

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

Received: 23 August 2023

Accepted: 12 June 2024

Published online: 02 July 2024

 Check for updates

Ran Tian^{1,12}, Yaolei Zhang^{2,3,4,12}, Hui Kang^{5,6,12}, Fan Zhang¹, Zhihong Jin¹, Jiahao Wang^{2,3}, Peijun Zhang⁵, Xuming Zhou^{7,8}, Janet M. Lanyon⁹, Helen L. Sneath⁹, Lucy Woolford¹⁰, Guangyi Fan^{2,3,4,11}✉, Songhai Li^{5,6}✉ & Inge Seim^{1,5}✉

Sirenians of the superorder Afrotheria were the first mammals to transition from land to water and are the only herbivorous marine mammals. Here, we generated a chromosome-level dugong (*Dugong dugon*) genome. A comparison of our assembly with other afrotherian genomes reveals possible molecular adaptations 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 confirm 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. Our analysis also uncovers 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. Genomes from nine Australian locations and the functionally extinct Okinawan population confirm and date a genetic break -10.7 thousand years ago on the Australian east coast and provide evidence of an associated ecotype, and highlight the need for whole-genome resequencing data from dugong populations worldwide for conservation and genetic management.

The terrestrial ancestors of the marine mammal groups Sirenia, Cetacea, and Pinnipedia independently transitioned from land to water¹. The ostensibly 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 pinnipeds². 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 elsewhere.

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 (Fig. S1a). 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

A full list of affiliations appears at the end of the paper. ✉ e-mail: fanguangyi@genomics.cn; lish@idsse.ac.cn; inge@seimlab.org

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 (Fig. S1b). Although dugongs are abundant along the tropical waters off Australia, their numbers elsewhere have dwindled in recent decades and some populations are now functionally extinct. The species is listed as *Vulnerable* globally by the International Union for Conservation of Nature (IUCN)^{6,7} and is threatened by habitat loss from human activities and climate change. In this study, we explored the adaptive evolution of sirenians and dugong diversity and demography. We highlight genes that may underlie sirenian adaptations, including convergent regression of integumentary system genes with cetaceans and aquatic herbivory, reconstruct demographic histories of dugong populations, and validate and characterize a genetic break in waters off the Australian East Coast.

Results

An annotated dugong genome for comparative and population genomic analyses

We used single-tube long fragment read (stLFR) and high-throughput chromosome conformation capture (Hi-C) sequencing to generate a 3.06 Gb 25-chromosome genome assembly of a female dugong with 18,663 annotated protein-coding genes and 1.51 Gb (49.2%) repetitive elements (Supplementary Note 1, Fig. S2, and Tables S1 and S2). The assembly and gene set BUSCO⁸ completeness scores (94.4% and 91.7%, respectively) were comparable to other afrotherian species and dugong assemblies that became available after the commencement of our assembly (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) (Fig. S1c). Our afrotherian data set, which included the West Indian manatee and the phylogenetically closest extant terrestrial species to sirenians (elephants and hyaxes) (Supplementary Note 2 and Fig. S3), was interrogated (see “Comparative genomics analysis strategy” in Methods, Table S5, and Supplementary Data 1–8) to illuminate features present since the sirenian crown ancestor (Fig. 1) that may underlie aquatic herbivory, sirenian circadian activity patterns, and typical marine mammal features such as modified cardiovascular (Supplementary Note 3), integumental (i.e., skin

and associated structures), and sensory (vision, smell, and taste) systems.

Nutrient uptake by fully aquatic herbivores

Sirenians are the only aquatic herbivorous mammals, and we observed gene losses consistent with a diet comprising few animal products (Supplementary Note 4 and Supplementary Data 8). Nearshore marine plants and aquatic plants from rivers and swamps are a rich source of iodine, a nutrient required to synthesize thyroid hormones (Fig. 2a) essential for systemic energy metabolism, thermoregulation, and the integrity of many tissues^{9–11}. Both inadequate and excess iodine uptake can result in thyroid dysfunction. Despite their high-iodine plant diet, the blood thyroid hormone levels of wild West Indian manatees are unremarkable compared to the tropical, carnivorous Amazon River dolphin (*Inia geoffrensis*)^{12,13}. The previous observations support the idea that genomic changes accompanied sirenian evolution from a terrestrial to an iodine-rich aquatic plant diet. In agreement, nearly all genes of the thyroid hormone pathway harbor sirenian-specific amino acid substitutions (Fig. 2b and Supplementary Data 5). 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; Fig. 2c), *KCNQ1* (potassium voltage-gated channel subfamily Q member 1), and *KCNE2* (potassium voltage-gated channel subfamily E regulatory subunit 2; Fig. 2d). We also identified positive selection of *ATP1B4* (ATPase Na⁺/K⁺ transporting family member β4), *DIO1* (iodothyronine deiodinase 1), and *DUOX2A*, and rapid evolution of *KCNE2* and the thyroid-hormone binding albumin (*ALB*) (Supplementary Data 1).

The transmembrane protein SLC5A5 (NIS) is the only known iodide transporter^{14,15}. After the 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^{16,17} (Fig. 2b), two proteins that also have multiple sirenian-specific amino acid substitutions. Mutations in human NIS result in congenital I⁻ deficiency disorders (IDDs)^{14,18}. We identified five sirenian-specific NIS mutations. Four are in transmembrane domains (TMDs), and one is in an extracellular loop (Fig. 2c). 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 4¹⁹; sirenian Ala-445 of TMD 12 flanks Asn-441, an extracellular region residue thought to mediate NIS structure via α-helix N-capping

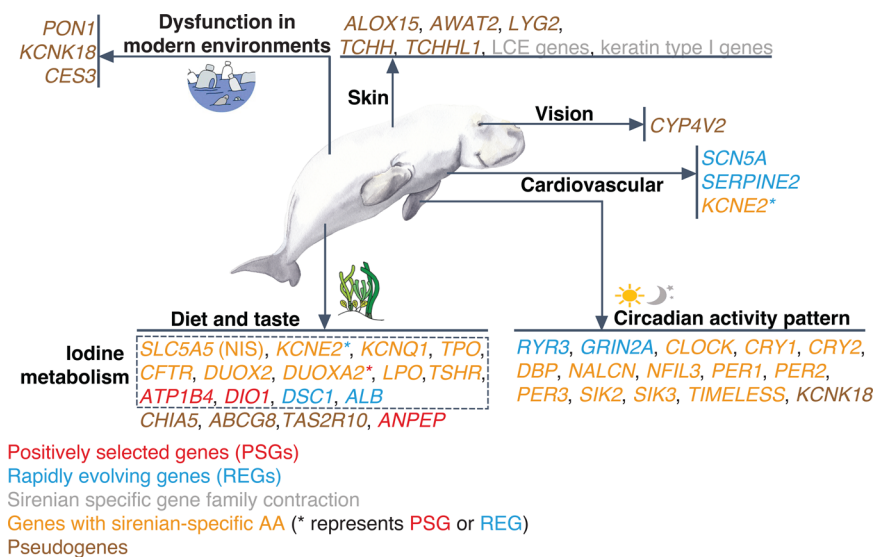
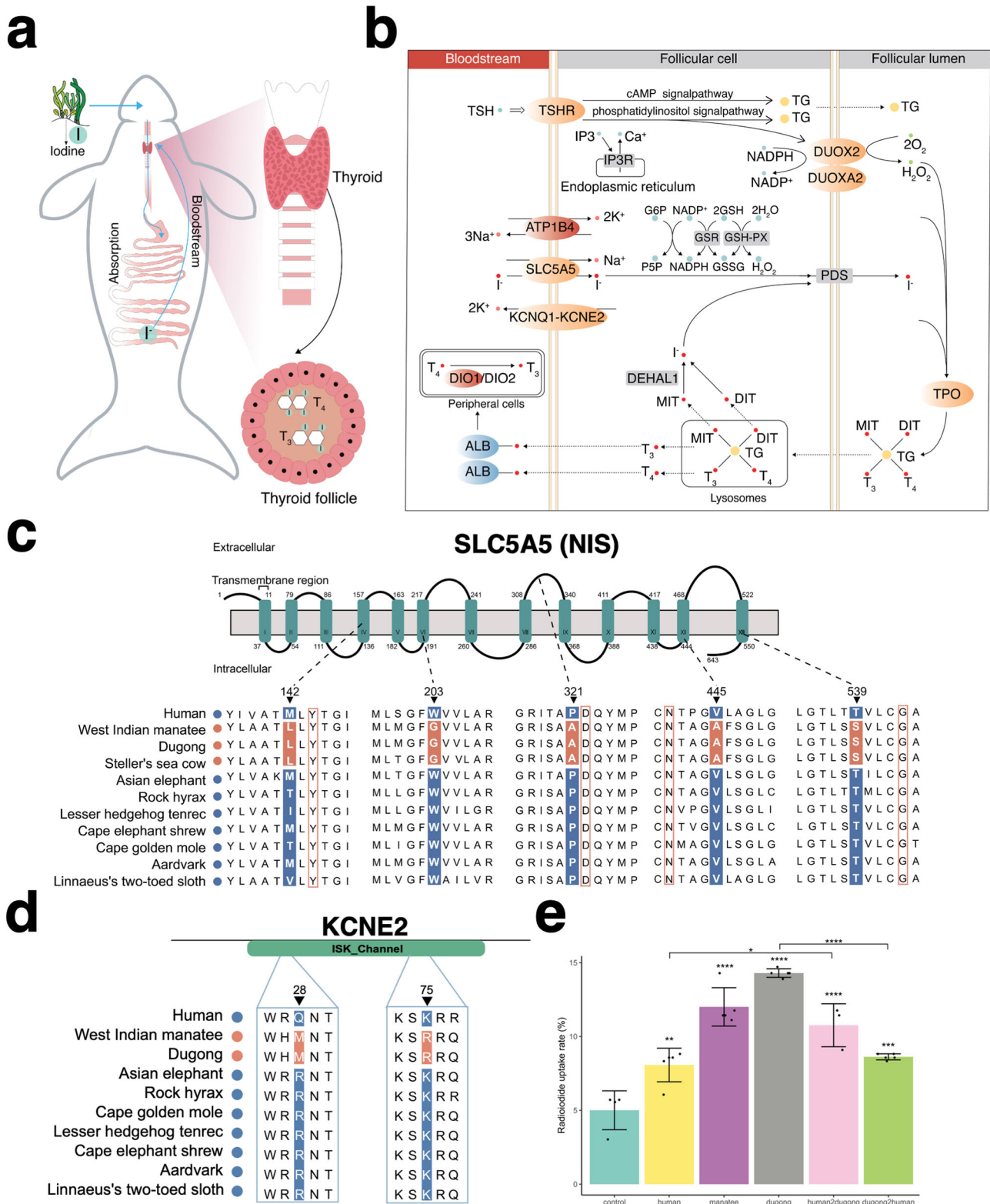


Fig. 1 | Genes with unique evolutionary signals in sirenians. Putative adaptive or maladaptive gene changes are shown. LCE denotes late cornified envelope. Image of dugong courtesy of Liudmila Kopecka/Shutterstock.com.



of TMD 12²⁰; and sirenian Ala-321 is next to Asp-322 in the extracellular loop of TMD 8 and 9²¹. Sirenian Ser-539 flanks Gly-543 in TMD 13, a residue required for NIS cell surface targeting²². 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 (Fig. 2e). 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 (Fig. 2e).

