

Sirenian genomes illuminate the evolution of fully aquatic species within the mammalian superorder Afrotheria

Ran Tian^{1,12}, Yaolei Zhang^{2,4,9,12}, Hui Kang^{3,11,12}, Fan Zhang¹, Zhihong Jin¹, Jiahao Wang^{2,4}, Peijun Zhang³, Xuming Zhou^{5,6}, Janet M. Lanyon⁷, Helen L. Sneath⁷, Lucy Woolford⁸, Guangyi Fan^{2,4,7,9,10,*}, Songhai Li^{3,11,*}, Inge Seim^{1,*}

¹ Integrative Biology Laboratory, Nanjing Normal University, Nanjing, China

² BGI Research, Qingdao 266555, China

³ Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya, China

⁴ BGI Research, Shenzhen 518083, China

⁵ Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁶ School of Life Sciences, University of Science and Technology of China, Anhui, 230027, China

⁷ School of the Environment, The University of Queensland, St Lucia 4072, Australia

⁸ School of Veterinary Sciences, The University of Adelaide, Roseworthy 5371, Australia

⁹ Qingdao Key Laboratory of Marine Genomics BGI Research, Qingdao 266555, China

¹⁰ State Key Laboratory of Agricultural Genomics, BGI Research, Shenzhen 518083, China

¹¹ Key Laboratory of Aquatic Biodiversity and Conservation, Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, China

¹² These authors contributed equally: Ran Tian, Yaolei Zhang, Hui Kang

* Correspondence: fanguangyi@genomics.cn (G.F.), lish@idsse.ac.cn (S.L.), and inge@seimlab.org (I.S.)

Abstract

Sirenians of the diverse superorder Afrotheria were the first mammals to retransition from land to water and are the sole herbivorous marine mammals, but little is known about the underlying genomic changes. Here, we generated a chromosome-level dugong (*Dugong dugon*) genome. A comparison of our assembly and a recent West Indian manatee genome with other afrotherian genomes revealed possible molecular solutions to aquatic life by sirenians, including a shift in daily activity patterns (circadian clock) and tolerance to a high-iodine plant diet mediated through changes in the iodide transporter NIS (*SLC5A5*) and its co-transporters. Functional *in vitro* assays confirmed that sirenian amino acid substitutions alter the properties of the circadian clock protein PER2 and NIS. Sirenians show evidence of convergent regression of integumentary system (skin and its appendages) genes with cetaceans, the only other fully aquatic mammal group. Our analysis also uncovered gene losses that may be maladaptive in a modern environment, including a candidate gene (*KCNK18*) for sirenian cold stress syndrome likely lost during their evolutionary shift in daily activity patterns. Finally, genomes from nine Australian locations and the functionally extinct Okinawan population confirm and date a genetic break on the Australian east coast and highlight the need for whole-genome resequencing data from small or isolated dugong populations worldwide for conservation and genetic management.

Keywords: sirenians; dugongs; manatees; comparative genomics; adaptations; population genomics

Main

The terrestrial ancestors of the marine mammal groups Sirenia, Cetacea, and Pinnipedia independently transitioned from land to water ¹. The first to leave land, sirenians, emerged around 60 million years ago within the afrotherian herbivorous clade Paenungulata, 10 and 30 million years before the emergence of cetaceans and pinnipeds ². Afrotherian mammals were isolated from other mammals until ~60 Mya when non-afrotherian mammals (ungulates from ~25 Mya) began to enter the African continent from Eurasia and displaced many local species ^{3,4}. This geographic isolation allowed the independent evolution of terrestrial mammals to an aquatic habitat in Africa (sirenians), paralleling the evolution of fully aquatic cetaceans from ungulates.

Dozens of sirenian species have existed in the past, but unlike cetaceans (about 90 extant species) and pinnipeds (about 30 extant species), sirenians are today far less diverse (**Figure 1a**). There are four extant sirenian species: the dugong (*Dugong dugon*) of the family Dugongidae (included the Steller's sea cow, *Hydrodamalis gigas*, that became extinct about 250 years ago) and manatees (family Trichechidae: the West Indian manatee, *Trichechus manatus* (includes the subspecies Florida manatee, *T. m. latirostris*, and Antillean manatee, *T. m. manatus*); the Amazonian manatee, *T. inunguis*; and the African manatee, *T. senegalensis*) ^{1,5}. The dugong and manatees have been found in distinct tropical and subtropical habitats since the middle Miocene (~12.2 Mya) ^{2,5}. The dugong originally dispersed into the Pacific from near Florida, and today inhabits the coastlines of the Indo-Pacific oceans, while the three species of manatee occupy the Atlantic Ocean and associated rivers (**Figure 1c**).

Many genome-scale studies have considered the evolution of marine mammals. However, compared to cetaceans and pinnipeds ^{6,7}, the lack of high-quality genomes has limited insights into sirenians, the only fully aquatic mammals within Afrotheria. To better understand the evolution to aquatic life by sirenians, we generated and juxtaposed a chromosome-level genome and gene set of the dugong with other afrotherian mammals (including a manatee). Our genome also enabled insights into the diversity and demography of dugongs, including the genomic impact of severe population declines on a recently functionally extinct Okinawan population.

Results

Molecular adaptations of sirenians

We used synthetic long-read (stLFR) and chromosome conformation capture (Hi-C) sequencing to generate a 3.06 Gb 25-chromosome genome assembly (Ddugon_BGI) of a female dugong (**Supplementary Note 1, Figure S1, and Tables S1 and S2**). The assembly and gene set BUSCO⁸ completeness scores (94.4% and 91.7%, respectively) were comparable to seven other afrotherian species (**Tables S3 and S4**).

Although similar in appearance, the dugong and manatees are not that closely related. They share a common ancestor (crown Sirenia) 31.2 Mya (95% CI: 27.4-37.0 Mya) (**Figure 1c**), providing a suitable backdrop to illuminate features present in the first fully aquatic sirenians. Our afrotherian dataset, which includes the phylogenetically closest extant species to sirenians (elephants and hyraxes) (**Supplementary Note 2, Figure 1c, and Figure S2**), was interrogated to identify protein-coding gene changes possibly associated with the evolution of all sirenians rather than focusing on their differences (e.g., see **Figure 1a**). Our analysis included gene family dynamics (**Tables S5 and S6, and Figure S3**), positively selected genes (PSGs) (**Table S7**), rapidly evolving genes (REGs) (**Table S8**), gene loss (pseudogenes) (**Tables S9 and Tables S10**), and amino acid substitutions (**Table S11**) shared by sirenians. Amino acid substitutions of interest were validated against 120 mammalian species in OrthoMaM⁹ and by BLAST searches of NCBI and Ensembl databases, while pseudogenes were validated by interrogating genome and transcriptome sequencing reads. This effort identified gene changes that may contribute to aquatic herbivory, sirenian circadian activity patterns, and changes that may underly typical marine mammal features, such as modified cardiovascular, integumental (i.e., skin and associated structures), and sensory (vision, smell, and taste) systems (**Figure 2**). We also report historical gene loss events that may have become maladaptive in modern environments.

Nutrient uptake of fully aquatic herbivores

Sirenians are the only aquatic herbivorous mammals, hence they have a common and misleading name, sea cows. Although sirenians can occasionally consume invertebrates, they rely primarily on aquatic plants, including marine seagrasses whose terrestrial ancestors independently returned to the sea during the Cretaceous period, from ~66 to ~100 Mya^{5,10}. Whilst dugongs are obligate, specialist seagrass grazers, all three species of manatee are more catholic feeders, taking a range of aquatic and semi-aquatic plant species, and in some cases, may consume emergent grassland species⁵. Adult sirenians consume 10-15% of their body weight daily^{5,11}. Similar to the distantly related elephants and hyraxes and some more distantly related mammals (e.g., rhinoceroses), sirenians are hindgut fermenters⁵. Manatees and the dugong have specialized mouthparts^{12,13}, long and capacious digestive tracts, and prolonged food retention times (up to a week in the dugong)¹⁴. These modifications, coupled with unique symbiotic microbes¹⁵, presumably maximize yield from nutrient-poor seagrass.

We observed gene losses consistent with a diet comprising few animal products (**Tables S9 and S10**). The lipid transporter *ABCG8*, a gene essential for transport of dietary cholesterol, is inactivated in the dugong and West Indian manatee. A recent study comparing 60 mammals, including non-herbivorous afrotherians, reported *ABCG8* loss in the West Indian manatee and elephants¹⁶. We also confirm a loss of chitinase 5 (*CHIA5*)¹⁷, a gene essential for the digestion of insect and crustacean exoskeletons, in the West Indian manatee and dugong. *ANPEP* (aminopeptidase N) is under positive selection in sirenians (**Figure 2 and Table S7**). It encodes an enzyme that finalizes the digestion of peptides after their hydrolysis in the intestines, brain, and other cells and tissues. *ANPEP* was previously reported to be under positive selection in sirenians (West Indian manatee), cetaceans (bottlenose dolphin, *Tursiops truncatus*, and killer whale, *Orcinus orca*), and pinnipeds (walrus, *Odobenus rosmarus*) and hypothesized to play a role in reducing oxidative stress in marine mammals by adjusting cysteine and glycine availability in the glutathione metabolism pathway¹⁸. This gene is also under positive selection in the ancestral branch of the

dietary-diverse bat family Phyllostomidae¹⁹ and has lower enzyme activity in herbivorous bird²⁰ and rodent²¹ species. We speculate that these observations hint at modifications of this digestive enzyme in marine mammals consistent with their dietary substrates, in addition to a role in oxidative stress resistance.

Nearshore marine plants and aquatic plants from rivers and swamps are a rich source of iodine, a nutrient required to synthesize thyroid hormones (**Figure 3a**) essential for systemic energy metabolism, thermoregulation, and the integrity of many tissues²²⁻²⁴. Both inadequate and excess iodine uptake can result in thyroid dysfunction. An elevated iodine intake is tolerated in humans, but long-term chronic intake can have adverse effects (e.g., a life-threatening thyroid storm from extreme levels of thyroid hormones)^{22,23}. Excess iodine-intake also increases the production of harmful oxygen species (ROS) by target tissues, including the thyroid gland^{25,26}. A study of captive West Indian manatees found that switching their diet from terrestrial plants to seagrass over 19 days increased blood thyroid hormone levels²⁷ – as expected given the difference in dietary iodine content. Wild manatees showed higher levels of thyroid hormones than any captive diet group²⁷, which we hypothesize resulted from a lifetime on a high-iodine seagrass diet and associated energy metabolism balance. Despite their high-iodine diet, the blood thyroid levels of wild West Indian manatees are unremarkable compared to the tropical, carnivorous Amazon River dolphin (*Inia geoffrensis*)^{27,28}. The above observations support the idea that genomic changes accompanied sirenian evolution from a terrestrial to an iodine-rich aquatic plant diet. In agreement, we observed that nearly all proteins in the thyroid hormone pathway harbor sirenian-specific amino acid substitutions (**Figure 3b** and **Table S11**). These include *TSHR* (thyroid-stimulating hormone receptor), *DUXO2* (dual oxidase 2), *DUOXA2* (dual oxidase maturation factor 2), *TPO* (thyroid peroxidase), *SLC5A5* (solute carrier family 5 member 5), *KCNQ1* (potassium voltage-gated channel subfamily Q member 1), and *KCNE2* (potassium voltage-gated channel subfamily E regulatory subunit 2). We also identified positive selection of *ATP1B4* (ATPase Na⁺/K⁺ transporting family member β 4), *DIO1* (iodothyronine deiodinase 1), and *DUOX2A*, and rapid evolution of *KCNE2* and the thyroid-hormone

binding albumin (*ALB*).

The transmembrane protein SLC5A5 (NIS) is the only known iodide transport protein^{29,30}. After reduction of iodine to iodide (I^-), iodide is transported from the bloodstream into thyroid follicular cells by NIS acting in concert with the potassium transporters KCNQ1 and KCNE2^{31,32} (**Figure 3b**), two proteins which also have multiple sirenian-specific amino acid substitutions. Mutations in human NIS result in congenital I^- deficiency disorders (IDDs)^{29,33}. We identified five sirenian-specific NIS mutations. Four are in transmembrane domains (TMDs), and one is in an extracellular loop (**Figure 3c**). While none of the sirenian-specific residues have been associated with IDD to date, they are close to residues conserved in mammals shown by site-directed mutagenesis to be important for NIS function. Sirenian Leu-142 flanks Tyr-144 of TMD 4³⁴; sirenian Ala-445 of TMD 12 flanks Asn-441, an extracellular region residue thought to mediate NIS structure via α -helix N-capping of TMD 12³⁵; and sirenian Ala-321 is next to Asp-322 in the extracellular loop of TMD 8 and 9³⁶. Sirenian Ser-539 flanks Gly-543 in TMD 13, a residue required for NIS cell surface targeting³⁷. We carried out I^- uptake assays of human and sirenian NIS in HEK293T cells, revealing higher uptake of I^- by dugong and West Indian manatee NIS (**Figure 3d**). Reciprocal site-directed mutagenesis of dugong and human NIS at the five residues further strengthens the evidence for more efficient iodide transport by the sirenian protein (**Figure 3d**). Our data indicate that evolutionary changes in thyroid hormone signaling, including the master regulator of iodine intake and thyroid hormone availability, played a crucial role in the secondary aquatic adaptation of sirenians.

The cardiovascular system

Sirenians forage in relatively shallow waters and perform short (typically a few minutes) but frequent dives⁵. Although, perhaps not as striking as that of cetaceans, the sirenian cardiovascular system can also sustain conditions that would prove dangerous or deadly to their terrestrial relatives. For example, the Amazonian manatee does not dramatically reduce its heart rate during foraging but shows a relatively

moderate 40-50 beats per minute both during dives and at the sea surface ³⁸.

We identified rapid evolution of *SERPINE2* (also known as PN-1 or protease nexin-1), a gene with an emerging role in the cardiovascular system, from cardiac fibrosis to protection against arteriosclerosis and aneurysms ³⁹ (**Figure 2**). *SCN5A* (cardiac Nav1.5 channel α subunit) and *KCNE2* (potassium voltage-gated channel subfamily E regulatory subunit 2) associated with heart rhythm disturbances are rapidly evolving in sirenians. Human mutations in *KCNE2* (also known as LQT6) result in long QT syndrome (LQTS) ⁴⁰. The QT interval is an electrocardiographic measurement that measures the time taken by the cardiac ventricles to depolarize and repolarize at every heartbeat, and its prolongation or shortening is associated with morbidity and mortality ⁴⁰. Two sirenian-specific amino acid changes in *KCNE2* were observed (**Figure 3e** and **Table S11**). The West Indian manatee has a prolonged QT interval compared to Asian and African elephants (*Elephas maximus* and *Loxodonta* spp., respectively) ^{41,42}. However, like cetaceans and pinnipeds, the sirenian cardiovascular system has many unique anatomical features compared to their closest terrestrial relatives and a direct comparison remains challenging ^{41,42}. Thus, although sirenian-specific *KCNE2* amino acid substitutions may be damaging in a human or elephant, such effects on the cardiovascular system may be less severe (or even adaptive) in a species with a cardiac anatomy modified for aquatic life. Instead, we propose that the unique amino acid substitution of sirenian *KCNE2* is primarily associated with their unique diet by serving as a critical constituent of iodide transport (see section above).

