Tropical ancient DNA reveals relationships of the extinct Bahamian giant tortoise *Chelonoidis alburyorum*

Christian Kehlmaier¹,†, Axel Barlow²,†, Alexander K. Hastings³, Melita Vamberger¹, Johanna L. A. Paijmans², David W. Steadman⁴, Nancy A. Albury⁵, Richard Franz⁵, Michael Hofreiter² and Uwe Fritz¹

¹Museum of Zoology, Senckenberg Dresden, A. B. Meyer Building, 01109 Dresden, Germany
²Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, OT Golm, Germany
³Virginia Museum of Natural History, 21 Starling Avenue, Martinsville, VA 24112, USA
⁴Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
⁵National Museum of The Bahamas, Marsh Harbour, Abaco, The Bahamas

Ancient DNA of extinct species from the Pleistocene and Holocene has provided valuable evolutionary insights. However, these are largely restricted to mammals and high latitudes because DNA preservation in warm climates is typically poor. In the tropics and subtropics, non-avian reptiles constitute a significant part of the fauna and little is known about the genetics of the many extinct reptiles from tropical islands. We have reconstructed the near-complete mitochondrial genome of an extinct giant tortoise from the Bahamas (*Chelonoidis alburyorum*) using an approximately 1 000-year-old humerus from a water-filled sinkhole (blue hole) on Great Abaco Island. Phylogenetic and molecular clock analyses place this extinct species as closely related to Galápagos (*C. niger* complex) and Chaco tortoises (*C. chilensis*), and provide evidence for repeated overseas dispersal in this tortoise group. The ancestors of extant *Chelonoidis* species arrived in South America from Africa only after the opening of the Atlantic Ocean and dispersed from there to the Caribbean and the Galápagos Islands. Our results also suggest that the anoxic, thermally buffered environment of blue holes may enhance DNA preservation, and thus are opening a window for better understanding evolution and population history of extinct tropical species, which would likely still exist without human impact.

### 1. Introduction

Post-mortem degradation of DNA is climate dependent, being greatly accelerated in warm tropical and subtropical regions [1,2]. As a result, extinct Late Pleistocene megafauna from cold climates has been widely studied using ancient DNA (aDNA) approaches [3], providing valuable insights in ecology, evolution, and biogeography, and causes of extinction of vanished species. By contrast, aDNA from tropical and subtropical environments remains largely unexplored, apart from some notable exceptions [4–8]. A further consequence is that aDNA studies are biased toward taxa that are abundant at higher latitudes, in particular, mammals. Other groups, such as non-avian reptiles, which are highly diverse in warm climates, remain little studied [9–13]. Subtropical and tropical islands are systems that would benefit greatly from information from aDNA because they have experienced substantial losses of both megafauna and small-bodied species after the Holocene arrival of humans [14–16]. The Bahamas are one such example, with much of the original vertebrate fauna (reptiles, birds, and mammals) having disappeared within a few centuries after the arrival of human settlers about 1 000 years before present (BP) [17].

Among the extinct Bahamian species is an endemic giant tortoise, *Chelonoidis alburyorum*, which is believed to have gone extinct around 780 BP [18]. Complete
tortoise fossils, with a shell length of up to 47 cm, have been discovered in Sawmill Sink, a deep inland blue hole and cave system, on Great Abaco Island in the northern Bahamas [19]. The fossils were retrieved from anoxic saltwater, and were found to contain substantial collagen [20], suggesting at least the potential for DNA preservation. Although samples from temperate saltwater deposits have yielded endogenous DNA [21,22], the validity of a recent report on aDNA from a Mexican underwater cave [23] has been questioned [24]. Thus, the preservation potential of DNA in tropical aquatic or water-logged environments is poorly understood. In this study, we have analyzed aDNA from an almost 1,000-year-old subfossil humerus of C. alburyorum from Sawmill Sink. Although DNA preservation in the sample is poor, we have recovered an almost complete mitochondrial genome sequence from the sample, which provides new insights into the origin and relationships of this enigmatic giant tortoise and contributes to a better understanding of the biogeography of the Bahamas.

2. Material and methods

(a) Studied specimens

The following specimens from the collections of the National Museum of The Bahamas, Marsh Harbour, Bahamas (NMB) and the Museum of Zoology, Senckenberg Dresden, Germany (MTD) were studied: C. alburyorum, subfossil humerus of specimen NMB.AB5.008 (Sawmill Sink, Abaco Island, Bahamas); C. carbonarius, fresh tissue sample MTD-T 5138 (Inacuco, French Guiana); C. chelinesis, fresh tissue sample MTD-T 5754 (~40.787778, −65.316389, Rio Negro Province, Argentina); C. denticulatus, fresh tissue sample MTD-T 7235 (from pet trade); C. vicina (‘Poldi’ kept at Reptile Zoo Happy, Klagenfurt, Austria), blood sample MTD-T 14174; Geochelone sulcata, fresh tissue sample MTD-T 872 (captive bred).