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 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

Fig. 2 | 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 T₃ and T₄) 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 function^{19–22}. **d** Alignment showing amino acid substitution

in the ISK_Channel domain of sirenian KCNE2. Sirenian-specific residues shown in red shading. **e** Radioiodide uptake rates of human and sirenian *SLC5A5* (NIS) constructs in HEK293T cells. The expression constructs human2dugong and dugong2human code for NIS proteins where the five sirenian-specific amino acids in (c) were replaced by their corresponding counterparts. The bar graph y-axis shows the percent of Na¹²⁵I uptake. Data derived from five independent experiments with single measurements (three for human2dugong), presented as mean ± S.D. relative to the control (empty *pcDNA3.1* vector), while dots represent individual data points per experiment (****P* < 0.0001, ***P* < 0.001, **P* < 0.01, *P* < 0.05, one-sided ANOVA with Tukey's multiple comparisons test). The exact *P* values compared to the control from left to right are 1.47 × 10⁻³, 0, 0, 1.50 × 10⁻⁶, and 1.90 × 10⁻⁴; human compared to human2dugong, 0.015; and dugong compared to dugong2human, 1.00 × 10⁻⁷. A *P* value of 0 indicates a number below 2.22 × 10⁻¹⁶.

sirenians²³. Consistent with previous reports^{23–26}, we observed a thin epidermis and a thick dermis in the dugong (Fig. 3a) and West Indian manatee (Fig. S4). We identified and validated, using dugong epidermis RNA-seq reads, the loss of multiple skin-associated genes (Fig. 3b and Supplementary Data 8). Notably, many of these genes are convergently lost in cetaceans, as revealed by manual literature searches and STRING²⁷ gene enrichment of the 15 shared sirenian pseudogenes identified in our analysis (Table S5 and Supplementary Note 5). In addition, our gene family screen revealed sirenian loss of late cornified envelope (LCE) gene family proteins expressed by the top layer of the epidermis (Fig. 3c). LCE gene numbers are also reduced in the afrotherian armadillo (*Orycteropus afer*; common ancestor ~80 Mya), suggesting a role in the evolution of its sparsely-haired skin²⁸.

Daily activity patterns

Most terrestrial animals rely on circadian rhythmicity, a molecular clock regulated by the sun's daily cycle, to regulate activity patterns^{29,30}. In contrast, many marine animals inhabiting coastal reef habitats and shallow waters rely heavily on the lunar (moon) cycle and its effect on water depth, food availability, and temperature²⁹. The pineal gland is absent or non-functional in sirenians, cetaceans, and some terrestrial mammals. These species have lost genes associated with the synthesis and reception of melatonin; a hormone mediating light stimulation of the circadian clock^{31–35}.

Sirenians do not show a “classical” diel (24-h) activity pattern but exhibit short episodes of sleep during respiratory pauses, 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^{5,36–38}. We identified sirenian-specific amino acid substitutions in most core circadian clock genes (Fig. 1, Fig. 4a, and Supplementary Data 5), 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)³⁹. A coimmunoprecipitation assay showed that dugong *PER2* bound CRY1 better than wildtype human *PER2* or human *PER2* with a sirenian substitution (I1220P) in the CRY-binding domain (Fig. 4b, c and Fig. S5). Thus, at least one of the 19 other sirenian-specific *PER2* residues may be required for the enhanced CRY-binding. Our analysis also revealed loss of *KCNK18* (also known as TRESK, TWIK-related spinal cord K⁺ channel) (Fig. 4d and Supplementary Data 8), a circadian clock-regulated ion channel. *KCNK18* is highly expressed in the SCN, and *Kcnk18*^{-/-} mice cannot use light to differentiate between day and night⁴⁰.

Gene loss and maladaptation in an altered environment

The loss of dozens of genes re-organized the skin of cetaceans⁴¹ and sirenians (see above) over millions of years, allowing their semi-aquatic

ancestors to become fully aquatic. Climate change and human activities can outpace natural selection. 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 long generation time⁴². Sirenians are long-lived (~70 years), with a generation time of ~20 years for the West Indian manatee and ~27 years for the dugong^{5,43}. Threats to sirenians (mainly 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 patterns^{44–46}. We identified three gene activation events that may be disadvantageous today: convergent loss of *PON1* and *CES3* in marine mammals (Supplementary Note 6 and Figs. S6 and S7) and sirenian-specific loss of *KCNK18*.

Shared and unique *KCNK18* inactivating mutations were observed in the dugong, Steller's sea cow, and West Indian manatee (Supplementary Data 8 and Fig. 4d). *KCNK18* loss was probably not inherently damaging (see Daily activity patterns section above) but may expose sirenians to natural and anthropogenic threats. Besides the hypothalamic SCN, skin sensory neurons express the gene, and *Kcnk18* knockout mice show elevated pain and avoidance behaviors when exposed to pyrethroid insecticides⁴⁷ or temperatures below 20 °C⁴⁸. Given its geographic range (Fig. S1c), the Florida manatee (a subspecies of the West Indian manatee) manifests cold stress syndrome (CSS), a condition resulting from prolonged exposure to water temperatures below 20 °C characterized by multiple physiological changes and comorbidities of unknown genetic cause^{5,49,50}. Although speculative, we hypothesize that loss of *KCNK18* decreases sirenian temperature tolerance (Fig. 4e) and that CSS is similar to semelparity in marsupials^{51,52} 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 5), 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 Florida manatees^{25,53}. 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 environment⁵⁴.

A dugong whole-genome resequencing data set

We next considered the population genomics of dugongs from ten locations (Fig. 5a). To this end, we generated short-read whole-genome resequencing data from seven locations (skin biopsies of 99 individuals) spanning 2000 km of the Australian east coast from Torres Strait to Moreton Bay, Queensland (Supplementary Data 9). We obtained 3.46 Tb of data, with an average sequencing depth of 11.41×, and identified 71.25 million high-quality SNPs (average SNP density 24.61 SNPs/kb). Publicly available resequencing data (one individual

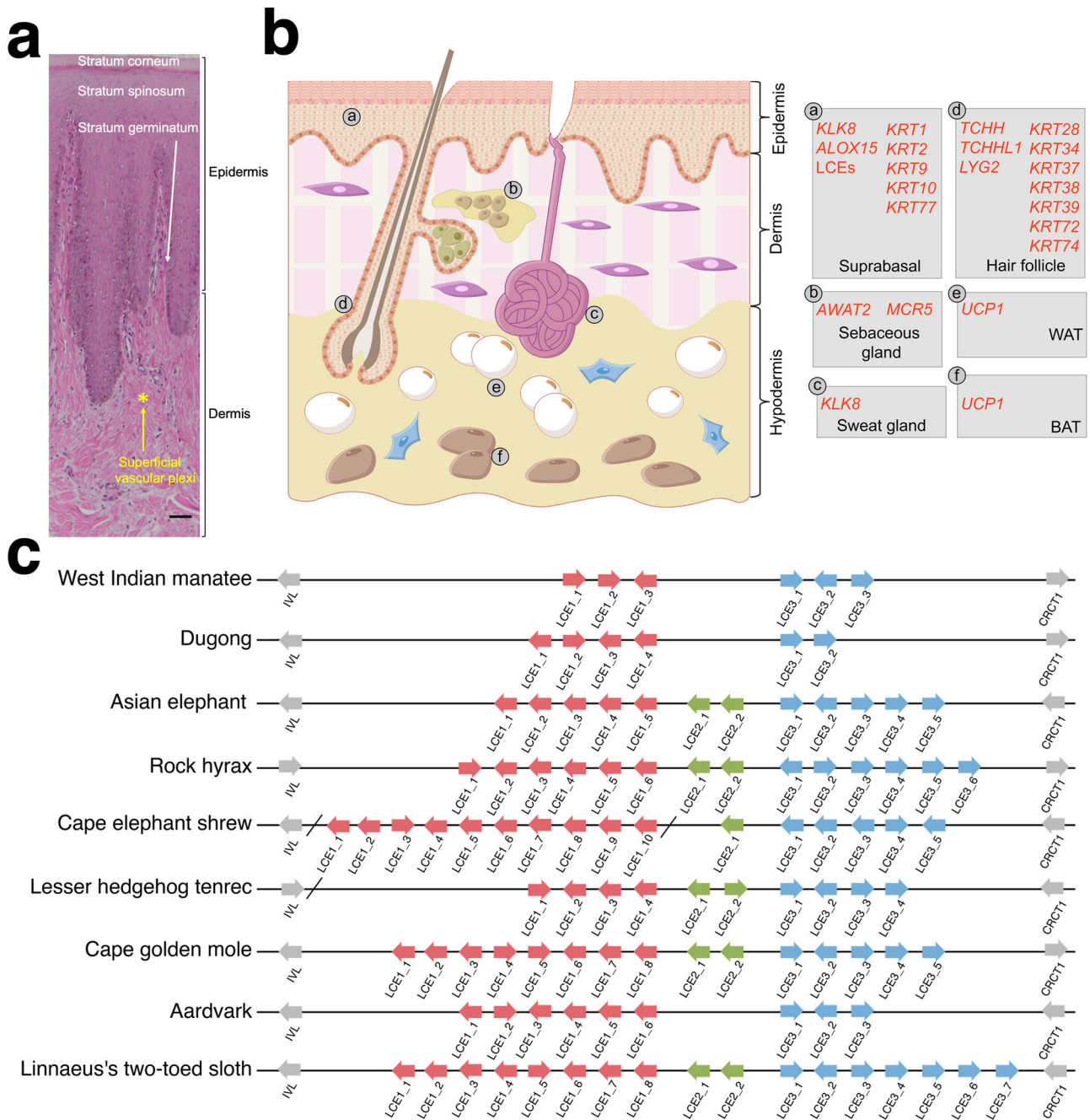


Fig. 3 | Anatomy and gene changes of the sirenian integumentary system. **a** Representative histology of hematoxylin and eosin (H&E)-stained dorsal dugong skin. The scale bar represents 50 μm . **b** Schematic drawing of H&E-stained human skin (as a model of mammals) and associated genes with changes in sirenians. 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 adipose tissue cell; BAT brown adipose tissue cell. **c** Schematic representation of late cornified envelope (LCE) gene clusters in afrotherian genomes. Arrows indicate genes and their direction of transcription; slashes indicate separate scaffolds; and different clusters are indicated by colored boxes.

each) was also obtained from two other Australian locations, Coogee Beach (New South Wales)⁵⁴ and Exmouth Gulf (Western Australia), and from waters off Okinawa (Japan). The average dugong genome-wide nucleotide diversity (π) and heterozygosity were 8.79×10^{-4} and 8.80×10^{-4} , respectively, higher than those of the killer whale (4.27×10^{-4} and 3.54×10^{-4})⁵⁵, northern elephant seal (2.04×10^{-4} and 1.78×10^{-4})⁵⁶, and Indo-Pacific humpback dolphin (1.55×10^{-4} and 1.79×10^{-4})⁵⁷. The average heterozygosity of Moreton Bay individuals ($n=32$) mirrored an estimate from a single individual from this locality⁵⁸ (1.40×10^{-3} vs. 1.60×10^{-3}).