The integumentary system

The skin of mammals consists of the epidermis (outer layer), the dermis (middle layer), and the hypodermis (deep layer) tissue. The sirenian epidermis, the outermost skin layer, is restructured compared to its terrestrial sister taxa. While skin appendages (hair follicles, sebaceous glands, and sweat glands) are associated with the epidermis, they project deep into the dermal layer and are absent (sweat glands) or sparsely distributed (e.g., vibrissae, specialized tactile hairs scattered over the body) in

sirenians⁴³. Consistent with previous reports⁴³⁻⁴⁶, we observed a thin epidermis and a thick dermis in the dugong (**Figure 4a**) and the West Indian manatee (**Figure S4**). We identified and validated using dugong epidermis RNA-seq reads the loss of multiple skin-associated genes (**Figure 4b** and **Table S9**). Many of these genes are also lost in cetaceans (**Supplementary Note 3**). In addition, our screen revealed sirenian loss of late cornified envelope (LCE) gene family proteins expressed by the top layer of the epidermis (**Figure 4c**). LCE gene numbers are also reduced in the afrotherian aardvark (*Orycteropus afer*; common ancestor ~80 Mya), suggesting a role in the evolution of its sparsely haired skin⁴⁷.

Daily activity patterns

Most terrestrial mammals rely on circadian rhythmicity, a molecular clock mediated by the sun's daily cycle, to mediate activity patterns^{48,49}. In contrast, many marine mammals inhabiting coastal reef habitats and shallow waters rely heavily on the lunar (moon) cycle and its effect on water depth, food availability, and temperature⁴⁸. The pineal gland is absent or non-functional in sirenians, cetaceans, and some terrestrial mammals. These species do not encode genes associated with the synthesis and reception of melatonin, a hormone essential for light stimulation of the circadian clock⁵⁰⁻⁵⁴. Sirenians do not show a 'classical' diel (24-hour) activity pattern, have unihemispheric slow wave sleep (i.e., one brain half is awake), and appear to respond behaviorally to tidal currents (tides may restrict foraging) and seasonal changes in water temperature^{5,55,56}.

We identified sirenian-specific amino acid substitutions in most core circadian clock genes (**Figure 5a**, **Figure 3** and **Table S11**), including numerous substitutions in all Period circadian regulator genes (five in *PER1*, 20 in *PER2*, and 13 in *PER3*). *PER2* is expressed by circadian pacemaker cells of the hypothalamic suprachiasmatic nucleus (SCN) and binds to cryptochromes (CRY) to regulate light-associated circadian rhythmicity and timing (including sleep patterns)⁵⁷. A coimmunoprecipitation assay showed that dugong *PER2* bound *CRY1* better than wildtype human *PER2* or human *PER2* with a sirenian substitution (C1220P) in the

CRY-binding domain (**Figure 5b,c**). The latter result suggests that at least one of the 19 other sirenian-specific PER2 residues is required for the enhanced CRY-binding. Our analysis also revealed loss of *KCNK18* (also known as TRESK, TWIK-related spinal cord K⁺ channel) (**Figure 5d** and **Tables S9** and **S10**), a circadian clock-regulated ion channel. *KCNK18* is highly expressed in the SCN, and *Kcnk18*^{-/-} mice cannot use light to differentiate between day and night⁵⁸. Taken together, the lack of a pineal gland and a genetic background where many circadian genes have unique amino acid changes or are lost strengthens the idea that the sirenian circadian clock has been recalibrated.

Gene loss and maladaptation in a human-altered environment

A gene can become lost (inactivated) by chance or because it no longer serves a relevant function. Such losses can be functionally compensated by other genes⁵⁹ or may be adaptive⁶⁰. For example, the loss of dozens of genes re-organized the skin of cetaceans⁶¹ and sirenians (see above) over millions of years, allowing semi-aquatic ancestors to become fully aquatic.

Climate change and human activities can influence marine life at all levels. Threats to sirenians, particularly through loss of aquatic plant habitats, include pollution of waterways, fishing operations, and coastal dredging and reclamation, all of which may be exacerbated by changing climate patterns⁶²⁻⁶⁴. While the herbivorous sirenians are at a much lower position in the food web and show relatively low levels of persistent organic pollutants compared to cetaceans at similar geographic locations⁶⁵, they are likely also sensitive to certain manmade chemicals. It follows that such conditions can far outpace natural selection and that historical gene loss that may have been adaptive or tolerated in an ancestral environment can become maladaptive in a modern environment, especially in species with a large generation time⁶⁶. Sirenians are long-lived (~70 years), with a generation time of ~20 years for the West Indian manatee and ~27 years for the dugong^{5,67}. We identified loss of three genes that may be disadvantageous today: *PONI*, *KCNK18*, and *CES3*.

Loss of *PONI* (paraoxonase 1) in most aquatic mammals – sirenians (including

the extinct Steller's sea cow *Hydrodamalis gigas*), cetaceans, pinnipeds, beavers, and the sea otter *Enhydra lutris* – is thought to reduce oxidative stress and associated inflammation induced by diving^{68,69}. *PON1* was also identified in our pseudogene analysis, confirming a single shared premature stop codon event in the dugong and West Indian manatee as well as several independent inactivation mutations (**Tables S9 and S10**, and **Figure S5**). The presumably adaptive loss of *PON1* likely became maladaptive in aquatic mammals after the industrialization of agriculture as the enzyme protects against organophosphate pesticides (the ozone forms of chlorpyrifos and diazinon)⁶⁸.

KCNK18 loss was probably not damaging (see above) and may similarly expose sirenians to natural and anthropogenic threats. Shared and unique *KCNK18* inactivating mutations were observed in the dugong, Steller's sea cow, and West Indian manatee (**Tables S9 and S10**, and **Figure 5d**). Besides the hypothalamic SCN, skin sensory neurons express the gene, and *Kcnk18* knockout mice show elevated pain and avoidance behaviors when exposed to pyrethroid insecticides⁷⁰ or temperatures below 20 °C⁷¹. Given its geographic range (**Figure 1c**), the Florida manatee (i.e., a subspecies of the West Indian manatee) manifests cold stress syndrome (CSS), a condition resulting from prolonged exposure to water temperatures below 20 °C, and characterized by multiple physiological changes and comorbidities of unknown genetic cause^{5,72,73}. We speculate that loss of *KCNK18* decreases manatee temperature stress tolerance (**Figure 5e**) and that CSS is similar to semelparity in marsupials^{74,75} in that a progressive and systemic deterioration of body condition and physiological function is mediated by an endocrine factor, perhaps from elevated levels of the stress hormone cortisol. Extant sirenians have relatively thin blubber compared to cetaceans and have lost *UCPI* (**Supplementary Note 3**), which could render them further susceptible to cold temperatures. Dugongs should arguably also be inherently sensitive to cold temperatures but have a more insulating integument, higher metabolic rate, and live in warmer waters throughout the year than manatees^{45,76}. The extinct Steller's sea cow related to the dugong further evolved a huge body size and thicker blubber to survive in the cold, sub-Arctic environment⁷⁷.

We identified loss of carboxylesterase 3 (*CES3*; also known as *ES3I*) in sirenians (**Table S9** and **S10**), cetaceans (loss of first nine coding exons), and phocids (10-bp deletion in *Phocidae*, the largest pinniped family⁷⁸) (**Figure S6**). Loss of the *CES3* – by the West Indian manatee and killer whale⁵¹, and by four cetaceans and two hippos⁷⁹ – has previously been reported but not discussed. A premature stop codon is shared by all sirenians, while the dugong and Steller’s sea cow share an additional stop codon (**Figure S6**). *CES3* and its homolog *CES1* have been associated with macrophage cholesterol ester metabolism and are both also expressed by the liver and intestines, where they break down ester bonds of endogenous compounds and xenobiotics⁸⁰⁻⁸². While *CES3* expression is much lower than *CES1* and its enzyme has several magnitudes lower catalytic efficiency than *CES1* for many compounds (and, thus, *CES3* loss in marine mammals is likely compensated), carboxylesterase 3 may show exclusive specificity against manufactured compounds such as pesticides.

Diversity and demography of vulnerable and recently extinct dugongs

Despite their numbers, the ~165,000 dugongs in Australian waters are listed as *Vulnerable* (at high risk of extinction in the medium-term future) by the International Union for Conservation of Nature (IUCN)^{83,84}. The number of dugongs in other locations is orders of magnitude lower. The dugong recently became functionally extinct in Chinese^{85,86} and Japanese waters⁸⁷ and is at risk elsewhere in Asia, Oceania, and eastern Africa^{84,88}.

We considered the genomic diversity and demography of dugongs from ten locations (**Figure 6a**). Short-read whole-genome resequencing data were generated from seven locations (99 individuals) spanning 2,000 km of the Australian east coast (from Torres Strait to Moreton Bay, Queensland) (**Table S12**). Resequencing data (one individual carcass each) was also obtained from two other Australian locations, Coogee Beach (New South Wales)⁷⁷ and Exmouth Gulf (Western Australia), and waters off Okinawa (Japan). Principal component analysis (PCA) indicated that the individuals from Exmouth Gulf on the Australian west coast and Okinawa are genetically distinct to dugongs from the Australian east coast (Queensland and

Coogee Beach) (**Figure 6b**). The Coogee Beach individual clustered with Moreton Bay and Hervey Bay individuals. Because this dugong stranded ~750 km from the accepted eastern Australian range during the summer (November), we speculate it represents one of the few instances⁸⁹ of seasonal long-distance ranging of a single individual from a population in close geographic proximity to Moreton Bay.

Pairwise sequential Markovian coalescent (PSMC)⁹⁰ analysis of dugong autosomes was used to track changes in N_e during the Pleistocene (about 2.6 Mya to 20 kya) (**Figure 6c** and **Figure S7**). All dugongs showed a N_e decline in the mid-Pleistocene until ~400-500 kya also observed in a Steller's sea cow population off the Arctic Commander Islands^{77,91}. Individuals from the seven eastern Queensland locations had near-identical demographic histories. The Coogee Beach individual's PSMC curve mirrored the Queensland individuals, agreeing with the above PCA. The Exmouth Gulf individual from the Indian Ocean showed a distinct N_e to the other Australian populations from the Pacific Ocean range but, nevertheless, a curve that likely reflects the similar environmental conditions across the Australian continent (**Figure 6c**). In contrast, the N_e of the individual from Japanese waters declined continuously over the last 400,000 years.

Indian Ocean and Pacific Ocean dugong populations off Australia became separated by a land bridge across the Torres Strait (**Figure 6a**) from ~115 kya until ~7 kya when sea levels rose to their present levels^{5,92} and allowed resumed panmixia with populations from the Indian Ocean and further southward migration. Genetic evidence of this event on dugong demography has been reported by several studies^{67,93,94}. Further interrogation of the 99-individual Queensland dataset, revealed pronounced genetic structure into a northern and a southern group since ~10.2 kya, consistent with a recently reported potential ecological barrier at or near the Whitsunday Islands on the Great Barrier Reef⁹⁵ (**Figure 6a, Supplementary Note 4** and **Figure S8**). We next assessed proxies of genomic diversity, relative nucleotide diversity (π) (**Figure S8e**) and genome-wide heterozygosity (**Figure S8f**). The northern group showed higher diversity, suggesting that individuals from these locations are more 'genetically healthy', consistent with effective population size

differences in the recent past reconstructed using SMC++ (**Figure 6d**). Four complementary methods (XP-EHH⁹⁶, XP-CLR⁹⁷, π , and F_{st}) were used to detect regions with putative selective sweeps between the northern and southern groups. We identified a ~2 megabase region (24.48-26.50 Mb) on chromosome 18 under positive selection in the northern dugong group (**Figure 6e**). This region contains six annotated protein-coding genes (**Table S13**). They comprise three immunoglobulin genes, the nuclear pore transporter *NUP42* (also known as *NUPL2*), ClpXP protease (*CLPX*), and an apparent *GLPX* paralog ('*GLPX-like*'). *NUP42* is expressed by stress granules, membraneless cell compartments formed in response to various environmental stressors⁹⁸, in organisms ranging from mammals⁹⁹ to yeast¹⁰⁰. The mitochondrial protein *CLPX* is required to synthesize oxygen-binding hemoglobin and buffers oxidative stress^{101,102}. While it is difficult to conclude the driving force behind the selective sweeps of *NUP42*, *CLPX*, and *CLPX-like* loci, sirenians are vulnerable to climate change and regional environmental conditions that dramatically alter their nearshore habitats (particularly seagrass growth)⁶³. Northern Queensland has seen distinct climate change events after the formation of the ostensible ecological barrier at the Whitsunday Islands ~10.2 kya. The sea level on the north-east Queensland coast has continuously decreased since about 6,000 years ago^{103,104}. In addition, organisms that inhabit nearshore habitats must also acclimate to and survive oxidative stress stemming from changes in salinity by high seasonal summer rainfall and coastal run-offs¹⁰⁵. Northern Queensland has such environmental conditions.

Extinction events can often be predicted from the genetic history of a species, with loss of genomic diversity reflecting a dwindling population. The 99 dugongs from the Queensland coast had a mean genome-wide heterozygosity of 9.83×10^{-4} (standard deviation (SD) = 6.75×10^{-5} heterozygotes per site) (**Figure S9a**). The heterozygosity of the Exmouth Gulf (7.73×10^{-4}) individual was slightly lower, and markedly lower in the individual from the now functionally extinct Okinawa population (5.65×10^{-4}). The heterozygosity values of the Australian dugongs were similar to a West Indian manatee from Florida (1.08×10^{-3})¹⁸ and the critically endangered Yangtze finless porpoise *Neophocaena asiaeorientalis* ($\sim 1 \times 10^{-3}$)^{106,107} but an order of magnitude

higher than the extinct baiji *Lipotes vexillifer* (1.23×10^{-4})¹⁰⁸ and much higher than the critically endangered vaquita *Phocoena sinu* (9.04×10^{-5})¹⁰⁹.

The length and number of runs of homozygosity (ROHs) in a genome reflect the level of inbreeding and can provide conservation management guidance^{110,111}. Longer ROHs suggest recent inbreeding. The total length and number of ROHs were generally small and similar across the 99 dugongs sampled on the Queensland coast (**Supplementary Note 4**), with no evidence of inbreeding in recent times (since 1,242 years ago) (**Figure 6f** and **Table S14**). The individual from Exmouth Gulf ($F_{\text{ROH}>1\text{Mb}} = 0.021$, nine ROHs spanning 60.1 Mb) showed evidence of inbreeding as recent as nine generations ago (243 years), consistent with a small or isolated population at this time. An illustration of how environmental events may affect population size in this locality, is tropical cyclone Vance (1999) which appears to have resulted in large-scale emigration from Exmouth Gulf, reducing the dugong population from 1,000 to less than 200 individuals in five years¹¹². It follows that other migrations on a population scale in the last two centuries may have improved the genetic diversity of the Exmouth Gulf ancestors. The historical demography (PSMC) of the Okinawa individual aligned with its ROH estimates. One-third of the genome is in ROH segments above one megabase (389 ROHs spanning 934.4 Mb), and there is evidence of inbreeding as recently as 135 years ago ($F_{\text{ROH}>10\text{Mb}} = 0.082$, four ROHs spanning 73.6 Mb). This pattern is consistent with an ancient population bottleneck and subsequent extensive inbreeding since ~135 years ago, mirroring a *Critically Endangered* Sumatran rhinoceros (*Dicerorhinus sumatrensis*) population on the Malay Peninsula¹¹³. Dugongs in Japanese waters were hunted for centuries, and the Okinawan dugong showed a further dramatic reduction in population size from the late 1900s until its likely functional extinction in 2019⁸⁷.

