

# The tuatara genome reveals ancient features of amniote evolution

<https://doi.org/10.1038/s41586-020-2561-9>

Received: 5 December 2019

Accepted: 26 June 2020

Published online: 05 August 2020

Open access



Neil J. Gemmell<sup>1</sup>✉, Kim Rutherford<sup>1</sup>, Stefan Prost<sup>2,3</sup>, Marc Tollis<sup>4,5</sup>, David Winter<sup>6</sup>, J. Robert Macey<sup>7</sup>, David L. Adelson<sup>8</sup>, Alexander Suh<sup>9,10</sup>, Terry Bertozzi<sup>8,11</sup>, José H. Grau<sup>12,13</sup>, Chris Organ<sup>14</sup>, Paul P. Gardner<sup>15</sup>, Matthieu Muffato<sup>16</sup>, Mateus Patricio<sup>16</sup>, Konstantinos Billis<sup>16</sup>, Fergal J. Martin<sup>16</sup>, Paul Fllice<sup>16</sup>, Bent Petersen<sup>17</sup>, Lin Kang<sup>18</sup>, Paweł Michalak<sup>18,19,20</sup>, Thomas R. Buckley<sup>21,22</sup>, Melissa Wilson<sup>4</sup>, Yuanyuan Cheng<sup>23</sup>, Hilary Miller<sup>24</sup>, Ryan K. Schott<sup>25</sup>, Melissa D. Jordan<sup>26</sup>, Richard D. Newcomb<sup>26</sup>, José Ignacio Arroyo<sup>27</sup>, Nicole Valenzuela<sup>28</sup>, Tim A. Hore<sup>1</sup>, Jaime Renart<sup>29</sup>, Valentina Peona<sup>9,10</sup>, Claire R. Peart<sup>9,30</sup>, Vera M. Warmuth<sup>9,30</sup>, Lu Zeng<sup>8</sup>, R. Daniel Kortschak<sup>8</sup>, Joy M. Raison<sup>8</sup>, Valeria Velásquez Zapata<sup>28</sup>, Zhiqiang Wu<sup>28</sup>, Didac Santestmasses<sup>31</sup>, Marco Mariotti<sup>31</sup>, Roderic Guigó<sup>31</sup>, Shawn M. Rupp<sup>4</sup>, Victoria G. Twort<sup>21,22</sup>, Nicolas Dussex<sup>1</sup>, Helen Taylor<sup>1</sup>, Hideaki Abe<sup>1</sup>, Donna M. Bond<sup>1</sup>, James M. Paterson<sup>32</sup>, Daniel G. Mulcahy<sup>33</sup>, Vanessa L. Gonzalez<sup>33</sup>, Charles G. Barbieri<sup>7</sup>, Dustin P. DeMeo<sup>7</sup>, Stephan Pabinger<sup>34</sup>, Tracey Van Stijn<sup>35</sup>, Shannon Clarke<sup>35</sup>, Oliver Ryder<sup>36</sup>, Scott V. Edwards<sup>37</sup>, Steven L. Salzberg<sup>38</sup>, Lindsay Anderson<sup>39</sup>, Nicola Nelson<sup>39</sup>, Clive Stone<sup>40</sup> & Ngatiwai Trust Board\*

The tuatara (*Sphenodon punctatus*)—the only living member of the reptilian order Rhynchocephalia (Sphenodontia), once widespread across Gondwana<sup>1,2</sup>—is an iconic species that is endemic to New Zealand<sup>2,3</sup>. A key link to the now-extinct stem reptiles (from which dinosaurs, modern reptiles, birds and mammals evolved), the tuatara provides key insights into the ancestral amniotes<sup>2,4</sup>. Here we analyse the genome of the tuatara, which—at approximately 5 Gb—is among the largest of the vertebrate genomes yet assembled. Our analyses of this genome, along with comparisons with other vertebrate genomes, reinforce the uniqueness of the tuatara. Phylogenetic analyses indicate that the tuatara lineage diverged from that of snakes and lizards around 250 million years ago. This lineage also shows moderate rates of molecular evolution, with instances of punctuated evolution. Our genome sequence analysis identifies expansions of proteins, non-protein-coding RNA families and repeat elements, the latter of which show an amalgam of reptilian and mammalian features. The sequencing of the tuatara genome provides a valuable resource for deep comparative analyses of tetrapods, as well as for tuatara biology and conservation. Our study also provides important insights into both the technical challenges and the cultural obligations that are associated with genome sequencing.

<sup>1</sup>Department of Anatomy, University of Otago, Dunedin, New Zealand. <sup>2</sup>LOEWE-Center for Translational Biodiversity Genomics, Senckenberg Museum, Frankfurt, Germany. <sup>3</sup>South African National Biodiversity Institute, National Zoological Garden, Pretoria, South Africa. <sup>4</sup>School of Life Sciences, Arizona State University, Tempe, AZ, USA. <sup>5</sup>School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA. <sup>6</sup>School of Fundamental Sciences, Massey University, Palmerston North, New Zealand. <sup>7</sup>Peralta Genomics Institute, Oakland, CA, USA. <sup>8</sup>School of Biological Sciences, The University of Adelaide, Adelaide, South Australia, Australia. <sup>9</sup>Department of Ecology and Genetics – Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden. <sup>10</sup>Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden.

<sup>11</sup>Evolutionary Biology Unit, South Australian Museum, Adelaide, South Australia, Australia. <sup>12</sup>Amedes Genetics, Amedes Medizinische Dienstleistungen, Berlin, Germany. <sup>13</sup>Museum für Naturkunde Berlin, Leibniz-Institut für Evolutions- und Biodiversitätsforschung an der Humboldt-Universität zu Berlin, Berlin, Germany. <sup>14</sup>Department of Earth Sciences, Montana State University, Bozeman, MT, USA. <sup>15</sup>Department of Biochemistry, University of Otago, Dunedin, New Zealand. <sup>16</sup>European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. <sup>17</sup>Section for Evolutionary Genomics, The GLOBE Institute, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. <sup>18</sup>Edward Via College of Osteopathic Medicine, Blacksburg, VA, USA. <sup>19</sup>Center for One Health Research, Virginia–Maryland College of Veterinary Medicine, Blacksburg, VA, USA. <sup>20</sup>Institute of Evolution, University of Haifa, Haifa, Israel. <sup>21</sup>Manaaki Whenua - Landcare Research, Auckland, New Zealand. <sup>22</sup>School of Biological Sciences, The University of Auckland, Auckland, New Zealand. <sup>23</sup>School of Life and Environmental Sciences, The University of Sydney, Sydney, New South Wales, Australia. <sup>24</sup>Biomatters, Auckland, New Zealand. <sup>25</sup>Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. <sup>26</sup>The New Zealand Institute for Plant and Food Research, Auckland, New Zealand. <sup>27</sup>Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile. <sup>28</sup>Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA. <sup>29</sup>Instituto de Investigaciones Biomédicas Alberto Sols' CSIC-UAM, Madrid, Spain. <sup>30</sup>Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilian University of Munich, Planegg-Martinsried, Germany. <sup>31</sup>Centre for Genomic Regulation (CRG), The Barcelona Institute for Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain. <sup>32</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. <sup>33</sup>Global Genome Initiative, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. <sup>34</sup>Austrian Institute of Technology (AIT), Center for Health and Bioresources, Molecular Diagnostics, Vienna, Austria. <sup>35</sup>AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand. <sup>36</sup>San Diego Zoo Institute for Conservation Research, Escondido, CA, USA. <sup>37</sup>Department of Organismic and Evolutionary Biology and the Museum of Comparative Zoology, Harvard University, Cambridge, MA, USA. <sup>38</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. <sup>39</sup>School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand. <sup>40</sup>Ngatiwai Trust Board, Whangarei, New Zealand. \*A list of members and their affiliations appears at the end of the paper. ✉e-mail: neil.gemmell@otago.ac.nz

# Article

The tuatara is an iconic terrestrial vertebrate that is unique to New Zealand<sup>2</sup>. The tuatara is the only living member of the archaic reptilian order Rhynchocephalia (Sphenodontia), which last shared a common ancestor with other reptiles at about 250 million years ago (Fig. 1); this species represents an important link to the now-extinct stem reptiles from which dinosaurs, modern reptiles, birds and mammals evolved, and is thus important for our understanding of amniote evolution<sup>2</sup>.

It is also a species of importance in other contexts. First, the tuatara is a *taonga* (special treasure) for Māori, who hold that tuatara are the guardians of special places<sup>2</sup>. Second, the tuatara is internationally recognized as a critically important species that is vulnerable to extinction owing to habitat loss, predation, disease, global warming and other factors<sup>2</sup>. Third, the tuatara displays a variety of morphological and physiological innovations that have puzzled scientists since its first description<sup>2</sup>. These include a unique combination of features that are shared variously with lizards, turtles and birds, which left its taxonomic position in doubt for many decades<sup>2</sup>. This taxonomic conundrum has largely been addressed using molecular approaches<sup>4</sup>, but the timing of the split of the tuatara from the lineage that forms the modern squamates (lizards and snakes), the rate of evolution of tuatara and the number of species of tuatara remain contentious<sup>2</sup>. Finally, there are aspects of tuatara biology that are unique within, or atypical of, reptiles. These include a unique form of temperature-dependent sex determination (which sees females produced below, and males above, 22 °C), extremely low basal metabolic rates and considerable longevity<sup>2</sup>.

To provide insights into the biology of the tuatara, we have sequenced its genome in partnership with Ngātiwai, the Māori *iwi* (tribe) who hold *kaitiakitanga* (guardianship) over the tuatara populations located on islands in the far north of New Zealand. This partnership—which, to our knowledge, is unique among the genome projects undertaken to date—had a strong practical focus on developing resources and information that will improve our understanding of the tuatara and aid in future conservation efforts. It is hoped that this work will form an exemplar for future genome initiatives that aspire to meet access and benefit-sharing obligations to Indigenous communities.

We find that the tuatara genome—as well as the animal itself—is an amalgam of ancestral and derived characteristics. Tuatara has  $2n = 36$  chromosomes in both sexes, consisting of 14 pairs of macrochromosomes and 4 pairs of microchromosomes<sup>5</sup>. The genome size, which is estimated to be approximately 5 Gb, is among the largest of the vertebrate genomes sequenced to date; this is predominantly explained by an extraordinary diversity of repeat elements, many of which are unique to the tuatara.

## Sequencing, assembly, synteny and annotation

Our tuatara genome assembly is 4.3 Gb, consisting of 16,536 scaffolds with an N50 scaffold length of 3 Mb (Extended Data Table 1, Supplementary Information 1). Genome assessment using Benchmarking Universal Single-Copy Orthologs (BUSCO)<sup>6</sup> indicates 86.8% of the vertebrate gene set are present and complete. Subsequent annotation identified 17,448 genes, of which 16,185 are one-to-one orthologues (Supplementary Information 2). Local gene-order conservation is high; 75% or more of tuatara genes showed conservation with birds, turtles and crocodilians. We also find that components of the genome, of 15 Mb in size and larger, are syntenic with other vertebrates; protein-coding gene order and orientation are maintained between tuatara, turtle, chicken and human, and strong co-linearity is seen between tuatara contigs and chicken chromosomes (Extended Data Figs. 1, 2).

## Genomic architecture

At least 64% of the tuatara genome assembly is composed of repetitive sequences, made up of transposable elements (31%) and

low-copy-number segmental duplications (33%). Although the total transposable element content is similar to other reptiles<sup>7</sup>, the types of repeats we found appear to be more mammal-like than reptile-like. Furthermore, a number of the repeat families show evidence of recent activity and greater expansion and diversity than seen in other vertebrates (Fig. 2).

L2 elements account for most of the long interspersed elements in the tuatara genome (10% of the genome), and some may still be active (Supplementary Information 4). CR1 elements—the dominant long interspersed element in the genomes of other sauropsids<sup>8</sup>—are rare. CR1 elements comprise only about 4% of the tuatara genome (Fig. 2a, Supplementary Table 4.1), but some are potentially active (Supplementary Fig. 4.4). L1 elements, which are prevalent in placental mammals, account for only a tiny fraction of the tuatara genome (<1%) (Supplementary Table 4.1). However, we find that an L2 subfamily that is present in the tuatara, but is absent from other lepidosaurs, is also common in monotremes<sup>9</sup> (Supplementary Figs. 4.3–4.5). Collectively, these data suggest that stem-sauropsid ancestors had a repeat composition that was very different from that inferred in previous comparisons using mammals, birds and lizards<sup>7</sup>.

Many of the short interspersed elements (SINEs) in the tuatara are derived from ancient common sequence motifs (CORE-SINEs), which are present in all amniotes<sup>10</sup>; however, at least 16 SINE subfamilies were recently active in the tuatara genome (Fig. 2b, Supplementary Information 5). Most of these SINEs are mammalian-wide interspersed repeats (MIRs), and the diversity of MIR subfamilies in the tuatara is the highest thus far observed in an amniote<sup>11,12</sup>. In the human genome, hundreds of fossil MIR elements act as chromatin and regulatory domains<sup>13</sup>; the very recent activity of diverse MIR subfamilies in the tuatara suggests these subfamilies may have influenced regulatory rewiring on rather recent evolutionary timescales.

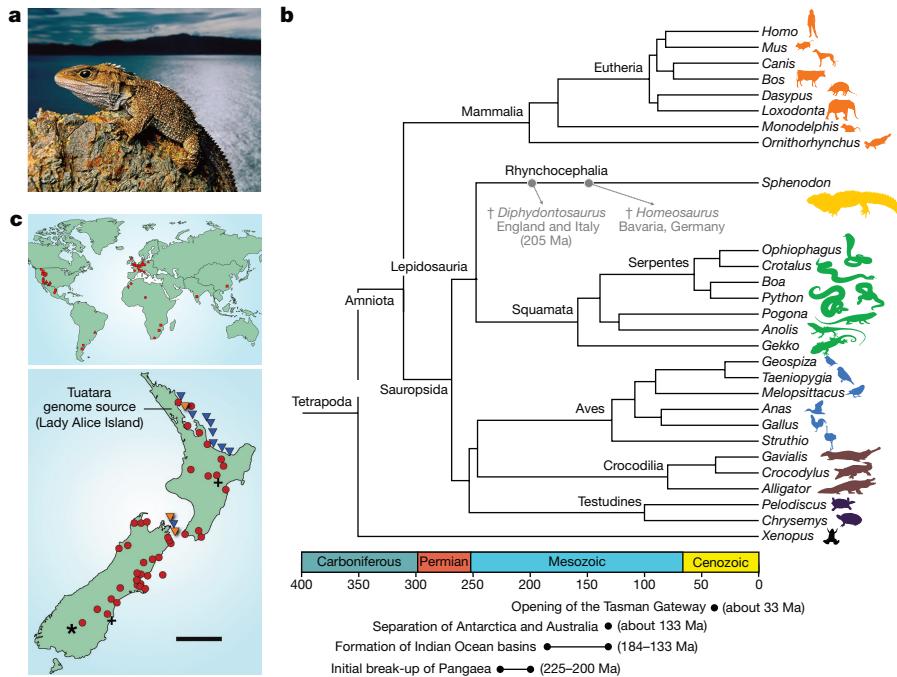
We detected 24 newly identified and unique families of DNA transposon, which suggests frequent germline infiltration by DNA transposons through horizontal transfer in the tuatara<sup>14</sup>. At least 30 subfamilies of DNA transposon were recently active, spanning a diverse range of cut-and-paste transposons and polintons (Supplementary Figs. 5.1, 5.2). This diversity is higher than that found in other amniotes<sup>15</sup>. Notably, we found thousands of identical DNA transposon copies, which suggests very recent—and/or ongoing—activity. Cut-and-paste transposition probably shapes the tuatara genome, as it does in bats<sup>15</sup>.

