Genomic insights into the secondary aquatic transition of penguins

Theresa L. Cole1,35✉, Chengran Zhou2,35, Miaoquan Fang2,35, Hailin Pan2, Daniel T. Ksepka3, Steven R. Fiddaman4, Christopher A. Emerling5, Daniel B. Thomas6, Xupeng Bi7, Qi Fang2, Martin R. Ellegaard8,9, Shaohong Feng2,7, Adrian L. Smith4, Tracy A. Heath10, Alan J. D. Tennyson11, Pablo García Borboroglu12,13,14, Jamie R. Wood15, Peter W. Hadden16, Stefanie Grosser17, Charles-André Bost18, Yves Cherel18, Thomas Mattern17, Tom Hart4, Mikkel-Holger S. Sinding8, Lara D. Shepherd11, Richard A. Phillips19, Petra Quillfeldt20, Juan F. Masello20, Juan L. Bouzat21, Peter G. Ryan22, David R. Thompson23, Ursula Ellenberg13,24,25, Peter Dann26, Gary Miller27,28, P. Dee Boersma12, Ruoping Zhao29, M. Thomas P. Gilbert8,9, Huanming Yang2,30,31, De-Xing Zhang32,33 & Guojie Zhang1,2,7,29,34✉

Penguins lost the ability to fly more than 60 million years ago, subsequently evolving a hyper-specialized marine body plan. Within the framework of a genome-scale, fossil-inclusive phylogeny, we identify key geological events that shaped penguin diversification and genomic signatures consistent with widespread refugia/recolonization during major climate oscillations. We further identify a suite of genes potentially underpinning adaptations related to thermoregulation, oxygenation, diving, vision, diet, immunity and body size, which might have facilitated their remarkable secondary transition to an aquatic ecology. Our analyses indicate that penguins and their sister group (Procellariiformes) have the lowest evolutionary rates yet detected in birds. Together, these findings help improve our understanding of how penguins have transitioned to the marine environment, successfully colonizing some of the most extreme environments on Earth.
Penguins are one of the most iconic groups of birds, serving as both a textbook example of the evolution of secondarily aquatic ecology and as sentinels for the impacts of global change on ecosystem health\(^1\). Although often associated with Antarctica in the popular imagination, penguins originated more than 60 million years ago (Mya), evolving wing-propelled diving and losing the capacity for aerial flight long before the formation of polar ice sheets\(^2\). Over time, penguins evolved the suite of morphological, physiological, and behavioral features that make them arguably the most uniquely specialized of all extant birds. These adaptations have allowed penguins to colonize some of the most extreme environments on Earth.

Previous phylogenetic studies have yielded insights into penguin evolution, yet have been limited by sampling issues (e.g., number of lineages incorporated and quality of molecular markers\(^3–7\)). Genomic studies have shed light on the diversification of extinct penguins\(^7–9\) but have not integrated extinct species. Because nearly three-quarters of known penguin species are represented only by fossils (e.g., \(^2\), sampling extinct species is crucial for improving phylogenetic resolution and dating accuracy, reconstructing biogeographic events, and understanding the environmental context in which key adaptations arose. While several studies have included fossil penguins, these utilized only mitochondrial genomes and/or small numbers of nuclear genes (e.g., \(^3–6\)), limiting their ability to disentangle confounding processes, such as historical and ongoing introgression and incomplete lineage sorting.

Here, we take a comprehensive approach to inferring the tempo and drivers of penguin diversification by combining genomes from all extant and recently-extinct penguin lineages (27 taxa) (Table 1), stratigraphic data from fossil penguins (47 taxa), and morphological and biogeographic data from all species (extant and extinct) (Fig. 1 and Supplementary Fig. 1; Supplementary Data 1) into a single framework for Bayesian phylogenetic analysis. This combined approach, using the fossilized birth-death process with sampled ancestors\(^4\) (see supplementary methods) offers a more complete understanding of speciation and biogeographic events over the entire history of penguin evolution.

It extends our insights beyond the ~15–20 million year (Ma) history of crown penguins to include the ~50 Ma interval during which only stem penguins existed. Within this phylogenetic framework, we highlight key genes involved in marine adaptations, compare evolutionary rates in penguins to those of other birds, and reconstruct the demographic histories of individual species. Together, these extensive datasets provide new insights into the evolution of extreme ecological preferences and the genetic basis for the adaptations that enabled penguins to occupy these niches.

### Results

**Climate change drove evolution, biogeography, and demography.** Phylogenetic results (Fig. 1 and Supplementary Fig. 2) confirm previous findings, recovering *Aptenodytes* (king and emperor penguins) as the sister clade to all other crown penguins, with brush-tailed (*Pygoscelis*) penguins in turn sister to two clades unifying the banded (*Spheniscus*) and little (*Eudyptula*) penguins and the yellow-eyed (Megadyptes) and crested (*Eudyptes*) penguins\(^7,8\). Biogeographical reconstructions (Fig. 1, Supplementary Figs. 3–4 and Supplementary Data 1) support a Zealandian origin for penguins\(^6,7\). Stem penguins radiated extensively in Zealandia before dispersing to South America and Antarctica multiple times, following the eastward-flowing direction of the Antarctic Circumpolar Current (ACC) (Fig. 1). Crown penguins most likely arose from descendant lineages in South America, before dispersing back to Zealandia at least three times. Interestingly, at least two such dispersals occurred before the inferred onset of the ACC system, suggesting that early stem penguins were not dependent on currents to disperse over long distances. A second pulse of speciation coincides with the onset of the ACC, though understanding whether this pattern is real or an artifact of fossil sampling requires more collecting from early Eocene localities. We infer an age of ~14 Ma for the origin of crown penguins, which is more recent than the ~24 Ma age recovered in genomic analyses, not including fossil taxa\(^7\) (Supplementary Fig. 2b) and coincides with the onset of global cooling during the middle Miocene climate transition\(^\text{8,10}\) (Supplementary Fig. 3a). This young age suggests that expansion of Antarctic ice sheets and the onset of dispersal vectors such as the Benguela Current\(^11\) during the middle to late Miocene facilitated crown penguin dispersal and speciation, as hinted at by fossil evidence\(^12\).