We conclude that while dugongs on the Australian east and west coasts have a genomic diversity comparable to other marine mammals of conservation concern, most populations (as indicated by our resequencing of 99 individuals from the Queensland coast) are in numbers that likely limit inbreeding. Viewed through a genetic lens, these populations can be considered relatively robust and healthy. In

contrast, the dugong from the functionally extinct Okinawan population showed a continuous decline in effective population size, lower genomic diversity, and an ROH pattern consistent with extensive inbreeding for millennia and, likely, climatic ages.

Discussion

By comparing afrotherian genomes and validating against 120 mammal species, we identified unique amino acid substitutions in protein-coding genes that may underlie sirenian-specific features such as tolerance to a high-iodine aquatic plant diet and an activity pattern heavily reliant on lunar tidal currents and water temperature fluctuations. Because the ancestors of extant manatees and the dugong diverged at crown Sirenia ~30 Mya², these gene changes were likely critical for the transition to an aquatic habitat. Many of the same genes of the integumentary system, particularly expressed in the outer layers of the skin^{51,79,114-119}, were lost by both sirenians and cetaceans. This further supports the idea that convergent gene loss occurs in species with similar ecological pressures¹²⁰ (here, the transition to a fully aquatic lifestyle by distantly related species over ~50-60 My of evolution). We also observed gene losses that may be maladaptive in a modern environment, including the previously reported paraoxonase 1 gene (*PON1*)⁶⁸. Sirenian *KCNK18* loss is possibly related to their shift in activity patterns and may render sirenians susceptible to impacts on water temperatures by climate change and human activities. Its role in cold stress syndrome (CSS) should be investigated. Our population genomics analysis offers insights into dugong diversity and demography. We confirm⁹⁵ and (for the first time) date a north-south genetic break on the Australian east coast and reveal a ~2 Mb genetic sweep region that may be associated with differences in salinity or other environmental factors between the north and south coast. The findings from a functionally extinct population (Okinawa, Japan) highlights an urgency to obtain whole-genome resequencing data from dugong populations worldwide to support conservation efforts. Recent human activities in the last 100 years, including overfishing and coastal development that reduced seagrass abundance, likely fast-tracked extinction of dugong populations in Chinese and Japanese waters^{64,85-87},

and is a considerable risk factor elsewhere. Our study reveals new insights into sirenian biology and the transition of terrestrial mammals to an aquatic lifestyle and provides a basis for future genomics explorations.

Online Methods

Sample collection and research ethics. Fetal liver tissue (sample D201106) for reference genome sequencing was obtained from a recently deceased near-term female dugong fetus recovered from a cow that was hunted illegally in the Burrum Heads region of Hervey Bay Queensland in November 2020. This fetal liver sample and a skin sample from a dugong (D110419) sampled after an indigenous subsistence hunt in Torres Strait in 2011 were used for RNA-sequencing.

Skin samples for whole-genome resequencing were collected from 99 dugongs from seven geographic locations on the east coast of Australia: Airlie Beach (AB, $n=3$), Bowling Green Bay (BG, $n=3$), Clairview (CV, $n=9$), Great Sandy Strait (GS, $n=23$), Hervey Bay (HB, $n=24$), Moreton Bay (MB, $n=32$), and Torres Strait (D, $n=5$; prefix ‘TS’ elsewhere in the manuscript) (**Table S12** and **Figure 6a**). Briefly, skin was collected from the dorsum of wild free-swimming dugongs at each location, except Torres Strait, using a handheld scraper device ¹²¹. In the Torres Strait, skin was excised from fresh dugong carcasses post-hunt by local Traditional Owners at Mabiug Island. Skin samples were stored in salt-saturated 20% DMSO and stored frozen at -20 °C until further genetic processing for resequencing (see ⁹⁵). Skin samples for histology were sub-sampled from the Torres Strait specimens and stored in 10% neutral buffered formalin until sectioned and stained.

The dugong skin samples were collected under the following permits issued to J.M.L.: the University of Queensland Animal Ethics Permits SBS/360/14, SBS/181/18, Scientific Purposes Permits WISP07255110 and WISP14654414, Moreton Bay Marine Parks Permit #QS2000 to #QS2010CV L228, Great Sandy Marine Parks Permit QS2010-GS043, and Great Barrier Reef Marine Park Permits #G07=23274:1 and G14/36987.1.

Sampling and species distribution maps were generated using the R package ‘OpenStreetMap’ ¹²²

Genome sequencing. High-molecular weight DNA was extracted from a fetal liver

sample (D201106). DNA quantity, purity and integrity were assessed by Qubit fluorometry (Invitrogen), Nanodrop spectrophotometry (Thermo Fisher Scientific), and pulse-field gel electrophoresis. Single tube long fragment read (stLFR) libraries¹²³ were sequenced on a MGISEQ-2000 sequencer. A total of ~358Gb (~100×) stLFR clean reads were generated after removing low-quality reads, PCR duplicates, and adaptors using SOAPnuke (v1.5)¹²⁴. Hi-C libraries (Lieberman-Aiden et al., 2009) were prepared from the same fetal liver sample. Hi-C data (200 Gb 150 bp paired-end reads) were generated on the BGISEQ-500 platform.

RNA sequencing. RNA from fetal liver (sample D201106) and skin (sample D110419) were sequenced on the BGISEQ-500 platform to generate 86.7 and 96.3 Gb of raw RNA-seq data, respectively.

Genome assembly. We carried out *k*-mer analysis with ~50× clean reads to estimate the genome size (*G*) of dugong using the formula $G = \text{num of } k\text{-mer} / \text{peak-depth}$.

The stLFRdenovo pipeline (<https://github.com/BGI-biotools/stLFRdenovo>), which is based on Supernova v2.11¹²⁵ and customized for stLFR data, was used to generate a *de novo* genome assembly. GapCloser v1.12¹²⁶ and clean stLFR reads (with the barcode removed with https://github.com/BGI-Qingdao/stLFR_barcode_split) were used to fill gaps. Redundans v0.14a¹²⁷ was used to remove heterozygous contigs. Clean paired-end Hi-C reads validated by HiC-Pro v3.2.0_devel¹²⁸ were used to construct chromosome clusters with the 3D de novo assembly (3D DNA) pipeline v170123¹²⁹. The assembly was further improved by interactive correction using Juicebox (v1.11.08)¹³⁰.

Assembly quality was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) BUSCO v5.4.3⁸, employing the gene predictor AUGUSTUS (v3.2.1)¹³¹ and a 9,226-gene BUSCO mammalian lineage data set (mammalia_odb10).

Genome annotation. The sex of the sequenced individual (i.e., to determine if we

could assemble the Y chromosome) was determined by visual inspection of the specimen prior to dissection, as well as BLAST interrogation of African elephant¹³² and published dugong¹³³ sex chromosome genes against our initial stLFR assembly and by comparing the mapping rate of clean stLFR sequencing reads against chromosome X and autosomes of the chromosome-level genome assembly (see the whole-genome resequencing section below for method). All approaches suggested that the sequenced individual was female.

We identified repetitive elements by integrating homology and de novo prediction data. Protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted prediction methods.

Homology-based transposable elements (TE) annotations were obtained by interrogating a genome assembly with known repeats in the Repbase database v16.02¹³⁴ using RepeatMasker v4.0.5 (DNA-level)¹³⁵ and RepeatProteinMask (protein-level; implemented in RepeatMasker). De novo TE predictions were obtained using RepeatModeler v1.1.0.4¹³⁶ and LTRharvest v1.5.8¹³⁷ to generate database for a RepeatMasker run. Tandem Repeat Finder (v4.07)¹³⁸ was used to find tandem repeats (TRs) in the genome. A non-redundant repeat annotation set was obtained by combining the above data.

Protein-coding genes were annotated using homology-based prediction, de novo prediction, and RNA-seq-assisted [generated from fetal liver and skin] prediction methods.

For homology-based prediction, proteins sequences from five mammals [African bush elephant (*Loxodonta africana*) assembly Loxafr3.0; Cape elephant shrew (*Elephantulus edwardii*) assembly EleEdw1.0; armadillo (*Dasypus novaeboracensis*) assembly OryAfe1.0; West Indian manatee (*Trichechus manatus latirostris*) assembly TriManLat1.0, and human (*Homo sapiens*) assembly GRCh38.p12)] were downloaded from NCBI. These protein sequences were aligned to the repeat-masked genome using BLAT v0.36¹³⁹. GeneWise v2.4.1¹⁴⁰ was employed to generate gene structures based on the alignments of proteins to a genome assembly. De novo gene prediction was performed using AUGUSTUS v3.2.3¹⁴¹, GENSCAN v1.0¹⁴², and

GlimmerHMM v3.0.1¹⁴³ with a human training set. For RNA-seq-assisted gene prediction, 150 bp PE reads from fetal liver and skin, generated on an BGI-SEQ 500 instrument, were filtered using Flexbar v3.4.0^{144,145} with default settings (removes reads with any uncalled bases). Any residual ribosomal RNA reads [the majority ostensibly removed by poly(A) selection prior to sequencing library generation] were removed using SortMeRNA v2.1b¹⁴⁶ against the SILVA v119 ribosomal database¹⁴⁷. Transcriptome data (clean reads) were mapped to the assembled genome using HISAT2 v2.1.0¹⁴⁸ and SAMtools v1.9¹⁴⁹, and coding regions were predicted using TransDecoder v5.5.0^{150,151}.

A final, non-redundant reference gene set was generated by merging the three annotated gene sets using EvidenceModeler v1.1.1 (EVM)¹⁵². The gene models were translated into amino acid sequences and used in local BLASTp¹⁵³ searches against the public databases Kyoto Encyclopedia of Genes and Genomes (KEGG; v89.1)¹⁵⁴, NCBI non-redundant protein sequences (NR; v20170924)¹⁵⁵, Swiss-Prot (release-2018_07)¹⁵⁶, TrEMBL (Translation of EMBL [nucleotide sequences that are not in Swiss-Prot]; release-2018_07)¹⁵⁷, and InterPro (v69.0)¹⁵⁸. The gene set was also examined using BUSCO v5.4.3 and its mammalia_odb10 gene set ('transcriptome mode').

Phylogeny and divergence time estimation. Genome and protein sequences of eight afrotherians and Linnaeus's two-toed sloth *Choloepus didactylus* (outgroup) (see **Table S2**) were downloaded from the NCBI or DNA Zoo databases.

We identified single-copy gene family orthologs using OrthoFinder (v2.5.4)^{159,160}. The coding sequences (CDS) for each species were aligned using PRANK v70427^{161,162} and filtered by Gblocks v0.91b¹⁶³ to identify conserved blocks (removing gaps, ambiguous sites, and excluding alignments less than 300 bp in size). Finally, 7,695 single-copy genes were concatenated into supergenes for phylogenetic analyses.

To identify conserved non-exonic elements (CNEs) we generated whole-genome alignments (WGAs) using LASTZ v1.04.22¹⁶⁴ with the parameters

‘H=2000 Y=3400 L=3000 K=2400’ and our dugong reference genome (Ddugon_BGI) as the reference. We then merged aligned sequences using MULTIZ (version 11.2) ¹⁶⁵. To estimate the non-conserved model, we employed phyloFit (v1.4) in the PHAST package ¹⁶⁶ with 4d sites in the Afrotheria alignments and the topology using the default parameters. We ran phastCons with the non-conserved model to estimate the conserved regions with the parameters ‘target-coverage = 0.3, expected-length = 45, rho = 0.31’. Exon regions were excluded from the highly conserved elements to generate 627,279 CNEEs with a total length of 103.9M longer than or equal to 50 bp.

Mitochondrial genomes and protein-coding sequences for the following species were obtained from NCBI GenBank: dugong (NC_003314.1), West Indian manatee (MN105083.1), Asian elephant (*Elephas maximus*; NC_005129_CDS), rock hyrax (*Procavia capensis*; NC_004919.1), lesser hedgehog tenrec (*Echinops telfairi*; NC_002631.2), golden mole (*Chrysochloris asiatica*; NC_004920.1), aardvark (NC_002078.1), Cape elephant shrew (NC_041486.1), and Linnaeus's two-toed sloth (*Choloepus didactylus*; NC_006924.1). We used MARS ¹⁶⁷ to rotate the mitochondrial sequences to the same origin as the dugong.

Separate maximum-likelihood (ML) phylogenetic trees of eight afrotherians and Linnaeus's two-toed sloth (outgroup) were generated with RaxML v8.2.9 ¹⁶⁸ (1,000 bootstrap iterations) using coding sequences from 7,695 genes, 1,127,156 fourfold degenerate sites in the 7,695 genes, 5,508 single-copy Benchmarking Universal Single-Copy Ortholog (BUSCO) genes ^{8,169}, 627,279 conserved non-exonic elements (CNEEs), mitochondrial genomes, and 13 mitochondrial coding sequences. The resulting tree with the highest GTRGAMMA likelihood score was selected as the best tree.

We also used ASTRAL-III v5.6.2 ¹⁷⁰ to generate a coalescent species tree from non-overlapping 20 kb windows (to minimize linkage between subsequent windows from the WGs generated above (see ¹⁷¹). Briefly, after excluding windows from alignments less than 2,000-bp in size and with more than 25% gaps, 152,478 windows were used to generate window-based gene trees (WGTs) using RaxML (with 1,000 bootstraps). WGTs with mean bootstrap support $\geq 80\%$ were input into ASTRAL-III

(with default parameters) to estimate an unrooted species tree. Apart from the mitochondrial trees (its whole genome and protein coding genes were not identical), the nuclear genome-derived ML trees showed the same topology as the ASTRAL WGT tree.

Divergence times between species was estimated using MCMCTree [a Bayesian molecular clock model implemented in PAML v4.8¹⁷²] with the JC69 nucleotide substitution model, and the concatenated whole-CDS supergenes as inputs. We used 100,000 iterations after a burn-in of 10,000 iterations. MCMCTree calibration points (million years ago; Mya) were obtained from TimeTree¹⁷³ [Cape golden mole-lesser hedgehog tenrec (58.4-81.7 Mya), Cape golden mole-Cape elephant shrew (63.0-87.5 Mya), Cape golden mole-aardvark (54.9-89.3 Mya), Cape golden mole-West Indian manatee (76.0-80.8 Mya), Linnaeus's two-toed sloth-West Indian manatee (84.0-97.9 Mya)] and a recent manuscript² that employed a total-evidence approach (incorporating morphological, molecular, temporal, and geographic data from living and fossil species) to estimate a divergence of Trichechidae (i.e., West Indian manatee ancestor) from Dugongidae (i.e., dugong) ancestors 31.7-36.7 Mya.