(b) Processing of the ancient sample

All stages of sample processing prior to polymerase chain reaction (PCR) amplification were carried out in dedicated aDNA facilities at the University of Potsdam, following established guidelines [25]. Negative controls (water blanks) were included during DNA extraction and library preparation and screened for evidence of contamination. Two 50 mg bone powder samples were obtained from the C. alburyorum humerus. DNA was extracted from each sample using a published protocol optimized for the recovery of short aDNA fragments [26]. DNA extracts were treated with uracil-DNA glycosylase (UDG) to remove uracil residues possibly resulting from DNA damage and then converted into Illumina sequencing libraries using a protocol based on double-stranded DNA amplification [27]. All stages of sample processing prior to polymerase chain reaction (PCR) amplification were carried out in dedicated aDNA facilities at the University of Potsdam, following established guidelines [25]. Negative controls (water blanks) were included during DNA extraction and library preparation and screened for evidence of contamination. Two 50 mg bone powder samples were obtained from the C. alburyorum humerus. DNA was extracted from each sample using a published protocol optimized for the recovery of short aDNA fragments [26]. DNA extracts were treated with uracil-DNA glycosylase (UDG) to remove uracil residues probably resulting from DNA damage and then converted into Illumina sequencing libraries using a protocol based on double-stranded DNA amplification [27].

An initial assessment of DNA preservation and contamination was made by low-level shotgun sequencing of the libraries on an Illumina NextSeq 500 sequencing platform generating 150 bp paired-end reads. Owing to low abundance of endogenous DNA fragments in the sequencing libraries, we performed two-rounds of in-solution hybridization capture to enrich for mitochondrial DNA fragments [28,29], using DNA baits generated from long-range PCR products of the congeneric species C. chelinesis (see below). Sequencing of enriched libraries was as described above.

(c) Processing of modern samples

DNA of extant relatives of C. alburyorum was extracted using commercial kits (Analytik Jena AG, Jena, Germany), and served as template for amplicon sequencing (C. chelinesis, C. vicina, and C. sulcata), or in-solution hybridization capture enrichment (C. carbonarius and C. denticulatus), depending on DNA quality. Amplicon sequencing involved PCR amplification of mitogenomes using standard methods (for primer sequences and PCRs see electronic supplementary material, Amplicon sequencing, and table S1). Amplification products were sheared and converted into Illumina sequencing libraries using a published protocol based on double-stranded DNA [30] with modifications [31]. Hybridization capture enrichment of degraded samples followed the procedures described previously for the ancient sample. All modern sample libraries were sequenced on an Illumina NextSeq 500 sequencing platform generating 150 bp paired-end reads.

(d) Assessment of endogenous and contaminant DNA content

Prior to analysis, adapter sequences were trimmed from the 3′ read ends, overlapping paired-end reads were merged, and any merged reads less than 20 bp discarded, using the program SeqPrep [32]. Analysis of shotgun data from the ancient C. alburyorum sample involved estimation of endogenous DNA content by calculating the proportion of sequence reads that could be mapped to the reference genome assembly of the painted turtle (Chrysemys picta bellii) [33] using bwa [34] with a mismatch value of 0.001. Reads with low mapping quality (less than 30) and likely PCR duplicates were removed from the alignment using SAMtools [35]. Cow, dog, cat, human, and mouse were then investigated as potential sources of contamination using fastqscreen [36]. To assess the authenticity of the ancient reads obtained, the shotgun data as well as the assembled reads from the enriched libraries were re-mapped to the reference nuclear genome assembly of Ch. picta and the newly generated mitogenome of C. alburyorum, respectively, in order to generate nucleotide misincorporation plots using mapDamage 2.0 [37]. Finally, we estimated the preservation of DNA in a bone sample deposited in the terrestrial environment of the Bahamas at 25.04829° latitude and −77.43284° longitude and buried under a 20 m layer of silt-loam soil using the online resource http://thermal-age.eu (Job 1337), for comparison to the empirical data obtained from the C. alburyorum sample.

(e) Assembly of mitogenome sequences

Assembly of mitogenome sequences from the enriched and amplicon libraries involved a two-step baiting and iterative mapping approach in MITObin [38]. Prior to assembly, duplicate read pairs were removed from each dataset using FastUniq [39] and the order of the remaining unique reads randomized using fastsort [36]. Only reads more than 31 bp were used for assembly, which corresponded to the k-mer size used for baiting. Various levels of coverage and mapping stringency were tested and optimal values selected based on visual assessment of the final alignments in Table v. 1.15.09.1 [40]. After assembly, PCR priming sites were removed from amplicon assemblies. Mitogenome annotation was performed using MITOS [41].

