The origin and diversification of pteropods precede past perturbations in the Earth’s carbon cycle

Katja T. C. A. Peijnenburgb,c,1, Ari W. Janssena, Deborah Wall-Palmera, Erica Goetzea, Amy E. Maasd,c, Jonathan A. Toddea,b, and Ferdinand Marléta,2,c

aPlankton Diversity and Evolution, Naturalis Biodiversity Center, 2300 RA Leiden, The Netherlands; bDepartment Freshwater and Marine Ecology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GE Amsterdam, The Netherlands; cDepartment of Oceanography, University of Hawai’i at Mānoa, Honolulu, HI 96822; dBermuda Institute of Ocean Sciences, St. Georges GE01, Bermuda; eDepartment of Earth Sciences, Natural History Museum, London SW7 5BD, United Kingdom; fCentre for Life’s Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom; and gMolecular Genetics Unit, Okinawa Institute of Science and Technology, Onna-son 904-0495, Japan

Pteropods are a group of planktonic gastropods that are widely regarded as biological indicators for assessing the impacts of ocean acidification. Their aragonitic shells are highly sensitive to acute changes in ocean chemistry. However, to gain insight into their potential to adapt to current climate change, we need to accurately reconstruct their evolutionary history and assess their responses to past changes in the Earth’s carbon cycle. Here, we resolve the phylogeny and timing of pteropod evolution with a phylogenomic dataset (2,654 genes) incorporating new data for 21 pteropod species and revised fossil evidence. In agreement with traditional taxonomy, we recovered molecular support for a diversification of the Paleocene–Eocene Thermal Maximum (PETM), the closest analog to modern-day ocean acidification and warming. Our findings indicate that planktonic aragonitic calcifiers have shown resilience to perturbations in the Earth’s carbon cycle over evolutionary timescales.

Significance

Pteropods are abundant aragonitic calcifiers, contributing up to 89% of total pelagic calcification. Because of their delicate shells, they are considered “canaries in the coalmine” to indicate impacts of ocean acidification. Their sensitivity to high CO2 levels and limited fossil record has led to the widely held view that pteropods only became abundant after the PETM. Based on phylogenomic analyses, we show that all major pteropod groups have Cretaceous origins and, hence, they must have survived past perturbations in the Earth’s carbon cycle. Although this suggests that pelagic aragonitic calcifiers have been more resilient to past ocean acidification than currently thought, it is unlikely that pteropods have experienced global change of the same magnitude and speed during their evolutionary history.

Author contributions: K.T.C.A.P. conceived this study and obtained funding; K.T.C.A.P., A.W.J., and F.M. performed research; K.T.C.A.P. and E.G. collected and identified specimens at sea; D.W.-P., A.E.M., and J.A.T. contributed new data/analytic tools; F.M. led the data analyses; K.T.C.A.P. and F.M. wrote the paper; A.W.J., D.W.-P., E.G., A.E.M., and J.A.T. contributed comments to the paper. The authors declare no competing interest.

This article is a PNAS Direct Submission. J.P.H. is a guest editor invited by the Editorial Board.

This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

1To whom correspondence may be addressed. Email: k.t.c.a.peijnenburg@uva.nl or ferninand.marleza@gmail.com.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1920918117/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1920918117

PNAS Latest Articles | 1 of 9
calcite (25). In addition, several groups of pteropods have only partial shells or are shell-less as adults and thus are rarely preserved in marine sediments. Hence, resolving the evolutionary history of pteropods requires a combination of molecular and fossil-based approaches to assess past diversification and timing.

While most researchers recognize pteropods as composed of two orders, Thecosomata (“sea butterflies”) and Gymnosomata (“sea angels”), a recent classification (26) identifies three suborders: (i) Euthecosomata, fully-shelled, omnivorous mucus-web feeders; (ii) Pseudothecosomata, a poorly known group with shelled, partially-shelled, and unshelled species that also use mucus webs for feeding; and (iii) Gymnosomata, with shell-less adults that are specialized predators, primarily on euthecosomes. Progressive evolution toward loss of shells as an adaptation to planktonic life has been proposed for the group (21, 27), but never fully tested.

