A high-resolution, chromosome-assigned Komodo dragon genome reveals adaptations in the cardiovascular, muscular, and chemosensory systems of monitor lizards

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

1. Gladstone Institutes, San Francisco, CA 94158, USA.

2. Cardiovascular Research Institute, University of California, San Francisco, CA 94143, USA.

3. Department of Biology, University of Florence, 50019 Sesto Fiorentino (FI), Italy

4. Institute for Human Genetics, University of California, San Francisco, CA 94143, USA.

5. Department of Comparative Biomedicine and Food Science, University of Padova, 35020 Legnaro (PD), Italy

6. Department of Ecology, Charles University, 128 00 Prague, Czech Republic

7. Institute of Animal Physiology and Genetics, The Czech Academy of Sciences, 277 21 Liběchov, Czech Republic

8. Institute of Molecular and Cellular Biology SB RAS, Novosibirsk 630090, Russia

9. B.U. Bioscience, Wageningen University, 6700 AA Wageningen, The Netherlands

10. Centre for Integrative Ecology, Deakin University, Waurn Ponds 3220, Australia

23 11. Department of Ecology and Evolutionary Biology, School of Biological Sciences, University of
24 California, Irvine, CA 92627, USA.
25 12. Institute for Conservation Research, San Diego Zoo, Escondido, CA 92027, USA
26 13. Zoo Atlanta, Atlanta GA 30315, USA.
27 14. School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, 30332, USA.
28 15. Department of Dermatology, University of California, San Francisco, CA 94143, USA.
29 16. Department of Epidemiology and Biostatistics, University of California, San Francisco, CA
30 94158, USA.
31 17. Institute for Computational Health Sciences, University of California, San Francisco, CA
32 94158, USA.
33 18. Chan-Zuckerberg BioHub, San Francisco, CA 94158, USA.
34 19. Department of Pediatrics, University of California, San Francisco, CA 94143, USA.

Summary

Monitor lizards are unique among ectothermic reptiles in that they have a high aerobic capacity and distinctive cardiovascular physiology which resembles that of endothermic mammals. We have sequenced the genome of the Komodo dragon (*Varanus komodoensis*), the largest extant monitor lizard, and present a high resolution *de novo* chromosome-assigned genome assembly for *V. komodoensis*, generated with a hybrid approach of long-range sequencing and single molecule physical mapping. Comparing the genome of *V. komodoensis* with those of related species showed evidence of positive selection in pathways related to muscle energy metabolism, cardiovascular homeostasis, and thrombosis. We also found species-specific expansions of a chemoreceptor gene family related to pheromone and kairomone sensing in *V. komodoensis* and several other lizard lineages. Together, these evolutionary signatures of adaptation reveal genetic underpinnings of the unique Komodo sensory, cardiovascular, and muscular systems, and suggest that selective pressure altered thrombosis genes to help Komodo dragons evade the anticoagulant effects of their own saliva. As the only sequenced monitor lizard genome, the Komodo dragon genome is an important resource for understanding the biology of this lineage and of reptiles worldwide.

52 **Introduction**

53 The evolution of form and function in the animal kingdom contains numerous examples
54 of innovation and diversity. Within vertebrates, non-avian reptiles are a particularly interesting
55 lineage. There are an estimated 10,000 reptile species worldwide, found on every continent
56 except Antarctica, with a diverse range of morphologies and lifestyles [1]. This taxonomic
57 diversity corresponds to a broad range of anatomic and physiological adaptations.
58 Understanding how these adaptations evolved through changes to biochemical and cellular
59 processes will reveal fundamental insights in areas ranging from anatomy and metabolism to
60 behavior and ecology.

61 The varanid lizards (genus *Varanus*, or monitor lizards) are an unusual group of
62 squamate reptiles characterized by a variety of traits not commonly observed within non-avian
63 reptiles. Varanid lizards vary in mass by close to five orders of magnitude (8 grams–100
64 kilograms), comprising the genus with the largest range in size [2]. Among the squamate reptiles,
65 varanids have a unique cardiopulmonary physiology and metabolism with numerous parallels
66 to the mammalian cardiovascular system. For example, their cardiac anatomy allows high
67 pressure shunting of oxygenated blood to systemic circulation [3]. Furthermore, varanid lizards
68 can achieve and sustain very high aerobic metabolic rates accompanied by elevated blood
69 pressure and high exercise endurance [4–6], which facilitates high-intensity movements while
70 hunting prey [7]. The specialized anatomical, physiological, and behavioral attributes of varanid
71 lizards are all present in the Komodo dragon (*Varanus komodoensis*). As the largest extant lizard
72 species, Komodo dragons can grow to 3 meters in length and run at speeds of up to 20
73 kilometers miles per hour, which allows them to hunt large prey such as deer and boar in their

74 native Indonesia [8]. Komodo dragons have a higher metabolism than predicted by allometric
75 scaling relationships for varanid lizards [9], which helps explains their extraordinary capacity for
76 daily movement to locate prey [10]. Their ability to locate injured or dead prey through scent
77 tracking over several kilometers is enabled by a powerful olfactory system [8]. Additionally,
78 serrate teeth, sharp claws, and saliva with anticoagulant and shock inducing properties aid
79 Komodo dragons in hunting prey [11,12]. Komodo dragons have aggressive intraspecific conflicts
80 over mating, territory, and food, and wild individuals often bear scars from previous conflicts [8].
81
82 To understand the genetic underpinnings of the specialized Komodo dragon physiology,
83 we sequenced its genome and used comparative genomics to discover how the Komodo dragon
84 genome differs from other species. We present a high quality *de novo* assembly, generated with
85 a hybrid approach of long-range sequencing using 10x Genomics, PacBio, and Oxford Nanopore
86 sequencing, and single-molecule physical mapping using the BioNano platform. This suite of
87 technologies allowed us to confidently assemble a high-quality reference genome for the
88 Komodo dragon, which can serve as a template for other varanid lizards. We used this genome
89 to understand the relationship of varanids to other reptiles using a phylogenomics approach.
90 We uncovered Komodo-specific positive selection for a suite of genes encoding regulators of
91 muscle metabolism, cardiovascular homeostasis, and thrombosis. Further, we discovered
92 multiple lineage-specific expansions of a family of chemoreceptor genes in several squamates,
93 including some lizards and a snake, as well as in the Komodo dragon genome. Finally, we
94 generated a high-resolution chromosomal map resulting from assignment of scaffold to

95 chromosomes, providing a powerful tool to address questions about karyotype and sex
96 chromosome evolution in squamates.
97
98 **Results**
99 *De novo genome assembly*
100 We obtained DNA from peripheral blood of two male individuals housed at Zoo Atlanta:
101 Slasher, a male offspring of the first Komodo dragons given as gifts in 1986 to US President
102 Reagan from President Suharto of Indonesia, and Rinca, a juvenile male (Figure 1A). The *V.*
103 *komodoensis* genome is distributed across 20 pairs of chromosomes, comprising eight pairs of
104 large chromosomes and 12 pairs of microchromosomes [13,14]. *De novo* assembly was performed
105 using 57X coverage of 10x Genomics linked-read sequencing data to generate an initial
106 assembly with a scaffold N50 of 10.2 Mb and a contig N50 of 95 kb. Separately, 80X coverage of
107 Bionano physical mapping data was *de novo* assembled to create an assembly with a scaffold
108 N50 of 1.2 Mb. These two assemblies were merged into a hybrid assembly and then scaffolded
109 further using 6.3X coverage from PacBio sequencing and 0.75X coverage from Oxford Nanopore
110 MinIon sequencing, for a total coverage of 144X. The final assembly contained 1,403 scaffolds
111 (>10 kb) with an N50 of 29 Mb (longest scaffold: 138 Mb) (Table 1). The assembly is 1.51 Gb in
112 size, ~32% smaller than the genome of the Chinese crocodile lizard (*Shinisaurus crocodilurus*),
113 the closest relative of the Komodo dragon for which a sequenced genome is available, and
114 ~15% smaller than the green anole (*Anolis carolinensis*), a model squamate lizard (Table S1). An
115 assembly-free error corrected *k*-mer counting estimate of genome size estimates the Komodo
116 dragon genome to be 1.69 Gb, making our assembly 89% complete. The GC content of the

A

B

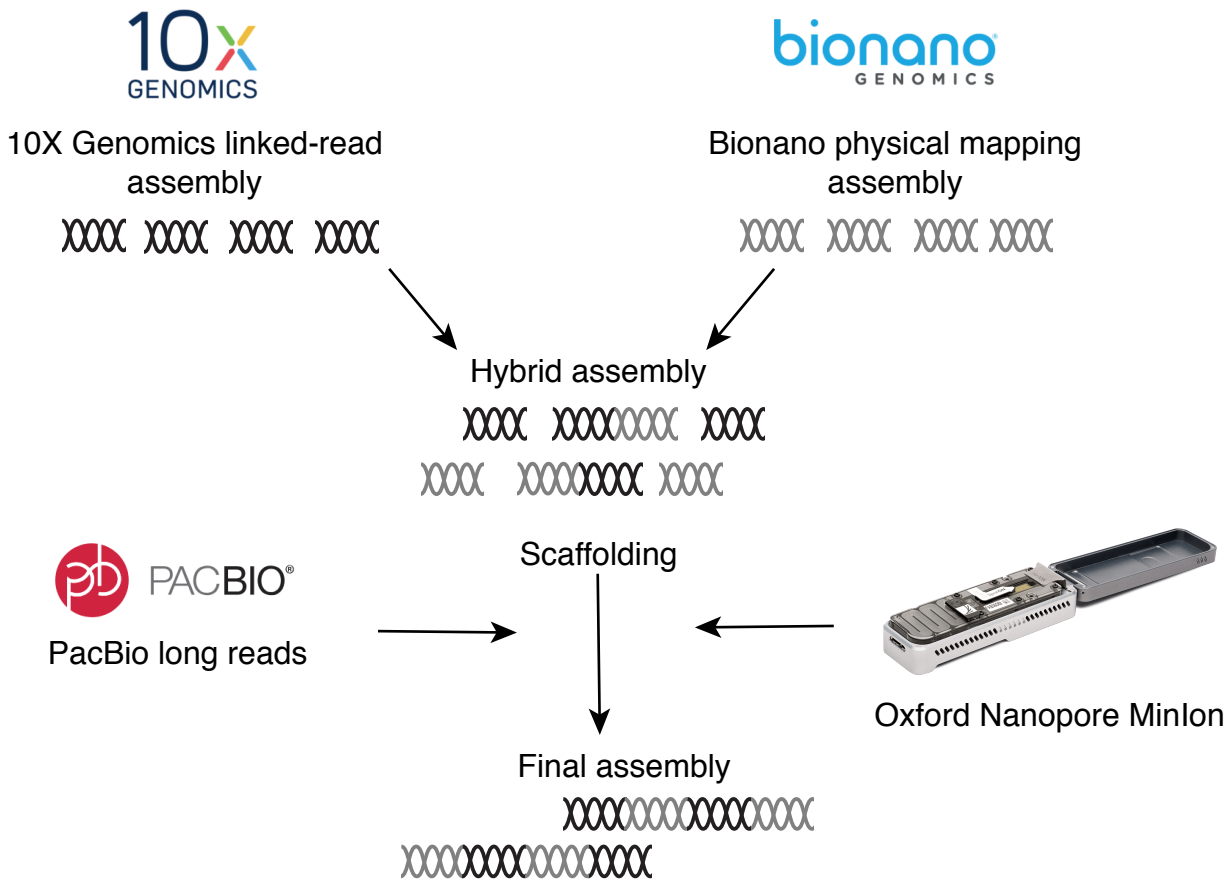

Figure 1. (A) Komodo dragons (left, Slasher; right, Rinca) sampled for DNA in this study. Photos courtesy of Adam K Thompson/Zoo Atlanta. (B) Genome assembly workflow. Two separate de novo assemblies were generated with 10x genomics and Bionano physical mapping data and merged into an intermediate hybrid assembly. Long reads from PacBio and Oxford Nanopore MinIon were used to scaffold the hybrid assembly into a final version.