Population structure and differentiation

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) (Fig. 5b). The Coogee Beach individual clustered with Moreton Bay and Hervey Bay individuals. Because this individual stranded ~750 km from the accepted eastern Australian range during the summer (November), we propose it represents one of the few instances⁵⁹ of seasonal long-distance ranging from a population in close geographic proximity to Moreton Bay.

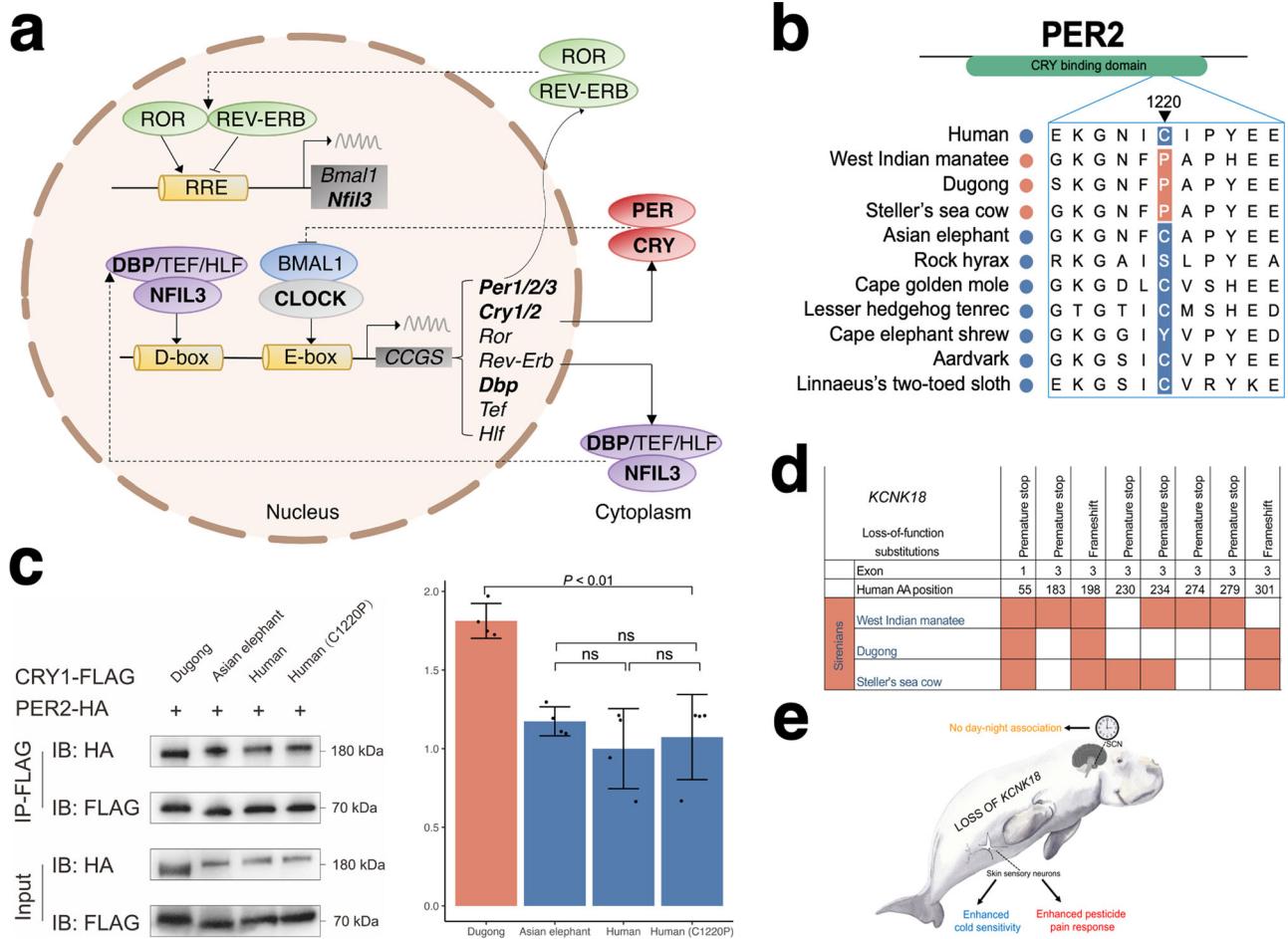


Fig. 4 | Multiple gene changes may underlie the circadian activity of sirenians.

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 representative co-immunoprecipitation assay reveals a stronger binding ability of dugong PER2 with its CRY1 compared to Asian elephant and human forms. Human C1220P denotes human PER2 mutated to match dugong residue 1220. Numbers on the right side of the gels represent molecular weight marker locations (values in kDa). Full images of western blots in Fig. S5. Right, bar graph showing ImageJ densitometry of western blot from four independent experiments with

single measurements, presented as mean arbitrary density units \pm S.D. relative to human, while dots represent individual data points per experiment. One-sided ANOVA with Tukey's multiple comparisons test (ns denotes not significant, a P value > 0.05). The exact P values were dugong-Asian elephant, 3.31×10^{-3} ; dugong-human, 4.48×10^{-4} ; dugong-human (C1220P), 1.02×10^{-3} ; Asian elephant-human, 0.62; Asian elephant-human (C1220P), 0.89; human-human (C1220P), 0.95. **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. SCN denotes the hypothalamic suprachiasmatic nucleus. Image of dugong courtesy of Liudmila Kopecka/ Shutterstock.com.

Interrogation of our 99-individual Queensland data set showed pronounced genetic structure into a northern and a southern group since -10.7 kya (95% CI: 9.1–12.2 kya), agreeing with a recently reported but undated potential ecological barrier at or near the Whitsunday Islands on the Great Barrier Reef⁶⁰ (Fig. 5a, c and Fig. S8). Summary statistics echoed the structure, revealing distinct genetic diversity (Fig. S9) and heterogeneity. The average pairwise fixation index (F_{ST}) values between the three northern and the four southern locations were around 0.1, indicating moderate differentiation (heterogeneity) between them. In contrast, there was no apparent within-group differentiation in the southern group (average F_{ST} 0.023) (Fig. 5d). A TreeMix consensus tree (Fig. 5e) was largely concordant with population clustering inferred by PCA (Fig. 5b), STRUCTURE (Fig. 5c), and a Neighbor-Joining tree (Fig. S8). While the TreeMix topology (Fig. 5e) and Patterson's D -statistics (ABBA-BABA-test) (Fig. S10 and Supplementary Data 10) showed admixture between the northern and southern populations, the low weights of the migration events inferred by TreeMix (see ref. 61) may reflect gene flow before the emergence of

the north-south barrier -10.7 kya. TreeMix and D -statistics showed no gene flow between Airlie Beach and the six other locations, indicating that dugongs at this location are genetically isolated.

Four complementary methods (XP-EHH⁶², XP-CLR⁶³, π , and F_{ST}) were used to detect regions with putative selective sweeps, SNPs under selection. We identified a two-megabase region (24.48–26.50 Mb) on chromosome 18 under positive selection in the northern dugong group (Fig. 5f). This region contains five annotated protein-coding genes (Table S6). They comprise three immunoglobulin genes, the nuclear pore transporter *NUP42* (also known as *NUPL2*), and ClpX protease (*CLPX*). Among these, only *CLPX* SNPs cause amino acid residue changes (Ile197Thr) unique to the northern dugong group and have an ortholog in afrotherians and other mammals. The threonine at residue 197 (predicted as “benign” by PolyPhen-2 and “tolerated” by SIFT; Table S6) was not found in other sirenians and the 120 mammalian species in OrthoMaM (Fig. S11). CLPX is a mitochondrial protein required to synthesize oxygen-binding hemoglobin that buffers oxidative stress^{64,65}. While it is difficult to conclude the driving force

