



Genome-wide analyses reveal drivers of penguin diversification

Juliana Vianna, Flávia Fernandes, María José Frugone, Henrique Figueiró, Luis Pertierra, Daly Noll, Ke Bi, Cynthia Wang-Claypool, Andrew Lowther, Patricia Parker, et al.

► To cite this version:

Juliana Vianna, Flávia Fernandes, María José Frugone, Henrique Figueiró, Luis Pertierra, et al.. Genome-wide analyses reveal drivers of penguin diversification. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117 (36), pp.22303-22310. 10.1073/pnas.2006659117 . hal-02999306

HAL Id: hal-02999306

<https://hal.science/hal-02999306>

Submitted on 17 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 BIOLOGICAL SCIENCES

2

3

4 Title: Genome-wide analyses reveal drivers of penguin diversification

Authors: Juliana A. Vianna^{1*}, Flávia A. N. Fernandes², María José Frugone³, Henrique Figueiró⁴, Luis R. Pertierra⁵, Daly Noll^{1,3}, Ke Bi⁶, Cynthia Y. Wang-Claypool⁶, Andrew Lowther⁷, Patricia Parker⁸, Celine Le Bohec^{9,10}, Francesco Bonadonna¹¹, Barbara Wienecke¹², Pierre Pistorius¹³, Antje Steinfurth¹⁴, Christopher Burridge¹⁵, Gisele P. M. Dantas¹⁶, Elie Poulin³, W. Brian Simison¹⁷, Jim Henderson¹⁷, Eduardo Eizirik⁴, Mariana F. Nery², Rauri C. K. Bowie^{6*}

11 Affiliations:

¹⁵ ²Universidade Estadual de Campinas (UNICAMP), Sao Paulo, Brazil.

¹⁶ ³Instituto de Ecología y Biodiversidad (IEB), Departamento de Ciencias Ecológicas,
¹⁷ Facultad de Ciencias, Universidad de Chile, Santiago, Chile.

¹⁸Pontifícia Universidade Católica do Rio Grande do Sul, Escola de Ciências, Brazil.

¹⁹ ⁵Museo Nacional de Ciencias Naturales, Departamento de Ecología Evolutiva, CSIC,
²⁰ C/José Gutiérrez Abascal, 2, 28006 Madrid, Spain.

21 ⁶Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life
22 Science Building, University of California, Berkeley, Berkeley, CA 94720, USA.

²³Norwegian Polar Institute, N-9297 Tromsø, Norway.

24 ⁸University of Missouri St Louis and Saint Louis Zoo, One University Blvd., St. Louis,
25 MO, 63121-4400, USA.

²⁶ ⁹Université de Strasbourg, CNRS, IPHC UMR 7178, 23 rue Becquerel, F-67000
²⁷ Strasbourg, France.

28 ¹⁰Centre Scientifique de Monaco, Département de Biologie Polaire, 8 quai Antoine 1er,
29 MC 98000 Monaco, Principality of Monaco.

30 ¹¹CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry
31 Montpellier, EPHE, 1919 route de Mende, 34293 Montpellier cedex 5, France.

32 ¹²Australian Antarctic Division, 203 Channel Highway Kingston, Tasmania 7050,
33 Australia.

34 ¹³DST/NRF Centre of Excellence at the Percy FitzPatrick Institute for African
35 Ornithology, Department of Zoology, Nelson Mandela University, Port Elizabeth, South
36 Africa.

37 ¹⁴RSPB Centre for Conservation Science, Royal Society for the Protection of Birds,
38 David Attenborough Building, Pembroke Street, Cambridge, Cambridgeshire, CB2 3QZ,
39 United Kingdom.

40 ¹⁵University of Tasmania, School of Natural Sciences, Private Bag 55, Room 320a Life
41 Sciences Building, Hobart, Tasmania 7001, Australia.

42 ¹⁶Pontifícia Universidade Católica de Minas Gerais, PPG in Vertebrate Zoology, Belo
43 Horizonte, Brazil.

44 ¹⁷Center for Comparative Genomics, California Academy of Sciences, San Francisco,
45 CA, 94118, USA.

46

47

48 **Author contributions:** J.A.V., F.A.N.F., M.N., R.C.K.B. conceived and designed the
49 study and contributed to preparing the manuscript. H.F., K.B., E.E., B.S. and J.H.
50 contributed bioinformatics. L.P. developed the ecological niche models. M.J.F., D.N.,
51 and E.P. facilitated analyses and contributed to the interpretation of results. M.A.L.,
52 P.P., C.L.B., F.B., C.G., B.W., P.P., A.S., C.B., G.P.M.D, and C.W. undertook field work,
53 provided samples, and contributed to the manuscript. All co-authors revised and
54 commented on the manuscript and approved the final version;

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

56 **Corresponding authors:**
57 Juliana A. Vianna, Pontificia Universidad Católica de Chile, Facultad de Agronomía e
58 Ingeniería Forestal, Departamento de Ecosistemas y Medio Ambiente, Vicuña Mackenna
59 4860, Macul, Santiago, Chile, phone 56-988199882, jvianna@uc.cl
60 Rauri C. K. Bowie, Museum of Vertebrate Zoology and Department of Integrative Biology,
61 3101 Valley Life Science Building, University of California, Berkeley, CA 94720, USA.
62 Phone 510-643-1617, bowie@berkeley.edu
63
64 **Keywords:** Penguin, Antarctica, Genome, Ancestral niche, Ancestral distribution.

65 ABSTRACT

66

67 Extant penguins are the only flightless family of diving birds. They comprise at least 18
68 modern species, distributed from polar to tropical environments in the Southern
69 Hemisphere. The history of their diversification and adaptation to these diverse
70 environments remains controversial. We employed 22 new genomes from 18 penguin
71 species to reconstruct the order, timing, and location of their diversification, track changes
72 in thermal niche through time, and test for associated adaptation across the genome. The
73 penguin crown-group originated during the Miocene in New Zealand and Australia, and
74 not in Antarctica as previously thought. *Aptenodytes* appears as sister taxon of all other
75 penguin species. We show that lineage diversification in penguins was driven by changing
76 climatic conditions and by the development of the Antarctica Circumpolar Current (ACC).
77 Changes in thermal niche were accompanied by adaptations in genes that govern
78 thermoregulation and oxygen metabolism. Our reconstruction of changes in ancestral
79 effective population sizes (N_e) recovered three different demographic trajectories in
80 deeper time, the most common (11/18 penguin species) being an increased N_e between
81 0.04 and 0.07 Mya; N_e drops precipitously after the Last Glacial Maximum (LGM) most
82 likely as a consequence of the overall decline in marine productivity following the last
83 glaciation.

84

85 Significance Statement

86 Penguins have long been of interest to the scientific community. However, their
87 evolutionary history is not well understood. Using genomes we investigate macro- and
88 microevolutionary drivers of penguin diversification. Crown-group penguins diverged in
89 the early Miocene in the Australia/New Zealand region. We found that *Aptenodytes*
90 (emperor and king penguins) form a separated clade, sister of all other penguin lineages,
91 in contrast with previous studies that placed them as sister taxa of *Pygoscelis*. Analyses
92 suggest that penguins first radiated from temperate environments and then occupied cold
93 Antarctic waters. Although the onset of the ACC (35 Mya) was suggested to have a role
94 on penguins diversification, our findings of more recent time estimates of divergence

95 suggest that the intensification of the ACC (11.6 Mya) may have promoted their
96 diversification and geographic expansion.

