

# Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards

Abigail L. Lind<sup>1</sup>, Yvonne Y. Y. Lai<sup>2</sup>, Yulia Mostovoy<sup>2</sup>, Alisha K. Holloway<sup>1</sup>, Alessio Iannucci<sup>3</sup>, Angel C. Y. Mak<sup>2</sup>, Marco Fondi<sup>3</sup>, Valerio Orlandini<sup>3</sup>, Walter L. Eckalbar<sup>4</sup>, Massimo Milan<sup>5</sup>, Michail Rovatsos<sup>6,7</sup>, Ilya G. Kichigin<sup>8</sup>, Alex I. Makunin<sup>8</sup>, Martina Johnson Pokorná<sup>6,7</sup>, Marie Altmanová<sup>6,7</sup>, Vladimir A. Trifonov<sup>8</sup>, Elio Schijlen<sup>9</sup>, Lukáš Kratochvíl<sup>6</sup>, Renato Fani<sup>3</sup>, Petr Velenský<sup>10</sup>, Ivan Reháč<sup>10</sup>, Tomaso Patarnello<sup>5</sup>, Tim S. Jessop<sup>11</sup>, James W. Hicks<sup>12</sup>, Oliver A. Ryder<sup>13</sup>, Joseph R. Mendelson III<sup>14,15</sup>, Claudio Ciofi<sup>3</sup>, Pui-Yan Kwok<sup>12,4,16</sup>, Katherine S. Pollard<sup>1,4,17,18,19\*</sup> and Benoit G. Bruneau<sup>1,2,20\*</sup>

**Monitor lizards are unique among ectothermic reptiles in that they have high aerobic capacity and distinctive cardiovascular physiology resembling that of endothermic mammals. Here, we sequence the genome of the Komodo dragon *Varanus komodoensis*, the largest extant monitor lizard, and generate a high-resolution de novo chromosome-assigned genome assembly for *V. komodoensis* using a hybrid approach of long-range sequencing and single-molecule optical mapping. Comparing the genome of *V. komodoensis* with those of related species, we find evidence of positive selection in pathways related to energy metabolism, cardiovascular homeostasis, and haemostasis. We also show species-specific expansions of a chemoreceptor gene family related to pheromone and kairomone sensing in *V. komodoensis* and other lizard lineages. Together, these evolutionary signatures of adaptation reveal the genetic underpinnings of the unique Komodo dragon sensory and cardiovascular systems, and suggest that selective pressure altered haemostasis genes to help Komodo dragons evade the anticoagulant effects of their own saliva. The Komodo dragon genome is an important resource for understanding the biology of monitor lizards and reptiles worldwide.**

The evolution of form and function in non-avian reptiles contains numerous examples of innovation and diversity. There are an estimated 10,000 reptile species worldwide, found on every continent except Antarctica, with a diversity of lifestyles and morphologies<sup>1</sup> corresponding to a broad range of anatomical and physiological adaptations. Understanding how these adaptations evolved through changes to biochemical and cellular processes will reveal fundamental insights into areas ranging from anatomy and metabolism to behaviour and ecology.

The varanid lizards (genus *Varanus*, or monitor lizards) are an unusual group within squamate reptiles (lizards and snakes). Varanids exhibit the largest range in size among reptiles, varying in mass by over five orders of magnitude (8 g–100 kg)<sup>2</sup>. Varanids have a unique cardiopulmonary physiology and metabolism with numerous parallels to the mammalian cardiovascular system. For example, their cardiac anatomy is characterized by well-developed

ventricular septa ('muscular ridge' and 'bulbus lamellae') resulting in a functionally divided heart<sup>3</sup>. This enables a dual-pressure cardiovascular system characterized by high systemic and low pulmonary blood pressures<sup>3</sup>. Furthermore, varanid lizards can achieve and sustain higher aerobic metabolic rates and endurance capacity than similar size non-varanid squamates, which enables intense, sustainable movements while hunting prey or in bouts of male–male combat. The largest of the varanid lizards, the Komodo dragon *Varanus komodoensis*, can grow to 3 m in length and run up to 20 km h<sup>-1</sup>, allowing them to hunt large prey including deer and boar<sup>4</sup>. Komodo dragons have a higher metabolism than predicted by allometric scaling relationships for varanid lizards<sup>5</sup>, which may explain their capacity for daily movement to locate prey<sup>6</sup>. Their ability to locate injured or dead prey through scent tracking over several kilometres is enabled by a powerful olfactory system<sup>4</sup>, and their hunting is aided by serrate teeth, sharp claws, and saliva with

<sup>1</sup>Gladstone Institutes, San Francisco, CA, USA. <sup>2</sup>Cardiovascular Research Institute, University of California, San Francisco, CA, USA. <sup>3</sup>Department of Biology, University of Florence, Sesto Fiorentino, Italy. <sup>4</sup>Institute for Human Genetics, University of California, San Francisco, CA, USA. <sup>5</sup>Department of Comparative Biomedicine and Food Science, University of Padova, Legnaro, Italy. <sup>6</sup>Department of Ecology, Charles University, Prague, Czech Republic. <sup>7</sup>Institute of Animal Physiology and Genetics, The Czech Academy of Sciences, Liběchov, Czech Republic. <sup>8</sup>Institute of Molecular and Cellular Biology SB RAS, Novosibirsk, Russia. <sup>9</sup>B.U. Bioscience, Wageningen University, Wageningen, The Netherlands. <sup>10</sup>Prague Zoological Garden, Prague, Czech Republic. <sup>11</sup>Centre for Integrative Ecology, Deakin University, Waurn Ponds, Victoria, Australia. <sup>12</sup>Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of California, Irvine, CA, USA. <sup>13</sup>Institute for Conservation Research, San Diego Zoo, Escondido, CA, USA. <sup>14</sup>Zoo Atlanta, Atlanta, GA, USA. <sup>15</sup>School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA. <sup>16</sup>Department of Dermatology, University of California, San Francisco, CA, USA. <sup>17</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA. <sup>18</sup>Institute for Computational Health Sciences, University of California, San Francisco, CA, USA. <sup>19</sup>Chan-Zuckerberg BioHub, San Francisco, CA, USA. <sup>20</sup>Department of Pediatrics, University of California, San Francisco, CA, USA. \*e-mail: [katherine.pollard@gladstone.ucsf.edu](mailto:katherine.pollard@gladstone.ucsf.edu); [benoit.bruneau@gladstone.ucsf.edu](mailto:benoit.bruneau@gladstone.ucsf.edu)

**Table 1 | Genome statistics of the Komodo dragon genome**

Assembly size	1.51 Gb (1,507,945,839 bp)
Number of scaffolds	1,411
Minimum scaffold length	10 kb
Maximum scaffold length	138 Mb
N50 scaffold length	29 Mb (29,129,838)
Number of protein-coding genes	18,457
GC content	44.04%

anticoagulant and shock-inducing properties<sup>7,8</sup>. Furthermore, Komodo dragons engage in aggressive intraspecific conflicts over mating, territory and food, and wild individuals often bear scars from previous conflicts<sup>4</sup>.

To understand the genetic underpinnings of Komodo dragon physiology, we sequenced its genome and present a de novo assembly, generated with a hybrid approach of Illumina short-read sequencing with long-range sequencing using 10x Genomics, PacBio and Oxford Nanopore sequencing, and single-molecule optical mapping using the Bionano platform. This suite of technologies allowed us to confidently assemble a high-quality reference genome for the Komodo dragon, which can serve as a template for other varanid lizards. We used this genome to understand the relationship of varanids to other reptiles using phylogenomics. We uncovered Komodo dragon-specific positive selection for genes encoding regulators of muscle metabolism, cardiovascular homeostasis, and haemostasis. Furthermore, we discovered multiple lineage-specific expansions of a family of chemoreceptor genes in several squamates. Finally, we generated a high-resolution chromosomal map by assigning genomic scaffolds to chromosomes, enabling us to address questions about karyotype and sex chromosome evolution in squamates.

## Results

**De novo genome assembly.** We sequenced the Komodo dragon genome principally from DNA isolated from peripheral blood of two male Komodo dragons housed at Zoo Atlanta: Slasher, offspring of the first Komodo dragons given to US President Reagan from President Suharto of Indonesia, and Rinca, an unrelated juvenile. A third individual from Gran Canaria was used for PacBio DNA sequencing. The Komodo dragon genome is distributed across 20 pairs of chromosomes, comprising eight pairs of large chromosomes and 12 pairs of microchromosomes<sup>3,10</sup>. De novo assembly was performed with a combination of 10x Genomics linked-read sequencing, Bionano optical mapping data, PacBio sequencing and Oxford Nanopore MinIon sequencing (Methods). The final assembly contained 1,411 scaffolds (>10 kb) with an N50 scaffold length of 24 Mb (longest scaffold: 138 Mb; Table 1). The assembly is 1.51 Gb in size, ~32% smaller than the genome of the Chinese crocodile lizard *Shinisaurus crocodilurus*<sup>11</sup>, the closest relative of the Komodo dragon for which a sequenced genome is available, and ~15% smaller than the green anole *Anolis carolinensis*<sup>12</sup>, a model squamate lizard (Supplementary Table 1). An assembly-free error corrected *k*-mer counting estimate of the Komodo dragon genome size<sup>13</sup> is 1.69 Gb, while a flow cytometry-based estimate of the Komodo dragon genome size is 1.89 Gb (ref. <sup>14</sup>; estimated 3.86 pg of DNA per nucleus, with a conversion factor of 978 Mb pg<sup>-1</sup> (ref. <sup>15</sup>)). Gaps comprise 0.97% of the assembly. We assessed the completeness of the Komodo dragon genome assembly by searching for 2,586 single-copy vertebrate genes using BUSCO<sup>16</sup>. The Komodo dragon genome has a similar distribution of single-copy (95.7%), duplicated (0.4%), fragmented (2%) and missing (1.9%) universal vertebrate genes as other reptile genomes (Supplementary Table 3). The GC content of the Komodo dragon genome is 44.0%, similar to that of the *S. crocodilurus*