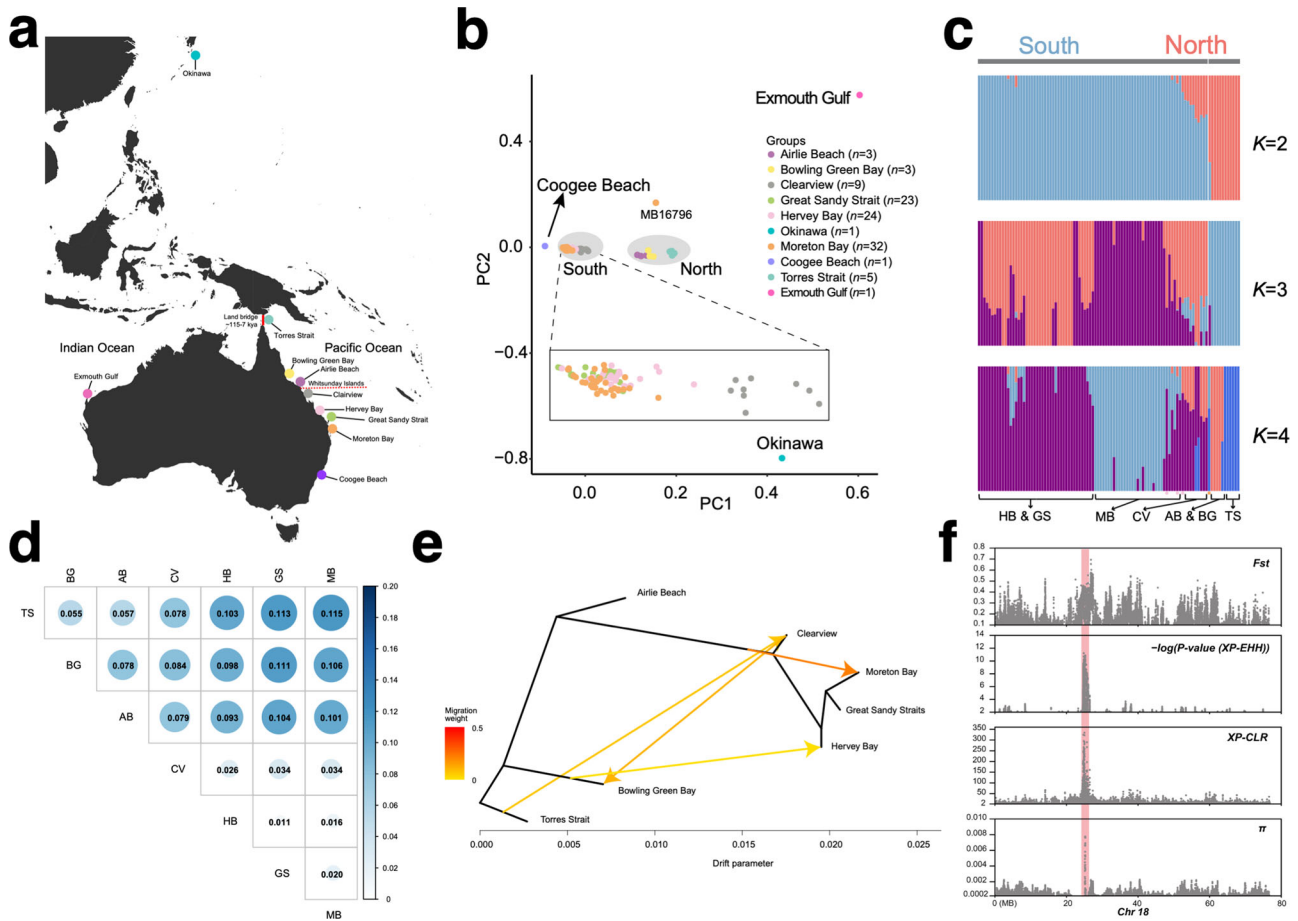


Fig. 5 | Evidence of population structure by dugongs on the Australian east coast. **a** Map indicating approximate sites of origin of re-sequenced dugongs. The red continuous line indicates a historical Torres Strait land border (115–7 kya); the dotted red line, a potential ecological barrier at or near the Whitsunday Islands on the Great Barrier Reef. Geographical locations are approximate. **b** Principal component (PC) analysis of the dugong individuals in (a). The gray highlighting shows clustering consistent with geographic locations south and north of the Whitsunday Islands also reported by ref. 60. **c** structure (ADMIXTURE) plots demonstrate substantial structure, supporting a break between population north and south of the Whitsunday Islands. Genetic structure inferred by varying the ancestry components (*K*) shown. **d** Pairwise population differentiation index (F_{ST}) differences between northern (TS, BG, and AB) and southern (CV, HB, GS, and MB)

population groups. Calculated in non-overlapping 10-kb sliding windows with 2-kb step sizes. **e** Gene flow among populations detected by TreeMix. A Maximum likelihood tree was inferred with four migration events. Terminal nodes are labeled by locality (see a). Arrows represent gene flow, and heat map colors reflect the intensity of gene flow, from low (yellow) to high (red). **f** 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. The $-\log(P\text{ value}(XP-EHH))$ is the \log_{10} of a two-sided *P* value for testing the null hypothesis that no selection occurred, with a value ≥ 5 equivalent to a *P* value of 0.00001. Note the positive values, indicating selection acting on the northern dugong population.

behind the selective sweep of *CLPX* and any functional effect of its amino acid substitution in northern Queensland dugongs, sirenians are vulnerable to climate change and regional environmental conditions that dramatically alter their nearshore habitats (particularly seagrass growth)⁴⁵. Northern Queensland has seen distinct climate change events after the formation of the ostensible ecological barrier at the Whitsunday Islands ~10.7 kya. The sea level on the north-east Queensland coast has continuously decreased since about 6000 years ago^{66,67}. Climate change in the region has also seen increasing frequency and intensity of weather events affecting seagrass habitats—including high seasonal summer rainfall and coastal runoffs, and cyclones. Coastal bathymetry is also more variable in northern Queensland, with seagrasses found from inshore shallow estuaries through to reef flats and deeper subtidal inter-reefal areas⁶⁸.

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 IUCN^{6,7}. The number of dugongs in other locations is

orders of magnitude lower. The dugong recently became functionally extinct in Chinese^{69,70} and Japanese waters⁷¹ and is at risk elsewhere in Asia, Oceania, and eastern Africa⁷². Pairwise sequential Markovian coalescent (PSMC)⁷³ analysis of dugong autosomes was used to track changes in effective population size (N_e ; the number of individuals that will contribute to the next population) during the Pleistocene (about 2.6 Mya to 20 kya) (Fig. 6a and Fig. S12). All examined dugongs showed a N_e decline in the mid-Pleistocene until ~400–500 kya that was also observed in a population of cold-resistant Steller’s sea cow off the Arctic Commander Islands^{54,74}, suggesting that lower seawater temperatures and sirenian cold stress syndrome did not drive this dugong population decline. 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 broadly similar environmental conditions across the Australian continent (Fig. 6a). Our PSMC curve of Moreton Bay individuals (Fig. 6a

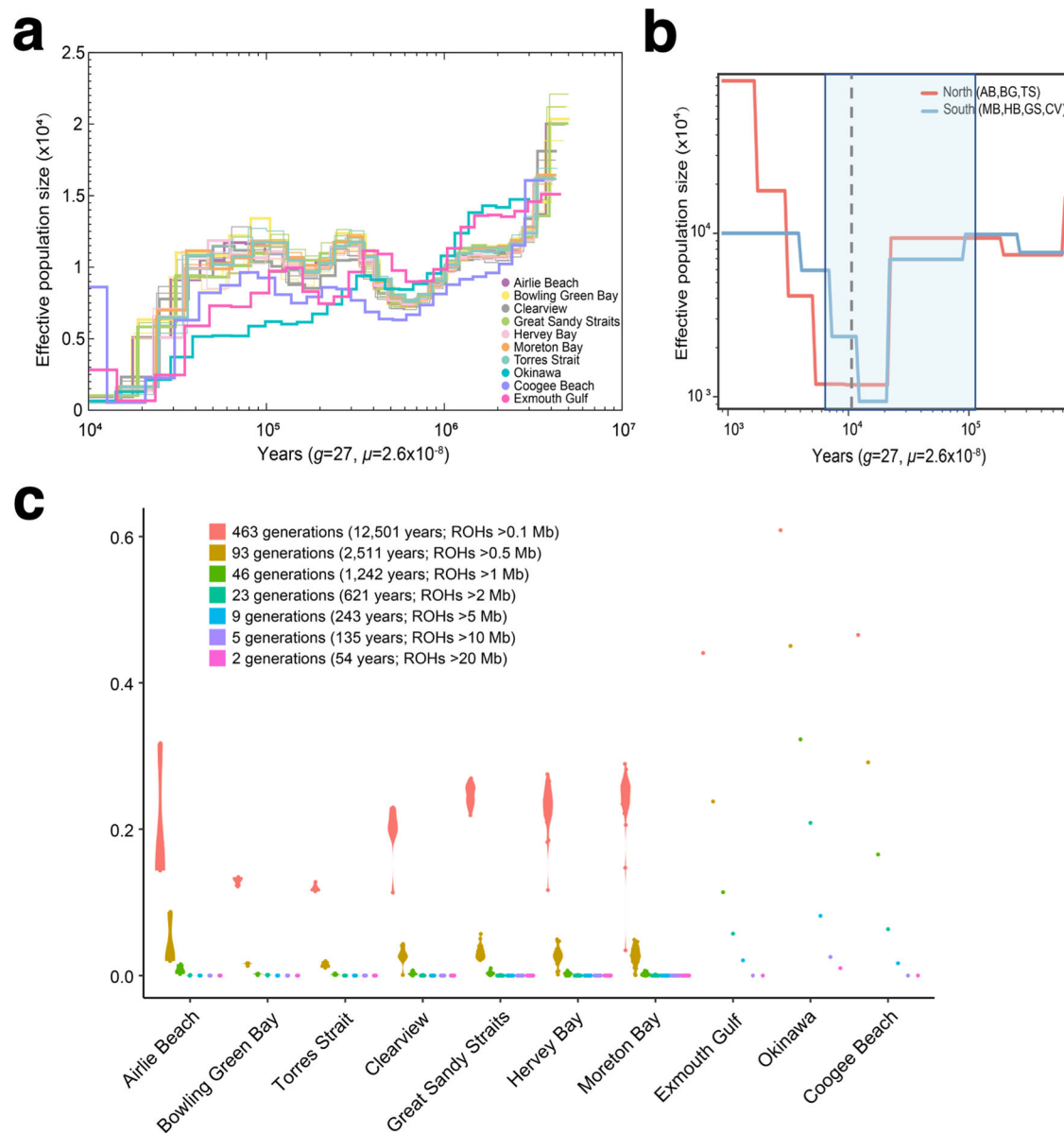


Fig. 6 | Demographic history of dugongs. **a** Demographic history of dugong individuals from 11 locations inferred using PSMC. Okinawa, Exmouth Gulf, and Coogee Beach are single individuals, while three individuals from the remaining Queensland locations are plotted. **b** Effective population size of populations from the Australian east coast (Queensland) inferred using SMC++. The most recent Torres Strait land border period is shaded in blue; the gray continuous line

indicates the estimated emergence time of a potential ecological barrier at or near the Whitsunday Islands. **c** Inbreeding in dugong individuals from 463 to two generations ago. The *y*-axis shows F_{ROH} , the sum length of all ROHs (runs of homozygosity) of a particular length cutoff divided by the autosomal genome length; the *x*-axis indicates various dugong populations.

and Fig. S12) was similar to a recent study⁵⁸ that examined a single individual from this location, but the effective population size was smaller in our dataset (e.g., ~600,000 vs. ~12,000 individuals about 100,000 years ago). We speculate that the much higher N_e in the recent study stems from different mutation rate parameters (6.25×10^{-9} vs. 2.60×10^{-8} per site per generation in our study) or failure to remove chromosome X before PSMC analysis—a step that can influence effective population size estimates (see refs. 73,75). In contrast to dugongs from Australian waters, the N_e of the individual from Japanese waters declined continuously over the last 400,000 years. The effective population size differences of dugongs on the Queensland coast in the recent past (i.e., <20 kya) reconstructed using SMC++ (Fig. 6b) were consistent with the distinct genetic diversity and heterogeneity associated with their genetic break ~10.7 kya. We speculate that the northern dugong populations are more “genetically healthy”

because they resumed panmixia with populations from the Indian Ocean once sea levels rose and covered a Torres Strait land bridge present from ~115–7 kya (see Fig. 5a).