117 Komodo dragon genome is 44.0%, similar to the GC content of the *S. crocodilurus* genome
118 (44.5%) but higher than the GC content of *A. carolinensis* (40.3%) (Table S1). Repeats were
119 annotated using RepeatMasker and the squamate repeat dabatase [15]. Repetitive elements
120 accounted for 32% of the genome, most of which were transposable elements (Table S2). As
121 repetitive elements in *S. crocodilurus* account for 49.6% of the genome [16], most of the
122 difference in size between the Komodo dragon genome and its closest sequenced relative
123 genome can be attributed to differences in repetitive element content.
124
125 *Chromosome scaffold content*
126 We isolated chromosome-specific DNA pools from a female embryo of *V. komodoensis*
127 (VKO) from Prague zoo stock through flow sorting [14]. We obtained Illumina short-read
128 sequences of these 15 DNA pools containing all chromosomes of *V. komodoensis* (Table S3). For
129 each chromosome we determined scaffold content and homology to *A. carolinensis* and *G.*
130 *gallus* chromosomes (Table 2 and Table S4). For each chromosome, we determined scaffold
131 content and homology to *A. carolinensis* and *G. gallus* chromosomes (Table 2 and Tables S4-5).
132 For those pools where chromosomes were mixed (VKO6/7, VKO8/7, VKO9/10, VKO11/12/W,
133 VKO17/18/Z, VKO17/18/19) we determined partial scaffold content of single chromosomes.
134 Homology to *A. carolinensis* microchromosomes was determined using scaffold assignments to
135 chromosomes performed in *Anolis* chromosome-specific sequencing project [17]. A total of 243
136 scaffolds containing 1.14 Gb (75% of total 1.51 Gb assembly) were assigned to 20 chromosomes
137 of *V. komodoensis*. As sex chromosomes shares homologous regions (pseudoautosomal
138 regions), scaffolds that were enriched in both 17/18/Z and 11/12/W samples most likely

139 contained sex chromosome regions of *V. komodoensis*. Considering that our reference genome
140 was from a male individual, they were assigned to the Z chromosome. All these scaffolds were
141 homologous to *A. carolinensis* chromosome 18, and mostly to chicken chromosome 28 as
142 recently determined by transcriptome analysis [18].
143
144 *Gene annotation*
145 As Komodo dragons have a unique cardiovascular physiology, we used heart tissue as
146 the source for RNA sequencing to increase the accuracy of cardiovascular gene prediction,
147 increasing our power to detect interesting changes to the cardiovascular system encoded in the
148 genome. RNA sequencing was assembled into transcripts with Trinity [19]. After soft masking
149 repetitive elements, genes were annotated using the MAKER pipeline with protein homology,
150 assembled transcripts, and de novo predictions as evidence, and stringently quality filtered (see
151 Methods). A total of 18,462 protein coding genes were annotated in the Komodo genome,
152 17,194 (93%) of which have one or more annotated Interpro functional domain (Table 1). Of
153 these protein-coding genes, 63% were expressed (RPKM > 1) in the heart. A total of 89% of
154 Komodo dragon protein-coding genes are orthologous to genes in the model lizard *A.*
155 *carolinensis* genome. The median percent identity of single-copy orthologs between Komodo
156 and *A. carolinensis* is 68.9%, whereas it is 70.6% between single-copy orthologs in Komodo and
157 *S. crocodilurus* (Figure S1).
158
159 *Phylogenetic placement of Komodo*

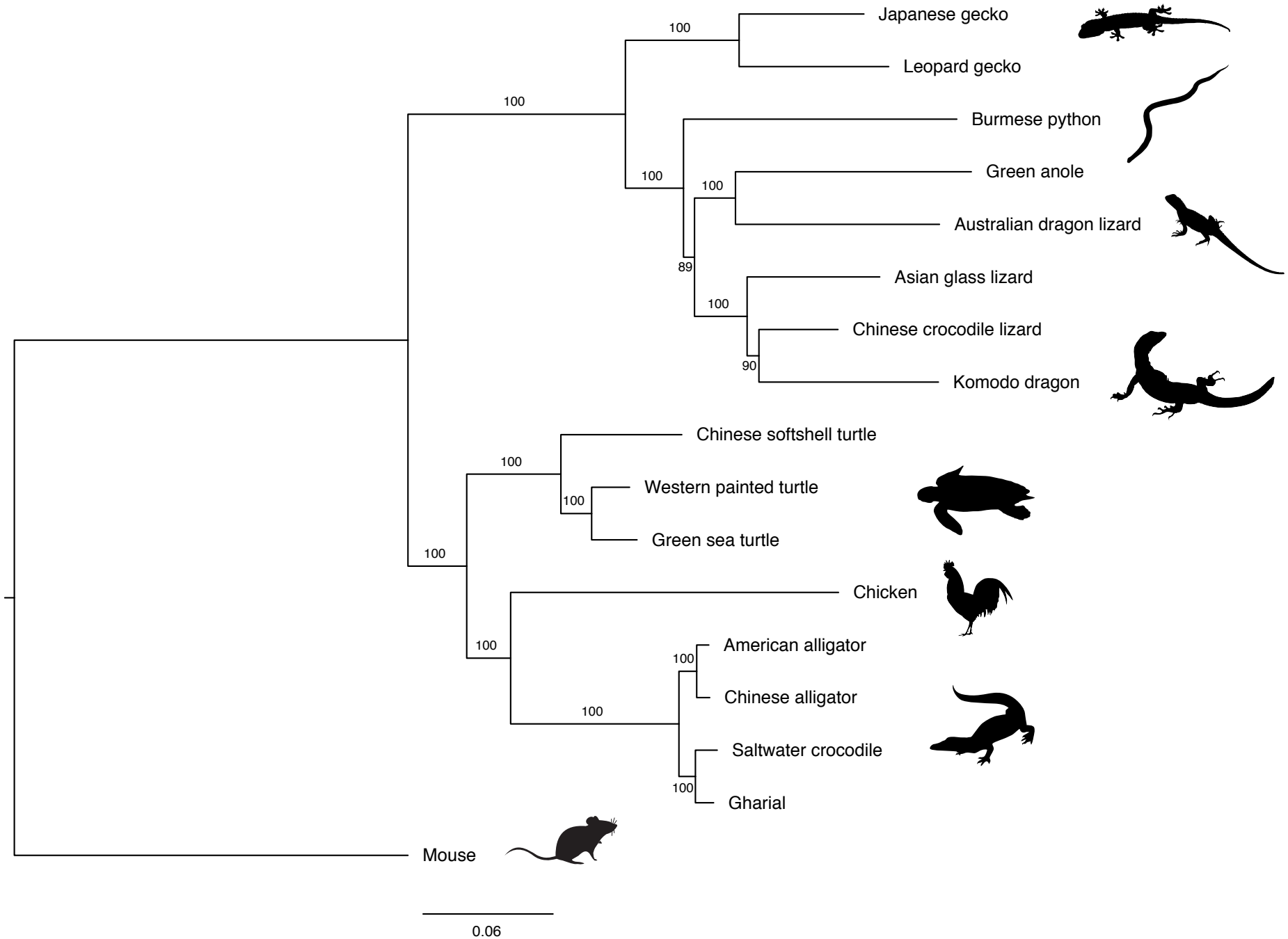

Figure 2. Estimated species phylogeny of 15 non-avian reptiles species and 2 additional vertebrates. Maximum likelihood phylogeny was constructed from 2,752 single-copy orthologous proteins. Support values from 10,000 bootstrap replicates are shown. All images obtained from PhyloPic.org.

160 As the Komodo dragon genome is the first monitor lizard (Family Varanidae) to have a
161 complete genome sequence, previous phylogenetic analyses of varanid lizards has been limited
162 to marker sequences [20,21]. We used the Komodo dragon genome to estimate a species tree
163 using 2,752 single copy orthologs (see Methods) present in the Komodo dragon and 14
164 representative non-avian reptile species, including 7 squamates, 3 turtles, and 4 crocodilians,
165 along with one avian species (chicken) and one mammalian species (mouse) (Figure 2). The
166 placement of Komodo dragon and the monitor lizard genus using this genome-wide dataset
167 agrees with previous marker gene studies [20,21].
168
169 *Expansion of vomeronasal genes across squamate reptiles*
170 The vomeronasal, or the Jacobson's, organ is a chemosensory tissue that detects
171 chemical cues such as pheromones and kairomones. It is shared across amphibians, mammals,
172 and reptiles though it has been secondarily lost in some groups, including birds [22,23]. Squamate
173 reptiles such as snakes and lizards have apparently functional vomeronasal organs with the
174 ability to sense prey-derived chemical signals, as well as specific associated behaviors such as
175 tongue-flicking to deliver olfactory cues to the sensory tissue, and it is clear that the
176 vomeronasal organ plays an important role in squamate reptile ecology [24]. Two types of
177 chemosensory receptors, both of which are seven-transmembrane G-protein coupled
178 receptors, function as sensors in the vomeronasal organ. The number of Type 1 vomeronasal
179 receptors (V1Rs) has expanded through gene duplications in certain mammalian lineages, while
180 the number of Type 2 receptors (V2Rs) has expanded in amphibians and some mammalian
181 lineages [22]. Crocodilian and turtle genomes contain few to no V1R and V2R genes [25]. Snakes, in

182 contrast, have a significantly expanded V2R repertoire that has arisen through gene duplication
183 [26].
184 To clarify the relationship between vomeronasal organ function and evolution of
185 vomeronasal-receptor gene families, we analyzed the coding sequences of 15 reptiles, including
186 Komodo, for presence of V1R and V2R genes (Figure 3A). We confirmed that there are few V1R
187 genes across reptiles generally and few to no V2R genes in crocodilians and turtles (Table S6).
188 The low number of V2R genes in green anole (*Anolis carolinensis)* and Australian dragon lizard
189 (*Pogona vitticeps)* suggest that V2R genes are infrequently expanded in iguanians, though more
190 iguanian genomes are needed to test this hypothesis. In contrast, we found a large repertoire
191 of V2Rs, comparable in size to that of snakes, in the Komodo dragon and other lizards.
192 To infer the details of the dynamic evolution of this gene family, we built a phylogeny of
193 all V2R gene sequences across squamates (Figure 3B). The topology of this phylogeny supports
194 that, as previously hypothesized, V2Rs expanded in the common ancestor of squamates, as
195 there are clades of gene sequences containing members from all species [26]. In addition, there
196 are a large number of well-supported single species clades (i.e., Komodo dragon only) dispersed
197 across the gene tree, which is consistent with multiple duplications of V2R genes later in
198 squamate evolution, including in the Komodo and gecko lineages (Figure 3B).
199 Because V2Rs have expanded in rodents through tandem gene duplications that
200 produced clusters of paralogs [27], we examined clustering of V2R genes in our Komodo assembly
201 to determine if a similar mechanism is likely driving these gene expansions. Of 151 V2Rs, 99 are
202 organized into 26 gene clusters ranging in size from 2 to 14 genes (Figure 4A, Table S7). To
203 understand if these gene clusters arose through tandem gene duplication, we constructed a