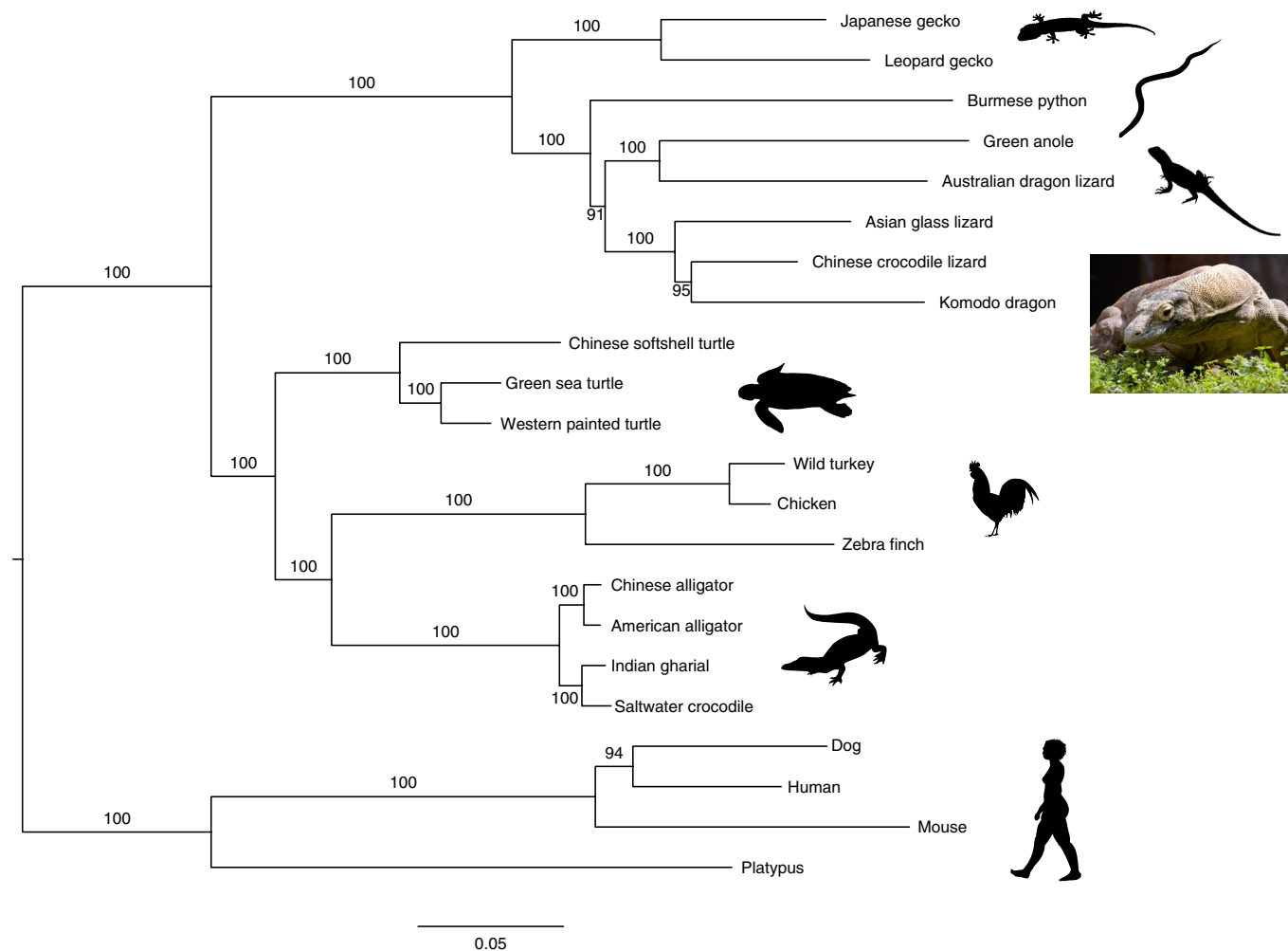
**Table 2 | Results of scaffold assignments to chromosomes of *V. komodoensis***

<i>V. komodoensis</i> chromosome	<i>G. gallus</i> homology	<i>A. carolinensis</i> homology	Total number of assigned scaffolds	Total length of assigned scaffolds (bp)
Chr1	Chr1, 3, 5, 18, Z	Chr1, 2, 3	94	245,019,529
Chr2	Chr1, 3, 5, 7	Chr1, 2, 6	14	156,023,568
Chr3	Chr1, 4	Chr3, 5	11	115,571,927
Chr4	Chr1, 2, 5, 27	Chr1, 4, 6	39	117,170,416
Chr5	Chr1	Chr3	6	75,951,376
Chr6, 7, 8	Chr2, 6, 8, 9, 20	Chr1, 2, 3, 4	25	200,178,831
Chr9, 10	Chr11, 22, 24	Chr7, 8	8	69,008,218
Chr11, 12	Chr4, 10	Chr10, 11	6	52,491,606
Chr13	Chr1, 5, 23	Chr9	9	19,625,567
Chr14	Chr14	Chr12	3	21,537,982
Chr15	Chr15	ChrX	4	14,821,201
Chr16	Chr17	Chr16	2	13,367,238
Chr17, 18	Chr1, 19, 21	Chr1, 9, 15, 17	10	17,262,365
Chr19	Chr1, 3, 25	Chr14	6	11,765,548
ChrZ	Chr1, 28	Chr18	6	10,642,498

*G. gallus* homology: homology of scaffolds to *G. gallus* chromosomes. *A. carolinensis* homology: homology of scaffolds to *A. carolinensis* chromosomes. Total number of assigned scaffolds: total number of scaffolds assigned to each chromosome. Total length of assigned scaffolds (bp): size in base pairs of the sum of all scaffolds assigned to each chromosome.

genome (44.5%) but higher than the GC content of *A. carolinensis* (40.3%; Supplementary Table 1). Repetitive elements accounted for 32% of the genome, most of which were transposable elements (Supplementary Table 2). As repetitive elements account for 49.6% of the *S. crocodilurus* genome<sup>11</sup>, most of the difference in size between the Komodo dragon genome and that of its closest sequenced relative can be attributed to repetitive element content.

**Chromosome scaffold content.** We isolated chromosome-specific DNA pools from a female Komodo dragon embryo from Prague zoo stock through flow sorting<sup>10</sup> and performed Illumina short-read sequencing on 15 DNA pools containing all Komodo dragon chromosomes (VKO1-20, VKOZ, VKOW; Supplementary Table 4). For each chromosome, we determined scaffold content and homology to *A. carolinensis* and chicken *Gallus gallus* chromosomes (Table 2 and Supplementary Tables 5 and 6). For pools where chromosomes were mixed, we determined partial scaffold content of single chromosomes. A total of 243 scaffolds containing 1.14 Gb (75% of total 1.51 Gb assembly) were assigned to 20 Komodo dragon chromosomes. As sex chromosomes share homologous pseudoautosomal regions, scaffolds enriched in both mixed 17/18/Z and 11/12/W chromosome pools most likely contained sex chromosome regions. As male varanid lizards are homogametic (ZZ) and the embryo used for flow sorting was female (ZW), scaffolds from the male-derived assembly enriched in these pools were assigned to the Z chromosome. Scaffold 79, which was assigned to the Z chromosome, contains an orthologue of the anti-Müllerian hormone (*amh*) gene, which plays a crucial role in testis differentiation in vertebrates<sup>17</sup>. Scaffolds assigned to the Z chromosome were homologous to



**Fig. 1 | Estimated species phylogeny of 15 non-avian reptile species, 3 avian species and 4 mammals.** Maximum likelihood phylogeny constructed from 1,394 one-to-one orthologous proteins. Support values from 10,000 bootstrap replicates are shown. All silhouettes reproduced from PhyloPic. Credits: python silhouette, V. Deepak under a Creative Commons licence [CC BY 3.0](https://creativecommons.org/licenses/by/3.0/); lizard silhouette, Ghedo and T. Michael Keesey under a Creative Commons licence [CC BY-SA 3.0](https://creativecommons.org/licenses/by-sa/3.0/). Photograph of Slasher, a Komodo dragon sampled for DNA in this study. Credit: photo courtesy of Adam Thompson/Zoo Atlanta.

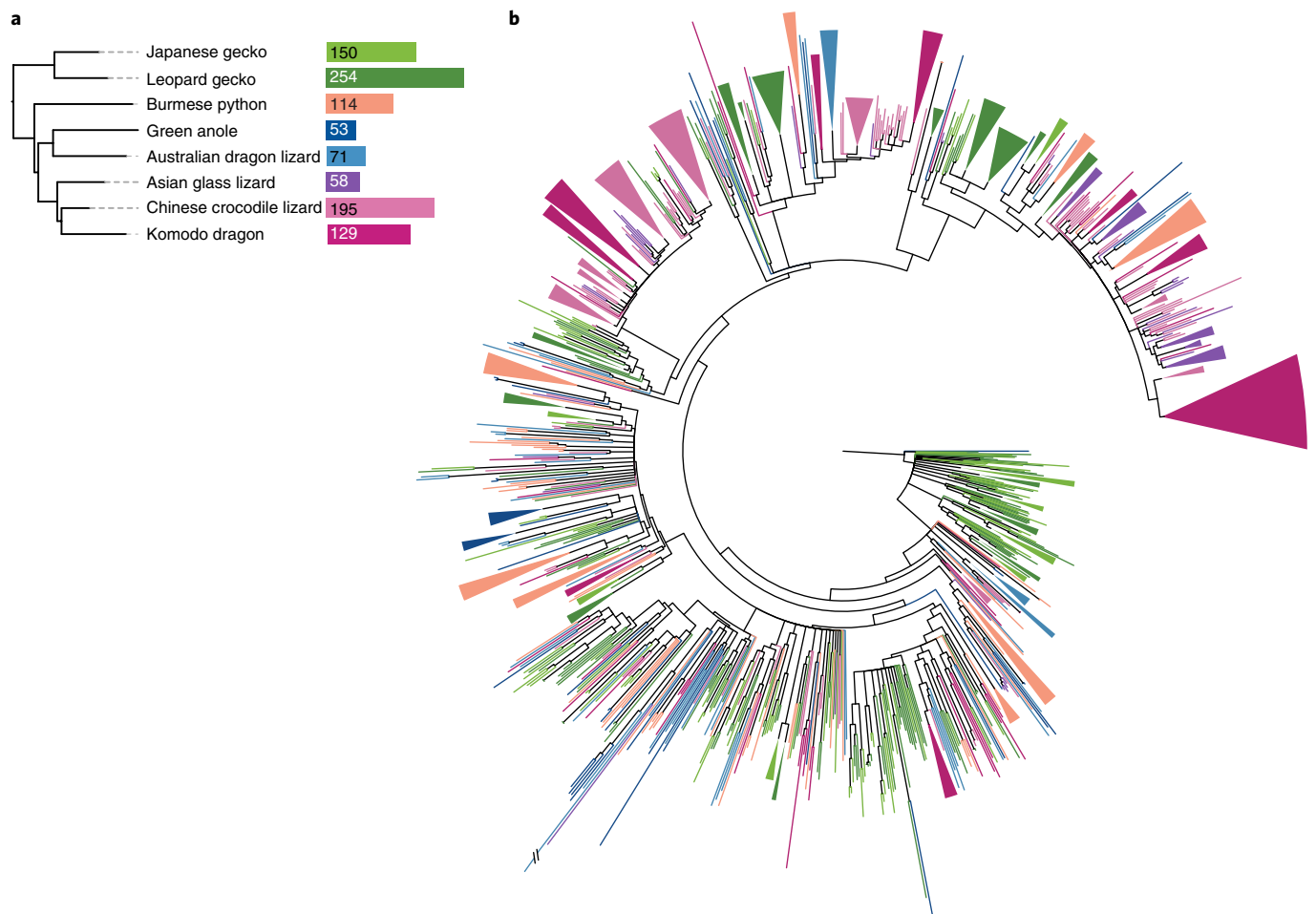
*A. carolinensis* chromosome 18, and mostly to *G. gallus* chromosome 28, in agreement with recent transcriptome analysis<sup>18</sup>.