97

98 **Introduction**

99

100 Few organisms have been as successful at colonizing the globe as seabirds, a large
101 ecological assemblage of oceanic and nearshore species that undergo some of the most
102 remarkable foraging and migratory journeys on Earth (1, 2). Despite their ubiquitous
103 presence, surprisingly little is known about the mechanisms that spurred their
104 diversification and allowed for their adaptation to diverse and often dynamic oceanic
105 habitats.

106 As the only living clade of flightless diving birds, penguins (order Sphenisciformes)
107 occupy both terrestrial and marine habitats. They forage across a wide range of ocean
108 temperatures and depths, from Antarctic to tropical waters (3). Our understanding of
109 penguin diversification and adaptation is hampered by disagreements about their
110 phylogenetic relationships (4–10) and the chronology of their radiation. When these
111 estimations are made using few genetic markers from different sections of the genome,
112 discordant results are to be expected as genomic regions vary in their mutation rates and
113 evolutionary history, including unknown patterns of gene introgression when different
114 species hybridize (11). Accordingly, estimates of ordinal crown-age vary between the
115 Miocene and Eocene (9.9–47.6 million years ago [Mya]) (4, 5, 8, 9). The earliest crown-
116 group fossil dates to the late Miocene (12).

117 In addition, reconstructions of ancestral distributions and climatic niches are critical
118 to extend our knowledge of penguin diversification, but existing hypotheses are
119 conflicting: (i) an Antarctic origin, with later expansion towards warmer areas (5) or (ii) a
120 sub-Antarctic origin, with subsequent colonization of Southern Ocean islands and
121 Antarctica (7, 13). Testing these alternative hypotheses on a broad scale requires
122 accurate knowledge of the evolutionary history of penguins, including branching patterns
123 and timing of diversification.

124 The extent of ice and changes in the currents of the Southern Ocean during
125 repeated glacial cycles likely played a significant role in structuring penguin populations
126 and lineage diversification (14–16). Biogeographic boundaries in the Southern Ocean,
127 particularly the Antarctic Polar Front (APF) and the sub-tropical Front (STF), serve as
128 barriers to dispersal of some penguin species (14, 15, 17, 18). Differences in abiotic (e.g.,
129 temperature and salinity) and biotic (e.g., types of food resources) variables on either side
130 of these two fronts may promote local adaptation and enable niche divergence among
131 penguins (19). Furthermore, associated currents have varied significantly over time in
132 latitude and strength in response to changing global circulation patterns (20). These
133 changes have been implicated in the colonization, isolation, and local extinction of some
134 penguin populations and species (4, 9).

135 We report here the reconstruction of the history of penguin diversification and
136 adaptation using 22 newly sequenced genomes representing 18 extant species and one
137 outgroup. We determined phylogenetic relationships and biogeographic histories, and
138 consequently the divergence of niche space through time. Finally, we also assessed the
139 extent of introgression between species that may have contributed to previous
140 disagreements on phylogenetic reconstructions, and adaptations of penguins across
141 environmental gradients.

142

143 **Results**

144 The penguin and petrel outgroup genomes were sequenced to ~30x coverage (Tables
145 S1–S3). All, but one genome sequences contained >90% of the conserved single-copy
146 genes identified across avian genomes (complete BUSCOs >90%; Table S4, Figs. S1–
147 4). From the assembled genomes we extracted 23,108 loci: Ultra Conserved Elements
148 (UCE: 4,057 loci), coding sequences (CDS: 11,011), and introns (8,040). We applied
149 species tree and concatenation methods to construct a phylogenetic hypothesis for extant
150 penguins using all loci together, and for each set of loci independently (Figs. 1, S5–8,
151 Table S5–S6).

152

153 **Phylogenetic history and patterns of introgression of crown-group penguins**

154 All analyses supported a similar phylogenetic hypothesis, all placing *Aptenodytes* as
155 sister taxon of all other extant penguin species (Fig. 1), distinct from other clades
156 comprising the genera *Pygoscelis*, *Spheniscus+Eudyptula*, and *Megadyptes+Eudyptes*.
157 The phylogeny based on mitogenomes was similar to that retrieved from the genomic
158 datasets, with a few minor differences in *Eudyptes* penguins that are likely explained by
159 genome-wide introgression among closely related species (Figs. 2, S7, S9, Table S7) as
160 we found evidence that some species may had hybridized during the course of their
161 diversification. Genomic introgressions were detected between i) erect-crested and the
162 ancestral rockhopper penguin species (17–23%); ii) erect-crested and macaroni/royal
163 (25%) penguins (Fig. 2), and iii) an ancestor of the Galápagos/humboldt and magellanic
164 penguins (11%; Fig. 2). Our divergence time estimated the divergence of the crown-group
165 penguins in the early Miocene at 21.9 Mya (95% CI 19.06–25.19; Figs. 1, S10, Table S8).
166

167 **Biogeographic history and ancestral niche reconstruction**

168 The reconstruction of ancestral geographical occurrence identified the coastlines of
169 Australia, New Zealand, and nearby islands as the most likely range of the ancestor of
170 extant penguins (Figs. 1, S11; Table S9). The first branching event (20.3–19.7 Mya) led
171 to the establishment of the genus *Aptenodytes* in the Antarctic, and reconstructions of the
172 ancestral *Pygoscelis* species indicate that they colonized the Antarctic Peninsula (area
173 C; Fig. 1, Table S9) soon after *Aptenodytes*, pointing to a long history of occupation in
174 Antarctica. In the mid-Miocene, the lineage leading to the *Spheniscus+Eudyptula* ancestor
175 colonized the South American coast (A), with members of the genera *Eudyptes*,
176 *Eudyptula*, *Megadyptes*, and *Spheniscus* progressively diversifying and colonizing
177 warmer oceanic environments (Fig. 1).

178 Consistent with the above findings, our reconstruction of sea surface temperatures
179 (SST) as a proxy for ecological niche disparity through time (DTT) suggests that penguins
180 originated from areas with a maximum SST of 9°C (Figs. 1, S12-S13) which is in
181 alignment with present-day temperatures of sub-Antarctic waters (21). Southern and
182 eastern rockhopper penguins have retained a thermal preference for a maximum SST of
183 9°C, while other closely related species (northern Rockhopper, fiordland, erect-crested)
184 have shifted towards warmer at-sea conditions (SST of 11–17°C; Figs. 1, S13). This

185 thermal shift was most likely driven by divergence across the STF over the past 4.5 Mya.
186 In contrast, Macaroni penguins have shifted to occupy colder at-sea conditions to feed in
187 the nutrient-rich waters off Antarctica (Fig. S13). Lower latitude geographical locations,
188 such as the South African continental coasts (G) and the Galápagos Islands (J), were
189 colonized in the Pleistocene (0.59 Mya). Galápagos penguins are present in Pacific
190 tropical waters near the equator, where SST reach up to 27°C, and exhibit a significant
191 increase in thermal tolerance compared to their sister-species, humboldt penguins (>6°C
192 higher; DTT results, Fig. 1).

193

194 **Genes under positive selection**

195 We detected 104 genes under positive selection (BEB value >0.95 and FDR q<0.1) using
196 a site model across all branches of the Spheniscidae (Table S10-12, Fig. S14). Using a
197 gene interaction network analysis, we found that many of these positively selected loci
198 are functionally connected and formed two major clusters (Fig. 3): one primarily related
199 to broad cellular functions, and the other containing genes that affect specific phenotypes
200 including immunity, renal function and cardiovascular activities (e.g., blood pressure,
201 oxygen metabolism, coagulation). Our gene ontology analysis revealed a concordant
202 pattern of pathway enrichment, including terms related to angiotensin regulation, blood
203 pressure, and oxygen metabolism (Fig. 3, Table S11–12). All of these adaptations are
204 related to diving and maintenance of high core body temperatures.