Previous attempts to resolve the molecular phylogeny of pteropods have relied on small subsets of genes, and resolution has been limited, especially at deeper nodes, due to large rate heterogeneity and insufficient taxonomic signal (28–31). Here, we used a phylogenomic approach based on transcriptome sequencing to fully resolve the phylogeny of pteropods. Using the pteropod fossil record to calibrate the timing of divergence, we estimate that two major groups of pteropods, sea butterflies and sea angels, diverged in the early Cretaceous, and thus both groups must have survived previous global perturbations to the ocean’s carbonate system.

Results and Discussion

Robust Phylogenomic Resolution. We generated transcriptome data for 21 pteropod species collected along two basin-scale transects in the Atlantic Ocean. The number of individuals used for RNA extraction varied from 1, in most cases, to 10 (SI Appendix, Tables S1 and S2). We also incorporated available data for three additional pteropod species (13, 15, 32) and three outgroup species: the sea hare Aplysia californica, within the proposed sister group of pteropods (Aplysiida), and two members of Cephalaspidea, Haminoea antillarum and Philine angasi, from ref. 32. Our taxonomic sampling included representatives of all extant families of Euthecosomata, two of three families of Pseudothecosomata, and two of six families of Gymnosomata. All superfamilies of Pteropoda were sampled except Hydromyloidea (Gymnosomata).

We inferred a set of single-copy orthologs using gene family reconstruction from assembled and translated transcripts and selected 2,654 single-copy nuclear genes for phylogenetic inference based on their taxonomic representation. These selected genes are well represented in our transcriptomes, with a median of 1,815 genes per species and the least being 682 genes (77.4% missing data) for Diacria trispinosa (SI Appendix, Table S2). We
combined these single-copy orthologs in a large data matrix of 834,394 amino acids with 35.75% missing data.

Using the large data matrix (2,654 genes) and a site-homogeneous model of evolution (LG+Γα), we recovered a fully resolved phylogenetic tree with maximal support values at all inter-specific nodes (Fig. 1). To account for the putative limitations of site-homogeneous models that could lead to systematic error, we also applied a site-heterogeneous model (CAT + GTR + Γ) using Bayesian inference. For this analysis, we used a reduced data matrix comprised of the 200 most informative genes (108,008 amino acids; Materials and Methods), and we recovered an identical topology except for a single terminal node (SI Appendix, Fig. S1).

**Pteropod Systematics Reappraised.** Our trees confirm that pteropods are a monophyletic group with sea harae (*Aphysis*) as their closest sister group (28, 30). Pteropods are split into two sister clades: Thecosomata and Gymnosomata, which is congruent with traditional classification but strongly supported by molecular evidence in the present study. An obvious shared character for pteropods are the wing-like structures or “parapodia” used for swimming. Histological and ultrastructural studies showed that the muscle arrangements in parapodia are very complex and look similar in Thecosomata and Gymnosomata, supporting a homologous origin (28). Thecosomata comprise the Euthecosomata and Pseudothecosomata clades, whose member species are all omnivorous mucus-web feeders (33). Their common feeding mechanism is reflected in the well-developed mucus-secreting palial gland that is shared among all thecosomes, as well as a muscular gizzard with which they can crush the hard exoskeletons of their prey (21). Gymnosomata are shell-less at the adult stage and are specialized carnivorous hunters. They have several morphological characters that set them apart from Thecosomata, including tentacle-like structures called “buccal cones” and hook sacs to grab and manipulate shell pteropod prey (28, 34). Hence, the recent revision by Bouchet et al. (26) with three separate suborders (Euthecosomata, Pseudothecosomata, Gymnosomata) should revert back to the original classification of the order Pteropoda with two main clades: suborder Thecosomata, comprising the Euthecosomata and Pseudothecosomata; and suborder Gymnosomata, as shown in Fig. 1. We named all suborders Euthecosomata which gives us greater confidence in our placement of the sister clade and the phylogenetic placement of the *Heliconoides* sp. calibra- tion, we tested five additional calibration schemes to assess the robustness of our analyses (Materials and Methods and SI Appendix). To ensure accurate reconstructions, we chose a realistic model of sequence evolution (CAT-GTR) using the reduced data matrix of 200 genes, with a birth–death prior on divergence times and a gamma prior on the age of the root. We performed cross-validation to select the best clock model (CIR) and prior parameters (birth–death; Fig. 24), and also verified that the divergence times reported under alternative choices were not markedly different for the major clades (Table 2 and SI Appendix, Figs. S2–S4).