Incongruences between species trees and gene trees were identified, e.g., alternate topologies occurred at high frequencies (>10%) for several internal branches (Fig. 1c; Supplementary Fig. 5). These patterns indicate that gene tree discordance may be caused by incomplete lineage sorting (ILS) or introgression events. By quantifying ILS and introgression via branch lengths from over 10,000 gene trees, we found that the rapid speciation within crown penguins was accompanied by >5% ILS content within the ancestors of *Spheniscus, Eudyptula, Eudyptes*, and several subgroups within *Eudyptes* (Fig. 2a). Our dated tree provides a temporal framework for this rapid radiation: the four extant *Spheniscus* taxa are all inferred to have split from one another within the last ~3 Ma, and likewise the nine extant *Eudyptes* taxa likely split from one another in that same time (Fig. 1b). Many closely related penguin species/lineages are known to hybridize in the wild (see supplementary methods). Consistent with this, multiple analyses suggest that introgression also contributes to species tree—gene tree incongruence (Supplementary Figs. 6–9 and Supplementary Data 2; also see Supplementary Methods for further details). This could explain the most notable conflict in previous phylogenetic results, which showed inconsistency over whether *Aptenodytes alone*\(^7\) or *Aptenodytes* and *Pygoscelis* together\(^4,5\) represent the sister clade to all other extant penguins. Introgression was detected between the ancestor

---

**Table 1 Penguin names.**

| Taxa                      | Common name                        |
|---------------------------|------------------------------------|
| *Aptenodytes forsteri*    | Emperor penguin                   |
| *Aptenodytes patagonicus* | King penguin                      |
| *Eudyptes chrysoceo*      | Southern rockhopper                |
| *Eudyptes filholi*        | Eastern rockhopper                 |
| *Eudyptes chrysrolophus chrysolophus* | Royal penguin             |
| *Eudyptes chrysrolophus schlegel* | Northern rockhopper            |
| *Eudyptes moseleyi*       | Fiordland penguin                 |
| *Eudyptes pachyrhynchus*  | Snares penguin                    |
| *Eudyptes robustus*       | Erect-crested penguin             |
| *Eudyptula minor*         | New Zealand little penguin        |
| *Eudyptula minor* Banks Peninsula | NZ white-flippered penguin (BAN) |
| *Eudyptula novaehollandiae* | Australian fairy penguin         |
| *Megadyptes antipodes antipodes* | Yellow-eyed penguin             |
| *Pygoscelis adeliae*      | Adélie penguin                    |
| *Pygoscelis antarctica*   | Chinstrap penguin                 |
| *Pygoscelis papua West Antarctic Peninsula* | Gentoo penguin WAP   |
| *Pygoscelis papua Falklands/Malvinas* | Gentoo penguin FAL    |
| *Pygoscelis papua Kerguelen* | Gentoo penguin KER             |
| *Pygoscelis papua South Georgia* | Gentoo penguin SG         |
| *Spheniscus demersus*     | African penguin                  |
| *Spheniscus humboldti*    | Humboldt penguin                 |
| *Spheniscus magellanicus* | Magellanic penguin               |
| *Spheniscus mendiculus*   | Galápagos penguin                |
| *Megadyptes antipodes richdalei* | Chatham Islands Megadyptes   |
| *Megadyptes antipodes waitaha* | Waitaha penguin         |
| *Eudyptes warhami*        | Chatham Islands crested penguin  |
of Aptenodytes and the ancestor of other extant penguins, and is inferred to have occurred when the range of these ancestors overlapped in South America (Fig. 2a and Supplementary Data 2). Introgression (>9%) was also detected between Eudyptula novaehollandiae and Eudyptula minor, and several introgression events were especially pervasive in Eudyptes (Fig. 2a and Supplementary Fig. 6).

Many extant penguin lineages began to diverge within the last 3 Ma (Fig. 1). To obtain insight into this recent phase of penguin diversification, we inferred post-speciation introgression events and estimated the time when gene flow from introgression ceased between 20 pairs of closely related lineages (see Supplementary Methods). Our results provide further evidence for recent introgression between all sampled pairings (Fig. 2b) except for
Eudyptes chrysocome and E. filholi, whose ranges are geographically disparate (Fig. 1a). Almost all species exhibit a genomic signature of a period of physical isolation during the Last Glacial Period (LGP) with increased climate fluctuation and environmental uncertainty, followed by postglacial contact and introgression as Earth warmed once again (Supplementary Figs. 8–9). This strongly supports the hypothesis that penguins were impacted by ecosystem-wide, climate-driven refugia/recolonization cycles in the Southern Ocean13,14, a pattern also observed in other marine taxa during the Last Glacial Maximum (e.g.,15). As ice volumes increased during the LGP high-latitude penguin species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia. As climate warmed from the late Pleistocene to Holocene, these species were likely forced into isolated mid-latitude refugia.
a window into long-term population histories (very recent trends cannot be accurately recovered with these methods\(^\text{16}\)). Four demographic patterns emerge for this critical time interval, illuminating disparate responses of penguins to glacial-interglacial cycles (Fig. 2c). The most prevalent pattern is shared by nine lineages (Aptenodytes patagonicus, Pygoscelis antarctica, P. papua "WAP", P. papua "SG", S. magellanicus, E. moseleyi, E. filholi, E. chrysophalus schlegeli, and E. sclateri) showing evidence of population decline coincident with the beginning of the LGP, followed by population decline towards the end of the LGP. In contrast to this pattern, nine lineages (A. forsteri, P. adeliae, E. moseleyi, E. filholi, E. chrysophalus schlegeli, and E. sclateri) show evidence of population decline coincident with the beginning of the LGP, followed by population expansion towards the end of the LGP. Almost all of the remaining lineages show strong evidence of persistent long-term declines in populations from the early LPG to the end of LPG. All three Eudyptula taxa and Eudyptes chrysophalus chrysophalus underwent a steep population decline spanning the LGP, while three taxa (P. papua "FA", S. mendiculus, and E. chrysocome) show evidence of continual population decline across the last 250 thousand years (ka).

