nature ecology & evolution

ARTICLES

DOI: 10.1038/s41559-017-0331-3

Ctenophore relationships and their placement as the sister group to all other animals

Nathan V. Whelan[1,2*], Kevin M. Kocot[3], Tatiana P. Moroz[4], Krishanu Mukherjee[4], Peter Williams[4], Gustav Paulay[5], Leonid L. Moroz[4,6*] and Kenneth M. Halanych[1*]

Ctenophora, comprising approximately 200 described species, is an important lineage for understanding metazoan evolution and is of great ecological and economic importance. Ctenophore diversity includes species with unique colloblasts used for prey capture, smooth and striated muscles, benthic and pelagic lifestyles, and locomotion with ciliated paddles or muscular propulsion. However, the ancestral states of traits are debated and relationships among many lineages are unresolved. Here, using 27 newly sequenced ctenophore transcriptomes, publicly available data and methods to control systematic error, we establish the placement of Ctenophora as the sister group to all other animals and refine the phylogenetic relationships within ctenophores. Molecular clock analyses suggest modern ctenophore diversity originated approximately 350 million years ago ± 88 million years, conflicting with previous hypotheses, which suggest it originated approximately 65 million years ago. We recover *Euplokamis dunlapae*—a species with striated muscles—as the sister lineage to other sampled ctenophores. Ancestral state reconstruction shows that the most recent common ancestor of extant ctenophores was pelagic, possessed tentacles, was bioluminescent and did not have separate sexes. Our results imply at least two transitions from a pelagic to benthic lifestyle within Ctenophora, suggesting that such transitions were more common in animal diversification than previously thought.

Ctenophores, or comb jellies, have successfully colonized nearly every marine environment and can be key species in marine food webs[1–6]. For example, invasive ctenophores have caused dramatic fisheries collapses by voraciously preying on native fish larvae and their food, resulting in the economic loss of millions of US dollars to impacted areas[4]. Understanding the morphological and life history diversity of ctenophores in a comparative context is essential for our knowledge of ctenophore and metazoan diversification as a whole[7]. Ctenophores have received considerable attention with regards to debate about whether they are the sister group to all other animals[3,5,8–11], but the relationships within Ctenophora have been the focus of only limited research[3,12,13].

Putative ctenophore fossils date back to the Ediacaran Period[14] and substantial morphological diversity is present in the Cambrian[15,16]. All ctenophores possess smooth muscles, and at least one genus, *Euplokamis*, has striated muscles[17]. Most ctenophores possess tentacles (Fig. 1), but species in the genus *Ocyropsis* lose tentacles as adults[18] and beroids lack them throughout their life cycle (Fig. 1)[1,6]. Many species are pelagic, but some are benthic or semi-benthic as adults and can have a relatively flattened body and lose the ciliary comb rows that otherwise characterize the phylum[6,19] (Fig. 1). Relationships among ctenophore lineages remain poorly resolved as past phylogenetic analyses have either included too few taxa to recover broad evolutionary patterns[3] or resulted in weak support for the deepest nodes, likely resulting from the use of only one or two genes[12,13]. Past researchers[12,13] have also hypothesized that Ctenophora has undergone a bottleneck in species diversity, possibly as recently as 65 million years ago (MYA). However, the age of crown group ctenophores has yet to be estimated using molecular dating methods. Here, we sequenced 27 transcriptomes from species across most of the known phylogenetic diversity of Ctenophora. New sequence data were combined with 10 ctenophore and 50 non-ctenophore publicly available transcriptomes (Supplementary Tables 1 and 2) to clarify the phylogenetic placement of Ctenophora[11,20–22]. Thus, we performed analyses to determine appropriate outgroups and ctenophore placement among other metazoans using more ctenophore taxa than previous studies[3,5,9–11,20] (Supplementary Table 2).

Results

Ctenophora is the sister lineage to all other extant metazoans. Using a variety of data-filtering schemes and different substitution models to control for systematic error (Supplementary Table 2), we recovered ctenophores as the sister group to all other extant metazoans (1.00 Bayesian posterior probability (PP), 100% bootstrap support (BS); Fig. 2 and Supplementary Figs. 1–14). The percentage of individual genes favouring the hypothesis of ctenophores as the sister group to all other animals was higher in every dataset (56.8–75.4%; Table 1) than the percentage of genes favouring the hypothesis of sponges as sister to all other animals (32.7–43.2%; Table 1). Datasets that were trimmed of the genes most likely to cause long-branch attraction had the highest percentage of genes supporting Ctenophora as the sister to all other animals, indicating that this hypothesis is not a result of long-branch attraction. Our recovered placement of ctenophores does not change when the concerns of Pisani et al.[20] about outgroup choice and the use of site-heterogeneous models are taken into account (see additional considerations in refs [22,23]).

