N-terminus in *MAPK14* in Galápagos iguanas, altering its interaction with upstream kinases *MAP2K3* and *MAP2K6* and potentially affecting cell stress signalling [54]. Furthermore, analysis of selective pressure variation revealed key genes involved in the main DNA repair mechanisms and tumorigenesis, including frequent targets of cancer therapies such as *BRAF* (a frequently mutated oncogene in melanomas). This would be of particular relevance to the pink iguana, which has skin patches lacking protective melanin, but experiences increased UV intensity (up to 600µWatt/cm^2) at altitudes exceeding 1,700m on Wolf Volcano [6]. The pink skin areas are devoid of pigment cells and result from blood flowing through a network of superficial dermal vases [55]. This depigmentation leads to higher rates of erythrocyte nuclear DNA damage in response to UV radiation compared to the other Galápagos iguana species [56]. To our knowledge, Galápagos iguanas have no reported incidences of cancer. It is possible that the observed variation in these genes represent an adaptive, enhanced cancer resistance mechanism, evolved in response to high UV exposure while basking.

The Galápagos iguanas provide an excellent system to test the molecular basis of pigmentation, as phenotypes range from amelanism in the pink iguana to extensive black colouration in the marine species. Our analysis revealed species-specific signatures of positive selection in key pigmentation genes. Marine iguanas are generally black and have a thick epidermal layer packed with melanocytes, which is reflected in the number of positively selected genes involved in melanocyte development and melanin synthesis and transport. In addition, males can present hues of red and green during the reproductive season, which could be associated to the positive selection in key carotene pigment metabolism genes. In land iguanas, genes involved in xanthophore differentiation, pteridine synthesis and carotenoid metabolism were under positive selection, particularly at the branch level. Xanthophores are pigment cells that contain carotene and pteridine pigments, which result in yellow to red colouration. Previous studies confirmed the presence of carotenoids in *Conolophus* blood, supporting their role in skin colour [57]. The pink iguana is one of the most unique reptiles owing to its colouration. Hatchlings exhibit a maculated pattern on a green background as an anti-predatory adaptation, although this pigmentation is lost during

ontogeny. Pigmentation loss could be mediated by transcription factors and signalling genes of key pigmentation pathways, such as the regulatory elements of *MITF* under positive selection. Moreover, dysfunctional myosins may contribute to depigmentation. These proteins are involved in pigment transport and have been implicated in the loss of pigmentation across various pigmented structures in distantly related species [58]. This is the first genetic evidence for the basis of pigmentation loss in the pink iguana, but the roles of different selective pressures and trade-offs, including sensitivity to UV damage, thermoregulation, vitamin D metabolism and mate choice in the evolution of the trait remain to be fully understood [37].

Overall, we provide a foundation for future population genomics research and conservation genomic management of these endangered reptiles. We contribute to the broader understanding of the genomic makeup of squamates, which remain relatively underrepresented in genomic datasets compared to other vertebrates. Our analyses showcase how sequencing data can be leveraged to reconstruct evolutionary and demographic histories as well as to identify the molecular basis of adaptive traits, some of which may be of broader biomedical interest. Finally, our results provide compelling evidence for the need to implement strong conservation measures to reduce the risk of erosion of genetic diversity in Galápagos iguanas.

Methods

Sampling and Extraction

Samples were collected from a single individual of *Amblyrhynchus cristatus* on Plaza Sur, *Conolophus pallidus* on Santa Fe, and both *C. marthae* and *C. subcristatus* on Wolf Volcano in Isabela. *Cyclura carinata* was sampled on Little Water Cay in the Turks and Caicos Islands. *Ctenosaura bakeri* blood was collected from an individual housed in the Bioparco di Roma in Italy that originates from the island of Útila in Honduras. The *Iguana delicatissima* sample was obtained from an individual housed at Rotterdam Zoo in The Netherlands that originates from the island of Sint Eustatius. Briefly, blood was collected from the caudal vein of adult males and preserved in a near-saturated ammonium sulphate solution buffered with sodium citrate and EDTA acid.

Blood lysates were centrifuged at 1,000 rpm for 3 minutes and the transparent supernatant was discarded. High molecular weight DNA was extracted for all species from the cell fraction using a phenol-chloroform protocol [59]. In addition, RNA was extracted for the marine iguana using the Direct-zol RNA MiniPrep kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol and eluted in a final volume of 50 µl.

Sequencing

Illumina paired-end and mate-paired sequencing of *A. cristatus* was carried out at the Malaysia Genome and Vaccine Institute (Kajang, Malaysia). Paired-end libraries were prepared based on 500 bp fragment size using Illumina TruSeq PE kit, and mate-paired libraries were prepared from 3 kbp and 8 kbp fragments using Nextera XT kit. Marine iguana libraries were sequenced on an Illumina HiSeq 2000. Illumina paired-end sequencing of *C. pallidus, C. marthae, C. subcristatus, C. carinata, C. bakeri* and *I. delicatissima* was performed at Novogene (Cambridge, UK) on an Illumina NovaSeq 6000, with paired-end libraries prepared based on 500 bp insert size using a NEBNext® DNA Library Prep. The sequencing library for long read sequencing was constructed using the SQK-LSK109 kit from Oxford Nanopore Technologies. Long-read ONT sequencing was performed at the Next Generation Sequencing Facility at the University of Leeds for *A. cristatus, C. pallidus, C. marthae, C. subcristatus, C. bakeri* and *I. delicatissima* on R9.4 minION flowcells using a MinION sequencer, and the raw fast5 data was basecalled with Guppy v.3.2.10 (Oxford Nanopore Technologies, UK). Additionally, RNA libraries for the marine iguana were constructed using a NEBNext® Ultra™ Directional RNA Library Prep Kit and sequenced on a NovaSeq 6000 at Novogene (Cambridge, UK).

Read quality assessment and filtering

Short and long read sequences were generated for all iguana species except for *C. carinata*, for which only Illumina short read sequences were produced. Short-read quality was examined using FastQC v.0.11.9 [60]. We used Trimmomatic v.0.39 [61] to remove lower quality reads with the following flags "-phred 33 ILLUMINACLIP:TruSeq3-PE.fa:2:30:10

SLIDINGWINDOW:4:30 LEADING:30 TRAILING:30 MINLEN:50". Long-read quality was analysed using NanoPlot [62]. Long ONT reads were left untrimmed as this resulted in higher quality assemblies.

Genome assembly

Eleven short, hybrid and long genome assembly pipelines were tested under a variety of assembly parameters for each method. For short-read assembly, MaSuRCA v.4.0.2 [63], SOAPdenovo v.2.04 [64] and Sparse v.20160205 [65] were tested. The hybrid assemblers that incorporate short and long read data used were dbg2olc v. 20180222 [66], MaSuRCA v.4.0.2 [63] and Wengan v.0.2 [67], as well as the long-read assemblers Flye v. 2.8.3 [68], MECAT v.1.0 [69], Miniasm v.0.3 [70], Shasta v.0.7.0 [71] and wtdbg2 v.2.5 [72]. The output from the best performing long-read assembler Flye was polished with short reads using Racon v.1.4.11 [73] and Pilon v.1.23 [74], as well as with long reads using Medaka v.1.2.3 (https://github.com/nanoporetech/medaka). The genome assembly statistics for all pipelines tested can be found in Supplementary Table 1.