(f) Phylogenetic analyses and molecular dating

Novel sequences were aligned with all Testudinidae mitochondrial genomes available on GenBank, plus representatives of the turtle genera Mauremys and Emys as outgroups (electronic supplementary material, Mitochondrial genomes from GenBank used for phylogenetic analyses), using the ClustalW algorithm [42] with default settings, resulting in 22 485 aligned positions. Alternative data partitioning schemes were compared using the software PartitionFinder [43] using the Bayesian Information Criterion (BIC).

Phylogenetic analysis using Bayesian Inference was conducted with MrBayes 3.2.1 [44] and optimal models selected by PartitionFinder (electronic supplementary material, table S2), with two parallel runs (each with four chains) and default parameters. Parameter convergence, sampling adequacy, and
appropriate burn-in was determined using the software Tracer 1.6 [45]. A 50% majority rule consensus tree was then generated from the posterior sample of trees. Phylogenetic analysis was additionally conducted under Maximum Likelihood using RAxML 7.2.8 [46] and the GTR + G substitution model. Clade support was assessed by bootstrap analysis, involving multiple independent runs using both fast and thorough bootstrap algorithms.

Molecular dating was conducted with BEAST 1.8.2 [47]. Two calibration points were specified using normally distributed priors. Based on the fossil species C. hesternus from the middle Miocene La Venta Fauna of Colombia, thought to be close to the last common ancestor of C. carbonarius and C. denticulatus [48], the split between these two species was identified with La Venta age, 13.5–11.8 million years ago (mya) [49,50]. Accordingly, the node age was set to a mean of 12.35 mya with a standard deviation of 0.6. The Geoemydidae (Mauremys) + Testudinidae node was dated to 50.3–66.99 mya, based on the fossil tortoise species Hadrianus majusculus [51], using a mean of 58.65 mya and a standard deviation of 5.08. Analyses involved the HKY substitution model, estimated base frequencies, an uncorrelated lognormal relaxed molecular clock, and the Yule tree prior. MCMC chains were inspected as described above, and the maximum clade credibility tree was extracted using TreeAnnotator and viewed in FigTree 1.4.2 [52].

3. Results
(a) DNA preservation of the Chelonoidis alburyorum sample

Analysis of the C. alburyorum shotgun data indicated high levels of degradation and contamination (electronic supplementary material, figure S1). Only 1.4% of reads could be mapped to the Ch. picta reference genome, although this is almost certainly an underestimate of endogenous DNA content due to the considerable evolutionary divergence of Ch. picta from C. alburyorum (approx. 86 mya) [51]. To corroborate the presence of ancient endogenous DNA molecules, misincorporation plots were generated for the 24,362 reads resulting from shotgun sequencing that mapped against the full genome of Ch. picta (electronic supplementary material, figure S1a), as well as for 25,913 captured reads of C. alburyorum that re-mapped to the assembled mitochondrial genome (electronic supplementary material, figure S1b). The observed C to T substitutions increase towards the ends of the fragments, which is consistent with the expectation for aDNA fragments [53,54]. The relatively low misincorporation rates for the re-mapped mitochondrial reads can be attributed to the use of UDG during library preparation, which removes the majority of deaminated cytosines. Overall, this result validates the ancient origin of the C. alburyorum mitogenome.

Contamination analysis using fastqscreen revealed multiple potential sources of contamination, in particular, human (electronic supplementary material, figure S1c). Yet, more reads could be uniquely assigned to the Chrysemys genome than to any of the alternative genomes tested. Predicted DNA preservation for a bone sample deposited in a terrestrial environment of the Bahamas indicated a mean fragment length of just 24 bp, and a probability of 0.012 for the survival of an intact 100 bp fragment (electronic supplementary material, figure S1d). However, DNA preservation appears to be substantially better in the C. alburyorum sample; the mean length of recovered mitochondrial fragments is 65 bp and 7.88% of recovered fragments are at least 100 bp long (electronic supplementary material, figure S1e).