We identified about 7,500 full-length, long-terminal-repeat retro-elements (including endogenous retroviruses), which we classified into 12 groups (Fig. 2c, Supplementary Information 6). The general spectrum of long-terminal-repeat retroelements in the tuatara is comparable to that of other sauropsids<sup>7,15</sup>. We found at least 37 complete spumaretroviruses, which are among the most ancient of endogenous retroviruses<sup>16</sup>, in the tuatara genome (Fig. 2c, Supplementary Figs. 6.1, 6.2).

The tuatara genome contains more than 8,000 elements related to non-coding RNA. Most of these elements (about 6,900) derive from recently active transposable elements, and overlap with a newly identified CR1-mobilized SINE (Fig. 2b, Supplementary Information 7). The remaining high-copy-number elements are sequences closely related to ribosomal RNAs, spliceosomal RNAs and signal-recognition particle RNAs.

Finally, a high proportion (33%) of the tuatara genome originates from low-copy-number segmental duplications; 6.7% of these duplications are of recent origin (on the basis of their high level of sequence identity (>94% identity)), which is more than seen in other vertebrates<sup>9</sup>. The tuatara genome is 2.4× larger than the anole genome, and this difference appears to be driven disproportionately by segmental duplications.

Overall, the repeat architecture of the tuatara is—to our knowledge—unlike anything previously reported, showing a unique amalgam of features that have previously been viewed as characteristic of either reptilian or mammalian lineages. This combination of ancient amniote



**Fig. 1 | The phylogenetic significance and distribution of the tuatara.** **a**, The tuatara, (*S. punctatus*) is the sole survivor of the order Rhynchocephalia. **b, c**, The rhynchocephalians appear to have originated in the early Mesozoic period (about 250–240 million years ago (Ma)) and were common, speciose and globally distributed for much of that era. The geographical range of the rhynchocephalians progressively contracted after the Early Jurassic epoch (about 200–175 Ma); the most recent fossil record outside of New Zealand is from Argentina in the Late Cretaceous epoch (about 70 Ma). **c**, The last bastions of the rhynchocephalians are 32 islands off the coast of New Zealand, which have recently been augmented by the establishment of about 10 new island or

mainland sanctuary populations using translocations. The current global population is estimated to be around 100,000 individuals. Rhynchocephalian and tuatara fossil localities are redrawn and adapted from ref.<sup>1</sup> with permission, and incorporate data from ref.<sup>2</sup>. In the global distribution map (**c**, top); triangle = Triassic; square = Jurassic; circle = Cretaceous; and diamond = Palaeocene. In the map of the New Zealand distribution (**c**, bottom); asterisk = Miocene; cross = Pleistocene; circle = Holocene; blue triangle = extant population; and orange triangle = population investigated in this study. Scale bar, 200 km. Photograph credit, F. Lanting.

features—as well as a dynamic and diverse repertoire of lineage-specific transposable elements—strongly reflects the phylogenetic position of this evolutionary relic.

Our low-coverage bisulfite-sequencing analysis found approximately 81% of CpG sites are methylated in tuatara (Fig. 3a)—the highest reported percentage of methylation for an amniote. This pattern differs from that observed in mouse, human (about 70%) and chicken (about 50%), and is more similar to that of *Xenopus* (82%) and zebrafish (78%). One possible explanation for this high level of DNA methylation is the large number of repetitive elements found in tuatara, many of which appear recently active and might be regulated via DNA methylation.

The low normalized CpG content of the tuatara suggests its genome has endured substantial historic methylation<sup>17</sup>. The tuatara has a significantly bimodal distribution of normalized CpG (Extended Data Fig. 3) in all of the genomic regions we examined, a similarity it shares with other reptiles that have temperature-dependent sex determination<sup>17</sup>. The low normalized CpG count of the tuatara in non-promoter regions may result from methylation silencing of repeat elements, and the bimodality of normalized CpG promoters suggests dual transcriptional regulation (Extended Data Fig. 3, Supplementary Information 8).

The mitochondrial genome in the tuatara reference animal is 18,078 bp in size, containing 13 protein-coding, 2 ribosomal RNA and 22 transfer (t)RNA genes, a gene content typical among animals (Extended Data Fig. 4). This contradicts previous reports<sup>18</sup> that the tuatara mitochondrial genome lacks three genes: *NDS*, *tRNA<sup>Thr</sup>* and *tRNA<sup>His</sup>*. These genes are found—with an additional copy of *tRNA<sup>Leu(CUN)</sup>* and an additional non-coding block (which we refer to as NC2)—in a single segment of the mitochondrial genome. Three non-coding areas (NC1, NC2 and NC3) with control-region (heavy-strand replication

origin) features, and two copies of *tRNA<sup>Leu(CUN)</sup>* adjacent to NC1 and NC2, possess identical or near-identical sequences that are unique to the tuatara mitochondrial genome. These three non-coding regions may be a result of concerted evolution.

## Genomic innovations

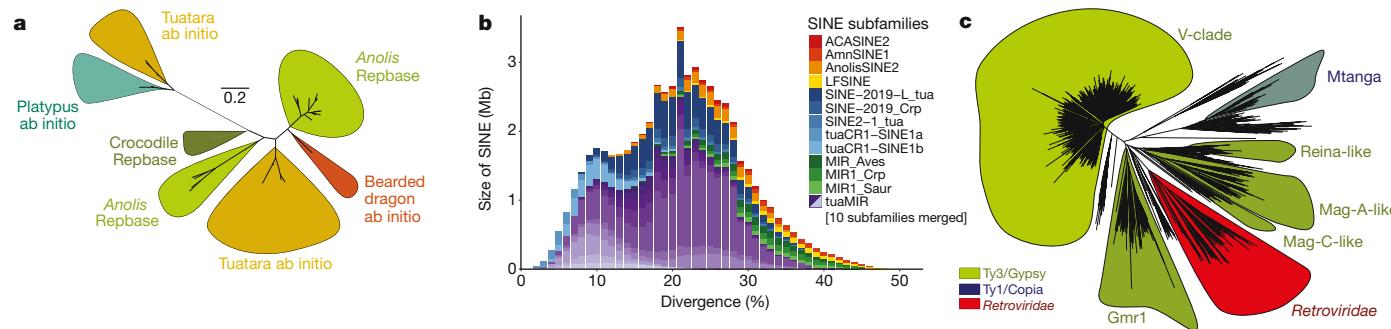
As befits the taxonomic distinctiveness of the tuatara, we find that its genome displays multiple innovations in genes that are associated with immunity, odour reception, thermal regulation and selenium metabolism.

Genes of the major histocompatibility complex (MHC) have an important role in disease resistance, mate choice and kin recognition, and are among the most polymorphic genes in the vertebrate genome. Our annotation of MHC regions in the tuatara, and comparisons of the gene organization with that of six other species, identified 56 MHC genes (Extended Data Fig. 5, Supplementary Information 9).

Of the six comparison species, the genomic organization of tuatara MHC genes is most similar to that of the green anole, which we interpret as typical for Lepidosauria. Tuatara and other reptiles show a gene content and complexity more similar to the MHC regions of amphibians and mammals than to the highly reduced MHC of birds. Although the majority of genes annotated in the tuatara MHC are well-conserved as one-to-one orthologues, we observed extensive genomic rearrangements among these distant lineages.

The tuatara is a highly visual predator that is able to capture prey under conditions of extremely low light<sup>2</sup>. Despite the nocturnal visual adaptation of the tuatara, it shows strong morphological evidence of an ancestrally diurnal visual system<sup>19</sup>. We identified all five of the

# Article



**Fig. 2 | Analysis of the repeat landscape in the tuatara genome identifies unique repeat families, evidence of recent activity and a greater expansion and diversity of repeats than any other amniote. a,** A phylogenetic analysis on the basis of the reverse transcriptase domain of L2 repeats identifies two L2 subfamilies; one typical of other lepidosaurs and one that is similar to platypus L2. This phylogeny is based on L2 elements >1.5-kb long with a reverse transcriptase domain of >200 amino acids. **b,** Landscape plot of SINE retrotransposons suggests the tuatara genome is dominated by MIR sequences that are most typically associated with mammals; the tuatara genome is now the amniote genome in which the greatest MIR diversity has been observed. Only SINE subfamilies that occupy more than 1,000 bp are shown. Definitions of the abbreviations of the SINE subfamilies follow: ACASINE2, *Anolis*

*carolinensis* SINE family; AmnSINE1, Amniota SINE1; AnolisSINE2, *A. carolinensis* SINE2 family, LFSINE, lobe-finned fishes SINE; SINE-2019-L\_tua, tuatara SINE; tuaCR1-SINE1a and b, tuatara CR1-mobilized SINES; MIR\_Aves, avian MIR sequence; MIR1\_Crp, *C. porosus* MIR sequence; MIR1\_Saur, Sauropsida MIR sequence; tuaMIR, tuatara MIRs. **c,** The tuatara genome contains about 7,500 full-length, long-terminal-repeat retro-elements, including nearly 450 endogenous retroviruses that span the five major retroviral clades. A Ty1/Copia element (Mtanga-like) is especially abundant, but Bel-Pao long-terminal-repeat retro-elements are absent. At least 37 complete spumaretroviruses are present in the tuatara genome.

vertebrate visual opsin genes in the tuatara genome (Supplementary Information 10).

Our comparative analysis revealed one of the lowest rates of visual-gene loss known for any amniote, which contrasts sharply with the high rates of gene loss observed in ancestrally nocturnal lineages (Extended Data Fig. 6). Visual genes involved in phototransduction showed strong negative selection and no evidence for the long-term shifts in selective pressures that have been observed in other groups with evolutionarily modified photoreceptors<sup>20</sup>. The retention of five visual opsins and the conserved nature of the visual system also suggests tuatara possess robust colour vision, potentially at low light levels. This broad visual repertoire may be explained by the dichotomy in tuatara life history: juvenile tuatara often take up a diurnal and arboreal lifestyle to avoid the terrestrial, nocturnal adults that may predate them<sup>2</sup>. Collectively, these results suggest a unique path to nocturnal adaptation in tuatara from a diurnal ancestor.

Odorant receptors are expressed in the dendritic membranes of olfactory receptor neurons and enable the detection of odours. Species that depend strongly on their sense of smell to interact with their environment, find prey, identify kin and avoid predators may be expected to have a large number of odorant receptors. The tuatara genome contains 472 predicted odorant receptors, of which 341 sequences appear intact (Supplementary Information 11). The remainder lack the initial start codon, have frameshifts or are presumed to be pseudogenes. Many odorant receptors were found as tandem arrays, with up to 26 genes found on a single scaffold.

The number and diversity of odorant receptor genes varies greatly in Sauropsida: birds have 182–688 such genes, the green anole lizard has 156 genes, and crocodilians and testudines have 1,000–2,000 genes<sup>21</sup>. The tuatara has a number of odorant receptors similar to that of birds, but contains a high percentage of intact odorant receptor genes (85%) relative to published odorant receptor sets from the genomes of other sauropsids. This may reflect a strong reliance on olfaction by tuatara, and therefore pressure to maintain a substantial repertoire of odorant receptors (Extended Data Fig. 7). There is some evidence that olfaction has a role in identifying prey<sup>2</sup>, as well as suggestions that cloacal secretions may act as chemical signals.

The tuatara is a behavioural thermoregulator, and is notable for having the lowest optimal body temperature of any reptile (16–21 °C). Genes that encode transient receptor potential ion channels (TRP genes) have

an important role in thermoregulation, as these channels participate in thermosensation and cardiovascular physiology<sup>22</sup>; this led us to hypothesize that TRP genes may be linked to the thermal tolerance of the tuatara. Our comparative genomic analysis of TRP genes in the tuatara genome identified 37 TRP-like sequences, spanning all 7 known subfamilies of TRP genes (Extended Data Fig. 8, Supplementary Information 12)—an unusually large repertoire of TRP genes.

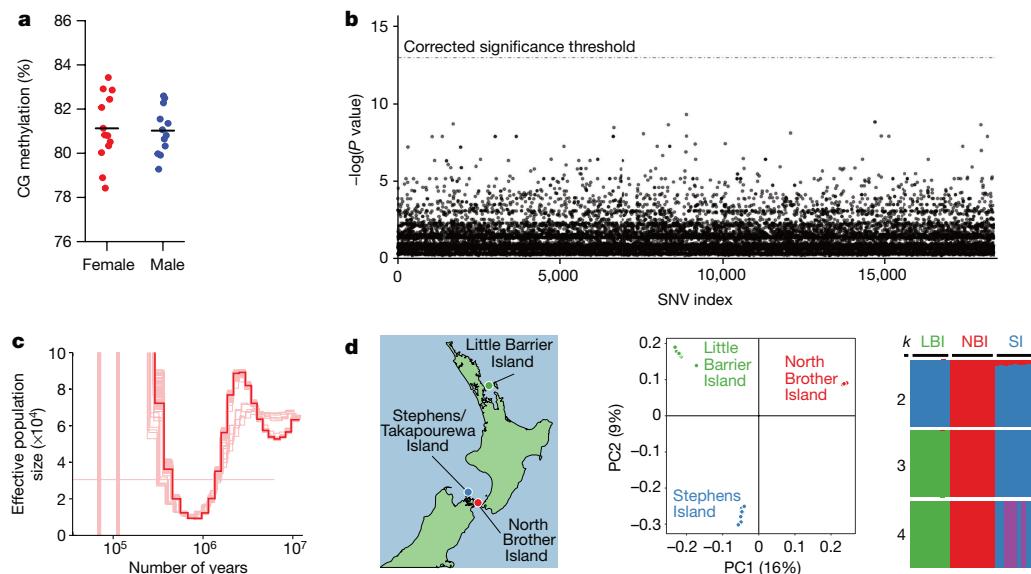
Among this suite of genes, we identified thermosensitive and non-thermosensitive TRP genes that appear to result from gene duplication, and have been differentially retained in the tuatara. For example, the tuatara is unusual in possessing an additional copy of a thermosensitive TRPV-like gene (*TRPV1/2/4*, sister to the genes *TRPV1*, *TRPV2* and *TRPV4*) that has classically been linked to the detection of moderate-to-extreme heat<sup>22</sup>—a feature it shares with turtles. A strong signature of positive selection among heat-sensitive TRP genes (*TRPA1*, *TRPM* and *TRPV*) was also observed.

In general, these results show a high rate of differential retention and positive selection in genes for which a function in heat sensation is well-established<sup>22</sup>. It therefore seems probable that the genomic changes in TRP genes are associated with the evolution of thermoregulation in tuatara.