Runs of homozygosity (ROHs) in a genome reflect the level of inbreeding and can provide conservation management guidance^{76,77}. 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 7 and Fig. S9d–f), with no evidence of inbreeding in recent times (since 1242 years ago) (Fig. 6c and Supplementary Data 11). The individual from Exmouth Gulf ($F_{ROH>1Mb} = 0.021$, 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) that appears to have resulted

in a large-scale emigration from Exmouth Gulf that reduced the dugong population from 1000 to less than 200 individuals in five years⁷⁸. 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 Okinawan individual aligned with its genome diversity estimates. It had one magnitude of order lower genome-wide heterozygosity (5.65×10^{-4}) compared with Australian dugongs ($\sim 1 \times 10^{-3}$). One-third of its genome was in ROH segments above one megabase (389 ROHs spanning 934.4 Mb), with evidence of inbreeding as recently as 135 years ago ($F_{\text{ROH}>10\text{Mb}} = 0.025$, four ROHs spanning 73.6 Mb) to 54 years ago ($F_{\text{ROH}>20\text{Mb}} = 0.010$, one ROH spanning 29.2 Mb). This pattern is consistent with an ancient population bottleneck and subsequent extensive inbreeding until recent times, mirroring a *Critically Endangered* Sumatran rhinoceros (*Dicerorhinus sumatrensis*) population on the Malay Peninsula⁷⁹. 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 2019⁷¹.

Discussion

Using comparative genomic approaches, we here report genetic changes that may underlie sirenian features. Because the ancestors of extant manatees and the dugong diverged at crown Sirenia ~ 30 Mya, these changes were likely critical for their transition to an aquatic habitat. The lack of a pineal gland and their genetic background support that the circadian clock (i.e., sleep-wake cycle) of sirenians has been recalibrated (e.g., see ref. 38), likely to facilitate an activity pattern in a more light-limited, fully aquatic environment heavily reliant on lunar tidal currents and water temperature fluctuations. We appreciate, however, that the function of sirenian-specific changes should be further assessed in vivo—as illustrated by a recent follow-up study on amino acid substitutions of panda DUOX2 in CRISPR-Cas9-edited mice⁸⁰. This thyroid hormone synthesis-associated protein also has unique residues in sirenians. Many of the same genes of the integumentary system, particularly those expressed in the outer layers of the skin^{32,81–87}, were lost by both sirenians and cetaceans. Our data support the idea that convergent gene loss occurs in species with similar ecological pressures⁸⁸ (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*)⁸⁹. Sirenian *KCNK18* loss is possibly related to their shift in activity patterns and may render them susceptible to water temperatures affected by climate change and human activities. Its role in sirenian cold stress syndrome (CSS) should be investigated.

Our population genomics analysis offers insights into dugong diversity and demography. We conclude that while dugongs on the Australian east and west coasts have a genetic 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 and an ROH pattern consistent with extensive inbreeding for millennia and, likely, climatic ages. 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 Japanese and Chinese waters and is a considerable risk factor elsewhere^{46,69–71}. Future studies should interrogate whole-genome resequencing data from extant and extinct dugong populations worldwide—including modern, historical, and ancient samples. Such efforts promise to shed further light on dugong evolution and inform their conservation. We also confirm⁶⁰ and date a north-south genetic

break that emerged approximately 10.7 thousand years ago on the Australian east coast and identify a two-megabase genetic sweep region that may be associated with historical and recurrent environmental differences between the north and south coast and formation of an ecologically distinct population group (i.e., ecotype⁹⁰). Our data set allows further study of genetic structure related to geographical regions and environmental variables.

In conclusion, our study reveals insights into sirenian biology and the transition of terrestrial mammals to an aquatic lifestyle and provides a basis for future genomic explorations.

Methods

Sample collection and research ethics

Dugong 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 WISPO7255110 and WISPI14654414, 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. All applicable institutional and/or national guidelines for the care and use of animals were followed.

Liver tissue (sample D201106) for reference genome sequencing was obtained by Queensland Parks and Wildlife Service 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. The fetus was transported frozen to The University of Queensland and dissected by J.M.L. This fetal liver sample and a skin sample from a dugong (D110419) sampled after an indigenous subsistence hunt (see below for sampling details) 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, $n = 5$; prefixed TS elsewhere in the manuscript) (Supplementary Data 9 and Fig. 5a). Briefly, skin was collected from the dorsum of wild free-swimming dugongs at each location, except Torres Strait, using a handheld scraper device⁹¹. 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% dimethyl sulfoxide (DSMO) and frozen at -20 °C until further processing for resequencing (see ref. 60). Skin samples for histology were sub-sampled from the Torres Strait specimens and stored in 10% neutral buffered formalin until sectioned and stained.

Sampling and species distribution maps were generated using the R package “OpenStreetMap”⁹².

Genome sequencing. High-molecular-weight DNA was extracted from a fetal liver sample (D201106) using a MagAttract HMW DNA Kit (QIAGEN). 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) libraries⁹³ were sequenced on an MGISEQ-2000 sequencer. A total of ~ 358 Gb ($\sim 100\times$) stLFR clean reads were generated after removing low-quality reads, PCR duplicates, and adaptors using SOAPnuke (v1.5)⁹⁴. Hi-C libraries (Lieberman-Aiden et al.⁹⁵) 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), extracted using an RNeasy Mini Kit (QIAGEN), was sequenced on the BGISEQ-500 platform to generate 86.7 and 96.3 Gb of 150-bp paired-end read RNA-seq data, respectively.

Genome assembly. The pipeline stLFRdenovo [<https://github.com/BGI-biotools/stLFRdenovo>], which is based on Supernova v2.11⁹⁶ and customized for stLFR data, was used to generate a de novo genome assembly. GapCloser v1.12⁹⁷ and clean stLFR reads (with the barcode removed with the stLFR_barcode_split tool [https://github.com/BGI-Qingdao/stLFR_barcode_split]) were used to fill gaps. Redundants v0.14a⁹⁸ was used to remove heterozygous contigs. Clean paired-end Hi-C reads validated by HiC-Pro v3.2.0_devel⁹⁹ were used to construct chromosome clusters with the 3D de novo assembly (3D DNA) pipeline v170123¹⁰⁰. The assembly was further improved by interactive correction using Juicebox (v1.11.08)¹⁰¹.

Assembly quality was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs) BUSCO v5.4.3⁸, employing the gene predictor AUGUSTUS (v3.2.1)¹⁰² 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 BLAST¹⁰³ interrogation of African elephant¹⁰⁴ and published dugong¹⁰⁵ 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. Homology-based transposable elements (TE) annotations were obtained by interrogating a genome assembly with known repeats in the Repbase database v16.02¹⁰⁶ using RepeatMasker v4.0.5 (DNA-level)¹⁰⁷ and RepeatProteinMask (protein-level; implemented in RepeatMasker). De novo TE predictions were obtained using RepeatModeler v1.1.0.4¹⁰⁸ and LTRharvest v1.5.8¹⁰⁹ to generate database for a RepeatMasker run. Tandem Repeat Finder (v4.07)¹¹⁰ 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, protein sequences from five mammals were downloaded from NCBI: African bush elephant (*Loxodonta africana*) assembly LoxAfr3.0 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000785645.1]; Cape elephant shrew (*Elephantulus edwardii*) assembly EleEdw1.0 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000299155.1]; aardvark (*Orycteropus afer*) assembly OryAfe1.0 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000298275.1]; West Indian manatee (*Trichechus manatus latirostris*) assembly TriManLat1.0 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000243295.1], and human (*Homo sapiens*) assembly GRCh38.p12 [https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.38]. These protein sequences were aligned to the repeat-masked genome using BLAT v0.36¹¹¹. GeneWise v2.4.1¹¹² was employed to generate gene structures based on the alignments of proteins to a genome assembly. De novo gene prediction was performed using AUGUSTUS v3.2.3¹¹³, GENSCAN v1.0¹¹⁴, and GlimmerHMM v3.0.1¹¹⁵ 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^{116,117} 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 v2.1b¹¹⁸ against the SILVA v119 ribosomal database¹¹⁹. Transcriptome data (clean reads) were mapped to the assembled genome using HISAT2 v2.1.0¹²⁰ and SAMtools v1.9¹²¹, and coding regions were predicted using TransDecoder v5.5.0^{122,123}. A final, non-redundant reference gene set was generated by merging the three annotated gene sets using

EvidenceModeler v1.1.1 (EVM)¹²⁴. The gene models were translated into amino acid sequences and used in local BLASTp¹⁰³ searches against the public databases Kyoto Encyclopedia of Genes and Genomes (KEGG; v89.1)¹²⁵, NCBI non-redundant protein sequences (NR; v20170924)¹²⁶, Swiss-Prot (release-2018_07)¹²⁷, TrEMBL (Translation of EMBL [nucleotide sequences that are not in Swiss-Prot]; release-2018_07)¹²⁸, and InterPro (v69.0)¹²⁹. 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)^{130,131}. The coding sequences (CDS) for each species were aligned using PRANK v70427^{132,133} and filtered by Gblocks v0.91b¹³⁴ to identify conserved blocks (removing gaps, ambiguous sites, and excluding alignments less than 300 bp in size). Finally, 7695 single-copy genes were concatenated into supergenes for phylogenetic analyses.

To identify conserved non-exonic elements (CNEEs), we generated whole-genome alignments (WGs) using LASTZ v1.04.22¹³⁵ 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 (v11.2)¹³⁶. To estimate the non-conserved model, we employed phyloFit (v1.4) in the PHAST package¹³⁷ with 4d sites in the Afrotheria alignments and the topology using default parameters. We ran phastCons with the non-conserved model to estimate 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 103.9 M 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_005129.2), 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 (NC_006924.1). We used MARS (Multiple circular sequence Alignment using Refined Sequences)¹³⁸ 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 v8.2.9¹³⁹ (1000 bootstrap iterations) using coding sequences from 7695 genes, 1,127,156 fourfold degenerate sites in the 7695 genes, 5508 single-copy Benchmarking Universal Single-Copy Ortholog (BUSCO) genes^{8,140}, 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 v5.6.2¹⁴¹ to generate a coalescent species tree from non-overlapping 20 kb windows (to minimize linkage between subsequent windows from the WGs generated above (see ref. 142)). Briefly, after excluding windows from alignments less than 2000-bp in size and with more than 25% gaps, 152,478 windows were used to generate window-based gene trees (WGTs) using RaxML (with 1000 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 gene trees 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

v4.8¹⁴³) 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 TimeTree¹⁴⁴: 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). We also included data from a recent manuscript² that employed a total-evidence approach (i.e., 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 mode^{145–148}. We employed a recently developed pipeline¹⁴⁷ that requires the 2000 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) test¹⁴⁹ was used to assess presence/absence matrixes.

Comparative genomics analysis strategy. To summarize our analysis strategy (see detailed methods below) and manuscript data, we first compared signatures of natural selection with literature searches (comprehensive reviews on the anatomical and physiological adaptations of sirenians to aquatic life, including refs. 5,150) to discover broad functional categories associated with sirenian adaptations. Enrichment analysis of positively selected (Supplementary Data 1 and 2) and rapidly evolving (Supplementary Data 3 and 4) genes using KOBAS revealed an over-representation (Benjamini–Hochberg $P < 0.05$) of terms related to thyroid hormone synthesis, the cardiovascular system, integumentary system (i.e., cornified envelope), and circadian activity. Sirenian-specific amino acid substitutions in the thyroid hormone pathway and circadian clock proteins were next identified using FasParser2^{151,152} (Supplementary Data 5) and validated against our genome resequencing data set of 99 dugongs, the 120 mammalian species in OrthoMaM¹⁵³, and by BLAST¹⁰³ searches of NCBI and Ensembl databases. Functional *in vitro* assays were used to evaluate selected substitutions. CAFE¹⁵⁴ revealed loss of gene families of the integumentary (i.e., cornification/keratinization) and olfactory systems (Supplementary Data 6 and 7). A recently described pipeline¹⁵⁵ confirmed reported pseudogenes (e.g., refs. 89,32) among the 15 shared by sirenians (Supplementary Data 8 and Table S5) but also gene inactivation events not previously described.

