Differential sensory and immune gene evolution in sea turtles with contrasting demographic and life histories

Blair P. Bentley[1,✉], Tomás Carrasco-Valenzuela[2,3], Elisa K. S. Ramos[2,3,4], Harvinder Pawar[5], Larissa Souza Arantes[2,3], Alana Alexander[6], Shreya M. Banerjee[1,7], Patrick Masterson[8], Martin Kuhlwilm[5,9,10], Martin Pippel[11,12], Jacquelyn Mountcastle[13,22], Bettina Haase[13], Marcela Uliano Silva[2,3], Giulio Formenti[13,14], Kerstin Howe[15], William Chow[15], Alan Tracey[15], Yumi Sims[15], Sarah Pelan[15], Jonathan Wood[15], Justin R. Perrault[16], Kelly Stewart[17], Scott R. Benson[18,19], Yaniv Levy[20], Erica V. Todd[21], H. Bradley Shaffer[22,23], Peter Scott[22,24], Brian T. Henen[25], Robert W. Murphy[26], David W. Mohr[27], Alan F. Scott[27], Neil J. Gemmell[28], Alexander Suh[29,30], Sylke Winkler[11,31], Françoise Thibaud-Nissen[8], Mariana F. Nery[4], Tomas Marques-Bonet[5,32,33,34], Agostinho Antunes[35,36], Yaron Tikochinski[37], Peter H. Dutton[17], Olivier Fedrigo[13], Eugene W. Myers[11,12,38], Erich D. Jarvis[13,14,39], Camila J. Mazzoni[2,3,*], and Lisa M. Komoroske[1,*]

[1]Department of Environmental Conservation, University of Massachusetts, Amherst, MA, USA; [2]Evolutionary Genetics Department, Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany; [3]Berlin Center for Genomics in Biodiversity Research, Berlin, Germany; [4]Department of Genetics, Evolution, Microbiology and Immunology, University of Campinas, Campinas, Brazil; [5]Institut de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra), PRBB, Doctor Aiguader 88, Barcelona, Catalonia, Spain; [6]Department of Anatomy, School of Biomedical Sciences, University of Otago, Dunedin, New Zealand; [7]Department of Biology, Stanford University, Stanford CA, USA; [8]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD, USA; [9]Department of Evolutionary Anthropology, University of Vienna, Vienna, Austria; [10]Human Evolution and Archaeological Sciences (HEAS), University of Vienna, Austria; [11]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany; [12]Center for Systems Biology, Dresden, Germany; [13]Vertebrate Genome Lab, The Rockefeller University, New York, NY, USA; [14]Laboratory of Neurogenetics of Language, The Rockefeller University, New York, NY, USA; [15]Tree of Life, Wellcome Sanger Institute, Cambridge, CB10 1SA, UK; [16]Loggerhead Marinelife Center, Juno Beach, FL, USA; [17]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, La Jolla, CA, United States; [18]Marine Mammal and Turtle Division, Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Moss Landing, CA, USA; [19]Moss Landing Marine Laboratories, Moss Landing, CA, USA; [20]National Sea Rescue Centre, Israel's Nature and Parks Authority, Mevoot Yam, Michmoret, Israel; [21]School of Life and Environmental Sciences, Deakin University, Queenscliff, VIC, Australia; [22]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA; [23]La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA; [24]Department of Life, Earth, and Environmental Sciences, West Texas A&M University, Canyon, TX, USA; [25]Environmental Affairs, Marine Air Ground Task Force and Training Command, Marine Corps Air Ground Combat Center, Twentynine Palms CA, USA; [26]Centre for Biodiversity, Royal Ontario Museum, Toronto, Ontario, Canada; [27]Genetic Resources Core Facility, McKusick-Nathans Dept of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore MD 21287, USA; [28]Allan Wilson Centre, Department of Anatomy, University of Otago, Dunedin, New Zealand; [29]School of Biological Sciences, University of East Anglia, Norwich NR4 7TU, UK; [30]Department of Organismal Biology, Evolutionary Biology Centre (EBC), Science for Life Laboratory, Uppsala University, Uppsala, Sweden; [31]DRESDEN concept Genome Center, Dresden, Germany; [32]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain; [33]Institucio Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia 08010, Spain; [34]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona; [35]CIIMAR/CIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Av. General Norton de Matos, s/n, 4450-208, Porto, Portugal; [36]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007, Porto, Portugal; [37]Faculty of Marine Sciences, Ruppin Academic Center, Michmoret, Israel; [38]Faculty of Computer Science, Technical University Dresden, Dresden, Germany; [39]Howard Hughes Medical Institute, Chevy Chase, MD, USA.

Classification: Genetics and Evolution

Keywords: marine turtle; gene evolution; conservation genomics; genetic diversity; demographic history

Abbreviations: TE - transposable element; RE - repetitive element; RRC - region of reduced collinearity; FP – Fibropapillomatosis; ROH – runs of homozygosity

[✉]Corresponding author: Blair P. Bentley, University of Massachusetts, Amherst USA. Email: bbentley@umass.edu.
* These authors contributed equally to the work: Camila J. Mazzoni and Lisa M. Komoroske.

Abstract

Sea turtles represent an ancient lineage of marine vertebrates that evolved from terrestrial ancestors over 100 MYA, yet the genomic basis of the unique physiological and ecological traits enabling these species to thrive in diverse marine habitats remain largely unknown. Additionally, many populations have declined drastically due to anthropogenic activities over the past two centuries, and their recovery is a high global conservation priority. We generated and analyzed high-quality reference genomes for green (*Chelonia mydas*) and leatherback (*Dermochelys coriacea*) turtles, representing the two extant sea turtle families (MRCA ~60 MYA). These genomes are highly syntenic and homologous, but localized non-collinearity was associated with higher copy numbers of immune, zinc-finger, or olfactory receptor (OR) genes in green turtles, and ORs related to waterborne odorants were greatly expanded in green turtles. These findings suggest that divergent evolution of these key gene families may underlie immunological and sensory adaptations assisting navigation, occupancy of neritic versus pelagic environments, and diet specialization. Reduced collinearity was especially prevalent in microchromosomes, with greater gene content, heterozygosity, and genetic distances between species, supporting their critical role in vertebrate evolutionary adaptation. Finally, diversity and demographic histories starkly contrasted between species, indicating that leatherback turtles have had a low yet stable effective population size and extremely low diversity compared to other reptiles, and a higher proportion of deleterious variants, reinforcing concern over their persistence under future climate scenarios. These genomes provide invaluable resources for advancing our understanding of evolution and conservation best practices in an imperiled vertebrate lineage.

Statement of significance

Sea turtles represent a clade whose populations have undergone recent global declines. We analyzed de novo genomes for both extant sea turtle families through the Vertebrate Genomes Project to inform their conservation and evolutionary biology. The highly conserved genomes were largely differentiated by localized gene-rich regions of divergence, particularly in microchromosomes, suggesting that these overlooked genomic elements may play key functional roles in sea turtle evolution. We further demonstrate that dissimilar evolutionary histories impact standing genomic diversity and genetic load, and are critical to consider when using these metrics to assess adaptive potential and extinction risk. Examination of these relationships may be important to reveal drivers of adaptation and diversity in sea turtles and other vertebrates with conserved genome synteny.

Introduction

Sea turtles recolonized marine environments over 100 MYA (1, 2) and are now one of the most widely distributed vertebrate groups on the planet (3). Leatherback turtles (*Dermochelys coriacea*) represent the only remaining species of the family Dermochelyidae, which diverged from the Cheloniidae (hard-shelled sea turtles) about 60 MYA (4). Unique morphological (Fig. 1a) and physiological traits allow leatherback turtles to exploit cool, highly productive pelagic habitats (5, 6), while green turtles (*Chelonia mydas*) and other hard-shelled chelonid species largely inhabit warmer nearshore habitats following an early pelagic life stage. Most previous research in this group has focused on organismal and ecological adaptations (7), but the genomic basis of traits that differentiate or unite these species is not well understood.

Anthropogenic pressures have caused substantial population declines in sea turtles, with contemporary populations currently representing mere fractions of their historical abundances (8, 9). Although sea turtles spend most of their life in the ocean, they also exhibit long-distance migrations to natal rookeries for terrestrial reproduction (7, 10, 11). Consequently, they are threatened by human activities in both terrestrial and marine environments, including direct harvest of meat and eggs (12), fisheries bycatch (13), coastal development (14, 15), pollution (16), disease (17), and climate change (18, 19), which is exacerbated by their temperature-dependent mechanism of sex determination (TSD) altering population dynamics (20, 21). The IUCN lists most sea turtle species as vulnerable or endangered, and while decades of conservation efforts have fueled positive trends for some populations (22), others continue to decline (23). In particular, leatherback turtles have undergone extensive declines (>95% in some populations) over the last century (24–27), including the extirpation of the Malaysian nesting population (28). Leatherback turtle recovery is also impeded by relatively low hatching success compared to other sea turtle species (29). In contrast, many green turtle populations have recently increased following conservation actions (22), but their continued recovery remains threatened by anthropogenic activities and high incidence of the neoplastic disease fibropapillomatosis (FP) (30).

Genomic data have been instrumental in advancing understanding of species' evolutionary histories and ecological adaptations (31–33), and providing critical information for conservation management (34–37). However, this research has been hampered in taxa where genomic resources remain limited. In particular, the lack of high-quality reference genomes, which are essential for accurate comparative evolutionary analyses (38, 39) and estimates of a wide range of metrics to inform conservation biology (36, 40), impede this work in many threatened species. An early draft genome for the green turtle was assembled almost a decade ago (41), and provided important insights into turtle evolution. However, errors, gaps, and misassemblies in draft genomes can lead to spurious inferences, potentially masking signals of interest (38, 42). Well-annotated, chromosomal-level reference genomes can resolve these issues, improving our understanding of genomic underpinnings of ecological and evolutionary adaptations (39, 43). For example, high-quality genomes with accurate annotations have enabled examination of gene changes associated with recolonization of the marine environment by terrestrial vertebrates, including the loss of olfactory receptor (OR) gene families (32, 44). Comparative genomic analyses also demonstrated adaptive diversity in genes underlying reptilian immunity (45), with high-quality genomes providing insights into disease susceptibility (33, 46, 47). This is critical for sea turtles, with diseases such as FP adversely impacting populations across the globe (30), and information on immune genes key for devising effective conservation strategies (48). The contiguity of high-quality genomes is also invaluable for conservation-focused analyses, especially runs of homozygosity (ROH)

and genetic load that provide insights into population demography and inbreeding depression, and are difficult to accurately quantify with fragmented, poorly-annotated genomes (49).

Here, we assembled chromosome-level reference genomes with species-specific annotations for leatherback and green turtles as part of the Vertebrate Genomes Project (VGP) to facilitate critical research centered around sea turtle evolutionary history and conservation. We conducted comparative analyses to explore the genomic basis of their shared and unique phenotypic traits, and compared their chromosomal organization, genomic diversity, and demographic histories. These genomes represent two of the most contiguous reptilian genomes assembled to date, providing invaluable resources for ongoing investigations into conservation and adaptation for this imperiled vertebrate lineage.