Barring tortoises, tuatara are the longest lived of the reptiles—probably exceeding 100 years of age<sup>2</sup>. This enhanced lifespan may be linked to genes that afford protection against reactive oxygen species. One class of gene products that affords such protection is the selenoproteins. The human genome encodes 25 selenoproteins, the roles of which include antioxidation, redox regulation, thyroid hormone synthesis and calcium signal transduction, among others<sup>23</sup>.

We identified 26 genes that encode selenoproteins in the tuatara genome, as well as 4 selenocysteine-specific tRNA genes; all of these appear to be functional (Supplementary Information 13). Although further work is needed, the additional selenoprotein gene (relative to the human genome) and the selenocysteine-specific tRNA genes may be linked to the longevity of tuatara or might have arisen as a response to the low levels of selenium and other trace elements in the terrestrial systems of New Zealand.

Tuatara has a unique mode of temperature-dependent sex determination, in which higher temperatures during egg incubation result in males<sup>2</sup>. We found orthologues for many genes that are known to act antagonistically in masculinizing (for example, *SFI* and *SOX9*) and feminizing (for



**Fig. 3 | Analysis of sex differences, demographic history and population structure.** **a**, Methylation levels in the tuatara genome are high (mean 81%), but show no significant differences among the sexes (female  $n=13$ , mean = 81.13, s.d. = 1.55; male  $n=12$ ; mean = 81.02, s.d. = 1.07). The black horizontal line represents the mean in each dataset. **b**, No single-nucleotide variant (SNV) is significantly differentiated with respect to sex in the tuatara genome. Each point represents a  $P$  value from a test of sexual differentiation for a single SNV. The dashed line represents the threshold for statistical significance after accounting for multiple testing ( $n=28$ ; 13 males and 15 females).  $P$  values calculated using Fisher's exact test, two-tailed test and corrected for multiple testing using the Bonferroni method. **c**, Pairwise sequential Markovian

coalescent plot of the demographic history of tuatara using a mutation rate of  $1.4 \times 10^{-8}$  substitutions per site per generation and a generation time of 30 years. **d**, We examined the three known axes of genetic diversity in tuatara: northern New Zealand (Little Barrier Island (LBI) ( $n=9$ )) and two islands in the Cook Strait (Stephens Island (SI) ( $n=9$ ) and North Brother Island (NBI) ( $n=10$ )), using genotype-by-sequencing methods. Principal component (PC) analysis and structure plots demonstrate substantial structure among tuatara populations, and strongly support previous suggestions that the tuatara on the North Brother Island are genetically distinct and warrant separate management.

example, *RSPO1* and *WNT4*) gene networks to promote testicular or ovarian development, respectively<sup>24</sup>. We also found orthologues of several genes that have recently been implicated in temperature-dependent sex determination, including *CIRBP*<sup>24</sup> (Supplementary Information 17, Supplementary Table 17.2). Tuatara possess no obviously differentiable sex chromosomes<sup>5</sup>, and we found no significant sex-specific differences in global CG methylation (Fig. 3a) and no sex-specific single-nucleotide variants between male and female tuatara (Fig. 3b). On a gene-by-gene basis, sex-specific differences in methylation and gene expression patterns probably exist, but this remains to be investigated.

## Phylogeny and evolutionary rates

Our phylogenomic analyses, which incorporated both whole-genome alignments and clusters of single-copy orthologues (Supplementary Information 14, 15) recapitulated many patterns that have been observed in the fossil record and corroborated during the genomic era (Fig. 1). After their appearance about 312 million years ago<sup>25</sup>, amniote vertebrates diversified into two groups: the synapsids (which include all mammals) and the sauropsids (which include all reptiles and birds). We obtained full phylogenomic support for a monophyletic Lepidosauria, marked by the divergence of the tuatara lineage from all squamates (lizards and snakes) during the early part of the Triassic period at about 250 million years ago, as estimated using a penalized likelihood method (Fig. 1, Supplementary Information 14–16).

The rate of molecular evolution in the tuatara has previously been suggested to be paradoxically high, in contrast to the apparently slow rate of morphological evolution<sup>26,27</sup>. However, we find that the actual divergence in terms of DNA substitutions per site per million years at fourfold degenerate sites is relatively low, particularly with respect to lizards and snakes; this makes the tuatara the slowest-evolving lepidosaur yet analysed (Extended Data Fig. 9a, b). We also find that in general

amniote evolution can be described by a model of punctuated evolution, in which the amount of genomic change is related to the degree of species diversification within clades<sup>28,29</sup>. The tuatara falls well below this trend, accumulating substitutions at a rate expected given the lack of rhynchocephalian diversity (Extended Data Fig. 9c, Supplementary Information 16). This suggests that rates of phenotypic and molecular evolution were not decoupled throughout the evolution of amniotes<sup>30</sup>.

## Patterns of selection

In two sets of analyses, we find that most genes exhibit a pattern of molecular evolution that suggests that the tuatara branch evolves at a different rate than the rest of the tree (Supplementary Information 17, Supplementary Table 4). Approximately 659 of the 4,284 orthologues we tested had significantly different  $\omega$  values (ratios of non-synonymous to synonymous substitutions, dN/dS) on the tuatara branch relative to the birds and other reptiles we tested (Supplementary Information 17). Although none of these orthologues had  $\omega$  values suggestive of strong positive selection (that is,  $>1$ ), the results do indicate that shifts in patterns of selection are affecting many genes and functional categories of genes across the tuatara genome, including genes involved in RNA regulation, metabolic pathways, general metabolism and sex determination.

## Population genetics

Once widespread across the supercontinent of Gondwana, Rhynchocephalia is now represented by a single species (the tuatara) found on a few islands offshore of New Zealand (Fig. 1c). Historically, tuatara declined in range and numbers because of introduced pests and habitat loss<sup>2</sup>. They remain imperilled owing to their highly restricted distribution, threats imposed by disease and changes in sex ratios induced by climate change that could markedly affect their survival<sup>31</sup>. Previous work has found that

# Article

populations in northern New Zealand are genetically distinct from those in the Cook Strait, and that the population on North Brother Island in the Cook Strait might be a distinct species<sup>3</sup>. Although subsequent studies have not supported species status for the population on North Brother Island<sup>32</sup>, it is managed as a separate conservation unit.

We used the tuatara reference genome to perform ancestral demographic and population genomic analyses of this species. First, we investigated genome-wide signals for demographic change using a pairwise sequentially Markovian coalescent method (Supplementary Information 18). Our reconstructed demography (Fig. 3c) reveals an increase in effective population size ( $N_e$ ) that is detectable around 10 million years ago, a marked decrease in  $N_e$  about 1–3 million years ago and a rapid increase in  $N_e$  between 500 thousand million years ago and 1 million years ago. These events correlate well with the known geological history of New Zealand<sup>33</sup>, and may reflect an increase in available landmass subsequent to Oligocene drowning, a period of considerable climatic cooling that probably reduced tuatara habitat and the formation of land bridges that facilitated population expansion.

Our population genomic analyses examined the major axes of genetic diversity in tuatara<sup>32,34</sup>, and revealed substantial genetic structure (Fig. 3d, Supplementary Information 19). Our genome-wide estimate of the fixation index ( $F_{ST}$ ) is 0.45, and more than two-thirds of variable sites have an allele that is restricted to a single island. All populations have relatively low genetic diversity (nucleotide diversity ranges from  $8 \times 10^{-4}$  for North Brother Island to  $1.1 \times 10^{-3}$  for Little Barrier Island). The low within-population diversity and marked population structure we observe in the tuatara suggests that the modern island populations were isolated from each other sometime during the Last Glacial Maximum at about 18 thousand years ago.

Our results also support the distinctiveness of the North Brother Island tuatara, which has variously been described as *S. punctatus* or *Sphenodon guntheri*<sup>3,32</sup>. This population is highly inbred and shows evidence of a severe bottleneck, which most probably reflects a founder event around the time of the last glaciation<sup>34</sup>. It is not clear whether the distinctiveness we observe is due to changes in allele frequency brought about by this bottleneck, or is reflective of a deeper split in the population history of tuatara. Regardless, this population is an important source of genetic diversity in tuatara, possessing 8,480 private alleles. Although we support synonymization of *S. punctatus* and *S. guntheri*<sup>32</sup>, the ongoing conservation of the North Brother Island population as an independent unit is recommended.

## A cultural dimension

The tuatara is a *taonga* for many Māori—notably Ngātiwai and Ngāti Koata who are the *kaitiaki* (guardians) of tuatara. We worked in partnership with Ngātiwai *iwi* to increase knowledge and understanding of tuatara, and aid in the conservation of this species in the long term. Ngātiwai were involved in all decision-making regarding the use of the genome data by potential collaborators; for each new project we proposed, we discussed the benefits that might accrue from this work and how these could be shared. The need to engage with—and protect the rights of—Indigenous communities in such a transparent way has seldom been considered in the genome projects published to date, but is a mandated consideration under the Nagoya Protocol (<https://www.cbd.int/abs/>). Our partnership is a step towards an inclusive model of genomic science, which we hope others will adopt and improve upon. Although each partnership is unique, we provide a template agreement (Supplementary Information 20) that we hope will be useful to others.

## Discussion

The tuatara has a genomic architecture unlike anything previously reported, with an amalgam of features that have previously been viewed as characteristic of either mammals or reptiles. Notable among these

features are unusually high levels of repetitive sequences that have traditionally been considered mammalian, many of which appear to have been recently active, and—to our knowledge—the highest level of genome methylation thus far reported. We also found a mitochondrial-genome gene content at odds with previously published reports that omitted the *NDS* gene<sup>18</sup>; this gene is present, nested within a repeat-rich region of the mitochondrial DNA.

Our phylogenetic studies provide insights into the timing and speed of amniote evolution, including evidence of punctuated genome evolution across this phylogeny. We also find that, in contrast to previous suggestions that the evolutionary rate for tuatara is exceptionally fast<sup>26</sup>, it is the slowest-evolving lepidosaur yet analysed.

Our investigations of genomic innovations identified genetic candidates that may explain the ultra-low active body temperature, longevity and apparent resistance to infectious disease in tuatara. Further functional exploration will refine our understanding of these unusual facets of tuatara biology, and the tuatara genome itself will enable many future studies to explore the evolution of complex systems across the vertebrates in a more complete way than has previously been possible.

Our population genomic work reveals considerable genetic differences among tuatara populations, and supports the distinctiveness of the North Brother Island tuatara.

Finally, this genome will greatly aid in future work on population differentiation, inbreeding and local adaptation in this global icon, the last remaining species of the once globally dominant reptilian order Rhynchocephalia.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-020-2561-9>.

1. Jones, M. E., Tennyson, A. J., Worthy, J. P., Evans, S. E. & Worthy, T. H. A sphenodontine (Rhynchocephalia) from the Miocene of New Zealand and palaeobiogeography of the tuatara (*Sphenodon*). *Proc. R. Soc. Lond. B* **276**, 1385–1390 (2009).
2. Cree, A. *Tuatara: Biology and Conservation of a Venerable Survivor* (Canterbury Univ. Press, 2014).
3. Daugherty, C. H., Cree, A., Hay, J. M. & Thompson, M. B. Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature* **347**, 177–179 (1990).
4. Jones, M. E. H. et al. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **13**, 208 (2013).
5. O’Meally, D., Miller, H., Patel, H. R., Graves, J. A. M. & Ezaz, T. The first cytogenetic map of the tuatara, *Sphenodon punctatus*. *Cytogenet. Genome Res.* **127**, 213–223 (2009).
6. 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).
7. Pasquesi, G. I. M. et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* **9**, 2774 (2018).
8. Suh, A. et al. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol. Evol.* **7**, 205–217 (2015).
9. Zeng, L., Kortschak, R. D., Raison, J. M., Bertozi, T. & Adelson, D. L. Superior ab initio identification, annotation and characterisation of TEs and segmental duplications from genome assemblies. *PLoS ONE* **13**, e0193588 (2018).
10. Gilbert, N. & Labuda, D. CORE-SINEs: eukaryotic short interspersed retrotransposons with common sequence motifs. *Proc. Natl Acad. Sci. USA* **96**, 2869–2874 (1999).
11. Mikkelsen, T. S. et al. Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447**, 167–177 (2007).
12. Warren, W. C. et al. Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008).
13. Wang, J. et al. MIR retrotransposon sequences provide insulators to the human genome. *Proc. Natl Acad. Sci. USA* **112**, E4428–E4437 (2015).
14. Schaack, S., Gilbert, C. & Feschotte, C. Promiscuous DNA: horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol. Evol.* **25**, 537–546 (2010).
15. Sotero-Caio, C. G., Platt, R. N. II, Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
16. Aiewsakun, P. & Katzourakis, A. Marine origin of retroviruses in the early Palaeozoic Era. *Nat. Commun.* **8**, 13954 (2017).
17. Radhakrishnan, S., Literman, R., Mizoguchi, B. & Valenzuela, N. MeDIP-seq and nCpG analyses illuminate sexually dimorphic methylation of gonadal development genes with high historic methylation in turtle hatchlings with temperature-dependent sex determination. *Epigenetics Chromatin* **10**, 28 (2017).

18. Rest, J. S. et al. Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol. Phylogenet. Evol.* **29**, 289–297 (2003).
19. Meyer-Rochow, V. B., Wohlfahrt, S. & Ahnelt, P. K. Photoreceptor cell types in the retina of the tuatara (*Sphenodon punctatus*) have cone characteristics. *Micron* **36**, 423–428 (2005).
20. Schott, R. K., Bhattacharyya, N. & Chang, B. S. W. Evolutionary signatures of photoreceptor transmutation in geckos reveal potential adaptation and convergence with snakes. *Evolution* **73**, 1958–1971 (2019).
21. Vandewege, M. W. et al. Contrasting patterns of evolutionary diversification in the olfactory repertoires of reptile and bird genomes. *Genome Biol. Evol.* **8**, 470–480 (2016).
22. Nilius, B. & Owsianik, G. The transient receptor potential family of ion channels. *Genome Biol.* **12**, 218 (2011).
23. Labunsky, V. M., Hatfield, D. L. & Gladyshev, V. N. Selenoproteins: molecular pathways and physiological roles. *Physiol. Rev.* **94**, 739–777 (2014).
24. Capel, B. Vertebrate sex determination: evolutionary plasticity of a fundamental switch. *Nat. Rev. Genet.* **18**, 675–689 (2017).
25. Blair Hedges, S. & Kumar, S. *The Timetree of Life* (Oxford Univ. Press, 2009).
26. Hay, J. M., Subramanian, S., Millar, C. D., Mohandesan, E. & Lambert, D. M. Rapid molecular evolution in a living fossil. *Trends Genet.* **24**, 106–109 (2008).
27. Miller, H. C., Moore, J. A., Allendorf, F. W. & Daugherty, C. H. The evolutionary rate of tuatara revisited. *Trends Genet.* **25**, 13–15, author reply 16–18 (2009).
28. Landis, M. J. & Schraiber, J. G. Pulsed evolution shaped modern vertebrate body sizes. *Proc. Natl Acad. Sci. USA* **114**, 13224–13229 (2017).
29. Webster, A. J., Payne, R. J. H. & Pagel, M. Molecular phylogenies link rates of evolution and speciation. *Science* **301**, 478 (2003).
30. Subramanian, S., Hay, J. M., Mohandesan, E., Millar, C. D. & Lambert, D. M. Molecular and morphological evolution in tuatara are decoupled. *Trends Genet.* **25**, 16–18 (2009).
31. Mitchell, N. J., Kearney, M. R., Nelson, N. J. & Porter, W. P. Predicting the fate of a living fossil: how will global warming affect sex determination and hatching phenology in tuatara? *Proc. R. Soc. Lond. B* **275**, 2185–2193 (2008).
32. Hay, J. M., Sarre, S. D., Allendorf, F. W. & Daugherty, C. H. Genetic diversity and taxonomy: a reassessment of species designation in tuatara (*Sphenodon*: Reptilia). *Conserv. Genet.* **11**, 1063–1081 (2010).
33. Cooper, A. & Cooper, R. A. The Oligocene bottleneck and New Zealand biota: genetic record of a past environmental crisis. *Proc. R. Soc. Lond. B* **261**, 293–302 (1995).
34. MacAvoy, E. S. et al. Genetic variation in island populations of tuatara (*Sphenodon* spp) inferred from microsatellite markers. *Conserv. Genet.* **8**, 305–318 (2007).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

**Ngatiwai Trust Board**

**Clive Stone<sup>40</sup>, Jim Smillie<sup>40</sup> & Haydn Edmonds<sup>40</sup>**

# Article

## Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

A full description of the methods can be found in the Supplementary Information.