**Gene annotation.** To annotate genes in the Komodo dragon genome, we performed RNA sequencing (RNA-seq) of heart tissue, and then used the MAKER pipeline with assembled RNA-seq transcripts, protein homology and de novo predictions as evidence (Methods). A total of 18,457 protein-coding genes were annotated in the Komodo genome, 17,189 (93%) of which have at least one annotated Interpro functional domain (Table 1). Of these protein-coding genes, 63% were expressed (reads per kilobase of transcript per million mapped reads > 1) in the heart. Most (89%) Komodo dragon protein-coding genes are orthologous to *A. carolinensis* genes. The median amino acid identity of single-copy orthologues between Komodo dragon and *A. carolinensis* is 68.9%, whereas it is 70.6% between one-to-one orthologues in Komodo dragon and *S. crocodilurus* (Supplementary Fig. 1).

**Phylogenetic placement of Komodo dragon.** Recent analyses estimate that varanid lizards and their closest extant relative, the earless monitor lizard of the *Lanthanotus* genus, diverged 62 Myr ago, and that varanid lizards and the *Shinisauridae* family diverged 115 Myr ago<sup>19–38</sup>. We used 1,394 orthologous proteins from the Komodo

dragon genome, 14 representative non-avian reptile species (seven squamates, three turtles and four crocodylians), three avian species (chicken, wild turkey and zebra finch) and four mammalian species (platypus, mouse, dog and human) to estimate a species tree (Fig. 1). Our analysis supports a sister relationship between anguimorphs (monitor lizards, anguids, Chinese crocodile lizards and relatives) and iguanians (dragon lizards, chameleons and iguanas), with snakes as sister to these two groups. This is in agreement with previously published analyses, including the most comprehensive marker gene-based molecular phylogenetic analyses<sup>39–41</sup>, and in disagreement with a proposed sister relationship between anguimorphs and snakes or other topologies<sup>42,43</sup>.

**Expansion of vomeronasal genes across squamate reptiles.** The vomeronasal organ is a chemosensory tissue shared across most amphibians, reptiles and mammals that detects chemical cues including pheromones and kairomones. There are two classes of vomeronasal chemosensory receptors, both of which have undergone repeated gene family expansions and contractions across vertebrate evolution. The gene family encoding vomeronasal type 2 receptors (V2Rs) has expanded in amphibians, snakes and some mammalian lineages<sup>44,45</sup>. In contrast, crocodylian and turtle genomes contain very few V1R and V2R genes, and birds have entirely



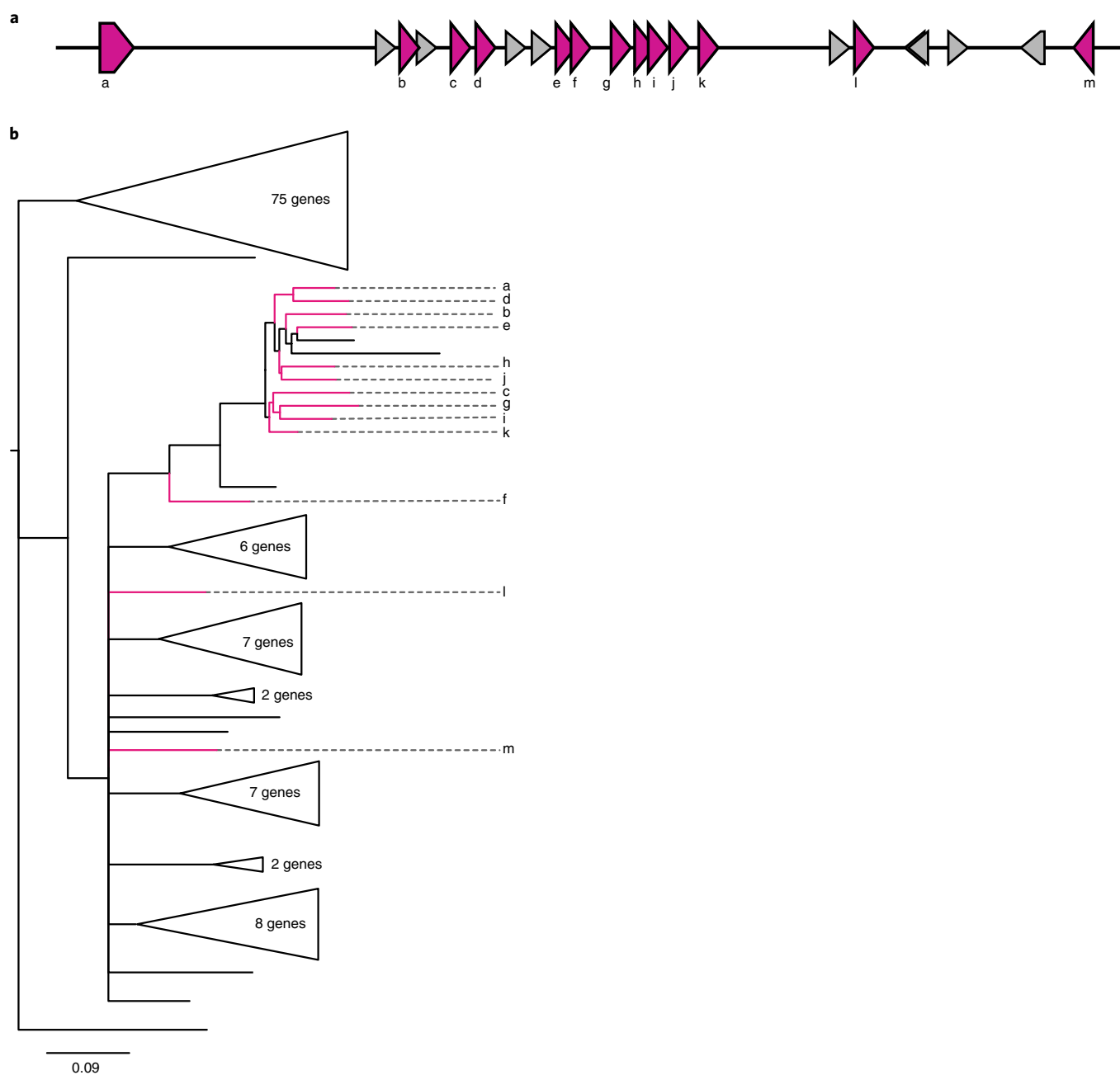
**Fig. 2 | V2Rs have expanded in Komodo dragons and several other squamate reptiles. a**, V2R gene counts in squamate reptiles. **b**, Unrooted gene phylogeny of 1,024 V2R transmembrane domains across squamate reptiles. The topology of the tree supports a gene expansion ancestral to squamates (that is, clades containing representatives from all species) as well as multiple species-specific expansions through gene duplication events (that is, clades containing multiple genes from one species). Branches with bootstrap support less than 60 are collapsed. Colours correspond to species in **a**. Clades containing genes from a single species are collapsed.

nonfunctional vomeronasal organs<sup>46,47</sup>. To clarify the relationship between vomeronasal organ function and the evolution of vomeronasal receptor gene families, we analysed the coding sequences of 15 reptiles, including the Komodo dragon, for presence of V1R and V2R genes (Fig. 2a). We found a large repertoire of V2Rs, comparable to that of snakes, in the Komodo dragon, other anguimorph lizard and geckos. We confirmed that there are few V1R genes across reptiles generally, and few to zero V2R genes in crocodylians and turtles (Supplementary Table 7). The low number of V2R genes in *A. carolinensis* and the Australian dragon lizard (*Pogona vitticeps*) suggests that V2R genes are infrequently expanded in iguanians, though more iguanian genomes are needed to test this hypothesis.

We next constructed a phylogeny of all V2R gene sequences across squamates (Fig. 2b) to understand the dynamic evolution of this gene family. The topology of this phylogeny supports the hypothesis that V2Rs expanded in the common ancestor of squamates, as there are clades of gene sequences containing members from all species<sup>45</sup>. In addition, there are many well-supported single-species clades (that is, Komodo dragon only or Burmese python only) dispersed across the gene tree, consistent with multiple duplications of V2R genes later in squamate evolution, including in the Komodo dragon and gecko lineages (Fig. 2b).

Because V2Rs expanded in rodents through tandem gene duplications that produced clusters of paralogues<sup>48</sup>, we examined clustering of V2R genes in our Komodo dragon assembly to determine whether a similar mechanism was at play. Of 129 V2 genes, 77 are organized into 21 gene clusters ranging from 2 to 13 paralogues (Fig. 3a and Supplementary Table 8). A phylogeny of all Komodo dragon V2R genes (Fig. 3b) showed that the genes in the largest 13-gene cluster group together in a gene tree of Komodo dragon V2R genes (Fig. 3). Of the remaining 52 V2R genes, 35 are on scaffolds less than 100 kb in size, so our estimate of V2R clustering is a lower bound due to fragmentation in the genome assembly (Supplementary Table 8). These results support the hypothesis that expansions of V2R genes in multiple squamate reptile lineages arose through tandem gene duplication.

**Positive selection.** To evaluate adaptive protein evolution in the Komodo dragon genome, we tested for positive selection across one-to-one orthologues in squamate reptiles using a branch-site model (Supplementary Table 9). Our analysis revealed 201 genes with signatures of positive selection in Komodo dragons (Supplementary Table 10). Of these, 188 had a one-to-one orthologue in humans, 93 mapped to pathways in the Reactome database and 34 had an annotated functional interaction with at least one other positively selected gene (Supplementary Fig. 2)<sup>49,50</sup>. These 34 genes are enriched for



**Fig. 3 | Gene clusters of V2Rs evolved through gene duplication. a**, Cluster of 13 V2R genes in the Komodo dragon genome. Pink genes are V2R genes and grey genes are non-V2R genes. Gene labels correspond to labels in **b**. **b**, Unrooted phylogeny of 129 V2R genes in the Komodo dragon. As most of the genes in this gene cluster group together in a gene phylogeny of all Komodo dragon V2R genes, it is likely that this cluster evolved through gene duplication events. Branches with bootstrap support less than 80 are collapsed. Clades without genes in this V2R gene cluster are collapsed. Genes in the V2R cluster are coloured pink and labelled as in **a**.

