Molecular Phylogenetics and Evolution 91 (2015) 178–193

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Interordinal gene capture, the phylogenetic position of Steller's sea cow based on molecular and morphological data, and the macroevolutionary history of Sirenia ☆

Mark S. Springer [a,*], Anthony V. Signore [b], Johanna L.A. Paijmans [c], Jorge Vélez-Juarbe [d], Daryl P. Domning [e,f], Cameron E. Bauer [b], Kai He [b], Lorelei Crerar [g], Paula F. Campos [h,i], William J. Murphy [j], Robert W. Meredith [k], John Gatesy [a], Eske Willerslev [h], Ross D.E. MacPhee [l], Michael Hofreiter [c,m], Kevin L. Campbell [b,*]

[a] Department of Biology, University of California, Riverside, CA 92521, USA
[b] Department of Biological Sciences, University of Manitoba, Winnipeg, MB R3T 2N2, Canada
[c] Department of Biology, The University of York, Wentworth Way, Heslington, York YO10 5DD, UK
[d] Department of Mammalogy, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, USA
[e] Laboratory of Evolutionary Biology, Department of Anatomy, Howard University, Washington, DC 20059, USA
[f] Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA
[g] Department of Biology, George Mason University, Fairfax, VA 22030, USA
[h] Center for GeoGenetics, Natural History Museum, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen K, Denmark
[i] CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas 289, 4050-123 Porto, Portugal
[j] Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843, USA
[k] Department of Biology and Molecular Biology, Montclair State University, Montclair, NJ 07043, USA
[l] Department of Mammalogy, American Museum of Natural History, New York, NY 10024, USA
[m] Adaptive and Evolutionary Genomics, Institute for Biochemistry and Biology, Faculty of Mathematics and Natural Sciences, University of Potsdam, Karl-Liebknecht-Strasse 24-24, 14476 Potsdam, Germany

ARTICLE INFO

Article history:
Received 4 April 2015
Revised 22 May 2015
Accepted 28 May 2015
Available online 4 June 2015

Keywords:
Steller's sea cow
Ancient DNA
Macroevolution
Sirenia
Teeth

ABSTRACT

The recently extinct (ca. 1768) Steller's sea cow (*Hydrodamalis gigas*) was a large, edentulous North Pacific sirenian. The phylogenetic affinities of this taxon to other members of this clade, living and extinct, are uncertain based on previous morphological and molecular studies. We employed hybridization capture methods and second generation sequencing technology to obtain >30 kb of exon sequences from 26 nuclear genes for both *H. gigas* and *Dugong dugon*. We also obtained complete coding sequences for the tooth-related enamelin (*ENAM*) gene. Hybridization probes designed using dugong and manatee sequences were both highly effective in retrieving sequences from *H. gigas* (mean = 98.8% coverage), as were more divergent probes for regions of *ENAM* (99.0% coverage) that were designed exclusively from a proboscidean (African elephant) and a hyracoid (Cape hyrax). New sequences were combined with available sequences for representatives of all other afrotherian orders. We also expanded a previously published morphological matrix for living and fossil Sirenia by adding both new taxa and nine new postcranial characters. Maximum likelihood and parsimony analyses of the molecular data provide robust support for an association of *H. gigas* and *D. dugon* to the exclusion of living trichechids (manatees). Parsimony analyses of the morphological data also support the inclusion of *H. gigas* in Dugongidae with *D. dugon* and fossil dugongids. Timetree analyses based on calibration density approaches with hard- and soft-bounded constraints suggest that *H. gigas* and *D. dugon* diverged in the Oligocene and that crown sirenians last shared a common ancestor in the Eocene. The coding sequence for the *ENAM* gene in *H. gigas* does not contain frameshift mutations or stop codons, but there is a transversion mutation (AG to CG) in the acceptor splice site of intron 2. This disruption in the edentulous Steller's sea cow is consistent with previous studies that have documented inactivating mutations in tooth-specific loci of a variety of edentulous and enamelless vertebrates including birds, turtles, aardvarks, pangolins, xenarthrans, and baleen whales. Further, branch-site dN/dS analyses provide evidence

☆ This paper was edited by the Associate Editor Liliana Davalos.

* Corresponding authors.
E-mail addresses: mark.springer@ucr.edu (M.S. Springer), Kevin.Campbell@umanitoba.ca (K.L. Campbell).

http://dx.doi.org/10.1016/j.ympev.2015.05.022

for positive selection in *ENAM* on the stem dugongid branch where extensive tooth reduction occurred, followed by neutral evolution on the *Hydrodamalis* branch. Finally, we present a synthetic evolutionary tree for living and fossil sirenians showing several key innovations in the history of this clade including character state changes that parallel those that occurred in the evolutionary history of cetaceans.

1. Introduction

Sirenia is one of two fully aquatic mammalian clades. Like Cetacea, Sirenia has a fossil record extending back to the early middle Eocene (Savage et al., 1994; Domning, 2000; Benoit et al., 2013) and includes quadrupedal forms that document the early evolution of this group from terrestrial ancestors (Domning, 2000, 2001). Recent sirenians include three manatee species (*Trichechus inunguis*, *T. manatus*, *T. senegalensis*), the dugong (*Dugong dugon*), and Steller's sea cow (*Hydrodamalis gigas*). The latter was discovered in the North Pacific by Georg Wilhelm Steller in 1741 and became extirpated by 1768 as a consequence of human predation on *H. gigas* and/or the indirect effects of human predation on sea otters (Domning, 1978; Forsten and Youngman, 1982; Anderson, 1995; Turvey and Risley, 2006; Domning et al., 2007). Turvey and Risley (2006) employed a modeling approach and concluded that overhunting was sufficient to exterminate Steller's sea cow and that the effects of sea otter removal on sea cow decline were minimal. In addition to their much larger size – up to 10 m in length and up to 11,000 kg in mass (Domning, 1978) – *H. gigas* are distinguished from extant sirenians by their lack of teeth, and instead used a pair of broad cornified horny pads to masticate kelp (algal seaweeds) (Steller, 1751, 1899; Domning, 1976; Forsten and Youngman, 1982). *D. dugon* exhibits an intermediate condition and possesses both horny pads and teeth (Lanyon and Sanson, 2006). However, the teeth are simple, peglike structures that lose their thin coat of surface enamel from wear shortly after eruption (Lanyon and Sanson, 2006).

Hydrodamalis has traditionally been placed in the family Dugongidae, along with *Dugong*, whereas manatees belong to the family Trichechidae (McKenna and Bell, 1997). Cladistic analyses of anatomical characters from the cranium provide some support for an association of *Hydrodamalis* and *Dugong* to the exclusion of Trichechidae (Domning, 1994; Buffrénil et al., 2010; Vélez-Juarbe et al., 2012). However, Voss (unpublished doctoral dissertation, 2013) concluded that *Hydrodamalis* is more closely related to trichechids than to dugongids based on a cladistic analysis of a matrix that included both cranial and postcranial characters. Thus, cladistic analyses of anatomical characters provide only mixed support for an association of *Hydrodamalis* and *Dugong* to the exclusion of Trichechidae. Molecular studies addressing this problem include immunological comparisons (Rainey et al., 1984) and analyses of partial sequences for the mitochondrial cytochrome b (*CYTB*) gene that were obtained with PCR and Sanger sequencing (Ozawa et al., 1997). Both of these studies recovered an association of *Hydrodamalis* and *Dugong* to the exclusion of *Trichechus*, although bootstrap values in the mitochondrial study were below 70% (Ozawa et al., 1997). By contrast, Crerar (unpublished doctoral dissertation, 2012) used PCR and Sanger sequencing to obtain partial sequences (several hundred base pairs) for two mitochondrial genes, *CYTB* and D-loop, with more extensive taxon sampling among paenungulates (Sirenia + Proboscidea + Hyracoidea) than Ozawa et al. (1997). Crerar's analyses provided only limited support for dugongid monophyly (i.e., *Hydrodamalis* + *Dugong*) and in some analyses grouped *Dugong* and *Trichechus* to the exclusion of *Hydrodamalis* or placed *Hydrodamalis* within *Trichechus* (Crerar, unpublished doctoral dissertation, 2012).

Here, we address the phylogenetic position of *Hydrodamalis gigas* with both molecular and morphological data. Molecular targets (26 nuclear loci; ~34 kb) were enriched via hybridization capture using probes predominantly designed from dugong and manatee DNA sequences. Probes using elephant and hyrax sequence as bait were also employed to capture the complete coding sequence of the tooth-specific gene enamelin (*ENAM*) from both modern (dugong) and ancient (*Hydrodamalis*) DNA samples. The morphological data set builds on Vélez-Juarbe et al.'s (2012) matrix and includes both cranial and postcranial characters for Recent and fossil Sirenia. We provide phylogenetic and timetree estimates based on molecular and morphological data sets, and outline important steps in the macroevolutionary history of Sirenia. We also provide evidence for an acceptor splice site mutation in the *ENAM* gene of *H. gigas*, thereby providing another example of molecular decay of *ENAM* that parallels morphological degeneration of enamel in the fossil record.

2. Materials and methods

2.1. Sampling

Six *Hydrodamalis gigas* specimens (ZI 6842, ZI 6844, ZI 6846, ZI 6852, ZI 6853, and ZI 17170(2)) collected in the mid-to late 1800s and housed in the Zoological Institute of the Russian Academy of Sciences (St. Petersburg, Russian Federation) were sampled. Extreme care was taken to minimize damage to the specimens, with sampling primarily conducted using a hand-held Dremel Moto-Tool. Cutting disks were replaced for each sample to prevent cross-contamination. Following collection, extracted samples were immediately placed in labeled bags. Accelerator mass spectrometry carbon-14 dating of ZI 6846 and three additional samples from the same collection ranged from ~680 to 1040 AD (R.D.E. MacPhee, unpublished data). Previously extracted DNA samples from two female dugongs (MD33 and MD118) that were collected along the coast of Australia in the Torres Strait in 1998 and 1999 were also included in the study (Table 1, Supplementary Table S1).

To minimize cross-species contamination, indexed DNA libraries suitable for Illumina sequencing were prepared from the Steller's sea cow extracts (see below) in a dedicated ancient DNA clean lab (University of York, UK), while the indexed dugong libraries were prepared at the University of Manitoba (Winnipeg, Canada). Two sets of experiments (in 2011 and 2013) were performed for most samples.

2.2. First round DNA extraction, library construction, enrichment and sequencing (2011)

Small (~250 mg) bone fragments were ground to powder in ancient DNA laboratories at the University of York (specimen ZI 6852) and at the University of Copenhagen (specimens ZI 6853 and ZI 17170(2)). DNA from the former sample was extracted following Rohland et al. (2009) while the latter samples were purified using the DNeasy Blood & Tissue kit (Qiagen, Valencia, CA). Extraction blanks, serving as negative controls, were treated in a similar manner throughout. Prior to DNA library construction, ~200 ng of DNA from each modern dugong sample was first

Table 1
Summary statistics of hybridization capture success for probes designed with *Dugong*, *Trichechus*, and *Loxodonta*/*Procavia* DNA sequences.

Probe: target	Base pairs captured	% Capture success	# Sequences aligning to target	Mean coverage depth
Dugong protein-coding				
ZI 6852	16,927	98.4	14,159	51.9
ZI 6853	6985	40.6	826	3.3
ZI 17170(2)	14,748	85.7	3645	11.1
Hydrodamalis consensus	17,084	99.3	18,630	66.4
Dugong MD33	17,204	99.9	74,174	291.1
Dugong MD118	17,206	100.0	61,410	225.1
Trichechus protein-coding				
ZI 6852	6921	97.4	6266	44.2
ZI 6853	3053	43.0	312	2.0
ZI 17170(2)	6005	84.5	1361	9.0
Hydrodamalis consensus	7005	98.6	7939	55.2
Dugong MD33	7097	99.9	18,815	141.7
Dugong MD118	7103	100.0	14,163	111.9
Trichechus UTRs				
ZI 6852	1466	89.6	622	19.5
ZI 6853	681	41.6	34	1.3
ZI 17170(2)	1139	69.6	179	5.9
Hydrodamalis consensus	1544	94.3	835	26.7
Dugong MD33	1593	97.3	1875	59.3
Dugong MD118	1591	97.2	1640	54.6
Loxodonta/*Procavia*				
ZI 6852	688	97.9	493	36.3
ZI 6853	369	52.5	63	5.9
ZI 17170(2)	559	79.5	140	10.9
Hydrodamalis consensus	696	99.0	696	53.1
Dugong MD33	695	98.9	1175	78.0
Dugong MD118	703	100.0	1066	76.6

fragmented to 100–400 base pair (bp) segments in 4.0 μl reactions, each containing 0.4 μl of NEBNext™ dsDNA Fragmentase (New England BioLabs), 1× Fragmentase Reaction Buffer and 100 μg/ml BSA. Each sample was incubated for 30 min at 37 °C, and 5 μl of 0.5 M EDTA added to stop the reaction. DNA samples were immediately purified with an Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Life Sciences).