A recent study by Simion et al.[21] recovered sponges as the sister lineage to all other animals, but methodological problems in their

---

[1]Molette Biology Laboratory for Environmental and Climate Change Studies, Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA. [2]Warm Springs Fish Technology Center, US Fish and Wildlife Service, 5308 Spring St, Warm Springs, GA 31830 USA. [3]Department of Biological Sciences, The University of Alabama, Box 870344, Tuscaloosa, AL 35487 USA. [4]The Whitney Laboratory for Marine Biosciences, University of Florida, St. Augustine, FL 32080, USA. [5]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. [6]Department of Neuroscience and McKnight Brain Institute, University of Florida, Gainesville, FL 32611, USA. *e-mail: nathan_whelan@fws.gov; moroz@whitney.ufl.edu; ken@auburn.edu

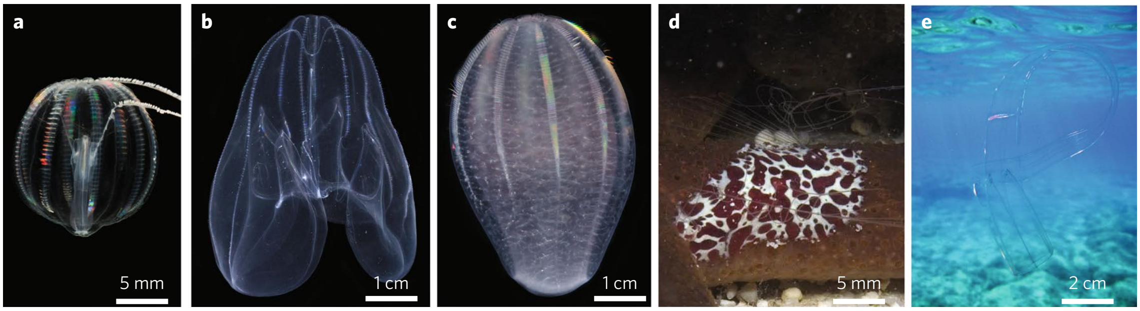

Fig. 1 | Exemplar morphological forms of Ctenophora. a, Cydippid morphology (ovate body and long tentacles); photograph taken by J. Townsend. **b**, Lobate morphology (reduced tentacles and large lobes). **c**, Beroida morphology (lacking tentacles and lobes). **d**, Platyctenida morphology (flattened and long tentacles). **e**, Cestida morphology (ribbon-like); photograph taken by R. Pillon and contrast adjusted in Adobe Photoshop.

analyses explain disagreement with our results. The placement of sponges as the sister lineage to all other animals was only recovered using the CAT-F81 substitution model (often referred to as 'CAT'), which has been shown to sometimes result in less accurate phylogenetic hypotheses than the models used here[24]. More problematically, not a single Bayesian analysis conducted by Simion et al.[21] converged (Simion et al. personal communication), rendering them statistically invalid. The use of other site-heterogeneous models that may not suffer from problems associated with CAT-F81 (see ref. [24] and Supplementary Discussion) resulted in Ctenophora as the sister to all other animals[21], which is consistent with our findings (Fig. 2 and Supplementary Figs. 1–14) and those of two recent papers that employed novel methods[25,26].

Wide consensus exists that Ctenophora is a hard lineage to place on the animal tree of life[8,11,20,21] and increased taxon sampling is broadly accepted to aid in the placement of difficult lineages[27–29]. Our datasets have greater ctenophore taxon sampling than past studies, including 27 novel ctenophore transcriptomes, and are arguably the most appropriate datasets generated to date for assessing the placement of Ctenophora. Using datasets with reasonably high ctenophore and other non-bilaterian taxon sampling, our results strongly reject the hypothesis that sponges are the sister lineage to all other extant metazoans.

Bayesian inference with a relaxed molecular clock also recovered ctenophores as the sister group to all other animals with maximum support (1.00 PP; Supplementary Fig. 15 and Supplementary Discussion). These analyses indicated that sampled ctenophores shared a common ancestor much more recently than either crown group sponges, cnidarians or bilaterians (Supplementary Fig. 15 and Supplementary Discussion). Thus, our findings are consistent with the hypothesis that Ctenophora has undergone a species-diversity bottleneck, but we acknowledge uncertainty in our absolute diversification timing (350 ± 88 MYA; Supplementary Discussion). Nevertheless, this bottleneck appears to have occurred between 456 and 261 MYA (Supplementary Fig. 15), which is much longer ago than the 65 MYA previously hypothesized[12,13]. Given our results, ancestral ctenophores likely experienced a drastic decline before or during the Permian–Triassic extinction (~250 MYA; ref. [30]). Early-to-mid-Paleozoic Ctenophore fossils display substantially greater morphological diversity (for example, more than eight comb rows) than seen today[14–16], supporting the hypothesis that the phylum underwent a major diversity decline during the Paleozoic.

Evolution of Ctenophora. The relationships among ctenophores were assessed using a novel set of ctenophore-centric core orthologs. Orthology determination, subject to paralog and contamination screening, resulted in a primary dataset of 350 genes and 98,844 amino acid positions (Supplementary Table 3). Potential causes of systematic error were controlled for by creating additional datasets that removed potentially problematic genes (Supplementary Table 3)[11]. Phylogenetic analyses were conducted with data partitioning under maximum likelihood and with the CAT-GTR site-heterogeneous substitution model[31] under Bayesian inference. All phylogenetic analyses focusing on intra-ctenophore relationships resulted in identical, highly-supported relationships (Fig. 3 and Supplementary Figs. 16–19).