### Sampling and sequencing

A blood sample was obtained from a large male tuatara from Lady Alice Island ( $35^{\circ} 53' 24.4''$  S  $174^{\circ} 43' 38.2''$  E) (New Zealand), with appropriate ethical permissions and *iwi* consultation and support (Supplementary Information 20). Total genomic DNA and RNA were extracted and sequenced using the Illumina HiSeq 2000 and MiSeq sequencing platforms (Illumina) supported by New Zealand Genomics (Supplementary Information 1).

### Genome, transcriptome and epigenome

Raw reads were de novo-assembled using Allpaths-LG (version 49856). With a total input data of 5,741,034,516 reads for the paired-end libraries and 2,320,886,248 reads of the mate-pair libraries, our optimal assembly used 85% of the fragment libraries and 100% of the jumping libraries (Supplementary Information 1.4). We further scaffolded the assembly using Chicago libraries and HiRise (Supplementary Information 1.3).

We assembled a de novo transcriptome as a reference for read-mapping using total RNA derived from the blood of our reference male tuatara, and a collection of transcriptomic data previously collected from early-stage embryos<sup>35</sup>. In total, we had 131,580,633 new 100-bp read pairs and 60,637,100 previous 50-bp read pairs. These were assembled using Trinity v.2.2.0 (Supplementary Information 1.4).

Low-coverage bisulfite sequencing was undertaken using a modified post-bisulfite adaptor tagging method to explore global patterns of methylation in the genome for 12 male and 13 female tuatara (Fig. 3d, Supplementary Information 1.5).

### Repeat and gene annotation

We used a combination of ab initio repeat identification in CARP/RepeatModeler/LTRharvest, manual curation of specific newly identified repeats, and homology to repeat databases to investigate the repeat content of the tuatara genome (Supplementary Information 1.6). From these three complementary repeat identification approaches, the CARP results were in-depth-annotated for long interspersed elements and segmental duplications (Supplementary Information 4), the RepeatModeler results were in-depth-annotated for SINEs and DNA transposons (Supplementary Information 5), and the LTRharvest results were in-depth-annotated for long-terminal-repeat retrotransposons (Supplementary Information 6).

For the gene annotation, we used RepeatMasker (v.4.0.3) along with our partially curated RepeatModeler library plus the Repbase sauropsid repeat database to mask transposable elements in the genome sequence before the gene annotation. We did not mask simple repeats at this point to allow for more efficient mapping during the homology-based step in the annotation process. Simple repeats were later soft-masked and protein-coding genes predicted using MAKER2. We used anole lizard (*A. carolinensis*, version AnoCar2.0), python (*Python bivittatus*, version bivittatus-5.0.2) and RefSeq ([www.ncbi.nlm.nih.gov/refseq](http://www.ncbi.nlm.nih.gov/refseq)) as protein homology evidence, which we integrated with ab initio gene prediction methods including BLASTX, SNAP and Augustus. Non-coding RNAs were annotated using Rfam covariance models (v.13.0) (Supplementary Information 7).

### Orthologue calling

We performed a phylogenetic analysis to infer orthology relationships between the tuatara and 25 other species using the Ensembl GeneTree method (Supplementary Information Tables 2.1, 2.2).

Multiple-sequence alignments, phylogenetic trees and homology relationships were extracted in various formats (<https://zenodo.org/record/2542571>). We also calculated the gene order conservation score, which uses local synteny information around a pair of orthologous genes to compute how much the gene order is conserved. For each of these species, we chose the parologue with the best gene order conservation score and sequence similarity, which resulted in a total set of 3,168 clusters of orthologues (Supplementary Information 2, Table 2.3).

### Gene tree reconstructions and substitution rate estimation

We constructed phylogenies using only fourfold-degenerate-site data derived from whole-genome alignments for 27 tetrapods, analysed as a single partition in RAxML v.8.2.3. Using the topology and branch lengths obtained from the best maximum likelihood phylogeny, we estimated absolute rates of molecular evolution in terms of substitution per site per million years and estimated the divergence times of amniotes via the semiparametric penalized likelihood method in r8s v.1.8 (Supplementary Information 14.5).

We also generated gene trees on the basis of 245 single-copy orthologues found between all species using a maximum-likelihood-based multi-gene approach (Supplementary Information 15). Sequences were aligned using the codon-based aligner PRANK. The FASTA format alignments were then converted to PHYLIP using the catfasta2phylm.pl script (<https://github.com/nylander/catfasta2phylm>). Next, we used the individual exon PHYLIP files for gene tree reconstruction using RaxML using a GTR + G model. Subsequently, we binned all gene trees to reconstruct a species tree and carried out bootstrapping using Astral (Supplementary Information 15, Supplementary Fig. 15.1).

### Divergence times and tests of punctuated evolution

We inferred time-calibrated phylogenies with BEAST v.2.4.8 using the CIPRES Science Gateway to explore divergence times (Supplementary Information 16.1). We then used Bayesian phylogenetic generalized least squares to regress the total phylogenetic path length (of fourfold-degenerate sites) on the net number of speciation events (nodes in a phylogenetic tree) as a test for punctuated evolution (Supplementary Information 16.2).

### Analysis of genomic innovations

We explored the genomic organization and sequence evolution of genes associated with immunity, vision, smell, thermoregulation, longevity and sex determination (Supplementary Information 8–13). Tests of selection were undertaken across multiple genes, including those linked to metabolism, vision and sex determination using multispecies alignments and PAML (Supplementary Information 17).

### Population genomics

Demographic history was inferred from the diploid sequence of our tuatara genome using a pairwise sequential Markovian coalescent method (Supplementary Information 18). We also sampled 10 tuatara from each of three populations that span the main axes of genetic diversity in tuatara (Supplementary Information 19, Table 19.1), and used a modified genotype-by-sequencing approach to obtain the SNVs that we used for population genomic analysis, investigations of loci associated with sexual phenotype and estimates of genetic load (Supplementary Information 19).

### Permits and ethics

This project was undertaken in partnership with Ngatiwai and in consultation with other *iwi* who are *kaitiaki* of tuatara (Supplementary Information 20). Samples were collected under Victoria University of Wellington Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347 and held and used under permits 45462-DOA and 32037-RES 32037-RES issued by the New Zealand Department of Conservation.

## Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

The Tuatara Genome Consortium Project whole-genome shotgun and genome assembly are registered under the umbrella BioProjects PRJNA418887 and PRJNA445603, which are associated with BioSamples SAMN08038466 and SAMN08793959. Transcriptome read data are submitted under SRR7084910 (whole blood), together with previous data (SRR485948). The transcriptome assembly is submitted to GenBank with ID GGNQ00000000.1. Illumina short-read, Oxford Nanopore and PacBio long-read sequences are in the Sequence Read Archive accessions associated with PRJNA445603. The genome assembly (GCA\_003113815.1) described in this paper is version QEPC00000000.1 and consists of sequences QEPC01000001–QEPC01016536. Maker gene predictions are available from Zenodo at <https://doi.org/10.5281/zenodo.1489353>. The repeat library database developed for tuatara is available from Zenodo at <https://doi.org/10.5281/zenodo.2585367>.

35. Miller, H. C., Biggs, P. J., Voelckel, C. & Nelson, N. J. De novo sequence assembly and characterisation of a partial transcriptome for an evolutionarily distinct reptile, the tuatara (*Sphenodon punctatus*). *BMC Genomics* **13**, 439 (2012).
36. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. *Palaeontol. Electron.* **18**, 1.1.FC (2015).
37. Head, J. J. Fossil calibration dates for molecular phylogenetic analysis of snakes 1: Serpentes, Alethinophidia, Boidae, Pythonidae. *Palaeontol. Electron.* **18**, 1.6.FC (2015).

**Acknowledgements** N.J.G. acknowledges the support of Ngatiwai iwi, Allan Wilson Centre, University of Otago, New Zealand Department of Conservation, New Zealand Genomics and

Illumina. J.I.A. was supported by CONICYT National Doctoral Scholarship No. 21130515. M.W. was supported by NIH grant R35 GM124827. Ensembl annotation was supported by the Wellcome Trust (WT108749/Z/15/Z) and the European Molecular Biology Laboratory. We thank Ngāti Koata, Te Ātiawa o Te Waka-a-Māui, and Ngāti Manuhiri iwi for granting permission to reuse tuatara samples obtained from Stephens Island (Takapourewa), North Brother Island and Little Barrier Island (Haururu), respectively; all of the people involved in obtaining and curating the samples held in the Victoria University of Wellington tuatara collection; A. Zimin, D. Puiu, G. Marcais, J. Yorke and R. Crowhurst for help with and discussions about genome assembly; I. Fiddes, J. Armstrong and B. Paten for help with comparative genome alignments and annotation; the National eScience Infrastructure (NeSI) and Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) for computational support; R. McPhee for help with figures; and T. Braisher for manuscript coordination and editing.

**Author contributions** N.J.G. designed the original concept and scientific objectives and oversaw the project and analyses. N.J.G., L.A., N.N., H.T., O.R., S.V.E., C.S. contributed samples or assisted in sample preparation and permitting. N.J.G., K.M.R., S.P., M.T., D.W., J.R.M., D.L.A., A.S., T.B., J.H.G., C.O., P.P.G., M.M. M.P., K.B., F.J.M., P.F., B.P., L.K., P.M., T.R.B., M.W., Y.C., H.M., R.K.S., M.D.J., R.D.N., J.I.A., N.V., T.A.H., J.R., V.P., C.R.P., V.M.W., L.Z., R.D.K., J.M.R., V.V.Z., Z.W., D.S., M. Mariotti, R.G., S.M.R., V.G.T., N.D., H.A., D.M.B., J.M.P., D.G.M., V.L.G., C.G.B., D.P.D., S. Pabinger, T.v.S., S.C., S.L.S. planned and carried out experiments or analyses. N.J.G., K.M.R., S.P., M.T., D.W., J.R.M., D.L.A., A.S., T.B., J.H.G., C.O., P.P.G., M. Muffato, M.P., K.B., F.J.M., B.P., L.K., P.M., T.R.B., M.W., Y.C., H.M., R.K.S., M.D.J., R.D.N., J.I.A., N.V., T.A.H., J.R., V.P., C.R.P., V.M.W., L.Z., R.D.K., J.M.R., V.V.Z., Z.W., D.S., M. Mariotti, R.G., S.M.R., V.G.T., N.D., H.T., H.A., J.M.P., D.G.M., V.L.G., C.G.B., D.P.D., S. Pabinger, S.L.S. contributed to the interpretation and presentation of results in the main manuscript and supplementary documents. N.J.G. wrote the first draft of the manuscript with input from all other authors.

**Competing interests** The authors declare no competing interests.

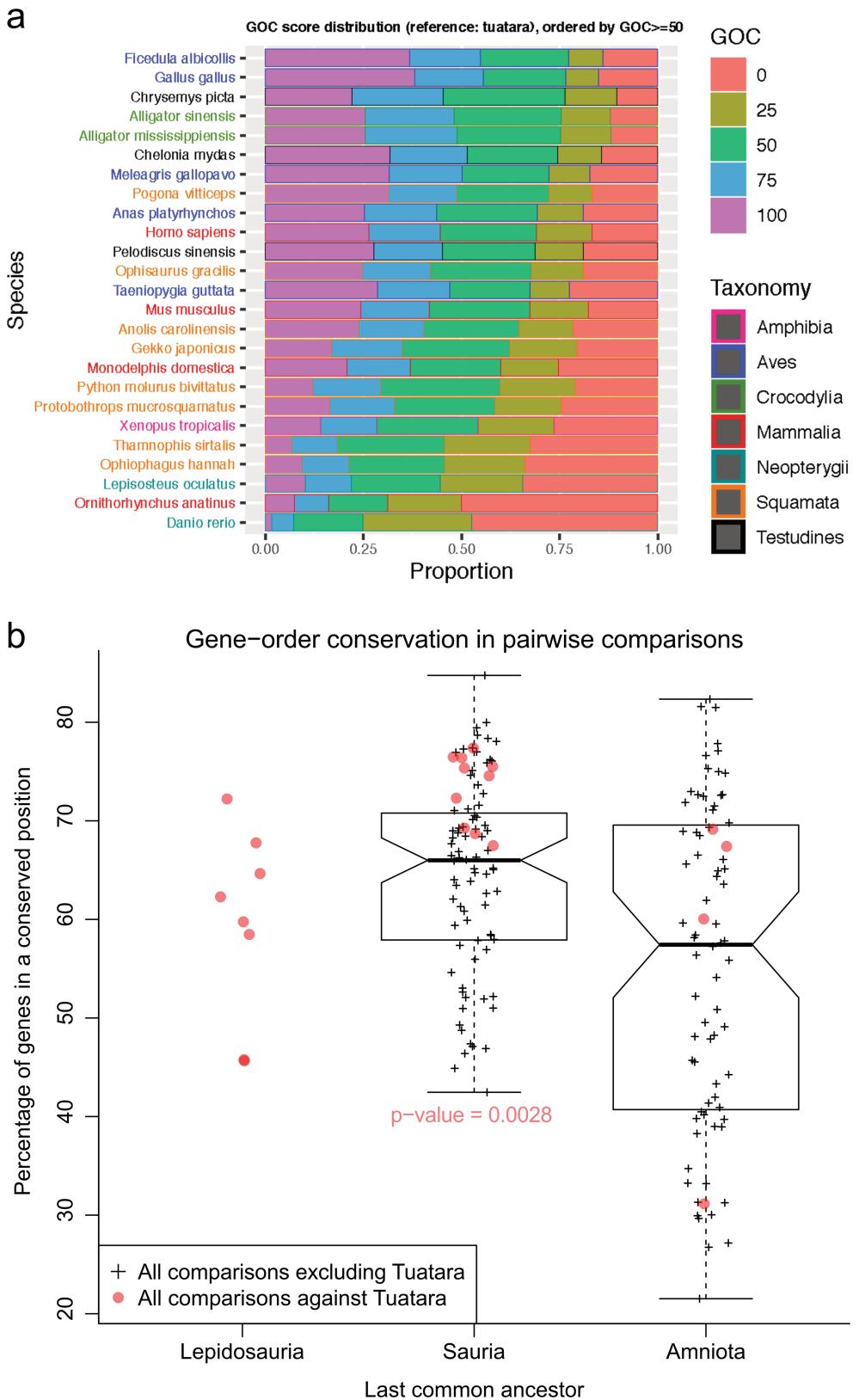
### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-020-2561-9>.