Interestingly, taxa that increased in population size towards the end of the LGP (e.g., A. forsteri, P. adeliae, S. magellanicus, E. filholi, E. moseleyi, E. sclateri, and E. schlegeli) are typically migratory, and tend to forage offshore (>50 km; see Supplementary Data 1\(^\text{17}\)), while taxa that decreased towards the end of the LGP (e.g., S. humboldti, M. a. antipodes and likely M. a. richdalei) tend to be residential, and forage inshore; see Supplementary Data 1. Taxa that disperse farther may have overcome local impacts of global climate cooling during the LPG (e.g., changes in sea-ice extent, prey abundance and terrestrial glaciation, however see\(^\text{18}\)) largely by relocating to lower latitudes (e.g.,\(^\text{14}\)), whereas locally-restricted taxa may have been more prone to sudden population collapses.

Penguins have the lowest evolutionary rates among birds. The integrated evolutionary speed hypothesis (IESH) proposes that temperature, water availability, population size, and spatial heterogeneity influence evolutionary rate\(^\text{19}\). Life history traits also impact the evolutionary rate, but such relationships remain incompletely understood in birds\(^\text{20}\). Penguins are long-lived, large-bodied, and produce few offspring, thus providing an ideal case study in how life history may impact evolutionary rate. We found that penguins and their sister group (Procellariiformes) had the lowest evolutionary rates of the 17 avian orders sampled within penguins (Supplementary Data 4). Extant penguin lineages show a wide range of individual rates, and phylogenetic correlation analyses (phylogenetic generalized least squares regression) shed light on potential factors influencing this disparity (Fig. 3c-e and Supplementary Data 3). Extant penguins showed a significant negative correlation between body mass and average sea surface temperature (Fig. 3d). Despite species from warmer regions having shorter generation times (Fig. 3d), a significant negative correlation was found between evolutionary rate and average sea surface temperature (Fig. 3e), suggesting that temperature may influence penguin evolutionary rates by regulating selective pressures, but not only through its effect on metabolism\(^\text{22}\). This result is in parallel with studies that show speciation rates to be higher in polar environments than in the tropics, pointing towards faster rates of evolution and more opportunities for divergence at high latitudes\(^\text{23,24}\). We propose that these patterns together reflect the signature of climate oscillations on high latitude species: polar penguins (e.g. A. forsteri/P. adeliae) were likely forced into more northerly refugia during ice ages, subsequently colonizing Antarctica during interglacials\(^\text{14}\). These events may have led to faster evolutionary rates as these lineages underwent population contraction-expansion cycles and were periodically forced to adapt to new environments.

Putative molecular adaptations unique to penguins. As penguins became increasingly adapted to a flightless diving ecology, they encountered novel selection pressures that required modifications to their locomotory strategy, thermoregulation, sensory perception, and diet. We tested whether these phenotypic changes have been facilitated through the evolution of the underlying protein-coding genes (Supplementary Data 4) by identifying positively selected genes (PSGs), rapidly evolving genes (REGs), and pseudogenes that relate to specific adaptations including thermoregulation, oceanic diving, oxygenation, underwater vision, shifts in diet and taste, body size and immunity (see Figs. 4, 5 and Supplementary Methods for additional details and citations). These genes either differ in all penguins compared with other birds, differ in the genus Aptenodytes compared with other penguins, or are under distinct selective pressures within penguins (Supplementary Data 4). In the branch leading to the last common ancestor (bLCA) of penguins, 27 PSGs (false discovery rate [FDR] \(q < 0.05\)) and 13 REGs (FDR \(q < 0.05\)) were detected. In the bLCA of Aptenodytes, 25 PSGs (FDR \(q < 0.05\)) and 3 REGs (FDR \(q < 0.05\)) were detected. In the bLCA of penguins and four flightless/nearly flightless birds (Nannopterum harrisi, Rynchochetos jubatus, Zapornia atra, and Laterallus rogersi, see Supplementary Fig. 16a), five PSGs (FDR \(q < 0.05\)) and 38 REGs (FDR \(q < 0.05\)) were detected. Within penguins, 275 PSGs (FDR \(q < 0.01\)) were detected (Supplementary Data 4). We related the gene pathways and known functions of 15 PSGs and six REGs to penguin-specific adaptations (Fig. 4a). We also highlight five genes containing penguin-specific substitutions, seven pseudogenes, and two gene expansions (Fig. 4a, Supplementary Figs. 14, 15).

We identified three REGs that are shared by penguins and other flightless/nearly flightless birds. These genes are likely associated with the shortening, rigidity, and increased density of the forelimb bones which contribute to the flipper-like wing of penguins (Fig. 4a). TBXT and FOXP1 are related to the development of articular cartilage, tendons, and limb bones. SMAD3 is involved in the transforming growth factor-beta signaling pathway, which is important for maintaining articular cartilage and stimulating osteogenesis and bone formation. Perhaps most interestingly, TNMD, a PSG, is expressed during the differentiation and developmental phase of limb tendon, ligament, and collagen fibrils, and loss of TNMD can result in reduced tenocyte density. We hypothesize that TNMD may be key to the nearly wholesale replacement of penguin distal wing musculature by tendons, which stiffens and reduces heat loss to the high surface area flipper (Supplementary Fig. 16a-d). We also identified two genes KCNU1 and KCNMA1 that are related to calcium sequestration to be expanded in the genomes of both
penguins and grebes (*Podiceps cristatus* and *Podilymbus podiceps*) (Fig. 4a, Supplementary Fig. 15). These genes likely contribute to the high bone density characteristic of these taxa, which helps reduce buoyancy for deep diving.

Penguins have densely-packed waterproof feathers, thick skin, and a layer of subcutaneous fat enabling them to thermoregulate in cold environments. We identified four genes under selective pressure in common ancestors of penguins that are related to thermoregulation (Supplementary Data 4). These genes (*APPL1, TRPC1, EVPL*) showed evidence of positive selection or rapid rates of evolution on the bLCA of extant penguins but not in other birds (Fig. 4a). The white adipose tissue of penguins is important for survival in the cold, acting as an insulative layer in part via myoglobin concentration and utilizing anaerobic metabolism. Myoglobin levels are high in penguins, and are related to the capacity for the deepest diving (>500 m depth) and body mass, as seen in extreme deep-diving cetaceans. We hypothesize that *TFRC* may have the capacity for the deepest diving (>500 m depth) and, thus, these gene variants may enable these species to dive to extreme depths. While none of the hemoglobin genes were PSGs (P-value >0.05), we observed that *HBB-aA* (A140S) and *HBB-bB* (L87M) genes (Fig. 4c and Supplementary Fig. 16e) and TRPC1 are related to glucose levels and fatty acid breakdown through adiponectin.