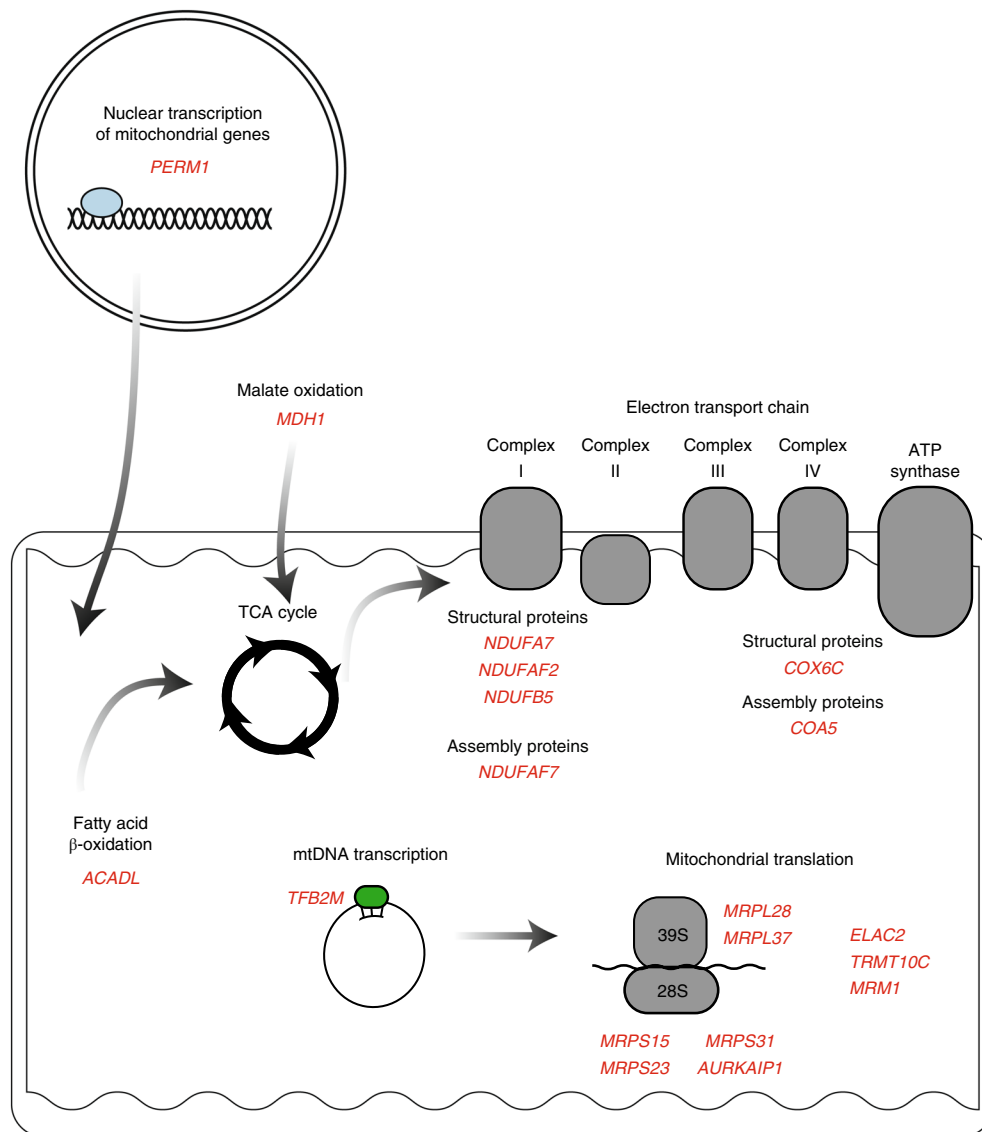
12 pathways (false discovery rate <5%), including three related to mitochondrial function, four related to coagulation and five related to immune function (Supplementary Table 11).

Many of the genes under positive selection point towards important adaptations of the Komodo dragon's mammalian-like cardiovascular and metabolic functions, which are unique among non-varanid ectothermic reptiles. These include mitochondrial function and cellular respiration, haemostasis and the coagulation cascade, and angiotensinogen (Supplementary Table 11 and Supplementary Fig. 2)<sup>51</sup>. Innate and adaptive immunity genes, which are frequently under positive selection in vertebrates, are well represented among positively selected genes<sup>52</sup>. Finally, 106 positively selected genes do not

have an annotated function and 25% of positively selected genes were not detectably expressed in the heart and likely represent adaptations in other aspects of Komodo dragon biology.

**Positive selection of genes regulating mitochondrial function.** In the Komodo dragon genome, we found evidence of positive selection of electron transport chain components including multiple subunits and assembly factors of the type 1 NADH dehydrogenase—*NDUFA7*, *NDUFA7*, *NDUFA2*, *NDUFB5*—as well as components of the cytochrome *c* oxidase protein complexes, *COX6C* and *COA5* (Fig. 4, Supplementary Fig. 3 and Supplementary Table 10). We also found signatures of positive





**Fig. 4 | Positive selection of mitochondrial genes in the Komodo dragon.** Genes in the Komodo dragon genome under positive selection include components of the electron transport chain, regulators of mtDNA transcription, regulators of mitochondrial translation and fatty acid  $\beta$ -oxidation. TCA, tricarboxylic acid.

selection for other elements of mitochondrial function in the Komodo dragon lineage (Fig. 4). For example, we detected positive selection for *ACADL*, which encodes a critical enzyme for mitochondrial fatty acid  $\beta$ -oxidation, the major postnatal metabolic process in cardiac myocytes<sup>53</sup>. Furthermore, two genes that promote mitochondrial biogenesis, *TFB2M* and *PERM1*, have undergone positive selection in the Komodo dragon. *TFB2M* regulates mitochondrial DNA (mtDNA) transcription and dimethylates mitochondrial 12S ribosomal RNA<sup>54,55</sup>, while *PERM1* regulates the expression of select *PPARGC1A/B* and *ESRRA/B/G* target genes with roles in glucose and lipid metabolism, energy transfer, contractile function, muscle mitochondrial biogenesis and oxidative capacity<sup>56</sup>. *PERM1* also enhances mitochondrial biogenesis, oxidative capacity and fatigue resistance when overexpressed in mice<sup>57</sup>. Finally, we observed positive selection of *MDH1*, encoding malate dehydrogenase, which together with mitochondrial *MDH2* regulates the supply of NADPH and acetyl coenzyme A to the cytoplasm, thus modulating fatty acid synthesis<sup>58</sup>.

Multiple factors regulating mitochondrial translation have also undergone positive selection in the Komodo dragon (Fig. 4).

These include four components of the mitochondrial 28S small ribosomal subunit (*MRPS15*, *MRPS23*, *MRPS31* and *AURKAIP1*) and two components of the mitochondrial 39S large ribosomal subunit (*MRPL28* and *MRPL37*). We also found positive selection of *ELAC2* and *TRMT10C*, which are required for maturation of mitochondrial transfer RNA, and *MRM1*, which encodes a mitochondrial rRNA methyltransferase<sup>59–61</sup>.

Overall, these instances of positive selection in genes encoding proteins important for mitochondrial function could underlie the remarkably high aerobic capacity in the Komodo dragon. Additional genome sequences are needed to determine whether these changes are specific to the Komodo dragon, shared across varanid lizards generally, or found in unsequenced reptiles.

**Positive selection of angiotensinogen.** We detected positive selection for angiotensinogen (*AGT*), which encodes the precursor of several peptide regulators of cardiovascular function, the most well-studied being angiotensin II (AngII) and angiotensin 1-7 (A1-7). AngII has a vasoactive function in blood vessels and inotropic effects on the heart<sup>62</sup>. In mammals, the level of AngII increases in response

to intense physical activity, contributing to arterial blood pressure and regional blood regulation<sup>63–66</sup>. Reptiles have a functional renin–angiotensin system (RAS) that is important for their cardiovascular response to aerobic activity<sup>67–69</sup>. The positive selection for *AGT* points to important adaptations in cardiovascular physiology and the renin–angiotensin system in the Komodo dragon.

**Positive selection of haemostasis-related genes.** We found evidence for positive selection across regulators of haemostasis, which reduces blood loss after injury. Four genes that regulate platelet activities, *MRVII*, *RASGRP1*, *LCP2* and *CD63*, have undergone positive selection in the Komodo dragon genome. *MRVII* is involved in inhibiting platelet aggregation<sup>70</sup>, *RASGRP1* coordinates calcium-dependent platelet responses<sup>71</sup>, *LCP2* is involved in platelet activation<sup>72</sup> and *CD63* has a role in controlling platelet spreading<sup>73</sup>. In addition, two coagulation factors, *F10* (factor X) and *F13B* (coagulation factor XIII B chain) have undergone positive selection in the Komodo dragon genome. Activation of factor X is the first step in initiating coagulation<sup>74</sup> and factor 13 is the final factor activated in the coagulation cascade<sup>75</sup>. Further, *FGB*, which encodes one of the three subunits of fibrinogen, the molecule converted to the clotting agent fibrin<sup>76</sup>, has undergone positive selection in the Komodo dragon genome.

## Discussion

We have sequenced and assembled a high-quality genome of the Komodo dragon that will be a template for analysis of other varanid genomes, and for further investigation of genomic innovations in the varanid lineage. We were able to assign 75% of the genome to chromosomes, providing a significant contribution to comparative genomics of squamates and vertebrates generally. As the number of squamate whole-genome sequences continues to grow, there will be opportunities to examine the evolution of noncoding DNA in these reptiles.

Varanid lizards have genotypic sex determination and share ZZ/ZW sex chromosomes with other anguimorph lizards<sup>10,18</sup>. Here, we were able to assign genomic scaffolds to the Z chromosome of the Komodo dragon. All Z chromosome scaffolds were homologous to *A. carolinensis* chromosome 18 and mostly to chicken chromosome 28, in agreement with a recent transcriptome-based analysis<sup>18</sup>. Within Iguania, the sister group of anguimorphs<sup>20,30,40</sup>, there are environmental sex-determination systems without sex chromosomes as well as conserved XX/XY sex chromosomes homologous to anguimorph autosomes<sup>77–81</sup>. Sex chromosomes in most snakes (pythons and all families of caenophidian snakes<sup>82</sup>) are homologous to chromosome 6 of *A. carolinensis* and thus to autosomes of the Komodo dragon, suggesting an independent origin of sex chromosomes in snakes and anguimorphs. However, the ancestral sex determination of snakes remains unresolved<sup>82,83</sup>. The regions of sex chromosomes shared by the common ancestor of varanids and several other lineages of anguimorph lizards contain the *amh* gene<sup>18</sup>, which has a crucial role in vertebrate testis differentiation. Homologues of *amh* are strong candidates for sex-determining genes in several lineages of teleost fishes and in monotremes<sup>84–87</sup>, and should be considered candidate sex-determining genes in varanids and other anguimorphs.

Our comparative genomic analysis identified previously undescribed species-specific expansion of V2Rs across multiple squamates, including lizards and at least one snake. Komodo dragons, like other squamates, are known to have a sophisticated lingual–vomeronasal system for chemical sampling of their environment<sup>88</sup>. This sensory apparatus allows Komodo dragons to perceive environmental chemicals for social and ecological activities, including kin recognition, mate choice<sup>89,90</sup>, predator avoidance<sup>91,92</sup> and hunting prey<sup>93,94</sup>. Komodo dragons are unusual as they adopt differential foraging tactics across ontogeny, with smaller juveniles preferring

active foraging for small prey and large adult dragons targeting larger ungulate prey via ambush predation<sup>6</sup>. However, utilization of the vomeronasal system across ontogeny seems likely, given the exceptional capacity for Komodo dragons of all sizes to locate prey. Future work will be able to explore the role of V2R expansion in the behaviour and ecology of Komodo dragons, including their ability to locate prey at long distances<sup>4</sup>.

We found evidence for positive selection in the Komodo dragon genome across genes involved in regulating mitochondrial biogenesis, cellular respiration and cardiovascular homeostasis. Komodo dragons and other monitor lizards have a high aerobic capacity and exercise endurance, and our results reveal selective pressures on biochemical pathways that are likely to be the source of this high aerobic capacity. Reptile muscle mitochondria typically oxidize substrates at a much lower rate than mammalian mitochondria, partly based on substrate-type use<sup>95</sup>. The findings that Komodo dragons experienced selection in several genes encoding mitochondrial enzymes, including one involved in fatty acid metabolism, points towards a more mammalian-like mitochondrial function. Future work on additional varanid species, and other squamate outgroups, will test these hypotheses. Selective pressures acting on these mitochondrial genes in Komodo dragons is consistent with the increased expression of genes associated with oxidative capacity found in pythons after feeding<sup>96,97</sup>.