205

206 **Ancestral effective populations sizes (N_e)**

207 Our reconstruction of changes in ancestral effective population sizes (N_e) recovered three
208 different demographic trajectories in deeper time. Eleven species of penguin (emperor,
209 king, Adélie, chinstrap, gentoo [Kerguelen Population], humboldt, magellanic, African,
210 eastern rockhopper, little) increased N_e between 0.04 and 0.07 Mya, with N_e dropping
211 precipitously during the Last Glacial Maximum. In contrast, three species (northern
212 rockhopper, southern rockhopper, erect-crested) increased N_e during the Penultimate
213 Glaciation Period (PGP) between 0.130–0.194, but were in decline by 0.07 Mya. Finally,

214 two species (gentoo [Antarctica, Falkland Is. populations], Galápagos) have been in
215 steady decline since at least the Naynayxungla Glaciation (0.50–0.72 Mya; Fig. 4, S16).
216 Why Galápagos penguins and southern populations of gentoo penguins have been in
217 decline over such a considerable period of time remains uncertain, as there are no large
218 life-history or ecological differences that distinguishes these two species from other
219 penguins (Table S2). One possibility is that both Galápagos penguins and southern
220 populations of gentoo penguins represent recent divergence events (Fig 1. S7, S10), and
221 hence, the demographic reconstructions may be reflective of a founder event with larger
222 deep time N_e reflecting ancestral population sizes.

223 For erect-crested, southern and northern rockhoppers, and magellanic penguins
224 distributed around the tip of South America and on islands north of the APF, N_e was
225 highest during the PGP with a decline starting in or shortly after the subsequent
226 interglacial. This pattern could reflect changes in ecosystem productivity over deep time,
227 but it could be an artifact of the inability of the PSMC method to account for introgression.
228 Our analyses suggest that for these four species ancestral introgression (Fig. S9) would
229 have elevated estimates of N_e and could have offset the time interval when a peak in N_e
230 occurred (Fig. S16). The largest ancestral N_e we observed was for Adélie penguins
231 ($>10^5$), a primarily cold-water species (Fig. 1) widely distributed south of the APF;
232 decreases in temperatures and increases in ice cover may have promoted the expansion
233 of that population. Smaller historical peaks for N_e ($\sim 2\text{--}3 \times 10^4$) were recovered for island
234 endemic species, such as fiordland (southern New Zealand) and northern rockhopper
235 penguins (Tristan da Cunha and Gough Island).

236

237 **Discussion**

238 Phylogenetic reconstructions recovered by this study show that the large emperor and
239 king penguins (i.e., *Aptenodytes*) are sister to all other extant penguins, refuting the
240 hypothesis that *Aptenodytes* and *Pygoscelis* are sister-taxa (4, 8, 9). While genome-wide
241 data used in our trees allowed for the detection of the deep lineage split between
242 *Aptenodytes* and all other penguins, the short internal branch recovered by our analysis
243 at this split likely indicates a rapid diversification event (<1 Mya) (22). Such rapid
244 speciation events may explain the discrepancies among the previous proposed

245 phylogenetic hypothesis generated using that few molecular markers, which are known
246 to be insufficient in resolving such deep and short branches (23). As the clades become
247 more divergent, the phylogenetic position of the remaining genera and species are
248 consistent with phylogenetic hypotheses generated in other studies (4–10). Reticulated
249 evolution may also have contributed to some inconstancies in penguin phylogenetic
250 reconstructions. Genome analyses detected deep and shallow introgression events
251 across the phylogeny of extant penguins. Species of crested penguins (*Eudyptes*) appear
252 to have exchanged genes throughout much of their evolutionary history. The directionality
253 of introgression among these linages is consistent with the clockwise direction of the
254 Antarctic Circumpolar Current (ACC) connecting sub-Antarctic islands and promoting
255 dispersal. We also detected extensive genomic introgression between an ancestor of the
256 galápagos/humboldt penguin and the magellanic penguin; such introgression is still
257 observed at present between magellanic and humboldt penguins for which hybridization
258 has been reported in the wild (24).

259 Our divergence time estimates are consistent with the fossil record (12), placing
260 the ancestor of crown-group penguins in the early Miocene. Our biogeographic
261 reconstruction for crown-group penguins pointed to the coastlines of Australia and New
262 Zealand and nearby islands as the most likely range of the ancestor of extant penguins,
263 supporting earlier suggestion (7), rather than Antarctica as place of origin (5). Because of
264 the recency of most diversification events (2–9 Mya), we propose an alternative
265 hypothesis regarding their geographic expansion, namely that the opening of the Drake
266 Passage around the tip of South America led eventually to the full development of a deep,
267 strong ACC by ~11.6 Mya (26) which in turn may have contributed to colonization of new
268 areas and diversification of penguins, rather than the initial formation of the ACC ~35 Mya
269 (26)(Fig. 1). Global climate cooling intensified with the strengthening of the ACC and
270 possibly led to the extinction of several species inhabiting Antarctica (20, 27). The ACC
271 also reinforced the isolation of taxa inhabiting Antarctica from those on continental
272 shelves and islands to the north (20, 27), while promoting eastward colonization of
273 available sub-Antarctic islands.

274 Acceleration of the decline in surface temperatures since the Pliocene (20) may
275 have facilitated the colonization of new areas such as islands in the Indian Ocean. This

276 hypothesis is supported by our analysis of the divergence of the four gentoo penguin
277 lineages across thermal and salinity gradients to the north and south of the APF (Figs.
278 S5–S8). The Pliocene-Pleistocene cooling, along with the expansion of ice shelves
279 across the Southern Ocean (20), would have reduced connectivity among penguin
280 populations and facilitated speciation across *Pygoscelis*, *Spheniscus*, *Eudyptes* and
281 *Aptenodytes* between 2 and 5.5 Mya (Fig. 1).

282 During the Quaternary glaciations (1.8–0.01 Mya), sea-ice is thought to have
283 reached c. 40°S at the coast of South America (28). This extent of sea-ice would have
284 promoted the northern expansion of *Spheniscus* to the subtropics, and subsequently
285 enabled colonization of the Galápagos Islands, home to rich feeding resources for
286 penguins due to the upwelling of cooler, nutrient rich waters. Strong north-flowing currents
287 – the Humboldt Current and the Benguela Current – would have further facilitated penguin
288 colonization of subtropical habitats in the Pacific (Galápagos) and Atlantic (southern
289 Africa), respectively. Neither of these current systems penetrate beyond the equator; this,
290 together with their sub-Antarctic origins and preference for cooler waters, may explain
291 why crown-group penguins never successfully colonized the Northern Hemisphere.

292 The genes for which signals of positive selection were detected are associated with
293 thermoregulation (e.g., vasoconstriction/vasodilation: ENPEP, MME, BDKRB2),
294 osmoregulation (e.g., balance of fluids and salt: SCL6A19, ACE2, AGT), and diving
295 capacity (e.g., oxygen storage: MB), thereby reflecting adaptations that enabled penguins
296 to colonize both colder and warmer habitats, away from their ancestral thermal maximum
297 of c. 9°C (Fig. 1, S12-13). In Antarctica, emperor penguins are exposed to temperatures
298 as low as -40°C and forage in waters of -1.8°C (29), whereas Galápagos penguins are
299 exposed to SST >27°C and air temperatures exceeding 40°C. These high temperatures
300 are mostly associated with El Niño events, which may cause heat stress, high mortality,
301 and low recovery of the penguin colonies (30, 31). Regulating blood pressure selectively
302 through the constriction of blood vessels can further conserve oxygen consumption and
303 facilitate the maintenance of high core body temperatures (32).