**Correspondence and requests for materials** should be addressed to N.J.G.

**Peer review information** *Nature* thanks Rebecca Johnson and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

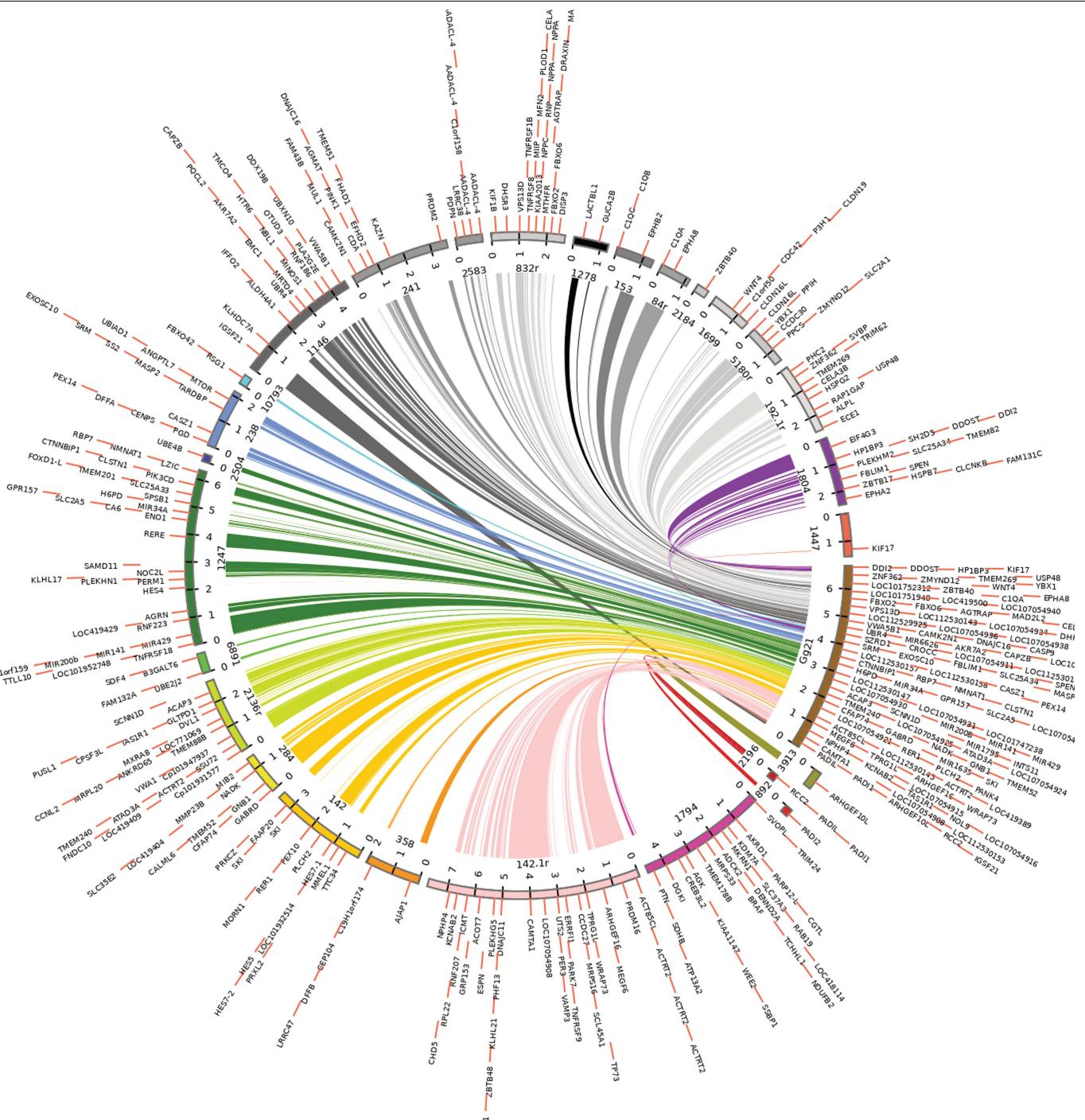
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.



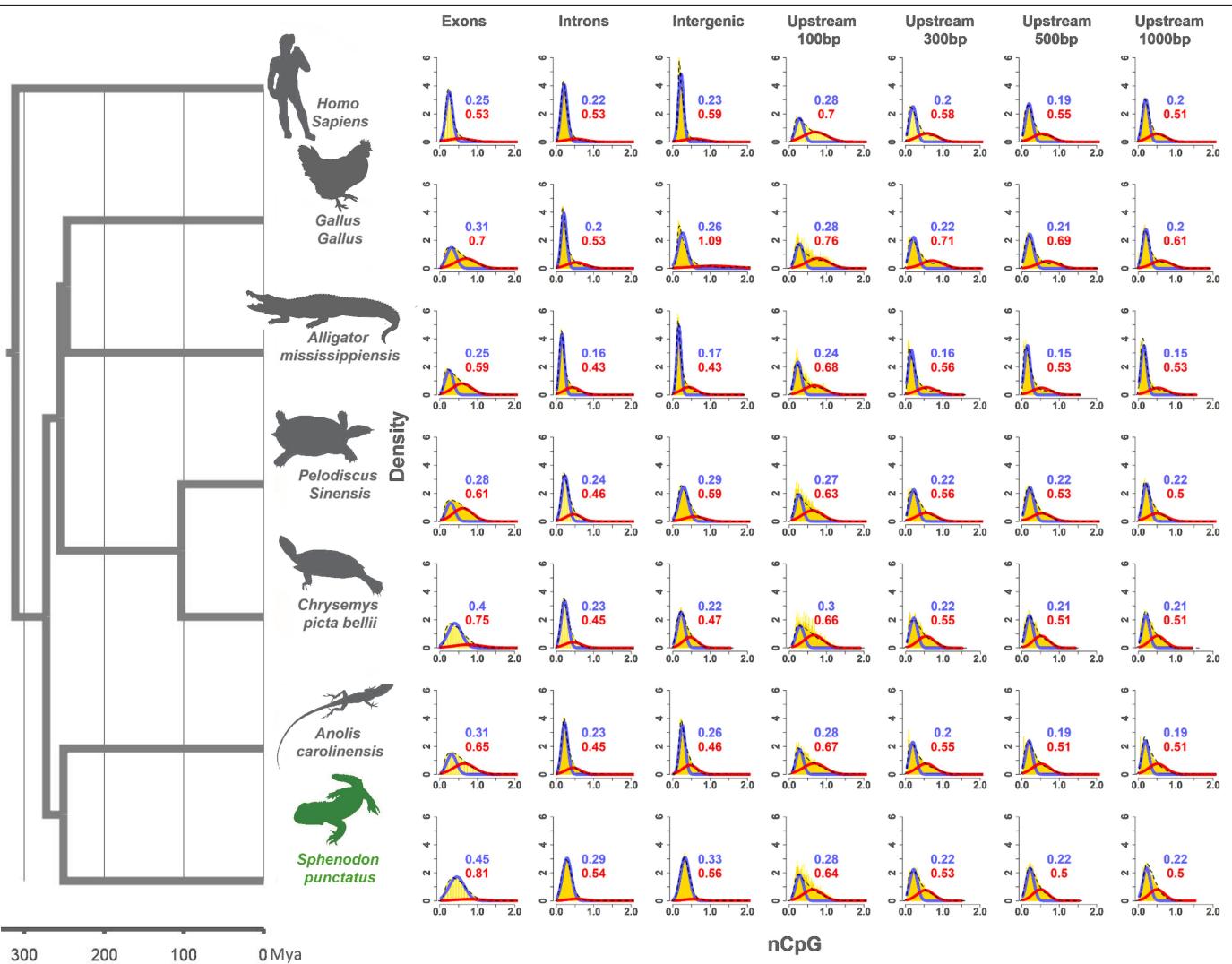
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Gene order conservation.** **a**, Gene-order conservation score distribution using the tuatara as a reference. Species are ordered by the proportion of top-scoring orthologues  $\geq 50$ . **b**, Gene-order conservation versus divergence time. For the three taxonomic groupings (Lepidosauria, Sauria and Amniota), we analysed the percentage of genes that are found in a conserved position across all pairs of genomes. Pairwise comparisons involving tuatara are shown in plain red circles (respectively,  $n=8$ ,  $n=10$ , and  $n=4$ ), and comparisons that do not involve tuatara are black (box plot and + signs; respectively,  $n=0$ ,  $n=80$  and  $n=72$ ). The conservation of gene order

between tuatara, birds and turtles is significantly higher (one-sided, two-sample Kolmogorov–Smirnov test,  $P$  value =  $2.8 \times 10^{-3}$ ,  $D = 0.575$ ) than that observed between squamates, birds and turtles. As the tuatara is the only remaining rhynchocephalian, there is no control distribution for the Lepidosauria ancestor. Box plot coordinates (minimum, first quartile, median, mean, third quartile, maximum) are 42.46%, 57.93%, 66.00%, 64.50%, 70.66% and 84.76% for the Sauria box plot, and 21.52%, 40.81%, 57.43%, 55.17%, 69.46% and 82.34% for the Amniota box plot.

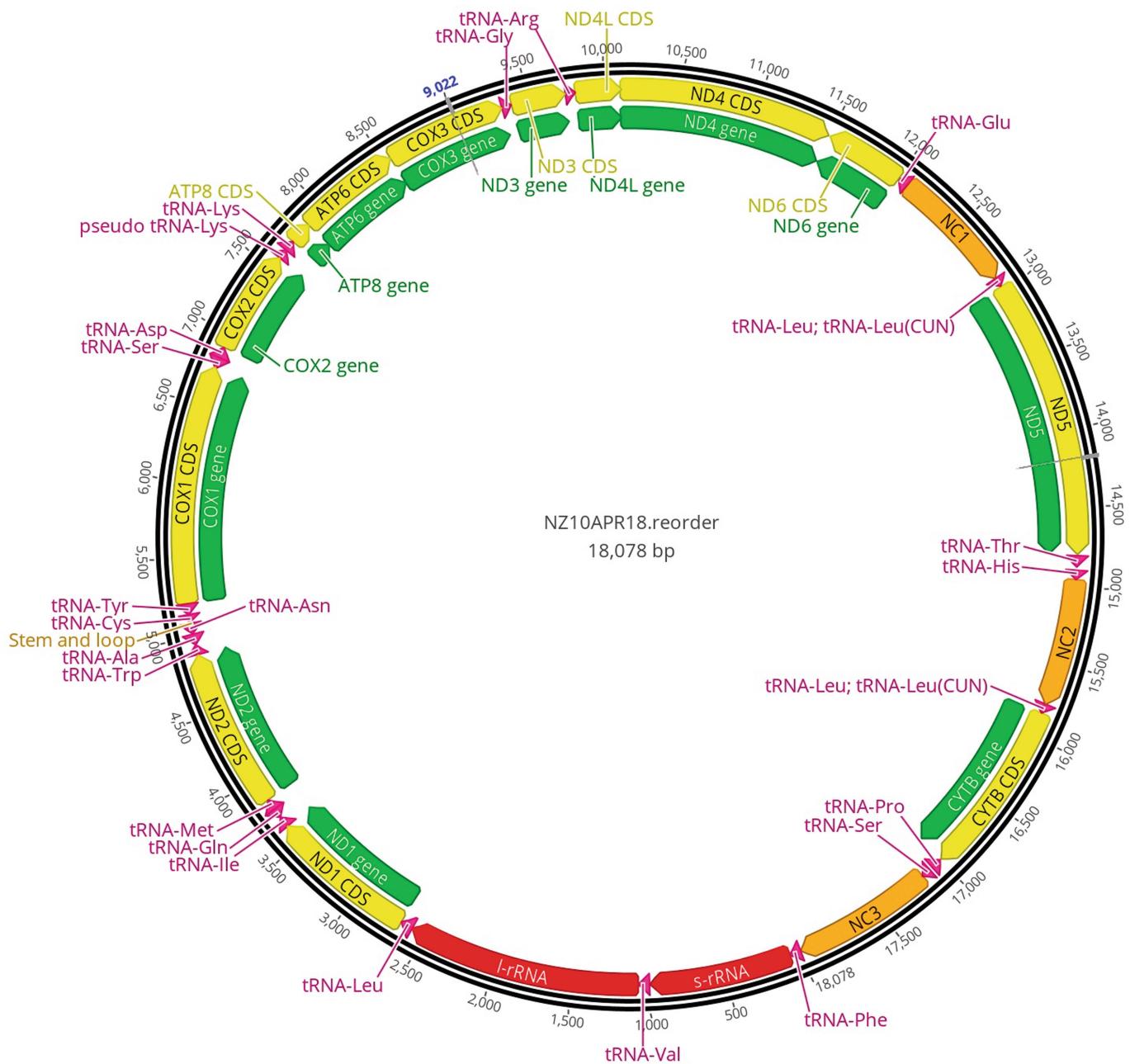


**Extended Data Fig. 2 | Co-linearity between *Gallus gallus* chromosome 21 and tuatara contigs.** Circos plot highlighting the gene-order conservation observed between chicken chromosome 21 (assembly GRCg6a) and multiple tuatara contigs. The gene names shown derive from the chicken assembly.