Our phylogenomic time-tree dates the origin of pteropods to the early Cretaceous (139.1 Ma with 95% credibility interval [CrI] 188.2 to 111.4; Fig. 24), which is substantially older than the estimates based on previous molecular phylogenies (Paleocene and late Cretaceous as estimated by refs. 30 and 31, respectively). Choosing alternative calibration schemes, priors, or clock models increases the variance around this estimate, but results still place the origin of pteropods firmly in the early Cretaceous (Table 2 and SI Appendix, Figs. S2–S4). Previous phylogenetic studies (30, 31) employed only two and three genetic markers, respectively, and failed to resolve deeper nodes because of large rate heterogeneity between taxa. Our analyses are based on a much more comprehensive gene set, resulting in a fully resolved phylogeny that does not have long branches and is more suitable for molecular dating. Furthermore, we employed realistic models of evolution and better-curated fossil calibration points, which gives us greater confidence in our obtained estimates. We also find that divergences of the two main lineages, Thecosomata and Gymnosomata, are placed in the Cretaceous period, dated at 129.4 (CrI, 175.8 to 103.3) and 100.7 (137.0 to 76.4) Ma, respectively. Of the fully shelled pteropods, all coiled species share a most recent common ancestor at 107 (145.5 to 85.1) Ma and the uncoiled species at 99.1 (134.7 to 78.2) Ma (Fig. 24). These estimates were robust to alternative calibration schemes and model applications (Table 2 and SI Appendix, Figs. S3 and S4). All major lineages of pteropods (sea butterflies and sea angels, species with coiled and uncoiled shells) were already present in the Cretaceous and thus must have survived previous major environmental changes and associated extinctions, such as the asteroid impact at the end of the Cretaceous (K–Pg, 66 Ma; refs. 47 and 48) and the Paleocene–Eocene Thermal Maximum (PETM, 56 Ma; refs. 16, 49, and 50; Fig. 2). Many pelagic groups disappeared during the well-known mass extinction at the K–Pg boundary (51, 52), and, though it is difficult to pinpoint a specific cause of these changes (17), recent evidence from boron isotopes in foraminifers shows that rapid surface ocean acidification was associated with the impact (53). The PETM also had an impact on marine calcifiers, as shown by a rapid decrease in calcium carbonate content in marine sediments (54) and the recording of one of the largest extinctions among deep-sea benthic foraminifers (49, 55). Planktonic calcifiers were also affected during the PETM, with shifts in species assemblages and changes in species...
Table 1. Nine fossil calibrations used in the molecular dating analyses as minimum ages with soft bounds

<table>
<thead>
<tr>
<th>Calibrated node</th>
<th>Fossil species</th>
<th>Epoch: stage</th>
<th>Minimum age, Ma</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a: Diacria</td>
<td>Diacria sangiorgii</td>
<td>Miocene: Tortonian</td>
<td>7.2</td>
<td>39</td>
</tr>
<tr>
<td>b: Cavolinia</td>
<td>Vaginella gaasenesis</td>
<td>Oligocene: Rupelian</td>
<td>28.1</td>
<td>40</td>
</tr>
<tr>
<td>c: Cavolinia+</td>
<td>Cavolinia microbestis</td>
<td>Miocene: Late Burdigalian</td>
<td>16</td>
<td>41</td>
</tr>
<tr>
<td>d: Cavolinioidea</td>
<td>Euchilotheca yenensis</td>
<td>Eocene: Ypresian</td>
<td>47.8</td>
<td>43</td>
</tr>
<tr>
<td>e: Limacina</td>
<td>Limacina gormani</td>
<td>Eocene: Ypresian</td>
<td>47.8</td>
<td>40</td>
</tr>
<tr>
<td>f: Limacinoidea</td>
<td>Helicinoides sp.</td>
<td>Cretaceous: Campanian</td>
<td>72.1</td>
<td>20</td>
</tr>
<tr>
<td>g: Pseudothecosomata</td>
<td>Peracle ammonae</td>
<td>Oligocene: Chattian-Burdigalian</td>
<td>16</td>
<td>41</td>
</tr>
<tr>
<td>h: Gymnosomata</td>
<td>Clione ? imindaisensis</td>
<td>Oligocene: Chattian</td>
<td>23</td>
<td>41</td>
</tr>
<tr>
<td>i: Pteropoda+Aplysia</td>
<td>Akeria neocomiensis</td>
<td>Cretaceous: Early Hauderivian</td>
<td>133</td>
<td>42</td>
</tr>
</tbody>
</table>

Letters refer to nodes labeled a to i in the chronogram of Fig. 2A. Ages are given in million years ago (Ma).