304 Penguins spend most of their lives at sea, often performing prolonged dives while
305 foraging. They store oxygen in their lungs, blood, and muscles (33), and their rates of

306 oxygen consumption can be very low (34). The two largest penguin species, the emperor
307 and king penguins, can achieve depths of >300 meters and maximum dive durations of
308 22 and 8 minutes, respectively (33, 35, 36). Smaller penguin species tend to dive in
309 shallow waters (<50 m) with dives of 1–2 minutes in duration (33), although the chinstrap
310 penguins often dive off-shelf (>200 m, (37)). In this sense, nucleotide differences in
311 Myoglobin (MB, overall positive selection, $Z=2.645$, $p=0.005$) across species groups
312 could be associated with differences in diving capacity. For example, we found several
313 non-synonymous substitutions that were common within *Pygoscelis*, *Eudyptes*, and
314 *Aptenodytes* penguins, but differed in their amino acid composition (Fig. S15). It is
315 possible that these non-synonymous mutations encode greater oxygen-binding capacity,
316 which would facilitate the deep and prolonged dives performed by *Aptenodytes* and some
317 species of *Pygoscelis* penguins compared with *Eudyptes* (higher dN/dS ratios, Table
318 S13).

319 Our results suggest that adaptive evolution, implicating genes involved in multiple
320 interconnected genetic pathways, has increased the foraging success and survival of
321 penguin species across diverse temperature and salinity gradients. Foraging success is
322 associated with reproductive success (38, 39) and also with survival during long periods
323 of fasting while caring for eggs and chicks (40). Collectively, such adaptations would have
324 enabled the radiation of penguin species across the Southern Hemisphere.

325 Penguins have a remarkable evolutionary history. Their radiation from the
326 temperate coasts of New Zealand and Australia into other parts of the Southern
327 Hemisphere was facilitated by changes in global circulation patterns over the past 20
328 million years. Our analysis detected positive selection across several gene networks,
329 suggesting that molecular adaptation enabled the establishment of penguin populations
330 in Antarctic and tropical regions, and enhanced the ability of some species to dive deeply.
331 Demographic reconstructions over the past million years show that most penguin species
332 have declined during the severe ice conditions during the LGM in the Southern Ocean, a
333 result concordant with that recovered for several other bird species (41, 42).

334 Our results suggest that penguins originated from areas with a maximum SST of
335 9°C and diversified over millions of years, occupying colder Antarctic and warmer tropical

336 waters. As such, it seems unlikely that locally adapted species will be able to keep pace
337 with rapid climate change, especially as marine species may be more vulnerable to global
338 warming than terrestrial species (43, 44). This vulnerability is especially pertinent for
339 penguins, as illustrated by the massive mortality of Adélie penguin chicks and by the
340 relocation of emperor penguins in response to suboptimal sea ice conditions (45, 46). As
341 large-scale genomic studies become increasingly feasible and data for more
342 sophisticated global climate models become available for niche modelling, the application
343 of approaches like those in the present study hold significant promise for revealing new
344 insights into the evolutionary history and climatic vulnerability of many of the world's
345 poorly understood taxonomic groups.

346

347 **Materials and Methods**

348 Detailed methods for each of the sections below are provided in the supplementary
349 documents.

350

351 **Genome sequencing and assembly**

352 The genomes of 18 extant penguin species (22 individuals) as well as the southern giant
353 petrel (*Macronectes giganteus*) were sequenced to ~30x coverage with 150 bp paired-
354 end reads using an Illumina HiSeq X platform at MedGenome (USA; Table S1). We used
355 the giant petrel as outgroup for the analyses reported below. Briefly, duplicate sequences
356 were removed using Super Deduper (<https://github.com/dstreett/Super-Deduper>) and the
357 reads were then filtered. We aligned the resulting cleaned reads of each individual to the
358 emperor penguin reference genome (<http://gigadb.org/dataset/100005>; scaffold-level
359 assembly) using LAST (<http://last.cbrc.jp/>). The extent to which genome assemblies were
360 complete was assessed using the Benchmarking Universal Single-Copy Orthologs,
361 BUSCO v2 dataset (47)(Fig. S2) and KAT spectra-cn plots (48)(Fig. S3, S4). The CDS,
362 intron and mitogenome sequences were extracted for each genome using the reference
363 genome GFF. We extracted 120 bp Ultra Conserved Element (UCE) loci with 750 bp of
364 length of the flanking sequence padded to each side with scripts from the PHYLUCE
365 pipeline (49). Sequences were aligned using MAFFT (50).

366

367 **Estimation of phylogeny, divergence times, and interspecific introgression**

368 To account for potential genome-wide incompatibilities between taxa and loci, we used
369 Astral III (48) to estimate species tree phylogenies for each of the UCE, intron, and CDS
370 data sets, and for all three data sets combined (Fig S8). We used RAxML-NG (v. 0.5.1b
371 BETA) (51) to generate independent gene trees for each locus of the UCE, intron, and
372 CDS alignments. As input to Astral III, we merged all of the “best” trees generated by
373 RAxML-NG into a single file (Supplement). For the phylogenomic concatenated analysis
374 using all the data from UCEs, CDS, intron, and the mitogenome we carried out maximum
375 likelihood (ML) analyses in IQ-TREE, (52), with 1000 bootstrap replicates (Fig. S5-S7).
376 We estimated divergence times in BEAST v2.5.2 (53) using the computer resources
377 available through the CIPRES Science Gateway (Fig S7, S10) (54) calibrating the
378 topology with five fossils (Table S8). This phylogeny was used for all subsequent analysis
379 that required this information. Analyses to investigate the extent of interspecific
380 introgression across the phylogeny were performed using a partitioned *D*-statistics
381 approach implemented in DFOIL (55) for possible taxa combinations (eight) respecting
382 the priors of a symmetrical tree composed of four taxa and an outgroup, one ingroup
383 clade being younger than the other (Fig. 2, S9, Table S7). To perform the tests, we split
384 the genome-wide alignments into 100 kb, non-overlapping windows with Bedtools and
385 custom scripts.

386

387 **Ancestral Distribution and Niche Reconstruction**

388 For the historical biogeographic analysis, we estimated the ancestral range of the extant
389 penguin species in the R package BioGeoBEARS (56) implementing three models of
390 ancestral area reconstruction with and without long distance dispersal (the parameter *j*:
391 “jump dispersal”). We subdivided the extant penguin geographic distribution into 10
392 different areas. Occurrence records for all penguin species and six marine variables were
393 used to created raw models with MaxEnt–Javascript (57). Niche overlap was estimated
394 between all penguin species for the set of variables considered. We used the package
395 ‘Phyloclim’ to create Predicted Niche Occupancy (PNO) profile values and plots.
396 Subsequently we combined this information with the phylogenetic tree to generate the

397 Divergence Through Time (DTT) plots and climatic tolerance chronograms depicted in
398 Figure 1.

399

400 **Detection of signatures of positive selection**

401 We performed a dN/dS ratio test using the CodeML algorithm implemented in ETE3 (58,
402 59) for all species and sub-groups in the phylogeny. Due to the large number of analyzed
403 genes, we performed a multiple comparisons (false-discovery rate [FDR]) test
404 implemented in R. Genes that persisted on the list (i.e. remained significantly different
405 from neutral expectations, supporting a positive selection regime) were then used to
406 perform a gene ontology analysis using WebGestalt (WEB-based GEne SeT AnaLysis
407 Toolkit) (60) and network analysis performed with StringDB (61) (Figure S14-S15).

408

409 **Demographic history**

410 To address questions about how climate may have influenced effective population size
411 for each penguin species, we performed a demographic analysis using a pairwise
412 sequential Markovian coalescent (PSMC) method (62). PSMC was run with parameters
413 “-N25 -t15 -r5 -p 4 + 25*2 + 4 + 6”, and an estimated generation time (g) for the different
414 penguin species (Figure 4, S16, Table S2) using a substitution rate derived from chicken
415 pedigrees (63).