To further examine the relationship between paenungulates, we also considered retroelements. Retroelements (i.e., LINEs, SINEs, and LTRs) are considered near homoplasy-free markers given their insertion mode¹⁷⁴⁻¹⁷⁷. We employed a recently developed pipeline¹⁷⁶ that requires the 2,000 bp flanking a retroelement insertion site to assign informative phylogenetic markers from pairwise whole-genome alignments (here, with *Dugong_BGI* as the reference genome). The Kuritzin-Kischka-Schmitz-Churakov (KKSC) test¹⁷⁸ was used to assess presence/absence matrixes.

Gene family analysis. Gene family expansion and contraction analysis was performed using CAFE v4.2¹⁷⁹ with the our consensus phylogenetic tree generated (also see **Figure 1c**) as the input. Expanded and contracted gene families on each branch of the tree were detected by comparing the cluster size of each branch with the

maximum-likelihood cluster size of the ancestral node leading to that branch. A smaller ancestral node indicates gene family expansion, whilst a larger ancestral node indicates family contraction. Gene families with a P -value < 0.01 were defined as significantly expanded or contracted in a branch of interest.

Sirenian gene selection. Selection signatures of the 7,695 single-copy gene family orthologs in our nine-species data set were identified using their coding sequences and PAML ‘codeml’ v4.8¹⁷².

We tested for positively selected genes (PSGs) on sirenian branches by comparing branch-site models, allowing a codon site class with (dN/dS ; also known as ‘omega’, ω) >1 along foreground branches, with branch-site null models. We identified sites under positive selection using Bayes’ Empirical Bayes (BEB) in PAML¹⁸⁰ and a Benjamini–Hochberg P -value cut-off set at 0.05.

Rapidly evolving genes [REGs, i.e., genes with an elevated dN/dS] in sirenians were identified using the PAML branch model. The two-ratio model (model=2) allows one ratio for background branches and another for foreground (sirenians) branches, while the one-ratio model (model=0) enforces the same ratio for all branches. Genes with a P -value (computed using the χ^2 statistic) less than 0.05 and a higher ω value in the foreground lineage were considered REGs.

Gene loss in the sirenian lineage. To screen for gene loss in sirenians, defined as genes harboring premature stop codons and/or frameshifts in a species, we employed a previously published approach¹⁸¹. Briefly, the longest human protein sequence for each gene was mapped to genomes of the dugong, West Indian manatee, rock hyrax, and Asian elephant using BLAT v36¹³⁹ and genBlastA v1.0.1¹⁸². Next, the mapped mapped genomic regions and 1,000 bp upstream and downstream were examined for disruptions (nonsense mutations and frameshifts) to the gene coding sequences using GeneWise v2.4.1¹⁴⁰. We removed loci hits belonging to large gene families (including olfactory receptors, zinc finger proteins, and vomeronasal receptors) and predicted

proteins or intronless cDNA/expressed sequences¹⁸³. False positives with disruptive mutations introduced by GeneWise, sequencing errors, or annotation errors were removed following the steps in¹⁸¹. Candidate pseudogenes with multiple disruptions were manually inspected to remove short or low-quality alignments. We also interrogated raw sequencing reads from dugong and West Indian manatee to confirm each disruption.

Gene loss events were dated using the same procedure employed by¹⁸⁴ (based on^{183,185}). Briefly, a gene is assumed to have evolved under selective pressure before it was inactivated and neutrally after being inactivated (i.e., it should accumulate synonymous and non-synonymous mutations at an equal rate). The K_a/K_s value (K) estimated for this entire branch is considered the average of the K_a/K_s value for the part of the branch where the gene was under selection (K_s) and the K_a/K_s value for the part of the branch where the gene evolved neutrally ($K_n=1$), weighted by the proportion of time for which the gene evolved under selection (T_s/T) and neutrally (T_n/T): $K = K_s \times T_s/T + K_n \times T_n/T$, where T represents the time since the split from the last common ancestor. Using the lower and upper bound of the confidence interval for the species divergence time T obtained from TimeTree¹⁷³ and K_a/K_s value for afrotherians with a functional gene, one can calculate a lower and upper bound for T_n as $T_n = T \times (K - K_s) / (1 - K_s)$ to estimate how long a gene has evolved neutrally. A limitation of this method is that sirenian gene loss cannot be dated if a gene of interest is missing or heavily fragmented in any non-sirenian species within our afrotherian dataset.

Sirenian lineage-specific amino acid changes. Amino acid alignments of the 7,695 single-copy gene family orthologs in our nine-species data set (eight afrotherians and Linnaeus's two-toed sloth) and FasParser v2^{186,187} (<https://github.com/Sun-Yanbo/FasParser>) was used to identify amino acid residues specific to the sirenian lineage. We employed KOBAS v3.0¹⁸⁸ (<http://kobas.cbi.pku.edu.cn>) gene enrichment and literature searches to identify genes that may underlie aquatic specializations of sirenians. Putative sirenian-specific

residues of interest were further validated using 120 mammalian sequences downloaded from OrthoMaM v10c⁹ (available at FigShare (<https://figshare.com/s/16ca82c4e5f0a9f97161>), as well as by interrogating NCBI and Ensembl databases using BLAST. Potential functional effects of substitutions were predicted by PolyPhen-2^{189,190} and SIFT^{191,192}. PolyPhen-2 predicts the possible impact of amino acid substitutions on the stability and function of human proteins using structural and comparative evolutionary considerations. SIFT predicts the potential impact of amino acid substitutions or indels on protein function.

NIS radioiodide uptake assay. DNA sequences containing the protein-coding region of dugong, manatee, and human *SLC5A5* (NIS), as well as dugong (L142M, G203W, A321P, A445V, and S539T) and human (M142L, W203G, P321A, V445A, and T539S) sequences where five amino acid substitutions were changed to their reciprocal residues, were synthesized by GenScript Biotech Corp. (Nanjing, China). Each DNA was individually subcloned into the *pcDNA3.1* vector (Invitrogen) to generate ‘NIS plasmids’. All constructs were sequenced to verify their nucleotide sequences. HEK293T cells were cultured in 6-well plates until reaching 50-60% confluency. Cells were transfected with 2 µg NIS plasmid. Empty *pcDNA3.1* vector was used as a control. One day after transfection, HEK293T cells were seeded into 24-well plates and treated with radiopharmaceuticals after 24 hours. Briefly, cells were incubated with Na¹²⁵I (Shanghai Xinke Pharmaceutical, Shanghai, China) at 37 °C with 5% CO₂ for one hour. The Na¹²⁵I added to each well of the culture plate was counted using a radioactivity meter (FJ-391A4) to obtain the total radioactive count T (µCi) per well. After discarding the radioactive supernatant, cells were washed twice with PBS solution. Radioactive Na¹²⁵I absorbed by the HEK293T cells was denoted the cell radioactive count C (µCi). The radioiodide uptake rate is shown as C/T%.

Coimmunoprecipitation and immunoblotting of PER2 and CRY1. DNA sequences containing the protein-coding region of dugong, Asian elephant, and human *PER2* and *CRY1* were synthesized by GenScript Biotech. In addition, a human *PER2*

sequence with a sirenian-specific proline at residue 1220 (C1220) was synthesized. The CRY1 sequences contained a C-terminal 3×FLAG tag, the *PER2* sequences a C-terminal human influenza hemagglutinin (HA) tag. Each of the seven DNA sequences was individually subcloned into the *pcDNA3.1* vector (Invitrogen) and sequenced to confirm their identity. HEK293T cells were cultured in 10-cm Petri dishes until reaching 70-80% confluency and then transfected with 4 µg *CRY1-3×FLAG-pcDNA3.1* and 6 µg *PER2-3×HA-pcDNA3.1*. Two days after transfection, cells were lysed in Western IP Cell Lysis buffer (Beyotime, Nantong, China) supplemented with 1 mM PMSF (Biosharp) and subjected to coimmunoprecipitation. In total, 10% of the cell extracts were retained for input. Cell lysates were incubated with Anti-FLAG M2 Magnetic Beads (Sigma) at 4 °C overnight. After washing three times, the precipitates were resuspended in SDS-PAGE sample buffer, boiled for 5 min, and run on a 6% SDS-PAGE gel. Immunoblotting was done with mouse monoclonal anti-HA or anti-Flag antibodies (Proteintech).

Whole-genome resequencing. Sequencing data from 99 dugongs sampled on the Australian east coast (see ‘Sample collection and research ethics’) were sequenced on a DNBSEQ-G400 RS instrument by BGI-Australia to generate 100-bp paired end reads. We also obtained public sequencing data from Okinawa (Japan; DRR251525; sampled 17 November 2019), Coogee Beach (New South Wales, Australia; ERR5621402; sampled 25 November 2009)⁷⁷, and Exmouth Gulf (Western Australia, Australia; SRR17870680; sampled 3 June 2018). We also obtained 101-bp paired-end WGS reads generated on the Illumina HiSeq 2000 platform (NCBI SRA SRR331137, SRR331139, and SRR331142) from a female West Indian manatee (Lorelei, born in captivity in Florida, USA)¹⁸. The manatee reads were employed as outgroups in population genomics analyses. Raw data were filtered using SOAPnuke v2.1.5¹²⁴ to remove adapters and low-quality reads. For comparative analyses, all samples were down-sampled to ~10× coverage using SAMtools v1.9¹⁴⁹

Because we had a chromosome-level genome of the dugong, we assigned sex to

samples by using the Rx ratio method, where sequencing reads are mapped to a genome with an assigned X chromosome and the number of reads mapping to autosomes are compared to the X-chromosome (normalized by chromosome length)^{193,194}. The Rx ratio approximates 1.0 for females and 0.5 for males. Briefly, we first identified the X chromosome of the West Indian manatee (assembly TriManLat1.0_HiC; chromosome-level genome based on an assembly reported by¹⁸) and dugong (assembly Dugong_BGI) by BLAST searches using the coding sequences of genes evenly distributed across the X chromosome of the African savanna elephant (*Loxodonta africana*)¹³² and dugong Y chromosome genes¹³³. This effort revealed that the ~167 Mb HiC_scaffold_7 and the ~149 Mb chr7 in West Indian manatee and dugong corresponds to their respective X chromosome. Next, we aligned reads (the reads generated prior to the X chromosome-removal step above) to the chromosome-level genomes using bowtie2 v2.3.4.3^{195,196} (parameter: --no-unal to only retained mapping reads), followed by conversion to a BAM file and filtering using SAMtools v1.7^{149,197} (removal of PCR duplicate and retaining reads with a quality score, Q, above 30). Index statistics for BAM files were generated using idxstats in SAMtools and parsed by modifying an R script available at¹⁹⁸. Average sequencing depth was estimated from indexed and sorted BAM files after reads using mosdepth v0.3.3¹⁹⁹ (parameters: -n --fast-mode --by 500). See **Table S12** for sample statistics.

Identification and characterization of SNPs. Including the X chromosome can interfere with demographic history estimates^{90,200,201} and other population genomics analyses^{202,203} (X chromosome SNP effects). Therefore, reads that could be mapped to the X chromosome were removed using bowtie2 (v2-2.2.5)¹⁹⁶ with default parameters, resulting in ~34.14 Tb of clean data with an average sequencing depth of around 11.41-fold. The filtered clean reads were aligned to our Dugong_BGI reference genome using BWA (v0.7.12-r1039)²⁰⁴ with default parameters. SAMtools (v1.2)²⁰⁵ was employed to convert SAM files to BAM format and to sort alignments, followed by the Picard package (v1.54; <https://broadinstitute.github.io/picard>) for duplicate removal. Next, GATK (v4.1.2.0)²⁰⁶ was utilized to realign reads around

InDels and detect SNPs. Briefly, we obtained the genomic variant call format (GVCF) in ERC mode based on read mapping with the parameters ‘-T HaplotypeCaller, -stand-call-conf 30.0 -ERC GVCF’. Joint variant calling was then conducted with the GATK *CombineGVCFs* module. Lastly, the GATK’s *VariantFiltration* module was used for hard filtering with the parameters ‘--filter-name LowQualFilter --filterExpression QD < 2.0 || MQ < 40.0 || FS > 60.0 || ReadPosRankSum < -8.0 || MQRankSum < -12.5 || SOR > 3.0’, as recommended by GATK. This process generated 61,741,769 SNPs.

Analysis of population structure. To quantify the genetic structure of dugong populations, we first carried out SNPs filtration using vcftools (v0.1.16)²⁰⁷ with the parameters ‘-max-missing 0.95’. Plink (v1.90b6.6)²⁰⁸ was used to perform SNP quality control with the parameters ‘-geno 0.1 --maf 0.01’. Linkage disequilibrium was also used as a criterion to filter SNPs for downstream analysis using Plink with the parameters ‘-indep-pairwise 50 5 0.2’. Next, ADMIXTURE (v1.3.0)²⁰⁹, with the parameters ‘--cv -j20 -B5’ for multiple repeats, was used to perform ancestry inference. To construct the population evolutionary tree, an identity by state (IBS) distance matrix was constructed using Plink with the parameters ‘--distance 1-ibs’. MEGA7²¹⁰ was used to construct the Neighbor-Joining Tree (NJ tree) based on the IBS matrix, and the evolutionary tree was visualized using the iTOL online tool²¹¹. In addition, PCA (principal component analysis) was performed using Plink with the parameters ‘--make-rel --pca 3’ and visualized by the R ggplot2 package²¹².

Estimation of genome heterozygosity and runs of homozygosity. To estimate the heterozygosity of each dugong individual, we used the Plink function ‘—het’ to detect heterozygous SNPs from the final SNP dataset. Additionally, we estimated runs of homozygosity (ROH) using the R package *detectRUNS*²¹³ with the parameters ‘windowSize = 50, minSNP = 30, maxGap= 1000000, minLengthBps = 100000, minDensity = 1/100000’ based on the same filtered SNPs set used in the ‘population structure analysis’ section.

ROH's per dugong generation was calculated as follows. FASTPRR (v2)²¹⁴ was used to estimate the recombination rate of the dugong population based on the 99 Queensland individuals. It was estimated to be ~1.08cM/Mb. The coalescent times of ROH for each subpopulation were calculated as $g = 100/(2rL)$ (see²¹⁵), where g is the expected time (in generations; 27 years⁶⁷) back to the parental common ancestor, r is the recombination rate and L is the length of the ROH in megabases. Thus, ROHs <100kb in length are estimated to result from inbreeding more than 12,501 years ago; ROHs>100 kb, up to 12,501 years ago (463 generations); ROHs>1 Mb, up to 1,242 years ago (46 generations); ROHs>5 Mb, up to 243 years ago (9 generations); and ROHs>10 Mb, up to 135 years ago (5 generations).