Evolutionary History of Pteropods. Our estimated divergence times for the major pteropod groups precede by far their oldest known fossils (Fig. 2A). The largest discrepancies are found for Gymnosomata and Pseudothecosomata, for which the oldest fossils are from the Chattian (probably a Dothicoidea and Pseudothecosomata, for which the oldest fossils are from the Chattian (probably a Cymbulia spp., 28 to 23 Ma) and Chattian-Burdigalian periods (Peracle ambersae, 28 to 16 Ma), respectively (ref. 41 and Dataset S1). This led Janssen et al. (2016; ref. 60) to suggest that rapid environmental changes during the PETM may have had a triggering effect on pteropod evolution.

The abundance of calcareous nanoplanктon, but no clear signal of increased extinction (56–59). Although we show that major pteropod lineages survived the K-Pg and PETM, this does not mean that pteropods were unaffected by these crises. It may be that many lineages were lost that have not been recovered in the fossil record and which cannot be sampled from extant populations. Interestingly, however, the pteropod fossil record shows a sudden increase from a single recorded species prior to PETM (Thanetian, Helicinoides mercinensis) to 33 species post-PETM (Ypresian; Fig. 2C and Dataset S1). This led Janssen et al. (2016; ref. 60) to propose that rapid environmental changes during the PETM may have triggered a significant change in pteropod evolution.

We do not see a clear trend toward gradual loss of shells in our phylogeny, though shelled groups have earlier origins than unshelled ones. This hypothesis should be further assessed in future work by sampling more species belonging to the elusive Pseudothecosomata, with species ranging from fully shelled (Peracle spp.) to partially shelled (Cymbulia, Corolla, Gleba spp.) and even entirely unshelled as adults (Desmopterus spp.). In our phylogeny, we only included one Peracle species possessing an external coiled calcareous shell and one Cymbulia species bearing a gelatinous “pseudoconch.” Since the pteropod outgroups are benthic gastropods (Aplysida, Cephalaspidea) with mostly coiled or reduced shells, the ancestor of pteropods most likely lived on the seafloor and had a coiled shell (28, 30). Some authors have proposed that pteropods evolved in a neotenic fashion, with larvae of benthic gastropods becoming sexually mature and living their full life cycle in the open water column, because of “juvenile” shell characters such as the sinistral spiral and aragonitic shell structure (69, 70). However, this hypothesis was questioned by Jägersten (1972; ref. 71), who argued that the foot—an adult feature—played a decisive role in the transition from a benthic to a holoplanktonic existence by evolving from a creeping organ to swimming fins. It is interesting to note that a similar hypothesis was proposed for the only other extant group of holoplanktonic gastropods, the heteropods (Pterotrachioidea). For this group, the earliest fossil, Coelodiscus minutus, occurs in the early Jurassic and represents the oldest known holoplanktonic gastropod (72). Colonization of the open water column required numerous adaptations in both groups of holoplanktonic gastropods independently: pteropods belonging to subclass Heterobranchia and heteropods belonging to subclass Caenogastropoda.

Fate of Pteropods from Cretaceous to Anthropocene. Pteropods evolved in the early Cretaceous and thus were contemporaries of other major calcifying groups in the open ocean, such as ammonites and foraminifers (Fig. 2B). Based on their fossil records,
Tajika et al. (2018; ref. 23) suggested that planktonic gastropods filled the ecological niche left empty by juvenile ammonites after their extinction at the end of the Cretaceous. However, this seems less likely given our earlier estimated origins of both sea butterflies (thecosomes) and sea angels (gymnosomes). Instead, we propose that sea angels evolved as specialized predators of planktonic gastropods.