Genome annotation

Repetitive elements were identified using RepeatModeler v.2.0.2a [75]. Custom repeat libraries were created for each species and were combined with repeat sequences from *Anolis carolinensis* and were classified using the module RepeatClassifier. Genomes were soft masked with RepeatMasker v.4.1.2 (http://www.repeatmasker.org/) using the repeat libraries. BRAKER v.2.1.6 [76,77] was used to predict protein-coding genes using RNA-seq data and protein homology information. Firstly, gene prediction was performed in BRAKER1 using RNA-seq data from marine iguanas, including the newly generated data and publicly available Ion Torrent RNA-seq reads (NCBI SRA at BioProject PRJNA602224). The RNA-seq datasets were aligned to the soft masked genomes using STAR [78]. Given the high alignment rate of marine iguana reads to the other iguanas' genomes (>85% uniquely mapped reads) and the lack of RNA-seq data for the other species, this data was used as evidence for this gene prediction step in all species. Secondly, an independent gene prediction was run in BRAKER2

559 using protein homology information from the OrthoDB v.10 database [79]. Subsequently, the
560 two gene predictions were integrated using TSEBRA v.1.0.3 [80] with default settings. TSEBRA
561 selects transcripts from the BRAKER1 and BRAKER2 annotations to generate a joint prediction
562 based on both RNA-seq and protein evidence.

563 Functional annotation of the protein-coding genes was performed combining the search for
564 full-sequence similarity using BLAST against the NCBI squamate protein database and
565 targeted characterisation of functional elements through the InterProScan pipeline [81]. The
566 BLAST and InterProScan outputs were loaded onto the genome annotations using the script
567 agat_sp_manage_functional_annotation.pl from the AGAT package
568 (https://github.com/NBISweden/AGAT). To obtain 'high-confidence' sequences from the
569 gene annotations, Diamond v.0.9.24 [82] searches were conducted against the *Anolis*
570 *carolinensis* reference proteome, the UniProtKB/Swiss-Prot database, and the
571 UniProtKB/TrEMBL database (The UniProt Consortium, 2021). The BUSCO [83] scores were
572 computed in the 'protein' mode for each of the subsets derived from the unique matches
573 obtained when searching against the three databases. The subsets with the highest BUSCO
574 scores across all species corresponded to the unique matches against the UniProtKB/TrEMBL
575 database.

576

577 *Mitogenome assembly and annotation*

578 *De novo* assembly of the mitochondrial genome was performed with the adapter-trimmed
579 Illumina sequencing reads using the GetOrganelle pipeline [84]. GetOrganelle was run under
580 default options with a range of k-mers of 21, 45, 65, 85 and 105. Mitogenome annotation and
581 visualisation was performed using the MITOS WebServer [85].

582

583 *Transcriptome assembly and annotation*

584 Both *de novo* and reference-based transcriptome assemblies were performed for *A. cristatus*.
585 For the reference-guided assembly, the RNA-seq reads were mapped onto the genome
586 assembly of *A. cristatus*. The two assembly strategies were performed using Trinity v.2.14.0
587 [86] under default settings with the 'trimmomatic' flag. Publicly available raw Ion Torrent RNA-

seq reads from heart, muscle, blood, skin and lung tissues of a marine iguana from Genovesa island were also added to the dataset (NCBI SRA at BioProject PRJNA602224). The *de novo* and genome-guided, trimmed and untrimmed assemblies were performed on both the newly generated data and on the concatenation with the publicly available reads. The assemblies were filtered to remove redundant and poorly constructed transcripts using TransDecoder v.5.5.0 (https://github.com/ TransDecoder/TransDecoder) and CD-Hit v.4.8.1 [87].

The transcriptome was annotated using Trinotate v.3.2.2 [88]. Coding sequences were translated into amino acid sequences using TransDecoder v.5.5.0. Trinotate used BLAST [89] for homology search and HMMER3 [90] for sequence feature annotation. rnammer v.1.2 [91] was run for RNA classification, SignalP v.5.0 [92] for signal peptide identification and tmHMM v.2.0c [93] for the prediction of transmembrane domains. The databases queried were the UniProtKB/Swiss-Prot database (The UniProt Consortium, 2021), the Pfam protein family database [94], the SignalP signal peptide database, the eggNOG database of nested orthologous gene groups [95], the tmHMM transmembrane domain database, the Gene Ontology (GO) knowledgebase (The Gene Ontology Consortium, 2021) and the Kyoto Encyclopedia of Genes and Genomes [96] (KEGG).

Ortholog identification and filtering

The predicted proteomes or coding sequences of all squamate reptile species with assembled genomes available in April 2023 were obtained (Supplementary Table). The coding sequences were 'cleaned' (to remove those sequences that were not divisible by 3) and translated into amino acid sequences using the vespa_clean.py and vespa_translate.py commands from VESPA [97]. The longest transcript variant per gene was extracted for each proteome using the custom script *p_LongestIsoformInFastaFile.py*. OrthoFinder v.2.5.4 was used to infer orthogroups among the squamate dataset [98].

The single-copy orthologs identified were filtered to obtain true orthologs with informative signals. Firstly, alignments were constructed for each single-copy orthogroup using MAFFT v. 7.505 [99], MUSCLE v.5.1 [100] and PRANK v. 170427 [101] using default settings. The best fitting alignment per orthogroup was selected based on the highest normalized mean distance score [102] and the lowest number of gaps. The alignments were then trimmed using TrimAL v. 1.4.1

[103] with the 'automated1' flag. Sequence saturation was calculated using the 'saturation' function in PhyKIT [104] and orthogroups with values below 0.8 were removed. Clan_check (https://github.com/ChrisCreevey/clan_check) was used to reduce noise from hidden paralogy and genes that violated uncontroversial splits at the suborder level were filtered out and not used in further analysis. Likelihood mapping was performed using the command 'lmap' in IQ-TREE [105] to remove genes with insufficient phylogenetic signal. Lastly, compositional heterogeneity was assessed for each orthogroup using a chi-squared test in IQ-TREE. Ortholog filtering resulted in a dataset of 229 single-copy orthologs across 28 squamates.

Phylogenomic reconstructions and divergence time estimation

Phylogenomic analyses on the squamate single-copy orthologs were performed using maximum likelihood in IQ-TREE with automatic model selection, Bayesian analysis in PhyloBayes under the CAT+GTR model with a gamma distribution of four categories and a supertree approach in ASTRAL-III [106,107].