Blunt end repair, adapter ligation, and adapter fill-in reactions were performed on both modern and ancient DNA samples following Meyer and Kircher (2010), with the following exceptions: all SPRI Bead reaction clean-up steps were replaced by spin column purification with the MinElute PCR purification kit (Qiagen) and less adapter mix was used during the ligation step of ancient samples (0.5 μM of each adapter, instead of 2.5 μM). A 10 μl aliquot of each modern library preparation was added to indexing PCR reactions consisting of 1× Phusion HF Buffer (Finnzymes), 200 μM of each dNTP, 200 nM of primer IS4_indPCR.P5 (Meyer and Kircher, 2010), 200 nM of the appropriate indexing primer (Supplementary Table S1) and 0.02 U/μl of Phusion High-Fidelity DNA Polymerase (Finnzymes) in a final volume of 50 μl. Reactions were cycled in an MJ Mini Gradient Thermal Cycler (Bio-Rad) with a temperature profile of: 98 °C for 30 s, followed by 12 cycles of 98 °C for 10 s, 60 °C for 20 s, and 72 °C for 20 s, ending with a final extension of 72 °C for 10 min. Amplified products were excised from a 2.5% agarose gel and purified with an Illustra™ GFX™ PCR DNA and Gel Band Purification Kit (GE Life Sciences). Purified PCR products were split into four equal volumes and re-amplified with primers IS5_reamp.P5 and IS6_reamp.P7 (Meyer and Kircher, 2010) as detailed above, and spin column purified with the MinElute PCR purification kit (Qiagen). Ancient library amplification was performed in 50 μl reaction volumes containing 1× AmpliTaq Gold Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 0.75 μM of each primer IS7 and IS8 (from Meyer and Kircher, 2010), and 0.05 U/μl AmpliTaq Gold (Applied Biosystems). Amplification was performed according to the following temperature profile: initial denaturation 94 °C for 10 min, followed by 35 cycles of 94 °C for 30 s, 60 °C for 45 s, and 72 °C for 45 s, with a final extension of 72 °C for 5 min. In order to increase starting template for the library while still maintaining library complexity, a subsequent re-amplification was performed in six parallel reactions with the same reaction mix and temperature profiles, with 5 μl of template library in each reaction. The resulting products were pooled and purified using Qiagen MinElute spin columns, according to the manufacturer's instructions.

Probes for the complete protein-coding region of the *ENAM* gene were designed from GenBank sequences for *Dugong dugon* (partial exon 9; 2751 bp), and Ensembl 60 sequences for *Procavia capensis* (exons 4, 6–8) and *Loxodonta africana* (exons 2–9; exon 9 included the entire coding region [2826 bp]). Probes designed from the latter two species included 25 bp of 5′ and 3′ flanking sequence for each exon. Probes were also designed from GenBank sequences for 20 additional exonic segments (dugong: *A2AB*, *APOB*, *BRCA1*, *BRCA2*, *DMP1*, *GHR*, *IRBP*, *RAG1* [partial cds between nucleotides 533-1333], *VWF*; manatee: *ADORA3*, *ADRB2*, *ATP7A*, *BCHE*, *BDNF*, *CNR1*, *EDG1*, *PNOC*, *RAG1* [partial cds between nucleotides 1762-2528)], *RAG2*, *TYR1*) and four untranslated regions (manatee: *APP*, *BMI1*, *CREM*, *PLCB4*). Repetitive elements were identified using RepeatMasker (www.repeatmasker.org) and excised, and Agilent's eArray web application (https://earray.chem.agilent.com) was then used to create a series of overlapping 60 bp oligos (tiled at 1 bp) across all nuclear targets. The final microarray design was imprinted on four identical 244K SureSelect microarrays (Agilent Technologies, Santa Clara, CA, USA).

Hybridization capture of samples ZI 6852, ZI 6853, and ZI 17170(2) was performed on individual arrays whereas the two dugong samples were pooled and hybridized to a single array following the protocol of Hodges et al. (2009), with the following modifications: (a) species-specific COT-I DNA was omitted from the hybridization mixture, and (b) after elution of the hybridized fragments, the mixture was purified using Qiagen MinElute columns. The resulting product was amplified in six parallel reactions for 20 cycles (reaction mix and temperature profiles as described above) with 10 μl starting template in each reaction. Final products were pooled and purified using Qiagen MinElute spin columns, according to the manufacturer's instructions.

After validating the success of the hybridization and amplification on a 2.5% agarose gel, indexing PCR was performed in four parallel reactions containing 1× AmpliTaq Gold Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 200 nM of primer IS4 (Meyer and Kircher, 2010), 200 nM of the appropriate indexing primer (Supplementary Table S1), and 0.05 U/μl AmpliTaq Gold (Applied Biosystems) in a final volume of 50 μl. Reactions were performed with the 20 cycle temperature profile described above, with 10 μl starting template in each reaction. The resulting products were pooled and purified using Qiagen MinElute spin columns, according to the manufacturer's instructions.

A 55-bp singleton sequencing protocol was performed on all products using two lanes of an Illumina GAIIx instrument (Ambry Genetics, Aliso Viejo, California), with raw reads subsequently trimmed of adapters and low quality bases using Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic). Dugong reads were mapped to manatee reference sequences using Geneious version R6.1 software (Biomatters Ltd., Auckland, New Zealand). Dugong and manatee sequences were then used separately as templates for the *Hydrodamalis* assemblies.

All assemblies were manually checked by eye to remove any remaining unedited adapter sequences. Assembled reads from the three extinct specimens generally exhibited low levels of DNA damage (C → U[T] and G → A) artifacts (Hofreiter et al., 2001; Briggs et al., 2007; Brotherton et al., 2007), especially specimen ZI 6852. Given the relatively high sequencing depth of most targets (Table 1), these artifacts were often manifested as polymorphic C/T or G/A positions that were subsequently scored as C or G; non-polymorphic C → T or G → A changes relative to dugong or manatee sequences were treated as genuine.

2.3. Second round DNA extraction, library construction, enrichment and sequencing (2013)

Two independent extractions were performed on five specimens ZI 6842, ZI 6844, ZI 6846, ZI 6852, and ZI 17170(2) following Rohland et al. (2009). An extraction blank, serving as a negative control, was treated in a similar manner throughout. Two independent libraries were constructed from each extraction (producing four independent libraries for each individual) and the extraction blank following the protocol of Meyer and Kircher (2010), with additional modifications to further facilitate the conversion of ancient DNA templates into libraries (Fortes and Paijmans, 2015) with double barcoding system as a means of detecting cross-contamination and PCR chimeras (Kircher et al., 2011).

Primary library amplification was performed in a 20 μl reaction volume, containing 1× AmpliTaq Gold Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 0.75 μM of each primer IS7 and IS8 (from Meyer and Kircher, 2010), and 0.05 U/μl AmpliTaq Gold (Applied Biosystems) and 6 μl starting template. Reactions were performed with the following temperature profile: initial denaturation 94 °C for 10 min, followed by 20 cycles of 94 °C for 30 s, 60 °C for 45 s, and 72 °C for 45 s, with a final extension of 72 °C for 5 min. Each library was amplified in two parallel reactions. Sample 17170(2) required additional re-amplification performed in eight parallel reactions for eight cycles.

The indexing PCR was performed in eight parallel reactions containing 1× AmpliTaq Gold Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 0.75 μM of primer IS4 (Meyer and Kircher, 2010), 0.75 μM of the appropriate indexing primer (Supplementary Table S2), and 0.05 U/μl AmpliTaq Gold (Applied Biosystems) in a final volume of 20 μl. Reactions were performed following the temperature profile described above using 10 cycles, with 4 μl starting template in each parallel reaction. Negative controls were carried throughout the library preparation process and, in addition to the extraction blank library, were subjected to hybridization capture and sequencing as described below.

Consensus *Hydrodamalis* sequences obtained from the first capture experiments were used to design a new set of probes for each target (plus 30 bp of 5′ and 3′ flanking sequence), with gaps filled using dugong (preferentially) and manatee sequences. Additional probes for *TTN*, *FBN1*, and *RAG1* [nucleotides: 1334-1761] designed from dugong and manatee sequences were also included. Target sequences were then examined for repetitive elements before imprinting on four identical 244K SureSelect microarrays using 1 bp tiling (see above).

Hybridization capture was performed according to Hodges et al. (2009), with the same modification as described above (Fortes and Paijmans, 2015). Both dugong samples (the same MD33 and MD118 libraries prepared for the first round of capture) were pooled on a single array, as were samples ZI 6852 and ZI 17170(2) and samples ZI 6844 and ZI 6846, while specimen ZI 6842 was captured individually. After elution of the hybridized fragments, amplification was performed without prior concentration in 24 parallel reactions containing 1× AmpliTaq Gold Buffer (Applied Biosystems), 2 mM $MgCl_2$, 0.1 mg/ml BSA, 0.25 mM of each dNTP, 0.75 μM of primer IS5 and IS6 (Meyer and Kircher, 2010), and 0.05 U/μl AmpliTaq Gold (Applied Biosystems), using 20 μl template in a final reaction volume of 40 μl. Reactions were performed with the following temperature profile: initial denaturation 94 °C for 10 min, followed by 20–25 cycles of 94 °C for 30 s, 60 °C for 45 s and 72 °C for 45 s, with a final extension of 72 °C for 5 min.

Hybridization capture was performed a second time to further enrich the libraries (Templeton et al., 2013). After the second hybridization experiment, samples were only amplified for 10 cycles. A 101-bp single-end sequencing protocol was performed on all products using two lanes of an Illumina HiSeq2500 instrument (University of California, Riverside). Post sequencing reads were demultiplexed by the P7 barcode using the script demultiplex.pl (https://code.google.com/p/gjl3-genome-diversity-tools/) and trimmed of the P7 adapter using Trimmomatic. *Hydrodamalis* reads were then demultiplexed by the P5 barcode using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), which was subsequently removed using Trimmomatic. Only sequence reads with matching P5 and P7 barcodes with insert sizes of >20 bp were used in further analyses (Supplementary Tables S2 and S3).

2.4. Afrotheria matrix

Gene and taxon sampling for the Afrotheria matrix employed the same gene segments as Meredith et al. (2011a) for the following 16 taxa: Sirenia (*Hydrodamalis gigas*, *Dugong dugon*, *Trichechus manatus*), Proboscidea (*Loxodonta africana*, *Elephas maximus*), Hyracoidea (*Heterohyrax brucei*, *Procavia capensis*), Tubulidentata (*Orycteropus afer*), Macroscelidea (*Elephantulus* [chimeric of *E. rufescens* and *E. edwardii*], *Rhynchocyon petersi*), Afrosoricida (*Amblysomus hottentotus*, *Chrysochloris asiatica*, *Echinops telfairi*, *Geogale aurita*, *Micropotamogale lamottei*, Oryzorictinae [chimeric of *Limnogale mergulus* and *Microgale talazaci*]). All nuclear sequences for *H. gigas* and *D. dugon* are new and have been deposited into GenBank under the accession numbers KR827244–KR827360. All of the non-sirenian sequences are from NCBI/Ensembl. Accession numbers for all sequences in the Afrotheria matrix are provided in Supplementary Table S4. The Afrotheria matrix in nexus format is available in Supplementary Table S5.

2.5. Enamelin matrix

Sequences for the complete coding region of the *ENAM* gene (exons 2–9) were extracted from NCBI and Ensembl 71 for 28 taxa and combined with new sequences for *Hydrodamalis gigas* and *Dugong dugon*. Sequence representation from NCBI and Ensembl included seven additional afrotherians (*Orycteropus afer* [ALYB01124786, ALYB01124787], *Loxodonta africana* [Ensembl 71], *Trichechus manatus* [AHIN01095756, AHIN01095757], *Elephantulus edwardii* [AMGZ01205762], *Procavia capensis* [ABRQ01122116, ABRQ01122117, Trace archives (ti 1295248989, ti 1297884274), ABRQ01122118], *Chrysochloris asiatica* [AMDV01278748, AMDV01278750], *Echinops telfairi* [AAIY02097299, GQ354864, AAIY02097300]), one xenarthran (*Dasypus novemcinctus* [AAGV03237580]), 12 laurasiatherians (*Condylura cristata* [AJFV01034075], *Erinaceus europaeus* [AMDU01090808, AMDU01090809], *Ceratotherium simum* [AKZM01002861], *Equus caballus* [Ensembl 71], *Bos taurus* [XM_002688339, Ensembl 71], *Orcinus orca* [ANOL02069129], *Camelus ferus* [AGVR01039102], *Pteropus alecto* [ALWS01074157], *Eptesicus fuscus* [ALEH01131032, ALEH01131031,

ALEH01131030], *Canis lupus* [XM_539305], *Ailuropoda melanoleuca* [Ensembl 71], *Felis catus* [Ensembl 71]), and eight taxa from Euarchontoglires (*Oryctolagus cuniculus* [Ensembl 71], *Ochotona princeps* [ALIT01098127], *Spermophilus tridecemlineatus* [Ensembl 71], *Heterocephalus glaber* [AHKG01099858], *Tupaia belangeri* [ALAR01020882, ALAR01020883], *Saimiri boliviensis* [PreEnsembl], *Homo sapiens* [NM_031889], *Otolemur garnettii* [Ensembl 71]). The enamelin matrix in nexus format is available in Supplementary Table S6.

2.6. Morphology matrix

The morphology matrix is an expanded version of Vélez-Juarbe et al.'s (2012) matrix and comprises 74 parsimony-informative characters and 42 taxa (Supplementary Table S7). The matrix includes a combination of binary, ordered multistate, unordered multistate, and stepmatrix characters (Supplementary Table S7). Taxon representation encompassed two proboscidean outgroups (*Phosphatherium* and Elephantidae [represented by *Loxodonta*]) and 40 sirenians including representatives of the extinct families Prorastomidae and Protosirenidae, and the extant families Trichechidae and Dugongidae. Ancestral character state reconstructions of morphological characters were performed with parsimony in Mesquite 2.75 (Maddison and Maddison, 2011).

2.7. Alignments

DNA sequences were aligned manually with Se–Al (Rambaut, 1996). Alignment-ambiguous regions of *BRCA2*, *ENAM*, *FBN1*, and *PLCB4* were excluded prior to phylogenetic analyses. The final alignment comprised 34,055 bp.