Results

Genome quality

The reference genomes of the leatherback and green turtles were generated using four genomic technologies following the VGP pipeline v1.6 (39), with minor modifications (see Methods). A total of 100% of the leatherback and 99.8% of the green turtle assembled sequences were placeable within chromosomes. The assembled genomes were near full-length (~2.1 GB), with annotations of all 28 known chromosomes for both species, composed of 11 macrochromosomes (>50 Mb) and 17 microchromosomes (<50 Mb) (Tables 1 & S1, Fig. S1). These genomes are among the highest quality genomes assembled for non-avian reptiles to date in terms of both contiguity and completeness (Table S2), with the leatherback turtle assembly representing the first reptile genome where all scaffolds have been assigned to chromosomes. Scaffold N50s were high for both genomes (Table 1). We annotated 18,775 protein-coding genes in leatherback turtle genome and 19,752 in the green turtle (see below for analysis of gene number differences). Of these, 96.9% and 97.5%, for leatherback and green turtles respectively, were supported over 95% of their length from experimental evidence and/or high-quality protein models from related species (see Methods). The number of protein-coding genes falls within the range for other reptilian genomes (Table S2) and includes 97.7% and 98.2% complete BUSCO copies when using Sauropsida models for leatherback and green turtles respectively (50), which are similar or higher proportions than all other assembled reptilian genomes to date (Fig. S2).

Genome architecture

Despite diverging over 60 million years ago (4), leatherback and green turtles have extremely high genome synteny and collinearity (Figs. 1b,c, S3, S4). After multiple rounds of manual curation to correct any artifacts of misassemblies, only a few larger structural rearrangements remained, including inversions of up to 7 Mb on chromosomes 12, 13, 24 and 28 (Fig. S3). The high collinearity between the two genomes included near complete end-to-end contiguous synteny for nine out of 28 chromosomes (Fig. S3). The remaining 19 chromosomes exhibited at least one small region of reduced collinearity (RRC) between the species, with RRCs representing a total of ~83.4 Mb (~3.9%) and ~110.5 Mb (~5.2%) of the leatherback and green turtle genome lengths, respectively. Eight chromosomes exhibited small RRCs (between 0.1–3 Mb), and 11 contained RRCs that were between 3–18 Mb in length (Figs. 2a-d & Table S3). Analyses of coding regions revealed a similar pattern of high collinearity between the two species at the gene level (Figs. 1c & S3), particularly within the macrochromosomes that contain more than 80% of the total length of the genomes.

The two genomes displayed similar percentages of repetitive elements (REs; 45.8% and 44.4%, respectively; Fig. S5 & Table S4), which were almost exclusively transposable elements (TEs; 30.5% and

27.4%) and unclassified repeats (14.6% and 16.5%, respectively). While both genomes carry similar proportions of REs, the leatherback turtle genome exhibited relatively longer TEs across all but two chromosomes when compared to the green turtle (Fig. S6a). The landscape of TE superfamily composition over evolutionary time is generally similar between the two species (Fig. S5), and consistent with other reptiles (51, 52). One striking difference, however, is seen in REs with low Kimura values (<5%), which appeared at much higher frequency in the leatherback turtle genome (Fig. S5), representing either relatively recent insertions or reflecting lower mutation rates in this species.

Gene families and gene functional analysis

Gene function analysis of localized RRCs revealed that most contain genes with higher copy numbers in the green turtle compared to the leatherback (Fig. 2a-d, Table S3). From the RRCs in the 19 chromosomes that had higher gene copy numbers in the green turtle, ten contained genes associated with immune system, olfactory reception and/or zinc-finger protein genes. In addition to localized RRCs, higher gene copy numbers in the green turtle occurred in many gene orthologous groups (orthogroups) across the entire genome, and generally in variable multicopy genes (Fig. 2f, g). Copy number variation accounted for most of the nearly one thousand more genes annotated in the green turtle genome relative to the leatherback (Fig. 2f, g; Table 1). We detected no evidence of collapsed multicopy genes in the leatherback turtle assembly across multiple analyses (see Methods), supporting this as a biological signal rather than technical artifact of assembly.

Olfactory receptors (ORs) represented the largest orthogroups in both genomes, and differences in copy numbers were tightly connected to observed RRCs. All OR Class I genes were clustered at the beginning of chromosome 1, and the green turtle had higher copy numbers in this region (Fig. 2a-d). This area also contains a cluster of OR Class I genes in at least three additional testudinid species, and is the only divergent region across the very large chromosome 1 in the turtles analyzed (Fig. S7). In contrast, OR Class II genes were spread across several chromosomes in both sea turtle species, but again higher copy numbers in the green turtle were all found within RRCs (Fig. 2b-d). The instability and rapid evolution of OR gene numbers in turtles is further illustrated in the expansion-contraction analysis of orthogroups (Fig. 2e, Table S6a-d), which showed that OR Class I genes underwent a modest contraction in the ancestral sea turtle lineage, followed by an expansion in the green turtle but a further contraction in the leatherback turtle. Similar trends were detected for OR Class II genes, but with a greater magnitude of contraction in the ancestral sea turtle lineage followed by a further contraction for the leatherback turtle and only a small expansion for the green turtle.

A second important RRC encompassed the major histocompatibility complex (MHC), which plays a critical role in vertebrate immunity and has particularly strong relevance to sea turtle conservation due to the threat of FP and other diseases (32). In addition to the MHC genes, this RRC (RRC14) includes several copies of OR Class II genes, zinc-finger protein genes and other genes involved with immunity, such as butyrophilin subfamily members and killer cell lectin-like receptors (Fig. 2d, Table S12). Invariably, the green turtle carried higher numbers of all the multicopy genes present in RRC14. RRCs on other chromosomes also showed increased levels of zinc-finger protein genes in the green turtle, including the RRCs labeled 6A, 11A, 14A, and 28 (Table S3). In particular, zinc-finger protein genes were highly prevalent on chromosomes 14 and 28 in both sea turtles, representing more than 50% of all the protein domains present on these chromosomes (Fig. S8).

Finally, given the critical importance of understanding the underlying mechanism of temperature-dependent sex determination (TSD) in the face of climate change, we analyzed genes known to be

associated with TSD across reptiles. Almost all 216 genes previously implicated in male- or female-producing pathways in reptilian species with TSD were single-copy genes in both sea turtle species (Table S7; 210 genes per species). Only three genes (*MAP3K3*, *EP300*, and *HSPA8*) were duplicated in both genomes, with the copies located on different chromosomes in all cases. Moreover, homologous genes were generally located in the same region of the genomes for both species (Fig. S9), and missing genes were typically absent in both species, with only four genes found in one species but not the other (Table S7).

Macro and microchromosomes

Microchromosomes contained higher proportions of genes than macrochromosomes (Fig. 3a,b), and gene content was strongly positively correlated with GC content (Fig. S10). These patterns were particularly apparent in small (<20 Mb) microchromosomes, where GC content reached 50%, compared to the 43 - 44% genome-wide averages. Within chromosome groups, larger proportions of multicopy genes were generally associated with higher total gene counts, and chromosomes with the highest multicopy genes numbers have increased proportions of RRCs (Fig. 3a,b).

Mean genetic distances for single-copy regions between the two sea turtles were also higher in small microchromosomes (0.053) compared to both intermediate (>20 Mb) microchromosomes (0.047), and macrochromosomes (0.045) (Fig. 3c). However, examination of intermediate microchromosome and macrochromosome RRCs revealed elevated genetic distances in these regions that approached values observed in small microchromosomes (Table S8). Genetic distances were also positively correlated with heterozygosity, which was higher in small microchromosomes for both species (Figs. 3d & S11-13).

Genome diversity

Although both species displayed similar patterns of higher heterozygosity in microchromosomes than macrochromosomes (Figs. 3d & S11-13), they differed in the genome-wide nucleotide diversity level by almost an order of magnitude (repeat masked $\pi = 3.19 \times 10^{-4}$ leatherback and 22.2×10^{-4} green turtle; Fig. S12, Table S9). Exonic regions exhibited lower levels of heterozygosity than non-coding regions (Fig. 4a, Table S9), with a greater reduction in heterozygosity within green turtle exons (~20%) than leatherback turtle exons (~10%) compared to genome-wide levels. In addition, the percentage of 100 Kb windows containing zero heterozygous sites was higher in the green turtle (6.60%) than the leatherback turtle (2.87%), suggesting that although diversity was lower overall in the leatherback turtle, it was more evenly spread across the genome than in the green turtle. Using a standardized heterozygosity pipeline (see Methods; Fig. 4b), we found the genomic diversity of the leatherback turtle was substantially lower than almost all other reptiles examined, including *Chelonoidis abingdonii*, where low diversity has been considered a contributing factor to their extinction (53), and only the critically endangered Chinese alligator (*Alligator sinensis*) showed lower diversity (54). In contrast, the genomic diversity of the green turtle fell in the mid-range for reptiles, as well as similar analyses conducted on mammals (55, 56). Finally, we identified high-diversity exonic regions using multiple approaches (see Methods), and found that many contained immune, OR, and zinc-finger protein genes in both species, but especially for the green turtle which showed a greater number of high-diversity windows (Fig 4c; Table S10). Given the striking similarity to the RRC analysis results above, these findings independently reinforced the importance of these gene families in the divergent evolution of these species.

Runs of homozygosity (ROH)

The leatherback turtle had a greater number of ROHs (>100 Kb) compared to the green turtle (N_{ROH} = 2,045 and 873, respectively), as well as higher accumulated length and proportion of the genome in ROH (S_{ROH} = 400.61 Mb (18.51% of genome) and 327.06 Mb (15.53%), respectively). The average length of ROHs was generally shorter in the leatherback turtle (L_{ROH} = 196 Kb and 375 Kb for the leatherback and green turtles, respectively; Fig. 4d), with the accumulated length of short (<500 Kb) ROHs highest when compared to medium (500 Kb-1 Mb) and long ROHs (>1 Mb) (Fig. 4d). The leatherback turtle genome only showed one ROH that was greater than 1 Mb in length, suggesting that recent bottlenecks or inbreeding are unlikely, rather that this species has maintained long-term low diversity. In contrast, the green turtle had 54 ROHs longer than 1 Mb, suggestive of a possible more recent population bottleneck or inbreeding events. The average lengths of ROHs were also higher in macrochromosomes than microchromosomes (Fig. S15).

Genetic load

Coding region variants of the leatherback turtle genome were found more likely to be impactful, with 0.10% and 0.07% of variants predicted to cause 'high impacts' (e.g., stop-codon gain or loss) for the leatherback and green turtles, respectively, and with 'moderately' and 'low' impact variants also higher in the leatherback turtle (Fig. 4e). Additionally, the missense to silent mutation ratio was higher in the leatherback (0.89) than green turtle (0.65), again indicating that genetic load is higher for the leatherback turtle. High-impact variants predicted by snpEff only occurred in one species for any given gene. The 103 and 357 nucleotide variants characterized as 'high' impact in the leatherback and green turtle were found within 59 and 171 unique genes, respectively. The functions of these genes were variable (Fig. S16). For the leatherback turtle, many of the genes impacted were linked to cell transport and demethylation, with DEAH-box helicase 40 and a *Hsp40* family member also impacted (Table S13). In contrast, for the green turtle, many of the genes were linked to immunity, including an MHC class I alpha chain gene, as well as B-cell receptors and killer-cell receptors. The green turtle also showed putative high impact variants within several OR genes.

Demographic history

Pairwise Sequential Markovian Coalescence (PSMC) analyses indicated different historical effective population sizes (N_e) between the two sea turtle species (Fig. 4f). The results indicate that the N_e for the leatherback turtle has been relatively small and sustained, ranging in size from approximately 2,000 to 21,000 over the last 10 million years, and at the lower end of this range for the last 5 million years. In comparison, the green turtle has experienced wider population fluctuations and a relatively higher overall N_e suggesting that N_e has fluctuated between approximately 44,000 and 83,000. While the N_e for the leatherback turtle is relatively low, it showed signs of increasing abundance prior to the Eemian warming period (Fig. 4f [H]), with a subsequent decrease during this period until the last glacial maximum (LGM). In contrast, the green turtle had three distinct peaks in N_e (Fig. 4f), potentially associated with ocean connectivity changes related to the closure of the Tethys Sea [A], the Pleistocene period [B], and a more pronounced peak that aligns with later marked temperature fluctuations [C]. We observed similar patterns for PSMC analyses conducted on additional individuals for both species (see Supplementary Appendix I).