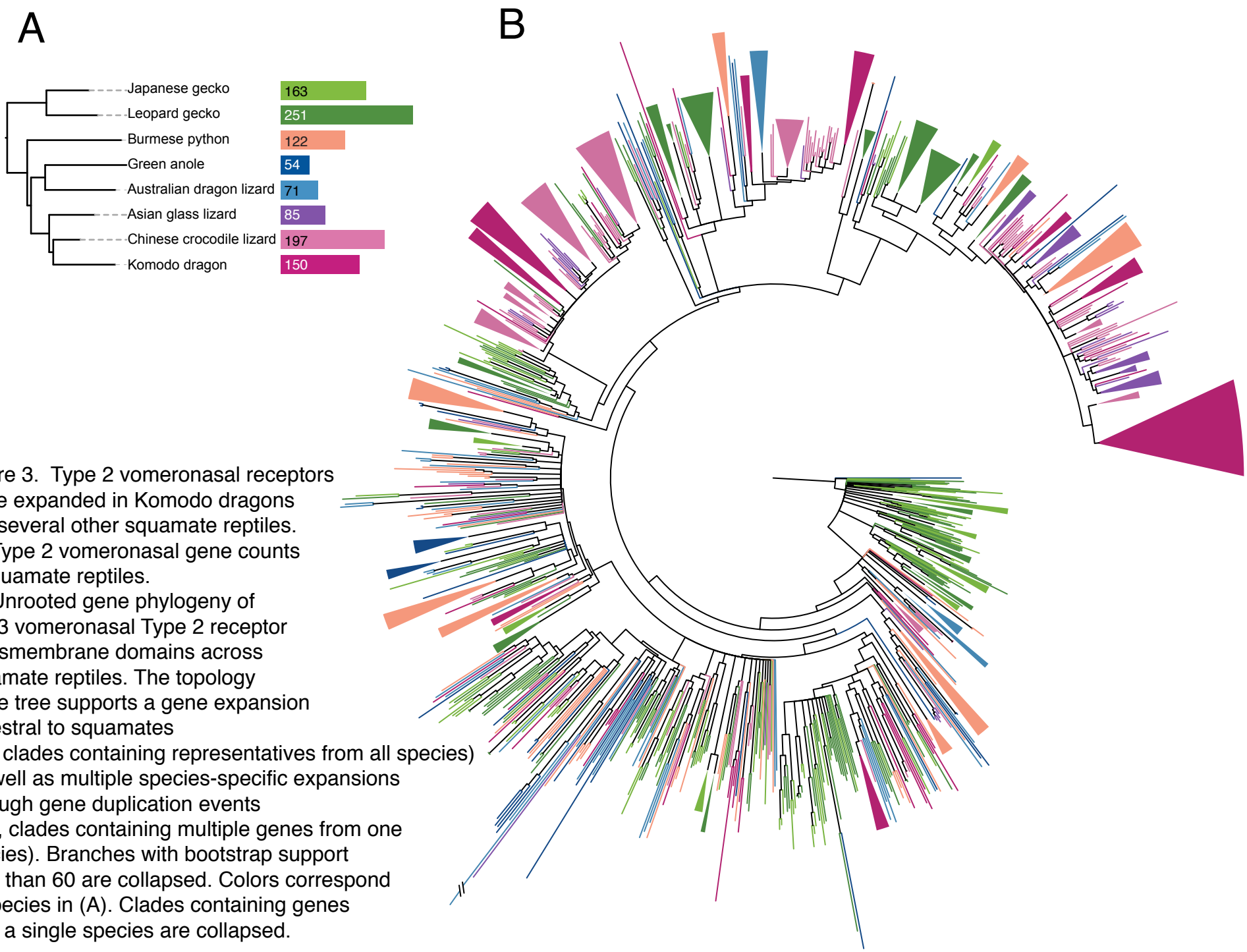

Figure 3. Type 2 vomeronasal receptors have expanded in Komodo dragons and several other squamate reptiles. (A) Type 2 vomeronasal gene counts in squamate reptiles. (B) Unrooted gene phylogeny of 1,093 vomeronasal Type 2 receptor transmembrane domains across squamate reptiles. The topology of the tree supports a gene expansion ancestral to squamates (i.e., clades containing representatives from all species) as well as multiple species-specific expansions through gene duplication events (i.e., clades containing multiple genes from one species). Branches with bootstrap support less than 60 are collapsed. Colors correspond to species in (A). Clades containing genes from a single species are collapsed.

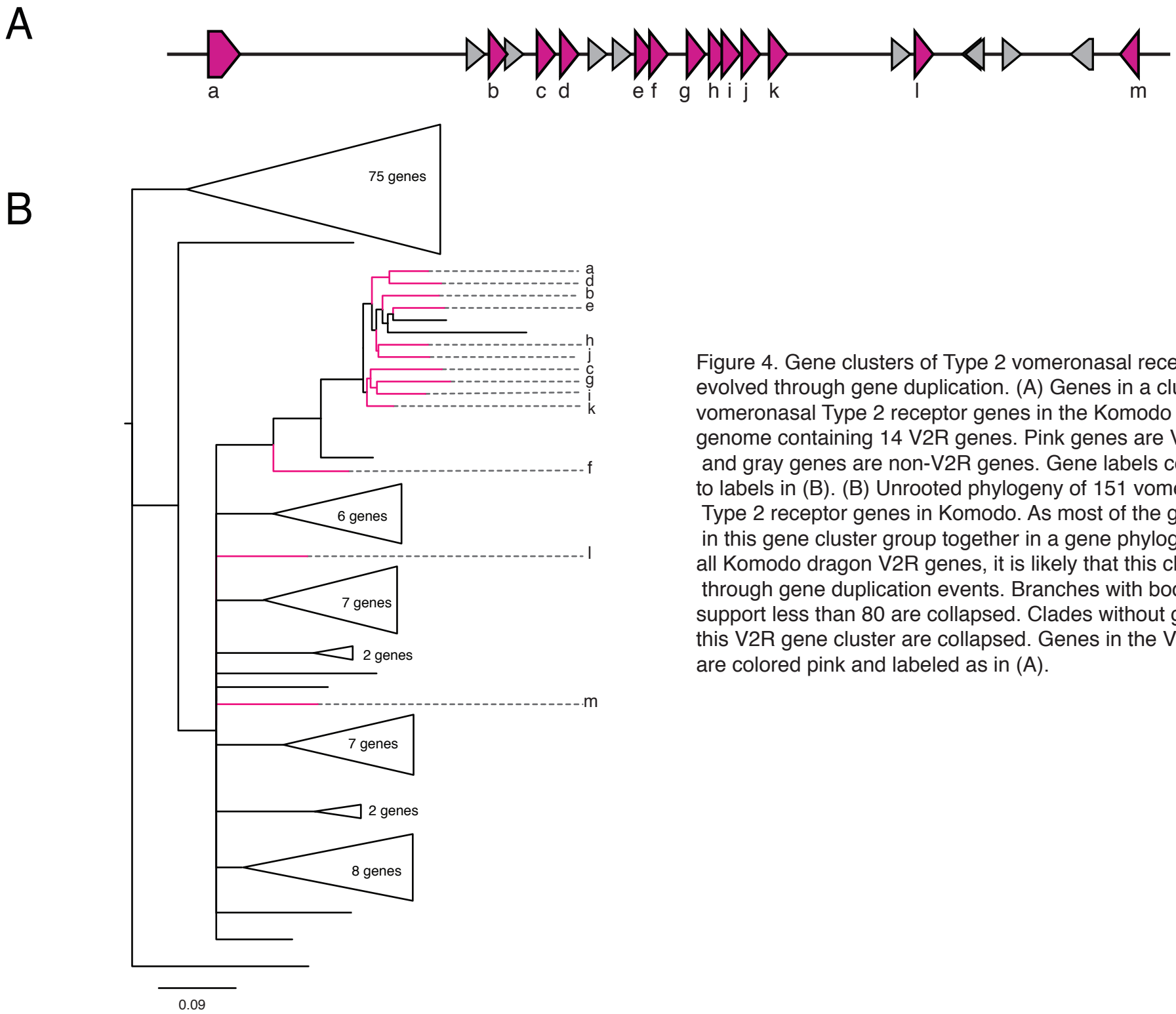

Figure 4. Gene clusters of Type 2 vomeronasal receptors evolved through gene duplication. (A) Genes in a cluster of vomeronasal Type 2 receptor genes in the Komodo dragon genome containing 14 V2R genes. Pink genes are V2R genes and gray genes are non-V2R genes. Gene labels correspond to labels in (B). (B) Unrooted phylogeny of 151 vomeronasal Type 2 receptor genes in Komodo. 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 colored pink and labeled as in (A).

204 phylogeny of all Komodo dragon V2R genes (Figure 4B). The largest V2R cluster contains 14 V2R
205 genes, which group together in a gene tree of Komodo V2R genes (Figure 4). Of the remaining
206 52 V2R genes, 38 are on scaffolds less than 10 Kb in size, so our estimate of V2R clustering is a
207 lower bound due to fragmentation in the genome assembly (Table S7).
208
209 *Positive selection*
210 To test for adaptive protein evolution in the Komodo dragon genome, we identified
211 single-copy orthologs across squamate reptiles, built codon alignments, and ran tests of
212 positive selection using a branch-site model to determine genes that have diversified in the
213 varanid lineage (see Methods and Table S8). Our analysis revealed 201 genes with signatures of
214 positive selection in Komodo dragons (Table S9). Many of the genes under positive selection
215 point towards important adaptations of the Komodo dragon's mammalian-like cardiovascular
216 and metabolic functions, which are unique among non-varanid ectothermic reptiles, though
217 25% of positively selected genes were not detectably expressed in the heart and likely
218 represent adaptations in other aspects of Komodo dragon biology. Pathways with positively
219 expressed genes include mitochondrial regulation and cellular respiration, hemostasis and the
220 coagulation cascade, innate and adaptive immunity, and angiotensinogen (a central regulator of
221 cardiovascular physiology). Many of these have implications for Komodo physiology, and for
222 varanid lizard physiology generally. We identified several functional categories with multiple
223 positively selected for more detailed analysis. In each case, the genes are located in different
224 parts of the Komodo genome and therefore likely represent recurrent selection on these
225 functions during Komodo evolution.

226

227 *Positive selection of genes regulating mitochondrial function*

228 Mitochondria regulate energy production in cells through the oxidative phosphorylation
229 process, which is mediated through the electron transport chain. Multiple subunits and
230 assembly factors of the Type 1 NADH dehydrogenase and cytochrome c oxidase protein
231 complexes, which perform the first and last steps of the electron transport chain respectively,
232 show evidence of positive selection in the Komodo dragon genome (Figure 5, Figure S2, Table
233 S9). These include the genes *NDUFA7*, *NDUFAF7*, *NDUFAF2*, *NDUFB5* from the Type 1 NADH
234 dehydrogenase complex and *COX6C* and *COA5* from the cytochrome c oxidase complex.

235 Beyond the electron transport chain, other elements of mitochondrial function have
236 signatures of positive selection in the Komodo lineage (Figure 5). Of note, we also detected
237 positive selection for *ACADL*, which encodes LCAD - acyl-CoA dehydrogenase, long chain—a
238 member of the acyl-CoA dehydrogenase family. LCAD is a critical enzyme for mitochondrial
239 fatty acid beta-oxidation, the major postnatal metabolic process in cardiac myocytes [28].
240 Further, two genes that promote mitochondrial biogenesis, *TFB2M* and *PERM1*, have
241 undergone positive selection in the Komodo dragon. *TFB2M* regulates mtDNA transcription and
242 dimethylates mitochondrial 12s rRNA [29,30]. *PERM1* regulates the expression of selective
243 PPARGC1A/B and ESRRA/B/G target genes with roles in glucose and lipid metabolism, energy
244 transfer, contractile function, muscle mitochondrial biogenesis and oxidative capacity [31].
245 *PERM1* also enhances mitochondrial biogenesis, oxidative capacity, and fatigue resistance when
246 over-expressed in mice [32]. Finally, we also identified *MDH1*, encoding malate dehydrogenase,

247 which together with the mitochondrial *MDH2*, regulates the supply of NADPH and acetyl-CoA to
248 the cytoplasm, thus modulating fatty acid synthesis [33].