416

417 **References**

- 418
- 419 1. Egevang C, *et al.* (2010) Tracking of Arctic terns Sterna paradisaea reveals
420 longest animal migration. *Proceedings of the National Academy of Sciences* 107(5):2078-
421 2081.
 - 422 2. Péron C & Grémillet D (2013) Tracking through Life Stages: Adult, Immature and
423 Juvenile Autumn Migration in a Long-Lived Seabird. *PLOS ONE* 8(8):e72713.
 - 424 3. Thomas DB & Fordyce RE (2012) Biological Plasticity in Penguin Heat-Retention
425 Structures. *Anatomical Record-Advances in Integrative Anatomy and Evolutionary
426 Biology* 295(2):249-256.
 - 427 4. Gavryushkina A, *et al.* (2017) Bayesian Total-Evidence Dating Reveals the Recent
428 Crown Radiation of Penguins. *Systematic Biology* 66(1):57-73.
 - 429 5. Baker AJ, Pereira SL, Haddrath OP, & Edge KA (2006) Multiple gene evidence for
430 expansion of extant penguins out of Antarctica due to global cooling. *Proc Biol Sci*
431 273(1582):11-17.

- 432 6. Clarke JA, *et al.* (2007) Paleogene equatorial penguins challenge the proposed
433 relationship between biogeography, diversity, and Cenozoic climate change. *Proc Natl
434 Acad Sci U S A* 104(28):11545-11550.
- 435 7. Ksepka D.T. BS, Giannini N. (2006) The phylogeny of the living and fossil
436 Sphenisciformes (penguins). *Cladistics* 22:412-441.
- 437 8. Subramanian S, Beans-Picon G, Swaminathan SK, Millar CD, & Lambert DM (2013)
438 Evidence for a recent origin of penguins. *Biol Lett* 9(6):20130748.
- 439 9. Cole TL, *et al.* (2019) Mitogenomes Uncover Extinct Penguin Taxa and Reveal Island
440 Formation as a Key Driver of Speciation. *Mol Biol Evol* 36(4):784-797.
- 441 10. Pan H, *et al.* (2019) High-coverage genomes to elucidate the evolution of penguins.
442 *GigaScience* 8(9).
- 443 11. Figueiró HV, *et al.* (2017) Genome-wide signatures of complex introgression and
444 adaptive evolution in the big cats. *Science Advances* 3(7):e1700299.
- 445 12. Degrange FJ, Ksepka DT, & Tambussi CP (2018) Redescription of the oldest crown
446 clade penguin: cranial osteology, jaw myology, neuroanatomy, and phylogenetic
447 affinities of *Madrynornis mirandus*. *Journal of Vertebrate Paleontology* 38(2):e1445636.
- 448 13. Bertelli S & Giannini NP (2005) A phylogeny of extant penguins (Aves:
449 Sphenisciformes) combining morphology and mitochondrial sequences. *Cladistics*
450 21(3):209-239.
- 451 14. Vianna JA, *et al.* (2017) Marked phylogeographic structure of Gentoo penguin reveals an
452 ongoing diversification process along the Southern Ocean. *Mol Phylogenet Evol* 107:486-
453 498.
- 454 15. Frugone MJ, *et al.* (2018) Contrasting phylogeographic pattern among *Eudyptes* penguins
455 around the Southern Ocean. *Sci Rep* 8(1):17481.
- 456 16. Munro KJ & Burg TM (2017) A review of historical and contemporary processes
457 affecting population genetic structure of Southern Ocean seabirds. *Emu - Austral
458 Ornithology* 117(1):4-18.
- 459 17. Frugone MJ, *et al.* (2019) More than the eye can see: Genomic insights into the drivers of
460 genetic differentiation in Royal/Macaroni penguins across the Southern Ocean.
461 *Molecular Phylogenetics and Evolution* 139:106563.
- 462 18. Clucas GV, *et al.* (2018) Comparative population genomics reveals key barriers to
463 dispersal in Southern Ocean penguins. *Molecular Ecology* 27(23):4680-4697.
- 464 19. Perttierra L, *et al.* (under review) Integrated phylogenomic and niche analyses reveal
465 cryptic speciation in Gentoo penguins driven by local adaptation. *Molecular Ecology*.
- 466 20. Halanych KM & Mahon AR (2018) Challenging Dogma Concerning Biogeographic
467 Patterns of Antarctica and the Southern Ocean. *Annual Review of Ecology, Evolution,
468 and Systematics* 49:355-378.
- 469 21. Dunstan PK, *et al.* (2018) Global patterns of change and variation in sea surface
470 temperature and chlorophyll a. *Scientific reports* 8(1):14624-14624.
- 471 22. Weisrock DW, Harmon LJ, & Larson A (2005) Resolving deep phylogenetic
472 relationships in salamanders: analyses of mitochondrial and nuclear genomic data. *Syst
473 Biol* 54(5):758-777.
- 474 23. Oliveros CH, *et al.* (2019) Earth history and the passerine superradiation. *Proceedings of
475 the National Academy of Sciences* 116(16):7916-7925.
- 476 24. Simeone A, *et al.* (2009) *Heterospecific Pairing and Hybridization between Wild
477 Humboldt and Magellanic Penguins in Southern Chile* (BIOONE) pp 544-550, 547.

- 478 25. Goldner A, Herold N, & Huber M (2014) Antarctic glaciation caused ocean circulation
479 changes at the Eocene-Oligocene transition. *Nature* 511(7511):574-577.
- 480 26. Dalziel IWD, *et al.* (2013) A potential barrier to deep Antarctic circumpolar flow until
481 the late Miocene? *Geology* 41(9).
- 482 27. Crame JA (2018) Key stages in the evolution of the Antarctic marine fauna. *Journal of*
483 *Biogeography* 45(5):986-994.
- 484 28. Glasser NF, Jansson KN, Harrison S, & Kleman J (2008) The glacial geomorphology and
485 Pleistocene history of South America between 38°S and 56°S. *Quaternary Science*
486 *Reviews* 27(3):365-390.
- 487 29. Williams CL, Hagelin JC, & Kooyman GL (2015) Hidden keys to survival: the type,
488 density, pattern and functional role of emperor penguin body feathers. *Proceedings of the*
489 *Royal Society B-Biological Sciences* 282(1817).
- 490 30. Boersma PD (1998) Population trends of the Galapagos penguin: Impacts of El Nino and
491 La Nina. *Condor* 100(2):245-253.
- 492 31. Boersma D (1975) Adaptations of Galapagos penguins for life in two different
493 environments. *The biology of penguins*, ed Stonehouse B (Macmillan, London), pp 101-
494 114.
- 495 32. Butler PJ & Jones DR (1997) Physiology of diving of birds and mammals. *Physiological*
496 *Reviews* 77(3):837-899.
- 497 33. Ponganis PJ & Kooyman GL (2000) Diving physiology of birds: a history of studies on
498 polar species. *Comparative Biochemistry and Physiology a-Molecular and Integrative*
499 *Physiology* 126(2):143-151.
- 500 34. Williams CL, Meir JU, & Ponganis PJ (2011) What triggers the aerobic dive limit?
501 Patterns of muscle oxygen depletion during dives of emperor penguins. *The Journal of*
502 *Experimental Biology* 214(11):1802-1812.
- 503 35. Wienecke B, Robertson G, Kirkwood R, & Lawton K (2007) Extreme dives by free-
504 ranging emperor penguins. *Polar Biology* 30(2):133-142.
- 505 36. Putz K & Cherel Y (2005) The diving behaviour of brooding king penguins (*Aptenodytes*
506 *patagonicus*) from the Falkland Islands: variation in dive profiles and synchronous
507 underwater swimming provide new insights into their foraging strategies. *Marine Biology*
508 147(2):281-290.
- 509 37. Kokubun N, Takahashi A, Mori Y, Watanabe S, & Shin HC (2010) Comparison of diving
510 behavior and foraging habitat use between chinstrap and gentoo penguins breeding in the
511 South Shetland Islands, Antarctica. *Marine Biology* 157(4):811-825.
- 512 38. Kowalczyk ND, Reina RD, Preston TJ, & Chiaradia A (2015) Environmental variability
513 drives shifts in the foraging behaviour and reproductive success of an inshore seabird.
514 *Oecologia* 178(4):967-979.
- 515 39. Berlincourt M & Arnould JPY (2015) Influence of environmental conditions on foraging
516 behaviour and its consequences on reproductive performance in little penguins. *Marine*
517 *Biology* 162(7):1485-1501.
- 518 40. Thiebot JB, *et al.* (2014) Adjustment of pre-moult foraging strategies in Macaroni
519 Penguins *Eudyptes chrysolophus* according to locality, sex and breeding status. *Ibis*
520 156(3):511-522.
- 521 41. Nadachowska-Brzyska K, Li C, Smeds L, Zhang G, & Ellegren H (2015) Temporal
522 Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome
523 Sequences. *Curr Biol* 25(10):1375-1380.