2.8. Phylogenetic analyses

The molecular data sets were analyzed with maximum likelihood and maximum parsimony, and the morphology data set was analyzed with maximum parsimony. Maximum likelihood analyses were performed with RAxML 8.0.9 (Stamatakis, 2006) on Cipres (Miller et al., 2010) with separate partitions for each gene in multigene analyses. Each gene was given its own GTR + Γ model of sequence evolution. Rapid bootstrap analyses (Stamatakis et al., 2008) were performed with 1000 replicates. The GTRGAMMA option in RAxML was used for both bootstrapping and final tree estimation. Maximum parsimony searches were performed with PAUP 4.0b10 (Swofford, 2002) and the minbrlen option for collapsing branches. A heuristic search for the shortest tree(s) employed 1000 randomized input orders with tree-bisection and reconnection (TBR) branch swapping. Parsimony bootstrap analyses on the molecular data set were performed with 1000 replicates, ten randomized input orders per replication, and TBR branch swapping. Bootstrap analyses with the morphological data set were performed with 100 replicates, ten randomized input orders per replicate, and TBR branch swapping.

2.9. Timetree analyses

Timetree analyses were performed with the mcmctree program in PAML 4.5 (Yang, 2007), which implements the relaxed clock MCMC algorithms of Rannala and Yang (2007). Analyses were performed with both autocorrelated and independent rates models. Each gene was allowed to have its own GTR + Γ model of sequence evolution. We set one time unit = 100 million years (Ma). Analyses were run with cleandata = 0. Shape (α) and scale (β) parameters for the gamma prior of the overall rate parameter μ (i.e., rgene_gamma in mcmctree) were 1 and 6.67, respectively. Calculations for the shape and scale parameters of the gamma prior for the rate-drift parameter assumed an age of 80.9 Ma for the most recent common ancestor of Afrotheria (average of eight analyses in Meredith et al., 2011a). RootAge was set at <0.809 in the control file. Chains were run for 100,000 generations after a burn-in of 10,000 generations, and were sampled every 20 generations. Analyses were performed with both hard-bounded and soft-bounded (SB) constraints and were run twice to check for convergence. Soft-bounded analyses allowed 2.5% of the prior distribution in each tail. Minimum ages were based on the oldest crown fossils that are assignable to each clade. Maximum ages were based on stratigraphic bounding, phylogenetic bracketing, and phylogenetic uncertainty (Reisz and Müller, 2004; Müller and Reisz, 2005; Benton and Donoghue, 2007; Meredith et al., 2010, 2011a; Springer et al., 2011). We followed Meredith et al. (2011a) for stratigraphic bounding except that individual stages from the Miocene were used instead of early, middle, and late Miocene. Stratigraphic bounds were extended by one stage for younger deposits (late Miocene) and by two stages for older deposits (middle Miocene and earlier) given that the fossil record becomes progressively less complete for earlier time periods. Phylogenetic bracketing (Reisz and Müller, 2004; Müller and Reisz, 2005; Meredith et al., 2010, 2011a; Springer et al., 2011) allowed for two successive outgroups following Meredith et al. (2011a). Stage boundaries are from the International Chronostratigraphic Chart v 2014/02 (www.stratigraphy.org, Cohen et al., 2013). We employed minimum and maximum constraints for nine nodes as outlined in Table 2.

2.10. Selection analyses

Branch and branch-site analyses that estimated the ratio (ω) of the non-synonymous substitution rate (dN) to the synonymous substitution rate (dS) were run with the Codeml program in PAML 4.5 (Yang, 2007). Analyses were performed with two different codon frequency models (CodonFreq = 2 [CF2] and CodonFreq = 3 [CF3]). CF2 employs equilibrium codon frequencies that are calculated from the average nucleotide frequencies at all three codon positions whereas equilibrium codon frequencies at the three codon positions are treated as free parameters with CF3. Branch analyses were performed with *ENAM* exon 9 sequences from the Afrotheria matrix and complete coding sequences for *ENAM* in the Enamelin matrix. We used the species tree in Fig. 1 for analyses with subsets of the Afrotheria matrix and a composite tree based on Meredith et al. (2011a) and Springer et al. (2012) for analyses with the complete coding sequences for *ENAM*. Branch analyses were performed with models M0 (one dN/dS ratio) and M2 with the following four (or five) branch categories: *Hydrodamalis*, stem dugongid branch, *Orycteropus*, *Dasypus* (only present in *ENAM* matrix), and all other branches (background). The dN/dS ratio was estimated separately for the stem dugongid branch because extensive tooth reduction is reconstructed to have evolved on this branch. Similarly, *Hydrodamalis*, *Orycteropus*, and *Dasypus* were given their own branch categories because these taxa lack teeth or enamel and have inactivating mutations in the *ENAM* gene (see below). Frameshift insertions were deleted prior to performing Codeml analyses. Similarly, stop codons were recoded as missing. We deleted a 540 bp repeated sequence in *Echinops ENAM* prior to running analyses with Codeml. We also performed dN/dS analyses on 20 protein-coding genes from the Afrotheria matrix (*ENAM* excluded) to determine if estimates of dN/dS on the *Hydrodamalis* and stem dugongid branches are upwardly biased owing to potential DNA damage artifacts. We used the same branch categories as above except that *Orycteropus* was included in the background category given that there are no pseudogenes in the Afrotheria matrix after excluding *ENAM*. Three taxa with high percentages (>50%) of missing

Table 2
Minimum and maximum ages (in millions of years) for nodes whose age was constrained in timetree analyses. Asterisks indicate nodes for which minimum and maximum ages are identical to Meredith et al. (2011a).

Calibrated node	Minimum age	Maximum age
1. Macroscelidea*	15.97 based on *Myohyrax* from the early Miocene (McKenna and Bell, 1997)	56.0 based on the phylogenetic bracketing/phylogenetic uncertainty (*Chambius* [Ypresian] is part of second outgroup to crown Macroscelidea in some phylogenetic analyses (Tabuce et al., 2001, 2008) although Cooper et al. (2014) recovered a deeper position for this taxon)
2. Chrysochloridae*	3.6 based on early Pliocene species of *Chrysochloris* (Asher and Avery, 2010)	33.9 based on phylogenetic bracketing (*Eochrysochloris* from the Rupelian is one of two chrysochlorid stem genera (Seiffert et al., 2007))
3. *Geogale* to Oryzorictinae	17.0 based on *Parageogale*, which is the oldest stem geogaline (Asher and Hofreiter, 2006)	28.1 based on stratigraphic bounding
4. Tenrecloidea	17.0 based on *Parageogale*, which is the oldest taxon with secure affinities in crown Tenrecoidea (Asher and Hofreiter, 2006)	59.2 Ma based on phylogenetic uncertainty, which allows for the possible inclusion of *Todralestes* (Selandian) in crown Tenrecoidea (Goswami et al., 2011); *Todralestes* is outside of Tenrecoidea in other analyses (Cooper et al., 2014; Manz et al., 2015)
5. Hyracoidea	6.08 based on *Dendrohyrax* fossils that have a minimum age of 6.08 (Ambrose et al., 2007; Pickford and Hlusko, 2007)	11.62 based on stratigraphic bounding
6. Proboscidea	6.8	11.62 based on stratigraphic bounding
7. Dugongidae	28.1 based on the inclusion of *Crenatosiren olsensi* (Rupelian) in crown Dugongidae (Vélez-Juarbe and Domning, 2015; this paper)	38.0 based on stratigraphic bounding
8. Sirenia	41.3 based on the inclusion of *Eotheroides aegyptiacum* (Lutetian) in crown Sirenia (Vélez-Juarbe et al., 2012; Vélez-Juarbe and Domning, 2014, 2015; this paper)	59.2 based on stratigraphic bounding
9. Paenungulata	56.0 Ma based on *Eritherium* (Thanetian), which is generally regarded as a stem proboscidean (Gheerbrant, 2009; Benton et al., 2015) although Cooper et al. (2014) recovered this taxon outside of Tethytheria in a paenungulate polytomy	66.0 based on stratigraphic bounding

sequences (*Geogale*, Oryzorictinae, *Micropotamogale*) were omitted from these analyses. Branch-site analyses (Yang et al., 2005; Zhang et al., 2005; Yang, 2007; Yang and dos Reis, 2011) on both *ENAM* data sets were performed with the stem dugongid and *Hydrodamalis* branches sequentially placed in the foreground. Branch-site analyses were performed with a modified version of model A and the corresponding null model (Yang et al., 2005; Zhang et al., 2005; Yang, 2007). Model A allows for a class of sites with dN/dS on the foreground branch, whereas dN/dS is fixed at one for these sites in the null model. The null distribution for this comparison is a 50:50 mixture of point mass 0 and X^2 with one degree of freedom, which yields critical values 2.71 at 5% and 5.41 at 1%. However, we followed Yang's (2007) recommendation and calculated P values with X^2 and one degree of freedom (i.e., no 50:50 mixture) to guard against possible violations of model assumptions.

3. Results

3.1. Hybridization capture results (2011)

A total of 24.7 and 40.0 million trimmed singleton reads were obtained from the two *Dugong dugon* and three *Hydrodamalis gigas* specimens, respectively. As expected, the percentage of these reads aligning to the manatee genome (TriManLat1.0) was much higher for modern (~70%) versus ancient (~8%) samples (Supplementary Table S8). Hybridization probes designed using dugong and manatee nuclear sequence were highly effective (96–100%) in capturing both modern (dugong) and ancient (Steller's sea cow) DNA sequences (Table 1). Mean sequencing depth and the number of sequences aligning to target, however, were much higher for the modern samples, while percent coverage, sequencing depth, and number of sequences aligning to target varied widely among the three *Hydrodamalis* samples. Consequently, 26,329 bp (of 26,649 bp targeted) and 26,640 bp of nuclear target sequence were obtained from *Hydrodamalis* and *Dugong*, respectively. Sequence coverage depth dropped sharply at the 5′ and 3′ ends of most gene targets (data not shown), which accounted for >50% (169 bp) of missing data for *Hydrodamalis*.

ENAM-specific probes exclusively designed from *Loxodonta africana* and *Procavia capensis* nucleotide sequences were successful in retrieving 99.0% (696 of 703 bp) and 100.0% coverage from the *Hydrodamalis* and *Dugong* specimens, respectively (Table 1); these gaps were subsequently filled by short-range PCR. Mean coverage depth varied greatly across the coding regions and was significantly correlated with both probe/target sequence similarity and probe GC content (Fig. 2).

3.2. Hybridization capture results (2013)

A total of 43.3 million reads were obtained from the two *Dugong* specimens, while 92.8 million reads contained P5 and/or P7 barcodes corresponding to the *Hydrodamalis* and blank libraries (Supplementary Table S2). As with the first run, the percentage of endogenous DNA content was much higher in the modern samples (Supplementary Table S3). The assembly for *Hydrodamalis* specimen ZI 6846 revealed evidence of contamination with dugong DNA and was thus excluded from all subsequent analyses. No cross-species contamination was evident for any of the other *Hydrodamalis* specimens. However, 4.4 million *Hydrodamalis* reads had mismatched P5 and P7 barcodes, while 263,755 reads were <20 bp, leaving 59.3 million reads available for subsequent analyses (Supplementary Table S3). As expected, sequence coverage depth for the two dugong samples was substantially higher (~8×) relative to the first sequencing run, and resulted in 100% coverage for all gene targets (Supplementary Table S1). Surprisingly, however, the number of *Hydrodamalis* sequences aligning to target was substantially reduced (7534 versus 28,100; *cf.*, Table 1 and Supplementary Table S1), resulting in a much lower percent coverage of all gene targets (34.1%; range = 0.0–87.2%). As with the first run, inter-individual variability was high with specimen ZI 6852 again exhibiting the highest number of sequences aligning to target (4957), and hence sequence coverage (30.0%). These differences are not attributable to average trimmed read

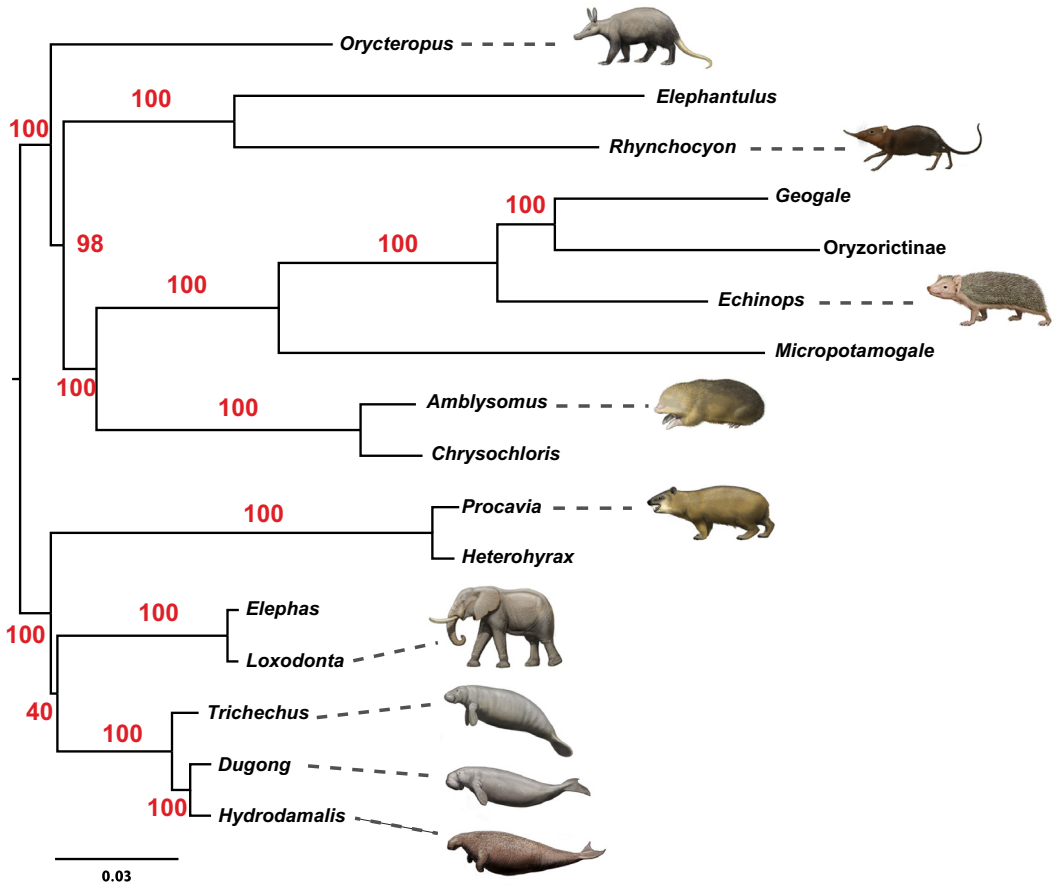

Fig. 1. RAxML tree with branch lengths in substitutions per site. Bootstrap support percentages are shown above or adjacent to branches.

lengths, which varied little among the four specimens (range: 50–59 bp); mean trimmed read lengths for the two dugongs were 84 and 88 bp, respectively (Supplementary Table S1). The extraction and library blanks produced 72,742 trimmed reads >20 bp of which only 57 (0.08%) aligned to the 26 target loci (data not shown).