We found pervasive non-monophyly among currently recognized ctenophore higher taxonomic groups, including Tentaculata, Cydippida, and Lobata (Fig. 3 and Supplementary Figs. 16–19), corroborating previous analyses[3,12,13]. Other traditional groups based on morphology, such as the benthic Platyctenida and atentaculate Beroida, were recovered as monophyletic (Fig. 3 and Supplementary Figs. 16–19), congruent with past analyses[3,12,13]. Lobata was paraphyletic by inclusion of Cestida, which was represented by the ribbon-like *Cestum veneris* (Fig. 3 and Supplementary Fig. 20). *Ocyropsis* species, which lose tentacles as adults, move by muscle propulsion and are dioecious (Supplementary Figs. 21 and 22), were monophyletic and sister to a clade with Cestida and all other lobates except the benthic *Lobatolampea tetragona*. These results indicate that the cydippid and lobate body plans are plesiomorphic (Fig. 3 and Supplementary Fig. 20).

We recovered *Euplokamis dunlapae* as the sister lineage to all other sampled ctenophores with maximum support (Fig. 3 and Supplementary Figs. 16–19), consistent with initial genomic analyses[3]. Previous studies also recovered *Mertensia ovum* and *Charistephane fugiens* with *E. dunlapae* as a united sister group to all other ctenophores[13]. Novel analyses based on 18 S ribosomal RNA, which included many more taxa than our transcriptome-based analyses, recovered Mertensiidae as non-monophyletic and a clade including *M. ovum*, *C. fugiens* and *Euplokamis* species as sister to all other extant ctenophores (see Supplementary Discussion and Supplementary Fig. 23). As we were unable to sample *M. ovum* and *C. fugiens*, we cannot reject that any three of these species, a clade of all three or a yet-to-be-discovered species could be the sister lineage to all other extant ctenophores. *E. dunlapae* is the only ctenophore species known to have striated muscles[2] and Bayesian ancestral state reconstruction suggests that striated muscles likely evolved after the split between *E. dunlapae* and other ctenophores (PP = 0.90; Supplementary Fig. 24), rather than being present in the most recent common ancestor (MRCA) of extant ctenophores. Striated muscles have evolved at least three times: after the split of the *E. dunlapae* lineage from other ctenophores, in select Cnidaria[32] and in bilaterians[32] (Supplementary Fig. 25). Given that all extant ctenophores have smooth muscles, the MRCA of all extant ctenophores almost certainly possessed smooth muscles (PP = 1.0; Supplementary Fig. 24). Given our inferred relationships among ctenophores, sponges, placozoans and cnidarians (Fig. 2), the MRCA to extant metazoans either possessed smooth muscles that were subsequently lost at least