Divergence time estimation was performed within a framework of Bayesian inference. Three sets of analyses were used to test the robustness of calibration choice. Strategy 1 modelled the calibrations as uniform distributions with soft upper bounds, strategy 2 modelled the nodes using a skewed-normal distribution, and strategy 3 used a truncated Cauchy distribution with a long tail. Then, ten fossil calibration points that covered most of the deep branches within the tree were identified (Supplementary Table 7). MCMCtree from the PAML v.4.10.6 package [108] was used to obtain the posterior time estimates on the fixed tree topology produced by the phylogenetic reconstructions. Two independent MCMC chains were run with a 2,000 generation burn-in, subsequently sampling every 10 generations until 20,000 samples were collected. MCMCtree estimated the posterior mean divergence times and the 95% highest posterior density credible intervals on divergence times for the three calibration strategies. Additionally, the analyses were run without sequence data to generate the time prior, and the samples from the prior and the posterior were compared. The divergence time estimates were plotted with the MCMC.tree.plot function from the MCMCtreeR package in R v.4.1.1 [109]. We found that the choice of prior calibration

distributions had an impact on the posterior estimates of divergence times. The concordance in age and substitution rate estimates between the uniform and skewed-normal distributions, along with the overestimations produced by the Cauchy distribution, suggest that the first distributions are a closer representation of the true ages of the splits.

Genome-wide diversity and inbreeding

Genome-wide heterozygosity levels and runs of homozygosity (ROH) were computed for each iguanid using ROHan [110]. Firstly, the raw Illumina reads were mapped onto each conspecific genome assembly using BWA-mem v.0.7.17 [111]. ROHan was run three independent times on the mapped bam files using three different window sizes (500 kbp, 1 Mb, 2 Mb), under default options for the other parameters.

The number of generations since inbreeding events was calculated using the formula G = 100 / 2rL, whereby G = number of generations, r = recombination rate and L = length of ROH in Mbp [112]. This model assumes that ROH length decreases over time and that recombination rate is constant across the genome. Published estimates of reptile recombination rates were used as an approximation since genome-wide recombination rates are unavailable for Iguanidae, including chicken [113], green anole [114] and steppe agama [39]. The minimum, average and maximum lengths of the ROH used to calculate the number of generations since inbreeding were estimated by ROHan.

Coalescent models of demographic history

PSMC analysis was applied to estimate the demographic history of the seven iguanid species [109]. Illumina reads were aligned to the corresponding reference genome assembly using Minimap2 [115]. Variants were called with SAMtools v.1.7 mpileup using the call command [116]. The resulting VCF files were converted into diploid consensus FASTSQ files using VCFtools [116], setting the minimum read depth '-d' to a third of the average read depth and the maximum read depth '-D' to twice the average depth. The FASTQ files were then converted to a 'psmc' file using the PSMC command fq2psmcfa with the '-q 20' flag 109. N_e was inferred across 64 atomic time intervals (4+25*2+4+6) using the '-p' parameter with the options '-N 25 –t20 -r5'.

100 bootstrap replicates were performed for each species by splitting the 'psmcfa' file and randomly sampling from these regions. The model was scaled using a generation time of ten years to account for the variability within the species and a substitution rate of 7.7 x 10-9 per year per site per generation, based on the average neutral substitution rate estimated for squamate reptiles [117].

Scanning for signatures of positive selection

The squamate database generated for phylogenomic reconstruction was filtered to remove the proteomes with BUSCO scores <85%, resulting in a set of 22 species (Supplementary Table 5). OrthoFinder v.2.5.4 was run under default settings to infer orthologous groupings among the dataset [98]. Orthogroups representing less than seven species were discarded, as likelihood ratio tests performed with less than seven species have low power in detecting positive selection at amino acid sites [118]. Multiple sequence alignments were constructed for each orthogroup using MAFFT v. 7.505, MUSCLE v.5.1 and PRANK v. 170427. Alignments for each orthogroup were selected based on the highest normalized mean distance score and the lowest number of gaps. Multi-gene orthogroups were paralog-filtered to derive subtrees with 1:1 orthologous relationships, this was done using phylogenetic tree-based orthology inference in PhyloPyPruner with the maximum inclusion method (https://github.com/fethalen/phylopypruner). Following, the isolation of putative single-copy orthologs, including those identified by OrthoFinder and those obtained by paralogy pruning, were tested for hidden paralogy. Firstly, IQ-TREE was run with automatic model selection on a gene-by-gene basis using ModelFinder [119] and 1,000 ultrafast bootstrap replicates. Robinson-Foulds distances between each gene tree and the canonical species tree were estimated using Clann [120] and only genes with distances below 0.5 were retained.

The single-copy orthologs were converted to nucleotide format using the map_alignments command from the VESPA package. The orthologs were further filtered using the custom python script *p_geneFilter.py* to ensure that each gene family retained for further analysis contained: all Galápagos species, at least two non-Galápagos iguanids, and two non-iguanid species. This yielded 6,682 single-copy genes for the selective pressure analysis. Signatures of positive selection were tested on the ortholog set using the aBSREL method from the HyPhy

package [121,122]. Selection was evaluated on a lineage-specific level for multiple lineages by annotating the input trees with LabelTrees (https://github.com/veg/hyphy-analyses/tree/master/LabelTrees). The foreground lineages specified were the four Galápagos species along with the three internal nodes within the Galápagos clade (Galápagos root node, land iguana node and *Conolophus subcristatus – C. pallidus* node). aBSREL adjusts p-values by applying the Holm-Bonferroni correction to account for multiple branch testing. To account for multiple gene testing, we applied an additional correction by calculating the false discovery rate (FDR) using the *p.adjust* function in R v.4.1.1, following the Benjamini-Hochberg procedure, with an FDR threshold of 0.05.

Gene family expansion and contraction were examined with the Computational Analysis of gene Family Evolution (CAFE) v.5 using a Poisson distribution and estimating the rate of change of evolution (lambda parameter) for each species [123].

Functional characterisation of candidate genes under positive selection

The R library biomaRt v.2.50.3 was used to retrieve the gene and gene ontology (GO) information from the BioMart databases for the green anole, common wall lizard, mainland tiger snake and eastern brown snake [124]. Genes were tested for functional enrichment of GO IDs, GO slims and KEGG terms. Fisher's exact tests were performed with Bonferroni multiple test correction to estimate GO ID and GO slim term over-representation with the script *GOEnrichment.R*. REVIGO was used to summarise the GO terms based on similarity [125]. The *enrichKEGG* function from the R package clusterProfiler v.4.2.2 was used to assess whether genes were significantly enriched for a particular KEGG pathway, under default Benjamini-Hochberg adjustment of p-values [126]. Lastly, the PANTHER (Protein ANalysis THrough Evolutionary Relationships) classification system was used to classify the HyPhy input genes and positively selected genes by biological process, molecular function and protein class [127]. The positively selected genes were searched in the literature to identify genotype-phenotype associations, prioritising the functionally enriched set and genes with demonstrated functional associations with the adaptive traits in other vertebrates. Gephebase was interrogated to identify genetic variants associated with trait variation [128]. Silhouettes for the figures were taken from PhyloPic (https://www.phylopic.org/).

737 Molecular model of *Conolophus marthae* MAPK14-MAP2K6 complex was generated using
738 AlphaFold3 (http://alphafoldserver.com), with high confidence in the interface between the
739 kinases (pairwise iPTM 0.72). Image was produced using PyMOL 3.7 (http://www.pymol.org).

740

741 **Data availability**

742 The sequencing data (genomic and transcriptomic raw reads) generated in this study are
743 available through NCBI repositories linked to BioProject accession number xxxxxxx. The
744 transcriptome annotation SQL database is available from the Research Data Repository at the
745 University of Leeds xxxxxxx.