Fig. 2. All major lineages of pteropods diverged in the Cretaceous and thus must have survived previous major global change events. (A) Chronogram of pteropods (25 taxa and 3 outgroups) based on 200 genes, concatenated as 108,008 amino acid positions, analyzed under a CAT-GTR+Γ evolutionary model and CIR relaxed clock model with a birth–death prior on divergence times and a gamma prior on the age of the root (150 ± 70 Ma). Red diamonds indicate fossil-calibrated nodes, with red dotted lines to show divergence from minimum ages of the fossils. (Right) Drawings of fossil species used for the calibrations of nodes: a, Diacria sangiorgii; b, Vaginella gaasensis; c, Cavolinia microsetias; d, Euchilotheca ganensis; e, Limacina gormani; f, Heliconoides sp.; g, P. amberae; and h, Clione ? imdinaensis (Table 1). The outgroup calibration is A. neocomiensis (no drawing available). Two major global change events are indicated as K–Pg (Cretaceous–Paleogene asteroid impact; 66 Ma) and PETM (Paleocene–Eocene Thermal Maximum; ∼56 Ma). (B) Curves for ammonite (red), planktonic foraminifer (yellow), and pteropod (blue) observed species diversity through time are based on sediment records from refs. 44 and 46 and our own database (Dataset S1), respectively. The blue arrow indicates the oldest known pteropod fossil. (C) More detailed records of observed species diversity through time for different pteropod groups: Cavolinioidea and Limacinoidea (green), Pseudothecosomata (orange), and Gymnosomata (purple).
sea butterflies during the Cretaceous, perhaps in an evolutionary arms race (73) in which thecosomes evolved stronger and ever more sophisticated shells (e.g., Diacria, Cavolinia), and gymno-
somes evolved adaptations for prey capture and extraction, specific to the shell shapes of their prey. For instance, Clione feed exclusively on certain Limacina species by manipulating the coiled shells with their flexible buccal cones and extracting the prey with their hooks, while Pseudodermapsis feed on long and straight-shelled Creseis species using their long and flexible proboscis for prey extraction (21, 74). We thus consider it plausible that different groups of pteropods coevolved during a period that is referred to as the Marine Mesozoic Revolution. In this major evolutionary episode, a series of ecological shifts took place on the seafloor because of the evolution of powerful, rel-
atively small, shell-destroying predators, including a massive rad-
diation of predatory gastropods in the Early Cretaceous (75, 76).

This forced benthic gastropods to develop heavily armored shells and perhaps also to escape into the open water column, as was suggested for earlier geological periods (72, 77). Furthermore, during the mid-late Cretaceous, major changes in ocean circu-
lation, stratification, and nutrient partitioning took place that were favorable for plankton evolution, particularly for planktonic calcifiers (78).

Although the open ocean may have been a refuge for gas-
tropods in the early Cretaceous, it is an increasingly challenging habitat to survive in during the Anthropocene. Planktonic gas-
tropods have evolved thin, fragile shells of aragonite that are sensitive to ocean acidification, and most species live in surface waters where CO2 is absorbed by the ocean. Incubation experi-
ments with shelled pteropods mimicking future ocean conditions have shown that elevated pCO2 and undersaturated aragonite conditions cause decreased calcification rates (79), shell disso-
lution (80), increased mortality (10, 81), and differential ex-
pression of genes involved in neurofunction, ion transport, and shell formation (13, 15). However, such experiments have pri-
marily assessed phenotypic responses in short-term, single-
generation studies, and thus cannot take into account the abil-
ities of organisms to acclimate or adapt to changing conditions over longer timescales. The ecological record, alternatively, can provide insight into long-term evidence of ocean acidification and the associated responses of marine calcifiers. However, the fossil record of pteropods is far from complete, and we need to rely on estimates of molecular divergence times to resolve the tempo and pattern of their evolution. The fact that pteropods have survived previous episodes of ocean acidification, such as during the PETM, does not, however, mean that they are infi-
nitely resilient to current changes. Current rates of carbon re-
lease from anthropogenic sources are at least an order of magnitude higher than we have seen for the past 66 million years (82). Although our results suggest resilience of pteropods to past ocean acidification, it is unlikely that they have ever, during their entire evolutionary history, experienced global change of the magnitude and speed that we see today.