In addition, we found positive selection for angiotensinogen, which encodes two potent vasoactive and inotropic peptides with central roles in cardiovascular physiology. In mammals, AngII contributes to the mean arterial blood pressure and to the redistribution of cardiac output<sup>65,66</sup>. A compelling hypothesis is that these changes to angiotensinogen may be an important component in the ability of the Komodo dragon to rapidly increase blood pressure and cardiac output as required for hunting, extended periods of locomotion including inter-island swimming, and male–male combat during the breeding season. Direct measures of cardiac function have not been made in Komodo dragons, but in other varanid lizards, a large aerobic scope during exercise is associated with a large factorial increase in cardiac output<sup>98</sup>. Future physiological studies measuring the hemodynamic responses to exercises with respect to AngII expression can test this hypothesis. Giraffes, which have evolved high blood pressure to maintain cardiovascular homeostasis in their elongated bodies, have experienced positive selection on several blood pressure regulators, including the angiotensin-converting enzyme<sup>99</sup>. It is possible that positive selection in animals with high blood pressures converges on angiotensin regulators. Overall, the evolution of these genes suggests a profoundly different cardiovascular and metabolic profile relative to other squamates, endowing the Komodo dragon with unique physiological properties.

We also found evidence for positive selection across genes that regulate blood clotting. Like other monitor lizards, the saliva of Komodo dragons contains anticoagulants, which is thought to aid in hunting<sup>7,8</sup>. During conflict with conspecifics over food, territories or mates, Komodo dragons use their serrate teeth to inflict bite wounds, raising the possibility that these anticoagulants may enter their bloodstream. The extensive positive selection of genes encoding their coagulation system may reflect a selective pressure for Komodo dragons to evade the anticoagulant and hypotensive effects of the saliva of conspecifics. While all monitor lizards tested contain anticoagulants in their saliva, the precise mechanism by which they act varies<sup>8</sup>. It is possible that different species of monitor lizards evolved adaptations that reflect the diversity of their anticoagulants, or that coevolution has occurred between monitor lizard coagulation systems and anticoagulant saliva. Further, as Komodo dragons have high blood pressure, changes to their coagulation system may reflect increased protection from vascular damage.

## Methods

**DNA isolation and processing for Bionano optical mapping.** Komodo dragon whole blood was obtained from one of two individuals housed at Zoo Atlanta (Rinca). Samples from the animals at Zoo Atlanta were collected with the approval of the Zoo's Scientific Research Committee. High-molecular-weight genomic DNA was extracted for genome mapping. Blood was centrifuged at 2,000 g for 2 min, plasma was removed and the sample was stored at 4°C. Then, 2.5 µl of blood was embedded in 100 µl agarose gel plugs to give ~7 µg DNA per plug, using the Bio-Rad CHEF Mammalian Genomic DNA Plug Kit (Bio-Rad Laboratories). Plugs were treated with proteinase K overnight at 50°C. The plugs were then washed, melted and solubilized with GELase (Epicentre). The purified DNA was subjected to 4 h of drop dialysis. DNA concentration was determined using Qubit 2.0 Fluorometer (Life Technologies), and the quality was assessed with pulsed-field gel electrophoresis.

The high-molecular-weight DNA was labelled according to commercial protocols using the IrysPrep Reagent Kit (Bionano Genomics). Specifically, 300 ng of purified genomic DNA was nicked with 7 U nicking endonuclease Nb.BbvCI (NEB) at 37°C for 2 h in NEB Buffer 2. The nicked DNA was labelled with a fluorescent-dUTP nucleotide analogue using Taq polymerase (NEB) for 1 h at 72°C. After labelling, the nicks were repaired with Taq ligase (NEB) in the presence of dNTPs. The backbone of fluorescently labelled DNA was stained with DNA stain (Bionano).

**Bionano mapping and assembly.** Using the Bionano Irys instrument, automated electrophoresis of the labelled DNA occurred in the nanochannel array of an IrysChip (Bionano Genomics), followed by automated imaging of the linearized DNA. The DNA backbone (outlined by YOYO-1 staining) and locations of fluorescent labels along each molecule were detected using the Irys instrument's software. The length and set of label locations for each DNA molecule defines an individual single-molecule map. Raw Bionano single-molecule maps were de novo assembled into consensus maps using the Bionano IrysSolve assembly pipeline (version 5134) with default settings, with noise values calculated from the 10x Genomics Supernova assembly. The Bionano optical mapping data comprised 80x genome coverage, and the scaffold N50 of the assembly was 1.2 Mb.

**DNA processing for 10x Genomics linked-read sequencing.** Blood samples from two individuals housed at Zoo Atlanta (Slasher and Rinca) were used. High-molecular-weight genomic DNA extraction, sample indexing and generation of partition barcoded libraries were performed by 10x Genomics according to the Chromium Genome User Guide and as published previously<sup>100</sup>. Then 1.2 ng of genomic DNA was used as input to the Chromium system.

**10x Genomics sequencing and assembly.** The 10x Genomics barcoded library was sequenced on the Illumina HiSeq2500. A total of 660 million of the raw reads comprising 57x genome coverage were assembled using the company's Supernova software (version 1.0) with default parameters. Output fasta files of the Supernova assemblies were generated in pseudohap format, which links phased and unphased regions of the assembly into 'pseudo-haplotype' scaffolds. This generated an initial assembly with a scaffold N50 length of 10.2 Mb and a contig N50 length of 95 kb.

**Oxford Nanopore sequencing.** DNA isolated from Slasher was sequenced to 0.75x coverage on an Oxford Nanopore MinIon sequencer following the manufacturer's instructions. MinKNOW was used for basecalling and output to FASTQ files.

**DNA processing for PacBio sequencing.** Komodo dragon whole blood collected in EDTA from an individual housed at Reptilandia zoo, Gran Canaria under institutional approval, stored at -20°C, was used to extract high-molecular-weight DNA for single-molecule real-time sequencing. Extraction was performed using gravity-flow, anion-exchange tips (Qiagen genomic-tip 100/G kit) to a final DNA concentration of 130 ng µl<sup>-1</sup> assessed using a Qubit 2.0 Fluorometer. Size of extracted DNA was determined by a 16 h pulse-field gel electrophoresis, which resolved high-molecular-weight fragments from 15 kb to 85 kb. We constructed a 10 kb and a 20 kb insert library using 20 µg of genomic DNA. The library was then size selected using a Blue Pippin (Sage Science) and resulting double-stranded DNA fragments were capped by hairpin loops at both ends to form a SMRTbell template. Single-molecule SMRTbell templates were then loaded in 150,000 Zero Mode Waveguides SMRT cells of a PacBio RS II sequencing system using paramagnetic beads (MagBeads, PacBio). Sequencing was performed using a total of 29 SMRT cells. We obtained a total of 2,061,804 subread filtered sequences for a total sequence length of 11,907,672,561 bp. Average, maximum and minimum sequence lengths were 5,775 bp, 48,338 bp and 35 bp, respectively. Median sequence length was 4,486 bp. N50 read length was 12,457 bp. In total, PacBio sequencing data represented 6.3x genome sequencing coverage.

**Merging datasets into a single assembly.** Sequencing and mapping data types were merged together as follows. First, Bionano assembled contigs and the 10x Genomics assembly were combined using Bionano's hybrid assembly tool with the -B2 -N2 options. SSPACE-LongRead (<https://doi.org/10.1186/1471-2105-15-211>) was used in series with default parameters to scaffold the hybrid assembly using PacBio reads, Nanopore reads, and unincorporated 10x Genomics Supernova scaffolds/contigs, resulting in the final assembly.

**Genome completeness assessment.** The BUSCO pipeline version 3.0.2 was used to determine the completeness of the Komodo dragon genome, using the 2,586 gene vertebrata gene set and the Augustus retraining parameters (-long)<sup>16</sup>. For comparison, BUSCO was also run with the same parameters on all reptile genomes used for comparative analyses (*S. crocodilurus*, *Ophisaurus gracilis*, *A. carolinensis*, *P. vitticeps*, *Python molurus bivittatus*, *Eublepharis macularius*, *Gekko japonicus*, *Pelodiscus sinensis*, *Chelonia mydas*, *Chrysemys picta bellii*, *Alligator sinensis*, *Alligator mississippiensis*, *Gavialis gangeticus* and *Crocodylus porosus*) (Supplementary Table 3).

We obtained an assembly-free estimate of the Komodo dragon genome size using an sequencing error corrected *k*-mer counting method implemented in the PreQC component of the SGA assembler<sup>13</sup>.

**Assignment of scaffolds to chromosomes.** Isolation of Komodo dragon chromosome-specific DNA pools was performed as previously described<sup>10</sup>. Briefly, fibroblast cultivation of a female *V. komodoensis* were obtained from tissue samples of an early embryo of a captive individual at the Prague National Zoo. Embryos are obtained under the laws of the Czech Republic and of the European Union. Chromosomes obtained by fibroblast cultivation were sorted using a Mo-Flo (Beckman Coulter) cell sorter. Fifteen chromosome pools were sorted in total. Chromosome-specific DNA pools were then amplified and labelled by degenerate oligonucleotide primed PCR and assigned to their respective chromosomes by hybridization of labelled probes to metaphases. *V. komodoensis* chromosome pools obtained by flow sorting were named according to chromosomes (for example majority of DNA of VKO6/7 belong to chromosomes 6 and 7 of *V. komodoensis*). *V. komodoensis* pools for macrochromosomes are each specific for one single pair of chromosomes, except for VKO6/7 and VKO8/7, which contain one specific chromosome pair each (pair 6 and pair 8, respectively), plus a third pair which overlaps between the two of them (pair 7). For microchromosomes, pools VKO9/10, VKO17/18/19, VKO11/12/W and VKO17/18/Z contained more than one chromosome each, while the rest are specific for one single pair of microchromosomes. The W and Z chromosomes are contained in pools VKO11/12/W and VKO17/18/Z, respectively, together with two pairs of other microchromosomes each.

Chromosome-pool-specific genetic material was amplified by GenomePlex Whole Genome Amplification Kit (Sigma) following manufacturer protocols. DNA from all 15 chromosome pools was used to prepare Illumina sequencing libraries, which were independently barcoded and sequenced 125 bp paired-end in a single Illumina HiSeq2500 lane.

Reads obtained from sequencing of flow-sorting-derived chromosome-specific DNA pools were processed with the dopseq pipeline (<https://github.com/lca-imcb/dopseq>)<sup>101,102</sup>. Illumina adaptors and whole-genome amplification primers were trimmed off by cutadapt v1.13<sup>103</sup>. Then, pairs of reads were aligned to the genome assembly of *V. komodoensis* using bwa mem<sup>104</sup>. Reads were filtered by mapping quality  $\geq 20$  and length  $\geq 20$  bp, and aligned reads were merged into positions using pybedtools 0.7.10<sup>105,106</sup>. Reference genome regions were assigned to specific chromosomes based on distance between positions. Finally, several statistics were calculated for each scaffold. Calculated parameters included: mean pairwise distance between positions on scaffold, mean number of reads per position on scaffold, number of positions on scaffold, position representation ratio (PRR) and *P* value of PRR. PRR of each scaffold was used to evaluate enrichment of given scaffold on chromosomes. PRR was calculated as ratio of positions on scaffold to positions in genome divided by ratio of scaffold length to genome length. PRRs >1 correspond to enrichment, while PRRs <1 correspond to depletion. As the PRR value is distributed lognormally, we use its logarithmic form for our calculations. To filter out only statistically significant PRR values, we used thresholds of logPRR >0 and its *P* value  $\leq 0.01$ . Scaffolds with logPRR >0 were considered enriched in the given sample. If one scaffold was enriched in several samples we chose the highest PRR to assign scaffold as top sample.