746

747 **Code availability**

748 The bioinformatic scripts are available from the Research Data Repository at the University of
749 Leeds xxxxxxx.

750

751 **Acknowledgements**

752 We thank the Galápagos National Park Directorate and park rangers for their invaluable
753 support. We are grateful to Morag Raynor and Carolina Lascelles for library preparation and
754 sequencing at the Next Generation Sequencing Facility of the University of Leeds. We thank
755 the Rotterdam Zoo for sourcing the Lesser Antillean iguana sample. We thank the Bioparco of
756 Rome for sourcing the Útila spiny-tailed iguana sample. We received support by the Italian
757 Ministry of Research and University (MIUR, 2003) with the “Brain Gain Program” grant to GG.
758 We received support from the San Diego Zoo Wildlife Alliance through a grant to GPG. We
759 acknowledge the MOSTI Science Fund (02-05-20-SF11119) funded to MNMI for generating
760 short read sequences of the marine iguana. JLD was supported by a PhD scholarship from the
761 Faculty of Biological Sciences of the University of Leeds. JLD thanks the Priestley Centre for
762 Climate Futures for a travel grant. JLD thanks Elizabeth Duncan from the University of Leeds
763 for constructive feedback. JLD also thanks Alan Beavan from the University of Nottingham,

Bede Constantinides from the University of Oxford and Peter Mulhair from Trinity College Dublin for sharing code for the phylogenomic analysis. CP was supported by a PhD scholarship and travel grant from the School of Doctorate in Ecology and Evolutionary Biology, University of Rome Tor Vergata. Bionformatics work was performed on the University of Leeds High Performance Computing ARC4 cluster.

Author contributions

JLD, MJO'C, GG and SJG designed the research. CP, GC, GG and GPG collected the samples. JLD and CP performed the DNA and RNA extractions. JLD assembled and annotated the reference genomes, with bioinformatics support from IMC. JLD performed the evolutionary, demographic and selection pressure analyses, with guidance from MJO'C, PG, GG and SJG. RB carried out the protein structure modelling. JLD wrote the primary manuscript, and all authors edited and revised the final version.

Ethics declaration

Animal manipulation and blood sampling were performed according to a protocol that minimized animal stress, in accordance with the European Community guidelines and with the approval of the Galápagos National Park Directorate and the Turks and Caicos Islands Department of Environment and Coastal Resources. *Conolophus* and *Amblyrhynchus* samples were exported under permits granted by the Galápagos National Park and the CITES-Authority to GG (permit no. 050-03-PNG and 004-05-PNG; EC9119940 and 0211914 CITES EXPORT). Samples were imported under the CITES permits IT/IM/2003/MCE/02328 and IT/IM/2006/MCE/07225 to GG. *Cyclura carinata* samples were exported under permits granted by the Turks and Caicos Department of Environment and Coastal Resources and the CITES authority to GG (permit no. PLS-W-2021 8); samples were imported under the CITES permit IT/IM/2022/MCE/03130 to GG.

References

1. Ali, J. R., & Fritz, U. Origins of Galápagos' land-locked vertebrates: What, whence, when, how? *Biological Journal of the Linnean Society* **134**, 261–284 (2021). https://doi.org/10.1093/biolinnean/blab085
2. Geist, D. J., Snell, H., Snell, H., Goddard, C., & Kurz, M. D. A Paleogeographic Model of the Galápagos Islands and Biogeographical and Evolutionary Implications. In *The Galápagos,* 145–166. (American Geophysical Union, 2014). https://doi.org/10.1002/9781118852538.ch8
3. Malone, C. L., Reynoso, V. H., & Buckley, L. Never judge an iguana by its spines: Systematics of the Yucatan spiny tailed iguana, *Ctenosaura defensor* (Cope, 1866). *Molecular Phylogenetics and Evolution* **115**, 27–39 (2017). https://doi.org/10.1016/j.ympev.2017.07.010
4. Paradiso, C., Gratton, P., Trucchi, E., López-Delgado, J., Gargano, M., Garizio, L., Carr, I. M., Colosimo, G., Sevilla, C., Welch, M. E., Firdaus-Raih, M., Mat-Isa, M. N., Goodman, S. J., & Gentile, G. Genomic insights into the biogeography and evolution of Galápagos iguanas. *Molecular Phylogenetics and Evolution* **204**, 108294 (2025). https://doi.org/10.1016/j.ympev.2025.108294
5. Wikelski, M., & Thom, C. Marine iguanas shrink to survive El Niño. *Nature* **403**, 37–38 (2000). https://doi.org/10.1038/47396
6. Di Giacomo, C., Pucillo, L., Sevilla, C., Fucci, G., Massoud, R., Bernardini, S., Fraziano, M., & Gentile, G. 25-Hydroxyvitamin D Plasma Levels in Natural Populations of Pigmented and Partially Pigmented Land Iguanas from Galápagos (*Conolophus* spp.). *BioMed Research International* **e7741397** (2022).
7. Selli, A., Ventura, R. V., Fonseca, P. A. S., Buzanskas, M. E., Andrietta, L. T., Balieiro, J. C. C., & Brito, L. F. (2021). Detection and Visualization of Heterozygosity-Rich Regions and Runs of Homozygosity in Worldwide Sheep Populations. *Animals* **11**. https://doi.org/10.3390/ani11092696
8. Vlashi, R., Zhang, X., Wu, M., & Chen, G. Wnt signaling: Essential roles in osteoblast differentiation, bone metabolism and therapeutic implications for bone and skeletal disorders. *Genes & Diseases* **10**, 1291–1317 (2023). https://doi.org/10.1016/j.gendis.2022.07.011

9. Wang, X., Wang, S., Mu, H., Yang, C., Dong, W., Wang, X., & Wang, J. Macrophage-derived amphiregulin promoted the osteogenic differentiation of chondrocytes through EGFR/Yap axis and TGF-β activation. *Bone* **190**, 117275 (2025). https://doi.org/10.1016/j.bone.2024.117275
10. Lee, D., Park, K. S., Yoon, G. J., Lee, H. J., Lee, J. Y., Park, Y. S., Park, J. C., Lee, G., Chung, C. P., & Park, Y. J. Identification of cell-penetrating osteogenic peptide from copine-7 protein and its delivery system for enhanced bone formation. Journal of *Biomedical Materials Research* **107**, 2392-2402 (2019).
11. Groelly, F. J., Fawkes, M., Dagg, R. A., Blackford, A. N., & Tarsounas, M. Targeting DNA damage response pathways in cancer. *Nature Reviews Cancer* **23**, 78–94 (2023). https://doi.org/10.1038/s41568-022-00535-5
12. Burguer, K., Ketley, R. F., & Gullerova, M. Beyond the Trinity of ATM, ATR, and DNA-PK: Multiple Kinases Shape the DNA Damage Response in Concert With RNA Metabolism. *Frontiers in Molecular Biosciences* **6** (2019). https://doi.org/10.3389/fmolb.2019.00061
13. Gong, X., Liu, A., Ming, X., Deng, P., & Jiang, Y. UV-induced interaction between p38 MAPK and p53 serves as a molecular switch in determining cell fate. *FEBS Letters* **584**, 4711-4716 (2010). https://doi.org/10.1016/j.febslet.2010.10.057
14. Tzika, A. C. On the role of TFEC in reptilian coloration. *Frontiers in Cell and Developmental Biology* **12**, 1-6 (2024). https://doi.org/10.3389/fcell.2024.1358828
15. Toews, D. P. L., Hofmeister, N. R., & Taylor, S. A. The Evolution and Genetics of Carotenoid Processing in Animals. *Trends in Genetics* **33**, 171–182 (2017). https://doi.org/10.1016/j.tig.2017.01.002
16. Burbrink, F. T., Grazziotin, F. G., Pyron, R. A., Cundall, D., Donnellan, S., Irish, F., Keogh, J. S., Kraus, F., Murphy, R. W., Noonan, B., Raxworthy, C. J., Ruane, S., Lemmon, A. R., Lemmon, E. M., & Zaher, H. Interrogating Genomic-Scale Data for Squamata (Lizards, Snakes, and Amphisbaenians) Shows no Support for Key Traditional Morphological Relationships. *Systematic Biology* **69**, 502–520 (2020). https://doi.org/10.1093/sysbio/syz062
17. Simões, T. R., Vernygora, O., Caldwell, M. W., & Pierce, S. E. Megaevolutionary dynamics and the timing of evolutionary innovation in reptiles. *Nature Communications* **11** (2020). https://doi.org/10.1038/s41467-020-17190-9