249 Multiple factors regulating translation within the mitochondria have also undergone
250 positive selection in the Komodo dragon (Figure 5). This includes the mitochondrial ribosome,
251 including four components of 28S small ribosomal subunit (*MRPS15*, *MRPS23*, *MRPS31*, and
252 *AURKAIP1*) and two components of the 39S large ribosomal subunit (*MRPL28* and *MRPL37*). We
253 also found evidence for positive selection on the *ELAC2* and *TRMT10C* genes, which are
254 required for maturation of mitochondrial tRNA, and MRM1, which encodes a mitochondrial
255 rRNA methyltransferase [34–36].

256 Overall, these instances of positive selection in a large range of genes encoding proteins
257 important for mitochondrial function and biogenesis clearly point to a coordinated genetic
258 pathway that could explain the remarkable aerobic capacity of the Komodo dragon. While it is
259 not possible to determine whether these adaptations are present in other monitor lizards in the
260 absence of a sequenced genome, changes in cellular respiration likely play a role in the high
261 aerobic capacity of most varanid lizards.

262

263 *Positive selection of angiotensinogen*

264 We detected positive selection for angiotensinogen (*AGT*), which encodes the precursor
265 of several important peptide regulators of cardiovascular function, the most well-studied being
266 angiotensin II (AII) and angiotensin1-7 (A1-7). AII has multiple important and potent activities in
267 cardiovascular physiology. The two most notable, and perhaps most relevant to Komodo
268 dragon physiology, are its vasoactive function in blood vessels, and its inotropic effects on the

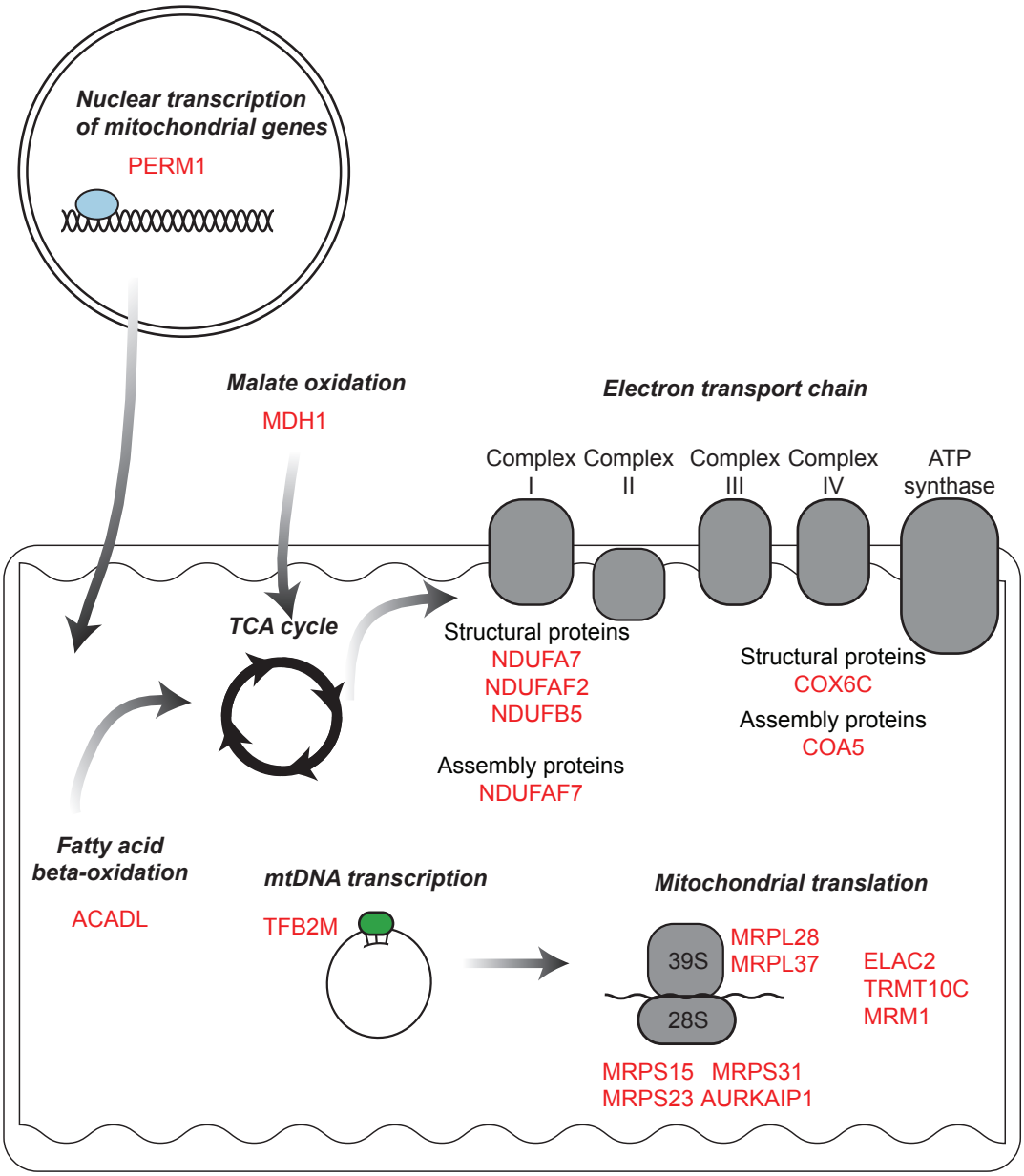

Figure 5. Mitochondrial genes under positive selection in the Komodo dragon. Genes in the Komodo dragon genome under positive selection include components of the electron transport chain, regulators of transcription, regulators of translation, and fatty acid beta-oxidation.

269 heart. In mammals during intense physical activity, All increases and contributes to arterial
270 blood pressure and regional blood regulation [37,38]. The positive selection for *AGT* points to
271 important adaptations in these physiological parameters. Reptiles have a functional renin-
272 angiotensin system that is important for their cardiovascular physiology [39–41]. It is likely that
273 positive selection for *AGT* is related to a mammalian-like cardiovascular function in the Komodo
274 dragon.
275
276 *Positive selection of thrombosis-related genes*
277 We find evidence for positive selection across different elements of the coagulation
278 cascade, including regulators of platelets and fibrin. The coagulation cascade controls
279 thrombosis, or blood clotting, preventing blood loss during injury. Four genes that regulate
280 platelet activities, *MRVI1*, *RASGRP1*, *LCP2*, and *CD63* have undergone positive selection in the
281 Komodo dragon genome. *MRVI1* is involved in inhibiting platelet aggregation [42], *RASGRP1*
282 coordinates calcium dependent platelet responses [43], *LCP2* is involved in platelet activation [44],
283 and *CD63* plays a role in controlling platelet spreading [45]. In addition to regulators of platelets,
284 two coagulation factors, *F10* (Factor X) and *F13B* (Coagulation factor XIII B chain) have
285 undergone positive selection in the Komodo genome. Factor X is centrally important to the
286 coagulation cascade and its activation is the first step in initiating coagulation [46]. Factor 13B is
287 the beta subunit of Factor 13, which is the final coagulation factor activated in the coagulation
288 cascade [47]. Further, *FGB*, which encodes one of the three subunits of fibrinogen, the molecule
289 which is converted to the clotting agent fibrin [48], has undergone positive selection in the
290 Komodo genome.

291 Komodo dragons, along with other species of monitor lizards, produce anticoagulants
292 and hypotension-inducing proteins in their saliva which are hypothesized to aid in hunting [11,12].
293 In addition to hunting, Komodo dragons use their serrate teeth during intraspecific conflict,
294 which can be aggressive and inflict serious wounds [8]. Because it is likely that their saliva enters
295 the bloodstream of Komodo dragons during these conflicts, we hypothesize that the positive
296 selection that we detected in many Komodo dragon coagulation genes may result from
297 selective pressure for Komodo dragons to evade the anticoagulant effects of conspecifics.
298
299 *Discussion*
300 We have sequenced and assembled a high-quality genome of the Komodo dragon. The
301 combination of platforms that we used allowed the de novo assembly of a genome that will
302 serve as a template for analysis of other varanid genomes, and for further investigation of
303 genomic innovations in the varanid lineage. Moreover, we assigned 75% of the genome to
304 chromosomes. Assignment of the Komodo dragon genome to chromosomes provides a
305 significant contribution to comparative genomics of squamates and vertebrates in general.
306
307 Our comparative genomic analysis identified previously undescribed species-specific
308 expansion of Type 2 vomeronasal receptors across multiple squamates, including lizards and at
309 least one snake. It will be exciting to explore the role this expansion of V2Rs plays in behavior
310 and ecology of Komodo dragons, including their ability to locate prey at long distances [8].
311 Komodo dragons, like other squamates, are known to possess a sophisticated lingual-
312 vomeronasal systems for chemical sampling of their environment [49]. This sensory apparatus

313 allows Komodo dragons to perceive chemicals from the environment for a variety of social and
314 ecological activities, including kin recognition, mate choice [50,51], predator avoidance [52,53],
315 hunting prey [54,55], and for locating and tracking injured or dead prey. Komodo dragons are
316 unusual as they adopt both foraging tactics across ontogeny with smaller juveniles preferring
317 active foraging for small prey and large adult dragons targeting larger ungulate prey via ambush
318 predation [10]. However, retention of a highly effective lingual-vomeronasal system across
319 ontogeny seems likely, given the exceptional capacity for Komodo dragons of all sizes to locate
320 injured or dead prey.
321
322 We find evidence for positive selection across many genes involved in regulating
323 mitochondrial biogenesis, cellular respiration, and cardiovascular homeostasis. Komodo
324 dragons, along with other monitor lizards, have a high aerobic capacity and exercise endurance,
325 and our results reveal selective pressures on biochemical pathways that are likely to be the
326 source of this high aerobic capacity. Future genomic work on additional varanid species, and
327 other squamate outgroups, will test these hypotheses. These selective processes are consistent
328 with the increased oxidative capacity in python hearts after feeding [28]. Reptile muscle
329 mitochondria typically oxidize substrates at a much lower rate than mammals, partly based on
330 substrate-type use [56]. The findings that Komodo have experienced selection for several genes
331 encoding mitochondrial enzymes, including one involved in fatty acid metabolism, points
332 towards a more mammalian-like mitochondrial function. In addition to a clear indication of
333 adaptive muscle metabolism, we found positive selection for AGT, which encodes two potent
334 vasoactive and inotropic peptides with central roles in cardiovascular physiology. A compelling