A total of 30,004 bp of novel target sequence was obtained from the five *Hydrodamalis* specimens across the two sequencing runs, while 34,493 bp were obtained from the two dugongs.

3.3. Phylogenetic analyses of Afrotheria matrix

Fig. 1 shows the RAxML tree (optimized $\ln L = -146066.375388$) based on the concatenated data set (34,055 bp) with separate GTR + Γ partitions for each gene. All clades were recovered with 100% bootstrap support except for Tethytheria, which was recovered with 40% bootstrap support, and Afroinsectivora (i.e., Afrosoricida + Macroscelidea), which received 98% bootstrap support. A maximum parsimony analysis with branch and bound recovered a single tree (21,270 steps, retention index = 0.71) that was identical to the ML tree. MP bootstrap analyses recovered all clades with 100% bootstrap support except for Tethytheria (65%) and Afroinsectivora (84%). Within Sirenia, *Dugong* and *Hydrodamalis* are sister taxa to the exclusion of *Trichechus* with 100% support in both ML and MP bootstrap analyses.

3.4. Phylogenetic analyses of morphology matrix

Parsimony analyses of the morphology matrix resulted in 12 trees at 239 steps. The strict consensus of these trees with bootstrap support percentages is shown in Fig. 3. Prorastomidae (*Prorastomus* and *Pezosiren*) are paraphyletic at the base of Sirenia and are excluded from more crownward sirenians, which are united together with 90% bootstrap support. Protosirenidae (*Ashokia* and *Protosiren*), in turn, are paraphyletic at the base of remaining Sirenia, which cluster together with 96% bootstrap support. Trichechidae includes the extant genus *Trichechus* and the extinct genera *Miosiren* and *Anomotherium*. Among *Trichechus* spp., *T. inunguis* is the sister taxon to *T. manatus* + *T. senegalensis*. Within Dugongidae, *Eotheroides*, *Halitherium*, and *Priscosiren* are stem taxa to crown Dugongidae, which includes the reciprocally monophyletic clades Dugonginae (with *Dugong*) and Hydrodamalinae (with *Hydrodamalis*) (Fig. 3). However, there is only 40% bootstrap support for the clade that includes crown Dugongidae, *Halitherium*, and *Eotheroides* to the exclusion of Trichechidae.

3.5. Timetree analyses

Fig. 4 shows the results of a molecular dating analysis with the autocorrelated rates model in conjunction with hard-bounded constraints. The divergence between *Hydrodamalis* and *Dugong* is placed at 28.6 Ma (95% credibility interval = 28.1–29.9 Ma), which is close to the minimum calibration time (28.1 Ma) for this node (Fig. 4, Table 3). Similarly, a divergence date of 41.6 Ma (95% credibility interval = 41.3–42.2 Ma) for the split between Trichechidae and Dugongidae is only slightly older than the minimum calibration time (41.3 Ma) for this cladogenic event (Fig. 4, Table 3). Finally, the basal split in Tethytheria (Sirenia to Proboscidea) was estimated at 65.0 Ma (95% credibility interval = 63.9–65.8 Ma) in

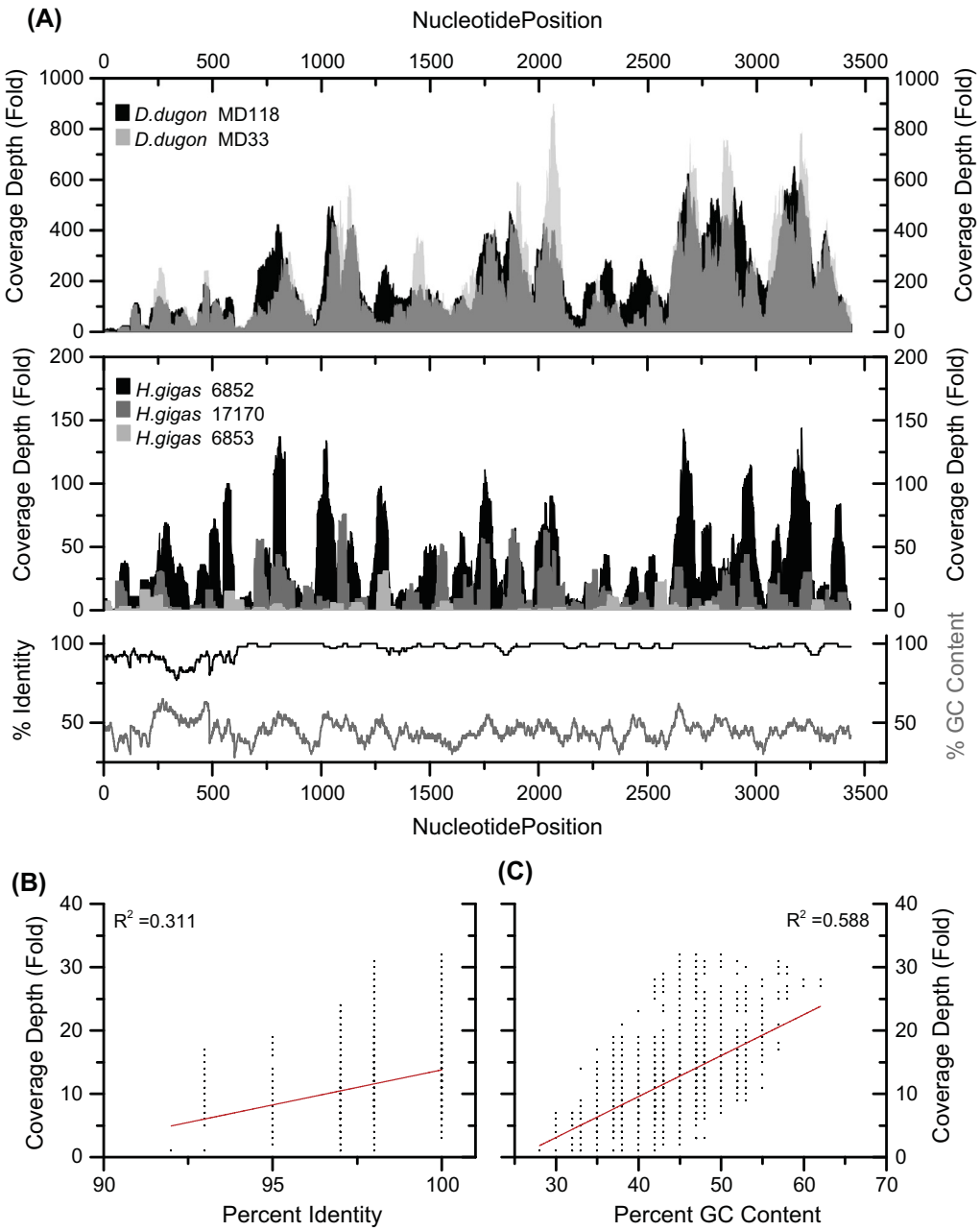

Fig. 2. (A) Sequence coverage depth of the complete coding sequence of *ENAM* for two *Dugong dugong* and three *Hydrodamalis gigas* specimens as a function of % *Hydrodamalis* sequence identity with probe sequence and probe % GC content. The latter two variables represent sliding 60 base-pair averages. The dependence of *Hydrodamalis* consolidated coverage on % identity (B) and % GC content (C) is shown only for the ninth exon of *ENAM* (nucleotides 601-3454). To eliminate potential post-capture PCR bias from plots (B) and (C), duplicate reads were removed from these analyses using Picard Tools v.1.128 (http://broadinstitute.github.io/picard/). Analyses of variance confirm that the linear regressions (red line) for (B) and (C) have a slope significantly different than zero ($p \leqslant 0.05$). This same result was recovered with PCR duplicates included in the analysis (data not shown). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the early Paleocene. Timetree analyses with independent rates and hard-bounded constraints resulted in divergence estimates for *Hydrodamalis* to *Dugong* (29.8 Ma, 95% credibility interval = 28.1–34.4 Ma), Trichechidae to Dugongidae (41.6 Ma, 95% credibility interval = 41.3–42.4 Ma), and Sirenia to Proboscidea (64.3 Ma, 95% credibility interval = 62.3–65.6 Ma) that are similar to the estimates that were obtained with autocorrelated rates and hard-bounded constraints (Table 3). Soft-bounded analyses resulted in dates that are 1.4–1.6 Ma younger for *Hydrodamalis* to *Dugong*, 0.1 Ma younger (independent rates) or 1.1 Ma older (autocorrelated rates) for Tethytheria, and equivalent (independent rates) or 1.1 Ma older (autocorrelated rates) for Paenungulata than dates with hard-bounded constraints (Table 3). Dates for Afrotheria and within Afroinsectiphilia are consistently older with soft-bounded constraints than with hard-bounded constraints.

3.6. Inactivating mutations

There were no frameshifts or stop codons in the *Hydrodamalis* coding sequence, but a transversion mutation (AG to CG) was detected in the acceptor splice site of intron 2. Although this type

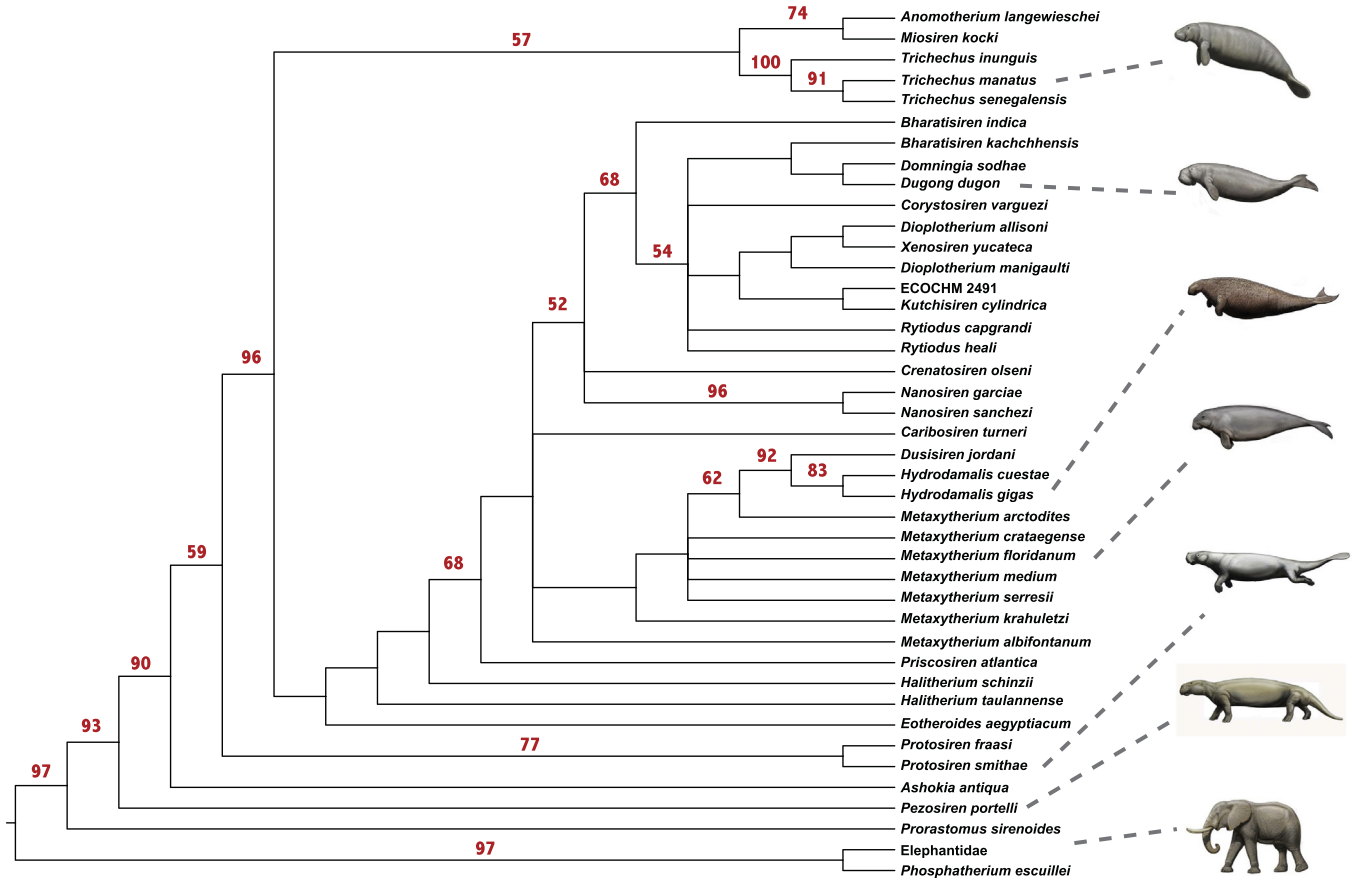

Fig. 3. Strict consensus of 12 trees (239 steps each) based on the morphology matrix. Bootstrap support percentages are shown for clades that were supported at or above 50%. The tree topology within Dugonginae (the smallest clade comprising *Bharatisiren*, *Crenatosiren*, *Nanosiren*, and other taxa) is unstable in recent analyses (Vélez-Juarbe et al., 2012; Vélez-Juarbe and Domning, 2015) and requires further investigation.