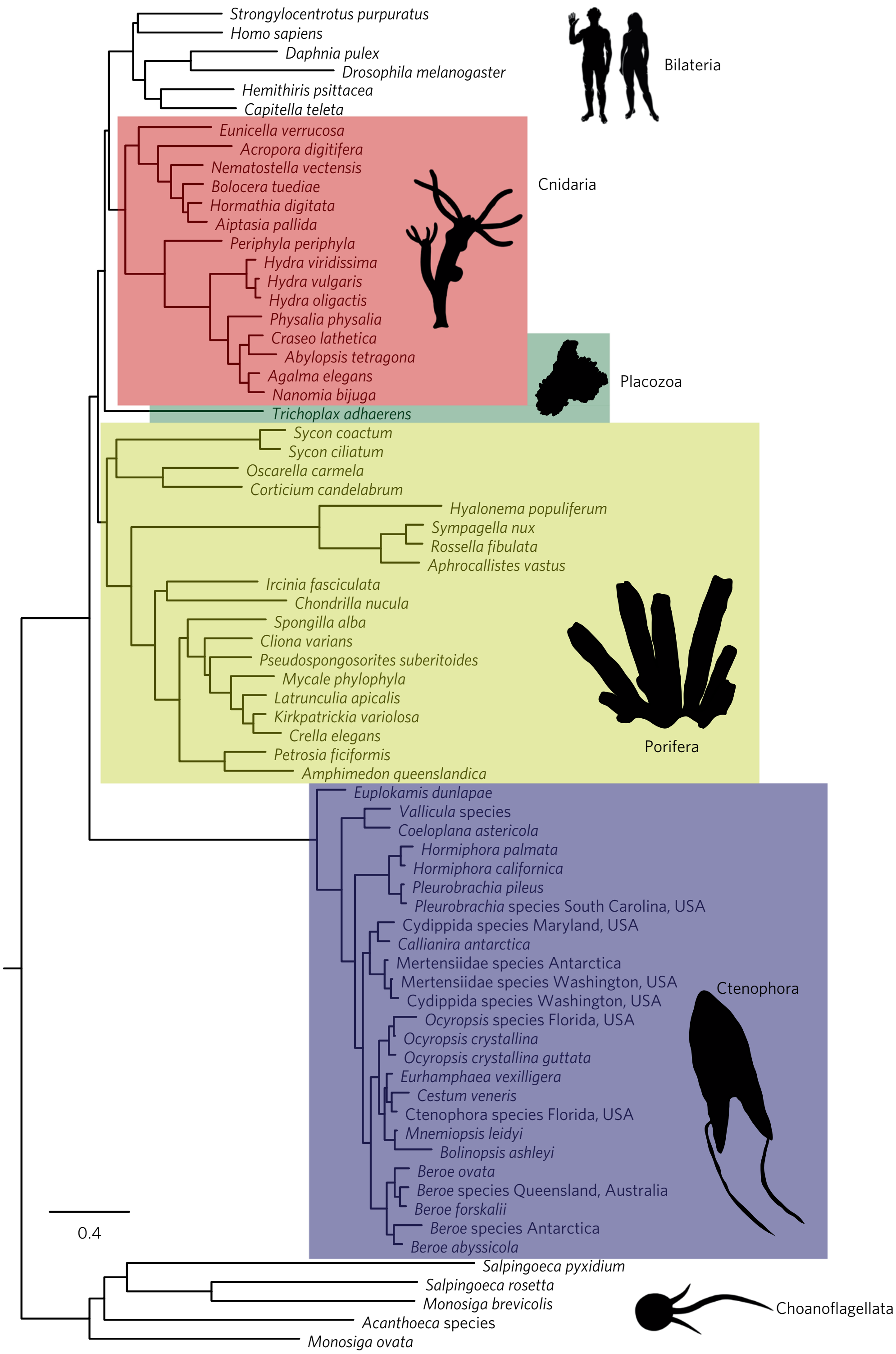

Fig. 2 | Relationships among metazoans inferred with the CAT-GTR substitution model and dataset Metazoa_Choano_RCFV_strict. All nodes have 100% PP. Inferred relationships among phyla are identical to those inferred by other models and datasets (Supplementary Figs. 1–15 and Supplementary Discussion). Scale bar is in expected substitutions per site. The silhouette images were downloaded from http://phylopic.org/.

twice (in Porifera and Placozoa) or, more parsimoniously, muscles evolved independently at least twice (in Ctenophora and the lineage leading to Cnidaria and Bilateria)[3].

The MRCA of extant ctenophores was most likely pelagic (PP = 0.91; Fig. 4 and Supplementary Fig. 24) with cydippid-like morphology (that is, an ovate body and branched tentacles; PP = 0.92; Fig. 3, Supplementary Fig. 20 and Supplementary Discussion) and a simultaneous hermaphrodite (PP = 0.99; Supplementary Fig. 22). Ancestral state reconstruction suggests plesiomorphy of the cydippid body plan, with most other morphotypes evolving from it (Fig. 3 and Supplementary Fig. 20). The one exception appears to be the ribbon-like Cestida, which evolved from a lobate-like ancestor.

Table 1 | Number of genes and sites in each dataset supporting alternative hypotheses of the sister lineage to all other metazoans

Dataset[a]	Genes supporting Ctenophora as sister lineage	Genes supporting Porifera as sister lineage	Sites supporting Ctenophora as sister lineage	Sites supporting Porifera as sister lineage
Metazoa_full	144 (64.3)	80 (35.7)	38,378 (56.4)	29,684 (43.6)
Metazoa_RCFV_relaxed	133 (64.8)	72 (35.2)	36,255 (55.5)	29,072 (44.5)
Metazoa_RCFV_strict	70 (60.3)	46 (39.7)	22,897 (52.9)	20,415 (47.1)
Metazoa_LB_relaxed	112 (68.3)	52 (31.7)	28,642 (55.9)	22,554 (44.1)
Metazoa_LB_strict	105 (69.5)	46 (30.5)	25,875 (55.1)	21,071 (44.9)
Metazoa_RCFV_LB_relaxed	97 (65.1)	52 (34.9)	26,647 (54.3)	22,389 (45.7)
Metazoa_RCFV_LB_strict	53 (71.6)	21 (28.4)	15,194 (52.8)	13,558 (47.2)
Metazoa_Choano	144 (61.5)	90 (38.5)	41,971 (55.3)	33,850 (44.7)
Metazoa_Choano_RCFV_relaxed	111 (68.9)	50 (31.3)	32,434 (54.3)	27,247 (45.7)
Metazoa_Choano_RCFV_strict	87 (68.5)	40 (31.5)	27,257 (55.2)	22,131 (44.8)
Metazoa_Choano_LB_relaxed	104 (56.8)	79 (43.2)	33,268 (54.8)	27,417 (45.2)
Metazoa_Choano_LB_strict	156 (75.4)	51 (32.7)	29,875 (59.2)	20,595 (40.8)
Metazoa_Choano_RCFV_LB_relaxed	83 (63.8)	47 (36.2)	26,586 (54.2)	22,493 (45.8)
Metazoa_Choano_RCFV_LB_strict	56 (68.3)	26 (31.7)	17,873 (57.3)	13,334 (42.7)

[a] See Supplementary Table 3 for more information on the datasets. LB, branch length heterogeneity; RCFV, relative composition frequency variability. Data are given as number (%).

Aside from beroids, which are atentaculate at all life stages, all ctenophores for which larval information is available have a free-swimming larval stage with cydippid-like morphology[1,33]. However, Platyctenids and to a lesser extent lobates and cestids undergo considerable morphological and functional changes during development[33]. Nevertheless, juvenile morphology among all ctenophores except the derived beroids resembles the inferred ancestral state of extant ctenophores (Fig. 3 and Supplementary Fig. 20).