Penguins function under hypoxic conditions during deep dives in part via myoglobin concentration and utilizing anaerobic metabolism. We identified seven genes related to oxygenation that are under positive selection or have penguin-specific substitutions in penguins. Transferrin Receptor 1 (*TFRC*) shows a positive selection in penguins (Supplementary Fig. 16f). Previous experimental work in cells has reported that *TFRC* messenger RNA is expressed in an oxygen-dependent manner. Importantly, *TFRC* is a top candidate gene for the hypoxia response of domesticated cattle. We hypothesize that *TFRC* has contributed to a convergent adaptation to surviving hypoxia in penguins. Interestingly, *FIBB* and *ANO6*, which are involved in blood coagulation, showed a signal of positive selection in *Aptenodytes*, but not in other genera (Supplementary Fig. 17). Among all penguins, *Aptenodytes* have the capacity for the deepest diving (>500 m depth) and, thus, these gene variants may enable these species to dive to extreme depths. While none of the hemoglobin genes were PSGs (P-value >0.05), we observed that *HBB-aA* (A140S) and *HBB-bB* (L87M) genes (Fig. 4c and Supplementary Fig. 16e) and TRPC1 are related to glucose levels and fatty acid breakdown through adiponectin.

Penguins frequently forage in low light, and exhibit specializations for vision in dim, blue-green marine environments.
Morphological research has shown that at least some penguins are cone trichromats with only three functional cone photoreceptor types, blue-shifted long-wavelength visual pigments, and no red oil droplets\(^1\). Genomic data support trichromatism in all penguins, plus rods (RH1) (Fig. 4a and Supplementary Fig. 19a). As all penguins share this deletion, reduced color vision must have occurred in the penguin stem lineage, similar to secondarily aquatic mammals\(^1\). Although penguins lack green cones, the functional orthologs of the remaining visual opsins in penguins strongly indicate the retention of violet (SWS1), blue (SWS2), and red (LWS) cones, plus rods (RH1) (Fig. 5a). This genetic signature is concordant with our experiments on Pygoscelis papua (see Supplementary Methods), which demonstrate a capacity for ultraviolet light perception at 365 nm, likely conferred by the SWS1 opsin. Furthermore, the peak wavelength sensitivity (\(\lambda_{\text{max}}\)) of penguin LWS opsins show evidence of shifts in spectral sensitivity to better match ambient underwater light. Relative to key avian model species (e.g., Taeniopygia guttata, Columba livia, Gallus gallus) and Procellariiformes, penguins possess substitutions at five key tuning sites in LWS, four of which (A180, F277,
A285, and S308) are associated with blue-shifting this pigment \(^5\) (Supplementary Fig. 19b). This suggests that this opsin has been fine-tuned for marine foraging, as observed in cetaceans \(^4\). CYP2J19, which encodes a carotenoid ketolase responsible for producing red oil droplets in avian cones \(^6\), has been inactivated in most penguins (Supplementary Data 4). Colored oil droplets are thought to fine-tune color vision \(^6\), though this comes at the cost of decreased visual sensitivity. Deactivation of CYP2J19 likely allows for higher retinal sensitivity when foraging in dim light conditions, as seen in nocturnal owls and kiwis \(^6\). Beyond these key genes, we note that two scotopic photoresponsive genes, TMEM30A (PSG) and KCNV2 (REG), show evidence of selection in penguins, and two others, CNGB1 and GNB3, each have a site mutation unique to penguins (Supplementary Fig. 19c, d). These genes play an important role in the transmission of light (Fig. 4b), and may further enhance visual sensitivity at low light levels, as mutations or loss of these genes impact in the reduced scotopic photoresponse \(^7,8\).

A wholesale reduction in gustation capacity appears to have accompanied the shift to underwater prey capture and consumption in penguins. We verified that penguins only retain genes associated with detecting sour and salty tastants, and lack functional copies of genes linked to umami, sweet and bitter tastants \(^9\) (Figs. 4a and 5a). The mutational loss of capacity for umami taste in penguins is puzzling, given the continued consumption of amino acid-rich prey. Intriguingly, the loss of umami genes is associated with detecting sour and salty tastants, and lack functional copies of genes linked to umami, sweet, and bitter tastants \(^9\). Potential explanations include a lower reliance on taste when swallowing food whole or weakened ability to taste prey due to cold temperatures and the sodium content of seawater (reviewed in \(^50\)).

A strong genomic indicator of diet is presented by chitinases that are expressed in the gastrointestinal tract \(^51\). The chitinase genes (CHIAs) exist as several paralogs, and the retention or loss of these paralogs in mammals has been correlated with diet \(^51\). Retention of intact CHIAs correlates with a higher degree of...
insectivory, and CHIA losses tend to occur in lineages that undergo dietary shifts to carnivory or herbivory. We examined CHIAs in penguins, and in contrast to most examined birds, which have one to four intact CHIAs, penguins have a single pseudogenized CHIA. At first glance, it is perplexing that penguins would lose CHIAs, as many species consume large amounts of crustaceans. Fossil evidence, however, reveals that stem penguins focused primarily on larger prey items like fish and squid, and that adaptations for capturing smaller planktonic prey arose as recently as the Pliocene. We propose that the two inactivating mutations shared by extant penguins (Fig. 5) evolved during a ~50 Ma interval during which stem penguins consumed little or no arthropod prey.