- 524 42. Cole TL, *et al.* (2019) Receding ice drove parallel expansions in Southern Ocean
525 penguins. *Proceedings of the National Academy of Sciences* 116(52):26690-26696.
526 43. Pinsky ML, Eikeset AM, McCauley DJ, Payne JL, & Sunday JM (2019) Greater
527 vulnerability to warming of marine versus terrestrial ectotherms. *Nature* 569(7754):108-
528 111.
529 44. Wiens JJ (2016) Climate-Related Local Extinctions Are Already Widespread among
530 Plant and Animal Species. *PLoS biology* 14(12):e2001104-e2001104.
531 45. Cimino MA, Lynch HJ, Saba VS, & Oliver MJ (2016) Projected asymmetric response of
532 Adélie penguins to Antarctic climate change. *Scientific Reports* 6:28785.
533 46. Fretwell PT & Trathan PN (2019) Emperors on thin ice: three years of breeding failure at
534 Halley Bay. *Antarctic Science*:1-6.
535 47. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, & Zdobnov EM (2015)
536 BUSCO: assessing genome assembly and annotation completeness with single-copy
537 orthologs. *Bioinformatics* 31(19):3210-3212.
538 48. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, & Clavijo BJ (2017) KAT:
539 a K-mer analysis toolkit to quality control NGS datasets and genome assemblies.
540 *Bioinformatics (Oxford, England)* 33(4):574-576.
541 49. Faircloth BC (2016) PHYLUCE is a software package for the analysis of conserved
542 genomic loci. *Bioinformatics* 32(5):786-788.
543 50. Katoh K & Standley DM (2013) MAFFT multiple sequence alignment software version
544 7: improvements in performance and usability. *Mol Biol Evol* 30(4):772-780.
545 51. Kozlov AM, Darriba D, Flouri T, Morel B, & Stamatakis A (2018) RAxML-NG: A fast,
546 scalable, and user-friendly tool for maximum likelihood phylogenetic inference.
547 *bioRxiv*:447110.
548 52. Nguyen L-T, Schmidt HA, von Haeseler A, & Minh BQ (2015) IQ-TREE: A Fast and
549 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies.
550 *Molecular Biology and Evolution* 32(1):268-274.
551 53. Bouckaert R, *et al.* (2014) BEAST 2: a software platform for Bayesian evolutionary
552 analysis. *PLoS Comput Biol* 10(4):e1003537.
553 54. Miller MA, Pfeiffer W, & Schwartz T (2010) Creating the CIPRES Science Gateway for
554 inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop*
555 (*GCE*), pp 1-8.
556 55. Pease JB & Hahn MW (2015) Detection and polarization of introgression in a five-taxon
557 phylogeny. *Systematic Biology* 64(4):651–662
558 56. Matzke N (2013) BioGeoBEARS: BioGeography with Bayesian (and Likelihood)
559 Evolutionary Analysis in R Scripts. *University of California, Berkeley, Berkeley, CA*. .
560 57. Phillips SJ, Anderson RP, & Schapire RE (2006) Maximum entropy modeling of species
561 geographic distributions. *Ecological Modelling* 190(3-4):231-259.
562 58. Huerta-Cepas J, Serra F, & Bork P (2016) ETE 3: Reconstruction, Analysis, and
563 Visualization of Phylogenomic Data. *Mol Biol Evol* 33(6):1635-1638.
564 59. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum
565 likelihood. *Computer applications in the biosciences : CABIOS* 13(5):555-556.
566 60. Wang J, Vasaikar S, Shi Z, Greer M, & Zhang B (2017) WebGestalt 2017: a more
567 comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit.
568 *Nucleic Acids Res* 45(W1):W130-W137.

- 569 61. Szklarczyk D, *et al.* (2019) STRING v11: protein-protein association networks with
570 increased coverage, supporting functional discovery in genome-wide experimental
571 datasets. *Nucleic Acids Res* 47(D1):D607-d613.
572 62. Li H & Durbin R (2011) Inference of human population history from individual whole-
573 genome sequences. *Nature* 475(7357):493-496.
574 63. Nam K, *et al.* (2010) Molecular evolution of genes in avian genomes. *Genome Biol*
575 11(6):R68.

576

577 **Acknowledgments:** We thanks to Andrea Polanowski, Claudia Godoy and Klemens
578 Pütz for assistance with sample acquisition; **Funding:** Financial support was provided
579 by INACH RT_12–14, Fondecyt Project 1150517, GAB PIA CONICYT ACT172065,
580 NSF DEB-1441652, French Polar Institute Paul-Emile Victor (IPEV, Progs. 137 and
581 354), and a Lakeside grant from the California Academy of Sciences;

582 **Data and materials availability:** Penguins and giant petrel raw fastq reads,
583 reconstructed genomes (BioProject PRJNA530615, BioSample accession
584 SAMN11566608-SAMN11566630) and mitogenomes (MK760983-MK761004,
585 MK761006) were deposited in Genbank. All UCE, CDS, intron and mitogenome
586 alignments, dated phylogenies and codes for all data analyses are available at Dryad
587 (<https://doi.org/10.5061/dryad.pk0p2ngj2>). All other data needed to evaluate the
588 conclusions in this paper are present either in the main text or the supplementary
589 materials.