Ancestral state reconstruction indicates that ctenophores have transitioned from a pelagic to benthic or semi-benthic adult lifestyle at least twice (Fig. 4 and Supplementary Fig. 25). These two transitions occurred on the branches leading to Platyctenida and *Lobatolampea*, but we cannot rule out additional transitions in undescribed benthic lineages. Interestingly, *Lobatolampea* was recovered as the sister lineage to a clade with all other lobates and Cestida, while Platyctenida was recovered as sister to all other ctenophores but *Euplokamis*. Thus, the two benthic lineages evolved separately. The transition between benthic and pelagic lifestyles has been studied in numerous invertebrate groups[34], with most documented transitions occurring from a benthic to pelagic existence. However, we found no evidence that any ancestrally benthic ctenophore lineage has evolved to occupy the water column (Fig. 4 and Supplementary Fig. 25).

Pleurobrachiidae is one of the most common and well-studied groups of ctenophores and is often used as a reference for the phylum[3,35]. However, Pleurobrachiidae lacks bioluminescence[35], and past uncertainty about the phylogenetic position of the family has limited our ability to fully analyse the evolution of bioluminescence in ctenophores[12,13]. We confidently recovered Pleurobrachiidae (that is, *Pleurobrachia* and *Hormiphora*) plus Pukiidae as a monophyletic lineage on a relatively long branch (Fig. 3 and Supplementary Figs. 16–19). Like Pleurobrachiidae, Pukiidae is incapable of bioluminescence. Ancestral state reconstruction suggests that the MRCA to extant ctenophores was bioluminescent ($PP = 0.96$; Supplementary Fig. 25 and Supplementary Discussion), and this trait has likely been lost only once within Ctenophora. Bioluminescence is generally considered advantageous in deep water[36], but most pleurobrachiids are found near the shore at shallow depths[1,37,38], which may have relaxed the selective pressures for maintaining bioluminescence.

The MRCA of extant ctenophores likely fed by capturing plankton with branched tentacles equipped with colloblasts—a unique synapomorphy of ctenophores. However, multiple transitions in the adult feeding mode have occurred (Fig. 5, Supplementary Fig. 20 and Supplementary Discussion). These transitions are associated with lineage-specific behavioural and morphological innovations[38]. For instance, the simplification of tentacles seen in *Dryodora*, followed by the complete loss of tentacles in *Beroe* (Fig. 5 and Supplementary Figs. 20 and 21), is associated with engulfing larger prey items, rather than using tentacles and/or lobes for food capture as in other lineages (Fig. 5 and Supplementary Fig. 20); although *Dryodora* has tentacles, they are likely used for sensing rather than capturing prey (Supplementary Discussion). The sister relationship between *Dryodora* and *Beroe* suggests a gradual transition from branched to reduced tentacles, followed by complete loss of tentacles. More broadly, ancestral state reconstruction of feeding behaviours produced three nodes where no character state had PPs of 90% or greater (Fig. 5 and Supplementary Fig. 20). Ambiguity at these nodes is associated with a clear shift away from using primarily, or only, tentacles for prey capture as adults and dramatic morphological transitions.

Discussion

Using greater ctenophore taxon sampling than previous studies, data filtering schemes to remove potential causes of systematic error and a variety of substitution models, we recovered Ctenophora as the sister lineage to all other animals. The debate surrounding the phylogenetic placement of Ctenophora has complicated studies on the evolution of complex characters such as muscles and neurons. Genomic components of these features suggest extensive convergent and parallel evolution across Metazoa[3], which is further supported by our phylogenetic results. However, events of independent origins of neural and muscular systems are not directly coupled with competing hypotheses of metazoan phylogeny[3,39,40]. Nevertheless, the placement of Ctenophora as the sister lineage to all other animals appears to be robust to error.

Our results suggest that Ctenophora has undergone a species-diversity bottleneck considerably longer ago than was previously hypothesized (Supplementary Fig. 15). Subsequent diversification resulted in numerous morphotypes evolving from a cydippid-like ancestor (Fig. 3). A benthic lifestyle has evolved convergently in at least two ctenophore lineages (Fig. 4), but the evolution of striated muscles, loss of bioluminescence and loss of tentacles throughout