Co-evolution between hosts and pathogens is pervasive in vertebrates. Given the range of different climatic niches occupied by penguins, and the differences in pathogen assemblages to which they are undoubtedly exposed, penguins may have undergone significant adaptation to local pathogen pressures. Accordingly, we detected 51 PSGs in penguins that have a role in immune system that sequesters and scavenges iron has also been detected in mammals, as many species consume large amounts of crustaceans. Fossil evidence, however, reveals that stem penguins focused primarily on larger prey items like fish and squid, and that adaptations for capturing smaller planktonic prey arose as recently as the Pliocene. We propose that the two inactivating mutations shared by extant penguins (Fig. 5) evolved during a ~50 Ma interval during which stem penguins consumed little or no arthropod prey.

Co-evolution between hosts and pathogens is pervasive in vertebrates. Given the range of different climatic niches occupied by penguins, and the differences in pathogen assemblages to which they are undoubtedly exposed, penguins may have undergone significant adaptation to local pathogen pressures. Accordingly, we detected 51 PSGs in penguins that have a role in immune system that sequesters and scavenges iron has also been detected in mammals, as many species consume large amounts of crustaceans. Fossil evidence, however, reveals that stem penguins focused primarily on larger prey items like fish and squid, and that adaptations for capturing smaller planktonic prey arose as recently as the Pliocene. We propose that the two inactivating mutations shared by extant penguins (Fig. 5) evolved during a ~50 Ma interval during which stem penguins consumed little or no arthropod prey.

Discussion

Our comprehensive study encompassing all extant and many fossil penguins provides a new window into the processes that have shaped penguins over the past 60 Ma of evolution. Our phylogenomic analyses confirm the Zealandian origin of penguins, extensive radiation before dispersal to South America and Antarctica, and the second pulse of speciation at the onset of the ACC. Our study reveals new evidence that penguin speciation events were driven by changes in global climate and oceanic dispersal, leading to allopatric speciation across the Southern Hemisphere. Recent speciation in Eudyptes, Megadyptes, Spheniscus, and Eudyptula has been rapid, with a complicated history of gene flow and ILS that make species boundaries within these taxa difficult to untangle (e.g., 5,14).

Importantly, the mechanisms that have shaped penguin diversification in the past (e.g., development of major current systems, geological uplift of oceanic islands) remain important for taxa that appear to still be in the process of speciation today (e.g., between Pygoscelis papua and between Eudyptes chrysolophus chrysolophus/E. c. schlegeli, E. pachyrhynchos/E. robustus, E. chrysoceome/E. filholi/E. mosleyi, and Spheniscus spp 5,14,68).

By comparing our penguin genomes to >300 other avian genomes, we demonstrate that penguins and Procellariiformes have the lowest evolutionary rates observed among birds to date. These low evolutionary rates seem to belie the profound adaptations penguins show for a secondary aquatic existence, but a synthetic reading of the fossil record and the genomic data suggests that penguins rapidly acquired many of the key features associated with their aquatic life very early in their diversification and rates of change slowed towards the present. Genomic signals of molecular adaptations with evidence of positive selection or penguin-specific substitutions were identified in a variety of genes, including genes associated with oceanic diving, thermoregulation, oxygenation, underwater vision, taste, and immunity. Though the overall evolutionary rate in penguins is slow, we identified higher evolutionary rates in crown penguin ancestors than in extant penguins and shifts in rates in individual lineages over the past 14 Ma.

While evolutionary rates and sea surface temperatures appear to be negatively correlated, evolutionary rates and body mass are positively correlated, suggesting that large-bodied species inhabiting colder climates are more equipped to adapt to new environments during climate events. Indeed, our demographic results reveal that penguins have had a complicated history, shaped by climatic oscillations, which has led to population crashes in those species reliant on restricted niches and ecologies. Genomic evidence highlights how some penguin populations collapsed during previous climatic shifts 13,14, and the risks of future collapses are ever-present as penguin populations across the Southern Hemisphere are faced with rapid anthropogenic climate change 69. While our analyses suggest that ocean temperature may regulate certain selection
pressures, the current pace of warming combined with limited refugia in the Southern Ocean will likely far exceed the adaptive capability of penguins\(^{[20]}\). Over 60 Ma these iconic birds have evolved to become highly specialized marine predators, and are now well adapted to some of the most extreme environments on Earth. Yet, as their evolutionary history reveals, they now stand as sentinels highlighting the vulnerability of cold-adapted fauna in a rapidly warming world.

**Methods**

**Genome sequencing, assembly, and annotation.** We analyzed 27 genomes comprising all extant and recently-extinct penguin species, subspecies, and major lineages. 21 of the high-coverage genomes have been published by members of our consortium for this project\(^{[20]}\). To supplement the dataset, we sequenced three high-coverage genomes from the remaining Pygoscelis papua lineages from Falkland Islands/Malvinas “FAL”, Kerguelen Island “KER” and South Georgia “SG” (see\(^{[66]}\)), and partial genomes from the recently-extinct Eudyptes warhami, M. a. richdalei and M. a. waitaha (see ref. \(^{[6]}\) and citations within). See Supplementary Methods for more detail on sample collection, extraction, sequencing, assembly and sex chromosomes. As such, we present the most comprehensive genomic dataset spanning all modern penguins, and to the best of our knowledge, present the first genomic dataset encompassing an entire multi-species vertebrate order. To compare our penguin genomes to other bird genomes, we obtained 361 bird genomes recently released\(^{[20]}\) as part of the BI10K project (https://bi10k.genomics.cn), representing 36 orders and 218 families.

**Additional data on modern and fossil penguins.** We expanded the morphological dataset\(^{[68]}\) by incorporating additional fossil penguin species and seven additional characters. The final matrix comprised 72 fossil and extant penguin taxa, two outgroup taxa, and 281 morphological characters (Supplementary Data 5). The average sea surface temperatures were obtained from spot locations from each lineage (Supplementary Data 3). Generation times of each extant lineage were obtained from the IUCN. For all M. a. antipodes generation time (Supplementary Data 2) (see Supplementary Methods).

**Phylogenomic inference and divergence time estimation.** We combined all penguin genomes with the morphological matrix to resolve the timing and drivers of >60 million years of penguin evolution. In so doing, we update previous phylogenies (e.g., \(^{45,79}\)) to include genomes and morphology from all penguin taxa, including all major P. papua lineages and recently-extinct taxa. To explore the diversification of penguins, we undertook multiple phylogenomic analyses encompassing different subsets of taxa (Fig. 1, Supplementary Figs. 2–4 and Supplementary Software).