335 hypothesis is that this positive selection is an important component in the ability of the
336 Komodo to rapidly increase blood pressure and cardiac output for attacks on prey, extended
337 periods of locomotion including inter-island swimming, and male-male combat during the
338 breeding season. Direct measures of cardiac function have not been made in Komodo dragons,
339 but in other varanid lizards, a large aerobic scope during exercise is associated with a large
340 factorial increase in cardiac output [57]. Overall, these cardiovascular genes suggest a profoundly
341 different cardiovascular and metabolic profile relative to other squamates, endowing the
342 Komodo dragon with unique physiological properties.
343
344 We also found evidence for positive selection across genes that regulate blood clotting.
345 Like other monitor lizards, the saliva of Komodo dragons contains anticoagulants. The extensive
346 positive selection on the genes encoding their coagulation system likely reflects that there is
347 selective pressure for Komodo dragons to evade the anticoagulant and hypotensive effects of
348 the saliva of conspecific rivals for food, territories, or mates. While all monitor lizards tested
349 contain anticoagulants in their saliva, the precise mechanism by which they act varies [12]. It is
350 likely that monitor lizards have evolved different types of adaptations that reflect the diversity
351 of their anticoagulants. Understanding how these systems have evolved has the potential to
352 further our understanding of the biology of thrombosis.
353
354 Varanids, including Komodo dragons, possess genotypic sex determination and share
355 ZZ/ZW sex chromosomes with other anguimorphan lizards [14,18]. Here, we were able to detail
356 the content of Z chromosome of *V. komodoensis*. The chromosome sequencing data provided

357 significant insights into the content of *V. komodoensis* Z chromosome. All scaffolds assigned to
358 Z chromosome were homologous to *A. carolinensis* chromosome 18 (ACA18) and to chicken
359 chromosome 28, as confirmed by comparison of blood transcriptome between sexes [18]. The
360 same syntenic blocks and genes appear to be implicated in different vertebrate lineages in sex
361 determination mechanisms [58]. In particular, the regions of sex chromosomes that are shared by
362 the common ancestor of varanids and several other lineages of anguimorphan lizards contain
363 the *amh* (anti-Müllerian hormone) gene [18], which plays a crucial role in the testis differentiation
364 pathway. Homologs of the *amh* gene are also strong candidates for being the sex-determining
365 genes in several lineages of teleost fishes and in monotremes [59–62].
366
367

368

369 **Materials and Methods**

370 *DNA isolation and processing for Bionano optical mapping*

371 Komodo dragon whole blood was used to extract high molecular weight genomic DNA
372 for genome mapping. Blood was centrifuged at 2000g for 2 minutes, plasma was removed, and
373 the sample was stored at 4°C. 2.5μl of blood was embedded in 100μl of agarose gel plug to give
374 ~7μg DNA/plug, using the BioRad CHEF Mammalian Genomic DNA Plug Kit (Bio-Rad
375 Laboratories, Hercules, CA, USA). Plugs were treated with proteinase K overnight at 50°C. The
376 plugs were then washed, melted, and then solubilized with GELase (Epicentre, Madison, WI,
377 USA). The purified DNA was subjected to four hours of drop-dialysis. DNA concentration was
378 determined using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA), and the quality
379 was assessed with pulsed-field gel electrophoresis.

380 The high molecular weight DNA was labeled according to commercial protocols using
381 the IrysPrep Reagent Kit (Bionano Genomics, San Diego, CA, USA). Specifically, 300 ng of
382 purified genomic DNA was nicked with 7 U nicking endonuclease Nb.BbvCI (NEB, Ipswich, MA,
383 USA) at 37°C for two hours in NEB Buffer 2. The nicked DNA was labeled with a fluorescent-
384 dUTP nucleotide analog using Taq polymerase (NEB) for one hour at 72°C. After labeling, the
385 nicks were repaired with Taq ligase (NEB) in the presence of dNTPs. The backbone of
386 fluorescently labeled DNA was stained with DNA stain (BioNano).

387

388 *DNA processing for 10x Genomics linked read sequencing*

389 High molecular weight genomic DNA extraction, sample indexing, and generation of
390 partition barcoded libraries were performed by 10x Genomics (Pleasanton, CA, USA) according
391 to the Chromium Genome User Guide and as published previously [64].
392
393 *Bionano mapping and assembly*
394 Using the Bionano Irys instrument, automated electrophoresis of the labeled DNA
395 occurred in the nanochannel array of an IrysChip (Bionano Genomics), followed by automated
396 imaging of the linearized DNA. The DNA backbone (outlined by YOYO-1 staining) and locations
397 of fluorescent labels along each molecule were detected using the Irys instrument's software.
398 The length and set of label locations for each DNA molecule defines an individual single-
399 molecule map. Raw Bionano single-molecule maps were de novo assembled into consensus
400 maps using the Bionano IrysSolve assembly pipeline (version 5134) with default settings, with
401 noise values calculated from the 10x Genomics Supernova assembly.
402
403 *10x Genomics sequencing and assembly*
404 The 10x Genomics barcoded library was sequenced on the Chromium machine, and the
405 raw reads were assembled using the company's Supernova software (version 1.0) with default
406 parameters. Output fasta files of the phased Supernova assemblies were generated in
407 pseudohap format.
408
409 *Merging datasets into a single assembly*

410 Sequencing and mapping data types were merged together as follows. First, Bionano
411 assembled contigs and the 10x Genomics assembly were combined using Bionano's hybrid
412 assembly tool with the -B2 -N2 options. SSPACE-LongRead (cite https://doi.org/10.1186/1471-
413 2105-15-211) was used in series with default parameters to scaffold the hybrid assembly using
414 PacBio reads, Nanopore reads, and unincorporated 10x Genomics Supernova scaffolds/contigs,
415 resulting in the final assembly.
416
417 *Assignment of scaffolds to chromosomes*
418 Isolation of *V. komodoensis* (VKO) chromosome-specific DNA pools as previously
419 described [14]. Briefly, fibroblast cultivation of a female *V. komodoensis* were obtained from
420 tissue samples of an early embryo of a captive individual. Chromosomes obtained by fibroblast
421 cultivation were sorted using a Mo-Flo (Beckman Coulter) cell sorter. Fifteen chromosome
422 pools were sorted in total. Chromosome-specific DNA pools were then amplified and labelled
423 by degenerate oligonucleotide primed PCR (DOP-PCR) and assigned to their respective
424 chromosomes by hybridization of labelled probes to metaphases. *V. komodoensis* chromosome
425 pools obtained by flow sorting were named according to chromosomes (e.g. majority of DNA of
426 VKO6/7 belong to chromosomes 6 and 7 of *V. komodoensis*). *V. komodoensis* pools for
427 macrochromosomes are each specific for one single pair of chromosomes, except for VKO6/7
428 and VKO8/7, which contain one specific chromosome pair each (pair 6 and pair 8, respectively),
429 plus a third pair which overlaps between the two of them (pair 7). For microchromosomes,
430 pools VKO9/10, VKO17/18/19, VKO11/12/W and VKO17/18/Z contained more than one
431 chromosome each, while the rest are specific for one single pair of microchromosomes. The W

432 and Z chromosomes are contained in pools VKO11/12/W and VKO17/18/Z, respectively,
433 together with two pairs of other microchromosomes each.

434 Chromosome-pool specific genetic material was amplified by GenomePlex® Whole
435 Genome Amplification (WGA) Kit (Sigma) following manufacturer protocols. DNA from all 15
436 chromosome pools was used to prepare Illumina sequencing libraries, which were
437 independently barcoded and sequenced 125 bp paired-end in a single Illumina Hiseq2500 lane.
438 Reads obtained from sequencing of flow-sorting-derived chromosome-specific DNA pools were
439 processed with the dopseq pipeline (https://github.com/lca-imcb/dopseq) [17,65]. Illumina
440 adapters and WGA primers were trimmed off by cutadapt v1.13 [66]. Then, pairs of reads were
441 aligned to the genome assembly of *V. komodoensis* using bwa mem [67]. Reads were filtered by
442 MAPQ ≥ 20 and length ≥ 20 bp, and aligned reads were merged into positions using pybedtools
443 0.7.10 [68,69]. Reference genome regions were assigned to specific chromosomes based on
444 distance between positions. Finally, several statistics were calculated for each scaffold.
445 Calculated parameters included: mean pairwise distance between positions on scaffold, mean
446 number of reads per position on scaffold, number of positions on scaffold, position
447 representation ratio (PRR) and p-value of PRR. PRR of each scaffold was used to evaluate
448 enrichment of given scaffold on chromosomes. PRR was calculated as ratio of positions on
449 scaffold to positions in genome divided by ratio of scaffold length to genome length. PRRs >1
450 correspond to enrichment, while PRRs <1 correspond to depletion. As the PRR value is
451 distributed lognormally, we use its logarithmic form for our calculations. To filter out only
452 statistically significant PRR values we used thresholds of logPRR >0 and its p-value <=0.01.

453 Scaffolds with logPRR > 0 were considered enriched in the given sample. If one scaffold was
454 enriched in several samples we chose highest PRR to assign scaffold as top sample.
455 We also assigned homology of *V. komodoensis* genome to genomes of *Anolis carolinensis*
456 (AnoCar2.0) and *Gallus gallus* (galGal3) generating alignment between genomes with LAST [70]
457 and subsequently using chaining and netting technique [71]. For LAST we used default scoring
458 matrix and parameters of 400 for gap existence cost, 30 for gap extension cost and 4500 for
459 minimum alignment score. For axtChain we used same distance matrix and default parameters
460 for other chain-net scripts.
461
462 *RNA sequencing*
463 RNA was extracted from heart tissue obtained from an adult male specimen that died of
464 natural causes. Trizol reagent was used to extract RNA following manufacturer's instructions.
465 RNAseq libraries were produced using a NuGen RNAseq v2 and Ultralow v2 kits, and sequenced
466 on an Illumina Nextseq 500.
467
468 *Genome annotation*
469 RepeatMasker was used to mask repetitive elements in the Komodo dragon genome
470 using the squamata repeat database as reference [15]. After masking repetitive elements,
471 protein-coding genes were annotated using the MAKER version 3.01.02 [72] pipeline, combining
472 protein homology information, assembled transcript evidence, and de novo gene predictions
473 from SNAP and Augustus version 3.3.1 [73]. Protein homology was determined by aligning
474 proteins from 15 reptile species (Table S10) to the Komodo dragon genome using exonerate

475 version 2.2.0 [74]. RNA-seq data was aligned to the Komodo genome with STAR version 2.6.0 [75]
476 and assembled into 900,722 transcripts with Trinity version 2.4.0 [19]. Protein domains were
477 determined using InterProScan version 5.31.70 [76]. Gene annotations from the MAKER pipeline
478 were filtered based on the strength of evidence for each gene, the length of the predicted
479 protein, and the presence of protein domains. Clusters of orthologous genes across 15 reptile
480 species were determined with OrthoFinder v2.0.0 [77]. A total of 284,107 proteins were clustered
481 into 16,546 orthologous clusters. In total, 96.4% of Komodo genes were grouped into
482 orthologous clusters. For estimating a species phylogeny only, protein sequences from *Mus*
483 *musculus* and *Gallus gallus* were added to the orthologous clusters with OrthoFinder. tRNAs
484 were annotated using tRNAscan-SE version 1.3.1[78], and other non-coding RNAs were annotated
485 using the Rfam database [79] and the Infernal software suite [80].
486
487 *Phylogenetic analysis*
488 A total of 2,752 single-copy orthologous proteins across 15 reptile species, including
489 *Varanus komodoensis, Shinisaurus crocodilurus, Ophisaurus gracilis, Anolis carolinensis, Pogona*
490 *vitticeps, Python molorus bivittatus, Eublepharis macularius, Gekko japonicus, Pelodiscus*
491 *sinensis, Chelonia mydas, Chrysemys picta bellii, Alligator sinensis, Alligator mississippiensis,*
492 *Gavialis gangeticus,* and *Crocodylus porosus*, along with the chicken *Gallus gallus* and mouse
493 *Mus musculus,* were each aligned using PRANK v.170427 [81] (Table S10). Aligned proteins were
494 concatenated into a supermatrix, and a species tree was estimated using IQ-TREE version
495 1.6.7.1 [82] with model selection across each partition [83] and 10,000 ultra-fast bootstrap
496 replicates [84].