18. Pyron, R. A., & Burbrink, F. T. Early origin of viviparity and multiple reversions to oviparity in squamate reptiles. *Ecology Letters* **17**, 13–21 (2014). https://doi.org/10.1111/ele.12168

19. Zheng, Y., & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Molecular Phylogenetics and Evolution* **94**, 537–547 (2016). https://doi.org/10.1016/j.ympev.2015.10.009

20. Caldwell, M. W., Nydam, R. L., Palci, A., & Apesteguía, S. The oldest known snakes from the Middle Jurassic-Lower Cretaceous provide insights on snake evolution. *Nature Communications* **6**, (2015). https://doi.org/10.1038/ncomms6996

21. Rassmann, K. Evolutionary Age of the Galápagos Iguanas Predates the Age of the Present Galápagos Islands. *Molecular Phylogenetics and Evolution* **7**, 158–172 (1997). https://doi.org/10.1006/mpev.1996.0386

22. Wyles, J. S., & Sarich, V. M. Are the Galápagos iguanas older than the Galápagos? Molecular evolution and colonization models for the archipelago. In *Patterns of Evolution in Galapagos organisms,* 177–185. (American Division for the Advancement of Science, 1983).

23. Christie, D. M., Duncan, R. A., McBirney, A. R., Richards, M. A., White, W. M., Harpp, K. S., & Fox, C. G. Drowned islands downstream from the Galapagos hotspot imply extended speciation times. *Nature* **355**, 246–248 (1992). https://doi.org/10.1038/355246a0

24. Orellana-Rovirosa, F., & Richards, M. Emergence/Subsidence Histories Along the Carnegie and Cocos Ridges and Their Bearing Upon Biological Speciation in the Galápagos. *Geochemistry, Geophysics, Geosystems* **19**, 4099–4129 (2018). https://doi.org/10.1029/2018GC007608

25. Sequeira, A. S., Lanteri, A. A., Scataglini, M. A., Confalonieri, V. A., & Farrell, B. D. Are flightless *Galapaganus* weevils older than the Galápagos Islands they inhabit? *Heredity* **85** (2000). https://doi.org/10.1046/j.1365-2540.2000.00690.x

26. Torres-Carvajal, O., Barnes, C. W., Pozo-Andrade, M. J., Tapia, W., & Nicholls, G. Older than the islands: Origin and diversification of Galápagos leaf-toed geckos (Phyllodactylidae: *Phyllodactylus*) by multiple colonizations. *Journal of Biogeography* **41**, 1883–1894 (2014). https://doi.org/10.1111/jbi.12375

27. Zaher, H., Yánez-Muñoz, M. H., Rodrigues, M. G., Graboski, R., Machado, F. A., Altamirano-Benavides, M., Bonatto, S. L., & Grazziotin, F. G. Origin and hidden diversity within the poorly known Galápagos snake radiation (Serpentes: *Dipsadidae*). *Systematics and Biodiversity* **16**, 614-642 (2018). https://doi.org/10.1080/14772000.2018.1478910

28. Gentile, G., Fabiani, A., Marquez, C., Snell, H. L., Snell, H. M., Tapia, W., & Sbordoni, V. An overlooked pink species of land iguana in the Galapagos. *Proceedings of the National Academy of Sciences* **106**, 507–511 (2009). https://doi.org/10.1073/pnas.0806339106

29. Steinfartz, S., Glaberman, S., Lanterbecq, D., Russello, M. A., Rosa, S., Hanley, T. C., Marquez, C., Snell, H. L., Snell, H. M., Gentile, G., Dell'Olmo, G., Powell, A. M., & Caccone, A. Progressive colonization and restricted gene flow shape island-dependent population structure in Galápagos marine iguanas (*Amblyrhynchus cristatus*). *BMC Evolutionary Biology* **9**, 297 (2009). https://doi.org/10.1186/1471-2148-9-297

30. MacLeod, A., Rodríguez, A., Vences, M., Orozco-terWengel, P., García, C., Trillmich, F., Gentile, G., Caccone, A., Quezada, G., & Steinfartz, S. Hybridization masks speciation in the evolutionary history of the Galápagos marine iguana. *Proceedings of the Royal Society B: Biological Sciences* **282**, 20150425 (2015). https://doi.org/10.1098/rspb.2015.0425

31. Filipski, A., Murillo, O., Freydenzon, A., Tamura, K., & Kumar, S. Prospects for Building Large Timetrees Using Molecular Data with Incomplete Gene Coverage among Species. *Molecular Biology and Evolution* **31**, 2542–2550 (2014). https://doi.org/10.1093/molbev/msu200

32. Rannala, B., & Yang, Z. Inferring Speciation Times under an Episodic Molecular Clock. *Systematic Biology* **56**, 453–466 (2007). https://doi.org/10.1080/10635150701420643

33. Saladin, B., Leslie, A. B., Wüest, R. O., Litsios, G., Conti, E., Salamin, N., & Zimmermann, N. E. Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. *BMC Evolutionary Biology* **17**, 1-15 (2017). https://doi.org/10.1186/s12862-017-0941-z

34. McCartney, J. A., Stevens, N. J., & O'Connor, P. M. The Earliest Colubroid-Dominated Snake Fauna from Africa: Perspectives from the Late Oligocene Nsungwe Formation of Southwestern Tanzania. *PLOS ONE* **9**, e90415 (2014). https://doi.org/10.1371/journal.pone.0090415.