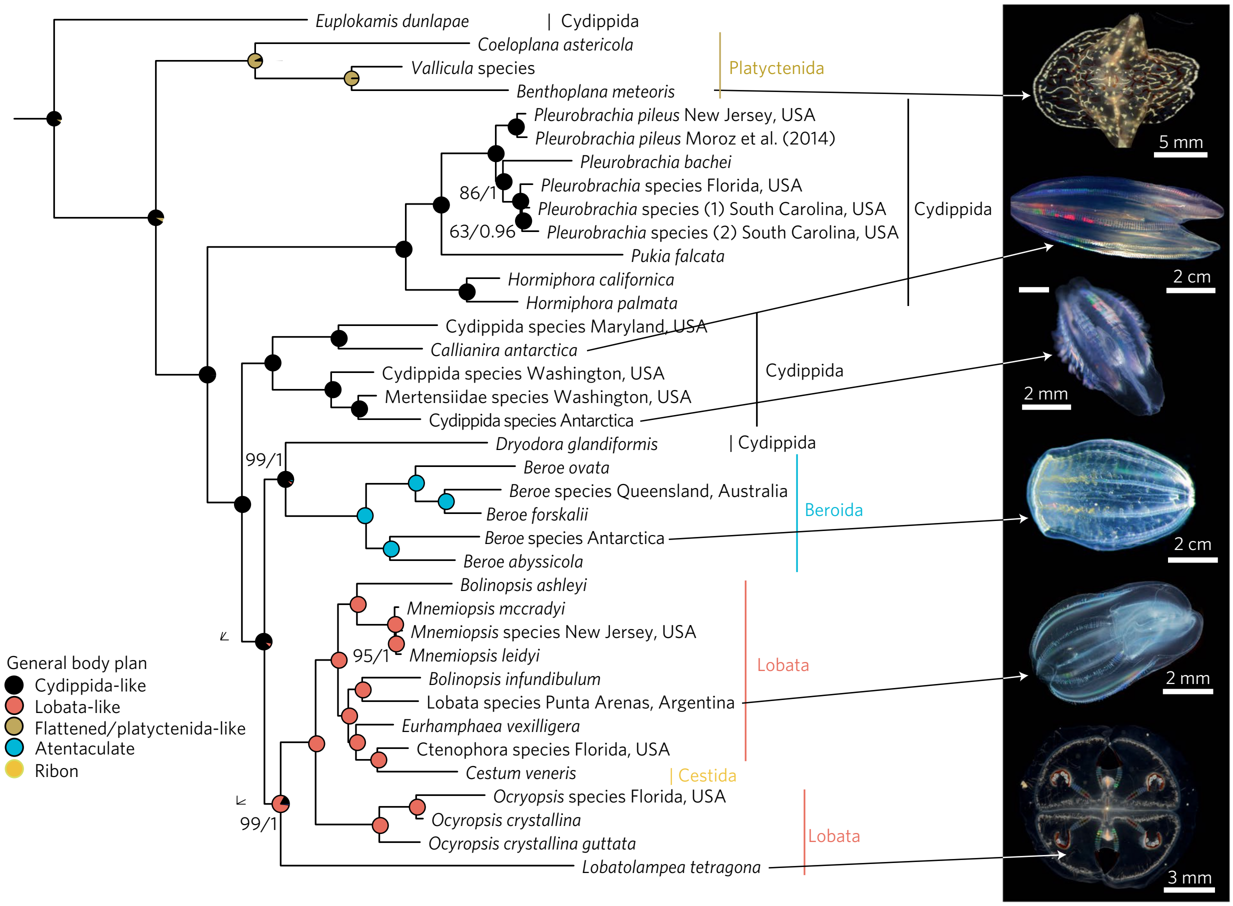

Fig. 3 | Evolutionary relationships among Ctenophora and ancestral character state reconstruction of the general body plan. Traditional orders are labelled with colours matching their corresponding body plan morphotype. Nodes are labelled with pie charts depicting the PP of character states. Phylogeny was inferred using dataset Ctenophore_RCFV_LB. Arrows connect the photographs of exemplars with species identities. Sponge and cnidarian outgroups that were used to root the tree have been removed for illustrative purposes. Nodes have 100% BS or 1.00 PP support unless otherwise noted (BS/PP).

all life cycles appears to have only occurred once (Supplementary Figs. 20–24). Ctenophora is in need of thorough taxonomic revision and we expect progress to be made on that front in the coming years. Ctenophora is one of the most morphologically diverse and under-studied metazoan groups, and our results provide a phylogenetic foundation for future studies on developmental, neuromuscular and tissue/organ evolution both within Ctenophora and among all metazoans.

Methods

Taxon sampling and sequencing. We sampled ctenophores from locations around the world (Table 1), mostly between 2013 and 2016. Ctenophore specimens were identified to as low a taxonomic level as possible (Table 1). Many newly sequenced species, particularly those sampled from Antarctica, were undescribed species. Complementary DNA libraries for newly collected ctenophores were constructed using a template-switch method using the SMART complementary DNA library construction (catalogue number 639537; Clontech). Full-length complementary DNA was amplified using the Advantage 2 PCR system (catalogue number 639201; Clontech) and the minimum number of polymerase chain reaction cycles necessary for single-end sequencing for Ion Proton or 2 × 100-bp paired-end sequencing with Illumina. Illumina and Ion Proton sequencing libraries were subsequently prepared using a NEBNext Ultra DNA Library Prep Kit for Illumina (catalogue number E7645S; New England Biolabs) or a NEBNext Fast DNA Library Prep Set for Ion Torrent (catalogue number E6270S; New England Biolabs). Each library was sequenced using either an Illumina NextSeq 500 or Ion Proton (see Table 1).

Publicly available ctenophore and non-ctenophore transcriptomes or gene models were retrieved from the National Center for Biotechnology Information and other databases (Supplementary Tables 1 and 2). *Bolinopsis infundibulum* from Moroz et al.[3] was determined to be misidentified based on our sequencing of a novel *B. infundibulum* transcriptome. Thus, we now use the name 'Cydippida species Washington, USA' for the transcriptome labelled '*Bolinopsis infundibulum*' in Moroz et al.[3].