Identification of selective sweep regions

Four different methods, including *Fst* (population differentiation), π (relative nucleotide diversity), *XP-EHH* (the cross-population extended haplotype homozygosity statistic)⁹⁶, and *XP-CLR* (the cross-population composite likelihood ratio test)⁹⁷ were employed to assess selective sweeps between the northern and southern groups. Briefly, vcfTools (v0.1.16)²⁰⁷ with the parameters ‘-max-missing 0.95 --maf 0.01’ was used to perform autosome SNPs quality control. Next, the pairwise *Fst* was calculated between the seven groups and between the whole northern and southern groups using vcfTools with the parameter ‘--fst-window-size 10000 --fst-window-step 2000’. π was calculated for each of the seven groups and the whole northern and southern groups using vcfTools with the parameters ‘- -window-pi 10000 --window-pi-step 2000’. XP-CLR scores were calculated using xpcr v1.1.2²¹⁶ with default parameters. Because XP-EHH requires the genetic distance between adjacent SNPs, we considered a chromosome segment of 1 Mb to be 1 cM. The filtered SNPs were phased using ‘vcf_phase.py’, with the parameter “--phase-algorithm beagle” from the PPP (v0.1.12) pipeline²¹⁷, followed by vcfTools with the parameter ‘—IMPUTE’. Next, the *xpehhbin* module from Hapbin (v1.3.0)²¹⁸ was used to calculate XP-EHH values (see²¹⁹). The XP-EHH scores were normalized and corresponding *P* values were calculated. If a *P* value is less than 0.01, we considered the

region as a candidate sweep region. An XP-EHH score is directional: a positive score suggests that selection occurred in the northern group; a negative score, the southern group. Genomic regions that overlapped between all four methods were considered candidate selective sweep regions.

Demographic history inference. The generation time (g) and mutation (μ) rate are necessary to infer the demographic history of populations. The estimated generation time of the dugong (~27 years) has been reported previously⁶⁷. We estimated the mutation rate for dugong using r8s²²⁰ and the single copy orthologous genes described in the ‘Phylogeny and divergence time estimation’ section. The final estimated mutation rate was 2.6×10^{-8} per site per generation.

The pairwise sequentially Markovian coalescent method (PSMC) v0.6.5-r67²²¹ was employed to infer historical effective population size (N_e) fluctuations in dugongs. PSMC can quantitatively reveal changes in N_e from approximately 1 million to 20 thousand years ago⁹⁰. We first constructed diploid consensus sequences for each sample using SAMtools (v1.9)²⁰⁵ mpileup and BCFtools (v1.4)²²² with the parameters ‘-C50’ and ‘-d 4 -D 24’. The consensus sequences were transformed to PSMC input format using fq2psmcfa with the parameter ‘-q20’. Finally, PSMC was used to infer the population history with the parameters ‘-N25 -t15 -r5 -p 4 + 25*2 + 4+6’ and 100 rounds of bootstrapping. SMC++ (v1.15.5)²²³ was used to infer more recent population history for Queensland dugong individuals grouped into two distinct populations (see ‘Analysis of population structure’ above). The SMC++ modules *vcf2smc*, *estimate*, and *plot* were used.

Data availability

The data for *D. dugon* including stLFR data, Hi-C data, RNA-seq data, resequencing data, genome assembly, and gene annotation in this study has been deposited at the China National GeneBank Nucleotide Sequence Archive (CNSA: <https://db.cngb.org/cnsa>) under accession number CNP0004405. Multiple sequence alignments (MSAs) of thyroid hormone pathway and circadian clock genes with

sirenian-specific amino acid substitutions are available on FigShare (<https://figshare.com/s/16ca82c4e5f0a9f97161>).

Acknowledgements

We thank those who assisted with Australian dugong tissue collection, including The University of Queensland Dugong Team (especially Alex McGowan, Erin Neal, and Rob Slade) members of the Mabuiag Island community of Torres Strait (especially Terrence Whap), and Steve Hoseck (Southern Marine Parks, Queensland Parks and Wildlife) for facilitating access to the dugong fetus used for stLFR and RNA-sequencing. We also thank the laboratory of Prof. Kai Yang at Soochow University (China) for their radioisotope expertise and access to equipment, and Dr. Erina Young at Murdoch University (Australia) for information on the dugong individual from Exmouth Gulf, Western Australia.

Support for this research was provided by the National Natural Science Foundation of China (grant 32270441 to R.T. and grant 42225604 to S.L.) National Key Programme of Research and Development, Ministry of Science and Technology (grant 2022YFF1301601 to R.T.); Sea World Foundation (Australia) and Winifred Violet Scott Foundation grants (to J.M.L.); the ‘One Belt and One Road’ Science and Technology Co-operation Special Program of the International Partnership Program of the Chinese Academy of Sciences (grant 183446KYSB20200016 to S.L.); the Specially-appointed Professor Program of Jiangsu Province (to I.S.), the Jiangsu Foreign Expert Bureau (to I.S.), and the Jiangsu Provincial Department of Technology (grant JSSCTD202142 to I.S.).

Author contributions

R.T., J.M.L., G.F., S.L., and I.S. conceived the study; Y.Z., P.Z., J.M.L., H.L.S. collected or curated samples/specimens; R.T., Z.J., and L.W. performed laboratory work; R.T., Y.Z., H.K., F.Z., J.W., and I.S. performed computational biology work; R.T., S.L., G.F., and I.S. managed the project; R.T. and I.S. wrote the original draft; all authors commented on and proof-read the manuscript, with significant contributions

from J.M.L., G.F., and S.L.

Competing interests

The authors declare no competing interests.

References

- 1 Jefferson, T. A., Webber, M. A. & Pitman, R. L. *Marine mammals of the world: a comprehensive guide to their identification*. (Elsevier, 2015).
- 2 Heritage, S. & Seiffert, E. R. Total evidence time-scaled phylogenetic and biogeographic models for the evolution of sea cows (Sirenia, Afrotheria). *PeerJ* **10**, e13886, doi:10.7717/peerj.13886 (2022).
- 3 Monadjem, A. *AFRICAN ARK: Mammals, Landscape and the Ecology of a Continent*. (NYU Press, 2023).
- 4 Springer, M. S. Afrotheria. *Curr Biol* **32**, R205-R210, doi:10.1016/j.cub.2022.02.001 (2022).
- 5 Marsh, H., O'Shea, T. J. & Reynolds III, J. E. *Ecology and conservation of the Sirenia: dugongs and manatees*. (Cambridge University Press, 2012).
- 6 Cabrera, A. A. *et al.* A genetic perspective on cetacean evolution. *Annual Review of Ecology, Evolution, and Systematics* **52**, 131-151 (2021).
- 7 Yuan, Y. *et al.* Comparative genomics provides insights into the aquatic adaptations of mammals. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2106080118 (2021).
- 8 Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol Biol* **1962**, 227-245, doi:10.1007/978-1-4939-9173-0_14 (2019).
- 9 Scornavacca, C. *et al.* OrthoMaM v10: Scaling-Up Orthologous Coding Sequence and Exon Alignments with More than One Hundred Mammalian Genomes. *Mol Biol Evol* **36**, 861-862, doi:10.1093/molbev/msz015 (2019).
- 10 Waycott, M., Biffin, E. & Les, D. H. in *Seagrasses of Australia: Structure, Ecology and Conservation* 129-154 (Springer, 2018).
- 11 Best, R. C. Foods and feeding habits of wild and captive Sirenia. *Mammal Review* **11**, 3-29 (1981).
- 12 Lanyon, J. M. & Sanson, G. D. Mechanical disruption of seagrass in the digestive tract of the dugong. *Journal of Zoology* **270**, 277-289 (2006).
- 13 Lanyon, J. M. & Sanson, G. D. Degenerate dentition of the dugong (*Dugong dugon*), or why a grazer does not need teeth: morphology, occlusion and wear of mouthparts. *Journal of Zoology* **268**, 133-152 (2006).
- 14 Lanyon, J. & Marsh, H. Digesta passage times in the dugong. *Australian Journal of Zoology* **43**, 119-127 (1995).
- 15 Eigeland, K. A. *et al.* Bacterial community structure in the hindgut of wild and captive dugongs (*Dugong dugon*). *Aquatic Mammals* **38** (2012).
- 16 Sharma, V. & Hiller, M. Losses of human disease-associated genes in placental mammals. *NAR Genom Bioinform* **2**, lqz012, doi:10.1093/nargab/lqz012 (2020).
- 17 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, doi:10.1126/sciadv.aar6478 (2018).
- 18 Foote, A. D. *et al.* Convergent evolution of the genomes of marine mammals. *Nat Genet* **47**, 272-275, doi:10.1038/ng.3198 (2015).
- 19 Potter, J. H. T. *et al.* Dietary Diversification and Specialization in Neotropical Bats Facilitated by Early Molecular Evolution. *Mol Biol Evol* **38**, 3864-3883, doi:10.1093/molbev/msab028 (2021).
- 20 Newman, J. *et al.* Low activities of digestive enzymes in the guts of herbivorous grouse (Aves: Tetraoninae). *Journal of Ornithology* **162**, 477-485 (2021).
- 21 Sabat, P., Lagos, J. A. & Bozinovic, F. Test of the adaptive modulation hypothesis in rodents: dietary flexibility and enzyme plasticity. *Comp Biochem Physiol A Mol Integr Physiol* **123**, 83-87, doi:10.1016/s1095-6433(99)00042-2 (1999).
- 22 Leung, A. M. & Braverman, L. E. Consequences of excess iodine. *Nat Rev Endocrinol* **10**, 136-142, doi:10.1038/nrendo.2013.251 (2014).
- 23 Yun, A. J. & Doux, J. D. Iodine in the Ecosystem: an Overview. *Comprehensive Handbook of Iodine*. Eds VR Preedy, GN Burrow and R. Watson. Elsevier Inc, 119-123 (2009).
- 24 Hohmann, G., Ortmann, S., Remer, T. & Fruth, B. Fishing for iodine: what aquatic foraging by bonobos tells us about human evolution. *BMC Zoology* **4**, 5, doi:10.1186/s40850-019-0043-z (2019).
- 25 Yu, M. *et al.* Excess iodide-induced reactive oxygen species elicit iodide efflux via beta-tubulin-associated CIC-3 in thyrocytes. *Biochem J* **479**, 629-640, doi:10.1042/BCJ20210709 (2022).
- 26 Arriagada, A. A. *et al.* Excess iodide induces an acute inhibition of the sodium/iodide symporter in thyroid male rat cells by increasing reactive oxygen species. *Endocrinology* **156**, 1540-1551, doi:10.1210/en.2014-1371 (2015).
- 27 Ortiz, R. M., Mackenzie, D. S. & Worthy, G. A. Thyroid hormone concentrations in captive and free-ranging West Indian manatees (*Trichechus manatus*). *Journal of Experimental Biology* **203**, 3631-3637 (2000).
- 28 Robeck, T. R. *et al.* Thyroid hormone concentrations associated with age, sex, reproductive status and apparent reproductive failure in the Amazon river dolphin (*Inia geoffrensis*). *Conserv Physiol* **7**, coz041, doi:10.1093/conphys/coz041 (2019).
- 29 Ravera, S., Reyna-Neyra, A., Ferrandino, G., Amzel, L. M. & Carrasco, N. The Sodium/Iodide Symporter (NIS): Molecular Physiology and Preclinical and Clinical Applications. *Annu Rev Physiol* **79**, 261-289, doi:10.1146/annurev-physiol-022516-034125 (2017).
- 30 Portulano, C., Paroder-Belenitsky, M. & Carrasco, N. The Na⁺/I⁻ symporter (NIS): mechanism and medical impact. *Endocr Rev* **35**, 106-149, doi:10.1210/er.2012-1036 (2014).
- 31 Purtell, K. *et al.* The KCNQ1-KCNE2 K(+) channel is required for adequate thyroid I(-) uptake. *FASEB J* **26**, 3252-3259, doi:10.1096/fj.12-206110 (2012).
- 32 Roepke, T. K. *et al.* Kcne2 deletion uncovers its crucial role in thyroid hormone biosynthesis. *Nat Med* **15**, 1186-1194, doi:10.1038/nm.2029 (2009).
- 33 Reyna-Neyra, A. *et al.* The Iodide Transport Defect-Causing Y348D Mutation in the Na(+)/I(-) Symporter Renders the Protein Intrinsically Inactive and Impairs Its Targeting to the Plasma Membrane. *Thyroid* **31**, 1272-1281, doi:10.1089/thy.2020.0931 (2021).
- 34 Ravera, S. *et al.* Structural insights into the mechanism of the sodium/iodide symporter.

- Nature* **612**, 795-801, doi:10.1038/s41586-022-05530-2 (2022).
- 35 Li, W., Nicola, J. P., Amzel, L. M. & Carrasco, N. Asn441 plays a key role in folding and function of the Na⁺/I⁻ symporter (NIS). *FASEB J* **27**, 3229-3238, doi:10.1096/fj.13-229138 (2013).
- 36 Li, C. C. *et al.* Conserved charged amino acid residues in the extracellular region of sodium/iodide symporter are critical for iodide transport activity. *J Biomed Sci* **17**, 89, doi:10.1186/1423-0127-17-89 (2010).
- 37 De la Vieja, A., Ginter, C. S. & Carrasco, N. Molecular analysis of a congenital iodide transport defect: G543E impairs maturation and trafficking of the Na⁺/I⁻ symporter. *Mol Endocrinol* **19**, 2847-2858, doi:10.1210/me.2005-0162 (2005).
- 38 Gallivan, G. J., Kanwisher, J. W. & Best, R. C. Heart rates and gas exchange in the Amazonian manatee (*Trichechus inunguis*) in relation to diving. *J Comp Physiol B* **156**, 415-423, doi:10.1007/BF01101104 (1986).
- 39 Madjene, C., Boutigny, A., Bouton, M. C., Arocas, V. & Richard, B. Protease Nexin-1 in the Cardiovascular System: Wherefore Art Thou? *Front Cardiovasc Med* **8**, 652852, doi:10.3389/fcvm.2021.652852 (2021).
- 40 Crump, S. M. & Abbott, G. W. Arrhythmogenic KCNE gene variants: current knowledge and future challenges. *Front Genet* **5**, 3, doi:10.3389/fgene.2014.00003 (2014).
- 41 Storlund, R. L., Rosen, D. A. S. & Trites, A. W. Electrocardiographic Scaling Reveals Differences in Electrocardiogram Interval Durations Between Marine and Terrestrial Mammals. *Front Physiol* **12**, 690029, doi:10.3389/fphys.2021.690029 (2021).
- 42 Siegal-Willott, J. *et al.* Electrocardiography in two subspecies of manatee (*Trichechus manatus latirostris* and *T. m. manatus*). *J Zoo Wildl Med* **37**, 447-453, doi:10.1638/05-086.1 (2006).
- 43 Berta, A., Sumich, J. L. & Kovacs, K. M. in *Marine Mammals (Third Edition)* (eds Annalisa Berta, James L. Sumich, & Kit M. Kovacs) 169-210 (Academic Press, 2015).
- 44 Kipps, E., Mclellan, W. A., Rommel, S. & Pabst, D. A. Skin density and its influence on buoyancy in the manatee (*Trichechus manatus latirostris*), harbor porpoise (*Phocoena phocoena*), and bottlenose dolphin (*Tursiops truncatus*). *Marine Mammal Science* **18**, 765-778 (2002).
- 45 Horgan, P., Booth, D., Nichols, C. & Lanyon, J. M. Insulative capacity of the integument of the dugong (*Dugong dugon*): thermal conductivity, conductance and resistance measured by in vitro heat flux. *Marine biology* **161**, 1395-1407 (2014).
- 46 Chernova, O., Kiladze, A. & Shpak, O. in *Doklady Biological Sciences*. 150-156 (Springer).
- 47 Shoshani, J., Goldman, C. A. & Thewissen, J. *Orycteropus afer*. *Mammalian species*, 1-8 (1988).
- 48 Andreatta, G. & Tessmar-Raible, K. The Still Dark Side of the Moon: Molecular Mechanisms of Lunar-Controlled Rhythms and Clocks. *J Mol Biol* **432**, 3525-3546, doi:10.1016/j.jmb.2020.03.009 (2020).
- 49 Hazlerigg, D. G. & Tyler, N. J. C. Activity patterns in mammals: Circadian dominance challenged. *PLoS Biol* **17**, e3000360, doi:10.1371/journal.pbio.3000360 (2019).
- 50 Emerling, C. A. *et al.* Genomic evidence for the parallel regression of melatonin synthesis and signaling pathways in placental mammals [version 2; peer review: 2 approved]. *Open Res Eur* **1**, 75, doi:10.12688/openreseurope.13795.2 (2021).