We also compared the genome organization at the chromosome level among *V. komodoensis*, *A. carolinensis* and *G. gallus*. We determined homology of each *V. komodoensis* scaffold to scaffolds of *A. carolinensis* (AnoCar2.0) and *G. gallus* (galGal3) genome generating alignment between genomes with LAST<sup>107</sup> and subsequently using chaining and netting technique<sup>108</sup>. Homology to *A. carolinensis* microchromosomes was determined using scaffold assignments from an *Anolis* chromosome-specific sequencing project<sup>101</sup>. For LAST, we used default scoring matrix and parameters of 400 for gap existence cost, 30 for gap extension cost and 4,500 for minimum alignment score. For actChain we used same distance matrix and default parameters for other chain-net scripts.

**RNA sequencing.** RNA was extracted from heart tissue obtained from an adult male specimen that died of natural causes at the San Diego Zoo. This was approved by the Institutional Animal Care and Use Committee and Biomaterials Review Group of San Diego Zoo. Trizol reagent was used to extract RNA following the manufacturer's instructions. Two RNA-seq libraries were produced using a NuGen RNAseq v2 and Ultralow v2 kits from 100 ng total RNA each, and sequenced on an Illumina Nextseq 500 with 150 bp paired-end strand-specific reads.

**Genome annotation.** RepeatMasker v4.0.7 was used to mask repetitive elements in the Komodo dragon genome using the Squamata repeat database from the



RepBase-Dfam combined database as reference<sup>109</sup>. After masking repetitive elements, protein-coding genes were annotated using the MAKER version 3.01.02<sup>110</sup> pipeline, combining protein homology information, assembled transcript evidence, and de novo gene predictions from SNAP and Augustus version 3.3.1<sup>111</sup>. Protein homology was determined by aligning proteins from 15 reptile species (Supplementary Table 1) to the Komodo dragon genome using exonerate version 2.2.0<sup>112</sup>. RNA-seq data were aligned to the Komodo dragon genome with STAR version 2.6.0<sup>113</sup> and assembled into 900,722 transcripts with Trinity version 2.4.0<sup>114</sup>. Protein domains were determined using InterProScan version 5.31.70<sup>115</sup>. Gene annotations from the MAKER pipeline were filtered based on the strength of evidence for each gene, the length of the predicted protein and the presence of protein domains. Clusters of orthologous genes across 15 reptile species were determined with OrthoFinder v2.0.0<sup>116</sup>. A total of 284,107 proteins were clustered into 16,546 orthologous clusters. In total, 96.4% of Komodo dragon genes were grouped into orthologous clusters. For estimating a species phylogeny only, protein sequences from *Mus musculus* and *G. gallus* were added to the orthologous clusters with OrthoFinder. Transfer RNAs were annotated using tRNAscan-SE version 1.3.1<sup>117</sup>, and other noncoding RNAs were annotated using the Rfam database<sup>118</sup> and the Infernal software suite<sup>119</sup>.

**Phylogenetic analysis.** A total of 1,394 one-to-one orthologous proteins across 15 reptile species, three birds and four mammals were used to estimate a species phylogeny. The 15 reptile species used were *V. komodoensis*, *S. crocodilurus*, *O. gracilis*, *A. carolinensis*, *P. vitticeps*, *P. molurus bivittatus*, *E. macularius*, *G. japonicus*, *P. sinensis*, *C. mydas*, *C. picta bellii*, *A. sinensis*, *A. mississippiensis*, *G. gangeticus* and *C. porosus*. The three birds used were *G. gallus*, *Meleagris gallopavo* and *Taeniopygia guttata*, and the four mammals were *Ornithorhynchus anatinus*, *M. musculus*, *Canis familiaris* and *Homo sapiens*. Each set of orthologous proteins were aligned using PRANK v.170427<sup>120</sup>. Aligned proteins were concatenated into a supermatrix, and a species tree was estimated using IQ-TREE version 1.6.7.1<sup>121</sup> with model selection across each partition<sup>122</sup> and 10,000 ultrafast bootstrap replicates<sup>123</sup>.

To identify genes in this dataset of 1,394 orthologues that are evolving in a clock-like manner and thus useful for phylogenomic analyses such as divergence time estimates, we created individual trees for each of the genes in our analysis using the procedure described above but without bootstrapping. We then used SortaDate<sup>124</sup> to calculate a number of informative metrics, including root-to-tip variance, tree length and bipartition support relative to the species phylogeny. This dataset can be used to find candidate marker genes for phylogenomic analyses and is available in a Figshare repository at <https://doi.org/10.6084/m9.figshare.7967300>.

**Gene family evolution analysis.** Gene family expansion and contraction analyses were performed with CAFE v4.2<sup>125</sup> for the squamate reptile lineage, with a constant gene birth and gene death rate assumed across all branches.

V2Rs were first identified in all species by containing the V2R domain InterPro domain (IPR004073)<sup>126</sup>. To ensure that no V2R genes were missed, all proteins were aligned against a set of representative V2R genes using BLASTp<sup>127</sup> with an *e*-value cutoff of  $1 \times 10^{-6}$  and a bitscore cutoff of 200 or greater. Any genes passing this threshold were added to the set of putative V2R genes. Transmembrane domains were identified in each putative V2R gene with TMHMM v2.0<sup>128</sup> and discarded if they did not contain seven transmembrane domains in the C-terminal region. Beginning at the start of the first transmembrane domain, proteins were aligned with MAFFT v7.310 (auto-alignment strategy)<sup>129</sup> and trimmed with trimAL (gappyout)<sup>130</sup>. A gene tree was constructed using IQ-TREE<sup>121–123</sup> with the JTT+ model of evolution with empirical base frequencies and 10 FreeRate model parameters, and 10,000 bootstrap replicates. Genes were discarded if they failed the IQ-TREE composition test.

**Positive selection analysis.** We analysed 4,047 genes that were universal and one-to-one across all squamate lineages tested (*V. komodoensis*, *S. crocodilurus*, *O. gracilis*, *A. carolinensis*, *P. vitticeps*, *P. molurus bivittatus*, *E. macularius* and *G. japonicus*) to test for positive selection (Supplementary Table 9). An additional 2,013 genes that were universal and one-to-one across a subset of squamate species (*V. komodoensis*, *A. carolinensis*, *P. molurus bivittatus* and *G. japonicus*) were also analysed (Supplementary Table 9). We excluded multicopy genes from all positive selection analyses to avoid confounding from incorrect paralogy inference. Proteins were aligned using PRANK<sup>120</sup> and codon alignments were generated using PAL2NAL<sup>131</sup>.

Positive selection analyses were performed with the branch-site model aBSREL using the HYPHY framework<sup>132,133</sup>. For the 4,081 genes that were single-copy across all squamate lineages, the full species phylogeny of squamates was used. For the 2,040 genes that were universal and single-copy across a subset of species, a pruned tree containing only those taxa was used. We discarded genes with unreasonably high dN/dS (defined as the ratio of the rate of nonsynonymous to synonymous substitutions) values across a small proportion of sites, as those were false positives driven by low-quality gene annotation in one or more taxa in the alignment. We used a cutoff of dN/dS of less than 50 across 5% or more of sites, and a *P* value of less than 0.05 at the Komodo dragon node. Each gene was first tested for positive selection only on the Komodo dragon branch. Genes undergoing positive selection

in the Komodo dragon lineage were then tested for positive selection at all nodes in the phylogeny. This resulted in 201 genes being under positive selection in the Komodo dragon lineage (Supplementary Table 10).

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

## Data availability

The assembled Komodo dragon genome is available in the National Center for Biotechnology Information (NCBI) under the accession SJPD00000000. All DNA sequencing used to generate the assembly is available in the NCBI Sequence Read Archive (SRA) database under accession PRJNA523222. Illumina sequencing data from chromosome pools is available in the NCBI SRA under accession PRJNA529483. RNA-seq data of heart tissue is available in the NCBI SRA under accession PRJNA527313. Original protein annotations, noncoding RNA annotations, all alignments for phylogenetic analyses and selection analyses, and newick files of phylogenetic trees are available in the following Figshare repositories: <https://doi.org/10.6084/m9.figshare.7961135.v1>, <https://doi.org/10.6084/m9.figshare.7955891.v1>, <https://doi.org/10.6084/m9.figshare.7955879.v1>, <https://doi.org/10.6084/m9.figshare.7949483.v1>, <https://doi.org/10.6084/m9.figshare.7759496.v1>, <https://doi.org/10.6084/m9.figshare.7967300>. The project folder for all Figshare data is available at [https://figshare.com/projects/Data\\_for\\_Komodo\\_dragon\\_genome\\_paper/61271](https://figshare.com/projects/Data_for_Komodo_dragon_genome_paper/61271).

Received: 21 February 2019; Accepted: 13 June 2019;

Published online: 29 July 2019

## References

- Chapman, A. D. *Numbers of Living Species in Australia and the World* (Australian Biological Resources Study, 2009).
- Collar, D. C., Schulte, J. A. & Losos, J. B. Evolution of extreme body size disparity in monitor lizards (*Varanus*). *Evolution* **65**, 2664–2680 (2011).
- Jensen, B., Wang, T., Christoffels, V. M. & Moorman, A. F. M. Evolution and development of the building plan of the vertebrate heart. *Biochim. Biophys. Acta Mol. Cell Res.* **1833**, 783–794 (2013).
- Auffenberg, W. *The Behavioral Ecology of the Komodo Monitor* (Univ. Presses of Florida, 1981).
- Green, B., King, D., Braysher, M. & Saim, A. Thermoregulation, water turnover and energetics of free-living komodo dragons, *Varanus komodoensis*. *Comp. Biochem. Physiol. A* **99**, 97–101 (1991).
- Purwandana, D. et al. Ecological allometries and niche use dynamics across Komodo dragon ontogeny. *Sci. Nat.* **103**, 27 (2016).
- Fry, B. G. et al. A central role for venom in predation by *Varanus komodoensis* (Komodo dragon) and the extinct giant *Varanus (Megalania) priscus*. *Proc. Natl Acad. Sci. USA* **106**, 8969–8974 (2009).
- Koludarov, I. et al. Enter the dragon: the dynamic and multifunctional evolution of anguimorpha lizard venoms. *Toxins* **9**, E242 (2017).
- Johnson Pokorná, M. et al. First description of the karyotype and sex chromosomes in the Komodo dragon (*Varanus komodoensis*). *Cytogenet. Genome Res.* **148**, 284–291 (2016).
- Iannucci, A. et al. Isolating chromosomes of the Komodo dragon: new tools for comparative mapping and sequence assembly. *Cytogenet. Genome Res.* **157**, 42–50 (2019).
- Gao, J. et al. Sequencing, de novo assembling, and annotating the genome of the endangered Chinese crocodile lizard *Shinisaurus crocodilurus*. *Gigascience* **6**, 1–6 (2017).
- Alföldi, J. et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011).
- Simpson, J. T. Exploring genome characteristics and sequence quality without a reference. *Bioinformatics* **30**, 1228–1235 (2014).
- Krishnan, A. et al. DNA index, genome size, and electronic nuclear volume of vertebrates from the Miami Metro Zoo. *Cytometry A* **65A**, 26–34 (2005).
- Doležel, J., Bartoš, J., Voglmayr, H. & Greilhuber, J. Letter to the editor. *Cytometry A* **51A**, 127–128 (2003).
- 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).
- Rey, R., Lukas-Croisier, C., Lasala, C. & Bedecarrás, P. AMH/MIS: what we know already about the gene, the protein and its regulation. *Mol. Cell. Endocrinol.* **211**, 21–31 (2003).
- Rovatsos, M., Reháč, I., Velenský, P. & Kratochvíl, L. Shared ancient sex chromosomes in varanids, beaded lizards and alligator lizards. *Mol. Biol. Evol.* **36**, 1113–1120 (2019).
- Welton, L. J., Travers, S. L., Siler, C. D. & Brown, R. M. Integrative taxonomy and phylogeny-based species delimitation of Philippine water monitor lizards (*Varanus salvator* Complex) with descriptions of two new cryptic species. *Zootaxa* **3881**, 201–227 (2014).