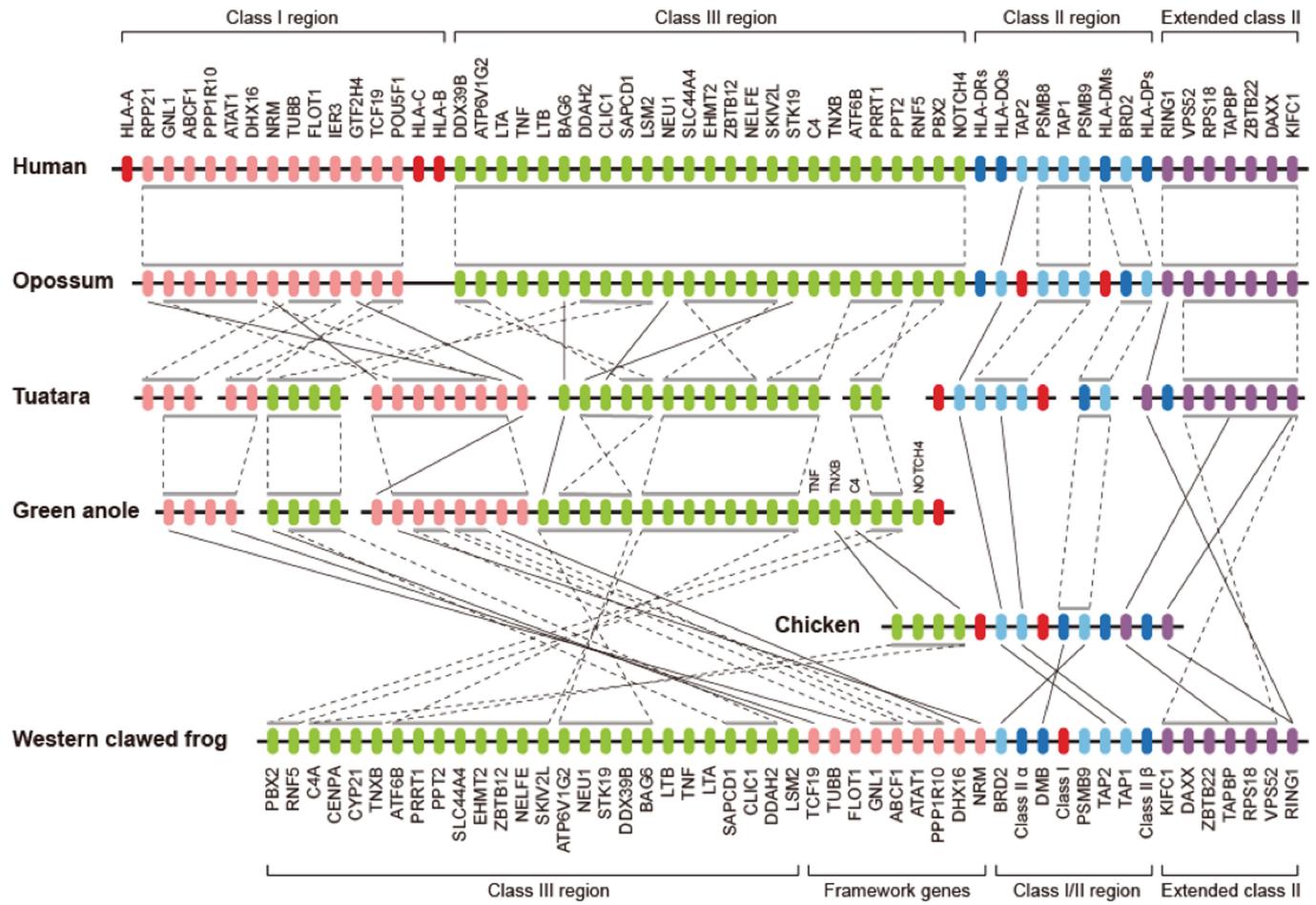
**Extended Data Fig. 3 | Normalized CpG distributions for tuatara and other vertebrates.** Tuatara show a statistically significant bimodal normalized CpG distribution in all genomic regions examined (gene promoters, exons, introns and intergenic sequences). Explicit testing using mixed models uncovered global normalized CpG bimodality across all taxa, indicating that bimodality is

ancestral and highly conserved across vertebrates irrespective of sex-determining mechanism. Further details are provided in Supplementary Information 8. Silhouettes are hand-drawn from the original photographs of A. Eckley, except for *Anolis* (D. Hobern, CC-BY-2.0), crocodile and hen (designed by Creazilla, <https://creazilla.com/>).



**Extended Data Fig. 4 | The mitochondrial genome of the tuatara.** In the Lady Alice Island reference tuatara, this molecule is 18,078 bp, containing 13 protein-coding, 2 rRNA and 22 tRNA genes, standard among animals and contradicting previous reports that 3 genes (*ND5*, *tRNA<sup>Thr</sup>* and *tRNA<sup>His</sup>*) were absent. Three non-coding areas with control region (heavy-strand replication origin) features (NC1, NC2 and NC3), and two copies of *tRNA<sup>Leu(CUN)</sup>* adjacent to NC1 and NC2 possess identical or near-identical sequence, possibly as a result of concerted evolution. A stem-and-loop structure is observed in the region encoding *tRNA<sup>Asn</sup>* and *tRNA<sup>Cys</sup>*, which may supplement for the origin of

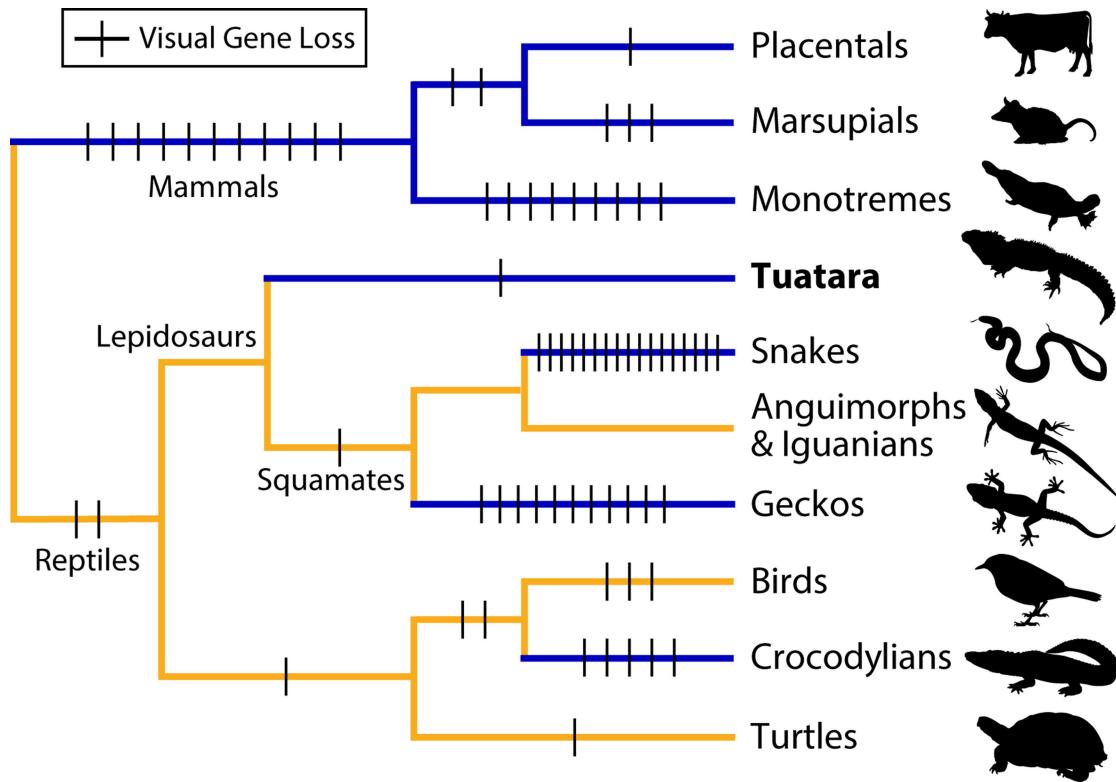
light-strand replication (O<sub>L</sub>) normally found in this location. The *tRNA<sup>Lys</sup>* gene is duplicated with the first copy, possibly a pseudo-gene. The *tRNA<sup>Cys</sup>* gene encodes a tRNA with a D-arm replacement loop. The gene and structure order is: *tRNA<sup>Phe</sup>*, 12S rRNA, *tRNA<sup>Val</sup>*, 16S rRNA, *tRNA<sup>Leu(UUR)</sup>*, ND1, *tRNA<sup>Ile</sup>*, *tRNA<sup>Gln</sup>*, *tRNA<sup>Met</sup>*, ND2, *tRNA<sup>Trp</sup>*, *tRNA<sup>Ala</sup>*, *tRNA<sup>Asn</sup>*, O<sub>L</sub>-like structure, *tRNA<sup>Cys</sup>*, *tRNA<sup>Tyr</sup>*, *COI*, *tRNA<sup>Ser(CUN)</sup>*, *tRNA<sup>Asp</sup>*, *COII*, pseudo-*tRNA<sup>Lys</sup>*, *tRNA<sup>Lys</sup>*, *ATP8*, *ATP6*, *COIII*, *tRNA<sup>Gly</sup>*, *ND3*, *tRNA<sup>Arg</sup>*, *ND4L*, *ND4*, *ND6*, *tRNA<sup>Glu</sup>*, NC1, *tRNA<sup>Leu(CUN)</sup>* copy one, *ND5*, *tRNA<sup>Thr</sup>*, *tRNA<sup>His</sup>*, NC2, *tRNA<sup>Leu(CUN)</sup>* copy two, *CYTB*, *tRNA<sup>Pro</sup>*, *tRNA<sup>Ser(AGY)</sup>* and NC3.



**Extended Data Fig. 5 | Comparative analysis of the MHC core region.** Only genes that were annotated in the tuatara genome were included in the analysis. Orthologues between species are connected by a solid line; the grey bars above and below genes indicate syntenic blocks and are linked by dashed lines between species. *Anolis* MHC class I/II and extended class II regions are not

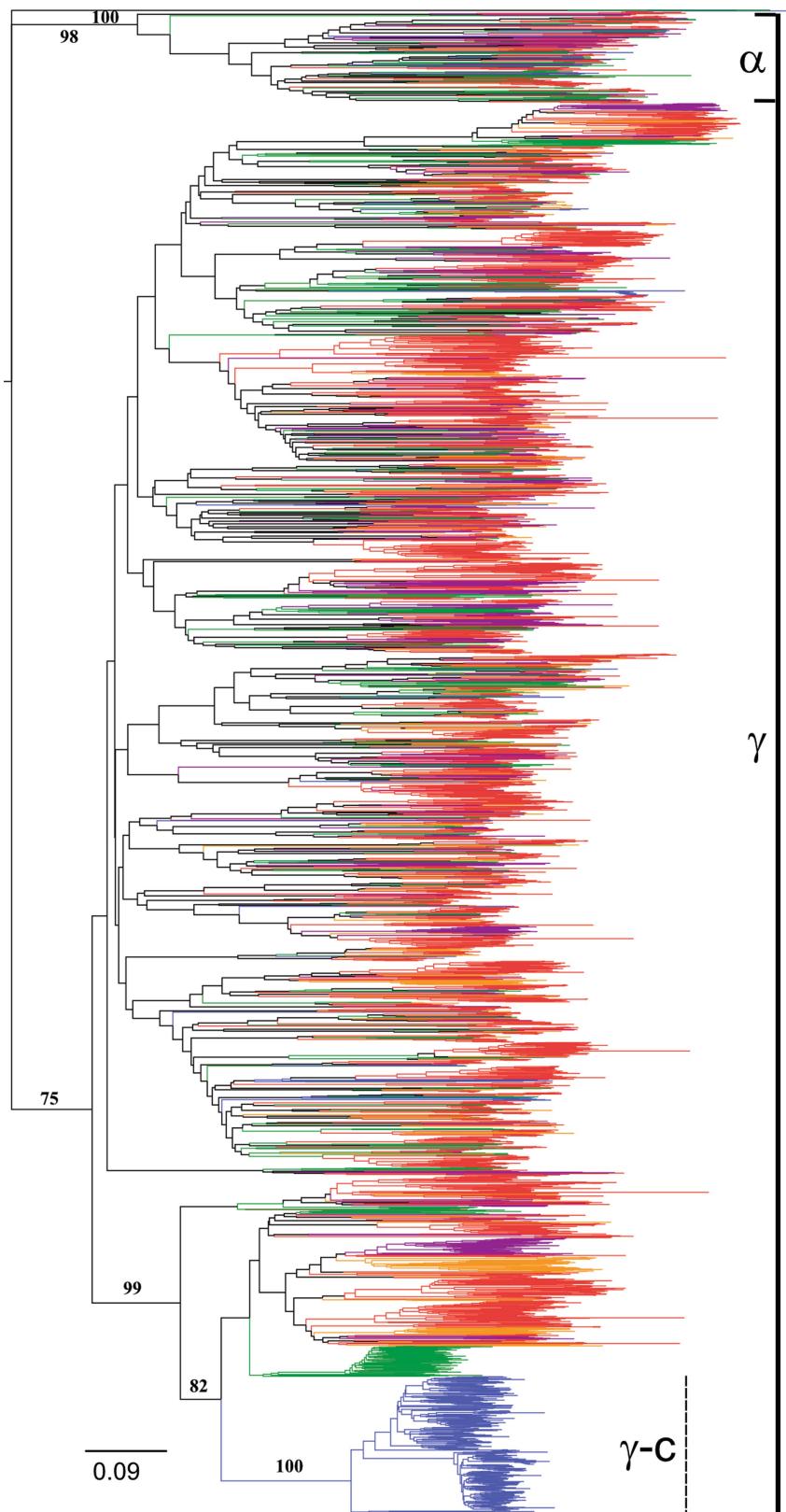
shown, owing to the high degree of genome assembly fragmentation in these regions. Red, class I genes; pink, class I region framework genes; green, class III genes; dark blue, class II genes; light blue, class II region framework genes; purple, extended class II genes.

# Article



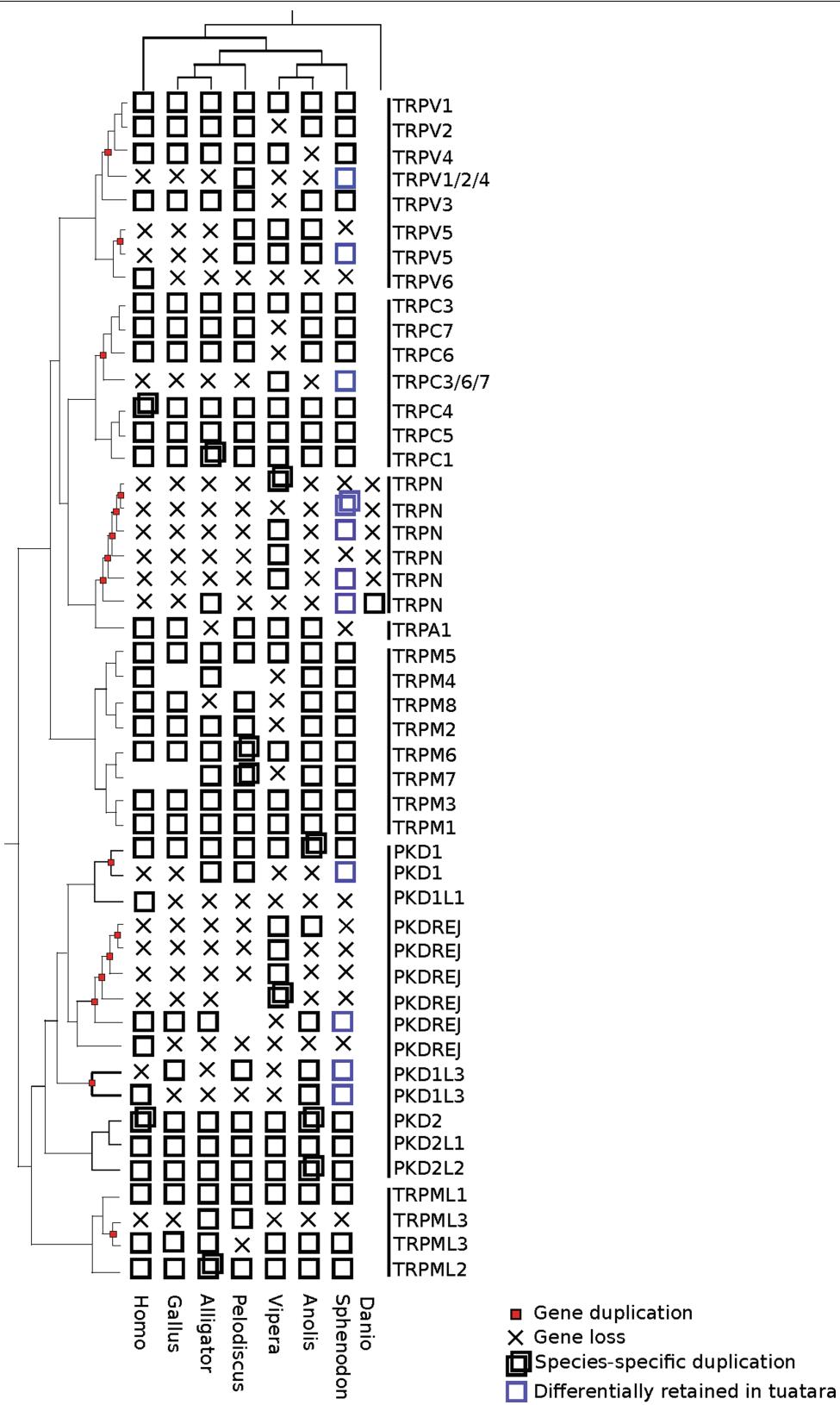
**Extended Data Fig. 6 | Phylogenetic tree of amniotes depicting inferred visual gene losses.** Lineages are coloured based on a rough approximation of their ancestral activity pattern (blue, nocturnal; yellow, diurnal). The tuatara lineage has experienced some of the lowest rates of gene loss despite a nocturnal ancestry, which in other lineages is associated with increased gene

loss. Supplementary Table 10.1 provides a summary of genes lost per branch. The complete table of visual genes analysed and their presence or absence from the various groups is available at <http://www.doi.org/10.5281/zenodo.2597599>. Silhouettes were designed by Creazilla, [www.creazilla.com](http://www.creazilla.com).