- 51 Huelsmann, M. *et al.* Genes lost during the transition from land to water in cetaceans highlight genomic changes associated with aquatic adaptations. *Sci Adv* **5**, eaaw6671, doi:10.1126/sciadv.aaw6671 (2019).
- 52 Yin, D. *et al.* Gene Duplication and Loss of AANAT in Mammals Driven by Rhythmic Adaptations. *Mol Biol Evol* **38**, 3925-3937, doi:10.1093/molbev/msab125 (2021).
- 53 Valente, R., Alves, F., Sousa-Pinto, I., Ruivo, R. & Castro, L. F. C. Functional or Vestigial? The Genomics of the Pineal Gland in Xenarthra. *J Mol Evol* **89**, 565-575, doi:10.1007/s00239-021-10025-1 (2021).
- 54 Lopes-Marques, M. *et al.* The Singularity of Cetacea Behavior Parallels the Complete Inactivation of Melatonin Gene Modules. *Genes (Basel)* **10**, doi:10.3390/genes10020121 (2019).
- 55 Mascetti, G. G. Unihemispheric sleep and asymmetrical sleep: behavioral, neurophysiological, and functional perspectives. *Nat Sci Sleep* **8**, 221-238, doi:10.2147/NSS.S71970 (2016).
- 56 Zeh, D. R. *et al.* Evidence of behavioural thermoregulation by dugongs at the high latitude limit to their range in eastern Australia. *Journal of Experimental Marine Biology and Ecology* **508**, 27-34 (2018).
- 57 Narasimamurthy, R. & Virshup, D. M. The phosphorylation switch that regulates ticking of the circadian clock. *Mol Cell* **81**, 1133-1146, doi:10.1016/j.molcel.2021.01.006 (2021).
- 58 Lalic, T. *et al.* TRESK is a key regulator of nocturnal suprachiasmatic nucleus dynamics and light adaptive responses. *Nat Commun* **11**, 4614, doi:10.1038/s41467-020-17978-9 (2020).
- 59 Szamecz, B. *et al.* The genomic landscape of compensatory evolution. *PLoS Biol* **12**, e1001935, doi:10.1371/journal.pbio.1001935 (2014).
- 60 Dubois-Mignon, T. & Monget, P. Gene essentiality and variability: What is the link? A within- and between-species perspective. *Bioessays* **44**, e2200132, doi:10.1002/bies.202200132 (2022).
- 61 Espregueira Themudo, G. *et al.* Losing genes: the evolutionary remodeling of cetacea skin. *Frontiers in Marine Science* **7**, 912 (2020).
- 62 Florida Fish and Wildlife Conservation Commission. *Manatee mortality event along the East Coast: 2020–2022*, Accessed Feb 15, 2023. <https://myfwc.com/research/manatee/rescue-mortality-response/ume>, (2022).
- 63 Marsh, H., Arraut, E. M., Diagne, L. K., Edwards, H. & Marmontel, M. Impact of climate change and loss of habitat on Sirenians. *Marine Mammal Welfare: Human Induced Change in the Marine Environment and its Impacts on Marine Mammal Welfare*, 333-357 (2017).
- 64 Du, J., Chen, B., Nagelkerken, I., Chen, S. & Hu, W. Protect seagrass meadows in China's waters. *Science* **379**, 447, doi:10.1126/science.adg2926 (2023).
- 65 Weijs, L., Leusch, F. & Covaci, A. Concentrations of legacy persistent organic pollutants and naturally produced MeO-PBDEs in dugongs (Dugong dugon) from Moreton Bay, Australia. *Chemosphere* **229**, 500-508, doi:10.1016/j.chemosphere.2019.05.033 (2019).
- 66 Whitehead, A., Clark, B. W., Reid, N. M., Hahn, M. E. & Nacci, D. When evolution is the solution to pollution: Key principles, and lessons from rapid repeated adaptation of killifish (*Fundulus heteroclitus*) populations. *Evol Appl* **10**, 762-783, doi:10.1111/eva.12470 (2017).
- 67 McDonald, B. J. *Population genetics of dugongs around Australia: implications of gene flow and migration*, James Cook University, (2005).
- 68 Meyer, W. K. *et al.* Ancient convergent losses of Paraoxonase 1 yield potential risks for

- modern marine mammals. *Science* **361**, 591-594, doi:10.1126/science.aap7714 (2018).
- 69 Graham, A. M. *et al.* Reduction of Paraoxonase expression followed by inactivation across independent semiaquatic mammals suggests stepwise path to pseudogenization. *Mol Biol Evol* **40**, doi:10.1093/molbev/msad104 (2023).
- 70 Castellanos, A. *et al.* Pyrethroids inhibit K2P channels and activate sensory neurons: basis of insecticide-induced paraesthesias. *Pain* **159**, 92-105, doi:10.1097/j.pain.0000000000001068 (2018).
- 71 Castellanos, A. *et al.* TRESK background K(+) channel deletion selectively uncovers enhanced mechanical and cold sensitivity. *J Physiol* **598**, 1017-1038, doi:10.1113/JP279203 (2020).
- 72 Hardy, S. K., Deutsch, C. J., Cross, T. A., de Wit, M. & Hostetler, J. A. Cold-related Florida manatee mortality in relation to air and water temperatures. *PLoS One* **14**, e0225048, doi:10.1371/journal.pone.0225048 (2019).
- 73 Bossart, G. D., Meisner, R. A., Rommel, S., Ghim, S.-j. & Jenson, A. B. Pathological features of the Florida manatee cold stress syndrome. *Aquatic Mammals* **29**, 9-17 (2003).
- 74 Tian, R. *et al.* A chromosome-level genome of *Antechinus flavipes* provides a reference for an Australian marsupial genus with male death after mating. *Mol Ecol Resour* **22**, 740-754, doi:10.1111/1755-0998.13501 (2022).
- 75 Naylor, R., Richardson, S. J. & McAllan, B. M. Boom and bust: a review of the physiology of the marsupial genus *Antechinus*. *J Comp Physiol B* **178**, 545-562, doi:10.1007/s00360-007-0250-8 (2008).
- 76 Lanyon, J. M., Horgan, P., Booth, D. & Nichols, C. Reply to the Comment of Owen et al. on “Insulative capacity of the integument of the dugong (*Dugong dugon*): thermal conductivity, conductance and resistance measured by in vitro heat flux” by Horgan, Booth, Nichols and Lanyon (2014). *Marine Biology* **162**, 1147-1149 (2015).
- 77 Le Duc, D. *et al.* Genomic basis for skin phenotype and cold adaptation in the extinct Steller's sea cow. *Sci Adv* **8**, eabl6496, doi:10.1126/sciadv.abl6496 (2022).
- 78 Paterson, R. S., Rybczynski, N., Kohno, N. & Maddin, H. C. A total evidence phylogenetic analysis of pinniped phylogeny and the possibility of parallel evolution within a monophyletic framework. *Frontiers in Ecology and Evolution* **7**, 457 (2020).
- 79 Springer, M. S. *et al.* Genomic and anatomical comparisons of skin support independent adaptation to life in water by cetaceans and hippos. *Curr Biol* **31**, 2124-2139 e2123, doi:10.1016/j.cub.2021.02.057 (2021).
- 80 Zhao, B., Bie, J., Wang, J., Marqueen, S. A. & Ghosh, S. Identification of a novel intracellular cholesteryl ester hydrolase (carboxylesterase 3) in human macrophages: compensatory increase in its expression after carboxylesterase 1 silencing. *Am J Physiol Cell Physiol* **303**, C427-435, doi:10.1152/ajpcell.00103.2012 (2012).
- 81 Sanghani, S. P. *et al.* Hydrolysis of irinotecan and its oxidative metabolites, 7-ethyl-10-[4-N-(5-aminopentanoic acid)-1-piperidino] carbonyloxycamptothecin and 7-ethyl-10-[4-(1-piperidino)-1-amino]-carbonyloxycamptothecin, by human carboxylesterases CES1A1, CES2, and a newly expressed carboxylesterase isoenzyme, CES3. *Drug Metab Dispos* **32**, 505-511, doi:10.1124/dmd.32.5.505 (2004).
- 82 Lian, J., Nelson, R. & Lehner, R. Carboxylesterases in lipid metabolism: from mouse to human. *Protein Cell* **9**, 178-195, doi:10.1007/s13238-017-0437-z (2018).

- 83 Cresswell, I., Janke, T. & Johnston, E. Australia state of the environment 2021: overview, independent report to the Australian Government Minister for the Environment, Commonwealth of Australia, Canberra. (2021).
- 84 Marsh, H. & Sobotzick, S. Dugong dugon (amended version of 2015 assessment). The IUCN Red List of Threatened Species 2019: e. T6909A160756767. en. Downloaded on 20 September 2021. doi: 10.2305/IUCN. UK. 2015-4. *RLTS. T6909A160756767. en* (2019).
- 85 Lin, M., Turvey, S. T., Liu, M., Ma, H. & Li, S. Lessons from extinctions of dugong populations. *Science* **378**, 148, doi:10.1126/science.ade9750 (2022).
- 86 Lin, M. *et al.* Functional extinction of dugongs in China. *R Soc Open Sci* **9**, 211994, doi:10.1098/rsos.211994 (2022).
- 87 Kayanne, H., Hara, T., Arai, N., Yamano, H. & Matsuda, H. Trajectory to local extinction of an isolated dugong population near Okinawa Island, Japan. *Sci Rep* **12**, 6151, doi:10.1038/s41598-022-09992-2 (2022).
- 88 Hamel, M. A., Marsh, H., Cleguer, C., Garrigue, C. & Oremus, M. Dugong dugon (New Caledonia subpopulation). The IUCN Red List of Threatened Species 2022: e. T218582754A218589361. en. Downloaded on 21 March 2023. (2022).
- 89 Allen, S., Marsh, H. & Hodgson, A. in *Proceedings of the Linnean Society of New South Wales*. 211-216.
- 90 Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493-496, doi:10.1038/nature10231 (2011).
- 91 Sharko, F. S. *et al.* Steller's sea cow genome suggests this species began going extinct before the arrival of Paleolithic humans. *Nat Commun* **12**, 2215, doi:10.1038/s41467-021-22567-5 (2021).
- 92 Voris, H. K. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of biogeography* **27**, 1153-1167 (2000).
- 93 Blair, D. *et al.* Pleistocene sea level fluctuations and the phylogeography of the dugong in Australian waters. *Marine Mammal Science* **30**, 104-121 (2014).
- 94 Tikel, D. *Using a genetic approach to optimise dugong (Dugong dugon) conservation management*, James Cook University, (1997).
- 95 McGowan, A. M. *et al.* Cryptic marine barriers to gene flow in a vulnerable coastal species, the dugong (Dugong dugon). *Marine Mammal Science* (2023).
- 96 Pickrell, J. K. *et al.* Signals of recent positive selection in a worldwide sample of human populations. *Genome Res* **19**, 826-837, doi:10.1101/gr.087577.108 (2009).
- 97 Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res* **20**, 393-402, doi:10.1101/gr.100545.109 (2010).
- 98 Escalante, L. E. & Gasch, A. P. The role of stress-activated RNA-protein granules in surviving adversity. *RNA* **27**, 753-762, doi:10.1261/rna.078738.121 (2021).
- 99 Zhang, K. *et al.* Stress Granule Assembly Disrupts Nucleocytoplasmic Transport. *Cell* **173**, 958-971 e917, doi:10.1016/j.cell.2018.03.025 (2018).
- 100 Guan, Q., Haroon, S., Bravo, D. G., Will, J. L. & Gasch, A. P. Cellular memory of acquired stress resistance in *Saccharomyces cerevisiae*. *Genetics* **192**, 495-505, doi:10.1534/genetics.112.143016 (2012).
- 101 Rondelli, C. M. *et al.* The ubiquitous mitochondrial protein unfoldase CLPX regulates erythroid heme synthesis by control of iron utilization and heme synthesis enzyme activation

- and turnover. *J Biol Chem* **297**, 100972, doi:10.1016/j.jbc.2021.100972 (2021).
- 102 Seo, J. H. *et al.* The Mitochondrial Unfoldase-Peptidase Complex ClpXP Controls Bioenergetics Stress and Metastasis. *PLoS Biol* **14**, e1002507, doi:10.1371/journal.pbio.1002507 (2016).
- 103 Chappell, J. Evidence for smoothly falling sea level relative to north Queensland, Australia, during the past 6,000 yr. *Nature* **302**, 406-408 (1983).
- 104 Lewis, S. E. *et al.* Rapid relative sea-level fall along north-eastern Australia between 1200 and 800 cal. yr BP: An appraisal of the oyster evidence. *Marine Geology* **370**, 20-30 (2015).
- 105 Rivera-Ingraham, G. A., Nommick, A., Blondeau-Bidet, E., Ladurner, P. & Lignot, J. H. Salinity stress from the perspective of the energy-redox axis: Lessons from a marine intertidal flatworm. *Redox Biol* **10**, 53-64, doi:10.1016/j.redox.2016.09.012 (2016).
- 106 Zhou, X. *et al.* Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. *Nat Commun* **9**, 1276, doi:10.1038/s41467-018-03722-x (2018).
- 107 Yuan, Y. *et al.* Genome Sequence of the Freshwater Yangtze Finless Porpoise. *Genes (Basel)* **9**, doi:10.3390/genes9040213 (2018).
- 108 Zhou, X. *et al.* Baiji genomes reveal low genetic variability and new insights into secondary aquatic adaptations. *Nat Commun* **4**, 2708, doi:10.1038/ncomms3708 (2013).
- 109 Robinson, J. A. *et al.* The critically endangered vaquita is not doomed to extinction by inbreeding depression. *Science* **376**, 635-639, doi:10.1126/science.abm1742 (2022).
- 110 Stanhope, M. J. *et al.* Genomes of endangered great hammerhead and shortfin mako sharks reveal historic population declines and high levels of inbreeding in great hammerhead. *iScience* **26**, 105815, doi:10.1016/j.isci.2022.105815 (2023).
- 111 Ceballos, F. C., Joshi, P. K., Clark, D. W., Ramsay, M. & Wilson, J. F. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet* **19**, 220-234, doi:10.1038/nrg.2017.109 (2018).
- 112 Gales, N., McCauley, R. D., Lanyon, J. & Holley, D. Change in abundance of dugongs in Shark Bay, Ningaloo and Exmouth Gulf, Western Australia: evidence for large-scale migration. *Wildlife Research* **31**, 283-290 (2004).
- 113 von Seth, J. *et al.* Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nat Commun* **12**, 2393, doi:10.1038/s41467-021-22386-8 (2021).
- 114 Hecker, N., Sharma, V. & Hiller, M. Transition to an Aquatic Habitat Permitted the Repeated Loss of the Pleiotropic KLK8 Gene in Mammals. *Genome Biol Evol* **9**, 3179-3188, doi:10.1093/gbe/evx239 (2017).
- 115 Liu, J. *et al.* Differential MC5R loss in whales and manatees reveals convergent evolution to the marine environment. *Dev Genes Evol* **232**, 81-87, doi:10.1007/s00427-022-00688-1 (2022).
- 116 Lopes-Marques, M. *et al.* Complete Inactivation of Sebum-Producing Genes Parallels the Loss of Sebaceous Glands in Cetacea. *Mol Biol Evol* **36**, 1270-1280, doi:10.1093/molbev/msz068 (2019).
- 117 Springer, M. S. & Gatesy, J. Evolution of the MC5R gene in placental mammals with evidence for its inactivation in multiple lineages that lack sebaceous glands. *Mol Phylogenet Evol* **120**, 364-374, doi:10.1016/j.ympev.2017.12.010 (2018).
- 118 Sun, X. *et al.* Comparative genomics analyses of alpha-keratins reveal insights into