| Table 1. Assembly of mitochondrial genomes using MITObim: settings and results. k-mer size (| length of bait) was 31 (default). Allowed mismatches per read was set to 8. Cropped contig length is the assembled contig cropped to 5’ and 3’ long-range PCR primers. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| taxon           | size of readpool| # MITObim iterations | # reads assembled| average consensus quality (max. = 90) | assembled contig length |
| Chelonoidis alburyorum | 19,929 (19.9%) | 203 | 19,929 (19.9%) | 90 | 15 | 150 bp |
| Chelonoidis carbonarius | 23,929 (59.8%) | 99 | 23,929 (59.8%) | 90 | 15 | 163 bp |
| Chelonoidis chilensis | 4,723 (94.6%) | 73 | 4,723 (94.6%) | 90 | 15 | 164 bp |
| Chelonoidis denticulatus | 5,000 | 79 | 5,000 | 79 | 15 | 146 bp |
| Chelonoidis vicina | 9,713 (97.1%) | 166 | 9,713 (97.1%) | 90 | 15 | 151 bp |
| Geocheleone sulcata | 4,683 (92.9%) | 74 | 4,683 (92.9%) | 90 | 15 | 141 bp |
(b) Mitochondrial phylogeny of Chelonoidis

The assembly of the C. alburyorum mitogenome comprised 19,929 reads, resulting in an average 85-fold read-depth and included the nearly complete mtDNA gene and tRNA complement, covering 15,328 bp and ranging from 12S to cyt b, but lacking the control region. Assemblies of modern relatives were all of a similar standard (table 1). Read information of sequenced voucher specimens, including European Nucleotide Archive (ENA) accession numbers and sequenced blanks, can be found in the electronic supplementary material, tables S3 and S4.

Phylogenetic analyses unambiguously placed C. alburyorum in a clade together with C. chilensis and C. vicina (figure 1; electronic supplementary material, figure S2), with the latter two suggested as weakly supported sister taxa. C. carbonarius and C. denticulatus together constituted the sister clade to the previous three taxa. The relationships of the remaining testudinid species corresponded to expectations from previous papers based on less sequence data [55,56]. According to our molecular clock calculations, and in agreement with the oldest record of a fossil tortoise in South America [55], the divergence of Chelonoidis from the African Geochelone sulcata and subsequent dispersal to South America would have occurred distinctly later than in the two other groups (Eocene), around the Oligocene–Miocene transition (figure 1). For the colonization of the Caribbean islands, two transoceanic routes have to be considered: directly from South America or via southern Central America. The originally wide Caribbean distribution of Chelonoidis is indicated by records of extinct species from 10 Bahamian islands as well as from Cuba, Hispaniola, Mona, Navassa, Barbados, Curacao, Grand Turk, Caicos, Anguilla, and Bermuda [19]. The extent to which Caribbean terrestrial ecosystems have been altered by the loss of these ‘ecosystem engineers’ is fertile ground for new research in palaeoecology and restoration ecology [59,60].

With a proposed divergence date of approximately 15.5 mya, this Caribbean island radiation postdates the divergence of South American Chelonoidis from African Geochelone...
by only approximately 7 mya and predates the divergence of Galápagos and Chaco tortoises by approximately 3.5 mya. Owing to human activities during the mid- to late Holocene, the entire Caribbean tortoise radiation was lost, as was the case for the sloths that once occupied the Greater Antilles [61]. This loss of the Caribbean tortoises is another example of the massive impoverishment of evolutionary diversity that accompanied human colonization of oceanic islands worldwide [15,16]. The extent of this depletion only increases as the insular fossil record continues to grow. Because these eliminated species and lineages would still exist if not for human interference, we should endeavour to incorporate

Figure 2. Extant and fossil occurrences of Chelonoidis and dispersals through time. For details, see electronic supplementary material, table S1.
them into studies of ‘modern’ biodiversity, including their genetic diversity. Until aDNA analyses are done on other Caribbean forms of Chelonoidis, we cannot evaluate, for example, how many dispersal events from South America were required to account for the Caribbean radiation of tortoises, or how much of the Holocene diversity of Chelonoidis was lost due to human activity.

(b) Implications for the study of tropical ancient DNA

The recovery of genetic information from tropical and subtropical fossils remains a challenge. A unique property of the C. alburyorum fossil analysed here is its deposition environment: the Sawmill Sink blue hole. It is well known that certain micro-environments can provide conditions that enhance DNA preservation, e.g. cave environments greatly improve the probability of DNA survival relative to the external landscape [2]. Marine environments in general are also known to provide promising potential for DNA preservation, as evidenced by studies of Late Pleistocene remains retrieved from temperate oceans [22,62]. Although estimated endogenous DNA content and preservation of the C. alburyorum sample is poor, it is nevertheless sufficient for mitogenome sequencing using methods optimized for the retrieval of aDNA. Moreover, preservation in this sample is substantially better than that predicted for a bone sample deposited for the same time in a terrestrial environment of the Bahamas. Although any conclusions based on this single sample are tentative, we propose that the anoxic, thermally buffered marine environment of blue holes and similar preservation contexts may provide conditions that enhance DNA preservation—even in tropical regions, where DNA recovery from ancient samples is often considered to be unachievable. These findings indicate a future direction with high potential for aDNA research in the tropics.

References