Gene family analysis. Gene family expansion and contraction analysis was performed using CAFE v4.2¹⁵⁴ with our consensus phylogenetic tree (also see Fig. S1c) 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 7695 single-copy gene family orthologs in our nine-species data set were identified using their coding sequences and PAML codeml v4.8¹⁴³.

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 PAML¹⁵⁶ 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.

We employed KOBAS v3.0 [<http://kobas.cbi.pku.edu.cn>]¹⁵⁷ gene enrichment with a Benjamini–Hochberg P value cut-off set at 0.05 to identify functional categories that may underlie aquatic specializations of sirenians. Gene sets were also interrogated using STRING v12.0 [<https://string-db.org>]¹²⁷, which includes “Reference publications” (i.e., publications with PubMed abstracts up to August 2022 and the PMC Open Access Subset up to April 2022) (false discovery rate cut-off set at 0.05).

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 approach¹⁵⁵. 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 v36¹¹¹ and genBlastA v1.0.1¹⁵⁸. Next, the mapped genomic regions and 1000 bp upstream and downstream were examined for disruptions (nonsense mutations and frameshifts) to the gene coding sequences using GeneWise v2.4.1¹¹². 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 sequences¹⁵⁹. False positives with disruptive mutations introduced by GeneWise, sequencing errors, or annotation errors were removed following the steps in ref. 155. 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 using BLAST¹⁰³ to confirm each disruption. Gene enrichment analysis was performed using KOBAS and STRING, as outlined above.

Sirenian lineage-specific amino acid changes. Amino acid alignments of the 7695 single-copy orthologs in our nine-species data set (eight afrotherians and Linnaeus's two-toed sloth) and FasParser v2^{151,152} was used to identify amino acid residues specific to the sirenian lineage. Putative sirenian-specific residues of interest were further validated using 120 mammalian sequences downloaded from OrthoMaM v10c¹⁵³ (available at FigShare [<https://doi.org/10.6084/m9.figshare.23975559>]), as well as by interrogating NCBI and Ensembl databases using BLAST. Potential functional effects of substitutions were predicted by PolyPhen-2^{160,161} and SIFT v6.2.1^{162,163}. 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. Each DNA was individually subcloned into the *pcDNA3.1* vector (Invitrogen) to generate NIS plasmids. All expression 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 μ g NIS plasmid using the PEI 40 K Transfection Reagent (Servicebio). 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 h. Briefly, cells were incubated with Na¹²⁵I (Shanghai Xinke Pharmaceutical Co.) for 1 h. The Na¹²⁵I added to each well of the culture plate was counted using a radioactivity meter (FJ-391A4) to obtain the total radioactivity 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* using the PEI 40 K Transfection Reagent (Servicebio). Two days after transfection, cells were lysed in Western IP Cell Lysis buffer (Beyotime) 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 performed using mouse monoclonal anti-HA (Proteintech cat. no 66006-2-Ig at 1:50,000 dilution) or anti-FLAG (Proteintech cat. no 66008-4-Ig at 1:25000 dilution) antibodies, and an anti-mouse secondary antibody (Proteintech cat. no SA00001-1 at 1:6000 dilution). A StarRuler broad-range (10–180 kDa) molecular weight marker (GenStar cat. no M221) was co-run to estimate protein weights.

Whole-genome resequencing. DNA from 99 dugongs sampled on the Australian east coast (see “Sample collection and research ethics”) was extracted using a QIAamp DNA Mini Kit (QIAGEN) and 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)⁵⁴, and Exmouth Gulf (Western Australia, Australia; [SRR17870680](#); sampled 3 June 2018). We 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)¹⁶⁴. The manatee reads were employed as the outgroup in population genomics analyses. Raw data were filtered using SOAPnuke v2.1.5⁹⁴ to remove adapters and low-quality reads. For comparative analyses, all samples were down-sampled to 10× coverage using SAMtools v1.9¹²¹.

Because we had a chromosome-level dugong genome, 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)^{165,166}. 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 by ref. 164) 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*)¹⁰⁴ and dugong Y chromosome genes¹⁰⁵. This effort revealed that the -167 Mb HiC_scaffold_7 and the -149 Mb chr7 in

West Indian manatee and dugong correspond to their respective X chromosome. Next, we aligned reads to the chromosome-level genomes using bowtie2 v2.3.4.3^{167,168} (parameter: –no-unal to only retained mapping reads), followed by conversion to a BAM file and filtering using SAMtools v1.7^{121,169} (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 via ref. 165. Average sequencing depth was estimated from indexed and sorted BAM files using mosdepth v0.3.3¹⁷⁰ (parameters: -n –fast-mode –by 500). See Supplementary Data 9 for sample statistics.

Identification and characterization of SNPs. Including the X chromosome can interfere with demographic history estimates^{73,171,172} and other population genomics analyses^{173,174} (X chromosome SNP effects). Therefore, reads that could be mapped to the X chromosome were removed using bowtie2 (v2-2.2.5)¹⁶⁸ 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-r1039¹⁷⁵ with default parameters. SAMtools v1.2¹⁷⁶ 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 v4.1.2.0¹⁷⁷ 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 v0.1.16¹⁷⁸ with the parameters “–max-missing 0.95”. Plink (v1.90b6.6)¹⁷⁹ 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 v1.3.0¹⁸⁰, 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 i-ibs” MEGA7¹⁸¹ was used to construct the Neighbor-Joining Tree (NJ tree) based on the IBS matrix, and the evolutionary tree was visualized using the Interactive Tree Of Life (iTOL) online tool v6 [<https://itol.embl.de>]¹⁸². In addition, PCA (principal component analysis) was performed using ggplot2 package¹⁸³. The divergence time between population groups was estimated using dadi v2.1¹⁸⁴.

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 data set. Additionally, we estimated runs of homozygosity (ROH) using the R package *detectRUNS* [<https://cran.r-project.org/web/packages/detectRUNS/vignettes/detectRUNS.vignette.html>] with the parameters “windowSize = 50, minSNP = 30, maxGap = 1,000,000, minLengthBps = 100,000, minDensity = 1/100,000” based on the same filtered SNPs set used in the “population structure analysis” section.

ROH’s per dugong generation was calculated as follows. FASTEPFR v2¹⁸⁵ was used to estimate the recombination rate of the dugong population based on the 99 Queensland individuals. It was estimated to be -1.08 cM/Mb. The coalescent times of ROH for each

subpopulation were calculated as $g = 100/(2rL)$ (see ref. 186), where g is the expected time (in generations; 27 years⁴³) back to the parental common ancestor, r is the recombination rate and L is the length of the ROH in megabases. Thus, ROHs <100 kb 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 >500 kb, up to 2511 years ago (93 generations); ROHs >1 Mb, up to 1242 years ago (46 generations); ROHs >2 Mb, up to 621 years ago (23 generations); ROHs >5 Mb, up to 243 years ago (9 generations); ROHs >10 Mb, up to 135 years ago (5 generations); and ROHs >20 Mb, up to 54 years ago (2 generations).

Identification of selective sweep regions. Four different methods, including F_{ST} (population differentiation), π (relative nucleotide diversity), XP-EHH (the cross-population extended haplotype homozygosity statistic)⁶², and XP-CLR (the cross-population composite likelihood ratio test)⁶³ were employed to assess selective sweeps between the northern and southern Queensland groups. Briefly, *vcftools* v0.1.16¹⁷⁸ with the parameters “-max-missing 0.95 -maf 0.01” was used to perform autosomal SNPs quality control. Next, the pairwise fixation index (F_{ST}) was calculated between the seven Queensland locations and between the whole northern (Torres Strait, Bowling Green Bay, and Airlie Beach) and southern (Moreton Bay, Great Sandy Straits, Hervey Bay, and Clairview) groups from Queensland 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 *xpclr* v1.1.2⁶³ 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) pipeline¹⁸⁷, followed by *vcftools* with the parameter “-IMPUTE”. Next, the *xpeshbin* module from Hapbin v1.3.0¹⁸⁸ was used to calculate XP-EHH values (see ref. 189). The XP-EHH scores were normalized and corresponding P values were calculated. If a P value was less than 0.01, we considered the region 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. 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 previously⁴³. We estimated the mutation rate for dugong using *r8s*¹⁹⁰ 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.

Pairwise sequentially Markovian coalescent (PSMC) v0.6.5-r67⁷³ 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 ago⁷³. We first constructed diploid consensus sequences for each sample using SAMtools v1.9¹⁷⁶ *mpileup* and *BCftools* v1.4¹⁶⁹ with the parameters “-C50” and “-d 4 -D 24”. The consensus sequences were transformed to PSMC input format using *fq2psmcf* 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.

Because of the insufficient resolution of PSMC in estimating demography more recently than ~20 kya¹⁹¹, we employed SMC++ v1.15.5¹⁹² to infer more recent population history for Queensland dugong individuals (see “Analysis of population structure” above). The SMC++ modules *vcf2smc*, *estimate*, and *plot* were used. One of the 32 individuals sampled from Moreton Bay (MB16796; an older female) clustered with the northern group in a PCA (Fig. 5b) and a neighbor-

joining tree (Fig. S8). It may reflect low-level individualistic movement rather than population migration (see refs. 60,193,194) between northern and southern populations and was excluded from the demographic history analysis.

To measure gene flow (i.e., migration) between populations, we interrogated our SNP data set with *Dsuite* (across all dugong populations with the West Indian manatee as the outgroup)¹⁹⁵ to obtain Patterson’s D (ABBA-BABA statistic) and *TreeMix* v1.13⁶¹ to visualize gene flow (migration) on a maximum likelihood tree of populations. *TreeView* was run with the parameters “-root TS -k 500000 -m 0-10”. Torres Strait, TS, was used to root (parameter *-root*) the tree, SNPs were grouped in 500,000-bp windows (parameter *k*). We estimated the optimal number of migration events (parameter *m*) for *TreeMix* analysis using the R package *optM*¹⁹⁶.