497

498 *Gene family evolution analysis*

499 Gene family expansion and contraction analyses were performed with CAFE v4.2 [85] for
500 the squamate reptile lineage, with a constant gene birth and gene death rate assumed across
501 all branches.

502 Vomeronasal type 2 receptors were first identified in all species by containing the V2R
503 domain InterPro domain (IPR004073) [86]. To ensure that no V2R genes were missed, all proteins
504 were aligned against a set of representative V2R genes using BLASTp [87] with an e-value cutoff of
505 1e-6 and a bitscore cutoff of 200 or greater. Any genes passing this threshold were added to the
506 set of putative V2R genes. Transmembrane domains were identified in each putative V2R gene
507 with TMHMM v2.0 [88] and discarded if they did not contain 7 transmembrane domains in the C-
508 terminal region. Beginning at the start of the first transmembrane domain, proteins were
509 aligned with MAFFT v7.310 (auto alignment strategy) [89] and trimmed with trimAL (gappyout) [90].
510 A gene tree was constructed using IQ-TREE [82–84] with the JTT+ model of evolution with empirical
511 base frequencies and 10 FreeRate model parameters, and 10,000 bootstrap replicates. Genes
512 were discarded if they failed the IQ-TREE composition test.

513

514 *Positive selection analysis*

515 We analyzed 4,081 genes that were universal and single-copy across all squamate
516 lineages tested (*Varanus komodoensis, Shinisaurus crocodilurus, Ophisaurus gracilis, Anolis*
517 *carolinensis, Pogona vitticeps, Python molorus bivittatus, Eublepharis macularius,* and *Gekko*
518 *japonicus*) to test for positive selection (Table S8). An additional 2,040 genes that were

519 universal and single-copy across a subset of squamate species (*Varanus komodoensis, Anolis*
520 *carolinensis, Python molurus bivittatus*, and *Gekko japonicus*) were also analyzed (Table S8). We
521 excluded multi-copy genes from all positive selection analyses to avoid confounding from
522 incorrect paralogy inference. Proteins were aligned using PRANK [81] and codon alignments were
523 generated using PAL2NAL [91].
524 Positive selection analyses were performed with the branch-site model aBSREL using the
525 HYPHY framework [92,93]. For the 4,081 genes that were single-copy across all squamate lineages,
526 the full species phylogeny of squamates was used. For the 2,040 genes that were universal and
527 single-copy across a subset of species, a pruned tree containing only those taxa was used. We
528 discarded genes with unreasonably high dN/dS values across a small proportion of sites, as
529 those were false positives driven by low quality gene annotation in one or more taxa in the
530 alignment. We used a cutoff of dN/dS of less than 50 across 5% or more of sites, and a p-value
531 of less than 0.05 at the Komodo node. Each gene was first tested for positive selection only on
532 the Komodo branch. Genes undergoing positive selection in the Komodo lineage were then
533 tested for positive selection at all nodes in the phylogeny. This resulted in 201 genes being
534 under positive selection in the Komodo lineage (Table S9).
535
536
537 **Acknowledgements**
538 Special thanks from B.G.B. to John Romano for inspiration and historical information. We are
539 grateful to staff at Zoo Atlanta for care of Slasher and Rinca and help obtaining samples, Jim
540 Pether from Reptilandia zoo in Gran Canaria in the Canary Islands, for additional samples, and

541 R. Chadwick and N. Carli (Gladstone Genomics Core) for DNA and RNA-seq library preparation.
542 We also thank Kristina Giorda, Rabeea Abbas, Deanna Church (10x Genomics) for 10x Genomics
543 Chromium sequencing and Supernova assembly. This work was supported by institutional
544 funding from the Gladstone Institutes to B.G.B and K.S.P.; an NHLBI grant to K.S.P and B.G.B
545 (HL098179); the Younger Family gift to B.G.B.; an NHGRI grant to P.-Y.K. (R01 HG005946); NIH
546 training grants (T32 AR007175) to A.C.Y.M and (T32 HL007731) to Y.M. M.A. and M.R. were
547 supported by GACR 17-22141Y, M.R. was additionally supported by Charles University projects
548 PRIMUS/SCI/46 and Research Centre (204069).
549

550 Author contributions: A.L.L. did genome annotation and all comparative genomics analyses.
551 Y.Y.Y.L. led the sequencing and assembly efforts with Y.M. and A.C.Y.M. A.I. sequenced isolated
552 chromosomes with M.R under supervision of L.K., and assigned sequences with A.M., I.K, and
553 V.T. M.F. and V.O. contributed to genome assembly the genome in the lab of R.F. with C.C.
554 A.K.H. led the initial development of the project. W.L.E. initially assembled the transcriptomes
555 and annotated the genome. O.A.R. provided frozen tissue samples. J.M. collected specimen
556 blood. M.M. and M.F. isolated samples and obtained PacBio sequence in the lab of T.P. and C.C.
557 E.S. performed PacBio sequencing. J.W.H., J.M., and C.C. provided direction on Varanid
558 physiology. T.J. and C.C. provided direction on Komodo dragon ecology. P.-Y.K. coordinated the
559 genomics efforts. K.S.P. directed comparative genomic analysis. B.G.B. initiated and
560 coordinated the project. A.L., K.S.P. and B.G.B. wrote the paper with input from all authors.
561

562 Correspondence: benoit.bruneau@gladstone.ucsf.edu (B.G.B.),

563 katherine.pollard@gladstone.ucsf.edu (K.S.P.)

564

565 **Tables.**

566 **Table 1.** Genome statistics of the Komodo dragon genome.

Assembly size	1.51 Gb (1,508,391,850 bp)
Number of scaffolds	1,403
Minimum scaffold length	10 Kb
Maximum scaffold length	138 Mb
N50	29 Mb (29,129,838)
Number of protein-coding genes	18,462
GC content	44.04%

567

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

***V. komodoensis* chromosome**	**GGA homology**	**ACA homology**	**No. of 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

569 GGA homology: homology of scaffolds to *G. gallus* chromosomes; ACA homology: homology of

570 scaffolds to *A. carolinensis* chromosomes; Total length of assigned scaffolds (bp): size in base

571 pairs of the sum of all scaffolds for each chromosome.

572
573

574

575

576 **Figure legends.**

577 Figure 1. (A) Komodo dragons (left, Slasher; right, Rinca) sampled for DNA in this study. Photos
578 courtesy of Adam K Thompson/Zoo Atlanta. (B) Genome assembly workflow. Two separate *de*
579 *novo* assemblies were generated with 10x genomics and Bionano physical mapping data and
580 merged into an intermediate hybrid assembly. Long reads from PacBio and Oxford Nanopore
581 MinIon were used to scaffold the hybrid assembly into a final version.

582

583 Figure 2. **Estimated species phylogeny of 15 non-avian reptiles species and 2 additional**
584 **vertebrates.** Maximum likelihood phylogeny was constructed from 2,752 single-copy
585 orthologous proteins. Support values from 10,000 bootstrap replicates are shown. All images
586 obtained from PhyloPic.org.

587

588 Figure 3. **Type 2 vomeronasal receptors have expanded in Komodo dragons and several other**
589 **squamate reptiles.** (A) Type 2 vomeronasal gene counts in squamate reptiles. (B) Unrooted
590 gene phylogeny of 1,093 vomeronasal Type 2 receptor transmembrane domains across
591 squamate reptiles. The topology of the tree supports a gene expansion ancestral to squamates
592 (i.e., clades containing representatives from all species) as well as multiple species-specific
593 expansions through gene duplication events (i.e., clades containing multiple genes from one
594 species). Branches with bootstrap support less than 60 are collapsed. Colors correspond to
595 species in (A). Clades containing genes from a single species are collapsed.

596

597 Figure 4. **Gene clusters of Type 2 vomeronasal receptors evolved through gene duplication.**
598 (A) Genes in a cluster of vomeronasal Type 2 receptor genes in the Komodo dragon genome

599 containing 14 V2R genes. Pink genes are V2R genes and gray genes are non-V2R genes. Gene
600 labels correspond to labels in (B). (B) Unrooted phylogeny of 151 vomeronasal Type 2 receptor
601 genes in Komodo. As most of the genes in this gene cluster group together in a gene phylogeny
602 of all Komodo dragon V2R genes, it is likely that this cluster evolved through gene duplication
603 events. Branches with bootstrap support less than 80 are collapsed. Clades without genes in
604 this V2R gene cluster are collapsed. Genes in the V2R cluster are colored pink and labeled as in
605 (A).
606
607 Figure 5. **Mitochondrial genes under positive selection in the Komodo dragon**. Genes in the
608 Komodo dragon genome under positive selection include components of the electron transport
609 chain, regulators of transcription, regulators of translation, and fatty acid beta-oxidation.
610
611 **Supplemental figure legends.**
612 Figure S1. Percent identities of single-copy orthologs between the Komodo dragon and the
613 green anole and the Komodo dragon and the Chinese crocodile lizard.
614
615 Figure S2. Positive selection on genes encoding structural proteins in the electron transport
616 chain. Dark gray genes were not tested for positive selection due to either missing data in one
617 or more species or difficulty resolving ortholog/paralog relationships. Pink genes have
618 signatures of positive selection, and light gray genes did not have signatures of positive
619 selection. Figure modified from WikiPathways [63].
620

621 **Supplemental file descriptions.**

622 Table S1. Genome statistics for non-avian reptiles used in this study.

623 Table S2. Repetitive elements in the Komodo dragon genome.

624 Table S3. Read statistics for chromosomal anchoring.

625 Table S4. Scaffold assignment and homologies of Komodo dragon scaffolds to green anole and

626 chicken chromosomes.

627 Table S5. Number of reads, scaffolds, and positions assigned to chromosomes.

628 Table S6. Number of V1R/V2R genes across non-avian reptiles.

629 Table S7. V2R gene clusters in the Komodo dragon genome.

630 Table S8. Genes assayed for positive selection in the Komodo dragon genome.

631 Table S9. Positively selected genes in the Komodo dragon genome.

632 Table S10. Sources and versions of genomes used for phylogenetic and comparative methods.

633 **References**

634 1. Chapman, A. D. *Numbers of Living Species in Australia and the World*. (Australian

635 Biological Resources Study, 2009).

636 2. Collar, D. C., Schulte, J. A. & Losos, J. B. Evolution of extreme body size disparity in

637 monitor lizards (Varanus). *Evolution (N. Y).* **65**, 2664–2680 (2011).

638 3. Jensen, B., Wang, T., Christoffels, V. M. & Moorman, A. F. M. Evolution and development

639 of the building plan of the vertebrate heart. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**,

640 783–794 (2013).

641 4. Clemente, C. J., Withers, P. C. & Thompson, G. G. Metabolic rate and endurance capacity

642 in Australian varanid lizards (Squamata: Varanidae: Varanus). *Biol. J. Linn. Soc.* **97**, 664–

643 676 (2009).

644 5. Burggren, W. & Johansen, K. Ventricular Haemodynamics in the Monitor Lizard Varanus
645 Exanthematicus: Pulmonary and Systemic Pressure Separation. *J. Exp. Biol.* **96**, (1982).

646 6. Ishimatsu, A., Hicks, J. W. & Heisler, N. Analysis of intracardiac shunting in the lizard,
647 Varanus niloticus: a new model based on blood oxygen levels and microsphere
648 distribution. *Respir. Physiol.* **71**, 83–100 (1988).