Table 2. Inferred divergence times of key pteropod clades under six calibration schemes, the CIR relaxed clock model, and birth–death priors based on 200 genes

<table>
<thead>
<tr>
<th>Calibration scheme</th>
<th>Pteropoda</th>
<th>Thecosomata</th>
<th>Euthecosomata</th>
<th>Pseudothecosomata</th>
<th>Gymnosomata</th>
</tr>
</thead>
<tbody>
<tr>
<td>s1: Akera 163.1 Ma/Euthecosomata</td>
<td>147.5 (121–195)</td>
<td>133.7 (109–177)</td>
<td>112.1 (92–148)</td>
<td>108.0 (80–144)</td>
<td>91.8 (64–132)</td>
</tr>
<tr>
<td>s2: Akera 123 Ma/Euthecosomata</td>
<td>142.5 (112–190)</td>
<td>129.6 (101–174)</td>
<td>111.8 (93–145)</td>
<td>108.0 (80–144)</td>
<td>91.4 (64–132)</td>
</tr>
<tr>
<td>s3: Akera 163.1 Ma/Limacinoidea</td>
<td>146.4 (119–200)</td>
<td>134.6 (107–181)</td>
<td>112.5 (89–153)</td>
<td>111.6 (82–151)</td>
<td>78.3 (62–123)</td>
</tr>
<tr>
<td>s4: Akera 133 Ma/Limacinoidea</td>
<td>139.1 (111–188)</td>
<td>129.4 (103–176)</td>
<td>110.2 (88–150)</td>
<td>110.9 (88–152)</td>
<td>100.7 (76–137)</td>
</tr>
<tr>
<td>s5: no Akera/Euthecosomata</td>
<td>135.4 (100–193)</td>
<td>125.4 (92–180)</td>
<td>107.2 (79–154)</td>
<td>105.5 (75–152)</td>
<td>90.1 (62–140)</td>
</tr>
<tr>
<td>s6: no Akera/Limacinoidea</td>
<td>141.9 (106–193)</td>
<td>131.5 (98–180)</td>
<td>111.1 (84–155)</td>
<td>110.3 (79–154)</td>
<td>99.9 (65–143)</td>
</tr>
</tbody>
</table>

Calibration schemes 1 to 6 (s1 to s6) differ in terms of the Akera calibration and the position of Helicinoida spp. as either the oldest Euthecosomata or Limacinoidea (Materials and Methods). Ages are given in million years ago (Ma), with 95% credibility intervals in brackets (see also Fig. 2).

Materials and Methods

Sample Collection. Pteropod specimens were collected on Atlantic Meridional Transect (AMT) cruises 22 and 24 (2012, 2014) using 0.71-m-diameter bongo and RMT1 nets towed obliquely between a median of 305 m depth and the sea surface (200-μm and 333-μm nets; SI Appendix, Table S1). Animals were sorted from bulk plankton and identified live under a light microscope to species level (when possible) based on the most recent taxonomic informa-
tion, following the World Register of Marine Species (www.marinespecies.
org). Specimens were preserved in RNALater (Invitrogen) and flash-frozen in liquid nitrogen, with storage at –80 °C.

Transcriptome Sequencing and Filtering. RNA was extracted using RNeasy micro or mini kits (Qiagen) after homogenization with a TissueLyser (Qiagen). Because of the size of individual specimens for some species (e.g., Helicinoida inflatus, Limacina bilimoides), the pooling of individuals (up to a maximum of 10) was needed to obtain sufficient RNA for sequencing (SI Appendix, Table S1). RNA quantity was determined by fluorometry using a Qubit (Invitrogen), and RNA integrity was assessed using the Xërion system (Bio-Rad). RNA-seq libraries were prepared using the TruSeq RNA Library preparation kit (Illumina), and between 12 and 33M paired-end reads per sample were sequenced for 100 cycles on a HiSeq2000 platform at the Wellcome Trust Centre for Human Genetics (Oxford). A total of 2.3 to 6.6 Gb of DNA sequence data was obtained from each of 22 samples, ensuring accurate de novo transcriptome assemblies (SI Appendix, Table S2). Reads were deposited to the SRA under the BioProject accession PRJNA591100. After quality assessment with Fastqc (83), reads were trimmed using sickle (84) and subsequently assembled using Trinity (v2.3.2) with default param-
eters and a k-mer of 25 (85).

To avoid cross-contamination between samples sequenced on the same Illumina lanes (possibly due to index “hopping”), we applied a filtering procedure based on relative transcript expression across datasets, similar to the one implemented in Croco (86). Briefly, for the assembled transcriptome of each dataset, we measured expression level using Kallisto (v0.42.4) (87) in each of the multiple datasets sequenced simultaneously. We calculated the ratio of TPMs between the putatively contaminating datasets and the original dataset and excluded transcripts with a contaminant enrichment greater than twofold enrichment and a minimal count lower than two in the original dataset. Across datasets, a median of 6% of transcripts were ex-
cluded on the first criterion and a median of 26% on the second.