We aligned and merged our genomes to the 363-bird alignments from the BI10K project\(^{[20]}\). The final alignments were extracted and multiple hits were filtered out for downstream analyses. We then created four alignments accounting for different subsets of taxa: (1) all putative species, subspecies and lineages (27 penguin taxa + 5 outgroups in total); (2) all extant lineages (24 penguin taxa + 5 outgroups in total), removing Eudyptes warhami, M. a. richdalei and M. a. waitaha from the former alignment; (3) all putative species and subspecies, removing P. papua “FAL”, “KER” and “SG” lineages (21 penguin taxa + 5 outgroups in total); and (4) only putative species (19 penguin taxa in total), further removing Eudyptula minor and BAN. For all taxa, we removed chrysopterus sclegeli. We also created one large genome alignment with all 385-bird taxa (not including Eudyptes warhami, M. a. richdalei and M. a. waitaha) (see Supplementary Methods).

To verify the phylogenetic relationships of modern penguins, we ran coalescent-based and concatenation-based phylogenies accounting for the different subsets of taxa described above (see Supplementary Methods). The topology for all clades was strongly supported and identical using all methods (Supplementary Fig. 2 and Supplementary Data 5), except for the placement of Eudyptes warhami among Eudyptes lineages in a single phylogeny. We estimated the divergence time between modern taxa using the calibration points in ref. \(^{[2]}\) (Supplementary Data 5), except we removed Pygoscelis calderensis based on recent revisions of topology\(^{[79]}\). We also added a “Crown Procellariiformes” (which is a sister to penguins) calibration point to calibrate the divergence between albatross and storm petrels. We also added three tip dates for extinct taxa, using the fossils Madrynornis mirandus, Spheniscus maizoni, and the fossil species MNNZ S.046318 (Eudyptes sp.) (see Supplementary Methods). All trees shared the same topology with our initial analyses, with the exception of the placement of the extinct Megadyptes antipodes waitaha, and had similar divergence times with each other (Supplementary Fig. 2b) We then generated a Bayesian total-evidence dating tree using the fossilized birth-death process (Fig. 1), expanding\(^{[68]}\) by including more species, genome data, and updating the morphology. We also calculated the genetic distances between our modern penguin genomes (Supplementary Fig. 12).

**Ancestral range estimation.** We estimated the ancestral distribution of penguins with the total-evidence dated phylogenomic tree and twelve models, expanding the ancestral areas for all extant and modern penguins. We used the biogeographical areas, distances and normalized distances against the shortest pairwise distance in the time slice in analysis. We then undertook standard model-testing (Likelihood Ratio Test and Akaike information criterion) to identify the best-fitting model for our data. We also used a Biogeographical Stochastic Mapping method to account for the appearance and dispersal/vicariance etc. See Supplementary Methods for more details.

**Quantifying introgression and ILS between taxa.** Controversy still remains regarding taxonomic boundaries between some closely related penguin taxa (see Supplementary Methods for more details). We undertook multiple analyses to assess the discordance of gene trees and levels of ILS and introgression (Supplementary Data 6). We first calculated the frequency of gene tree discordance for each internal branch and summarized the topologies for three different gene tree data sets. We assessed levels of ILS and introgression by quantifying them via internal branch lengths between all species (Supplementary Software). We tested the direction of introgression among lineages and assessed what genomic regions last coalesced, by analyzing phylogenetic groups near likely symmetric phylogenies (Supplementary Data 2). We also examined introgression, by selecting different taxa from different genera and some closely related lineages/species. Finally, we assessed the cessation of gene flow between six closely related penguin groups. See Supplementary Methods for more details.

**Demographic history of penguins.** We undertook analyses of demographic history by profiling heterozygosity across each genome (Supplementary Fig. 10), and undertaking analyses of effective population size \((N_e)\) over the last 1 Ma. As the number of heterosexual sites for M. a. waitaha and Eudyptes warhami remained too low, we only present analyses for M. a. richdalei. We used the species divergence tree as an estimation of the mutation rate and detailed the divergence times in Supplementary Data 2. We focussed on the last 500 Kya, a period encompassing dramatic glacial/interglacial cycles (see Supplementary Methods).

**Comparison of evolutionary rate.** The evolutionary rate among penguins and other birds was compared using both genomic distance and rate comparisons (Supplementary Fig. 12). We calculated P and K2P distances between taxa following the formulas: \(P\) distance \(= p - q\) and \(K2P\)-distance \(= \sqrt{\frac{1}{2}ln(1-2p-q)\). We also created one large genome alignment with all 385-bird taxa (not including Eudyptes warhami, M. a. richdalei and M. a. waitaha) (see Supplementary Methods). All these analyses expand on previous analyses that have only examined A. forsteri and P. adeliae (e.g., \(^{[8,99]}\)), or those that have relied on only on-site analysis for penguins (e.g., \(^{[7]}\)).

To understand the adaptive evolution of specific phenotypes in the branch leading to the last common ancestor of penguins, we identified positively selected genes, rapidly evolving genes, and evolutionarily conserved genes for extant penguins under a branch model and a branch-site model (see Supplementary Methods). We obtained a robust ancient genome for 44 bird species, with a complete assembly of the last common ancestor retaining a total of 8716 high-confidence orthologous genes. These genes were used to conduct a multiple sequence alignment. We then detected positively selected genes/rapidly evolving genes in the branch leading to the last common ancestor of penguins and detected positively selected genes/rapidly evolving genes in the branches of the last common ancestor of penguins. Our findings further demonstrated that these lineages (see Supplementary Methods for more details). Genes with a false discovery rate adjusted \(P\)-value less than 0.05 were treated as candidates for positive selection or rapid evolution (Supplementary Data 4). To reveal more characteristics in penguins, we predicted whether an amino acid substitution site may have an impact on the functional properties of all extant penguins, by comparing penguins to the 83 other birds, and scanning for premature stop codons in each gene alignment. We also examined specific genes individually. In addition, we annotated and undertook further qualitative comparisons of these genes identified in penguins with over 300 other avian species to explore what happens in other birds (Supplementary Data 7). See Supplementary Methods for more details. To support adaptive inferences is highly important, such data remains unrealistic in our study due to cultural and ethical hurdles.