(A → C) of replacement is inconsistent with ancient DNA damage artifacts (Briggs et al., 2007; Brotherton et al., 2007), this finding was based on only two reads (one of which was a PCR duplicate) that spanned this region from a single specimen (ZI 6852). We thus confirmed this splice site mutation via PCR on a ZI 6852 DNA library (data not shown). Multiple inactivating mutations occur in the *Orycteropus ENAM* sequence (Meredith et al., 2014). Inactivating mutations in exon 7 include an AfroSINE (Nikaido et al., 2003), a single bp frameshift deletion, and a stop codon (NCBI ALYB01124786); inactivating mutations in exon 9 include two frameshift deletions and one frameshift insertion (NCBI ALYB01124787). Finally, there are three single-base frameshift mutations in exon 9 of *Dasypus*, although all three mutations occur near the 3′ end of this exon (NCBI AAGV03237580).

3.7. Selection analyses

Branch analyses with *ENAM* and two codon frequency models (CF2, CF3) provide statistically significant support for the M2 model with four branch categories (exon 9, Afrotheria matrix) or five branch categories (complete *ENAM*, 30 placentals) relative to the M0 model with a single ω value (Table 4). dN/dS values on the *Hydrodamalis*, stem dugongid, *Orycteropus*, and *Dasypus* branches are all elevated above the background ω value, although only the stem dugongid branch has an ω value that is consistently >1 (Table 4). Results for protein-coding sequences of the Afrotheria matrix excluding *ENAM* indicate that dN/dS ratios on the *Hydrodamalis* (0.51, 0.51) and stem dugongid (0.48, 0.47) branches are only slightly elevated relative to the median ω value (0.39, 0.39) on other branches of the tree and are similar to or lower than some of the other dN/dS values, e.g., *Procavia* (0.54, 0.54), stem Afroinsectivora (0.48, 0.44), and stem Afrosoricida (1.08, 1.10). Branch-site analyses with exon 9 (Afrotheria matrix) and full protein-coding sequences (Enamelin matrix) of *ENAM* both provide support for positive selection on the stem dugongid branch (Table 5) including one site with a significant probability ($p > 0.95$) of membership in the positive selection ($\omega > 1$) bin.

4. Discussion

4.1. Gene capture with phylogenetically divergent probes

To our knowledge, this is the first study to demonstrate both interfamilial and interordinal gene capture from ancient DNA samples, though successful confamilial capture has been reported for both older primate (50,000 year old Neanderthal; Burbano et al., 2010) and more recent (47–170 year old museum specimens) dermopteran samples (Mason et al., 2011). Most of the bait sequences that were employed to capture gene segments from *Hydrodamalis gigas* were designed using homologous coding segments from *Dugong dugon* and *Trichechus manatus*, which diverged from *H. gigas* at least as far back as the mid-Oligocene and mid-Eocene, respectively. However, owing to the overall high sequence similarity between *Hydrodamalis* and these extant species (>98%), we were able to retrieve 98.8% of targeted sequence in the first run. Although the wide variability in sequence coverage/depth among the *Hydrodamalis* specimens led us to modify our sampling (e.g., multiple extractions per individual) and hybridization capture

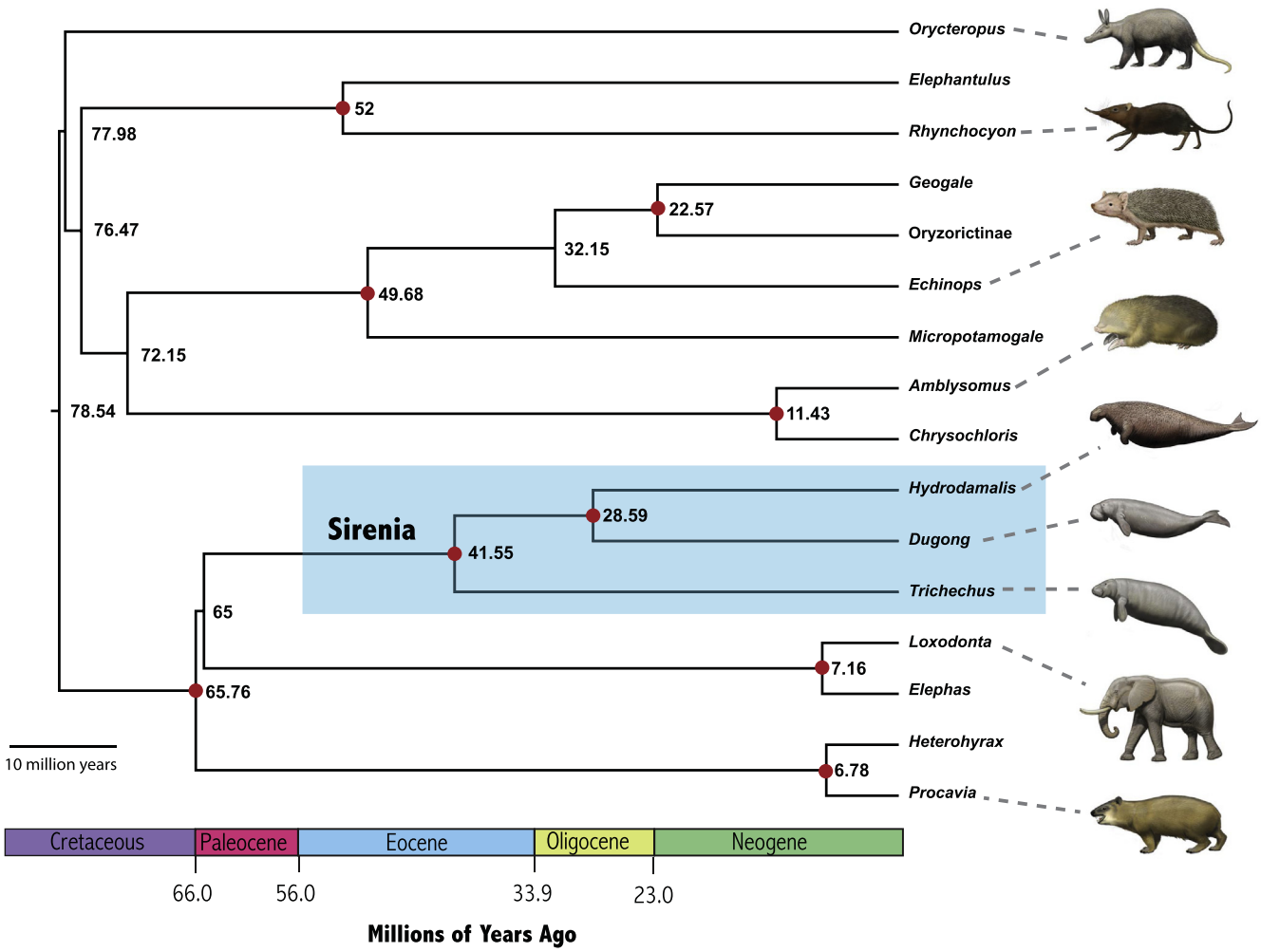

Fig. 4. MCMCTREE timetree based on autocorrelated rates and hard-bounded constraints for nine nodes (red circles) (see Table 2). Divergence dates at nodes are in millions of years. Credibility intervals (95%) and timetree dates with different combinations of evolutionary rate model (autocorrelated, independent) and constraint type (hard-bounded, soft-bounded) are provided in Table 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

procedures, the much lower sequence coverage (~34%) obtained in the second run is surprising and indicates that additional research is required to better understand the parameters influencing hybridization capture success with ancient samples. The presence of dugong contamination in a single sample (ZI 6846) of the second round capture experiments was also unexpected given that extraction and library construction of modern and ancient samples were conducted in separate institutions (on separate continents). The lack of dugong barcode adapter sequences in any of these reads suggests this contamination was introduced prior to specimen sampling (e.g., during previous handling of modern and extinct specimens in the museum). This conclusion is bolstered further by the absence of noticeable dugong contamination in the other five *Hydrodamalis* specimens included in the first and second round capture experiments, and by the very low number of reads in the blank libraries aligning to target.

To optimize target sequence retrieval from degraded samples we recommend the use of dense sequence tiling (i.e., 1 bp spacing between probes) (Ávila-Arcos et al., 2011) and inclusion of flanking intron sequences (or oligo stacking at the 5′ and 3′ ends of each target). This strategy was successful in obtaining all but seven bp of coding sequence (exons 2–8) of *ENAM* from ~1000 year old Steller's sea cow DNA using bait that was designed from *Procavia capensis* (rock hyrax) and *Loxodonta africana* (African elephant) sequences, despite sequence divergences of up to 13% between target and bait. Proboscidea and Sirenia have a most recent common ancestor that is at least as old as the early late Paleocene based on the age of the fossil proboscidean *Eritherium* (Gheerbrant, 2009), and demonstrate the utility of employing bait sequences from taxa in different families or even different orders when more closely related reference genomes are unavailable (see also Hedtke et al., 2013; Li et al., 2013). Notably, extant proboscideans and sirenians have large body sizes and slow rates of molecular evolution relative to many other mammals (Meredith et al., 2011a) and it is probable that interordinal capture success will be lower using bait designed from lineages containing smaller taxa with faster rates of molecular evolution (e.g., elephant shrew baits to capture tenrec sequences). Nevertheless, taxa such as Perissodactyla and Cetacea have relatively slow rates of molecular evolution and probe sequences from a single reference genome in each clade may be effective for capturing homologous sequences from even distantly related members within each group. In the case of taxa with faster rates of molecular evolution, it may still be possible to obtain high nuclear and mitochondrial sequence coverage from confamilial and congeneric relatives, respectively, especially in view of higher GC content in smaller species (Romiguier et al., 2010).

Table 3
Divergence time estimates (posterior mean and 95% credible intervals) in millions of years based on four different combinations of evolutionary rate model (autocorrelated rates [AUTO], independent rates [IR]) and calibration type (hard bounded [HARD], soft bounded [SOFT]). All of the divergence times reported below are based on analyses with constraints for nine nodes (Table 2).

Clade	AUTO, HARD	IR, HARD	AUTO, SOFT	IR, SOFT
Hydrodamalis + *Dugong*	28.6 (28.1–29.9)	29.8 (28.1–34.4)	27.0 (24.0–28.6)	28.4 (27.0–30.4)
Sirenia	41.6 (41.3–42.2)	41.6 (41.3–42.4)	35.3 (32.0–38.7)	34.1 (31.0–38.1)
Proboscidea	7.2 (6.8–8.1)	7.0 (6.8–7.6)	6.1 (5.4–7.0)	5.8 (4.9–6.5)
Tethytheria	65.0 (63.9–65.8)	64.3 (62.3–65.6)	66.1 (64.3–68.8)	64.2 (61.4–66.1)
Hyracoidea	6.8 (6.1–8.4)	6.6 (6.1–7.9)	7.3 (6.5–8.9)	7.1 (6.4–8.2)
Paenungulata	65.8 (65.2–66.0)	65.3 (63.5–66.0)	66.9 (65.4–69.5)	65.3 (62.7–66.8)
Macroscelidea	52.0 (46.6–55.7)	54.3 (50.4–56.0)	53.0 (47.0–57.1)	56.8 (52.5–63.3)
Geogale + Oryzoryctinae	22.6 (17.6–27.5)	25.3 (19.9–28.0)	22.8 (17.4–27.8)	25.7 (20.3–28.5)
Tenrecinae	32.2 (27.0–37.4)	38.0 (30.3–45.4)	32.4 (27.9–37.2)	37.9 (30.5–46.0)
Tenrecidae	49.7 (43.4–56.0)	56.5 (50.2–59.1)	50.2 (43.9–56.7)	58.2 (51.1–63.4)
Chrysochloridae	11.4 (7.6–16.9)	15.6 (10.7–21.2)	11.5 (7.6–16.9)	15.1 (11.1–20.2)
Afrosoricida	72.2 (68.7–75.3)	82.6 (75.5–90.4)	73.7 (69.4–77.7)	85.1 (77.4–93.2)
Afroinsectivora	76.5 (73.2–79.3)	88.7 (81.4–96.8)	78.1 (73.7–82.2)	91.2 (82.8–99.4)
Afroinsectiphilia	78.0 (74.7–80.8)	90.7 (83.3–98.9)	79.7 (75.3–83.8)	93.1 (85.6–101.2)
Afrotheria	78.5 (75.7–81.1)	91.6 (84.4–99.6)	80.3 (76.6–84.2)	93.9 (86.5–101.9)

Table 4
Summary of dN/dS analyses (branches) on *ENAM* with codon frequency models CF2 and CF3. In each comparison below, M0 is the null model and M2 is the model that allows selected individual branches to have their own dN/dS ratio.