- evolutionary adaptation of marine mammals. *Front Zool* **14**, 41, doi:10.1186/s12983-017-0225-x (2017).
- 119 Zhang, X. *et al.* Parallel Independent Losses of G-Type Lysozyme Genes in Hairless Aquatic Mammals. *Genome Biol Evol* **13**, doi:10.1093/gbe/evab201 (2021).
- 120 Albalat, R. & Canestro, C. Evolution by gene loss. *Nat Rev Genet* **17**, 379-391, doi:10.1038/nrg.2016.39 (2016).
- 121 Lanyon, J. M., Sneath, H. L. & Long, T. Three skin sampling methods for molecular characterisation of free-ranging dugong (*Dugong dugon*) populations. *Aquatic Mammals* **36**, 298 (2010).
- 122 Haklay, M. & Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive computing* **7**, 12-18 (2008).
- 123 Wang, O. *et al.* Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res* **29**, 798-808, doi:10.1101/gr.245126.118 (2019).
- 124 Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6, doi:10.1093/gigascience/gix120 (2018).
- 125 Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res* **27**, 757-767, doi:10.1101/gr.214874.116 (2017).
- 126 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18, doi:10.1186/2047-217X-1-18 (2012).
- 127 Pryszcz, L. P. & Gabaldon, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* **44**, e113, doi:10.1093/nar/gkw294 (2016).
- 128 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
- 129 Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92-95, doi:10.1126/science.aal3327 (2017).
- 130 Robinson, J. T. *et al.* Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst* **6**, 256-258 e251, doi:10.1016/j.cels.2018.01.001 (2018).
- 131 Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19 Suppl 2**, ii215-225, doi:10.1093/bioinformatics/btg1080 (2003).
- 132 Delgado, C. L., Waters, P. D., Gilbert, C., Robinson, T. J. & Graves, J. A. Physical mapping of the elephant X chromosome: conservation of gene order over 105 million years. *Chromosome Res* **17**, 917-926, doi:10.1007/s10577-009-9079-1 (2009).
- 133 McHale, M., Broderick, D., Ovenden, J. R. & Lanyon, J. M. A PCR assay for gender assignment in dugong (*Dugong dugon*) and West Indian manatee (*Trichechus manatus*). *Mol Ecol Resour* **8**, 669-670, doi:10.1111/j.1471-8286.2007.02041.x (2008).
- 134 Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6**, 11, doi:10.1186/s13100-015-0041-9 (2015).
- 135 Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **Chapter 4**, Unit 4 10, doi:10.1002/0471250953.bi0410s25 (2009).
- 136 Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. 2008-2015. Available at <http://www.repeatmasker.org>. (2010).

- 137 Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de
novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18,
doi:10.1186/1471-2105-9-18 (2008).
- 138 Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*
27, 573-580, doi:10.1093/nar/27.2.573 (1999).
- 139 Kent, W. J. BLAT--the BLAST-like alignment tool. *Genome Res* **12**, 656-664,
doi:10.1101/gr.229202 (2002).
- 140 Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res* **14**, 988-995,
doi:10.1101/gr.1865504 (2004).
- 141 Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids*
Res **34**, W435-439, doi:10.1093/nar/gkl200 (2006).
- 142 Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J Mol*
Biol **268**, 78-94, doi:10.1006/jmbi.1997.0951 (1997).
- 143 Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source
ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878-2879,
doi:10.1093/bioinformatics/bth315 (2004).
- 144 Roehr, J. T., Dieterich, C. & Reinert, K. Flexbar 3.0 - SIMD and multicore parallelization.
Bioinformatics **33**, 2941-2942, doi:10.1093/bioinformatics/btx330 (2017).
- 145 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-Flexible Barcode and Adapter
Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* **1**, 895-905,
doi:10.3390/biology1030895 (2012).
- 146 Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal
RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211-3217,
doi:10.1093/bioinformatics/bts611 (2012).
- 147 Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing
and web-based tools. *Nucleic Acids Res* **41**, D590-596, doi:10.1093/nar/gks1219 (2013).
- 148 Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment
and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**, 907-915,
doi:10.1038/s41587-019-0201-4 (2019).
- 149 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 150 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a
reference genome. *Nat Biotechnol* **29**, 644-652, doi:10.1038/nbt.1883 (2011).
- 151 Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity
platform for reference generation and analysis. *Nat Protoc* **8**, 1494-1512,
doi:10.1038/nprot.2013.084 (2013).
- 152 Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and
the Program to Assemble Spliced Alignments. *Genome Biol* **9**, R7,
doi:10.1186/gb-2008-9-1-r7 (2008).
- 153 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421,
doi:10.1186/1471-2105-10-421 (2009).
- 154 Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids*
Res **28**, 27-30 (2000).
- 155 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status,

- taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).
- 156 UniProt Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* **40**, D71-75, doi:10.1093/nar/gkr981 (2012).
- 157 O'Donovan, C. *et al.* High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief Bioinform* **3**, 275-284, doi:10.1093/bib/3.3.275 (2002).
- 158 Mitchell, A. L. *et al.* InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **47**, D351-D360, doi:10.1093/nar/gky1100 (2019).
- 159 Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238, doi:10.1186/s13059-019-1832-y (2019).
- 160 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).
- 161 Loytynoja, A. Phylogeny-Aware Alignment with PRANK and PAGAN. *Methods Mol Biol* **2231**, 17-37, doi:10.1007/978-1-0716-1036-7_2 (2021).
- 162 Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol Biol* **1079**, 155-170, doi:10.1007/978-1-62703-646-7_10 (2014).
- 163 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-577, doi:10.1080/10635150701472164 (2007).
- 164 Harris, R. S. *Improved pairwise alignment of genomic DNA*. (The Pennsylvania State University, 2007).
- 165 Blanchette, M. *et al.* Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res* **14**, 708-715, doi:10.1101/gr.1933104 (2004).
- 166 Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinform* **12**, 41-51, doi:10.1093/bib/bbq072 (2011).
- 167 Ayad, L. A. & Pissis, S. P. MARS: improving multiple circular sequence alignment using refined sequences. *BMC Genomics* **18**, 86, doi:10.1186/s12864-016-3477-5 (2017).
- 168 Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690, doi:10.1093/bioinformatics/btl446 (2006).
- 169 Khalturin, K. *et al.* Polyzoa is back: The effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci Adv* **8**, eabo4400, doi:10.1126/sciadv.abo4400 (2022).
- 170 Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153, doi:10.1186/s12859-018-2129-y (2018).
- 171 Edelman, N. B. *et al.* Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594-599, doi:10.1126/science.aaw2090 (2019).
- 172 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).
- 173 Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**, 1812-1819, doi:10.1093/molbev/msx116 (2017).
- 174 Mason, V. C. *et al.* Genomic analysis reveals hidden biodiversity within colugos, the sister

- group to primates. *Sci Adv* **2**, e1600633, doi:10.1126/sciadv.1600633 (2016).
- 175 Springer, M. S., Molloy, E. K., Sloan, D. B., Simmons, M. P. & Gatesy, J. ILS-Aware Analysis of Low-Homoplasmy Retroelement Insertions: Inference of Species Trees and Introgression Using Quartets. *J Hered* **111**, 147-168, doi:10.1093/jhered/esz076 (2020).
- 176 Feng, S. *et al.* Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* **185**, 1646-1660 e1618, doi:10.1016/j.cell.2022.03.034 (2022).
- 177 Doronina, L., Reising, O., Clawson, H., Ray, D. A. & Schmitz, J. True Homoplasmy of Retrotransposon Insertions in Primates. *Syst Biol* **68**, 482-493, doi:10.1093/sysbio/syy076 (2019).
- 178 Kuritzin, A., Kischka, T., Schmitz, J. & Churakov, G. Incomplete Lineage Sorting and Hybridization Statistics for Large-Scale Retroposon Insertion Data. *PLoS Comput Biol* **12**, e1004812, doi:10.1371/journal.pcbi.1004812 (2016).
- 179 De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271, doi:10.1093/bioinformatics/btl097 (2006).
- 180 Yang, Z., Wong, W. S. & Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**, 1107-1118, doi:10.1093/molbev/msi097 (2005).
- 181 Zheng, Z., Hua, R., Xu, G., Yang, H. & Shi, P. Gene losses may contribute to subterranean adaptations in naked mole-rat and blind mole-rat. *BMC Biol* **20**, 44, doi:10.1186/s12915-022-01243-0 (2022).
- 182 She, R., Chu, J. S., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res* **19**, 143-149, doi:10.1101/gr.082081.108 (2009).
- 183 Zhang, Z. D., Frankish, A., Hunt, T., Harrow, J. & Gerstein, M. Identification and analysis of unitary pseudogenes: historic and contemporary gene losses in humans and other primates. *Genome Biol* **11**, R26, doi:10.1186/gb-2010-11-3-r26 (2010).
- 184 Sharma, V. *et al.* A genomics approach reveals insights into the importance of gene losses for mammalian adaptations. *Nat Commun* **9**, 1215, doi:10.1038/s41467-018-03667-1 (2018).
- 185 Chou, H. H. *et al.* Inactivation of CMP-N-acetylneuraminic acid hydroxylase occurred prior to brain expansion during human evolution. *Proc Natl Acad Sci U S A* **99**, 11736-11741, doi:10.1073/pnas.182257399 (2002).
- 186 Sun, Y. B. FasParser2: a graphical platform for batch manipulation of tremendous amount of sequence data. *Bioinformatics* **34**, 2493-2495, doi:10.1093/bioinformatics/bty126 (2018).
- 187 Sun, Y. B. FasParser: a package for manipulating sequence data. *Zool Res* **38**, 110-112, doi:10.24272/j.issn.2095-8137.2017.017 (2017).
- 188 Bu, D. *et al.* KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res* **49**, W317-W325, doi:10.1093/nar/gkab447 (2021).
- 189 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature methods* **7**, 248-249 (2010).
- 190 Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, Unit7 20, doi:10.1002/0471142905.hg0720s76 (2013).
- 191 Sim, N.-L. *et al.* SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic acids research* **40**, W452-W457 (2012).

- 192 Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants
on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-1081,
doi:10.1038/nprot.2009.86 (2009).
- 193 de Flamingh, A., Coutu, A., Roca, A. L. & Malhi, R. S. Accurate Sex Identification of Ancient
Elephant and Other Animal Remains Using Low-Coverage DNA Shotgun Sequencing Data.
G3 (Bethesda) **10**, 1427-1432, doi:10.1534/g3.119.400833 (2020).
- 194 Mittnik, A., Wang, C. C., Svoboda, J. & Krause, J. A Molecular Approach to the Sexing of the
Triple Burial at the Upper Paleolithic Site of Dolni Vestonice. *PLoS One* **11**, e0163019,
doi:10.1371/journal.pone.0163019 (2016).
- 195 Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of
threads on general-purpose processors. *Bioinformatics* **35**, 421-432,
doi:10.1093/bioinformatics/bty648 (2019).
- 196 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**,
357-359, doi:10.1038/nmeth.1923 (2012).
- 197 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**,
doi:10.1093/gigascience/giab008 (2021).
- 198 Flamingh, A. d., Coutu, A., Roca, A. L. & Malhi, R. S. (GSA Journals, 2020).
- 199 Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and
exomes. *Bioinformatics* **34**, 867-868, doi:10.1093/bioinformatics/btx699 (2018).
- 200 Gottipati, S., Arbiza, L., Siepel, A., Clark, A. G. & Keinan, A. Analyses of X-linked and
autosomal genetic variation in population-scale whole genome sequencing. *Nat Genet* **43**,
741-743, doi:10.1038/ng.877 (2011).
- 201 Palkopoulou, E. *et al.* Complete genomes reveal signatures of demographic and genetic
declines in the woolly mammoth. *Curr Biol* **25**, 1395-1400, doi:10.1016/j.cub.2015.04.007
(2015).
- 202 Song, Y., Biernacka, J. M. & Winham, S. J. Testing and estimation of X-chromosome SNP
effects: Impact of model assumptions. *Genet Epidemiol* **45**, 577-592, doi:10.1002/gepi.22393
(2021).
- 203 Wang, Z., Sun, L. & Paterson, A. D. Major sex differences in allele frequencies for X
chromosomal variants in both the 1000 Genomes Project and gnomAD. *PLoS Genet* **18**,
e1010231, doi:10.1371/journal.pgen.1010231 (2022).
- 204 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform.
Bioinformatics **26**, 589-595 (2010).
- 205 Guindon, S., Delsuc, F., Dufayard, J.-F. & Gascuel, O. in *Bioinformatics for DNA Sequence
Analysis* (ed D Posada) 113-137 (Humana Press, 2009).
- 206 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
- 207 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).
- 208 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based
linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).
- 209 Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual
ancestry estimation. *BMC bioinformatics* **12**, 1-6 (2011).
- 210 Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis
version 7.0 for bigger datasets. *Molecular biology and evolution* **33**, 1870-1874 (2016).

211 Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new
developments. *Nucleic Acids Research* **47**, W256-W259 (2019).

212 Villanueva, R. A. M. & Chen, Z. J. (Taylor & Francis, 2019).

213 Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. detectRUNS: Detect runs of homozygosity
and runs of heterozygosity in diploid genomes. (2018).

214 Gao, F., Ming, C., Hu, W. & Li, H. New software for the fast estimation of population
recombination rates (FastEPRR) in the genomic era. *G3: Genes, Genomes, Genetics* **6**,
1563-1571 (2016).

215 Khan, A. *et al.* Genomic evidence for inbreeding depression and purging of deleterious genetic
variation in Indian tigers. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2023018118
(2021).

216 Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps.
Genome research **20**, 393-402 (2010).

217 Webb, A. *et al.* The pop-gen pipeline platform: a software platform for population genomic
analyses. *Molecular Biology and Evolution* **38**, 3478-3485 (2021).