De novo assembled transcriptomes usually include a high degree of re-
dundancy, as alternative transcripts derived from the same genes are dis-
tinguished in the assembly. As this might constitute a problem for orthology assign-
ment, we clustered transcripts based on the fraction of remapped reads that they share. To do so, we mapped reads from transcriptomes back to transcripts using Bowtie2 enabling up to 50 multimers (-k 50) and pro-
cessed the resulting alignments with Corset (v1.06) (88). Then, we estimated transcript expression using Kallisto (v0.43.1) (87), and we selected the most highly expressed transcript with each Corset cluster as a reference transcript for subsequent steps. The best open reading frame (ORF) for each of these selected transcripts was predicted using TransDecoder (v5.0.2) using a BLAST against a version of the UniProt Knowledgebase limited to metazoan taxa (e-value 10−5) (89).

Phylogenetic Analyses. We conducted orthology inference using the OMA package (v2.3.0), which performs Smith–Waterman alignment and identifies orthologous groups based on evolutionary distance (90). We selected single-copy orthologs (OMA groups) represented by at least one of the three out-
group taxa (A. californica, H. antillarum, or P. angasi) and at least half of the 28 ingroup taxa. These cutoffs yielded 2,652 single-copy orthologs suitable
for phylogenetic analysis. For each ortholog family, protein sequences were aligned using Mafft, and poorly aligned regions trimmed using Trimal (91) for sites with a gap in more than 75% of taxa (-gt 0.15) and applying a low minimum similarity threshold (-st 0.001). The concatenation of these 2,654 orthologous genes yields a supermatrix of 834,394 positions with a total fraction of 35.75% missing data. We performed maximum-likelihood reconstruction with ExaML (v3.0.17) using a partitioned site-homogeneous model (one partition per ortholog) and an LG+Γ4 model (92). Node support was calculated using 100 bootstrap replicates inferred with independently generated starting trees (Fig. 1). To compare results with a site-heterogeneous model, we used Bayesian inference and a reduced dataset (because the complete dataset would be computationally intractable). We selected a subset of 200 genes which showed the maximal average bootstrap support when analyzed independently (using RAxML v8.1.18; LG+Γ4 model and 100 rapid bootstraps) (93, 94). The concatenation of these 200 genes generated a 108,008-amino acid alignment. This alignment was analyzed using PhyloBayes-MP (v1.6) assuming a CAT + GTR+Γ4 model with chains run for more than 1,500 generations with 500 discarded as burn-in. Chain convergence was checked, and maxdiff found at 0 (95).

Molecular Divergence Time Inference. We used PhyloBayes (v4.1c) to infer molecular divergence times using the reduced 200-gene supermatrix and the topologies from Fig. 1 (ML analysis) (96). We used the CAT + GTR+Γ4 model of sequence evolution because it is the most appropriate for dealing with across-site heterogeneities while minimizing long-branch attraction. We evaluated three relaxed clock models, the lognormal autocorrelated process, the CIR process, and the uncorrelated gamma multipliers (reviewed in ref. 97; SI Appendix, Fig. 52). Using a 10-fold cross-validation procedure, we found a best fit of the CIR relaxed clock model and of the birth–death prior on divergence time for the starting points of calibration points (Fig. 2 and SI Appendix, Fig. 52), and therefore chose these parameters for our main divergence time inferences (Fig. 2A). We also applied a gamma prior on the root with a mean age of 150 Myr and SD of 70 Myr. This root prior was meant to be broad and was chosen based on the oldest known crown heterobranch (i.e., Sinularbulla, belonging to the extant superfamily Acteonidae) dated to 240 Ma; refs. 98 and 99) and species referred to as Akera (Akera mediojaurus and A. neocomiensis) dated at 163 to 166 Ma and 133 Ma, respectively; refs. 42 and 100 and SI Appendix). The root prior was tested by running the analysis without calibrations to verify that the priors did not overestimate the constrained ages. This analysis returned a posterior distribution on the root of mean 154 ± 32 Myr, supporting the priors as appropriate.