Model	Branches	CF2		CF3	
		ln*L*	dN/dS	ln*L*	dN/dS
1. Partial exon 9 of ENAM (Afrotheria matrix)					
M0	All branches	−12810.25	0.43	−12790.14	0.49
M2		−12800.25[a]		−12779.98[b]	
	Background		0.40		0.46
	Orycteropus		0.71		0.79
	Stem Dugongidae		4.08		6.07
	Hydrodamalis		1.16		1.30
2. Complete ENAM coding sequence (ENAM matrix)					
M0	All branches	−41154.38	0.47	−41303.42	0.53
M2		−41146.28[c]		−41295.50[d]	
	Background		0.46		0.52
	Orycteropus		0.61		0.70
	Stem Dugongidae		1.25		1.75
	Hydrodamalis		0.74		0.84
	Dasypus		0.67		0.74

[a] M2 significantly better than M0 (DF = 3, $p = 0.00017$).
[b] M2 significantly better than M0 (DF = 3, $p = 0.00015$).
[c] M2 significantly better than M0 (DF = 4, $p = 0.0028$).
[d] M2 significantly better than M0 (DF = 4, $p = 0.0033$).

Table 5
Summary of branch-site dN/dS analyses on the stem dugongid branch for *ENAM* with codon frequency models CF2 and CF3. Model A allows for a class of sites with dN/dS on the foreground branch, whereas dN/dS is fixed at one for these sites in the null model.

Model	CF2		CF3	
	ln*L*	Positively selected sites (BEB)	ln*L*	Positively selected sites (BEB)
1. Partial exon 9 of ENAM (Afrotheria matrix)				
Null	−12746.48		−12723.68	
Model A	−12743.18[a]		−12719.62[b]	
		25A		25A
		75P		75P
		226V		226V
		228N		228N
		334Q		334Q
		392S		392S
		651D		651D
		696S*		696S*
		805E		805E
		886K		886K
		892E		892E
		916V		916V
2. Complete ENAM coding sequence (ENAM matrix)				
Null	−40612.41	120K	−40746.70	120K
Model A	−40610.29[c]	466G	−40743.71[d]	264A
		468N		466G
		687S		468N
		950D		592Q
		995S		
		1105E		
		1185K		687S
		1191G		950D
		1215V		995S*
				1105E
				1185K
				1191G
				1215V

[a] Model A significantly better than null model (DF = 1, $p = 0.010$).
[b] Model A significantly better than null model (DF = 1, $p = 0.0043$).
[c] Model A significantly better than null model (DF = 1, $p = 0.039$).
[d] Model A significantly better than null model (DF = 1, $p = 0.015$).
* Significant at 0.05.

4.2. Phylogenetic analyses

Analyses with a molecular data set that includes representatives of all afrotherian orders resulted in a phylogenetic tree that is in excellent agreement with previous analyses (Murphy et al., 2001; Meredith et al., 2011a). Paenungulata is strongly supported, but the paenungulate trichotomy (Amrine and Springer, 1999) is still not resolved by molecular data. Within Sirenia, we find robust support for the monophyly of crown Dugongidae (i.e., *Dugong* and *Hydrodamalis*) to the exclusion of Trichechidae (*Trichechus*). This result agrees with previous molecular studies on the basis of albumin immunology (Rainey et al., 1984) and *CYTB* sequences (Ozawa et al., 1997), but goes beyond these studies in providing robust bootstrap support for Dugongidae monophyly. Analyses of the morphological data matrix also support an association of *Dugong* and *Hydrodamalis* to the exclusion of *Trichechus*. Morphological analyses may be misleading if there is a strong signature of ecomorphological convergence, as occurs across diverse placental orders (Springer et al., 2007, 2008, 2013), but our data provide no evidence of conflict between molecules and morphology for sirenian genera. A caveat is that direct comparisons between molecular and morphological data for Sirenia are limited to three genera. By contrast with our morphological data set, Voss (2013)

recovered an association of *Trichechus* and *Hydrodamalis* to the exclusion of *Dugong* based on a different morphological character matrix. However, Voss' (2013) result is strongly contradicted by our molecular results, which provide 100% bootstrap support for *Hydrodamalis* and *Dugong* to the exclusion of *Trichechus*.

Analyses with the morphological dataset suggest that Prorastomidae (*Prorastomus* and *Pezosiren*) are paraphyletic at the base of Sirenia. Protosirenids, in turn, are crownward of prorastomids and paraphyletic at the base of the remaining sirenians. The affinities of *Eotheroides* and *Halitherium*, which lie further crownward, are on the stem dugongid branch on the most parsimonious morphological trees, albeit with weak bootstrap support. Placement of these taxa on the dugongid stem suggests that the minimum age for the split between Dugongidae and Trichechidae is middle Eocene (Lutetian) based on the age of *E. aegyptiacum*, which predates our timetree analysis by 5–10 Ma (see below). Finally, *Anomotherium* and *Miosiren* were recovered as stem trichechids.

Another phylogenetic analysis of Sirenia (Sagne, unpublished doctoral dissertation, 2001) obtained different results that nested Trichechidae within a paraphyletic Protosirenidae to the exclusion of Dugongidae. None of the previous morphological analyses by Domning or Vélez-Juarbe (Domning, 1994; Vélez-Juarbe et al., 2012; Vélez-Juarbe and Domning, 2014, 2015), which employed character sets somewhat different from Sagne's, recovered a similar result. For example, Domning (1994) found Trichechidae nested well within a paraphyletic Dugongidae. The present phylogenetic hypothesis, wherein a Trichechidae + Dugongidae clade is rooted within a paraphyletic Protosirenidae, is a step closer to Sagne's phylogeny in that trichechids are no longer an offshoot of early dugongids. Given the nonexistent fossil record of pre-late Oligocene trichechids, and the limited sampling of protosirenids, future fossil discoveries may reveal an association of trichechids with at least some protosirenids to the exclusion of dugongids, as suggested by Sagne (2001). Diedrich (2013) suggested an even deeper split for trichechids and dugongs and hypothesized that fully aquatic trichechids and dugongids evolved independently from quadrupedal prorastomids in the New World and protosirenids in the Old World, respectively. However, this hypothesis was not based on a formal cladistic analysis and broadly conflicts with published studies including analyses presented here.

4.3. ENAM evolution

Branch analyses suggest that *ENAM* evolved under positive selection on the stem dugongid branch. The distribution of enamel types between the enamel–dentine junction and the enamel surface in an individual tooth ("schmelzmuster") is known to vary along the tooth row (Koenigswald and Clemens, 1992; Koenigswald, 1997; Mathur and Polly, 2000). Positive selection on the stem dugongid branch may have occurred in conjunction with changes in the feeding apparatus that included extensive tooth reduction, i.e., loss of the incisors (except for I1), canines, and permanent premolars. Branch-site analyses suggest that positive selection occurred at 12 codon sites, although only one site has a probability >0.95 for inclusion in the positive selection bin. Al-Hashimi et al. (2009) identified 19 codon sites in *ENAM* that have evolved under positive selection in Mammalia, but there is no overlap between Al-Hashimi et al.'s (2009) positively selected sites and the 12 sites that may have evolved under positive selection on the stem dugongid branch. However, this lack of overlap is perhaps not surprising given that Al-Hashimi et al. (2009) performed a dN/dS site analysis on a mammalian data set that included only three afrotherians (tenrec, elephant, hyrax). By contrast, we performed branch-site analyses that targeted the stem dugongid branch, which was missing from Al-Hashimi et al.'s (2009) study. We also note that the lack of data on the precise relationship between enamel proteins and the structure of mature enamel precludes pinpointing correlations, if they exist, between positively selected sites and enamel or diet (Al-Hashimi et al., 2009). By contrast with the stem dugongid branch, we did not find any evidence for positive selection on the *Hydrodamalis gigas* branch. Rather, the dN/dS ratio is not significantly different than 1, which suggests that *ENAM* has evolved neutrally on this branch. There are no frameshift mutations or stop codons, but the occurrence of a splice site mutation (AG to CG) may inactivate this gene by abrogating the production of a functional mRNA. By contrast with the "CG" splice site sequence in the edentulous *H. gigas*, the canonical AG splice site is widely conserved across placental taxa with enamel-capped teeth for which genome sequences are available (data not shown). The occurrence of a splice site mutation in *H. gigas* is not unexpected given that teeth were presumably lost in this lineage during the Miocene, and are absent in both adult and juvenile Steller's sea cows (Domning, 1978). The occurrence of inactivating mutations in the *ENAM* gene of two afrotherians (*Hydrodamalis gigas* and *Orycteropus afer*) provides additional support for the congruence of genomic and fossil data pertaining to patterns of tooth loss/enamel loss in mammals and other edentulous vertebrates (Meredith et al., 2009, 2011b, 2013, 2014).

4.4. Timetree analyses

Timetree estimates are generally consistent with previous molecular dating analyses (Meredith et al., 2011a), although in some cases our dates are slightly younger or slightly older than previous estimates. Dates for the last common ancestor of Paenungulata are in the range of 65.3–66.9 Ma and are similar to Meredith et al.'s (2011a) mean estimate (64.3 Ma) based on eight different analyses. However, our dates for crown Sirenia range from 34.1 to 41.6 Ma and are older than Meredith et al.'s (2011a) mean estimate of 31.4 Ma (range = 29.3–32.2 Ma). Rainey et al. (1984) suggested an even younger split for Trichechidae and Dugongidae (17–20 Ma) based on an albumin molecular clock. Rainey et al. (1984) also suggested a relatively young date for *Hydrodamalis* to *Dugong* (4–8 Ma) whereas our dates (27.0–29.8 Ma) are more in line with Ozawa et al.'s (1997) estimate of 22 Ma based on a *CYTB* clock.

Molecular dating analyses with our Afrotheria supermatrix also demonstrate the importance of employing multiple fossil calibrations in relaxed clock analyses with taxa that have a wide range of body sizes and molecular rates of evolution (Meredith et al., 2011a). Timetree analyses wherein all calibrations were omitted except for a single constraint, either Dugongidae (minimum = 28.1 Ma, maximum = 38.0 Ma) or *Geogale* + Oryzorictinae (minimum = 17.0 Ma, maximum = 28.1 Ma), resulted in a wide range of dates for all nodes within Afrotheria (Supplementary Table S9). When Dugongidae was the only constrained node, timetree estimates were unrealistically old across most other nodes in Afrotheria (e.g., Paenungulata = 138.9–142.7 Ma, Macroscelidea = 108.7–126.6 Ma, Tenrecidae = 102.5–130.0 Ma). By contrast, nodes within Paenungulata were consistently younger when the only constraint was within Afrosoricida. For example the dates for Dugongidae (*Dugong* and *Hydrodamalis*) and Sirenia were estimated at 2.7–5.5 Ma and 4.7–8.4 Ma, respectively, when the only constraint was *Geogale* + Oryzorictinae. The general pattern is one in which divergence estimates become too young in larger-bodied clades when the only constrained node is in a smaller-bodied clade (i.e., *Geogale* + Oryzorictinae), whereas divergence estimates become too old in smaller-bodied clades when the only constrained node is in a larger-bodied clade (i.e., Dugongidae). This result may be expected given that rates of molecular evolution

are generally faster in smaller mammals with shorter generation times than larger mammals with longer generations times (Martin and Palumbi, 1993). This problem is partly mitigated in analyses with multiple constraints that are spread through the tree (Meredith et al., 2011a), but there is still a tendency for timetree estimates at constrained nodes to push up against (hard-bounded analyses) or even through (soft-bounded analyses) minimum constraints in large-bodied clades. This finding also allows for the possibility that the middle Eocene taxon *Eotheroides aegyptiacum* belongs to crown Sirenia as suggested by cladistic analyses of the morphological data set even though timetree estimates for crown Sirenia suggest this taxon is on the sirenian stem. We employed a minimum age for crown Sirenia based on the age of *E. aegyptiacum*, which is known from the Lutetian (47.8–41.3 Ma), but as for other constraints used the top of this stage (i.e., 41.3 Ma) rather than the base of this stage for the minimum age.

The impact of widely varying rates of molecular evolution on estimates of deep divergences within Placentalia (e.g., placental root, Afrotheria, Boreoeutheria, Euarchontoglires) remains unclear. The inclusion of multiple constraints that are spread throughout the tree is essential for improving the accuracy of timetree estimates with relaxed clock methods when there is strong variation in rates of molecular evolution. Also, it has commonly been assumed that early placental mammals had small body sizes that were similar to mice or shrews, and by implication fast rates of molecular evolution (Feldhamer et al., 2007). By contrast, Romiguier et al. (2013) suggested that early placental mammals were larger than mice or shrews, and were ~1 kg based on ancestral reconstructions of genome properties that are highly correlated with life history traits in extant species. Romiguier et al.'s (2013) results suggest that the limitations of relaxed molecular clocks for dating Placentalia and its subclades may be most pronounced in crownward clades with larger body sizes and slower rates of molecular evolution or smaller body sizes and faster rates of evolution than the placental ancestor.