590

591

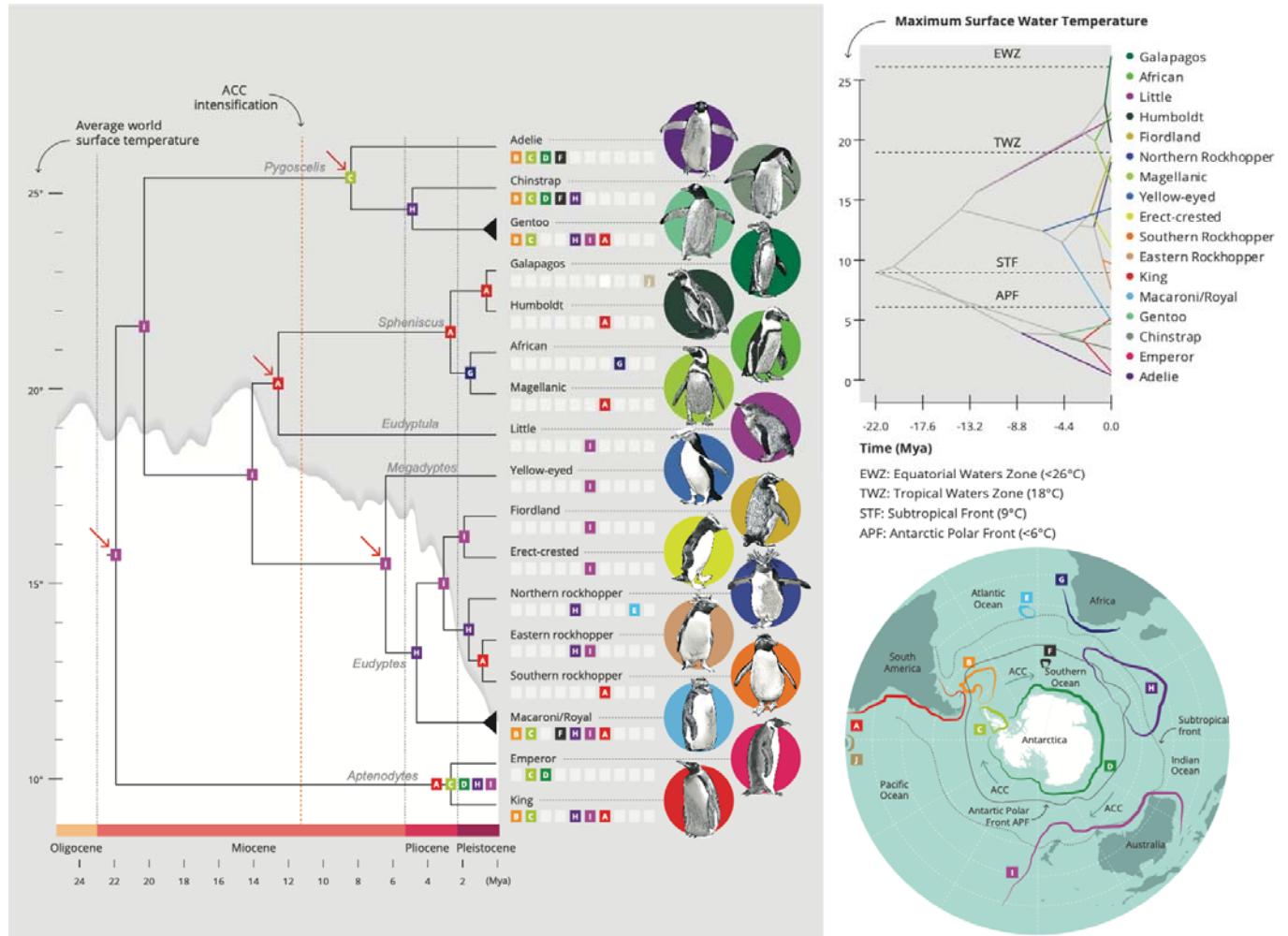
592 **Supplementary Materials:**

593 Materials and Methods

594 Figures S1-S16

595 Tables S1-S13

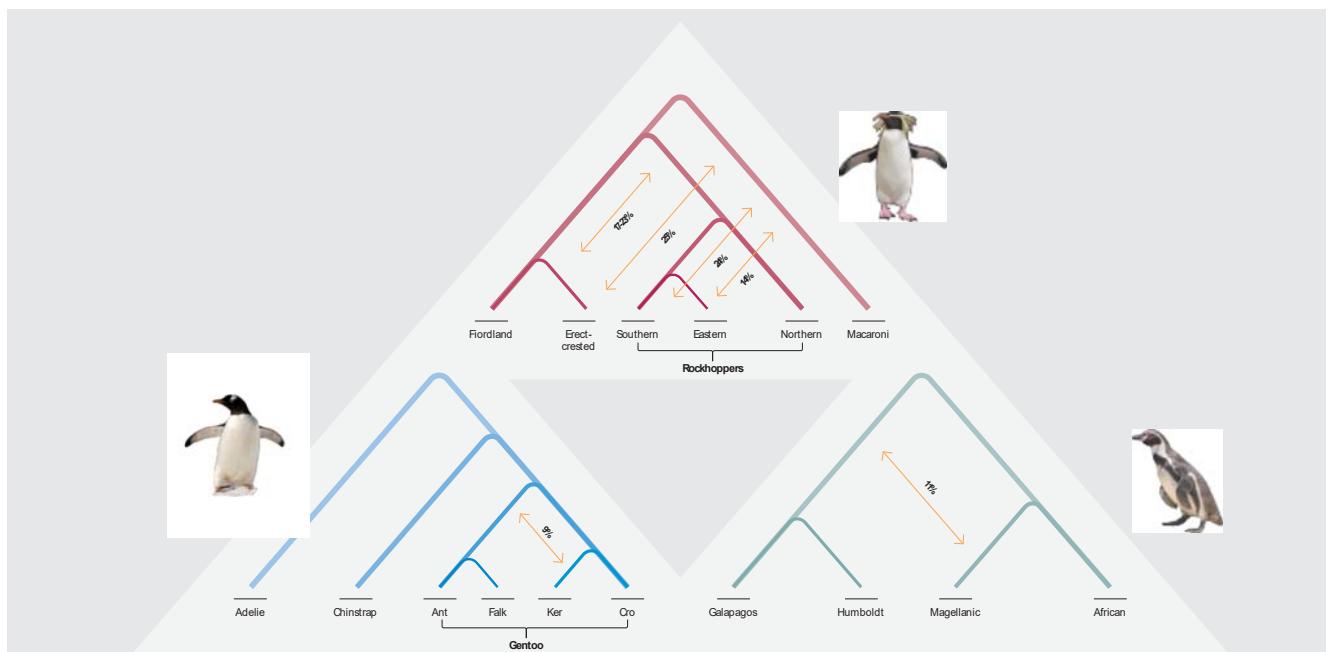
596 References (1-42)



599 **Fig. 1. Evolutionary history of penguins.** Phylogenetic hypothesis of penguin species
 600 and divergence time estimates using UCE dataset. Red arrows represent the four fossil
 601 calibration points (fifth point corresponds to the node with the outgroup which not
 602 represented). Each node is represented by the ancestral distribution before the
 603 cladogenesis event using Ancestral range reconstruction based on the best-fit model
 604 (DIVALIKE+J) and is associated with one or more of the ten geographic locations
 605 depicted on the map in the lower right (different letters A–J and colors); areas at branch
 606 tips represent the current range of species. Average world surface temperature in the
 607 past is represented by the white graph behind the phylogeny (41) and onset of the
 608 strengthening of the Antarctic Circumpolar Current (ACC) by a dashed red line. At the
 609 top right: Ecological niche disparity through time (DTT) for penguins (top right), with the
 610 phylogeny projected onto niche parameter space on the y-axis (maximum surface water

611 temperature) with predicted niche occupancy (PNO) over time (x-axis) reconstructed for
612 internal nodes.