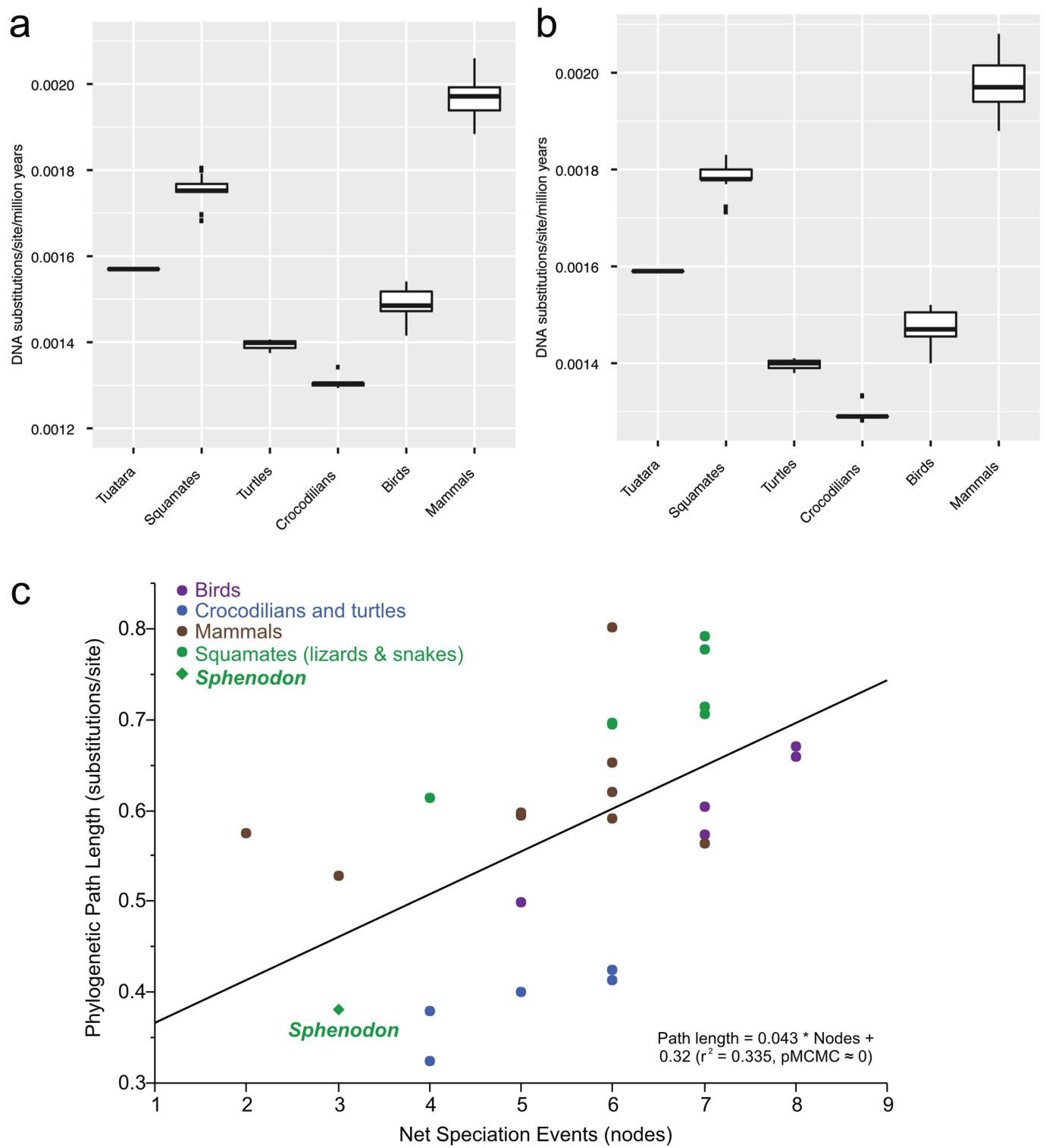
**Extended Data Fig. 7 | The evolutionary history of odorant receptors in terrestrial sauropsids.** The relationship among odorant receptors was inferred using the neighbour-joining method. The unrooted tree contains 3,213 amino acid sequences. Branches are coloured according to the following categories: green, tuatara; blue, birds (*G. gallus* and *Taenioptygia guttata*); red, snakes (*Notechis scutatus*, *Ophiophagus hannah*, *Probobothrops*

*mucrosquamatus*, *Pseudonaja textilis*, *P. bivittatus* and *Thamnophis sirtalis*); orange, lizards (*A. carolinensis* and *Pogona vitticeps*); and purple, gecko (*Gekko japonicus*). Bootstrap support values above 75% (1,000 replicates) are indicated for major branch splits relating to the different odorant receptor groups and branches leading to the species-specific odorant receptor expansions in birds (group  $\gamma$ -c) and tuatara (\*).


**Extended Data Fig. 8 | The repertoire of the TRP genes identified in tuatara.**

We compared the TRP genes identified in tuatara (*S. punctatus*) to those found in other six vertebrate species: lizard (*A. carolinensis*), viper (*Vipera berus*), turtle (*Pelodiscus sinensis*), alligator (*Alligator mississippiensis*), chicken (*G. gallus*), and human (*Homo sapiens*). Small red squares on nodes indicate

gene duplications, duplicated boxes indicate species-specific duplications and blue boxes indicate differential gene retentions in tuatara. Crosses indicate genes that were lost after duplication and empty spaces genes that have most probably been lost, but are not yet confirmed.



Extended Data Fig. 9 | See next page for caption.

# Article

**Extended Data Fig. 9 | Estimated DNA substitution rates of amniote clades on the basis of fourfold-degenerate sites and a test for evidence of punctuated evolution.** **a**, Box plots showing distribution of estimated substitution rates by clade using semiparametric penalized likelihood in r8s with fossil constraints from ref.<sup>36</sup> and ref.<sup>37</sup> for tuatara ( $n=1$ , median = 0.00159), squamates ( $n=13$ , minimum = 0.00171, maximum = 0.00183, median = 0.00178, 25th percentile = 0.00178, 75th percentile = 0.00183), turtles ( $n=3$ , minimum = 0.00138, maximum = 0.00141, median = 0.00140, 25th percentile = 0.00139, 75th percentile = 0.00140), crocodilians ( $n=5$ , minimum = 0.00129, maximum = 0.0134, median = 0.00130, 25th percentile = 0.00130, 75th percentile = 0.00130), birds ( $n=11$ , minimum = 0.00142, maximum = 0.00154, median = 0.00149, 25th percentile = 0.00147, 75th percentile = 0.00152) and mammals ( $n=15$ , minimum = 0.00188, maximum = 0.00206, median = 0.00197, 25th percentile = 0.00194, 75th percentile = 0.00199). **b**, Box plots showing distribution of estimated substitution rates with median time to most recent common ancestor estimates from [www.timetree.org](http://www.timetree.org) for tuatara ( $n=1$ , median = 0.00157), squamates ( $n=13$ , minimum = 0.00168, maximum = 0.00180, median = 0.00175, 25th percentile = 0.00178, 75th percentile = 0.00177), turtles ( $n=3$ , minimum = 0.00138, maximum = 0.00141, median = 0.00140, 25th percentile = 0.00139, 75th percentile = 0.00140), crocodilians ( $n=5$ , minimum = 0.00129, maximum = 0.0134, median = 0.00130, 25th percentile = 0.00130, 75th percentile = 0.00130), birds ( $n=11$ , minimum = 0.00142, maximum = 0.00154, median = 0.00149, 25th percentile = 0.00147, 75th percentile = 0.00152) and mammals ( $n=15$ , minimum = 0.00188, maximum = 0.00206, median = 0.00197, 25th percentile = 0.00194, 75th percentile = 0.00199). **c**, A test for evidence of punctuated evolution. The process of punctuated genome evolution predicts that the amount of evolution in the genome of a given species should correlate with the net number of speciation events. We used Bayesian phylogenetic generalized least squares to regress the total phylogenetic path length (of fourfold-degenerate sites) on the net number of speciation events (nodes in a phylogenetic tree). We find strong evidence for punctuated evolution, which accounts for 33.5% ( $R^2$ ; 95% credible interval = 0.34 to 0.38) of deviation from the molecular clock at fourfold-degenerate sites.

**Extended Data Table 1 | Assembly statistics and quality metrics for the tuatara genome**

	ALLPATHS-LG assembly	Final assembly
Number of scaffolds	48545	16537
Total size of scaffolds	4338416404	4272217537
Longest scaffold	6644062	29987930
N50 scaffold length	322768	3052611
Number of contigs	580919	270321
Longest contig	141552	347348
N50 contig length	10517	27421
CEGMA complete count	89	98
CEGMA complete %	35.89%	39.52%
CEGMA partial count	209	210
CEGMA partial %	84.27%	84.68%
Complete vertebrata BUSCOs	2688	2911
Complete %	80.1	86.8
Complete and single-copy BUSCOs	2667	2882
Complete and duplicated BUSCOs	21	29
Fragmented BUSCOs	353	218
Missing BUSCOs	313	225
Total BUSCO groups searched	3354	3354

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

## Software and code

Policy information about [availability of computer code](#)

### Data collection

Standard genome assembly and bioinformatic tools were employed and are detailed in full in the methods and supplementary materials for the paper. Genome assembly was undertaken using AllPaths-LG (version 49856, [http://software.broadinstitute.org/allpaths-lg/blog/?page\\_id=12](http://software.broadinstitute.org/allpaths-lg/blog/?page_id=12)) and Dovetail Genomics HiRise scaffolding software. Transcriptomes were assembled using Trinity v2.2.0. Bisulphite sequencing data were trimmed using TrimGalore v0.4.0 and reads mapped using Bismark v0.14.311 to identify methylated sites. Repeat annotation was undertaken using CARP, RepeatModeler and LTRharvest. Gene annotation used RepeatMasker (4.0.3) and MAKER2. Genotype-by-sequencing was undertaken using FastQC v0.10.1 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) followed by a QC analysis pipeline: “Deconvolute and quality control” <https://github.com/AgResearch/DECONVQC> ] and subsequent demultiplexing using GBSX, read mapping using BWA MEM, and SNV calling using STACKS and GATK.

### Data analysis

Standard bioinformatic tools were employed for our analyses. These are detailed in the methods and supplementary materials for the paper. Where custom code was utilized this is also specified and either available from GitHub or directly from the authors of the relevant section of our manuscript. All attributions to each component of our work are clearly signalled.

For completeness the full list is provided here also:

Repeat and gene annotation  
 RepeatMasker (v4.0.5), <http://www.repeatmasker.org/>  
 MAKER2 (v2.31.8), <http://www.yandell-lab.org/software/maker.html>  
 BLAST (v2.2.30+), <https://blast.ncbi.nlm.nih.gov/Blast.cgi>  
 SNAP (v2.4.7), <http://snap.cs.berkeley.edu>  
 Augustus (v3.3), <http://augustus.gobics.de>  
 BUSCO (v3.0), <https://busco.ezlab.org>

Ortholog calling

Ensembl GeneTree pipeline: <https://github.com/Ensembl/ensembl> and <https://github.com/Ensembl/ensembl-compara> branch “release/87”

Ensembl Hive workflow management system <https://github.com/Ensembl/ensembl-hive> branch “version/2.3”

Plotting script: <https://github.com/Ensembl/ensembl-compara/blob/release/89/scripts/homology/plotGocData.r>  
 hccluster 0.5.0 <https://sourceforge.net/projects/treesoft>

TCoffee 9.03.r1318 <http://www.tcoffee.org/>

Mafft 7.221 <https://mafft.cbrc.jp/alignment/software/mafft-7.221-with-extensions-src.tgz>

Treebest <https://github.com/Ensembl/treebest>

QuickTree 1.1 <https://github.com/khowe/quicktree>

NCBI Blast 2.2.30+ <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.2.30/>

HMMER 2.3.2 <http://eddylab.org/software/hmmer/2.3.2/hmmer-2.3.2.tar.gz>PantherScore 1.0.3 <https://github.com/Ensembl/pantherScore>PAML 4.3 <http://abacus.gene.ucl.ac.uk/software/SoftOld/paml4.3.tar.gz>Ktreedist 1.0.0 [http://molevol.cmima.csic.es/castresana/Ktreedist/Ktreedist\\_v1.tar.gz](http://molevol.cmima.csic.es/castresana/Ktreedist/Ktreedist_v1.tar.gz)CAFE 2.2 [http://downloads.sourceforge.net/project/cafehahnlab/Previous\\_versions/cafehahnlab-code\\_v2.2.tgz](http://downloads.sourceforge.net/project/cafehahnlab/Previous_versions/cafehahnlab-code_v2.2.tgz)

Ensembl annotation

RepeatMasker, <http://repeatmasker.systemsbiology.net/>RepeatModeler, <http://www.repeatmasker.org/RepeatModeler/>GenBlast, <http://genome.sfu.ca/genblast/>BWA, <http://bio-bwa.sourceforge.net/>MUSCLE, <http://www.drive5.com/muscle>BLAST, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>ensembl-analysis, <https://github.com/Ensembl/ensembl-analysis.git>ensembl code, [https://www.ensembl.org/info/docs/api/api\\_installation.html](https://www.ensembl.org/info/docs/api/api_installation.html)exonerate, <https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate>rnafold, <https://github.com/choener/RNAfold>Inferno, <http://eddylab.org/infernal/>