4.5. Sirenian macroevolution

Fossil and living cetaceans provide a model system for understanding macroevolutionary changes associated with the transition from a terrestrial environment to an aquatic environment. For example, fossil cetaceans document hind limb loss and the transformation of the front limbs into flippers on the Cetacea stem, as

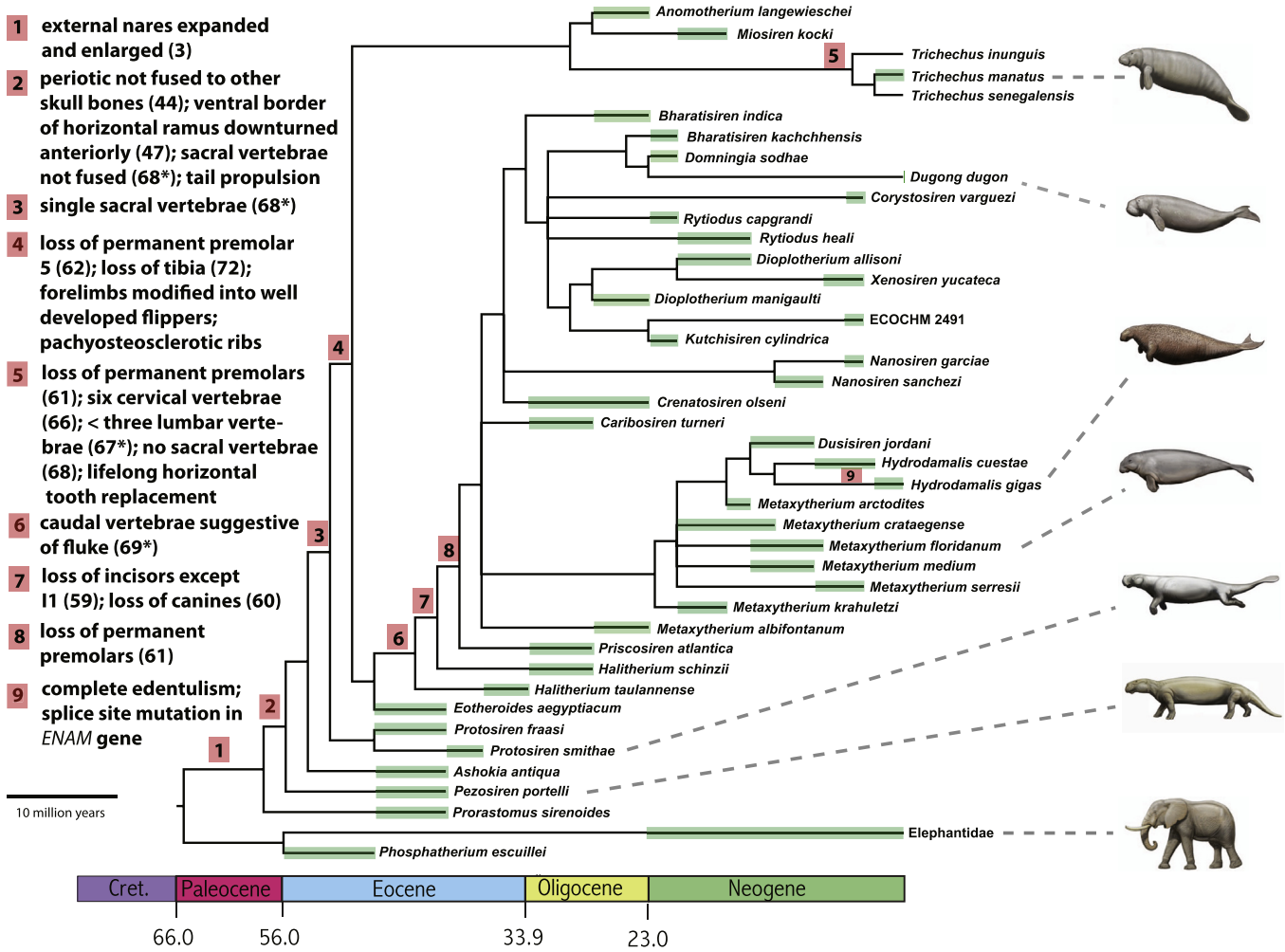

Fig. 5. Overview of macroevolutionary changes in Sirenia associated with the transition from terrestrial to fully aquatic forms. Key characters are mapped onto the strict consensus tree shown in Fig. 3. Approximate taxon ranges are shown in green bars. Black lines without green bars represent ghost lineages that are implied by known fossil ranges. These ghost lineages were arbitrarily extended by ~2 million years, when necessary, to avoid artificial polytomies, e.g., the common ancestral branch leading to *Trichechus manatus* + *T. senegalensis* and the temporally equivalent portion of the external branch leading to *T. inunguis*. The divergence date for Sirenia to Proboscidea was fixed to agree with the divergence date in Fig. 4. Ancestral character states were reconstructed with Mesquite 2.75 (Maddison and Maddison, 2011). Asterisks denote transformations that may have occurred at a deeper node that was reconstructed as ambiguous for that character. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

well as the acquisition of unique innovations within crown Cetacea including echolocation in toothed whales (odontocetes) and revamping of the feeding apparatus in baleen whales (mysticetes) that involved tooth loss and the evolution of baleen (Gatesy et al., 2013; McGowen et al., 2014). Extant and fossil sirenian species are less diverse than their cetacean counterparts, but the sirenian fossil record nevertheless includes key transitional forms that document important macroevolutionary changes leading from prorastomids to protosirenids (*Protosiren*) to fully aquatic sirenians (Domning and Gingerich, 1994; Domning, 2000, 2001) (Fig. 5).

The oldest sirenian fossils include Prorastomidae (e.g., *Prorastomus* and *Pezosiren*) and are known from the middle Eocene of Jamaica, Florida, and Africa (Savage et al., 1994; Domning, 2000, 2001; Benoit et al., 2013). Prorastomids are the most primitive sirenians and were amphibious quadrupeds that employed dorsoventral spinal undulation and bilateral thrusts of the hind limbs for locomotion (Domning, 2000, 2001). Presumed aquatic adaptations that evolved at or near the base of the sirenian tree in prorastomids include retracted nasal openings (character 3), acoustic isolation of the periotic, which is no longer fused with other skull bones in *Pezosiren* and later sirenians (e.g., character 44) (Benoit et al., 2013) and is convergent with acoustic isolation of the periotic in cetaceans (Nummela et al., 2007), incipient anterior down-turning of the ventral border of the horizontal ramus, which is a feature possibly related to bottom feeding that is seen in most sirenians (character 47), sacral vertebrae that are unfused in adult animals (character 68), and reduction of canines from double-rooted to single rooted (character 60) (Fig. 5).

The next stage in sirenian evolution is represented by protosirenids (*Protosiren*), which are first known from the middle Eocene of Pakistan and Egypt (Zalmout et al., 2003). Protosirenids were aquatic quadrupeds that employed dorsoventral undulations of the enlarged tail with assistance from bilateral thrusts of the hind limbs (Domning, 2000; Buffrénil et al., 2010). There is a single sacral vertebra (character 68), the ilium is more rodlike, and the obturator foramen is reduced (Domning, 2000) (Fig. 5). Additional changes occurred in the ancestry of fully aquatic, crown sirenians (Dugongidae, Trichechidae) including loss of permanent premolar 5 (character 62), hind limb reduction and loss (character 72, loss of tibia), increased reliance on the tail for locomotion (Domning, 2000), continued reduction of the sacrum (character 68), and fully pachyosteosclerotic ribs that enhance ballast (Buffrénil et al., 2010) (Fig. 5).

Additional changes occur within crown Sirenia (Domning, 2000, 2001). Modifications on the stem *Trichechus* branch include lifelong horizontal replacement of the molar teeth (Savage, 1976), continued reduction of the sacrum (character 68), and shortening of the neck and lumbar regions (characters 66, 67) (Fig. 5). Almost all extant mammals have seven cervical vertebrae and a reduction from seven to six in the ancestry of *Trichechus* resulted in more anteriorly positioned flippers that are capable of greater turning moments (Domning, 2000).

Changes in the caudal vertebrae that are suggestive of a fluke (character 69) occurred in the ancestry of *Halitherium* and more derived dugongids. There are also modifications to the dentition on a series of successive branches that resulted in complete edentulism in *Hydrodamalis gigas* (characters 55, 59-61) (Fig. 5). At the molecular level, pseudogenization of *ENAM* maps onto the branch leading to *H. gigas*. Additional fossil discoveries and genome sequencing will provide the basis for a more complete understanding of sirenian macroevolution.

4.6. Conclusions

The phylogenetic affinities of the recently extinct Steller's sea cow (*Hydrodamalis gigas*) have remained controversial in view of conflicting evidence based on both morphology (Domning, 1994; Vélez-Juarbe et al., 2012; Voss, 2013, unpublished doctoral dissertation) and molecules (Rainey et al., 1984; Crerar, 2012, unpublished doctoral dissertation). We used hybridization capture methods and second generation sequencing to assemble the first dataset comprised of nuclear gene sequences for *H. gigas* including the coding sequence for the *ENAM* gene. Phylogenetic analyses show conclusively that *H. gigas* belongs to Dugongidae and is more closely related to living dugongs (*Dugong dugon*) than manatees (*Trichechus* spp.). Cladistic analyses of a morphological data set that includes both cranial and postcranial characters provide additional support for this conclusion, and further document important character state transformations in the macroevolutionary history of Sirenia. The protein-coding sequence for the complete *ENAM* gene in *H. gigas* is intact, but a transversion mutation (AG to CG) in the acceptor splice site of intron 2 is consistent with loss of function of *ENAM* in this edentulous species.

Acknowledgments

This research was supported by NSF (United States) Grants EF0629860 (M.S.S., J.G.), EAR-PF1249920 (J.V.J.), and DEB1132229 (M.H.), NSERC (Canada) Discovery and Accelerator Supplement Grants (K.L.C.), and an NSERC Alexander Graham Bell Canada Graduate Scholarship (A.V.S.). We thank two anonymous referees for constructive comments on an earlier version of this manuscript.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.05.022.

References

Al-Hashimi, N., Sire, J.-Y., Delgado, S., 2009. Evolutionary analysis of mammalian enamelin, the largest protein, supports a crucial role for the 32-kDa peptide and reveals selective adaptation in rodents and primates. J. Mol. Evol. 69, 635–656.
Ambrose, S.H., Nyamai, C.M., Mathu, E.M., Williams, M.A.J., 2007. Geology, geochemistry, and stratigraphy of the Lemudong'o Formation, Kenya Rift Valley. Kirtlandia 56, 53–64.
Amrine, H.A., Springer, M.S., 1999. Maximum-likelihood analysis of the Tethythere hypothesis based on a multigene data set and a comparison of different models of sequence evolution. J. Mammal. Evol. 6, 161–176.
Anderson, P.K., 1995. Competition, predation, and the evolution and extinction of Steller's sea cow, *Hydrodamalis gigas*. Mar. Mammal Sci. 11, 391–394.
Asher, R.J., Avery, D.M., 2010. New golden moles (Afrotheria, Chrysochloridae) from the Early Pliocene of South Africa. Palaeontol. Electron. 13, 1–12.
Asher, R.J., Hofreiter, M., 2006. Tenrec phylogeny and the noninvasive extraction of nuclear DNA. Syst. Biol. 55, 181–194.
Ávila-Arcos, M.C., Cappellini, E., Romero-Navarro, J.A., Wales, N., Moreno-Mayar, J.V., Rasmussen, M., Fordyce, S.L., Montiel, R., Vielle-Calzada, J.-P., Willerslev, E., Gilbert, M.T.P., 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Sci. Rep. 1, 74.
Benoit, J., Adnet, S., Mabrouk, E.E., Khayati, H., Ali, M.B.H., Marivaux, L., Merzeraud, G., Merigeaud, S., Vianey-Liaud, M., Tabuce, R., 2013. Cranial remain from Tunisia provides new clues for the origin and evolution of Sirenia (Mammalia, Afrotheria) in Africa. PLoS ONE 8, e54307.
Benton, M.J., Donoghue, P.C.J., 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24, 26–53.
Benton, M.J., Donoghue, P.C.J., Asher, R.J., Friedman, M., Near, T.J., Vinther, J., 2015. Constraints on the timescale of animal evolutionary history. Palaeontol. Electron., 181.1FC.
Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prüfer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., Pääbo, S., 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proc. Natl. Acad. Sci. U.S.A. 104, 14616–14621.
Brotherton, P., Endicott, P., Sanchez, J.J., Beaumont, M., Barnett, R., Austin, J., Cooper, A., 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Res. 35, 5717–5728.
Buffrénil, V. de, Canoville, A., D'Anastasio, V., Domning, D.P., 2010. Evolution of sirenian pachyosteosclerosis, a model-case for the study of bone structure in aquatic tetrapods. J. Mammal. Evol. 17, 101–120.