Discussion

Divergence in localized RRCs and microchromosomes amidst high global genome synteny. The lineages leading to leatherback and green turtles diverged over 60 MYA (4), giving rise to species that are adapted to dissimilar habitats, diets, and modes of life. Despite high overall levels of genome synteny across both the macro- and microchromosomes between the sea turtle families, RRCs and small microchromosomes were particularly associated with high concentrations of multicopy gene families, as well as heightened genomic diversity and genetic distances between species, suggesting that these genomic elements may be important sources of variation underlying phenotypic differentiation. Though our results here do not demonstrate direct causality, we have identified candidate regions and gene families that can be targeted in further studies quantifying evidence for positive selection and their roles in sea turtle adaptation and speciation.

The high global stability of both macro- and microchromosomes between sea turtle families also aligns with recent work showing similar patterns across reptiles including birds, emphasizing the important roles of microchromosomes in vertebrate evolution (57). However, it is not yet clear if the characteristics of microchromosomes and RRCs we observed are unique to sea turtles, or, more likely, commonly observed in other vertebrates. Our detailed analyses of RRCs, microchromosomes, and their associated genes were only possible due to the high-quality of the assembled sea turtle genomes because these analyses can be sensitive to genome fragmentation and misassemblies (39). The prevalence or importance of such localized genomic differentiation among other closely or more distantly related reptiles or other vertebrate groups has not been evaluated due to a lack of equivalent genomic resources, but this is rapidly changing. As chromosomal-level genomes across all vertebrate lineages soon become available, our work provides a roadmap for identifying genomic regions harboring contrasting expansion/contractions of gene families and diversity levels in different vertebrate lineages. For taxa with highly conserved genomes like sea turtles, analyses of RRCs and microchromosomes are likely important to understand their divergent evolutionary histories and the phenotypic connections of the genes within them.

Contrasting sensory and immune gene evolution between sea turtle families. Sea turtles have complex sensory systems and can detect both volatile and water-soluble odorants, which are imperative for migration, reproduction and identification of prey, conspecifics, and predators (58–62). However, leatherback and green turtles occupy dissimilar ecological niches that depend on different sensory cues. While leatherback turtles inhabit the pelagic environment their entire lives post-hatching, performing large horizontal and vertical migrations to seek out prey patches of jellyfish and ctenophores (63), green turtles recruit as juveniles to neritic coastal and estuarine habitats and can have highly variable diets (64, 65). Substantial differences have been detected in the morphology of sea turtle nasal cavities, with leatherback turtle cavities relatively shorter, wider, and more voluminous than chelonids (66–68), suggesting reduced requirements for olfactory reception. OR genes encode proteins used to detect chemical cues, with the number of OR genes present in a species' genome strongly correlated to the number of odorants that it can detect (69), and linked to the chemical complexity of its environment (70). The two major groups of ORs in amniote vertebrates are separated by their affinities with hydrophilic molecules (Class I) or hydrophobic molecules (Class II) (71). Class I OR genes may be particularly important in aquatic adaptation (32), and expansions of Class I ORs in testudines, including green turtles, have been previously reported, although with some uncertainty due to the use of short-read assemblies

(32, 41, 72). Our reconstruction of both Class I and Class II OR gene evolution throughout the sea turtle lineage revealed that after ancestral contractions, gene copy evolution diverged in opposite directions between the sea turtle families. The greater loss of Class II compared to Class I ORs in the ancestral sea turtle lineage likely reflects relaxed selection for detection of airborne odorants, as has been observed in other lineages that recolonized marine environments, including marine mammals (73). However, as sea turtles continue to use terrestrial habitats for reproduction, they need to retain some of these capabilities, which could explain why the contraction was weaker than observed in fully marine species (e.g., the vaquita *Phocaena sinus*; Fig. 2e).

The strong Class I OR expansion in the green turtle may be related to its distribution in complex neritic habitats and variable diet, requiring detection of a high diversity of waterborne odorants, while the continued loss of ORs in the leatherback turtle could be a consequence of its more specialized diet and lower complexity of pelagic habitats. Although leatherback turtles can detect jellyfish chemical cues, sensory experiments have indicated that visual cues are more important for food recognition in this species (74). Additionally, while the precise mechanisms underpinning philopatry in sea turtles still remain unclear, green turtles are thought to use olfactory cues to reach natal nesting beaches following long-distance navigation guided by magnetoreception (60, 62). Leatherback turtles exhibit more 'straying' from natal rookeries than other species, and such relaxed philopatry may be related to reduced capabilities or reliance on olfactory cues to hone in on specific beaches.

The diversity of the highly-complex MHC region is a key component in the vertebrate immune response to novel pathogens, with greater gene copy numbers and heterozygosity linked to lower disease susceptibility (75). While both sea turtle species contained most of the core MHC-related genes, the green turtle had more copies of genes involved in adaptive as well as innate immunity. Pathogen prevalence and persistence is often greater in neritic habitats than open ocean habitats (76), so green turtles may be exposed to higher pathogen loads and diversity than leatherback turtles (77). However, reptilian immune systems are understudied compared to other vertebrates, and very few studies of MHC genes have been conducted in turtles (78). Thus, it is not yet understood how immune gene diversity translates into disease susceptibility or ecological adaptation in sea turtles, which is particularly critical for their conservation as FP continues to threaten the recovery of populations around the globe (30). Although this viral-mediated tumor disease occurs in all sea turtle species, there is high variation between species and populations in disease prevalence and recovery, making it plausible that harboring certain genes, copy numbers, or specific alleles may play important roles in disease dynamics. Despite decades of research there have been no studies of the immunogenomic factors governing FP susceptibility or resilience, in part due to difficulty in accurately quantifying hypervariable and complex MHC loci with short-read sequencing technologies (79). Our reference genomes now enable studies accurately interrogating MHC and other immune loci to close this critical research gap, and advance our fundamental understanding of immune gene evolution in sea turtles.

Conservation of reproductive genes and repetitive elements. In contrast to olfactory and immune genes, almost all genes with *a priori* linkages to TSD pathways (80–82) occurred as single copy orthologs with highly conserved chromosomal locations between the two species. This is likely indicative of strong selection for conservation of this reproductive pathway, but our understanding of the specific roles these genes play in sea turtle TSD remains limited. Resolving whether inter- (83) and intra-specific (84) variations in thermal thresholds are due to the few genes that diverged from the general pattern we observed, functional sequence variation between orthologs, or other factors (e.g., epigenetic processes) is

of high conservation concern for sea turtles (85), as climate warming is expected to skew sex ratios and alter population demographics (86) in the absence of substantial plasticity or adaptation. Our results serve as the foundation for these much-needed studies to quantify genomic mechanisms of TSD in sea turtles and determine their adaptive capacity to persist under climate change.

While REs in turtles have been investigated for over 30 years (87, 88), few studies have directly addressed the distribution and diversity of REs within testudine genomes (89). Both sea turtle genomes have substantially larger RE compositions (>40%) than previous estimates for other turtle species (41, 89, 90), including the draft genome of the green turtle (10% of the genome (41)). Interestingly, more recent reptile genome assemblies show higher proportions of REs (90, 91), with results similar to our estimates. The benefits of whole-genome approaches are further highlighted in the tuatara, where initial RE estimates suggested <10% of the genome was composed of REs (92), yet a subsequent whole-genome assembly increased this estimate to 64% (45). Collectively, these results support the notion that RE patterns could be more conserved across non-avian reptiles than previously believed, and the continued application of recent advances in genome sequencing, assembly methods, and analyses are needed to better understand the RE patterns and the processes that generate them (39, 43).

Differential genomic diversity and demographic histories. Genomic diversity is a critical metric for evaluating extinction risk and adaptive potential to environmental perturbation (93–95), with heterozygosity positively correlated with individual fitness (see reviews by (96, 97). Understanding the causes and consequences of genomic diversity is imperative for sea turtles, and for leatherback turtles in particular, where contemporary populations have experienced recent sharp declines due to human activities (25). The leatherback turtle genome exhibited exceptionally low diversity relative to the green turtle, other reptiles and mammals, broadly aligning with previous estimates (98, 99). However, factors influencing genomic diversity can vary among species (100), and our PSMC and ROH results indicate that low diversity in the leatherback turtle is likely a consequence of long-term low effective population sizes rather than recent population reductions. This is consistent with mitochondrial analyses suggesting that contemporary populations radiated from a small number of matriarchal lineages within a single refugium following the Pleistocene (99). The low, relatively evenly spread heterozygosity is also congruent with sustained low population sizes with frequent outbreeding similar to that observed in several mammal species (101, 102). Also encouragingly, although the reference genome was generated from an individual from the West Pacific leatherback turtle population that has suffered precipitous declines (103), we did not detect patterns consistent with recent inbreeding. This suggests that if ongoing anthropogenic threats are mitigated, the population may still be large enough to avoid complications arising through inbreeding depression during recovery. However, the possibility that population declines have occurred too rapidly for the impacts of inbreeding to yet be detected warrants cautious optimism and the need for continued genomic monitoring of the population. In contrast, the higher genomic diversity with some long ROHs observed in the green turtle likely reflects their radiation from many refugia (104), as well as relatively recent inbreeding events. This is potentially because the green turtle genome was generated from an individual from the Mediterranean population where nesting populations are relatively small (105) which, combined with strong natal philopatry (106), may increase the chance of inbreeding.

Regardless of the causes of current genomic diversity levels in sea turtles, the amount of standing variation may have important implications for their future persistence (107), especially given the adaptive capacity likely required to keep pace with rapid anthropogenic global change. With extremely low genomic diversity and a higher genetic load compared to the green turtle, the risks are presumably of

greater concern for leatherback turtles. Additionally, leatherback turtles have substantially lower hatching success compared to other sea turtle species (29) that is potentially related to the heightened genetic load and low heterozygosity (108, 109), and may combine with other factors to slow population recoveries following conservation measures. However, recent studies have documented low genome diversity in a number of species with wide geographic distributions and relatively large census population sizes, including some long-lived marine vertebrates (101, 110–113). Other species with low diversity have rebounded following population declines and/or appear to have purged deleterious alleles through long-term low population sizes (111, 114, 115), thereby limiting the impact of low genomic diversity on viability (56, 111, 116). Although our results of a greater genetic load despite long-term low N_e suggest this is not the scenario for leatherback turtles, further assessments of more individuals over greater spatial and temporal are needed. Studies enabled by the reference genomes presented here quantifying diversity and genetic load within and among global populations will clarify these relationships for leatherback turtles and other sea turtle species to guide conservation recommendations.

In contrast to the long-term low N_e of leatherback turtles, our demographic reconstructions showed the N_e of the green turtle has fluctuated widely over the same period. These fluctuations appear correlated with climatic events, beginning with the closure of the Tethys Sea, which altered ocean connectivity and represented a period of increasing temperatures that may have opened more suitable habitat. As temperatures subsequently decreased, N_e also decreased, however temperature fluctuations during the Pleistocene were associated with an additional increase in N_e, and a subsequent increase in N_e was associated with warmer temperatures following the Eemian period. While warmer temperatures presumably allowed for larger population sizes of green turtles, spikes in N_e (e.g., ~100KYA) are most likely associated with mixing of previously isolated populations due to warm-water corridors allowing movement between populations and ocean basins (117). The lower, long-term N_e of leatherback turtles may reflect a reduced census size associated with this species' greater mass and trophic position. Following an initial decrease in N_e associated with declining temperatures, the N_e of leatherback turtles remained relatively constant throughout the fluctuations of the Pleistocene. Although reptiles are generally sensitive to climatic thermal fluctuations, leatherback turtles exhibit unique physiological adaptations that produce regional endothermy and facilitate exploitation of cold-water habitats (6) that potentially led them to being less susceptible to periods of cooler temperatures. While our overall estimates and trends for both species were broadly concordant with previous studies (99, 118, 119), a recent study using MSMC found steep declines in N_e for green turtles >100,000 years before present (119), which was not detected in our PSMC analyses. Since this decline was also not detected in a prior study using PSMC on the draft the green turtle genome (118), this is likely a consequence of the different methods, with MSMC analyses inferring large N_e for more ancient time scales (120).