35. Green, R. E., Braun, E. L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M. W., John, J. A. S., Capella-Gutiérrez, S., Castoe, T. A., Kern, C., Fujita, M. K., Opazo, J. C.,

Jurka, J., Kojima, K. K., Caballero, J., Hubley, R. M., Smit, A. F., Platt, R. N., … Ray, D. A. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346** (2014). https://doi.org/10.1126/science.1254449

36. Iannucci, A., Benazzo, A., Natali, C., Arida, E. A., Zein, M. S. A., Jessop, T. S., Bertorelle, G., & Ciofi, C. Population structure, genomic diversity and demographic history of Komodo dragons inferred from whole-genome sequencing. *Molecular Ecology* **30**, 6309–6324 (2021). https://doi.org/10.1111/mec.16121
37. Garizio, L., Gargano, M., Colosimo, G., Gratton, P., Gerber, G. P., Lewbart, G., Sevilla, C., & Gentile, G. First evidence of recruitment in critically endangered Galápagos pink land iguanas (*Conolophus marthae*). *Conservation Science and Practice* **6**, e13108 (2024). https://doi.org/10.1111/csp2.13108
38. Gentile, G., Marquez, C., Snell, H. L., Tapia, W., & Izurieta, A. Conservation of a New Flagship Species: The Galápagos Pink Land Iguana (*Conolophus marthae* Gentile and Snell, 2009). In *Problematic Wildlife,* 315–336. (Springer International Publishing, 2016). https://doi.org/10.1007/978-3-319-22246-2_15
39. Dodge, T. O., Farquharson, K. A., Ford, C., Cavanagh, L., Schubert, K., Schumer, M., Belov, K., & Hogg, C. J. Genomes of two Extinct-in-the-Wild reptiles from Christmas Island reveal distinct evolutionary histories and conservation insights. *Molecular Ecology Resources* **00** (2023). https://doi.org/10.1111/1755-0998.13780
40. Robinson, J. A., Bowie, R. C. K., Dudchenko, O., Aiden, E. L., Hendrickson, S. L., Steiner, C. C., Ryder, O. A., Mindell, D. P., & Wall, J. D. Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology* **31**, 2939-2946.e5 (2021). https://doi.org/10.1016/j.cub.2021.04.035
41. Gargano, M., Colosimo, G., Gratton, P., Marta, S., Brilli, M., Giustini, F., Sevilla, C., & Gentile, G. Nitrogen and carbon stable isotope analysis sheds light on trophic competition between two syntopic land iguana species from Galápagos. *Scientific Reports* **12**, 16897 (2022). https://doi.org/10.1038/s41598-022-21134-2
42. Day, G., Fox, G., Hipperson, H., Maher, K., Tucker, R., Horsburgh, G., Waters, D., Durrant, K. L., Burke, T., Slate, J., & Arnold, K. Revealing the Demographic History of the European nightjar (*Caprimulgus europaeus*). *Ecology and Evolution* **14**, e70460 (2024). https://doi.org/10.1002/ece3.70460

948 43. Ludington, A. J., Hammond, J. M., Breen, J., Deveson, I. W., & Sanders, K. L. New
949 chromosome-scale genomes provide insights into marine adaptations of sea snakes
950 (Hydrophis: *Elapidae*). *BMC Biology* **21**, 284 (2023). https://doi.org/10.1186/s12915-023-
951 01772-2
952 44. Lawrence, E. S., Gu, W., Bohlender, R. J., Anza-Ramirez, C., Cole, A. M., Yu, J. J., Hu, H.,
953 Heinrich, E. C., O'Brien, K. A., Vasquez, C. A., Cowan, Q. T., Bruck, P. T., Mercader, K.,
954 Alotaibi, M., Long, T., Hall, J. E., Moya, E. A., Bauk, M. A., Reeves, J. J., … Simonson, T. S.
955 Functional EPAS1/HIF2A missense variant is associated with hematocrit in Andean
956 highlanders. *Science Advances* **10**, eadj5661 (2024).
957 https://doi.org/10.1126/sciadv.adj5661
958 45. Liu, X., Zhang, Y., Li, Y., Pan, J., Wang, D., Chen, W., Zheng, Z., He, X., Zhao, Q., Pu, Y., Guan,
959 W., Han, J., Orlando, L., Ma, Y., & Jiang, L. EPAS1 Gain-of-Function Mutation Contributes
960 to High-Altitude Adaptation in Tibetan Horses. *Molecular Biology and Evolution* **36**, 2591–
961 2603 (2019). https://doi.org/10.1093/molbev/msz158
962 46. vonHoldt, B., Fan, Z., Ortega-Del Vecchyo, D., & Wayne, R. K. EPAS1 variants in high
963 altitude Tibetan wolves were selectively introgressed into highland dogs. *PeerJ* **5**, e3522
964 (2017). https://doi.org/10.7717/peerj.3522
965 47. Syafiq, M, R. Molecular Insights into Diving Ability Adaptation of the Galápagos marine
966 iguana (*Amblyrhynchus cristatus*). PhD Thesis in Evolutionary Biology and Ecology,
967 University of Rome Tor Vergata, Italy (2024).
968 48. Chiou, K. L., Janiak, M. C., Schneider-Crease, I. A., Sen, S., Ayele, F., Chuma, I. S., Knauf, S.,
969 Lemma, A., Signore, A. V., D'Ippolito, A. M., Abebe, B., Haile, A. A., Kebede, F., Fashing, P.
970 J., Nguyen, N., McCann, C., Houck, M. L., Wall, J. D., Burrell, A. S., … Snyder-Mackler, N.
971 Genomic signatures of high-altitude adaptation and chromosomal polymorphism in
972 geladas. *Nature Ecology & Evolution* **6** (2022). https://doi.org/10.1038/s41559-022-
973 01703-4
974 49. Noh, H. J., Turner-Maier, J., Schulberg, S. A., Fitzgerald, M. L., Johnson, J., Allen, K. N.,
975 Hückstädt, L. A., Batten, A. J., Alfoldi, J., Costa, D. P., Karlsson, E. K., Zapol, W. M., Buys, E.
976 S., Lindblad-Toh, K., & Hindle, A. G. The Antarctic Weddell seal genome reveals evidence
977 of selection on cardiovascular phenotype and lipid handling. *Communications Biology* **5**,
978 1–12 (2022). https://doi.org/10.1038/s42003-022-03089-2

50. Genah, S., Monici, M., & Morbidelli, L. The Effect of Space Travel on Bone Metabolism: Considerations on Today's Major Challenges and Advances in Pharmacology. *International Journal of Molecular Sciences* **22** (2021). https://doi.org/10.3390/ijms22094585

51. Pereira, R. M. R., Delany, A. M., & Canalis, E. Cortisol inhibits the differentiation and apoptosis of osteoblasts in culture. *Bone* **28**, 484–490 (2001). https://doi.org/10.1016/S8756-3282(01)00422-7

52. Romero, L. M., & Wikelski, M. Corticosterone levels predict survival probabilities of Galápagos marine iguanas during El Niño events. *Proceedings of the National Academy of Sciences* **98**, 7366–7370 (2001). https://doi.org/10.1073/pnas.131091498