20. Zheng, Y. & Wiens, J. J. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol. Phylogenet. Evol.* **94**, 537–547 (2016).
21. Douglas, M. E., Douglas, M. R., Schuett, G. W., Beck, D. D. & Sullivan, B. K. Conservation phylogenetics of helodermatid lizards using multiple molecular markers and a supertree approach. *Mol. Phylogenet. Evol.* **55**, 153–167 (2010).
22. Castoe, T. A. et al. Dynamic nucleotide mutation gradients and control region usage in squamate reptile mitochondrial genomes. *Cytogenet. Genome Res.* **127**, 112–127 (2009).
23. Townsend, T. M. et al. Phylogeny of iguanian lizards inferred from 29 nuclear loci, and a comparison of concatenated and species-tree approaches for an ancient, rapid radiation. *Mol. Phylogenet. Evol.* **61**, 363–380 (2011).
24. Alfaro, M. E. et al. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc. Natl Acad. Sci. USA* **106**, 13410–13414 (2009).
25. Sanders, K. L. & Lee, M. S. Y. Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae, Colubroidea). *Mol. Phylogenet. Evol.* **46**, 1165–1173 (2008).
26. Okajima, Y. & Kumazawa, Y. Mitogenomic perspectives into iguanid phylogeny and biogeography: Gondwanan vicariance for the origin of Madagascan oplurines. *Gene* **441**, 28–35 (2009).
27. Kumazawa, Y. Mitochondrial genomes from major lizard families suggest their phylogenetic relationships and ancient radiations. *Gene* **388**, 19–26 (2007).
28. Hugall, A. F., Foster, R. & Lee, M. S. Y. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene *RAG-1*. *Syst. Biol.* **56**, 543–563 (2007).
29. Wiens, J. J., Brandley, M. C. & Reeder, T. W. Why does a trait evolve multiple times within a clade? Repeated evolution of snakelike body form in squamate reptiles. *Evolution* **60**, 123–141 (2006).
30. Pyron, R. A., Burbrink, F. T. & Wiens, J. J. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **13**, 93 (2013).
31. Zheng, Y. & Wiens, J. J. Do missing data influence the accuracy of divergence-time estimation with BEAST? *Mol. Phylogenet. Evol.* **85**, 41–49 (2015).
32. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
33. Hsiang, A. Y. et al. The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. *BMC Evol. Biol.* **15**, 87 (2015).
34. Tolley, K. A., Townsend, T. M. & Vences, M. Large-scale phylogeny of chameleons suggests African origins and Eocene diversification. *Proc. R. Soc. B* **280**, 20130184 (2013).
35. Jones, M. E. et al. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). *BMC Evol. Biol.* **13**, 208 (2013).
36. Portik, D. M. & Papenfuss, T. J. Monitors cross the Red Sea: the biogeographic history of *Varanus yemenensis*. *Mol. Phylogenet. Evol.* **62**, 561–565 (2012).
37. Pyron, R. A. A likelihood method for assessing molecular divergence time estimates and the placement of fossil calibrations. *Syst. Biol.* **59**, 185–194 (2010).
38. Vidal, N. et al. Molecular evidence for an Asian origin of monitor lizards followed by Tertiary dispersals to Africa and Australasia. *Biol. Lett.* **8**, 853–855 (2012).
39. Xiong, Z. et al. Draft genome of the leopard gecko, *Eublepharis macularius*. *Gigascience* **5**, 47 (2016).
40. Streicher, J. W. & Wiens, J. J. Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families. *Biol. Lett.* **13**, 20170393 (2017).
41. Wiens, J. J. et al. Combining phylogenomics and fossils in higher-level squamate reptile phylogeny: molecular data change the placement of fossil taxa. *Syst. Biol.* **59**, 674–688 (2010).
42. Fry, B. G. et al. Early evolution of the venom system in lizards and snakes. *Nature* **439**, 584–588 (2006).
43. Lee, M. S. Y. Hidden support from unpromising data sets strongly unites snakes with anguimorph 'lizards'. *J. Evol. Biol.* **22**, 1308–1316 (2009).
44. Silva, L. & Antunes, A. Vomeronasal receptors in vertebrates and the evolution of pheromone detection. *Annu. Rev. Anim. Biosci.* **5**, 353–370 (2017).
45. Brykczynska, U., Tzika, A. C., Rodriguez, I. & Milinkovitch, M. C. Contrasted evolution of the vomeronasal receptor repertoires in mammals and squamate reptiles. *Genome Biol. Evol.* **5**, 389–401 (2013).
46. Green, R. E. et al. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* **346**, 1254449 (2014).
47. Zippel, H. P. The ecology of vertebrate olfaction. *Behav. Process.* **7**, 198–199 (2002).
48. Yang, H., Shi, P., Zhang, Y. & Zhang, J. Composition and evolution of the V2r vomeronasal receptor gene repertoire in mice and rats. *Genomics* **86**, 306–315 (2005).
49. Fabregat, A. et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
50. Fabregat, A. et al. Reactome pathway analysis: a high-performance in-memory approach. *BMC Bioinformatics* **18**, 142 (2017).
51. Wu, G. & Haw, R. Functional interaction network construction and analysis for disease discovery. *Methods Mol. Biol.* **1558**, 235–253 (2017).
52. Shultz, A. J. & Sackton, T. Immune genes are hotspots of shared positive selection across birds and mammals. *Elife* **8**, e41815 (2019).
53. Riquelme, C. A. et al. Fatty acids identified in the Burmese python promote beneficial cardiac growth. *Science* **334**, 528–531 (2011).
54. Falkenberg, M. et al. Mitochondrial transcription factors B1 and B2 activate transcription of human mtDNA. *Nat. Genet.* **31**, 289–294 (2002).
55. Cotney, J., McKay, S. E. & Shadel, G. S. Elucidation of separate, but collaborative functions of the rRNA methyltransferase-related human mitochondrial transcription factors B1 and B2 in mitochondrial biogenesis reveals new insight into maternally inherited deafness. *Hum. Mol. Genet.* **18**, 2670–2682 (2009).
56. Cho, Y., Hazen, B. C., Russell, A. P. & Kralli, A. Peroxisome proliferator-activated receptor  $\gamma$  coactivator 1 (PGC-1)- and estrogen-related receptor (ERR)-induced regulator in muscle 1 (PERM1) is a tissue-specific regulator of oxidative capacity in skeletal muscle cells. *J. Biol. Chem.* **288**, 25207–25218 (2013).
57. Cho, Y. et al. Perm1 enhances mitochondrial biogenesis, oxidative capacity, and fatigue resistance in adult skeletal muscle. *FASEB J.* **30**, 674–687 (2016).
58. Zhao, S. et al. Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000–1004 (2010).
59. Brzezniak, L. K., Bijata, M., Szczesny, R. J. & Stepien, P. P. Involvement of human ELAC2 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol.* **8**, 616–626 (2011).
60. Holzmann, J. et al. RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell* **135**, 462–474 (2008).
61. Lee, K.-W. & Bogenhagen, D. F. Assignment of 2'-O-methyltransferases to modification sites on the mammalian mitochondrial large subunit 16 S ribosomal RNA (rRNA). *J. Biol. Chem.* **289**, 24936–24942 (2014).
62. Cingolani, H. E. et al. The positive inotropic effect of angiotensin II. *Hypertension* **47**, 727–734 (2006).
63. Forrester, S. J. et al. Angiotensin II signal transduction: an update on mechanisms of physiology and pathophysiology. *Physiol. Rev.* **98**, 1627–1738 (2018).
64. Kim, S. & Iwao, H. Molecular and cellular mechanisms of angiotensin II-mediated cardiovascular and renal diseases. *Pharmacol. Rev.* **52**, 11–34 (2000).
65. Symons, J. D. & Stebbins, C. L. Effects of angiotensin II receptor blockade during exercise: comparison of losartan and saralasin. *J. Cardiovasc. Pharmacol.* **28**, 223–231 (1996).
66. Stebbins, C. L. & Symons, J. D. Role of angiotensin II in hemodynamic responses to dynamic exercise in miniswine. *J. Appl. Physiol.* **78**, 185–190 (1995).
67. WILSON, J. X. The renin-angiotensin system in nonmammalian vertebrates. *Endocr. Rev.* **5**, 45–61 (1984).
68. Fournier, D., Luft, F. C., Bader, M., Ganten, D. & Andrade-Navarro, M. A. Emergence and evolution of the renin-angiotensin-aldosterone system. *J. Mol. Med.* **90**, 495–508 (2012).
69. Mueller, C. A., Eme, J., Tate, K. B. & Crossley, D. A. Chronic captopril treatment reveals the role of ANG II in cardiovascular function of embryonic American alligators (*Alligator mississippiensis*). *J. Comp. Physiol. B* **188**, 657–669 (2018).
70. Antl, M. et al. IRAG mediates NO/cGMP-dependent inhibition of platelet aggregation and thrombus formation. *Blood* **109**, 552–559 (2007).
71. Puetz, J. & Boudreaux, M. K. Evaluation of the gene encoding calcium and diacylglycerol regulated guanine nucleotide exchange factor I (CalDAG-GEFI) in human patients with congenital qualitative platelet disorders. *Platelets* **23**, 401–403 (2012).
72. Bezman, N. A. et al. Requirements of SLP76 tyrosines in ITAM and integrin receptor signaling and in platelet function in vivo. *J. Exp. Med.* **205**, 1775–1788 (2008).
73. Israels, S. & McMillan-Ward, E. CD63 modulates spreading and tyrosine phosphorylation of platelets on immobilized fibrinogen. *Thromb. Haemost.* **93**, 311–318 (2005).
74. Cooper, D. N., Millar, D. S., Wacey, A., Pemberton, S. & Tuddenham, E. G. Inherited factor X deficiency: molecular genetics and pathophysiology. *Thromb. Haemost.* **78**, 161–172 (1997).