Reporting summary

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

Data availability

Dugong sequencing reads (including stLFR, Hi-C, RNA-seq, resequencing) and the *Dugong_BGI* genome assembly are available at NCBI BioProject [PRJNA114306](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA114306). Dugong SNP data in VCF format are available at the European Nucleotide Archive (ENA) and linked to the NCBI BioProject. Multiple sequence alignments (MSAs) of thyroid hormone pathway and circadian clock genes with sirenian-specific amino acid substitutions are available on FigShare [<https://doi.org/10.6084/m9.figshare.23975559>]. Public datasets used in this study are available from NCBI RefSeq, NCBI SRA, and DNA Zoo. Table S2 lists afrotherian nuclear genome assemblies used in protein ortholog prediction and whole-genome alignments. Linnaeus’s two-toed sloth was used as an outgroup species (NCBI assembly *mChoDid1.pri*). Dugong genome annotation employed genes from the NCBI assemblies of West Indian manatee (ASM3001377v1), African bush elephant (*LoxAfr3.0*), Cape elephant shrew (*EleEdw1.0*), aardvark (*OryAfe1.0*), and human (GRCh38.p12). Mitochondrial genome assemblies and protein-coding sequences were obtained from NCBI dugong (NC_003314.1), West Indian manatee (MN105083.1), Asian elephant (NC_005129.2), rock hyrax (NC_004919.1), lesser hedgehog tenrec (NC_002631.2), golden mole (NC_004920.1), aardvark (NC_002078.1), Cape elephant shrew (NC_041486.1), and Linnaeus’s two-toed sloth (NC_006924.1). Gene loss was validated using NCBI SRA reads from West Indian manatee (SRR8616893, SRR331138, SRR24090881, SRR24090877, and SRR24090880) and Steller’s sea cow (ERR5559486, SRR12067494, and SRR12067500). Population genomic analyses employed SRA reads from dugong (DRR251525, ERR5621402, and SRR17870680) and West Indian manatee (SRR331137, SRR331139, and SRR331142).

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 (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 (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. 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).

7. Marsh, H. & Soltzick, S. Dugong dugon (amended version of 2015 assessment). The IUCN Red List of Threatened Species 2019: e.T6909A160756767. en. Downloaded on 20 September 2021. <https://doi.org/10.2305/IUCN>. UK. 2015-4. RLTS. T6909A160756767 (2019).
8. Seppey, M., Manni, M. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).
9. Leung, A. M. & Braverman, L. E. Consequences of excess iodine. *Nat. Rev. Endocrinol.* **10**, 136–142 (2014).
10. Yun, A. J. & Doux, J. D. Iodine in the ecosystem: an overview. (eds. Preedy, V. R., Burrow, G. N. & Watson, R.) *Comprehensive Handbook of Iodine*. 119–123 (Elsevier Inc, 2009).
11. Hohmann, G., Ortmann, S., Remer, T. & Fruth, B. Fishing for iodine: what aquatic foraging by bonobos tells us about human evolution. *BMC Zool.* **4**, 5 (2019).
12. 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**, cozo41 (2019).
13. Ortiz, R. M., Mackenzie, D. S. & Worthy, G. A. Thyroid hormone concentrations in captive and free-ranging West Indian manatees (*Trichechus manatus*). *J. Exp. Biol.* **203**, 3631–3637 (2000).
14. 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 (2017).
15. Portulano, C., Paroder-Belenitsky, M. & Carrasco, N. The Na⁺/I⁻ symporter (NIS): mechanism and medical impact. *Endocr. Rev.* **35**, 106–149 (2014).
16. Purtell, K. et al. The KCNQ1-KCNE2 K(+) channel is required for adequate thyroid I(-) uptake. *FASEB J.* **26**, 3252–3259 (2012).
17. Roepke, T. K. et al. Kcne2 deletion uncovers its crucial role in thyroid hormone biosynthesis. *Nat. Med.* **15**, 1186–1194 (2009).
18. 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 (2021).
19. Ravera, S. et al. Structural insights into the mechanism of the sodium/iodide symporter. *Nature* **612**, 795–801 (2022).
20. 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 (2013).
21. 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 (2010).
22. 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 (2005).
23. Berta, A., Sumich, J. L. & Kovacs, K. M. In *Marine Mammals* (Third Edition) (eds Berta, A., Sumich, J. L., & Kovacs, K. M.) 169–210 (Academic Press, 2015).
24. 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*). *Mar. Mammal. Sci.* **18**, 765–778 (2002).
25. 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. *Mar. Biol.* **161**, 1395–1407 (2014).
26. Chernova, O., Kiladze, A. & Shpak, O. In *Doklady Biological Sciences*. 150–156 (Springer).
27. Szklarczyk, D. et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
28. Shoshani, J., Goldman, C. A. & Thewissen, J. *Orycteropus afer*. *Mamm. Species* (300) 1–8 (1988).
29. 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 (2020).
30. Hazlerigg, D. G. & Tyler, N. J. C. Activity patterns in mammals: circadian dominance challenged. *PLoS Biol.* **17**, e3000360 (2019).
31. 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 (2021).
32. 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 (2019).
33. Yin, D. et al. Gene duplication and loss of AANAT in mammals driven by rhythmic adaptations. *Mol. Biol. Evol.* **38**, 3925–3937 (2021).
34. 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 (2021).
35. Lopes-Marques, M. et al. The singularity of cetacea behavior parallels the complete inactivation of melatonin gene modules. *Genes* **10**, 121 (2019).
36. Mascetti, G. G. Unihemispheric sleep and asymmetrical sleep: behavioral, neurophysiological, and functional perspectives. *Nat. Sci. Sleep.* **8**, 221–238 (2016).
37. Zeh, D. R. et al. Evidence of behavioural thermoregulation by dugongs at the high latitude limit to their range in eastern Australia. *J. Exp. Mar. Biol. Ecol.* **508**, 27–34 (2018).
38. Mukhametov, L. M., Lyamin, O. I., Chetyrbok, I. S., Vassilyev, A. A. & Diaz, R. P. Sleep in an Amazonian manatee, *Trichechus inunguis*. *Experientia* **48**, 417–419 (1992).
39. Narasimamurthy, R. & Virshup, D. M. The phosphorylation switch that regulates ticking of the circadian clock. *Mol. Cell* **81**, 1133–1146 (2021).
40. Lalic, T. et al. TRESK is a key regulator of nocturnal suprachiasmatic nucleus dynamics and light adaptive responses. *Nat. Commun.* **11**, 4614 (2020).
41. Espregueira Themudo, G. et al. Losing genes: the evolutionary remodeling of cetacea skin. *Front. Mar. Sci.* **7**, 912 (2020).
42. 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 (2017).
43. McDonald, B. J. Population Genetics of Dugongs around Australia: Implications of Gene Flow and Migration (Ph.D. Thesis) (James Cook University 2005).
44. Florida Fish and Wildlife Conservation Commission. Manatee mortality event along the East Coast: 2020–2022, Accessed 15 Feb 2023. <https://myfwc.com/research/manatee/rescue-mortality-response/ume> (2022).
45. 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 (Springer, Cham, 2017).
46. Du, J., Chen, B., Nagelkerken, I., Chen, S. & Hu, W. Protect seagrass meadows in China’s waters. *Science* **379**, 447 (2023).
47. Castellanos, A. et al. Pyrethroids inhibit K2P channels and activate sensory neurons: basis of insecticide-induced paraesthesias. *Pain* **159**, 92–105 (2018).

48. Castellanos, A. et al. TRESK background K(+) channel deletion selectively uncovers enhanced mechanical and cold sensitivity. *J. Physiol.* **598**, 1017–1038 (2020).
49. 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 (2019).
50. Bossart, G. D., Meisner, R. A., Rommel, S., Ghim, S.-J. & Jenson, A. B. Pathological features of the Florida manatee cold stress syndrome. *Aquat. Mamm.* **29**, 9–17 (2003).
51. 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 (2022).
52. 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 (2008).
53. 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). *Mar. Biol.* **162**, 1147–1149 (2015).
54. Le Duc, D. et al. Genomic basis for skin phenotype and cold adaptation in the extinct Steller’s sea cow. *Sci. Adv.* **8**, eabl6496 (2022).
55. Kardos, M. et al. Inbreeding depression explains killer whale population dynamics. *Nat. Ecol. Evol.* **7**, 675–686 (2023).
56. Hoelzel, A. R. et al. Genomics of post-bottleneck recovery in the northern elephant seal. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-024-02337-4> (2024).
57. Zhang, P. et al. An Indo-Pacific humpback dolphin genome reveals insights into chromosome evolution and the demography of a vulnerable species. *iScience* **23**, 101640 (2020).
58. Baker, D. N. et al. A chromosome-level genome assembly for the dugong (*Dugong dugon*). *J. Hered.* <https://doi.org/10.1093/jhered/esae003> (2024).
59. Allen, S., Marsh, H. & Hodgson, A. Occurrence and Conservation of the Dugong (*Sirenia: Dugongidae*) in New South Wales. *Proc. Linn. Soc. N.S.W.* **125**, 211–216 (2004).
60. McGowan, A. M. et al. Cryptic marine barriers to gene flow in a vulnerable coastal species, the dugong (*Dugong dugon*). *Mar. Mammal Sci.* **39**, 918–939 (2023).
61. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
62. Pickrell, J. K. et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res.* **19**, 826–837 (2009).
63. Chen, H., Patterson, N. & Reich, D. Population differentiation as a test for selective sweeps. *Genome Res.* **20**, 393–402 (2010).
64. 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 (2021).
65. Seo, J. H. et al. The mitochondrial unfoldase-peptidase complex ClpXP controls bioenergetics stress and metastasis. *PLoS Biol.* **14**, e1002507 (2016).
66. Chappell, J. Evidence for smoothly falling sea level relative to north Queensland, Australia, during the past 6000 yr. *Nature* **302**, 406–408 (1983).
67. 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. *Mar. Geol.* **370**, 20–30 (2015).
68. Carter, A. B. et al. A spatial analysis of seagrass habitat and community diversity in the Great Barrier Reef World Heritage Area. *Sci. Rep.* **11**, 22344 (2021).
69. Lin, M., Turvey, S. T., Liu, M., Ma, H. & Li, S. Lessons from extinctions of dugong populations. *Science* **378**, 148 (2022).
70. Lin, M. et al. Functional extinction of dugongs in China. *R. Soc. Open Sci.* **9**, 211994 (2022).
71. 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 (2022).
72. 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).
73. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
74. 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 (2021).
75. Cousins, T., Tabin, D., Patterson, N., Reich, D. & Durvasula, A. Accurate inference of population history in the presence of background selection. *bioRxiv*, <https://doi.org/10.1101/2024.01.18.576291> (2024).
76. 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 hammerheads. *iScience* **26**, 105815 (2023).
77. 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 (2018).
78. 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. *Wildl. Res.* **31**, 283–290 (2004).
79. von Seth, J. et al. Genomic insights into the conservation status of the world’s last remaining Sumatran rhinoceros populations. *Nat. Commun.* **12**, 2393 (2021).
80. Rudolf, A. M. et al. A single nucleotide mutation in the dual-oxidase 2 (DUOX2) gene causes some of the panda’s unique metabolic phenotypes. *Natl Sci. Rev.* **9**, nwab125 (2022).
81. 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 (2017).
82. Liu, J. et al. Differential MC5R loss in whales and manatees reveals convergent evolution to the marine environment. *Dev. Genes Evol.* **232**, 81–87 (2022).
83. Lopes-Marques, M. et al. Complete inactivation of sebump-producing genes parallels the loss of sebaceous glands in cetacea. *Mol. Biol. Evol.* **36**, 1270–1280 (2019).
84. 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 (2018).
85. 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 (2021).
86. Sun, X. et al. Comparative genomics analyses of alpha-keratins reveal insights into evolutionary adaptation of marine mammals. *Front. Zool.* **14**, 41 (2017).
87. Zhang, X. et al. Parallel independent losses of G-type lysozyme genes in hairless aquatic mammals. *Genome Biol. Evol.* **13**, evab201 (2021).
88. Albalat, R. & Canestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).
89. Meyer, W. K. et al. Ancient convergent losses of Paraoxonase 1 yield potential risks for modern marine mammals. *Science* **361**, 591–594 (2018).