Enabling future research and conservation applications. In addition to the insights reported here, the reference genomes for both extant sea turtle families provide invaluable resources to enable a wide breadth of previously unattainable fundamental and applied research. Combined with other forthcoming chromosomal-level vertebrate genomes, in-depth comparative genomics analyses can further investigate ecological adaptation related to immune and sensory gene evolution, as well as the genomic basis for traits of interest such as adaptation to saltwater, diving capacity, and long-distance natal homing. Studies leveraging these reference genomes alongside whole-genome sequencing of archival sample collections can assess how genomic erosion, inbreeding and mutational load are linked to population size,

trajectories, and conservation measures in global sea turtle populations. For instance, the fact that leatherback turtles have persisted with low diversity and N_e for long time periods offers hope for their recovery, but given that some populations have now been reduced to only a few hundred individuals (103), research quantifying purging of deleterious alleles, inbreeding depression and adaptive capacity within populations is urgently needed (121). Many conservation applications that may not require whole-genome data can also benefit from the utility of these reference genomes, including the development of amplicon panels and molecular assays to investigate TSD mechanisms and adaptive capacity under climate change, and assessing linkages between immune genes and disease risk. Finally, with global distributions and long-distance migratory connectivity, sea turtle conservation requires international collaboration that has been previously hampered by difficulty comparing datasets between laboratories. Existing anonymous markers can now be anchored to these genomes, and new ones can be optimized for conservation-focused questions and shared across the global research community, facilitating large-scale syntheses and equitable capacity building for genomics research. While ongoing anthropogenic impacts continue to threaten the viability of sea turtles to persist over the coming century, combined with the important work of reducing major threats such as fisheries bycatch and habitat loss, these genomes will enable research that make critical contributions to recovering imperiled populations.

Methods

Sample collection, genome assembly and annotation

Blood was collected from leatherback and green turtles using minimally invasive techniques for isolation of ultra-high molecular weight DNA, and tissue samples of internal organs for RNA were collected opportunistically from recently deceased or euthanized animals. Full details of sample collection, storage, and laboratory processing prior to sequencing can be found in Supplementary Appendix I. Resulting raw data were deposited into the VGP Genome Ark and NCBI Short-Read Archive (SRA) (see Data Accessibility Statement). We assembled both genomes using four genomic technologies following the VGP pipeline v1.6 (39) with a few modifications detailed in Supplementary Appendix I. Briefly, PacBio Continuous Long Reads were assembled into haplotype phased contigs, with contigs scaffolded into chromosome-level super scaffolds using a combination of 10X Genomics linked reads, Bionano Genomics optical maps, and Arima Genomics Hi-C 3D chromosomal interaction linked reads. Base call errors were corrected to achieve high quality (>Q40). The assemblies were manually curated, with structural errors corrected according to the Hi-C maps (Fig. S1), and the 28 super scaffolds (hereinafter referred to as chromosomes) numbered in both species according to sequence lengths in the leatherback turtle assembly, and synteny between the two species. A manual inspection comparing the sequence collinearity between the first curated versions of the genomes revealed a small number of artefactual sequence rearrangements that were corrected in a second round of manual curation (see Supplementary Appendix I).

To enable accurate, species-specific annotations for each genome, both short and long-read transcriptomic data (RNA-Seq and Iso-Seq) were generated from tissues known for their high transcript diversity in each species. These data, plus homology-based mapping from other species, were used to annotate the genomes using the standardized NCBI pipeline (122). Briefly, we performed annotation as previously described (39, 123), using the same RNA-Seq, Iso-Seq, and protein input evidence for the prediction of genes in the leatherback and green turtles. We aligned 3.5 billion RNA-Seq reads from eight green turtle tissues (blood, brain, gonads, heart, kidney, lung, spleen and thymus) and 427 million reads from four leatherback turtle tissues (blood, brain, lung and ovary) to both genomes, in addition to 144,000

leatherback turtle and 1.9 million green turtle PacBio IsoSeq reads, and all Sauropsida and *Xenopus* GenBank proteins, known RefSeq Sauropsida, *Xenopus*, and human RefSeq proteins, and RefSeq model proteins for *Gopherus evgoodei* and *Mauremys reevesii*.

Genome quality analysis

We used the pipeline assembly-stats from https://github.com/sanger-pathogens/assembly-stats to estimate the scaffolds N50, size distributions and assembly size. BUSCO analysis (115) and QV value estimations (116) were conducted to assess the overall completion, duplication, and relative quality of the assemblies. We used D-GENIES (118) with default parameters to conduct dot plot mapping of the entire genomes and each individual chromosomes to evaluate the synteny between leatherback and green turtle genomes, and Haibao Tang JCVI utility libraries following the MCScan pipeline (119) to verify the contiguity of the genomes. Incongruences in gene synteny blocks were manually investigated using Artemis Comparative Tool (120), identifying possible regions of inversion that could be caused by artifacts during assembly. These regions were then identified and corrected in the latest version of the assembly for both species. Only a few structural rearrangements between the two species remained after two rounds of manual curation with support of sequencing data. The final curated assemblies were analyzed using the Genome Evaluation Pipeline (https://git.imp.fu-berlin.de/cmazzoni/GEP) to obtain all final QC plots and summary statistics.

Identification and analysis of RRCs and REs

Leatherback and green turtle genomes were mapped to each other using Minimap2 with a dot plot of the mapping generated using D-GENIES (124). Using windows of 20 Mb, the dot plot was screened visually with regions larger than 1 Mb showing reduced collinearity (i.e., one or more breaks in the diagonal indicating homology), as well as smaller regions with obvious signals of genomic rearrangements (e.g., inversions), cataloged as regions of reduced collinearity (RRCs). Several genomic features were examined within these regions and compared to regions of the same length directly up- and down-stream of the RRCs (Table S3). We identified the functions of the genes present in RRCs using genome annotations and identified protein domains using Interproscan (125). The proportion of GO terms in each chromosome was estimated for each species using PANTHER (126); Fig. S16). To examine if RRCs presented differential patterns of sequence and/or gene duplication between the species, we aligned the genomes of the sea turtles against each other using Progressive Cactus (127, 128), and all homologous genes that presented more than one copy for one of the two species were isolated using an inhouse script (*IdentifyDupsReciprocalBlast.sh*) to retrieve duplicated genes (see Supplementary Appendix I for further details on Cactus alignments). Repetitive elements (REs) were identified by creating a *de novo* database of transposable elements using RepeatModeller2 (129), followed by running RepeatMasker (130, 131) to calculate Kimura values for all REs (see full analysis details in Supplementary Appendix I).

Gene families and gene functional analysis

To estimate the timing of gene family evolution for the OR gene families on sea turtles we used Computational Analysis of gene Family Evolution v5 (132). Briefly, CAFE5 uses phylogenomics and gene family sizes to identify gene family expansions and contractions, we used a dataset containing 8 species of turtle, 4 non-turtle reptiles, 3 mammals and 1 amphibian using OrthoFinder (133, 134). OR orthogroups were grouped based on subfamily (Class I and Class II; see (72)), and an ultrametric

phylogeny was generated by gathering 1:1 orthologs. We then aligned amino acid sequences for each orthogroup and generated a phylogenetic tree (see Supplementary Appendix I for details).

Like many reptile species, sea turtles possess TSD. We compiled a list of 217 genes that have been implicated in TSD in reptiles (see Table S7). To determine if these genes were present in our assembled genomes, we employed two methods of investigation. We firstly searched the genome annotations for gene identifiers and protein names, followed by a BLAST search of homologous sequences to account for variations in gene identifiers between taxonomic groups (see Supplementary Appendix I for details). Resultant locations on both genomes were plotted on a Circos plot CIRCA (http://omgenomics.com/circa).

To identify genes related to immunity, and the MHC in particular, we searched the genome for the list of core MHC genes provided in Gemmell et al. (2020) (45). Genes were searched for in a similar way to the method used for the TSD genes, with initial searches of gene identifications, followed by a search of protein identifiers. As genes associated with the MHC are diverse, and vary substantially among species, we did not use a BLAST search for these genes. Locations of the genes were then compared between species to determine which genes were annotated, and where the core MHC region is located within the genomes.

Genetic distance, genome diversity, runs of homozygosity, and historical demography

In order to estimate the genetic distance between the leatherback and green turtle genomes, we used the halSnps pipeline (135) to compute interspecific single variants based on genome alignments obtained with Progressive Cactus (127, 128) using the leatherback turtle genome as the reference. Genetic distances were calculated for windows across the genome where each window included exactly 10,000 positions presenting single alignments against the green turtle genome in the Cactus output. Positions with zero, or more than one alignment were ignored, and if this occurred over more than 50% of a given window, it was skipped entirely (i.e., each window analyzed covered between 10 and 20 Kb of the genome). Interspecific distances per bp were calculated by dividing the number of variants found within a window by 10,000.

We calculated genome-wide heterozygosity using a method adapted from Robinson et al. (2019) (102). Briefly, we used the Genome Analysis Toolkit (GATK) (136) to call genotypes at every site across the genome using the 10X reads sourced from the reference individual mapped back to the reference genomes using BWA-mem (137). Heterozygosity was calculated within 100 Kb non-overlapping windows, with only sites that had a depth of between ⅓× and 2× mean coverage retained for genotype scoring. Heterozygosity was calculated within these windows for (1) the entire genome, (2) the genome with repeat-regions masked, (3) only exon regions, (4) and for regions that were classified as ‘non-exons.’ We also adapted this pipeline to generate genome-wide heterozygosity for a number of additional reptilian and outgroup species with sequences sourced from the NCBI SRA where species-specific reference genomes were available (see details in Supplementary Appendix I).

ROHs were calculated by initially generating an additional SNP-list using whole-genome re-sequenced information from five additional individuals for each species (Table S11). This SNP-list was generated through the Analysis of Next Generation Sequencing Data (ANGSD; (138) pipeline due to the low- to moderate-coverage of the additional samples (~2-13×). ANGSD was parameterized to output files that were configured for use as input for the ROH analysis incorporated in PLINK (139). ROHs were then further characterized as ‘short’ (100-500 Kb), ‘medium’ (500Kb-1 Mb), or ‘long’ (>1 Mb) based on their length. ROHs for only the reference individuals are presented.

The alignments of the 10X reads for the reference individuals were also used as input for Pairwise Sequential Markovian Coalescence (PSMC; (140)) analysis of demographic history for both species. We used SAMtools (141) and BCFtools (142) to call genotypes with base and mapping quality filters of >Q30. We also filtered for insert size (50-5,000bp) and allele balance (AB) by retaining only biallelic sites with an AB of <0.25 and >0.75. We then ran PSMC analysis using the first 10 scaffolds, which constituted over 84% of the total length of the genome, using a generation time of 30 years (mid-way between reported generation times for both species; see Supplementary Appendix I), and a mutation rate of 1.2×10^{-8} (118).