649 7. King, D. & Green, B. *Goannas: The Biology of Varanid Lizards*. (University of New South
650 Wales, 1998).

651 8. Auffenberg, W. *The Behavioral Ecology of the Komodo Monitor*. (University Presses of
652 Florida, 1981).

653 9. Green, B., King, D., Braysher, M. & Saim, A. Thermoregulation, water turnover and
654 energetics of free-living komodo dragons, Varanus komodoensis. *Comp. Biochem.*
655 *Physiol. Part A Physiol.* **99**, 97–101 (1991).

656 10. Purwandana, D. *et al.* Ecological allometries and niche use dynamics across Komodo
657 dragon ontogeny. *Sci. Nat.* **103**, 27 (2016).

658 11. Fry, B. G. *et al.* A central role for venom in predation by Varanus komodoensis (Komodo
659 Dragon) and the extinct giant Varanus (Megalania) priscus. *Proc. Natl. Acad. Sci. U. S. A.*
660 **106**, 8969–8974 (2009).

661 12. Koludarov, I. *et al.* Enter the Dragon: The Dynamic and Multifunctional Evolution of
662 Anguimorpha Lizard Venoms. *Toxins* **9**, (2017).

663 13. Johnson Pokorná, M. *et al.* First Description of the Karyotype and Sex Chromosomes in
664 the Komodo Dragon (Varanus komodoensis). *Cytogenet. Genome Res.* **148**, 284–291

665 (2016).

666 14. Iannucci, A. *et al.* Isolating Chromosomes of the Komodo Dragon: New Tools for
667 Comparative Mapping and Sequence Assembly. *Cytogenet. Genome Res.* **157**, 42–50
668 (2019).

669 15. Smit, A., Hubley, R. & Green, P. Repeatmasker Open-4.0. (2013). Available at:
670 http://www.repeatmasker.org. (Accessed: 10th January 2015)

671 16. Gao, J. *et al.* Sequencing, de novo assembling, and annotating the genome of the
672 endangered Chinese crocodile lizard Shinisaurus crocodilurus. *Gigascience* **6**, 1–6 (2017).

673 17. Kichigin, I. G. *et al.* Evolutionary dynamics of Anolis sex chromosomes revealed by
674 sequencing of flow sorting-derived microchromosome-specific DNA. *Mol. Genet.*
675 *Genomics* **291**, 1955–1966 (2016).

676 18. Rovatsos, M., Rehák, I., Velenský, P. & Kratochvíl, L. Shared ancient sex chromosomes in
677 varanids, beaded lizards and alligator lizards. *Mol. Biol. Evol.* (2019).
678 doi:10.1093/molbev/msz024

679 19. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the
680 Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–512 (2013).

681 20. Pyron, R., Burbrink, F. T. & Wiens, J. J. A phylogeny and revised classification of
682 Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **13**, 93 (2013).

683 21. Welton, L. J., Travers, S. L., Siler, C. D. & Brown, R. M. Integrative taxonomy and
684 phylogeny-based species delimitation of Philippine water monitor lizards (Varanus
685 salvator Complex) with descriptions of two new cryptic species. *Zootaxa* **3881**, 201
686 (2014).

687 22. Silva, L. & Antunes, A. Vomeronasal Receptors in Vertebrates and the Evolution of
688 Pheromone Detection. *Annu. Rev. Anim. Biosci.* **5**, 353–370 (2017).
689 23. Stoddart, D. M. *The Ecology of Vertebrate Olfaction*. (Springer Netherlands, 1980).
690 24. Mason, R. T. & Parker, M. R. Social behavior and pheromonal communication in reptiles.
691 *J. Comp. Physiol. A* **196**, 729–749 (2010).
692 25. Green, R. E. *et al.* Three crocodilian genomes reveal ancestral patterns of evolution
693 among archosaurs. *Science (80-.).* **346**, 1254449–1254449 (2014).
694 26. Brykczynska, U., Tzika, A. C., Rodriguez, I. & Milinkovitch, M. C. Contrasted evolution of
695 the vomeronasal receptor repertoires in mammals and squamate reptiles. *Genome Biol.*
696 *Evol.* **5**, 389–401 (2013).
697 27. Yang, H., Shi, P., Zhang, Y. & Zhang, J. Composition and evolution of the V2r vomeronasal
698 receptor gene repertoire in mice and rats. *Genomics* **86**, 306–315 (2005).
699 28. Riquelme, C. A. *et al.* Fatty Acids Identified in the Burmese Python Promote Beneficial
700 Cardiac Growth. *Science (80-.).* **334**, 528 LP-531 (2011).
701 29. Falkenberg, M. *et al.* Mitochondrial transcription factors B1 and B2 activate transcription
702 of human mtDNA. *Nat. Genet.* **31**, 289–294 (2002).
703 30. Cotney, J., McKay, S. E. & Shadel, G. S. Elucidation of separate, but collaborative
704 functions of the rRNA methyltransferase-related human mitochondrial transcription
705 factors B1 and B2 in mitochondrial biogenesis reveals new insight into maternally
706 inherited deafness. *Hum. Mol. Genet.* **18**, 2670–2682 (2009).
707 31. Cho, Y., Hazen, B. C., Russell, A. P. & Kralli, A. Peroxisome Proliferator-activated Receptor
708 γ Coactivator 1 (PGC-1)- and Estrogen-related Receptor (ERR)-induced Regulator in

709 Muscle 1 (PERM1) Is a Tissue-specific Regulator of Oxidative Capacity in Skeletal Muscle
710 Cells. *J. Biol. Chem.* **288**, 25207–25218 (2013).
711 32. Cho, Y. *et al.* Perm1 enhances mitochondrial biogenesis, oxidative capacity, and fatigue
712 resistance in adult skeletal muscle. *FASEB J.* **30**, 674–687 (2016).
713 33. Zhao, S. *et al.* Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science*
714 *(80-.).* **327**, 1000–1004 (2010).
715 34. Brzezniak, L. K., Bijata, M., Szczesny, R. J. & Stepien, P. P. Involvement of human ELAC2
716 gene product in 3' end processing of mitochondrial tRNAs. *RNA Biol.* **8**, 616–626 (2011).
717 35. Holzmann, J. *et al.* RNase P without RNA: Identification and Functional Reconstitution of
718 the Human Mitochondrial tRNA Processing Enzyme. *Cell* **135**, 462–474 (2008).
719 36. Lee, K.-W. & Bogenhagen, D. F. Assignment of 2'-O-Methyltransferases to Modification
720 Sites on the Mammalian Mitochondrial Large Subunit 16 S Ribosomal RNA (rRNA). *J. Biol.*
721 *Chem.* **289**, 24936–24942 (2014).
722 37. Forrester, S. J. *et al.* Angiotensin II Signal Transduction: An Update on Mechanisms of
723 Physiology and Pathophysiology. *Physiol. Rev.* **98**, 1627–1738 (2018).
724 38. Kim, S. & Iwao, H. Molecular and Cellular Mechanisms of Angiotensin II-Mediated
725 Cardiovascular and Renal Diseases. *Pharmacol. Rev.* **52**, 11 LP-34 (2000).
726 39. WILSON, J. X. The Renin-Angiotensin System in Nonmammalian Vertebrates. *Endocr. Rev.*
727 **5**, 45–61 (1984).
728 40. Fournier, D., Luft, F. C., Bader, M., Ganten, D. & Andrade-Navarro, M. A. Emergence and
729 evolution of the renin–angiotensin–aldosterone system. *J. Mol. Med.* **90**, 495–508
730 (2012).

731 41. Mueller, C. A., Eme, J., Tate, K. B. & Crossley, D. A. Chronic captopril treatment reveals
732 the role of ANG II in cardiovascular function of embryonic American alligators (Alligator
733 mississippiensis). *J. Comp. Physiol. B* **188**, 657–669 (2018).
734 42. Antl, M. *et al.* IRAG mediates NO/cGMP-dependent inhibition of platelet aggregation and
735 thrombus formation. *Blood* **109**, 552–559 (2007).
736 43. Puetz, J. & Boudreaux, M. K. Evaluation of the gene encoding calcium and diacylglycerol
737 regulated guanine nucleotide exchange factor I (CalDAG-GEFI) in human patients with
738 congenital qualitative platelet disorders. *Platelets* **23**, 401–403 (2012).
739 44. Bezman, N. A. *et al.* Requirements of SLP76 tyrosines in ITAM and integrin receptor
740 signaling and in platelet function in vivo. *J. Exp. Med.* **205**, 1775–88 (2008).
741 45. Israels, S. & McMillan-Ward, E. CD63 modulates spreading and tyrosine phosphorylation
742 of platelets on immobilized fibrinogen. *Thromb. Haemost.* **93**, 311–318 (2005).
743 46. Cooper, D. N., Millar, D. S., Wacey, A., Pemberton, S. & Tuddenham, E. G. Inherited factor
744 X deficiency: molecular genetics and pathophysiology. *Thromb. Haemost.* **78**, 161—172
745 (1997).
746 47. Takahashi, N., Takahashi, Y. & Putnam, F. W. Primary structure of blood coagulation
747 factor XIIIa (fibrinoligase, transglutaminase) from human placenta. *Proc. Natl. Acad. Sci.*
748 **83**, 8019 LP-8023 (1986).
749 48. Mosesson, M. W. The roles of fibrinogen and fibrin in hemostasis and thrombosis. *Semin.*
750 *Hematol.* **29**, 177—188 (1992).
751 49. Halpern, M. Nasal chemical senses in reptiles: structure and function. Pp 423–523in Gans
752 C, Crews D (eds) Biology of the Reptilia, Vol. 18, Brain. *Horm. Behav. Chicago/IL Univ.*

753 *Chicago Press Google Sch.* (1992).

754 50. Martin, J. & Lopez, P. Chemoreception, symmetry and mate choice in lizards. *Proc. R. Soc.*
755 *B Biol. Sci.* **267**, 1265–1269 (2000).

756 51. Baeckens, S., Martín, J., García-Roa, R. & van Damme, R. Sexual selection and the
757 chemical signal design of lacertid lizards. *Zool. J. Linn. Soc.* **183**, 445–457 (2018).

758 52. van Damme, R., Bauwens, D., Thoen, C., Vanderstighelen, D. & Verheyen, R. F. Responses
759 of Naive Lizards to Predator Chemical Cues. *J. Herpetol.* **29**, 38 (1995).

760 53. van Damme, R. & Castilla, A. M. Chemosensory predator recognition in the lizard
761 Podarcis hispanica: effects of predation pressure relaxation. *J. Chem. Ecol.* **22**, 13–22
762 (1996).

763 54. Cooper, W. E. Correlated evolution of prey chemical discrimination with foraging, lingual
764 morphology and vomeronasal chemoreceptor abundance in lizards. *Behav. Ecol.*
765 *Sociobiol.* **41**, 257–265 (1997).

766 55. Cooper, W. Tandem evolution of diet and chemosensory responses in snakes. *Amphibia-*
767 *Reptilia* **29**, 393–398 (2008).

768 56. Hulbert, A. J. & Else, P. L. Evolution of mammalian endothermic metabolism:
769 mitochondrial activity and cell composition. *Am. J. Physiol. Integr. Comp. Physiol.* **256**,
770 R63–R69 (1989).

771 57. Gleeson, T. T., Mitchell, G. S. & Bennett, A. F. Cardiovascular responses to graded activity
772 in the lizards Varanus and Iguana. *Am. J. Physiol. Integr. Comp. Physiol.* **239**, R174–R179
773 (1980).