90. Stronen, A. V., Norman, A. J., Vander Wal, E. & Paquet, P. C. The relevance of genetic structure in ecotype designation and conservation management. *Evol. Appl.* **15**, 185–202 (2022).
91. Lanyon, J. M., Sneath, H. L. & Long, T. Three skin sampling methods for molecular characterisation of free-ranging dugong (*Dugong dugon*) populations. *Aquat. Mamm.* **36**, 298 (2010).
92. Haklay, M. & Weber, P. Openstreetmap: User-generated street maps. *IEEE Pervasive Comput.* **7**, 12–18 (2008).
93. Wang, O. et al. Efficient and unique cobarcode 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 (2019).
94. 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 (2018).
95. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
96. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
97. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
98. Pryszcz, L. P. & Gabaldon, T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.* **44**, e113 (2016).
99. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
100. Dudchenko, O. et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
101. Robinson, J. T. et al. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258.e251 (2018).
102. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215–225, (2003).
103. Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
104. 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 (2009).
105. 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 (2008).
106. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11 (2015).
107. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* <https://doi.org/10.1002/0471250953.bi0410s25> (2009).
108. Smit, A. F. & Hubley, R. RepeatModeler Open-1.0. 2008–2015. Available at <http://www.repeatmasker.org>. (2010).
109. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinform.* **9**, 18 (2008).
110. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
111. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
112. Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).
113. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–439, (2006).
114. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
115. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
116. Roehr, J. T., Dieterich, C. & Reinert, K. Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* **33**, 2941–2942 (2017).
117. Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-flexible barcode and adapter processing for next-generation sequencing platforms. *Biology* **1**, 895–905 (2012).
118. Kopylova, E., Noe, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
119. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–596, (2013).
120. 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 (2019).
121. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
122. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
123. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
124. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
125. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
126. O’Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
127. UniProt Consortium Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40**, D71–75, (2012).
128. O’Donovan, C. et al. High-quality protein knowledge resource: SWISS-PROT and TrEMBL. *Brief. Bioinform.* **3**, 275–284, (2002).
129. Mitchell, A. L. et al. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360 (2019).
130. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
131. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
132. Loytynoja, A. Phylogeny-aware alignment with PRANK and PAGAN. *Methods Mol. Biol.* **2231**, 17–37 (2021).
133. Loytynoja, A. Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155–170 (2014).
134. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577 (2007).
135. Harris, R. S. Improved Pairwise Alignment of Genomic DNA (The Pennsylvania State University, 2007).
136. Blanchette, M. et al. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
137. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
138. Ayad, L. A. & Pissis, S. P. MARS: improving multiple circular sequence alignment using refined sequences. *BMC Genom.* **18**, 86 (2017).
139. Stamatakis, A. RAXML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

140. Khalturin, K. et al. Polyzoa is back: The effect of complete gene sets on the placement of Ectoprocta and Entoprocta. *Sci. Adv.* **8**, eabo4400 (2022).
141. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinforma.* **19**, 153 (2018).
142. Edelman, N. B. et al. Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
143. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
144. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
145. Mason, V. C. et al. Genomic analysis reveals hidden biodiversity within colugos, the sister group to primates. *Sci. Adv.* **2**, e1600633 (2016).
146. Springer, M. S., Molloy, E. K., Sloan, D. B., Simmons, M. P. & Gatesy, J. ILS-aware analysis of low-homoplasy retroelement insertions: inference of species trees and introgression using quartets. *J. Hered.* **111**, 147–168 (2020).
147. Feng, S. et al. Incomplete lineage sorting and phenotypic evolution in marsupials. *Cell* **185**, 1646–1660.e1618 (2022).
148. Doronina, L., Reising, O., Clawson, H., Ray, D. A. & Schmitz, J. True homoplasy of retrotransposon insertions in primates. *Syst. Biol.* **68**, 482–493 (2019).
149. 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 (2016).
150. Berta, A., Sumich, J. L. & Kovacs, K. M. *Marine Mammals* (Elsevier, 2015).
151. Sun, Y. B. FasParser2: a graphical platform for batch manipulation of tremendous amount of sequence data. *Bioinformatics* **34**, 2493–2495 (2018).
152. Sun, Y. B. FasParser: a package for manipulating sequence data. *Zool. Res.* **38**, 110–112 (2017).
153. 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 (2019).
154. 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 (2006).
155. 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 (2022).
156. Yang, Z., Wong, W. S. & Nielsen, R. Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.* **22**, 1107–1118 (2005).
157. Bu, D. et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* **49**, W317–W325 (2021).
158. She, R., Chu, J. S., Wang, K., Pei, J. & Chen, N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res.* **19**, 143–149 (2009).
159. 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 (2010).
160. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
161. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **20**, <https://doi.org/10.1002/0471142905.hg0720s76> (2013).
162. Sim, N.-L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40**, W452–W457 (2012).
163. 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 (2009).
164. Foote, A. D. et al. Convergent evolution of the genomes of marine mammals. *Nat. Genet.* **47**, 272–275 (2015).
165. 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 (2020).
166. 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 (2016).
167. Langmead, B., Wilks, C., Antonescu, V. & Charles, R. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* **35**, 421–432 (2019).
168. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
169. Danecek, P. et al. Twelve years of SAMtools and BCftools. *Giga-science* **10**, giab008 (2021).
170. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
171. 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 (2011).
172. Palkopoulou, E. et al. Complete genomes reveal signatures of demographic and genetic declines in the woolly mammoth. *Curr. Biol.* **25**, 1395–1400 (2015).
173. 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 (2021).
174. 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 (2022).
175. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
176. Guindon, S., Delsuc, F., Dufayard, J.-F. & Gascuel, O. In *Bioinformatics for DNA Sequence Analysis* (ed. Posada, D.) 113–137 (Humana Press, 2009).
177. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
178. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
179. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
180. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 1–6 (2011).
181. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
182. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
183. Villanueva, R. A. M. & Chen, Z. J. (Taylor & Francis, 2019).
184. Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).

185. 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 Genet.* **6**, 1563–1571 (2016).
186. Khan, A. et al. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc. Natl Acad. Sci. USA* **118**, <https://doi.org/10.1073/pnas.2023018118> (2021).
187. Webb, A. et al. The pop-gen pipeline platform: a software platform for population genomic analyses. *Mol. Biol. Evol.* **38**, 3478–3485 (2021).
188. 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. *Mol. Biol. Evol.* **32**, 3027–3029 (2015).
189. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
190. 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).
191. Nishihara, H., Hasegawa, M. & Okada, N. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl Acad. Sci. USA* **103**, 9929–9934 (2006).
192. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
193. Sheppard, J. K. et al. Movement heterogeneity of dugongs, *Dugong dugon* (Müller), over large spatial scales. *J. Exp. Mar. Biol. Ecol.* **334**, 64–83 (2006).
194. Deutsch, C. J., Castellanico-Martínez, D. N., Groom, R. & Cleguer, C. In *Ethology and Behavioral Ecology of Sirenia* 155–231 (Springer, 2022).
195. Malinsky, M., Matschiner, M. & Svoldal, H. Dsuite—Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
196. Fitak, R. R. OptM: estimating the optimal number of migration edges on population trees using Treemix. *Biol. Methods Protoc.* **6**, bpab017 (2021).
- National Natural Science Foundation of China (grant 42225604 to S.L. and grant 32270441 to R.T.), the Young Elite Scientists Sponsorship Program of the China Association for Science and Technology (grant 2023QNR001 to R.T.), the Sea World Foundation (Australia) and Winifred Violet Scott Foundation (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. J.M.L. and 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.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-49769-x>.

Correspondence and requests for materials should be addressed to Guangyi Fan, Songhai Li or Inge Seim.

Peer review information *Nature Communications* thanks Xin Liu and Sankar Subramanian for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

Acknowledgements

We thank those who assisted with the Australian dugong tissue collection, including The University of Queensland Dugong Team (especially Alex McGowan, Erin Neal, and Rob Slade) members of the Mabuig 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, Dr. Erina Young at Murdoch University (Australia) for information on the dugong individual from Exmouth Gulf, Western Australia, Dr. Shaohong Feng (BGI Research) for access to scripts associated with their retroelement phylogenetic marker pipeline¹⁴⁷, and Prof. Harold H. Zakon (The University of Texas at Austin) for helpful feedback on the revised manuscript. Unpublished genome assemblies and sequencing data for the West Indian manatee, Asian elephant, rock hyrax, and aardvark were used with permission from the DNA Zoo Consortium [<https://www.dnazoo.org>]. Support for this research was provided by the Chinese Ministry of Science and Technology National Key Programme of Research and Development (grant 2022YFF1301601 to R.T. and S.L.), the

¹Integrative Biology Laboratory, Nanjing Normal University, Nanjing 210023, China. ²BGI Research, Qingdao 266555, China. ³BGI Research, Shenzhen 518083, China. ⁴Qingdao Key Laboratory of Marine Genomics BGI Research, Qingdao 266555, China. ⁵Marine Mammal and Marine Bioacoustics Laboratory, Institute of Deep-sea Science and Engineering, Chinese Academy of Sciences, Sanya 572000, China. ⁶The Innovation Research Center for Aquatic Mammals, and Key

Laboratory of Aquatic Biodiversity and Conservation of the Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China. ⁷Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. ⁸School of Life Sciences, University of Science and Technology of China, Hefei 230027, China. ⁹School of the Environment, The University of Queensland, Lucia 4072, Australia. ¹⁰School of Veterinary Sciences, The University of Adelaide, Roseworthy 5371, Australia. ¹¹State Key Laboratory of Agricultural Genomics, BGI Research, Shenzhen 518083, China. ¹²These authors contributed equally: Ran Tian, Yaolei Zhang, Hui Kang. ✉ e-mail: fanguangyi@genomics.cn; lish@idsse.ac.cn; inge@seimlab.org