Genetic load

Estimates of deleterious allele accumulation were conducted using the snpEff variant annotation software (143). We estimated the impacts of variants from coding regions using the species-specific genome annotations generated for both species, with a total of 18,775 genes for the leatherback turtle genome, and 19,752 genes for the green turtle genome used in the analysis. Variants were only included in the analyses if they met stringent quality requirements, with loci filtered during genotyping based on depth of coverage (⅓× - 2× mean coverage) and base quality metrics (Q < 20). The snpEff program predicts variant impacts and bins them into 'high', 'moderate', or 'low' impact categories, and outputs a list of genes that have predicted variant effects.

Acknowledgments

We are grateful for the assistance with the (1) leatherback turtle sample collection from the St. Croix Sea Turtle Program and the US Fish and Wildlife Service, the NOAA-SWFSC California in-water leatherback research team, and the New England Aquarium; (2) green turtle sample collection from the Israel National Sea Turtle Rescue Centre, the NOAA PIFSC-MTBAP team, and Thierry Work (USGS). We also thank Estefany Argueta and Jamie Stoll for assistance with literature searches for TSD and immune genes, and Phillip Morin, Andrew Foote, Anna Brüniche-Olsen, Annabel Beichman, Morgan McCarthy, David L. Adelson, and Yuanyuan Cheng for their invaluable discussions surrounding analysis approaches and comments on the manuscript. Sequencing of the green turtle has been performed by the Long Read Team of the DRESDEN-concept Genome Center, DFG NGS Competence Center, part of the Center for Molecular and Cellular Bioengineering (CMCB), Technische Universität Dresden and MPI-CBG.

Funding

Funding was provided by the University of Massachusetts Amherst, NSF-IOS (grant #1904439 to LMK), NOAA-Fisheries, Vertebrate Genomes Project, Rockefeller University, to EDJ, HHMI to EDJ, the Sanger Institute, Max-Planck-Gesellschaft, as well as grant contributions from Tom Gilbert, Paul Flicek, Robert Murphy, Karen A. Bjorndal, Alan B. Bolten, Ed Braun, Neil Gemmell, Tomas Marques-Bonet, and Alan Scott. We also acknowledge CONICYT-DAAD for scholarship support to TCV, and EKSR was supported by São Paulo Research Foundation - FAPESP (grant #2020/10372-6). BeGenDiv is partially funded by the German Federal Ministry of Education and Research (BMbF, Förderkennzeichen 033W034A). The work of FT-N and PM was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health. The work of MP was partially funded through the Federal Ministry of Education and Research (grant 01IS18026C). HP was supported by a Formació de Personal Investigador fellowship from Generalitat de Catalunya (FI_B100131). MK was

supported by “la Caixa” Foundation (ID 100010434), fellowship code LCF/BQ/PR19/11700002 and the Vienna Science and Technology Fund (WWTF) and the City of Vienna through project VRG20-001.

Data Accessibility Statement

Assemblies for both species have been deposited on NCBI GenBank. The NCBI GenBank accession numbers for the leatherback turtle genome assembly (rDerCor1) are GCF_009764565.3 and GCA_009762595.2 for the annotated primary and original alternate haplotypes in BioProject PRJNA561993, and for the green turtle assembly (rCheMyd1) are GCF_015237465.2 and GCA_015220195.2 for primary and alternate haplotypes respectively in BioProject PRJNA561941. The raw data used for assemblies are available on the Vertebrate Genome Ark (https://vgp.github.io/genomeark/). The leatherback turtle RNA-Seq data generated for the purpose of assembly annotation was deposited in the SRA under accession numbers SRX8787564-SRX8787566 (RNA-Seq) and SRX6360706-SRX6360708 (ISO-Seq). Green turtle RNA-Seq data generated for annotation were deposited in SRA under accessions SRX10863130-SRX10863133 (RNA-Seq) and as SRX11164043-SRX11164046 (ISO-Seq). All scripts used for downstream analyses following genome assembly and annotation have been deposited on GitHub under repository https://github.com/bpbentley/sea_turtle_genomes.

668 **Tables and Figures**

669

670 **Table 1 |** Quality statistics for the genome assemblies and annotations for leatherback (*Dermochelys coriacea*) and green (*Chelonia mydas*)
671 turtles.

	Leatherback turtle (*Dermochelys coriacea*)	Green turtle (*Chelonia mydas*)
Genome ID	rDerCor1	rCheMyd1
Assembly accession	GCA_009764565.3	GCA_015237465.1
Assembly level	Chromosome	Chromosome
Total genome length	2,164,762,090 bp	2,134,358,617 bp
Contig N50	7,029,801	39,415,510
Scaffold N50	137,568,771	134,428,053
Number of scaffolds	40	92
Number of chromosomes	28	28
Quality Value (QV)	38.9	47.7
Annotated protein-coding genes	18,775	19,752

BUSCO Assembly and Annotation Completeness Statistics (based on Vertebrate core BUSCOs) and Annotation BUSCO scores

BUSCO category	Assembly	Annotation	Assembly	Annotation
Complete genes	91.6%	97.2	94.2%	97.9%
Complete + fragmented	95.4%	97.7	96.7%	98.2%
Missing	4.10%	1.3%	2.8%	0.7%
Duplicated	0.5%	1.0%	0.5%	1.1%

672

673

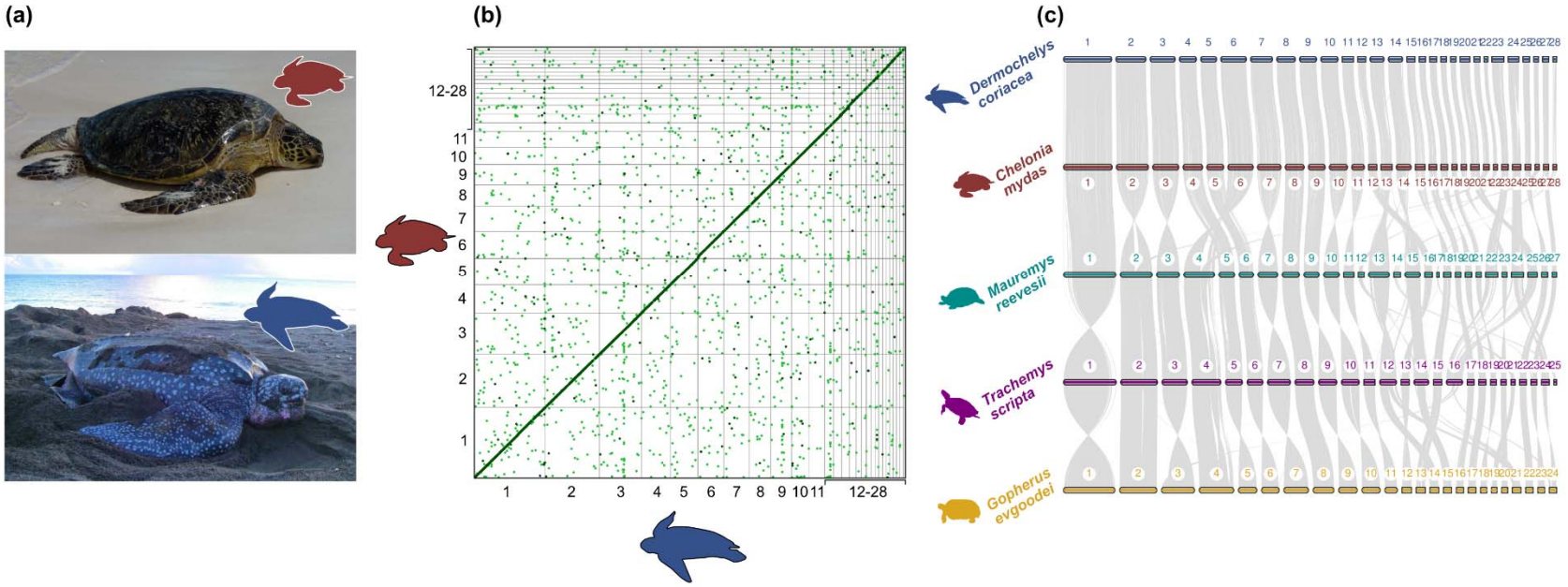

Fig. 1 | (a) Photographs of green turtle (*Chelonia mydas*)*;* photo credit: NOAA NMFS PIFSC under USFWS Permit #TE-72088A-3, and leatherback turtle (*Dermochelys coriacea*); photo credit: Ricardo Tapilatu. **(b)** Dot plot showing regions with an identity greater than 0.5 across the entire genomes of green (red) and leatherback (blue) turtles. **(c)** Gene synteny and collinearity per chromosome among five species of turtles: leatherback turtle (blue), green turtle (red), Chinese pond turtle (*Mauremys reevesii*; green), pond slider turtle (*Trachemys scripta*; purple) and Goode's thornscrub tortoise (*Gopherus evgoodei*; yellow). Each bar represents chromosomes with respective numbers and gray lines represent homolog gene connections among species.