**Putative molecular adaptations.** We undertook comparative genomic analyses across all extant penguin taxa to identify genes and regulatory changes contributing to the remarkable morphological and physiological variation within penguins. We do not include Eudyptes warhami, M. a. richdalei, and M. a. waitaha or additional P. papua lineages (“FAL”, “SG”, “KER”) in these analyses. Our analyses expand on previous analyses that have only examined A. forsteri and P. adeliae (e.g., \(^{[8,99]}\)), or those that have relied on only on-site analysis for penguins (e.g., \(^{[7]}\)). We also estimated the evolutionary rate of penguins using the substitution rate (substitution per site per year) as a summary statistic per site/divergence time. The correlation relationship between the substitution rate and sea surface temperature for extant penguins was tested using a phylogenetic generalized least squares (PGLS) regression (Fig. 3 and Supplementary Data 3). We also conducted PGLS regression analysis to determine the correlation relationship between sea surface temperature and body mass or generation time (Supplementary Software). We also compared the genome size among birds to check whether the genome size has a correlation with the proportion of repeat elements (Supplementary Data 3). See Supplementary Methods for more details.
Behavioral study of gentoo penguin vision. As a representative of penguins, we undertook a behavioral study on captive *P. papua* at SEALIFE Kelly Tarlton’s Aquarium, Auckland, New Zealand to examine their ability to see in the ultraviolet (UV) spectrum. A Tank007 TS566 black OEM 365 nm torch (Shenzhen Grandoor Electronic Co., Ltd., China) was projected onto the snow in the enclosure, and penguins were observed to determine whether they would follow the movements of the torch’s UV projection. At least five penguins appeared to be able to follow the torch’s projection. No such interest was displayed when the torch was turned off, demonstrating that *P. papua* are able to see in the near UV spectrum (Supplementary Movie 1).

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

**Data availability**
The sequencing data and genome assemblies generated in this study have been deposited in the NCBI database under BioProject PRJNA723815 and PRJNA567035, as well as the CNSA of the CNGBdb database under the accession number CNPN0000605. Appendix datasets (BioGenBEARS results and PSMC results) have been deposited on Figshare [https://doi.org/10.6084/m9.figshare.c.5535243.v1]. Supplementary data files and source data generated in this study are provided in the Supplementary Information and Source Data file. The following datasets were also used in this study: CNSA accession number CNPN0000595, and NCBI Genbank accession number NP_990272, NP_001071646, NP_001071647. Source data are provided in this paper.

**Code availability**
Analyses were performed using open-source software tools and the detailed parameters for each tool are shown in the relevant methods in Supplementary Information. The analyses were performed using open-source software tools and the detailed parameters for each tool are shown in the relevant methods in Supplementary Information.
47. Höltér, P. et al. The retinal clock drives the expression of Kcnv2, a channel essential for visual function and cone survival. *Invest. Ophthalmol. Vis. Sci.* 53, 6947–6954 (2012).

48. Vincent, A. et al. Biallelic mutations in GNFB cause a unique form of autosomal-recessive congenital stationary night blindness. *Am. J. Hum. Genet.* 98, 1011–1019 (2016).

49. Zhao, H., Li, J. & Zhang, J. Molecular evidence for the loss of three basic tastes in penguins. *Curr. Biol.* 25, R141–R142 (2015).

50. Wolsan, M. & Sato, J. I. Parallel loss of sweet and umami taste receptor function from phocids and otarioids suggests multiple colonizations of the marine realm by pinnipeds. *J. Biogeogr.* 47, 235–249 (2020).

51. Emerling, C. A., Delsuc, F. & Nachman, M. W. Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. *Sci. Adv.* 4, eaar6478 (2018).

52. Chen, Y. H. & Zhao, H. Evolution of digestive enzymes and dietary diversification in birds. *PeerJ* 5, e8840 (2019).

53. Levy, H. et al. Evidence of pathogen-induced immunogenetic selection across the large geographic range of a wild seabird. * Mol. Biol. Evol.* 37, 1708–1726 (2020).

54. Fiddaman, R. et al. Adaptation and cryptic pseudogenization in penguin toll-like receptors. *Mol. Biol. Evol.* 39, msab354 (2021).

55. Velová, H., Gutowska-Ding, M. W., Burt, D. W. & Vinkler, M. Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. *Mol. Biol. Evol.* 35, 2170–2184 (2018).

56. Yoon, S. I. et al. Structural basis of TLR5-flagellin recognition and signaling. *Science* 335, 859–866 (2012).

57. Zhang, B., Liu, X., Chen, W. & Chen, L. IFIT5 potentiates anti-viral response through enhancing innate immune signaling pathways. *Acta Biochim. Biophys. Sin.* 45, 867–874 (2013).

58. Katibah, G. E. et al. tRNA binding, structure, and localization of the human interferon-induced protein IFIT5. *Mol. Cell* 59, 743–750 (2015).

59. Katibah, G. E. et al. Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. *Proc. Natl Acad. Sci. USA* 111, 12025–12030 (2014).

60. Zhang, J. et al. CD81 is required for hepatitis C virus glycoprotein-mediated viral infection. *J. Virol.* 78, 1448–1453 (2004).

61. Forni, D. et al. Evolutionary analysis provides insight into the origin and adaptation of HCV. *Front. Microbiol.* 9, 854 (2018).

62. Cassat, J. E. & Skaar, E. P. Iron in infection and immunity. *Cell Host Microbe* 13, 509–519 (2013).

63. Sauderson, S. C. et al. Phylogenomic characterization of a novel corinobacterium species associated with fatal diphtheric stomatitis in endangered yellow-eyed penguins. *mSystems* 6, e0032021 (2021).

64. Barber, M. F. & Elde, N. C. Escape from bacterial iron piracy through rapid evolution of transferrin. *Science* 346, 1362–1366 (2014).

65. Mayr, G., Scofield, R. P., De Pietri, V. L. & Tennyson, A. J. A Paleocene fossil Sphenisciformes. *J. Biogeogr.* 46, R141–R142 (2019).