Investigation of gene co-linearity

BLAST+ 2.8.1 using Geneious version 10.2.6 (<https://www.geneious.com>).Circos software <http://circos.ca/software/>

Ab initio repeat annotation

CARP, <https://github.com/carp-te/carp-documentation> RepeatModeler, <http://www.repeatmasker.org/RepeatModeler/>CENSOR, which requires wu-blast and bioperl, <https://girinst.org/downloads/software/censor/> RepeatMasker, <http://www.repeatmasker.org/>BLASTN, <https://blast.ncbi.nlm.nih.gov/Blast.cgi> RepBase, <https://www.girinst.org/server/RepBase/>MUSCLE, <https://www.drive5.com/muscle/>MrBayes, <https://nbisweden.github.io/MrBayes/download.html>FastTree, <http://www.microbesonline.org/fasttree/>USEARCH, <https://www.drive5.com/usearch/>HMMER, <http://hmmer.org/> PILER, <https://www.drive5.com/piler/>

Repeat annotation of SINEs and DNA transposons

RepeatModeler version 1.0.8 <http://www.repeatmasker.org/RepeatModeler/>renameRMDLconsensi.pl script <https://github.com/genomicrocasm/physaliaTEcourse/blob/master/>

Practical2\_Computational\_annotation/renameRMDLconsensi.pl

repeatModelerPipeline4.pl script [https://github.com/genomicrocasm/physaliaTEcourse/blob/master/Practical3\\_Manual\\_curation/](https://github.com/genomicrocasm/physaliaTEcourse/blob/master/Practical3_Manual_curation/)

repeatModelerPipeline4.pl

BLASTN version 2.2.28+ [https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=Download](https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download)MAFFT version 6 standalone and version 7 webserver <https://mafft.cbrc.jp/alignment/software/>BioEdit version 7.2.6.1 <https://bioedit.software.informer.com/download/>CENSOR webserver <http://www.girinst.org/censor/index.php>RAxML version 8.0.0 in CIPRES Science Gateway <https://www.phylo.org/portal2/login!input.action>MEGA version 5.2 <https://www.megasoftware.net/>RepeatMasker version 4.0.7 <http://www.repeatmasker.org/RMDownload.html>calcDivergenceFromAlign.pl script in RepeatMasker package <http://www.repeatmasker.org/RMDownload.html>R scripts for making landscape plots <https://github.com/ValentinaBoP/TuataraTELandscapes/blob/master/>

Tuatara\_DNA\_SINE\_landscape\_figures.Rmd

LTR analyses

Genometools, <http://genometools.org/pub/genometools-1.5.8.tar.gz>, used for indexing genome and running LTRharvest.Blastn, <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.5.0.ncbi-blast-2.5.0+x64-linux.tar.gz>, used for blasting retroviral proteins.MAFFT 7, <https://mafft.cbrc.jp/alignment/software/>, used for multiple alignment of peptide sequences.CD-HIT-V4.6.5, <https://github.com/weizhongli/cdhit>, used to reduce number of similar retroelement copies in file.RAxML 8.2, <https://github.com/stamatak/standard-RAxML>, used to infer phylogenetic trees under maximum likelihood.NINJA 1.2.2, [http://nimbletwist.com/software/ninja/old\\_distros/ninja\\_1.2.2.tgz](http://nimbletwist.com/software/ninja/old_distros/ninja_1.2.2.tgz), used for large-scale neighbor-joining phylogeny inference.Inkscape 0.92, <https://inkscape.org/release/inkscape-0.92/>, used for preparing figures of repetitive elements.

RNA annotation

Rfam (version 13.0) covariance models, <ftp://ftp.ebi.ac.uk/pub/databases/Rfam/13.0/>tRNAscan-SE (version 1.3.1), <http://lowelab.ucsc.edu/software/>Infernal (version 1.1), <http://eddylab.org/infernal/>

- Mitochondrial genome sequence and assembly  
Illumina data: Bowtie 2, MaSuRCA, Minimus, Jellyfish, Geneious v10.2.4 (<https://www.geneious.com>).  
Oxford Nanopore data: Nanopolish, Guppy, Blastrn, Megablast, Discontiguous Megablast.
- MHC  
BLAST+ 2.3.0 <https://blast.ncbi.nlm.nih.gov/Blast.cgi> Geneious 9.1, (<https://www.geneious.com>).
- Opsin gene analysis  
BLAST+ 2.2.30, PAML 4.8, MEGA 5.2, PhyML 3.0, BLASTPHYME <https://github.com/ryankschott/BlastPhyMe>
- Odorant receptors  
tBLASTn in Geneious 10.0.3 (<https://www.geneious.com>), MAFFT (v7.338), MEGA 7.0.21, FigTree v1.4.4
- Transient receptor potential (TRP) ion channel gene analysis  
tBLASTn, <https://blast.ncbi.nlm.nih.gov/Blast.cgi> MAFFT v7.450 / <https://mafft.cbrc.jp/alignment/software/>  
FastTree2 as implemented in the CIPRES portal <https://www.phylo.org/>
- Selenoprotein analysis  
Selenoprofiles v. 3.6, <https://github.com/marco-mariotti/selenoprofiles>  
Secmarker v. 0.4, <https://secmarker.crg.cat>,
- Phylogeny and evolutionary rates  
LASTZ-chaining-netting/source code from the avian phylogenomics project (Zhang et al. 2014; GigaScience) [https://github.com/gigascience/paper-zhang2014/tree/master/Whole\\_genome\\_alignment/](https://github.com/gigascience/paper-zhang2014/tree/master/Whole_genome_alignment/)  
msa\_view in PHAST v1.3 [http://compgen.cshl.edu/phast/oldversions.php/](http://compgen.cshl.edu/phast/oldversions.php)  
RAxML v8.2.3/<https://github.com/stamatak/standard-RAxML>  
phyloFit in PHAST v1.3 <http://compgen.cshl.edu/phast/oldversions.php/>  
r8s v1.8 / <https://sourceforge.net/projects/r8s/>  
PRANK v1.7 / <http://wasabiapp.org/download/prank/>  
FASconCAT-G v.1.02 / <https://github.com/PatrickKueck/FASconCAT-G>  
RAxML v8.2.3/<https://github.com/stamatak/standard-RAxML>  
ASTRAL v4 / <https://github.com/smirarab/ASTRAL>  
MAFFT v7.450 / <https://mafft.cbrc.jp/alignment/software/>  
AMAS / <https://github.com/marekborowiec/AMAS>  
IQ-TREE v1.6.12 / <http://www.iqtree.org/>
- Punctuated evolution  
Time-calibrated phylogeny: BEAST v2.4.8 on the CIPRES Science Gateway: <https://www.phylo.org/>  
Punctuated evolution: BayesTraits V3.0.2 Nov 2019: <http://www.evolution.rdg.ac.uk/BayesTraitsV3.0.2/BayesTraitsV3.0.2.html>  
Node-density artefact: Test for Punctuational Evolution and the Node-Density Artifact. v1: <http://www.evolution.reading.ac.uk/pe/index.html>
- Patterns of selection  
TranslatorX, <http://translatorx.co.uk/>, Translation of nucleotide sequences  
MAFFT 7.310, <https://mafft.cbrc.jp/alignment/software/> , amino acid alignment  
PAL2NALv14, <https://github.com/drostlab/orthologr/tree/master/inst/pal2nal/pal2nal.v14> , conversion of protein alignments into codon-based DNA alignments  
TrimAI 1.2, <http://trimal.cgenomics.org/>, Alignment correction  
Garli 2.0.1, [https://www.nescent.org/wg\\_garli/Main\\_Page](https://www.nescent.org/wg_garli/Main_Page), maximum likelihood phylogenetic reconstruction  
PAML 4.5, <http://abacus.gene.ucl.ac.uk/software/paml.html> , infer branch specific evolutionary patterns  
QVALUE, <https://www.bioconductor.org/packages/release/bioc/html/qvalue.html> , false discovery rate correction  
Kobas 2.0, <http://kobas.cbi.pku.edu.cn/kobas3/?t=1>, Gene set enrichment analysis  
Galaxy's Stitch Gene Blocks Tool, <http://www.bioinformatics.nl/galaxy>  
AlignmentProcessor0.12, <https://github.com/WilsonSayresLab/AlignmentProcessor>  
PAML, CodeML, <http://abacus.gene.ucl.ac.uk/software/paml.html>  
R v3.3.1, <https://cran.r-project.org/>
- Reconstruction of the demographic history of the tuatara  
BWA mem, <http://bio-bwa.sourceforge.net/>  
Samtools mpileup, <http://www.htslib.org/>  
PSMC, <https://github.com/lh3/psmc>
- Population genomics  
FastQC v0.10.1, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> QC analysis pipeline, <https://github.com/AgResearch/DECONVQC>  
GBSX, <https://github.com/GenomicsCoreLeuven/GBSX>  
BWA mem, <http://bio-bwa.sourceforge.net/>  
STACKS 1.4.4, <http://catchenlab.life.illinois.edu/stacks/>  
GATK haplotypecaller, <https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data are freely available. The Tuatara Genome Consortium Project Whole Genome Shotgun and genome assembly are registered under the umbrella BioProject PRJNA418887 and BioSample SAMN08038466. Transcriptome read data are submitted under SRR7084910 (whole blood) together with prior data SRR485948. The transcriptome assembly is submitted to GenBank with ID GGNQ00000000.1. Illumina short-read and nanopore long read sequence are in SRAs associated with PRJNA445603. The assembly (GCA\_003113815.1) described in this paper is version QEPC00000000.1 and consists of sequences QEPC01000001-QEPC01016536. Maker gene predictions are available from Zenodo, DOI: 10.5281/zenodo.1489353. The repeat library database developed for tuatara is available from Zenodo, DOI: 10.5281/zenodo.2585367. Other data for analyses in specific sections of our paper have been uploaded to Zenodo and DOIs are clearly indicated in the paper.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description

This is the study of the genome, transcriptome and methylome of one exemplar male tuatara from Lady Alice Island in the far North of New Zealand. We have subsequently compared this genome to other published genomes, while also undertaking population genomic investigations of tuatara using 30 individuals that span the main axes of genetic diversity for tuatara.

Research sample

1 exemplar tuatara (*Sphenodon punctatus*) from Lady Alice Is. New Zealand. 30 other tuatara samples (approximately half male, half female) spanning three main populations (Lady Alice, Stephens, and Brothers Is.) that encompass the main axes of genetic diversity for this species.

Sampling strategy

Sampling was undertaken using venipuncture. Because this species is special to Maori and highly protected, our sampling strategy was ad hoc relying heavily on samples collected previously for other studies. While no tests of statistical power were undertaken past investigations of population structure in this species suggested that 10 samples per population was likely adequate to capture much of the variation present. The use of equal numbers of males and females provided a reasonable opportunity to explore obvious sex differences, should these be present.

Data collection

Blood samples for the population-level analysis were collected during field work for research on other projects, for example, investigating population size and genetic diversity of island populations of tuatara. Islands were searched for tuatara at night when they were active above ground. Tuatara were captured by hand, and ~1ml blood samples were obtained from the caudal artery. Samples were stored in liquid nitrogen then once back at the lab, transferred to a -80 °C freezer. The exemplar tuatara sample was collected during a survey trip to Lady Alice Island, but otherwise procedures were similar to earlier population samples.

Timing and spatial scale

Samples were collected at each of three sites (Lady Alice, Stephens, and Brothers Is.) from 1984 to 2011.

Data exclusions

Several samples (2 out of 30) failed in downstream genotype by sequencing and bisulphite methylation profiling due to DNA quality/quantity issues.

Reproducibility

Data reproducibility was verified using repeat sequencing and independent analyses using alternative pipelines. e.g. genome assembly used at least three independent pipelines, each of which had high concordance.

Randomization

There was no need to randomise our study given the focus on genomic and population genomics

Blinding

Blinding was irrelevant given the focus on genomic and population genomics

Did the study involve field work?

Yes       No

## Field work, collection and transport

Field conditions

Data not available, but field sites are all temperate offshore Islands in New Zealand

Location

Lady Alice, Stephens, and Brothers Is. New Zealand

Access and import/export

Samples were collected with the permission and support of local iwi and the NZ DOC, under Victoria University of Wellington

## Access and import/export

Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347 and held and used under permits 45462-DOA (1/9/15) and 32037-RES (11/11/11) issued by the New Zealand Department of Conservation. Samples from the exemplar Lady Alice animals were shipped from NZ to ZSD in USA using CITES Permit to Export – Permit # 13NZ000096 (25/7/13), NZ Dept. of Conservation Authority to Export – Permit # 36830-RES (11/7/2013), and CITES Import Permit # 13US727416/9 and US Federal Fish and Wildlife Permit # LE736007-0 (15/7/13). The ethics application and other permitting processes ensures minimal numbers of individuals are used in research, and that their use is justified under New Zealand and international law. All approvals issued by the Department of Conservation ensure the research complies with relevant acts of Parliament for access to collection sites and handling and research on native species of New Zealand. All Department of Conservation permits for capture, sampling, require consultation with local indigenous people affiliated with the islands.

## Disturbance

Animals were handled minimally and returned to the site of capture for release.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

- |                                     |                             |
|-------------------------------------|-----------------------------|
| n/a                                 | Involved in the study       |
| <input checked="" type="checkbox"/> | Antibodies                  |
| <input checked="" type="checkbox"/> | Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | Palaeontology               |
| <input type="checkbox"/>            | Animals and other organisms |
| <input checked="" type="checkbox"/> | Human research participants |
| <input checked="" type="checkbox"/> | Clinical data               |

## Methods

- |                                     |                        |
|-------------------------------------|------------------------|
| n/a                                 | Involved in the study  |
| <input checked="" type="checkbox"/> | ChIP-seq               |
| <input checked="" type="checkbox"/> | Flow cytometry         |
| <input checked="" type="checkbox"/> | MRI-based neuroimaging |

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

## Laboratory animals

This study did not involve the use of laboratory animals

## Wild animals

Adult tuatara, (*Sphenodon punctatus*) from Lady Alice, Stephens, and Brothers Is. New Zealand were captured and blood samples taken using established venipuncture approaches. Animals were captured while active outside their burrows at night. Blood samples were taken upon capture and animals were released at their site of capture. If animals were held while others were being sampled (<1h), they were placed into a cloth bag. No animals died as a result of this study.

## Field-collected samples

Bloods from one exemplar male and a further 30 animals, previously collected for other purposes, were utilised for our work. Sex ratios among samples were approximately 50:50 and equal numbers of samples were obtained from all sites. One sample was collected specifically for this study - the exemplar - but during another research project, so it did not require separate arrangements solely for this sample. All others arose from frozen samples from previous studies.

## Ethics oversight

Samples were collected under Victoria University of Wellington Animal Ethics approvals 2006R12; 2009R12; 2012R33; 22347.

Note that full information on the approval of the study protocol must also be provided in the manuscript.