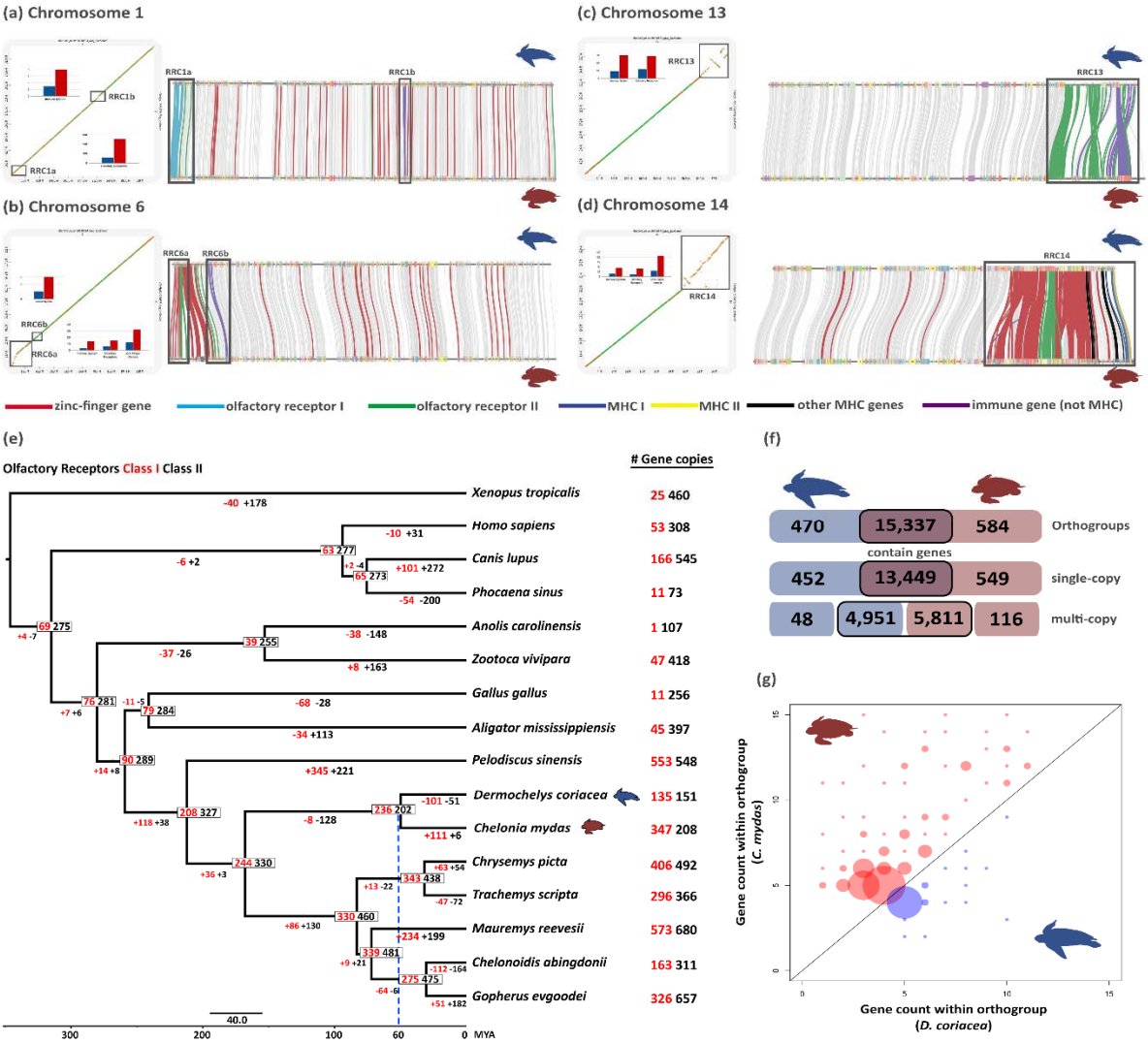

Fig. 2 | **(a-d)** Dotplots (identity values as color; dark green=1-0.75, green=0.75-0.5, orange=0.5-0.25 and yellow=0.25-0) depicting four of the regions with reduced collinearity (RRC) identified within chromosomes and associated with higher copy numbers of immune system, olfactory receptor, or zinc finger domain genes in the green turtle (*Chelonia mydas*) relative to leatherback (*Dermochelys coriacea*) turtle (see also Fig. S3, Tables S3 & S5 for full details of all RRCs). Positions of each RRC are marked with gray squares on the dot plots (left) and gene collinearity maps (right) for each chromosome highlighting the connections among specific gene families in different colors. **(e)** Gene family evolution of olfactory receptors Class I (red) and Class II (black) for amniote phylogeny. Gene numbers are presented on the nodes and gain/loss along each branch are presented below branches. Small scale bar represents substitutions/site and big scale bar represents divergence times (MA). The blue dashed line shows the estimated divergence between the two sea turtle families. **(f)** Number of unique and shared orthogroups and single and multi-copy genes between the two sea turtles (coding genes including genes with rearrangement). The boxes outlined in black denote shared orthogroups, with the higher multi-copy in the green turtle due to greater gene copies within orthogroups. **(g)** Comparison of gene counts between both species per multigenic orthogroup, depicting only those orthogroups where both species have different numbers of genes and a minimum number of five genes for one of the species. Bubbles above the diagonal represent higher counts for the green turtle and below for the leatherback turtle. The size of the bubbles represents the number of orthogroups with the same gene count combination.

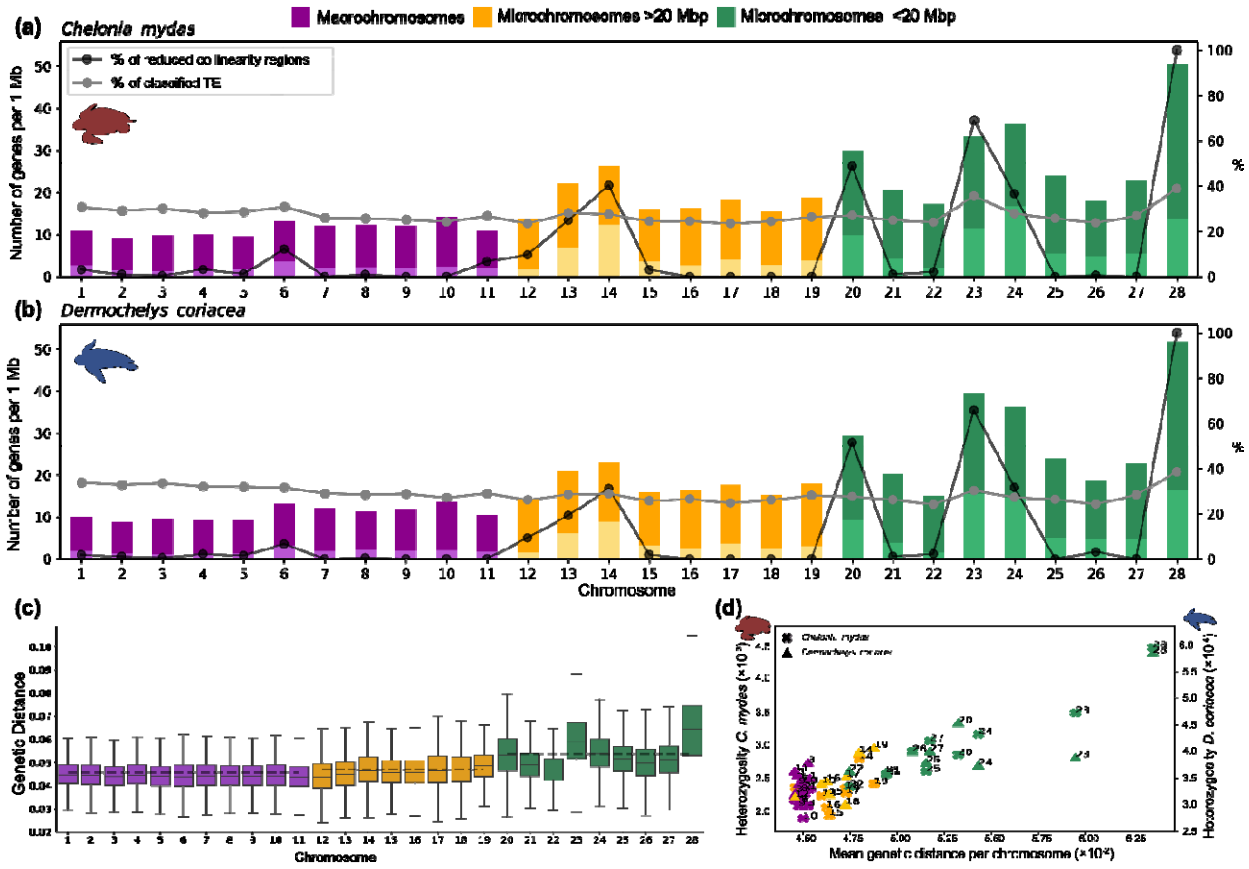

Fig. 3 | Number of genes, genetic distance between species and heterozygosity within species in macrochromosomes, small (<20 Mb) and intermediate (>20 Mb) microchromosomes. **(a)** Relation between the number of genes, percentage of reduced collinearity regions (RRCs), and classified TE per chromosome for the green (*Chelonia mydas*) and **(b)** leatherback (*Dermochelys coriacea*) turtles. Dark colors indicate to the total number of genes and light colors indicate to the number of multicopy genes. **(c)** Average genetic distance between green and leatherback turtles per chromosome. **(d)** Relation between genetic distance and heterozygosity per chromosome for each species.

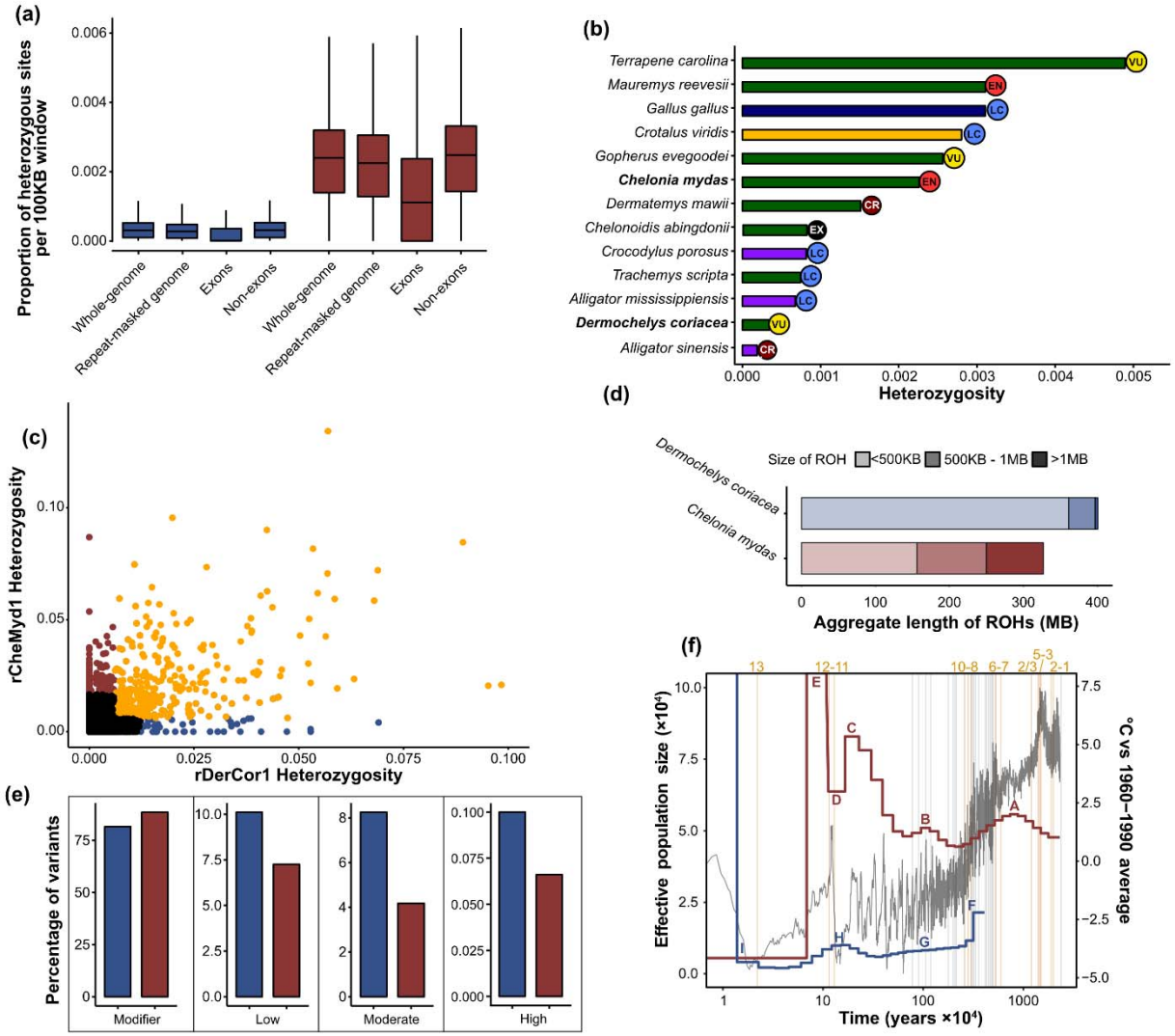

Fig. 4 | Data is presented for the leatherback (*Dermochelys coriacea*; blue) and green (*Chelonia mydas*, red) turtle genomes. **(a)** estimates of heterozygosity across the genome calculated with 100 Kb non-overlapping windows for the entire genome, repeat-masked genome, exons and non-exon regions. **(b)** comparison of genome-wide diversity (π) between the two sea turtle species and 11 other reptile species of varying conservation status using a standardized pipeline for heterozygosity estimation. Bars are colored by taxonomic groups: testudines (green), avians (blue), squamates (yellow), and crocodilians (purple). **(c)** Correlation between heterozygosity in 100 Kb windows containing only exons, generated through alignment to a common reference genome. Windows with higher than mean diversity in leatherbacks (blue), higher in greens (red), and generally high diversity (orange) are highlighted. **(d)** accumulated lengths of runs of homozygosity (ROH). **(e)** predicted impacts of variants from within coding regions. **(f)** Pairwise sequential Markovian coalescent plot (PSMC) of demographic history of both species using a mutation rate of 1.2×10^{-8} and generation time of 30 years, overlayed with temperature (dark grey), magnetic reversals (light grey vertical lines), and selected geological events (orange vertical lines) numbered above as follows: 1. Tethys Sea connectivity changes; 2. [Tethys Sea closure]; 3. [Panama deep water channels close]/Antarctic ice cap starts to grow; 4. Middle Miocene disruption; 5. East Antarctic ice sheet growth; 6. Strait of Gibraltar closes; 7. Zanclean Flood/land bridge between Alaska and Siberia floods; 8. Greenland ice cap starts to grow; 9. Isthmus of Panama formation; 10. Quaternary ice age begins; 11. Onset of Last Glacial Period; 13. Last Glacial Maximum. Repeated numbers/events in square brackets denote uncertainty in timing. Letters indicating portions of the PSMC curves (A-I) are referred to in the text. Data sources are given in supplementary material.