218 Maclean, C. A., Chue Hong, N. P. & Prendergast, J. G. Hapbin: an efficient program for
performing haplotype-based scans for positive selection in large genomic datasets. *Molecular
biology and evolution* **32**, 3027-3029 (2015).

219 Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human
populations. *Nature* **449**, 913-918 (2007).

220 Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in
the absence of a molecular clock. *Bioinformatics* **19**, 301-302 (2003).

221 Li, H. & Durbin, R. Inference of human population history from individual whole-genome
sequences. *Nature* **475**, 493-496 (2011).

222 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

223 Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history
from hundreds of unphased whole genomes. *Nature genetics* **49**, 303-309 (2017).

Figure legends

Figure 1 | Overview of extant sirenian taxa. a, Features of extant sirenians. In addition to common sirenian features, the dugong (*Dugong dugon*) and manatees (genus *Trichechus*; here the West Indian manatee, *T. manatus*) show various morphological differences. By courtesy of Encyclopædia Britannica, Inc. **b,** Species distribution of sirenians. Species ranges derived from spatial data on marine mammals from the International Union for the Conservation of Nature Red List of Threatened Species, December 2022 [version 6.3]. <https://www.iucnredlist.org>; Downloaded on 12 January 2023. Note that the dugong was very recently (2022) declared extinct in Chinese and Japanese waters. **c,** Maximum likelihood (ML) phylogenetic tree from the coding sequences of 7,695 orthologs from eight afrotherian species, rooted with a xenarthran outgroup (the Linnaeus's two-toed sloth). 1,000 bootstraps were used, and all nodes have 100% support. The number in each node represents the divergence time among species a red circle indicates the fossil record used for calibration in the node. The gray bars on the nodes indicate the 95% credibility intervals of the estimated posterior distributions of the divergence times. Afrotherian orders are shown to the right of each representative species.

Figure 2 | Genes with unique evolutionary signals in sirenians. Putative adaptive or maladaptive gene changes are shown.

Figure 3 | Sirenian-specific changes in iodine uptake and organification genes. a, Schematic diagram showing iodine uptake by a dugong, from seagrass ingestion to fermentation in the hindgut and reduction of iodine to iodide (I^-), to iodide transport in the bloodstream to the thyroid. The thyroid is an endocrine gland consisting of follicles, structural units of epithelial cells enclosing a lumen (the colloid) harboring thyroglobulin (TG). TG is a precursor for thyroid hormones (indicated by T_3 and T_4) biogenesis and is also an iodide storage and carrier protein. **b,** Iodide is transported

from the bloodstream to thyroid follicular cells by NIS (encoded by *SLC5A5*) interacting with KCNQ1 and KCNE2. Genes with sirenian-specific amino acid substitutions in orange and genes rapidly evolving and positively selected in sirenians in blue and red, respectively. **c**, Alignment showing an amino acid substitution in transmembrane domains of sirenian *SLC5A5* (NIS). Sirenian-specific residues shown in red shading. Red boxes indicate residues shown to be important for NIS function³⁴⁻³⁷. **d**, Radioiodide uptake rates of human and sirenian *SLC5A5* (NIS) constructs in HEK293T cells. The y-axis shows the percentage uptake of Na¹²⁵I absorbed. The cDNA constructs human2dugong and dugong2human code for NIS proteins where the five sirenian-specific amino acids shown in (c) are replaced by their corresponding counterparts. Control is empty *pcDNA3.1* vector. Data are derived from three independent experiments and presented as mean \pm S.D. **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.01$, * $P < 0.05$, ANOVA with Tukey's test. **e**, Alignment showing amino acid substitution in the ISK_Channel domain of sirenian KCNE2 (also known as LQT6). Sirenian-specific residues shown in red shading.

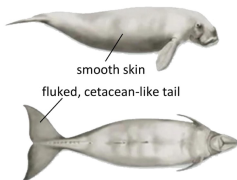
Figure 4 | Anatomy and gene changes of the sirenian integumentary system. a, Histology of dorsal dugong skin **b**, Schematic drawing of human skin and associated genes with changes in (as a model of mammals). Left, overview of major skin anatomical structures: epidermis, dermis, hypodermis, and ectodermal appendages (hair follicles, sebaceous glands, and sweat glands). Right, genes and their site of expression. Inactivated genes are in red (see *main text*). WAT denotes white adipocyte; BAT, brown adipocyte. **c**, Schematic representation of late cornified envelope (LCE) gene clusters in afrotherian genomes. Arrows indicate genes and their direction of transcription; different clusters are indicated by colored boxes. The slashes on each line indicate the ends of scaffolds in the assembled genomes.

Figure 5 | Multiple gene changes may underly the circadian activity of sirenians. a, Diagram of core circadian clock network. Genes with evolutionary changes in sirenians are bolded. D-box and E-box denote promoter elements found upstream of

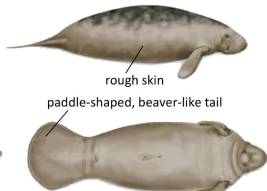
clock-controlled genes regulated by the transcription factors NFIL3 and CLOCK. **b**, Alignment showing an amino acid substitution in the CRY binding domain of sirenian PER2. A sirenian-specific proline at residue 1220 is shown in red shading. **c**, Left, A coimmunoprecipitation assay reveals the stronger binding ability of dugong PER2 with CRY1, as compared with human. Human C1220P denotes human PER2 mutated to match the dugong residue 1220. Right, quantification of densitometry analysis of blot using ImageJ. Shown as arbitrary density units (AU). ** $P < 0.01$, * $P < 0.05$, ANOVA with Tukey's test. **d**, Overview of inactivating substitutions in sirenian *KCNK18*. Colored cells indicate substitutions. **e**, Illustration of anatomical sites where *KCNK18* plays important roles and potential functional outcomes of its loss in sirenians.

Figure 6 | Demographic history of dugongs from 11 locations. **a**, Map indicating approximate sites of origin of resequenced dugongs. The red continuous line indicates a historical Torres Strait land border (115–7 kya); the dotted red line, a potential ecological barrier at or near the Whitsunday Islands on the Great Barrier Reef. Geographical locations are approximate. **b**, Principal component (PC) analysis of the dugong individuals. The gray highlighting shows clustering consistent with geographic locations south and north of the Whitsunday Islands also reported by ⁹⁵. **c**, Demographic history of dugong individuals inferred using PSMC. Okinawa, Exmouth Gulf, and Coogee Beach are single individuals, while three individuals from the remaining Queensland locations are plotted. **d**, Effective population size of populations inferred using SMC++. The most recent Torres Strait land border period is shaded in blue; the grey continuous line indicates the estimated emergence time of a potential ecological barrier at or near the Whitsunday Islands. **e**, A ~2 Mb selective sweep region on chromosome 18 identified by comparing population groups north and south of the Whitsunday Islands. Population differentiation (F_{st}), normalized XP-EHH, XP-CLR, and relative nucleotide diversity (π) are plotted. Note the positive XP-EHH value, indicating selection acting on the northern dugong population. **f**, Inbreeding in dugong individuals up to 463 generations ago (12,501 years, ROHs>100 kb); 46

generations ago (1,242 years; ROHs>1 Mb); nine generations ago (243 years ago; ROHs>5 Mb); and five generations ago (135 years; ROHs >10 Mb). The y-axis shows F_{ROH} , the sum length of all ROHs of a particular length cutoff divided by the autosomal genome length; the x-axis indicates various dugong populations.

a**Dugong**

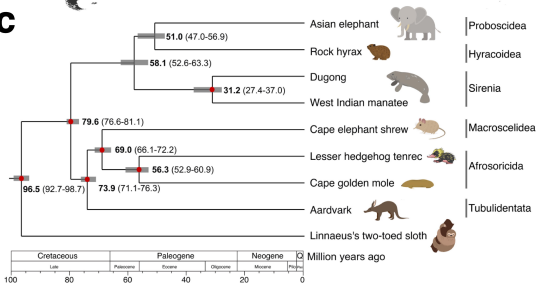
average length 2.7 m
average weight 250-300 kg

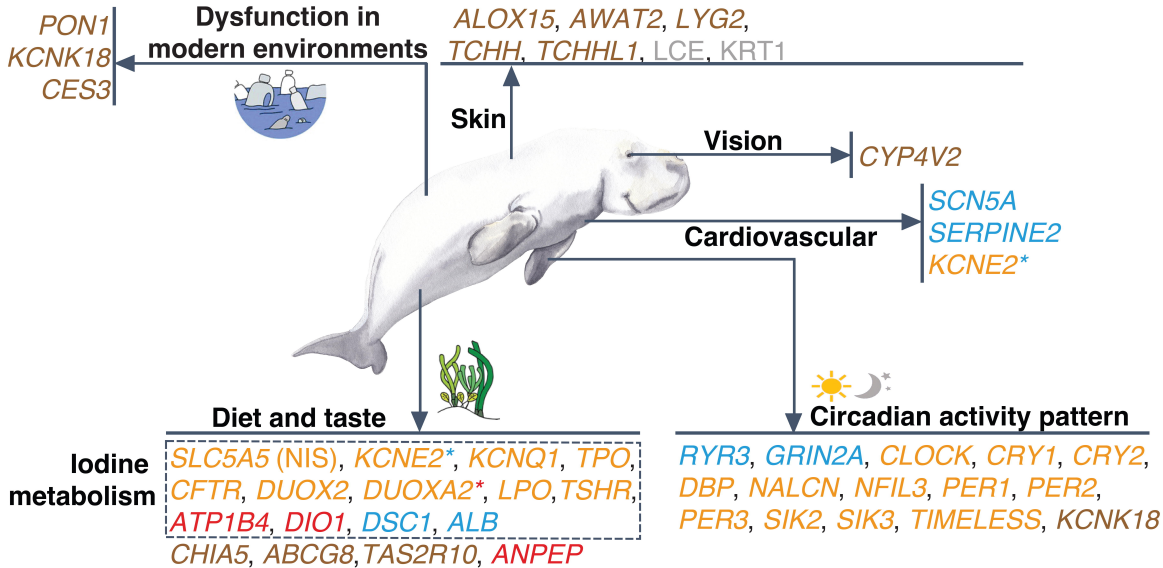
Manatee

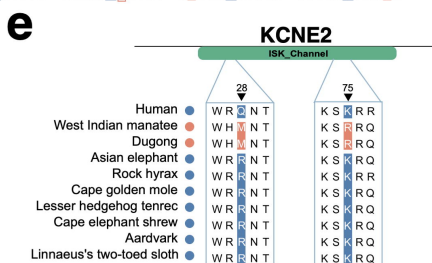
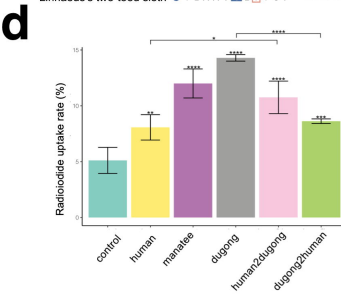
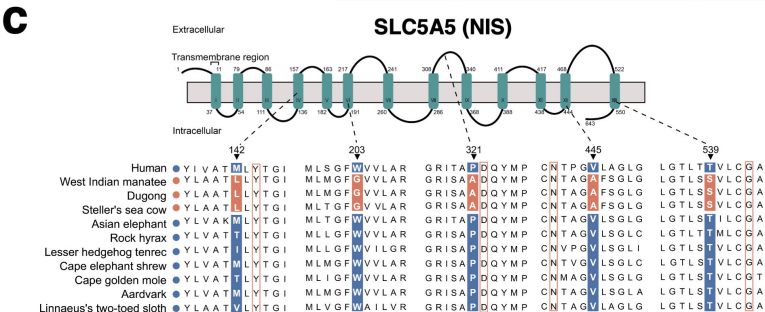
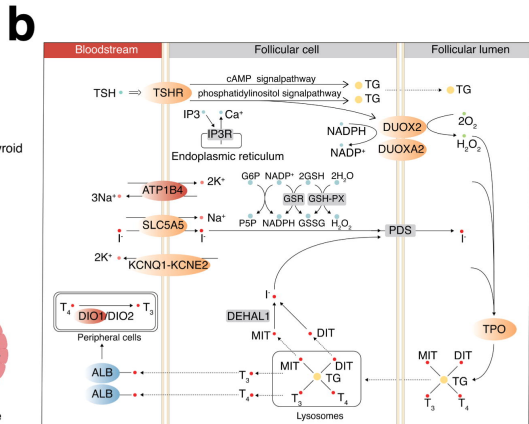
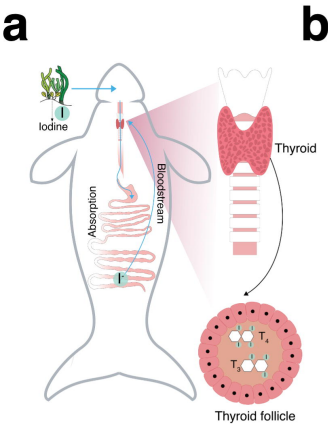
average length 3.0 m
average weight 500 kg

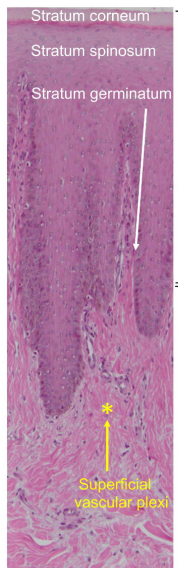
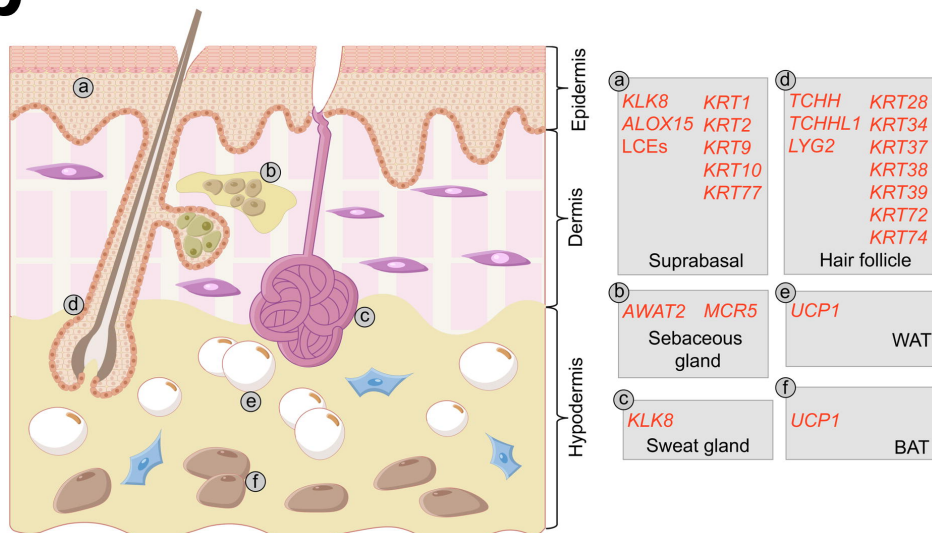
b

■ Dugong (*Dugon dugon*)
 ■ West Indian manatee (*Trichechus manatus*)
 ■ Amazonian manatee (*Trichechus inunguis*)
 ■ West African manatee (*Trichechus senegalensis*)

**c**





a**b****c**