Evaluating Alternative Schemes, Models, and Priors. We ran analyses to assess the impact of six calibration schemes, three clock models, and three different priors on divergence time estimates using PhyloBayes. We applied eight calibration points as minimum ages based on the oldest representatives for different clades of pteropods (Table 1 and Fig. 2A). These calibrations are based on a thorough review of all described pteropod taxa in the fossil record as well as from recent times (Dataset S1). Additionally, we assessed five alternative calibration schemes, evaluating the impact of different Akera calibrations and the position of the Helicoideidae sp. calibration on divergence time estimates.

Although we chose the A. neocomiensis fossil (42) from the Early Hauterivian (133 Ma) in France to date the node Pteropoda + Aplysia, there have been older fossils described in the Jurassic (A. mediojaurus, 163.1 Ma; refs. 100 and 101). However, these fossils either lack some of the features of Akera and may not belong here or are doubtful due to poor preservation (SI Appendix). It is also possible that true Akera representatives are even younger; for example, Akera striatella is known from the middle Eocene, and is poorly distinguishable from the recent species A. bullata (45). Further research, including analysis of fossil morphological characters, is necessary to resolve the exact phylogenetic relationships of Akera within Aplysioidea and to Pteropoda. Helicoideidae sp. is the oldest pteropod fossil found thus far, with a coiled shell and clearly belonging to the Limacinoidea (schemes 3, 4, and 6). However, we also tested the impact of a different placement of this fossil as the oldest Eucosmata (schemes 1, 2, and 5). These considerations resulted in six different calibration schemes (Table 2 and SI Appendix, Figs. 53 and 54).

We further tested the interplay of the three clock models (UGAM, normal, and CIR; SI Appendix, Fig. 53) and three different priors on divergence time estimates (uniform, birth–death, and Dirichlet; SI Appendix, Fig. 54) using the same other parameters as previously described. These analyses show that, although the overall variance around divergence time estimates increases when including different calibration schemes, clock models, and priors, the overall conclusions remain valid. All main pteropod groups (Eucosmata, Pseudoeucosmata, Gymnosomata) originate in the Cretaceous with Eucosmata and Pteropoda, Helicoideidae sp. is the oldest pteropod fossil found thus far, with a coiled shell and clearly belonging to the Limacinoidea (schemes 3, 4, and 6). However, we also tested the impact of a different placement of this fossil as the oldest Eucosmata (schemes 1, 2, and 5). These considerations resulted in six different calibration schemes (Table 2 and SI Appendix, Figs. 53 and 54).

Data Availability. Transcriptome data have been deposited in the NCBI database under BioProject accession PRJNAS91100. Alignments and files generated during analyses including Bayesian sample and bootstrap replicates are available as a Zenodo dataset: doi:10.5281/zenodo.3479131. All data are included in the article and SI Appendix.

ACKNOWLEDGMENTS. We appreciate the comments of J. Pap, D. Gaviouchkina, M. Malinsky, and three anonymous reviewers on earlier versions of this manuscript. We acknowledge the Oxford Genomics Centre, Wellcome Trust Centre for Human Genetics, for performing the transcriptome sequencing, and the Scientific Computing and Research Support Division at OIST for supporting computational analysis. We thank A. K. Burridge for collecting samples and T. Smythe, G. Tarran, and A. Rees for cruise leadership and support for participation in the AMT programme. This study contributes to the international IMBeR project and is contribution number 339 of the AMT programme. This research and K.T.C.A.P. were supported by Vidi Grant 016.161351 from the Netherlands Organisation for Scientific Research (NWO), a Tera van Benthem Jutting Grant from the Amsterdam University Fund, and KNAW Ecology Fund UPS/297/EC014031. D.W.-P. received funding from the European Union’s Horizon 2020 Research and Innovation Programme under Marie Skłodowska-Curie Grant Agreement 746186 (POSEIDON). E.G. was supported by National Science Foundation (USA) Grants OCE-1338959 and OCE-1253681. F.M. was supported by internal funding of the Okinawa Institute of Science and Technology to the Molecular Genetics Unit and by a Royal Society University Research Fellowship. The Atlantic Meridional Transect is funded by the UK Natural Environment Research Council through its National Capability Long-term Single Centre Science Programme, Climate Linked Atlantic Sector Science (Grant NE/R015953/1).