References

1. R. Hirayama, Oldest known sea turtle. *Nature* **392**, 705–708 (1998).

2. H. B. Shaffer, E. McCartney-Melstad, T. J. Near, G. G. Mount, P. Q. Spinks, Phylogenomic analyses of 539 highly informative loci dates a fully resolved time tree for the major clades of living turtles (Testudines). *Mol. Phylogenet. Evol.* **115**, 7–15 (2017).

3. D. A. Pike, Climate influences the global distribution of sea turtle nesting: Sea turtle nesting distributions. *Glob. Ecol. Biogeogr.* **22**, 555–566 (2013).

4. R. C. Thomson, P. Q. Spinks, H. B. Shaffer, A global phylogeny of turtles reveals a burst of climate-associated diversification on continental margins. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2012215118 (2021).

5. J. Davenport, Temperature and the life-history strategies of sea turtles. *J. Therm. Biol.* **22**, 479–488 (1997).

6. W. Frair, R. G. Ackman, N. Mrosovsky, Body Temperature of *Dermochelys coriacea*: Warm Turtle from Cold Water. *Science* **177**, 791–793 (1972).

7. P. C. H. Pritchard, "Evolution, Phylogeny, and Current Status" in *The Biology of Sea Turtles*, P. L. Lutz, J. A. Musick, Eds. (CRC Press, 1996), pp. 1–28.

8. L. McClenachan, J. B. C. Jackson, M. J. H. Newman, Conservation implications of historic sea turtle nesting beach loss. *Front. Ecol. Environ.* **4**, 290–296 (2006).

9. J. B. Jackson, *et al.*, Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–637 (2001).

10. M. A. Grassman, D. W. Owens, J. P. McVey, R. M. M, Olfactory-based orientation in artificially imprinted sea turtles. *Science* **224**, 83–84 (1984).

11. K. J. Lohmann, C. M. F. Lohmann, There and back again: natal homing by magnetic navigation in sea turtles and salmon. *J. Exp. Biol.* **222**, jeb184077 (2019).

12. P. S. Tomillo, V. S. Saba, R. Piedra, F. V. Paladino, J. R. Spotila, Effects of illegal harvest of eggs on the population decline of leatherback turtles in Las Baulas Marine National Park, Costa Rica. *Conserv. Biol.* **22**, 1216–1224 (2008).

13. S. Fossette, *et al.*, Pan-atlantic analysis of the overlap of a highly migratory species, the leatherback turtle, with pelagic longline fisheries. *Proc. Biol. Sci.* **281**, 20133065 (2014).

14. Y. Kaska, *et al.*, Natural and anthropogenic factors affecting the nest-site selection of Loggerhead Turtles, *Caretta caretta*, on Dalaman-Sarıgerme beach in South-west Turkey: (Reptilia: Cheloniidae). *Zool. Middle East* **50**, 47–58 (2010).

15. B. Von Holle, *et al.*, Effects of future sea level rise on coastal habitat. *Journal of Wildlife Management* **83**, 694–704 (2019).

16. N. Mrosovsky, G. D. Ryan, M. C. James, Leatherback turtles: the menace of plastic. *Mar. Pollut. Bull.* **58**, 287–289 (2009).

17. M. Chaloupka, G. H. Balazs, T. M. Work, Rise and fall over 26 years of a marine epizootic in Hawaiian green sea turtles. *J. Wildl. Dis.* **45**, 1138–1142 (2009).

18. L. A. Hawkes, A. C. Broderick, M. H. Godfrey, B. J. Godley, Climate change and marine turtles. *Endanger. Species Res.* **7**, 137–154 (2009).

19. B. P. Wallace, *et al.*, Global conservation priorities for marine turtles. *PLoS One* **6**, e24510 (2011).

20. C. L. Yntema, N. Mrosovsky, Incubation temperature and sex ratio in hatchling loggerhead turtles: a preliminary report. *Mar. Turtle Newsl.* **11**, 9–10 (1979).

21. M. P. Jensen, *et al.*, Environmental Warming and Feminization of One of the Largest Sea Turtle Populations in the World. *Curr. Biol.* **28**, 154–159.e4 (2018).

22. A. D. Mazaris, G. Schofield, C. Gkazinou, V. Almpanidou, G. C. Hays, Global sea turtle conservation successes. *Sci Adv* **3**, e1600730 (2017).

23. IUCN, The IUCN Red List of Threatened Species (2021) (April 16, 2021).

24. L. S. Martínez, *et al.*, Conservation and Biology of the Leatherback Turtle in the Mexican Pacific. *Chelonian Conserv. Biol.* **6**, 70–78 (2007).

25. P. Santidrián Tomillo, *et al.*, Reassessment of the leatherback turtle (*Dermochelys coriacea*) nesting population at Parque Nacional Marino Las Baulas, Costa Rica: effects of conservation efforts. *Chelonian Conserv. Biol.* **6**, 54–62 (2007).

26. J. R. Spotila, R. D. Reina, A. C. Steyermark, P. T. Plotkin, F. V. Paladino, Pacific leatherback turtles face extinction. *Nature* **405**, 529–530 (2000).

27. Laúd OPO Network, Enhanced, coordinated conservation efforts required to avoid extinction of critically endangered Eastern Pacific leatherback turtles. *Sci. Rep.* **10**, 4772 (2020).

28. E. Chan, H. Liew, Decline of the leatherback population in Terengganu, Malaysia, 1956-1995. *Chelonian Conserv. Biol.* **2**, 196–203 (1996).

29. K. L. Eckert, B. P. Wallace, J. G. Frazier, S. A. Eckert, P. C. H. Pritchard, "Synopsis of the biological data on the leatherback sea turtle, *Dermochelys coriacea*" (U.S. Department of Interior, Fish and Wildlife Service, 2012).

30. K. Jones, E. Ariel, G. Burgess, M. Read, A review of fibropapillomatosis in Green turtles *(Chelonia mydas)*. *Vet. J.* **212**, 48–57 (2016).

31. G. Zhang, *et al.*, Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).

32. I. Khan, *et al.*, Olfactory Receptor Subgenomes Linked with Broad Ecological Adaptations in Sauropsida. *Mol. Biol. Evol.* **32**, 2832–2843 (2015).

33. D. Jebb, *et al.*, Six reference-quality genomes reveal evolution of bat adaptations. *Nature* **583**, 578–584 (2020).

34. B. J. McMahon, E. C. Teeling, J. Höglund, How and why should we implement genomics into conservation? *Evolutionary Applications* **7**, 999–1007 (2014).

35. M. A. Supple, B. Shapiro, Conservation of biodiversity in the genomics era. *Genome Biol.* **19**, 131 (2018).

36. P. Brandies, E. Peel, C. J. Hogg, K. Belov, The Value of Reference Genomes in the Conservation of Threatened Species. *Genes* **10**, 864 (2019).

37. P. A. Hohenlohe, W. C. Funk, O. P. Rajora, Population genomics for wildlife conservation and management. *Mol. Ecol.* (2020) https:/doi.org/10.1111/mec.15720.

38. X. Zhang, J. Goodsell, R. B. Norgren Jr, Limitations of the rhesus macaque draft genome assembly and annotation. *BMC Genomics* **13**, 206 (2012).

39. A. Rhie, *et al.*, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746 (2021).

40. A. P. Fuentes-Pardo, D. E. Ruzzante, Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Mol. Ecol.* **26**, 5369–5406 (2017).

41. Z. Wang, *et al.*, The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **45**, 701–706 (2013).

42. A. Whibley, J. L. Kelley, S. R. Narum, The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Mol. Ecol. Resour.* **21**, 641–652 (2021).

43. V. Peona, *et al.*, Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol. Resour.* **21**, 263–286 (2021).

44. Y. Yuan, *et al.*, Comparative genomics provides insights into the aquatic adaptations of mammals. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

45. N. J. Gemmell, *et al.*, The tuatara genome reveals ancient features of amniote evolution. *Nature* **584**, 403–409 (2020).

46. R. N. Johnson, *et al.*, Adaptation and conservation insights from the koala genome. *Nat. Genet.* **50**, 1102–1111 (2018).

47. J.-N. Hubert, T. Zerjal, F. Hospital, Cancer- and behavior-related genes are targeted by selection in the Tasmanian devil (*Sarcophilus harrisii*). *PLoS One* **13**, e0201838 (2018).

48. L. M. Zimmerman, The reptilian perspective on vertebrate immunity: 10 years of progress. *J. Exp. Biol.* **223**, jeb214171 (2020).

49. H. Lee, J. Gurtowski, S. Yoo, M. Nattestad, S. Marcus, Third-generation sequencing and the future of genomics. *BioRxiv* (2016).

50. M. Seppey, M. Manni, E. M. Zdobnov, BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods Mol. Biol.* **1962**, 227–245 (2019).

51. G. I. M. Pasquesi, *et al.*, Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* **9**, 2774 (2018).

52. D. Kordis, Transposable elements in reptilian and avian (sauropsida) genomes. *Cytogenet. Genome Res.* **127**, 94–111 (2009).

53. V. Quesada, *et al.*, Giant tortoise genomes provide insights into longevity and age-related disease. *Nat Ecol Evol* **3**, 87–95 (2019).

54. Q.-H. Wan, *et al.*, Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res.* **23**, 1091–1105 (2013).

55. P. A. Morin, *et al.*, Reference genome and demographic history of the most endangered marine mammal, the vaquita. *Mol. Ecol. Resour.* **21**, 1008–1020 (2021).

56. J. A. Robinson, *et al.*, Genomic Flatlining in the Endangered Island Fox. *Curr. Biol.* **26**, 1183–1189 (2016).

57. P. D. Waters, *et al.*, Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

58. C. S. Endres, K. J. Lohmann, Detection of coastal mud odors by loggerhead sea turtles: a possible mechanism for sensing nearby land. *Mar. Biol.* **160**, 2951–2956 (2013).

59. C. S. Endres, N. F. Putman, K. J. Lohmann, Perception of airborne odors by loggerhead sea turtles. *J. Exp. Biol.* **212**, 3823–3827 (2009).

60. M. Manton, A. Karr, D. W. Ehrenfeld, Chemoreception in the migratory sea turtle, *Chelonia mydas*. *Biol. Bull.* **143**, 184–195 (1972).

61. C. Kitayama, *et al.*, Behavioral effects of scents from male mature Rathke glands on juvenile green sea turtles (*Chelonia mydas*). *J. Vet. Med. Sci.* **82**, 1312–1315 (2020).

62. C. S. Endres, *et al.*, Multi-Modal Homing in Sea Turtles: Modeling Dual Use of Geomagnetic and Chemical Cues in Island-Finding. *Front. Behav. Neurosci.* **10**, 19 (2016).

63. K. L. Dodge, J. M. Logan, M. E. Lutcavage, Foraging ecology of leatherback sea turtles in the Western North Atlantic determined through multi-tissue stable isotope analyses. *Mar. Biol.* **158**, 2813–2824 (2011).