66. Murakami, T. et al. Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. *Nat. Cell Biol.* 11, 1205–1211 (2009).

67. Schoor, M., Schuster-Gossler, K., Roopenian, D. & Gossler, A. Skeletal dysplasia, growth retardation, reduced postnatal survival, and impaired fertility in mice lacking the SNP2/5W1 family member ETIL. *Mech. Dev.* 85, 73–83 (1999).

68. Clucas, G. V. et al. Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. *Mol. Ecol.* 27, 4680–4697 (2018).

69. Forcada, J. & Trathan, P. N. Penguin responses to climate change in the Southern Ocean. *Glob. Chang. Biol.* 15, 1618–1630 (2009).

70. Forcada, J., Trathan, P. N., Reid, K., Murphy, E. J. & Croxall, J. P. Contrasting population changes in sympatric penguin species in association with climate warming. *Glob. Chang. Biol.* 12, 411–423 (2006).

Acknowledgements

We thank the British Antarctic Survey, Institut Polaire Français (IPEV), Laura Seaman, and staff at SEALIFE Kelly Tarlton’s Aquarium, Simone Giovannardi, Misha Vorobyev, David Ainley, Jason Turuwhehua, Nic Dussex, Kieren Mitchell, Damien Fordham, Stuart Brown, James Cahill, Shalanin Liu, Yun Zhao, Fang Li, Min Wu, Yun Wang, Guangli Chen, and B10K members for sample/data collection and discussions. This project was supported by the National Key Research and Development Program of China (MOST) grant (no. 2018YFC1406901) to D.-X.Z. and the International Partnership Program of Chinese Academy of Sciences (no. 152453KYSB20170002) to GZ. This project was also supported by the National Natural Science Foundation of China grant (no. 31901214 and No. 32170626) to S.F. and a Villum Investigator grant (no. 25900) from The Villum Foundation to G.Z. This project was also funded by the China National GeneBank.

Author contributions

Conceptualization: T.L.C., C.Z., M.F., H.P., D.T.K., D.-X.Z. and G.Z.; methodology: T.L.C., C.Z., M.F., H.P., D.T.K., S.R.F., C.A.E., D.R.T., M.R.E., A.L.S., T.A.H., J.R.W., S.G., P.W.H., Y.C., T.M., T.H., M.H.S.S., L.L.B., P.G.R., D.R.T. and G.Z.; Formal analysis: C.Z., M.F., S.R.F., C.A.E., X.B., Q.F., J.R.W. and P.H.; investigation: T.L.C., C.Z., M.F., H.P., S.R.F., C.A.E. and P.B.; Data Curation: C.Z., M.F., D.T.K., S.R.F., C.A.E., D.R.T., X.B., Q.F., M.R.E., S.F. and P.G.B.; resources: M.R.E., A.L.S., A.J.D.T., C.-A.B., Y.C., M.T., T.H., M.H.S.S., L.D.S., R.A.P., P.Q., J.E.M., J.L.B., P.G.R., D.R.T., U.E., P.D., G.M. and P.D.B.; writing—original draft: T.L.C., C.Z., M.F., D.T.K., S.R.F., C.A.E. and G.Z.; writing—review and editing: all authors have reviewed and approved the manuscript; visualization: T.L.C., C.Z. and M.F.; supervision: H.P., A.L.S., R.Z., M.F., H.P. and G.Z.; funding acquisition: D.-X.Z. and G.Z.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-31508-9.

Correspondence and requests for materials should be addressed to Theresa L. Cole or Guoqie Zhang.

Peer review information Nature Communications thanks Claudia Tambussi, Chi-Hing Cheng, Edward Braun, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Biology, University of Washington, Seattle, WA 98195, USA. 13Global Penguin Society, Puerto Madryn 9120, Argentina. 14CESIMAR CCT Cenpat-CONICET, Puerto Madryn 9120 Chubut, Argentina. 15School of Biological Sciences, Faculty of Science, Engineering and Technology, University of Adelaide, North Terrace Campus, Adelaide, SA 5005, Australia. 16Department of Ophthalmology, New Zealand National Eye Centre, Faculty of Medical and Health Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. 17Department of Zoology, University of Otago, Dunedin 9054, New Zealand. 18Centre d'Etudes Biologiques de Chizé (CEBC), UMR 7372 du CNRS-La Rochelle Université, 79360 Villiers-en-Bois, France. 19British Antarctic Survey, Natural Environment Research Council, High Cross, Cambridge CB3 OET, UK. 20Department of Animal Ecology and Systematics, Justus-Liebig-Universität Giessen, Heinrich-Buff-Ring 26, 35392 Giessen, Germany. 21Department of Biological Sciences, Bowling Green State University, Bowling Green, OH 43403, USA. 22FitzPatrick Institute of African Ornithology, University of Cape Town, Rondebosch 7701, South Africa. 23National Institute of Water and Atmospheric Research Ltd., Private Bag 14901, Kilbirnie, Wellington 6241, New Zealand. 24Department of Ecology, Environment and Evolution, La Trobe University, Melbourne, VIC 3086, Australia. 25Department of Marine Science, University of Otago, Dunedin 9016, New Zealand. 26Research Department, Phillip Island Nature Parks, PO Box 97, Cowes, Phillip Island, Cowes, VIC 3922, Australia. 27Division of Pathology and Laboratory Medicine, University of Western Australia, Crawley, WA 6009, Australia. 28Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, TAS 7001, Australia. 29State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China. 30James D. Watson Institute of Genome Sciences, Hangzhou 310029, China. 31Guangdong Provincial Academician Workstation of BGI Synthetic Genomics, BGI-Shenzhen, Shenzhen 518120, China. 32Center for Computational and Evolutionary Biology & State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Beijing 100101, China. 33College of Life Sciences, University of Chinese Academy of Sciences, 19 Yuquan Road, Beijing 100049, China. 34Innovation Center of Yangtze River Delta, Zhejiang University, 314102 Jiashan, China. 35These authors contributed equally: Theresa L. Cole, Chengran Zhou, Miaquan Fang.

✉email: tesscole1990@gmail.com; guojiezhang@zju.edu.cn