613

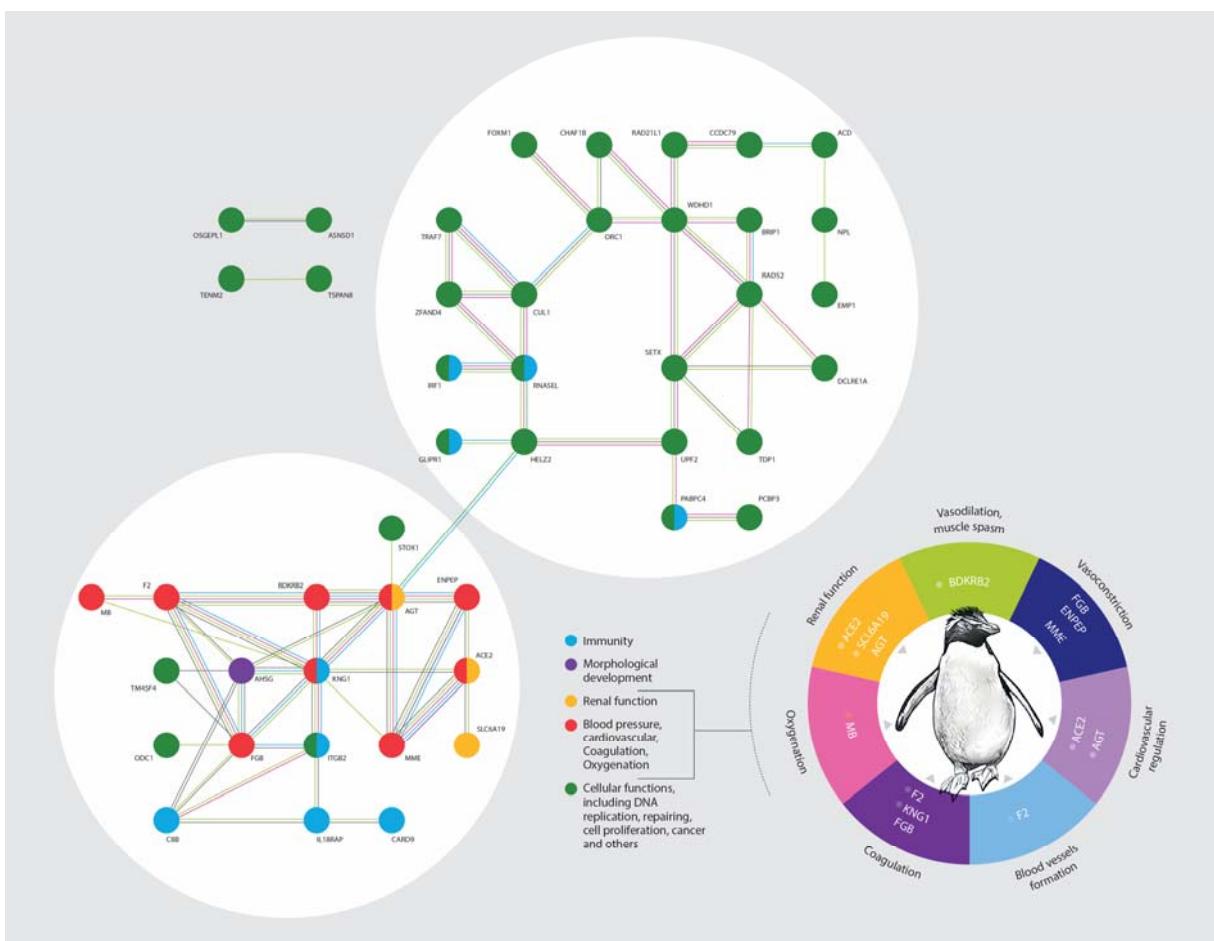


614

615 **Fig. 2. Summary of introgression among penguin taxa.** Arrows represent the
616 percentage of introgression (>2%) between taxa for all 8 combinations of the five-taxa
617 statements evaluated (see Fig. S9, Table S7).

618

619

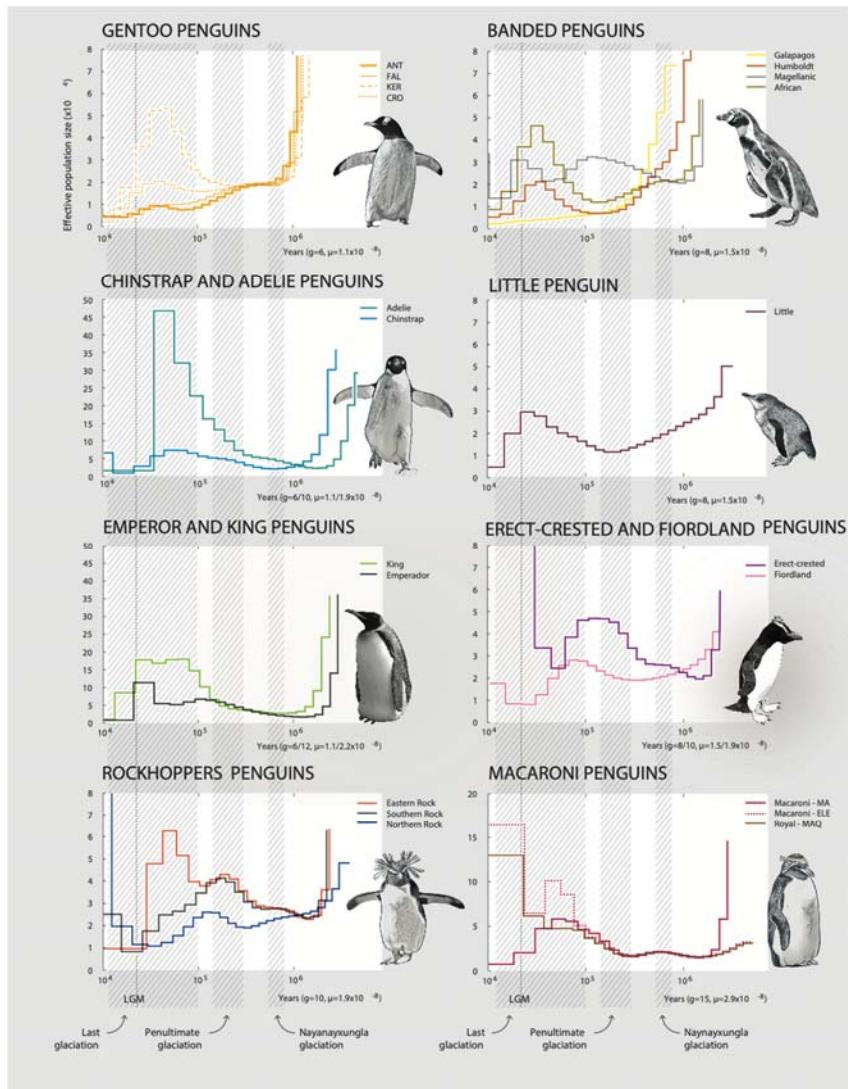


620

621 **Fig. 3. Analysis of genes under positive selection in extant penguin lineages.** The
622 main panel depicts results from the network analysis of positively selected genes,
623 retrieving two connected main clusters, one associated with general cellular functions
624 (green) and another grouped genes related to functions associated with osmoregulation
625 (renal function), immunity, thermoregulation (e.g. blood pressure) and diving ability (e.g.
626 oxygen metabolism) which are classified at the right pie-chart around the penguin.

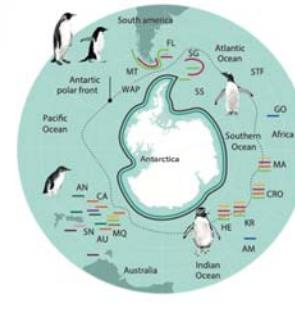
627

628



Legend for penguin lineages:

- Emperor
- Yellow-eyed
- King
- Little
- Royal
- Galapagos
- Macaroni
- Humboldt
- African
- Magellanic
- Gentoo
- Chinstrap
- Adelie



629

630 **Fig. 4. Demographic history of penguins.** Pairwise Sequentially Markovian
 631 Coalescent (PSMC) plot depicting the demographic history of each lineage inferred from
 632 genomic data represented by different colored lines (see Fig. S16 for bootstrapped
 633 curves). The period of the Naynayxungla glaciation, Penultimate glaciation and the Last
 634 Glaciation are represented in the plots, as well as the Last Glacial Maximum (LGM).