64. K. E. Arthur, M. C. Boyle, C. J. Limpus, Ontogenetic changes in diet and habitat use in green sea turtle (*Chelonia mydas*) life history. *Mar. Ecol. Prog. Ser.* **362**, 303–311 (2008).

65. J. A. Seminoff, *et al.*, Large□scale patterns of green turtle trophic ecology in the eastern Pacific Ocean. *Ecosphere* **12**, e03479 (2021).

66. C. Kitayama, *et al.*, Morphological features of the nasal cavities of hawksbill, olive ridley, and black sea turtles: Comparative studies with green, loggerhead and leatherback sea turtles. *PLoS One* **16**, e0250873 (2021).

67. D. Kondoh, C. Kitayama, Y. K. Kawai, The nasal cavity in sea turtles: adaptation to olfaction and seawater flow. *Cell Tissue Res.* **383**, 347–352 (2021).

68. Y. Yamaguchi, *et al.*, Computed tomographic analysis of internal structures within the nasal cavities of green, loggerhead and leatherback sea turtles. *Anat. Rec.* **304**, 584–590 (2021).

69. Y. Niimura, M. Nei, Evolutionary dynamics of olfactory and other chemosensory receptor genes in vertebrates. *J. Hum. Genet.* **51**, 505–517 (2006).

70. L. R. Yohe, M. Fabbri, M. Hanson, B.-A. S. Bhullar, Olfactory receptor gene evolution is unusually rapid across Tetrapoda and outpaces chemosensory phenotypic change. *Curr. Zool.* **66**, 505–514 (2020).

71. H. Saito, Q. Chi, H. Zhuang, H. Matsunami, J. D. Mainland, Odor coding by a Mammalian receptor repertoire. *Sci. Signal.* **2**, ra9 (2009).

72. M. W. Vandewege, *et al.*, Contrasting Patterns of Evolutionary Diversification in the Olfactory Repertoires of Reptile and Bird Genomes. *Genome Biol. Evol.* **8**, 470–480 (2016).

73. A. Liu, *et al.*, Convergent degeneration of olfactory receptor gene repertoires in marine mammals. *BMC Genomics* **20**, 977 (2019).

74. M. A. Constantino, M. Salmon, Role of chemical and visual cues in food recognition by leatherback posthatchlings (*Dermochelys coriacea* L). *Zoology* **106**, 173–181 (2003).

75. H. V. Siddle, J. Marzec, Y. Cheng, M. Jones, K. Belov, MHC gene copy number variation in Tasmanian devils: implications for the spread of a contagious cancer. *Proc. Biol. Sci.* **277**, 2001–2006 (2010).

76. L. E. Escobar, *et al.*, A global map of suitability for coastal Vibrio cholerae under current and future climate conditions. *Acta Trop.* **149**, 202–211 (2015).

77. L. Zhang, *et al.*, Massive expansion and functional divergence of innate immune genes in a protostome. *Sci. Rep.* **5**, 8693 (2015).

78. J. P. Elbers, S. S. Taylor, Others, Major histocompatibility complex polymorphism in reptile conservation. *Herpetol. Conserv. Biol.* **11**, 1–12 (2016).

79. X. Vekemans, *et al.*, Whole-genome sequencing and genome regions of special interest: Lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. *Mol. Ecol.* **30**, 6072–6086 (2021).

80. T. Rhen, A. Schroeder, Molecular mechanisms of sex determination in reptiles. *Sex Dev.* **4**, 16–28 (2010).

81. M. Czerwinski, A. Natarajan, L. Barske, L. L. Looger, B. Capel, A timecourse analysis of systemic and gonadal effects of temperature on sexual development of the red-eared slider turtle *Trachemys scripta elegans*. *Dev. Biol.* **420**, 166–177 (2016).

82. H. Merchant-Larios, V. Díaz-Hernández, A. Marmolejo-Valencia, Gonadal morphogenesis and gene expression in reptiles with temperature-dependent sex determination. *Sex Dev.* **4**, 50–61 (2010).

83. T. Wibbels, "Critical Approaches to Sex Determination in Sea Turtles" in *Biology of Sea Turtles*, P. L. Lutz, J. A. Musick, J. Wyneken, Eds. (CRC Press, 2003), pp. 103–134.

84. B. P. Bentley, J. L. Stubbs, S. D. Whiting, N. J. Mitchell, Variation in thermal traits describing sex determination and development in Western Australian sea turtle populations. *Funct. Ecol.* **34**, 2302–2314 (2020).

85. M. Hamann, *et al.*, Global research priorities for sea turtles: informing management and conservation in the 21st century. *Endanger. Species Res.* **11**, 245–269 (2010).

86. J. J. Bull, Sex Determination in Reptiles. *Q. Rev. Biol.* **55**, 3–21 (1980).

87. H. Endoh, N. Okada, Total DNA transcription in vitro: a procedure to detect highly repetitive and transcribable sequences with tRNA-like structures. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 251–255 (1986).

88. M. Kajikawa, K. Ohshima, N. Okada, Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. *Mol. Biol. Evol.* **14**, 1206–1217 (1997).

89. C. G. Sotero-Caio, R. N. Platt 2nd, A. Suh, D. A. Ray, Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).

90. H. B. Shaffer, *et al.*, The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013).

91. M. Tollis, *et al.*, The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PLoS One* **12**, e0177708 (2017).

92. Z. Wang, T. Miyake, S. V. Edwards, C. T. Amemiya, Tuatara (Sphenodon) genomics: BAC library construction, sequence survey, and application to the DMRT gene family. *J. Hered.* **97**, 541–548 (2006).

93. C. Moritz, Strategies to protect biological diversity and the evolutionary processes that sustain it. *Syst. Biol.* **51**, 238–254 (2002).

94. P. Fernandez-Fournier, J. M. M. Lewthwaite, A. Ø. Mooers, Do We Need to Identify Adaptive Genetic Variation When Prioritizing Populations for Conservation? *Conserv. Genet.* **22**, 205–216 (2021).

95. Y. J. Borrell, *et al.*, Heterozygosity-fitness correlations in the gilthead sea bream *Sparus aurata* using microsatellite loci from unknown and gene-rich genomic locations. *J. Fish Biol.* **79**, 1111–1129 (2011).

96. J. A. DeWoody, A. M. Harder, S. Mathur, J. R. Willoughby, The long-standing significance of genetic diversity in conservation. *Mol. Ecol.* **30**, 4147–4154 (2021).

97. Y. Willi, *et al.*, Conservation genetics as a management tool: The five best-supported paradigms to assist the management of threatened species. *Proc. Natl. Acad. Sci. U. S. A.* **119** (2022).

98. L. M. Komoroske, *et al.*, A versatile Rapture (RAD-Capture) platform for genotyping marine turtles. *Mol. Ecol. Resour.* **19**, 497–511 (2019).

99. P. H. Dutton, B. W. Bowen, D. W. Owens, A. Barragan, S. K. Davis, Global phylogeography of the leatherback turtle (*Dermochelys coriacea*). *J. Zool.* **248**, 397–409 (1999).

100. J. Romiguier, *et al.*, Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* **515**, 261–263 (2014).

101. M. V. Westbury, B. Petersen, E. Garde, M. P. Heide-Jørgensen, E. D. Lorenzen, Narwhal Genome Reveals Long-Term Low Genetic Diversity despite Current Large Abundance Size. *iScience* **15**, 592–599 (2019).

102. J. A. Robinson, *et al.*, Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Sci Adv* **5**, eaau0757 (2019).

103. United States National Marine Fisheries Service, U.S. Fish and Wildlife Service, "Endangered Species Act status review of the leatherback turtle (*Dermochelys coriacea*)" (United States National Marine Fisheries Service, 2020).

104. M. P. Jensen, *et al.*, The evolutionary history and global phylogeography of the green turtle (*Chelonia mydas*). *J. Biogeogr.* **46**, 860–870 (2019).

105. Y. Tikochinski, *et al.*, Mitochondrial DNA short tandem repeats unveil hidden population structuring and migration routes of an endangered marine turtle. *Aquat. Conserv.* (2018) https:/doi.org/10.1002/aqc.2908.

106. P. J. Bradshaw, *et al.*, Defining conservation units with enhanced molecular tools to reveal fine scale structuring among Mediterranean green turtle rookeries. *Biological Conservation* **222**, 253–260 (2018).

107. M. Kardos, *et al.*, The crucial role of genome-wide genetic variation in conservation. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2104642118 (2021).

108. P. Dobrynin, *et al.*, Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol.* **16**, 277 (2015).

109. A. L. K. Mattila, *et al.*, High genetic load in an old isolated butterfly population. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E2496–505 (2012).

110. M. V. Westbury, *et al.*, Extended and Continuous Decline in Effective Population Size Results in Low Genomic Diversity in the World's Rarest Hyena Species, the Brown Hyena. *Molecular Biology and Evolution* **35**, 1225–1237 (2018).

111. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Curr. Biol.* **28**, 3487–3494.e4 (2018).

112. Y. Xue, *et al.*, Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).

113. A. D. Foote, *et al.*, Killer whale genomes reveal a complex history of recurrent admixture and vicariance. *Mol. Ecol.* **28**, 3427–3444 (2019).

114. N. Dussex, *et al.*, Population genomics of the critically endangered kākāpō. *Cell Genomics* **1**, 100002 (2021).

115. C. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evol Lett* **5**, 33–47 (2021).

116. Y. D. Dewoody, J. A. Dewoody, On the estimation of genome-wide heterozygosity using molecular markers. *J. Hered.* **96**, 85–88 (2005).

117. J. P. van der Zee, *et al.*, The population genomic structure of green turtles (*Chelonia mydas*) suggests a warm-water corridor for tropical marine fauna between the Atlantic and Indian oceans during the last interglacial. *Heredity* **127**, 510–521 (2021).

118. R. R. Fitak, S. Johnsen, Green sea turtle (*Chelonia mydas*) population history indicates important demographic changes near the mid-Pleistocene transition. *Mar. Biol.* **165**, 110 (2018).

119. S. T. Vilaça, *et al.*, Divergence and hybridization in sea turtles: Inferences from genome data show evidence of ancient gene flow between species. *Mol. Ecol.* (2021) https:/doi.org/10.1111/mec.16113.

120. A. H. Patton, *et al.*, Contemporary Demographic Reconstruction Methods Are Robust to Genome Assembly Quality: A Case Study in Tasmanian Devils. *Mol. Biol. Evol.* **36**, 2906–2921 (2019).

121. A. Khan, *et al.*, Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

122. N. A. O'Leary, *et al.*, Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45 (2016).

123. K. D. Pruitt, *et al.*, RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* **42**, D756–63 (2014).

124. F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

125. M. Blum, *et al.*, The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).

126. H. Mi, *et al.*, PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. *Nucleic Acids Res.* **49**, D394–D403 (2021).

127. B. Paten, *et al.*, Cactus graphs for genome comparisons. *J. Comput. Biol.* **18**, 469–481 (2011).

128. J. Armstrong, *et al.*, Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587**, 246–251 (2020).

129. J. M. Flynn, *et al.*, RepeatModeler2: automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).

130. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **Chapter 4**, Unit 4.10 (2009).

131. A. Smit, R. Hubley, P. Green, RepeatMasker Open-4.0. 2013-2015http. *repeatmasker. org* (2015).

132. F. K. Mendes, D. Vanderpool, B. Fulton, M. W. Hahn, CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* (2020) https:/doi.org/10.1093/bioinformatics/btaa1022.

133. D. M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).

134. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

135. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013).

136. A. McKenna, *et al.*, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

137. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013).

138. T. S. Korneliussen, A. Albrechtsen, R. Nielsen, ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).

139. S. Purcell, *et al.*, PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

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

141. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

142. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

143. P. Cingolani, *et al.*, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).