Burbano, H.A., Hodges, E., Green, R.E., Briggs, A.W., Krause, J., Meyer, M., Good, J.F., Maricic, T., Johnson, P.L.F., Xuan, Z., Rooks, M., Bhattacharjee, A., Brizuela, L., Albert, F.W., de la Rasilla, M., Fortea, J., Rosas, A., Lachmann, M., Hannon, G.J., Pääbo, S., 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328, 723–725.
Cohen, K.M., Finney, S.C., Gibbard, P.L., Fan, J.-X., 2013. The ICS international chronostratigraphic chart. Episodes 36, 199–204 (updated).
Cooper, L.N., Seiffert, E.R., Clementz, M., Madar, S.I., Bajpai, S., Hussain, S.T., Thewissen, J.G.M., 2014. Anthracobunids from the Middle Eocene of India and Pakistan are stem perissodactyls. PLoS ONE 9, e109232.
Crerar, L.D., 2012. Genetics of the Steller's Sea Cow (*Hydrodamalis gigas*): A Study of Ancient Bone Material. Doctoral Dissertation, George Mason University.
Diedrich, C.G., 2013. The most northerly record of the sirenian *Protosiren* and the possible polyphyletic evolution of manatees and dugongs. Nat. Sci. 5, 1154–1164.
Domning, D.P., 1976. An ecological model for Late Tertiary sirenian evolution in the North Pacific Ocean. Syst. Zool. 25, 352–362.
Domning, D.P., 1978. Sirenian evolution in the North Pacific Ocean. Univ. Calif. Publ. Geol. Sci. 118, 1–176 (i–xi).
Domning, D.P., 1994. A phylogenetic analysis of the Sirenia. Proc. San Diego Soc. Nat. Hist. 29, 177–189.
Domning, D.P., 2000. The readaptation of Eocene sirenians to life in water. Hist. Biol. 14, 115–119.
Domning, D.P., 2001. The earliest known fully quadrupedal sirenian. Nature 413, 625–627.
Domning, D.P., Gingerich, P.D., 1994. *Protosiren smithae*, new species (Mammalia, Sirenia), from the late middle Eocene of Wadi Hitan, Egypt. Contrib. Mus. Paleontol. Univ. Mich. 29, 69–87.
Domning, D.P., Thomason, J., Corbett, D.G., 2007. Steller's sea cow in the Aleutian Islands. Mar. Mammal Sci. 23, 976–983.
Feldhamer, G.A., Drickamer, L.C., Vessey, V.C., Merritt, J.F., Krajewski, C., 2007. Mammalogy: Adaptation, Diversity, Ecology, third ed. John Hopkins University Press, Baltimore, MD.
Forsten, A., Youngman, P.M., 1982. *Hydrodamalis gigas*. Mammal. Species 165, 1–3.
Fortes, G.G., Paijmans, J.L.A., 2015. Analysis of whole mitogenomes from ancient samples. In: Kroneis, T. (Ed.), Whole Genome Amplification. Humana Press, USA. ArXiv pre-print: <http://arxiv.org/abs/1503.05074>.
Gatesy, J., Geisler, J.H., Chang, J., Buell, C., Berta, A., Meredith, R.W., Springer, M.S., McGowen, M.R., 2013. A phylogenetic blueprint for a modern whale. Mol. Phylogenet. Evol. 66, 479–506.
Gheerbrant, E., 2009. Paleocene emergence of elephant relatives and the rapid radiation of African ungulates. Proc. Natl. Acad. Sci. U.S.A. 106, 10717–10721.
Goswami, A., Prasad, G.V.R., Upchurch, P., Boyer, D.M., Seiffert, E.R., Verma, O., Gheerbrant, E., Flynn, J.J., 2011. A radiation of arboreal basal eutherian mammals beginning in the Late Cretaceous of India. Proc. Natl. Acad. Sci. U.S.A. 108, 16333–16338.
Hedtke, S.M., Morgan, M.J., Cannatella, D.C., Hillis, D.M., 2013. Targeted enrichment: maximizing orthologous gene comparisons across deep evolutionary time. PLoS ONE 8, e67908.
Hodges, E., Rooks, M., Xuan, Z., Bhattacharjee, A., Gordon, D., Brizuela, L., McCombie, W., Hannon, G., 2009. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nat. Protoc. 4, 960–974.
Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., Pääbo, S., 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucl. Acids Res. 29, 4793–4799.
Kircher, M., Sawyer, S., Meyer, M., 2011. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucl. Acids Res. 2011, 1–8.
Koenigswald, W. von, 1997. Brief survey of enamel diversity at the schmelzmuster level in Cenozoic placental mammals. In: Koenigswald, W. von, Sander, P.M. (Eds.), Tooth Enamel Microstructure. Balkema, Rotterdam, pp. 137–162.
Koenigswald, W. von, Clemens, W.A., 1992. Levels of complexity in the microstructure of mammalian enamel and their application in studies of systematics. Scan. Microsc. 6, 195–218.
Lanyon, J.M., Sanson, G.D., 2006. Degenerate dentition of the dugong (*Dugong dugon*), or why a grazer does not need teeth: morphology, occlusion and wear of mouthparts. J. Zool. 268, 133–152.
Li, C., Hofreiter, M., Straube, N., Corrigan, S., Naylor, G.J.P., 2013. Capturing protein-coding genes across highly divergent species. Biotechniques 54, 321–326.
Maddison, W.P., Maddison, D.R., 2011. Mesquite: A Modular System for Evolutionary Analysis. Version 2.75. <http://mesquiteproject.org>.
Manz, C.L., Chester, S.G.B., Bloch, J.I., Silcox, M.T., Sargis, E.J., 2015. New partial skeletons of Palaeocene Nyctitheriidae and evaluation of proposed euarchontan affinities. Biol. Lett. 11, 20140911.
Martin, A.P., Palumbi, S.R., 1993. Body size, metabolic rate, generation time, and the molecular clock. Proc. Natl. Acad. Sci. U.S.A. 90, 4087–4091.
Mason, V.C., Li, G., Helgen, K.M., Murphy, W.J., 2011. Efficient cross-species capture hybridization and next-generation sequencing of mitochondrial genomes from noninvasively sampled museum specimens. Genome Res. 21, 1695–1704.
Mathur, A.K., Polly, P.D., 2000. The evolution of enamel microstructure: how important is amelogenin? J. Mammal. Evol. 7, 23–42.
McGowen, M.R., Gatesy, J., Wildman, D.E., 2014. Molecular evolution tracks macroevolutionary transitions in Cetacea. Trends Ecol. Evol. 29, 336–346.
McKenna, M.C., Bell, S.K., 1997. Classification of Mammals above the Species Level. Columbia Univ. Press, New York.
Meredith, R., Gatesy, J., Murphy, W.J., Ryder, O.A., Springer, M.S., 2009. Molecular decay of the tooth gene enamelin (*ENAM*) mirrors the loss of enamel in the fossil record of placental mammals. PLoS Genet. 5, 1–12.
Meredith, R.W., Mendoza, M.A., Roberts, K.K., Westerman, M., Springer, M.S., 2010. A phylogeny and timescale for the evolution of Pseudocheiridae (Marsupialia: Diprotodontia) in Australia and New Guinea. J. Mammal. Evol. 17, 75–99.
Meredith, R.W., Janecka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011a. Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334, 521–524.
Meredith, R.W., Gatesy, J., Cheng, J., Springer, M.S., 2011b. Pseudogenization of the tooth gene enamelysin (*MMP20*) in the common ancestor of extant baleen whales. Proc. Roy. Soc. B 278, 993–1002.
Meredith, R.W., Gatesy, J., Springer, M.S., 2013. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. BMC Evol. Biol. 13, 20.
Meredith, R.W., Zhang, G., Gilbert, M.T.P., Jarvis, E.D., Springer, M.S., 2014. Evidence for a single loss of mineralized teeth in the common avian ancestor. Science 346, 1254390.
Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc., 2010: pdbprot5448.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop, 1–8.
Müller, J., Reisz, R.R., 2005. Four well-constrained calibration points from the vertebrate fossil record for molecular clock estimates. BioEssays 27, 1069–1075.
Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E., Ryder, O.A., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science 294, 2348–2351.
Nikaido, M., Nishihara, H., Fukumoto, Y., Okada, N., 2003. Ancient SINEs from African endemic mammals. Mol. Biol. Evol. 20, 522–537.
Nummela, S., Thewissen, J.G.M., Bapjai, S., Hussain, T., Kumar, K., 2007. Sound transmission in archaic and modern whales: anatomical adaptations for underwater hearing. Anat. Rec. 290, 716–733.
Ozawa, T., Hayashi, S., Mikhelson, V.M., 1997. Phylogenetic position of mammoth and Steller's sea cow within Tethytheria demonstrated by mitochondrial DNA sequences. J. Mol. Evol. 44, 406–413.
Pickford, M., Hlusko, L.J., 2007. Late Miocene procaviid hyracoids (*Hyracoidea: Dendrohyrax*) from Lemudong'o, Kenya. Kirtlandia 56, 106–111.
Rainey, W.E., Lowenstein, J.M., Sarich, V.M., Magor, D.M., 1984. Sirenian molecular systematics including the extinct Steller's sea cow (*Hydrodamalis gigas*). Naturwissenschaften 71, 586–588.
Rambaut, A., 1996. Se–Al: Sequence Alignment editor.
Rannala, B., Yang, Z., 2007. Inferring speciation times under an episodic molecular clock. Syst. Biol. 56, 453–466.
Reisz, R.R., Müller, J., 2004. Molecular timescales and the fossil record: a paleontological perspective. Trends Genet. 5, 237–241.
Rohland, N., Siedel, H., Hofreiter, M., 2009. A rapid column-based ancient DNA extraction method for increased sample throughput. Mol. Ecol. Resour. 10, 677–683.
Romiguier, J., Ranwez, V., Douzery, E.J., Galtier, N., 2010. Contrasting GC-content dynamics across 33 mammalian genomes: relationship with life-history traits and chromosome sizes. Genome Res. 20, 1001–1009.
Romiguier, J., Ranwez, V., Douzery, E.J.P., Galtier, N., 2013. Genomic evidence for large, long-lived ancestors to placental mammals. Mol. Biol. Evol. 30, 5–13.
Sagne, C., 2001. La diversification des siréniens à l'Éocène (Sirenia, Mammalia): Étude morphologique et analyse phylogénétique du sirénien de Taulanne, *Halitherium taulannense*. Doctoral dissertation, Muséum National d'Histoire Naturelle (Paris), vols. 2.
Savage, R.J.G., 1976. Review of early Sirenia. Syst. Zool. 25, 344–351.
Savage, R.J.G., Domning, D.P., Thewissen, J.G.M., 1994. Fossil Sirenia of the West Atlantic and Caribbean region. V. The most primitive known sirenian, *Prorastomus sirenoides* Owen, 1855. J. Vertebr. Paleontol. 14, 427–449.
Seiffert, E.R., Simons, E.L., Ryan, T.M., Bown, T.M., Attia, Y., 2007. New remains of Eocene and Oligocene Afrosoricida (Afrotheria) from Egypt, with implications for the origin(s) of afrosoricid zalambdodonty. J. Vertebr. Paleontol. 27, 963–972.
Springer, M.S., Burk-Herrick, A., Meredith, R., Eizirik, E., Teeling, E., O'Brien, S.J., Murphy, W.J., 2007. The adequacy of morphology for reconstructing the early history of placental mammals. Syst. Biol. 56, 673–684.
Springer, M.S., Meredith, R.W., Eizirik, E., Teeling, E., Murphy, W.J., 2008. Morphology and placental mammal phylogeny. Syst. Biol. 57, 499–503.
Springer, M.S., Meredith, R.W., Janecka, J.E., Murphy, W.J., 2011. The historical biogeography of Mammalia. Philos. Trans. Roy. Soc. B 366, 2478–2502.
Springer, M.S., Meredith, R.W., Gatesy, J., Emerling, C.A., Park, J., Rabosky, D.L., Stadler, T., Steiner, C., Ryder, O.A., Janečka, J.E., Fisher, C.A., Murphy, W.J., 2012. Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. PLoS ONE 7, e49521.
Springer, M.S., Meredith, R.W., Teeling, E.C., Murphy, W.J., 2013. Technical comment on "The placental mammal ancestor and the Post-K-Pg radiation of placentals". Science 341, 613-b.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.
Steller, G.W., 1751. De Bestiis Marinis [On Beasts of the Sea]. Novi Commentarii Acad. Sci. Imp. Petropoli 2, 289–398.
Steller, G.W., 1899. The beasts of the sea. In: Jordan, D.S. (Ed.), The Fur Seals and Fur-Seal Islands of the North Pacific Ocean. Gov. Printing Office, 629 pp. (Abridged transl. of Steller, 1751, by W. Miller and J.E. Miller).
Swofford, D.L., 2002. PAUP∗. Phylogenetic Analysis using Parsimony (∗ and Other Methods). Sinauer Associates, Sunderland, Massachusetts.
Tabuce, R., Asher, R.J., Lehmann, T., 2008. Afrotherian mammals: a review of current data. Mammalia 72, 2–14.
Tabuce, R., Coiffait, B., Coiffait, P.-E., Mahboudi, M., Jaeger, J.-J., 2001. A new genus of Macroscelidea (Mammalia) from the Eocene of Algeria: a possible origin for elephant shrews. J. Vertebr. Paleontol. 21, 535–546.
Templeton, J.E.L., Brotherton, P.M., Llamas, B., Soubrier, J., Haak, W., Cooper, A., Austin, J.J., 2013. DNA capture and next-generation sequencing can recover whole mitochondrial genomes from highly degraded samples for human identification. Invest. Genet. 4, 26.
Turvey, S.T., Risley, C.L., 2006. Modelling the extinction of Steller's sea cow. Biol. Lett. 2, 94–97.
Vélez-Juarbe, J., Domning, D.P., 2014. Fossil Sirenia of the West Atlantic and Caribbean region. IX. *Metaxytherium albifontanum*, sp. nov. J. Vertebr. Paleontol. 34, 444–464.
Vélez-Juarbe, J., Domning, D.P., 2015. Fossil Sirenia of the West Atlantic and Caribbean region. XI. *Callistosiren boriquensis*, gen. et sp. nov. J. Vertebr. Paleontol. http://dx.doi.org/10.1080/02724634.2014.885034.
Vélez-Juarbe, J., Domning, D.P., Pyenson, N.D., 2012. Iterative evolution of sympatric seacow (Dugongidae, Sirenia) assemblages during the past ∼26 million years. PLoS ONE 7, e31294.
Voss, M., 2013. Revision of the *Halitherium*-species complex (Mammalia, Sirenia) from the late Eocene to early Miocene of Central Europe and North America. Doctoral dissertation, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät I.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., dos Reis, M., 2011. Statistical properties of the branch-site test of positive selection. Mol. Biol. Evol. 28, 1217–1228.
Yang, Z., Wong, W.S.W., Nielsen, R., 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22, 1107–1118.
Zalmout, I.S., Haq, M., Gingerich, P.D., 2003. New species of *Protosiren* (Mammalia, Sirenia) from the early Middle Eocene of Balochistan (Pakistan). Contrib. Mus. Paleontol. Univ. Mich. 31, 79–87.
Zhang, J., Nielsen, R., Yang, Z., 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22, 2472–2479.

Glossary

AfroSINE: a novel family of short interspersed nuclear elements (SINEs) whose distribution is restricted to the genomes of afrotherian mammals

Relaxed molecular clock: a molecular clock model that relaxes the equal rates assumption of a strict molecular clock and allows for rate variation across lineages

Sirenia: an order of placental mammals that includes the first ancestor of *Dugong dugon* that is not also an ancestor of *Loxodonta africana* (African elephant), *Cornwallius sookensis* (desmostylian), or *Procavia capensis* (Cape hyrax), and all descendants of that ancestor