53. Vind, A. C., Wu, Z., Firdaus, M. J., Snieckute, G., Toh, G. A., Jessen, M., Martínez, J. F., Haahr, P., Andersen, T. L., Blasius, M., Koh, L. F., Maartensson, N. L., Common, J. E. A., Gyrd-Hansen, M., Zhong, F. L., Bekker-Jensen, S. The ribotoxic stress response drives acute inflammation, cell death, and epidermal thickening in UV-irradiated skin *in vivo*. *Molecular Cell* **84**, 4774-4789 (2024). https://doi.org/10.1016/j.molcel.2024.10.044

54. Juyoux, P, Galdadas, I., Gobbo, D., Von Velsen, J., Pelosse, M., Tully, M., Vadas, O., Gervasio, F. L., Pellegrini, E., & Bowler, M. W. Architecture of the MKK6-p38α complex defines the basis of MAPK specificity and activation. *Science* **381**, 1217-1225 (2023). https://doi.org/10.1126/science.add7859

55. Lewbart, G. A., Colosimo, G., Gaudette, C., Negrão Watanabe, T. T., Parker, J., Sevilla, C., Gerber, G. P., & Gentile, G. When pink is a question: Comparative gross and microscopic skin structure analyses reveal the histological basis of skin colour in Galápagos pink land iguanas (*Conolophus marthae*). *Acta Zoologica* **105** (2023). https://doi.org/10.1111/azo.12488

56. Gustavino, B., Maruccia, F., Williams, A., Di Giacomo, C., Franziano, M., Pucillo, L., Carrión Tacuri, J. E., & Gentile, G. When pink is a question: Comparative gross and microscopic skin structure analyses reveal the histological asis of skin colour in Galápagos pink land iguanas (*Conolophus marthae*). *Acta Zoologica* **105** (2023). https://doi.org/10.1111/azo.12488

57. Costantini, D., Dell'Omo, G., Casagrande, S., Fabiani, A., Carosi, M., Bertacche, V., Marquez, C., Snell, H., Snell, H., Tapia, W., & Gentile, G. Inter-population variation of carotenoids in Galápagos land iguanas (*Conolophus subcristatus*). *Comparative*

Biochemistry and Physiology Part B: Biochemistry and Molecular Biology **142**, 239–244 (2005). https://doi.org/10.1016/j.cbpb.2005.07.011

58. Liakath-Ali, K., Vancollie, V. E., Sequeira, I., Lelliott, C. J., & Watt, F. J. Myosin 10 is involved in murine pigmentation. *Experimental Dermatology* **28**, 391-394 (2018). https://doi.org/10.1111/exd.13528

59. Green, M. R., & Sambrook, J. Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harbor Protocols*, **2017**, pdb.prot093450 (2017). https://doi.org/10.1101/pdb.prot093450

60. Andrews, S. FastQC: A Quality Control tool for High Throughput Sequence Data (Computer software, (2010). https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

61. Bolger, A. M., Lohse, M., & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014). https://doi.org/10.1093/bioinformatics/btu170

62. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018). https://doi.org/10.1093/bioinformatics/bty149

63. Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. The MaSuRCA genome assembler. *Bioinformatics* **29**, 2669–2677 (2013). https://doi.org/10.1093/bioinformatics/btt476

64. Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., … Wang, J. SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1** (2012). https://doi.org/10.1186/2047-217X-1-18

65. Ye, C., Ma, Z. S., Cannon, C. H., Pop, M., & Yu, D. W. Exploiting sparseness in *de novo* genome assembly. *BMC Bioinformatics* **13** (2012). https://doi.org/10.1186/1471-2105-13-S6-S1

66. Ye, C., Hill, C. M., Wu, S., Ruan, J., & Ma, Z. (Sam). DBG2OLC: Efficient Assembly of Large Genomes Using Long Erroneous Reads of the Third Generation Sequencing Technologies. *Scientific Reports* **6**, 1–9 (2016). https://doi.org/10.1038/srep31900

67. Di Genova, A., Buena-Atienza, E., Ossowski, S., & Sagot, M.-F. Efficient hybrid de novo assembly of human genomes with WENGAN. *Nature Biotechnology* **39**, 422-430 (2020). https://doi.org/10.1038/s41587-020-00747-w

68. Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nature Biotechnology* **37**, 540–546 (2019). https://doi.org/10.1038/s41587-019-0072-8

69. Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., Luo, F., & Xie, Z. MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nature Methods* **14** (2017). https://doi.org/10.1038/nmeth.4432

70. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110 (2016). https://doi.org/10.1093/bioinformatics/btw152

71. Shafin, K., Pesout, T., Lorig-Roach, R., Haukness, M., Olsen, H. E., Bosworth, C., Armstrong, J., Tigyi, K., Maurer, N., Koren, S., Sedlazeck, F. J., Marschall, T., Mayes, S., Costa, V., Zook, J. M., Liu, K. J., Kilburn, D., Sorensen, M., Munson, K. M., … Paten, B. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nature Biotechnology* **38**, (2020). https://doi.org/10.1038/s41587-020-0503-6

72. Ruan, J., & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nature Methods* **17** (2020). https://doi.org/10.1038/s41592-019-0669-3

73. Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* **27**, 737–746 (2017). https://doi.org/10.1101/gr.214270.116

74. Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. *PLOS ONE* **9**, e112963 (2014). https://doi.org/10.1371/journal.pone.0112963

75. Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020). https://doi.org/10.1073/pnas.1921046117

76. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics* **3** (2021). https://doi.org/10.1093/nargab/lqaa108

77. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* **32**, 767–769 (2016). https://doi.org/10.1093/bioinformatics/btv661

78. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013). https://doi.org/10.1093/bioinformatics/bts635

79. Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., & Zdobnov, E. M. OrthoDB v10: Sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Research* **47**, D807–D811 (2019). https://doi.org/10.1093/nar/gky1053

80. Gabriel, L., Hoff, K. J., Brůna, T., Borodovsky, M., & Stanke, M. TSEBRA: Transcript selector for BRAKER. *BMC Bioinformatics* **22**, 566 (2021). https://doi.org/10.1186/s12859-021-04482-0

81. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014). https://doi.org/10.1093/bioinformatics/btu031

82. Buchfink, B., Reuter, K., & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods* **18** (2021). https://doi.org/10.1038/s41592-021-01101-x

83. Simão, 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). https://doi.org/10.1093/bioinformatics/btv351

84. Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., & Li, D.-Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* **21**, 241 (2020). https://doi.org/10.1186/s13059-020-02154-5

85. Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. MITOS: Improved *de novo* metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* **69**, 313–319 (2013). https://doi.org/10.1016/j.ympev.2012.08.023

86. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8** (2013). https://doi.org/10.1038/nprot.2013.084

87. Li, W., & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006). https://doi.org/10.1093/bioinformatics/btl158

88. Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., … Whited, J. L. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Reports* **18**, 762–776 (2017). https://doi.org/10.1016/j.celrep.2016.12.063

89. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990). https://doi.org/10.1016/S0022-2836(05)80360-2

90. Finn, R. D., Clements, J., & Eddy, S. R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research* **39**, W29–W37 (2011). https://doi.org/10.1093/nar/gkr367

91. Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H.-H., Rognes, T., & Ussery, D. W. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research* **35**, 3100–3108 (2007). https://doi.org/10.1093/nar/gkm160

92. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology* **37** (2019). https://doi.org/10.1038/s41587-019-0036-z

93. Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: Application to complete genomes. J*ournal of Molecular Biology* **305**, 567–580 (2001). https://doi.org/10.1006/jmbi.2000.4315

94. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. Pfam: The

protein families database in 2021. *Nucleic Acids Research* **49**, D412–D419 (2021). https://doi.org/10.1093/nar/gkaa913

95. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research* **47**, D309–D314 (2019). https://doi.org/10.1093/nar/gky1085

96. Kanehisa, M., & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**, 27–30 (2000). https://doi.org/10.1093/nar/28.1.27

97. Webb, A. E., Walsh, T. A., & O'Connell, M. J. VESPA: Very large-scale Evolutionary and Selective Pressure Analyses. *PeerJ Computer Science* **3**, e118 (2017). https://doi.org/10.7717/peerj-cs.118

98. Emms, D. M., & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**, 238 (2019). https://doi.org/10.1186/s13059-019-1832-y

99. Katoh, K., Kuma, K., Toh, H., & Miyata, T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**, 511–518 (2005). https://doi.org/10.1093/nar/gki198

100. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004). https://doi.org/10.1186/1471-2105-5-113

101. Löytynoja, A. Phylogeny-aware alignment with PRANK. In *Multiple Sequence Alignment Method*s, 155–170. (Humana Press, 2014). https://doi.org/10.1007/978-1-62703-646-7_10

102. Thompson, J. D., Plewniak, F., Ripp, R., Thierry, J.-C., & Poch, O. Towards a reliable objective function for multiple sequence alignments. *Journal of Molecular Biology* **314**, 937–951 (2001). https://doi.org/10.1006/jmbi.2001.5187

103. Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). https://doi.org/10.1093/bioinformatics/btp348

104. Steenwyk, J. L., Buida, T. J., III, Labella, A. L., Li, Y., Shen, X.-X., & Rokas, A. PhyKIT: A broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics* **37**, 2325–2331 (2021). https://doi.org/10.1093/bioinformatics/btab096

1168 105. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. IQ-TREE: A Fast and
1169 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
1170 *Molecular Biology and Evolution* **32**, 268–274 (2015).
1171 https://doi.org/10.1093/molbev/msu300
1172 106. Lartillot, N., Rodrigue, N., Stubbs, D., & Richer, J. PhyloBayes MPI: Phylogenetic
1173 Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Systematic*
1174 *Biology* **62**, 611–615 (2013). https://doi.org/10.1093/sysbio/syt022
1175 107. Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. ASTRAL-III: Polynomial time species
1176 tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
1177 https://doi.org/10.1186/s12859-018-2129-y
1178 108. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
1179 *Evolution* **24**, 1586–1591 (2007). https://doi.org/10.1093/molbev/msm088
1180 109. Puttick, M. N. MCMCtreeR: Functions to prepare MCMCtree analyses and visualize
1181 posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
1182 https://doi.org/10.1093/bioinformatics/btz554
1183 110. Renaud, G., Hanghøj, K., Korneliussen, T. S., Willerslev, E., & Orlando, L. Joint Estimates
1184 of Heterozygosity and Runs of Homozygosity for Modern and Ancient Samples. *Genetics*
1185 **212**, 587–614 (2019). https://doi.org/10.1534/genetics.119.302057
1186 111. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler
1187 transform. *Bioinformatics* **25**, 1754–1760 (2009).
1188 https://doi.org/10.1093/bioinformatics/btp324
1189 112. Thompson, E. A. Identity by Descent: Variation in Meiosis, Across Genomes, and in
1190 Populations. *Genetics* **194**, 301–326 (2013).
1191 https://doi.org/10.1534/genetics.112.148825
1192 113. Elferink, M. G., van As, P., Veenendaal, T., Crooijmans, R. P., & Groenen, M. A. Regional
1193 differences in recombination hotspots between two chicken populations. *BMC Genetics*
1194 **11** (2010). https://doi.org/10.1186/1471-2156-11-11
1195 114. Lisachov, A. P., Trifonov, V. A., Giovannotti, M., Ferguson-Smith, M. A., & Borodin, P.
1196 M. Immunocytological analysis of meiotic recombination in two anole lizards (Squamata,
1197 *Dactyloidae*). *Comparative Cytogenetics* **11** (2017).
1198 https://doi.org/10.3897/CompCytogen.v11i1.10916

115. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018). https://doi.org/10.1093/bioinformatics/bty191

116. Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021). https://doi.org/10.1093/gigascience/giab008

117. Perry, B. W., Card, D. C., McGlothlin, J. W., Pasquesi, G. I. M., Adams, R. H., Schield, D. R., Hales, N. R., Corbin, A. B., Demuth, J. P., Hoffmann, F. G., Vandewege, M. W., Schott, R. K., Bhattacharyya, N., Chang, B. S. W., Casewell, N. R., Whiteley, G., Reyes-Velasco, J., Mackessy, S. P., Gamble, T., … Castoe, T. A. Molecular Adaptations for Sensing and Securing Prey and Insight into Amniote Genome Diversity from the Garter Snake Genome. *Genome Biology and Evolution* **10**, 2110–2129 (2018). https://doi.org/10.1093/gbe/evy157

118. Anisimova, M., Bielawski, J. P., & Yang, Z. Accuracy and Power of the Likelihood Ratio Test in Detecting Adaptive Molecular Evolution. *Molecular Biology and Evolution* **18**, 1585–1592 (2001). https://doi.org/10.1093/oxfordjournals.molbev.a003945

119. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* **14** (2017). https://doi.org/10.1038/nmeth.4285

120. Creevey, C. J., & McInerney, J. O. Clann: Investigating phylogenetic information through supertree analyses. *Bioinformatics* **21**, 390–392 (2005). https://doi.org/10.1093/bioinformatics/bti020

121. Pond, S. L. K., Poon, A. F. Y., Velazquez, R., Weaver, S., Hepler, N. L., Murrell, B., Shank, S. D., Magalis, B. R., Bouvier, D., Nekrutenko, A., Wisotsky, S., Spielman, S. J., Frost, S. D. W., & Muse, S. V. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. *Molecular Biology and Evolution*, **37**, 295–299 (2019). https://doi.org/10.1093/molbev/msz197

122. Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Molecular Biology and Evolution* **32**, 1342–1353 (2015). https://doi.org/10.1093/molbev/msv022

123. Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021). https://doi.org/10.1093/bioinformatics/btaa1022

124. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., & Kasprzyk, A. BioMart – biological queries made easy. *BMC Genomics* **10**, 22 (2009). https://doi.org/10.1186/1471-2164-10-22

125. Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE* **6**, e21800 (2011). https://doi.org/10.1371/journal.pone.0021800

126. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2** (2021). https://doi.org/10.1016/j.xinn.2021.100141

127. Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., & Thomas, P. D. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Research* **49**, D394–D403 (2021). https://doi.org/10.1093/nar/gkaa1106

128. Courtier-Orgogozo, V., Arnoult, L., Prigent, S. R., Wiltgen, S., & Martin, A. Gephebase, a database of genotype–phenotype relationships for natural and domesticated variation in Eukaryotes. *Nucleic Acids Research* **48**, D696–D703 (2020). https://doi.org/10.1093/nar/gkz796