We performed phylogenetic analyses at two scales to achieve different goals. First, we inferred relationships among non-bilaterian metazoan phyla (with other opishtokonts as outgroups) to determine the sister lineage to Ctenophora. Second, we analysed relationships and trait evolution within Ctenophora using appropriate outgroups as identified in the broader Metazoa analyses. Depending on the focal taxonomic scale, different taxon-sampling schemes were used (see Supplementary Tables 1 and 2). Datasets designed to examine the relationships between metazoan phyla are named with the prefix 'Metazoa_' followed by more specific information about the dataset as appropriate. For example, datasets with only choanoflagellates as outgroups are named 'Metazoa_choano_'. Datasets designed to test relationships among ctenophores are named in a similar fashion except they have the prefix 'Ctenophore_'. See Supplementary Table 3 and below for additional information about dataset naming conventions.

When testing the relationships among metazoan phyla, the taxon sampling used was similar to that of Whelan et al.[11] with three exceptions. First, fewer bilaterians were included to decrease the computational time. Second, a larger number of choanoflagellates was sampled, which we expected to result in more robust rooting of Metazoa than in past analyses[3,5,8–11,24,41–43]. Finally, more ctenophores were sampled than in previous studies[3,5,8–11,24,41–43] in an attempt to increase the accuracy of ctenophore placement[28,29,44].

For analyses that focused on the relationships among metazoan phyla, we generated datasets that only had choanoflagellate outgroups and datasets that had Ichthyosporea, Filasterea and choanoflagellate outgroups (Supplementary Tables 2 and 3). These datasets included fewer ctenophores than the datasets generated to test the relationships within Ctenophora because we did not include individuals that were repetitive at or near the species level (for example, only one individual identified as *Pleurobrachia bachei* was included in the broader analyses; Supplementary Tables 1 and 2). This was done to decrease the required computational time. *Pukia falcata* was also not included in the broad metazoan analyses, despite its inclusion in ctenophore-centric phylogenetic inference,

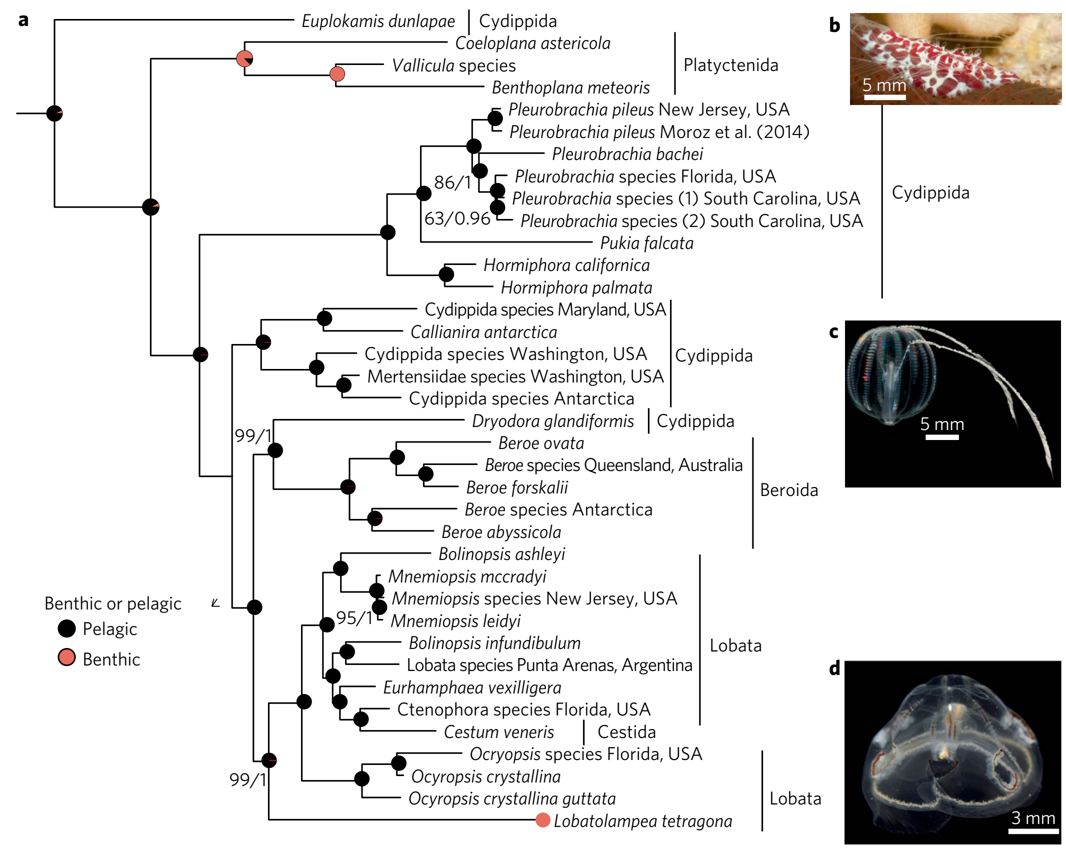

Fig. 4 | Evolutionary relationships of Ctenophora and ancestral character state reconstruction of benthic versus pelagic lifestyle. Nodes (and unique taxa) are labelled with pie charts depicting the PP of character states. Traditional orders are labelled. **a**, Phylogeny was inferred using the dataset Ctenophore_RCFV_LB. Sponge and cnidarian outgroups that were used to root the tree have been removed for illustrative purposes. Nodes have 100% BS or 1.00 PP support unless otherwise noted (BS/PP). **b**, Benthic Platyctenida; *C. astericola* on a seastar. **c**, Pelagic *P. bachei*. **d**, Benthic Lobata; *L. tetragona*.