774 58. Marshall Graves, J. A. & Peichel, C. L. Are homologies in vertebrate sex determination

775 due to shared ancestry or to limited options? *Genome Biol.* **11**, 205 (2010).

776 59. Hattori, R. S. *et al.* A Y-linked anti-Müllerian hormone duplication takes over a critical role
777 in sex determination. *Proc. Natl. Acad. Sci.* **109**, 2955 LP-2959 (2012).

778 60. Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals.
779 *Nature* **508**, 488 (2014).

780 61. Bej, D. K., Miyoshi, K., Hattori, R. S., Strüssmann, C. A. & Yamamoto, Y. A Duplicated,
781 Truncated amh Gene Is Involved in Male Sex Determination in an Old World Silverside.
782 *G3 Genes|Genomes|Genetics* **7**, 2489–2495 (2017).

783 62. Ieda, R. *et al.* Identification of the sex-determining locus in grass puffer (Takifugu
784 niphobles) provides evidence for sex-chromosome turnover in a subset of Takifugu
785 species. *PLoS One* **13**, e0190635 (2018).

786 63. Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging
787 metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667 (2018).

788 64. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of
789 diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).

790 65. Makunin, A. I. *et al.* Contrasting origin of B chromosomes in two cervids (Siberian roe
791 deer and grey brocket deer) unravelled by chromosome-specific DNA sequencing. *BMC*
792 *Genomics* **17**, 618 (2016).

793 66. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing
794 reads. *EMBnet.journal* **17**, 10 (2011).

795 67. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
796 (2013).

797 68. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
798 features. *Bioinformatics* **26**, 841–2 (2010).
799 69. Quinlan, A. R., Pedersen, B. S. & Dale, R. K. Pybedtools: a flexible Python library for
800 manipulating genomic datasets and annotations. *Bioinformatics* **27**, 3423–3424 (2011).
801 70. Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. Adaptive seeds tame genomic
802 sequence comparison. *Genome Res.* (2011). doi:10.1101/gr.113985.110
803 71. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron:
804 Duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl.*
805 *Acad. Sci.* **100**, 11484–11489 (2003).
806 72. Cantarel, B. L. *et al.* MAKER: an easy-to-use annotation pipeline designed for emerging
807 model organism genomes. *Genome Res.* **18**, 188–96 (2008).
808 73. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes
809 that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465-7 (2005).
810 74. Slater, G. & Birney, E. Automated generation of heuristics for biological sequence
811 comparison. *BMC Bioinformatics* **6**, 31 (2005).
812 75. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
813 (2013).
814 76. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
815 *Bioinformatics* **30**, 1236–1240 (2014).
816 77. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome
817 comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157
818 (2015).

819 78. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA
820 genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–64 (1997).
821 79. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A. & Eddy, S. R. Rfam: an RNA
822 family database. *Nucleic Acids Res.* **31**, 439–41 (2003).
823 80. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches.
824 *Bioinformatics* **29**, 2933–5 (2013).
825 81. Löytynoja, A. Phylogeny-aware alignment with PRANK. in *Methods in molecular biology*
826 *(Clifton, N.J.)* **1079**, 155–170 (2014).
827 82. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective
828 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**,
829 268–274 (2015).
830 83. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S.
831 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**,
832 587–589 (2017).
833 84. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:
834 Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
835 85. Han, M. V., Thomas, G. W. C., Lugo-Martinez, J. & Hahn, M. W. Estimating Gene Gain and
836 Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3.
837 *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
838 86. Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to
839 protein sequence annotations. *Nucleic Acids Res.* (2018). doi:10.1093/nar/gky1100
840 87. 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).

88. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. . Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. *J. Mol. Biol.* **305**, 567–580 (2001).

89. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

90. 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).

91. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, (2006).

92. 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**, 1342–53 (2015).

93. Pond, S. L. K., Frost, S. D. W. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679 (2005).

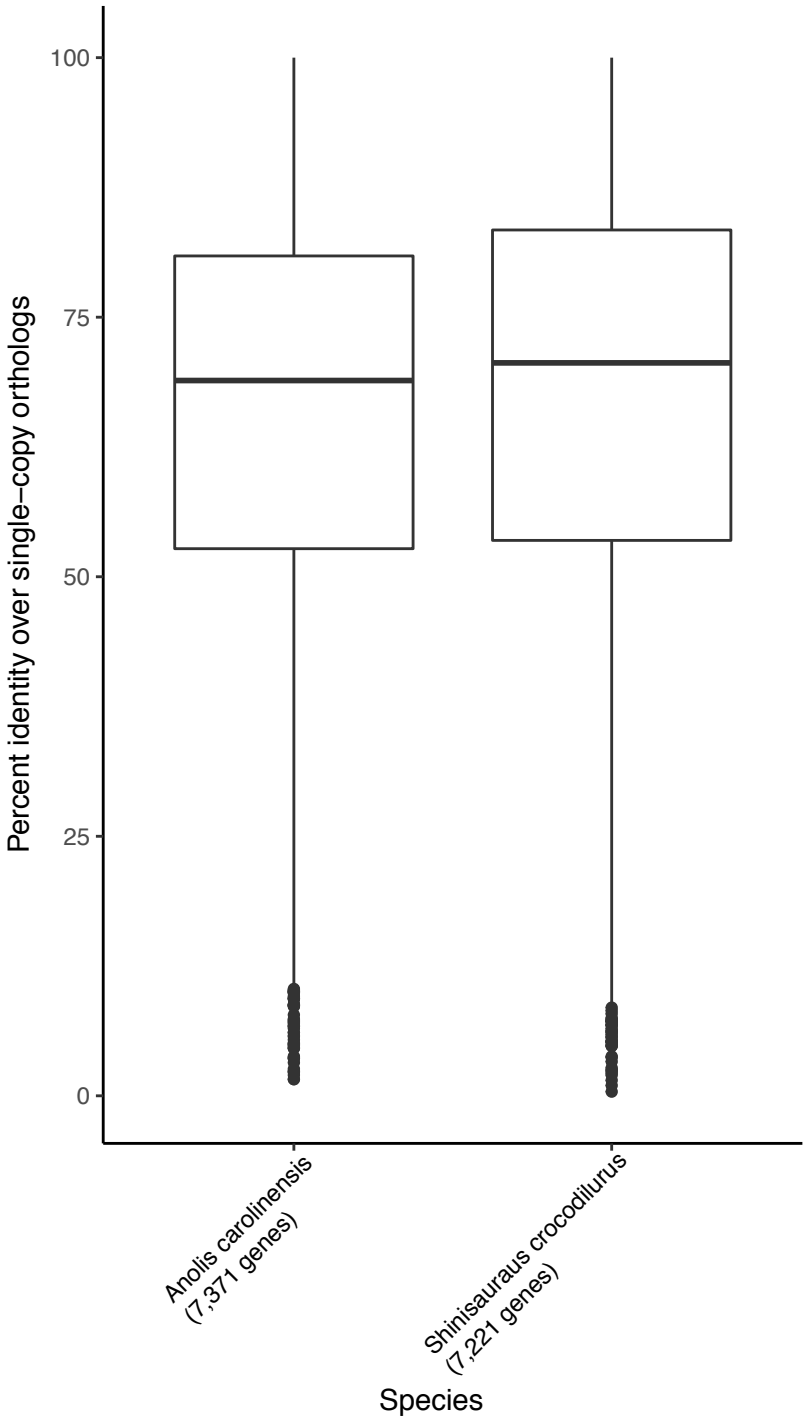

Figure S1. Percent identities of single-copy orthologs between the Komodo dragon and the green anole and the Komodo dragon and the Chinese crocodile lizard.

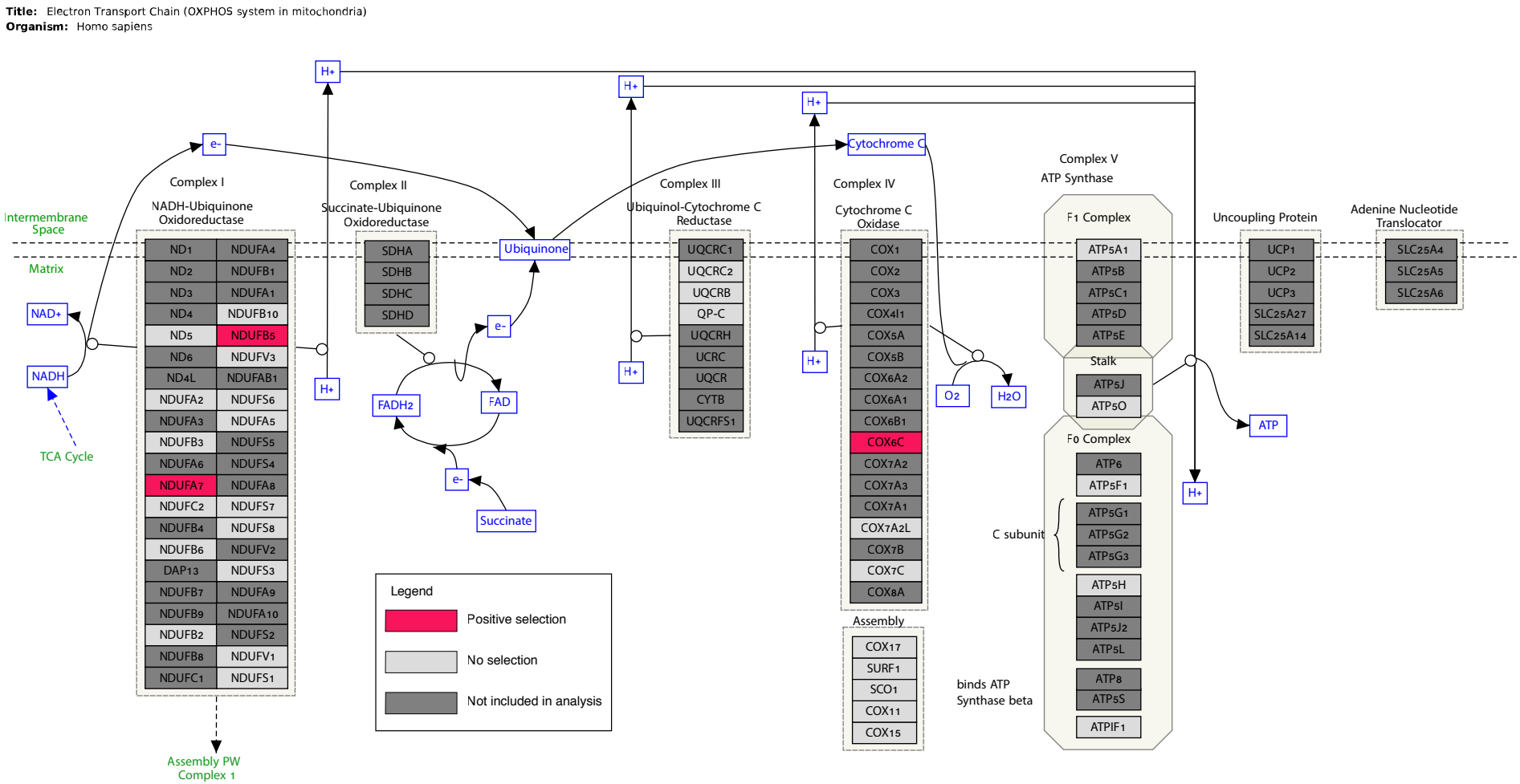

Figure S2. Positive selection on genes encoding structural proteins in the electron transport chain. Dark gray genes were not tested for positive selection due to either missing data in one or more species or difficulty resolving ortholog/paralog relationships. Pink genes have signatures of positive selection, and light gray genes did not have signatures of positive selection. Figure modified from WikiPathways 63.