75. Takahashi, N., Takahashi, Y. & Putnam, F. W. Primary structure of blood coagulation factor XIIIa (fibrinogenase, transglutaminase) from human placenta. *Proc. Natl Acad. Sci. USA* **83**, 8019–8023 (1986).
76. Mosesson, M. W. The roles of fibrinogen and fibrin in hemostasis and thrombolysis. *Semin. Hematol.* **29**, 177–188 (1992).
77. Pokorná, M. & Kratochvíl, L. Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zool. J. Linn. Soc.* **156**, 168–183 (2009).
78. Rovatsos, M., Pokorna, M., Altmanova, M. & Kratochvil, L. Cretaceous park of sex determination: sex chromosomes are conserved across iguanas. *Biol. Lett.* **10**, 20131093 (2014).
79. Gamble, T. et al. Restriction site-associated DNA sequencing (RAD-seq) reveals an extraordinary number of transitions among gecko sex-determining systems. *Mol. Biol. Evol.* **32**, 1296–1309 (2015).
80. Nielsen, S. V., Banks, J. L., Diaz, R. E., Trainor, P. A. & Gamble, T. Dynamic sex chromosomes in Old World chameleons (Squamata: Chamaeleonidae). *J. Evol. Biol.* **31**, 484–490 (2018).
81. Rovatsos, M., Altmanová, M., Pokorná, M. & Kratochvíl, L. Conserved sex chromosomes across adaptively radiated anolis lizards. *Evolution* **68**, 2079–2085 (2014).
82. Gamble, T. et al. The discovery of XY sex chromosomes in a boa and python. *Curr. Biol.* **27**, 2148–2153.e4 (2017).
83. Emerson, J. J. Evolution: a paradigm shift in snake sex chromosome genetics. *Curr. Biol.* **27**, R800–R803 (2017).
84. Hattori, R. S. et al. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. *Proc. Natl Acad. Sci. USA* **109**, 2955–2959 (2012).
85. Cortez, D. et al. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014).
86. Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A. & Yamamoto, Y. A duplicated, truncated *amh* gene is involved in male sex determination in an Old World silverside. *G3* **7**, 2489–2495 (2017).
87. Ieda, R. et al. Identification of the sex-determining locus in grass puffer (*Takifugu niphobles*) provides evidence for sex-chromosome turnover in a subset of *Takifugu* species. *PLoS ONE* **13**, e0190635 (2018).
88. Halpern, M. in *Biology of the Reptilia: Vol. 18, Physiology E: Hormones, Brain, and Behavior* (eds Gans, C. & Crews, D.) 423–523 (Univ. Chicago Press, 1992).
89. Martin, J. & Lopez, P. Chemoreception, symmetry and mate choice in lizards. *Proc. R. Soc. B* **267**, 1265–1269 (2000).
90. Baeckens, S., Martín, J., Garcia-Roa, R. & van Damme, R. Sexual selection and the chemical signal design of lacertid lizards. *Zool. J. Linn. Soc.* **183**, 445–457 (2018).
91. van Damme, R., Bauwens, D., Thoen, C., Vanderstighelen, D. & Verheyen, R. F. Responses of naive lizards to predator chemical cues. *J. Herpetol.* **29**, 38–43 (1995).
92. van Damme, R. & Castilla, A. M. Chemosensory predator recognition in the lizard *Podarcis hispanica*: effects of predation pressure relaxation. *J. Chem. Ecol.* **22**, 13–22 (1996).
93. Cooper, W. E. Correlated evolution of prey chemical discrimination with foraging, lingual morphology and vomeronasal chemoreceptor abundance in lizards. *Behav. Ecol. Sociobiol.* **41**, 257–265 (1997).
94. Cooper, W. Tandem evolution of diet and chemosensory responses in snakes. *Amphib. Reptil.* **29**, 393–398 (2008).
95. Hulbert, A. J. & Else, P. L. Evolution of mammalian endothermic metabolism: mitochondrial activity and cell composition. *Am. J. Physiol. Integr. Comp. Physiol.* **256**, R63–R69 (1989).
96. Castoe, T. A. et al. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl Acad. Sci. USA* **110**, 20645–20650 (2013).
97. Duan, J. et al. Transcriptome analysis of the response of burmese python to digestion. *Gigascience* **6**, 1–18 (2017).
98. Gleeson, T. T., Mitchell, G. S. & Bennett, A. F. Cardiovascular responses to graded activity in the lizards *Varanus* and *Iguana*. *Am. J. Physiol. Integr. Comp. Physiol.* **239**, R174–R179 (1980).
99. Agaba, M. et al. Giraffe genome sequence reveals clues to its unique morphology and physiology. *Nat. Commun.* **7**, 11519 (2016).
100. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
101. Kichigin, I. G. et al. Evolutionary dynamics of anolis sex chromosomes revealed by sequencing of flow sorting-derived microchromosome-specific DNA. *Mol. Genet. Genomics* **291**, 1955–1966 (2016).
102. Makunin, A. I. et al. Contrasting origin of B chromosomes in two cervids (Siberian roe deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. *BMC Genomics* **17**, 618 (2016).
103. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10 (2011).
104. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997> (2013).
105. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
106. Quinlan, A. R., Pedersen, B. S. & Dale, R. K. Pybedtools: a flexible python library for manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
107. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
108. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100**, 11484–11489 (2003).
109. Smit, A., Hubley, R. & Green, P. RepeatMasker Open-4.0. (2013–2015); <http://www.repeatmasker.org>
110. Cantarel, B. L. et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
111. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
112. Slater, G. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
113. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
114. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
115. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
116. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
117. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
118. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA family database. *Nucleic Acids Res.* **31**, 439–441 (2003).
119. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
120. Löytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
121. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
122. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
123. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
124. Smith, S. A., Brown, J. W. & Walker, J. F. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLoS ONE* **13**, e0197433 (2018).
125. Han, M. V., Thomas, G. W. C., Lugo-Martinez, J. & Hahn, M. W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
126. Mitchell, A. L. et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2018).
127. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
128. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
129. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
130. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
131. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
132. Smith, M. D. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* **32**, 1342–1353 (2015).
133. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

## Acknowledgements

Special thanks from B.G.B. to J. Romano for inspiration and historical information. We are grateful to staff at Zoo Atlanta for care of Slasher and Rinca and help obtaining samples, J. Pether from Reptilandia Zoo in Gran Canaria in the Canary Islands for additional samples, and R. Chadwick and N. Carli (Gladstone Genomics Core) for DNA and RNA-seq library preparation and MinION sequencing. We also thank K. Giorda, R. Abbas, D. Church (10x Genomics) for 10x Genomics Chromium sequencing and Supernova assembly. This work was supported by institutional funding from the Gladstone Institutes to B.G.B and K.S.P.; an NHLBI grant to K.S.P and B.G.B (UM1 HL098179); the Younger Family gift to B.G.B.; an NHGRI grant to P.-Y.K. (R01 HG005946); NIH training grants (T32 AR007175) to A.C.Y.M and (T32 HL007731) to Y.M. M.A. and M.R. were supported by GACR 17-22141Y, M.R. was additionally supported by Charles University projects PRIMUS/SCI/46 and Research Centre (204069).

## Author contributions

A.L.L. did genome annotation and all comparative genomics analyses. Y.Y.Y.L. led the sequencing and assembly efforts with Y.M. and A.C.Y.M. A.K.H. led the initial development of the project. A.I. sequenced isolated chromosomes with M.R., M.J.P. and M.A. under supervision of L.K., and assigned sequences with A.I.M., I.G.K. and V.A.T. M.F. and V.O. contributed to genome assembly in the lab of R.F. with C.C. W.L.E. initially assembled the transcriptomes and annotated the genome. M.M. and M.F. isolated samples and obtained PacBio sequence in the lab of T.P. and C.C. E.S. performed PacBio sequencing. P.V. and I.R. provided embryos for cell line establishment. O.A.R. provided frozen tissue samples. J.R.M. collected specimen blood. J.W.H., J.R.M. and C.C. provided direction on Varanid physiology. T.S.J. and C.C. provided direction on Komodo dragon ecology. P.-Y.K. coordinated the genomics efforts. K.S.P. directed comparative genomic

analysis. B.G.B. initiated and coordinated the project. A.L.L., K.S.P. and B.G.B. wrote the paper with input from all authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41559-019-0945-8>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to K.S.P. or B.G.B.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019



**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/>.



## 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

- |                                     |                                     |                                                                                                                                                                                                                                                            |
|-------------------------------------|-------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                           |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                     |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | 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

No software was used for data collection.

Data analysis

Bionano IrysSolve assembler version 5134, 10x Genomics Supernova assembler v1.0, SSPACe-LongRead, dopseq, cuptadapt v1.13, pybedtools v0.7.10, LAST, RepeatMasker, STAR version 2.6.0, Trinity version 2.4.0, exonerate version 2.2.0, MAKER version 3.01.02, SNAP, Augustus version 3.1.1, InterProScan version 5.31.70, OrthoFinder version 2.0.0, tRNAscan-SE version 1.3.1, Infernal software suite, PRANK v.170427, IQ-TREE version 1.6.7.1, CAFE v4.2, Protein-Protein BLAST 2.6.0+, TMHMM v2.0, MAFFT v7.310, trimAL, PAL2NAL, HYPHY, aBSREL, EMBOSS needle version 6.6.0.0

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### 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

The assembled Komodo dragon genome is available in NCBI under the accession SJPD000000000. All DNA sequencing used to generate the assembly is available in the NCBI SRA database under accession PRJNA523222. Illumina sequencing data from chromosome pools is available in the NCBI SRA under accession PRJNA529483. RNA-seq data of heart tissue is available in the NCBI SRA under accession PRJNA527313. Original protein annotations, noncoding RNA annotations, all alignments for phylogenetic analyses and selection analyses, and newick files of phylogenetic trees are available in the following Figshare repositories: doi:10.6084/m9.figshare.7961135.v1, doi:10.6084/m9.figshare.7955891.v1, doi:10.6084/m9.figshare.7955879.v1, doi:10.6084/m9.figshare.7949483.v1, doi:10.6084/m9.figshare.7759496.v1, doi:10.6084/m9.figshare.7967300. The project folder for all Figshare data is available at <https://figshare.com/projects/>

Data\_for\_Komodo\_dragon\_genome\_paper/61271.

## 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://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

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

Sample size    The comparative genomics work in this study was performed using single genomes from individual species, so no sample size calculations were performed.

Data exclusions    No data were excluded.

Replication    No biological experiments were performed in this work that could be replicated.

Randomization    Randomization is not relevant to this paper, as the main hypotheses in this work refer to individual species.

Blinding    No blinding was possible for analyses carried out in this study as hypotheses were tested on single genomic sequences.

## 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 & experimental systems

n/a    Involved in the study

Antibodies

Eukaryotic cell lines

Palaeontology

Animals and other organisms

Human research participants

Clinical data

### Methods

n/a    Involved in the study

ChIP-seq

Flow cytometry

MRI-based neuroimaging

## Animals and other organisms

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

Laboratory animals    Varanus Komodoensis, ages 20 and 6, house at Zoo Atlanta, one individual at Reptilandia zoo in Gran Canaria in the Canary Islands, and one deceased individual (tissue previously collected) and San Diego Zoo. All maintained under ambient conditions.

Wild animals    No wild animals

Field-collected samples    No field collected samples

Ethics oversight    Zoo Atlanata IACUC

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