because preliminary phylogenetic inference (not shown) revealed that its inclusion caused unstable relationships among metazoan phyla. Presumably, this was due to the comparably high amount of missing data for *P. falcata*. 'Mertensiidae species (Antarctica)' was inadvertently not included in the ctenophore-specific dataset generation. However, inclusion of this species would likely not have affected the overall conclusions about ctenophore evolution given its inferred placement from analyses using the metazoan datasets (Fig. 2 and Supplementary Figs. 1–14).

Informatics and data matrix assembly. Before assembly, raw transcriptome reads were digitally normalized to a target of 30× coverage using normalize-by-median.py (ref. [45]) and assembled using Trinity 20140717 (ref. [46]) with default parameters. After assembly, open reading frames and putative protein sequences were identified with TransDecoder[46] using default parameters. We used HaMStR 13.2 (ref. [47]) and two core ortholog sets to recover orthologous groups (OGs) for phylogenomic analyses (Supplementary Table 3). The model organism core ortholog set packaged with HaMStR 13.2 was used for testing the relationships among metazoan phyla because it was designed to be of broad taxonomic utility. For reconstructing ctenophore phylogeny, we designed a ctenophore-centric core ortholog set to increase the number of OGs in our datasets (Supplementary Table 3).

The ctenophore-centric core ortholog set was created by first performing an all-versus-all blastp search[48] among the transcriptomes of *Beroe abyssicola*, *Coeloplana astericola*, *E. dunlapae*, *Mnemiopsis leidyi*, *Ocyropsis* species from Florida, USA, and *P. bachei*. These species were chosen because they were hypothesized to represent a wide swathe of ctenophore phylogeny and had relatively deeply sequenced transcriptomes. An e-value cut-off of 10^5 was used for the blastp searches. The blastp results were used to perform Markov clustering using OrthoMCL 2.0 (ref. [49]) with an inflation parameter of 2.1 following Hejnol et al.[10] and Kocot et al.[50]. Markov clustering resulted in 55,433 putative OGs. These OGs were further filtered to remove possible paralogs and low-quality OGs. First, any sequence that was fewer than 100 amino acids in length was removed. Each OG was then aligned with MAFFT (ref. [51]) using an automatically chosen alignment strategy and a 'maxiterate' value of 1,000. After alignment, an approximately maximum likelihood tree was generated for each OG using FastTree 2 (ref. [52]) with 'slow' and 'gamma' options. Each tree and corresponding OG was processed using PhyloTreePruner[53] to screen for paralogs; a bootstrap value of 90 was used for collapsing nodes. If more than one sequence for any of the six respective species was present after the paralog pruning step, the longest sequence for that species was retained and others were discarded. Lastly, we removed OGs that had sequences for fewer than four species and any OG that did not have an *M. leidyi* sequence because it was chosen as the HaMStR primer taxon. The 2,354 remaining OG alignments were used to build protein hidden Markov models using the HMMER tools hmmbuild and hmmcalibrate[54]. Our ctenophore core ortholog set has been deposited on figshare (https://figshare.com/).

Transcriptomes and gene models were processed using HaMStR with one or both core ortholog sets (that is, model organism or ctenophore) depending on which analyses each taxon was included in (Supplementary Tables 1 and 2). Post-HaMStR orthology filtering followed Whelan et al.[11] with slight script modifications to increase speed and accuracy. For datasets generated to infer relationships among Bilateria and non-Bilateria phyla, OGs were discarded if they had fewer than 42 taxa present for datasets generated with all outgroups and fewer than 38 taxa present for datasets generated with only choanoflagellate outgroups (that is, the datasets Metazoa_full and Metazoa_Choano, respectively; Supplementary Table 2). For the datasets designed for testing the relationships among ctenophores, OGs were discarded if they had fewer than 27 taxa present.

After orthology filtering of each dataset, single gene trees were generated using RAxML 8.2.4 (ref. [55]) with a gamma distribution to model rate heterogeneity and amino acid substitution models identified by model testing implemented in RAxML. We performed 100 fast bootstrap replicates for each gene tree to assess nodal support. The resulting gene trees were used with TreSpEx (ref. [56]) for more thorough screening of paralogs and contamination that may have passed through initial orthology determination. Briefly, we used the BLAST-associated method in TreSpEx with the packaged *Capitella teleta* and *Helobdella robusta* blast databases following Struck[56] and Whelan et al.[11]. All sequences identified as certain or uncertain paralogs by TreSpEx—such sequences may also be non-target sequence contamination—were removed from the OGs. Subsequently, OGs that then had fewer than 42 taxa for the dataset Metazoa_Full, 38 taxa for the dataset Metazoa_Choano and 27 taxa for the dataset Ctenophore_full after paralog pruning with TreSpEx were also discarded to minimize the missing data. For clarity, the datasets

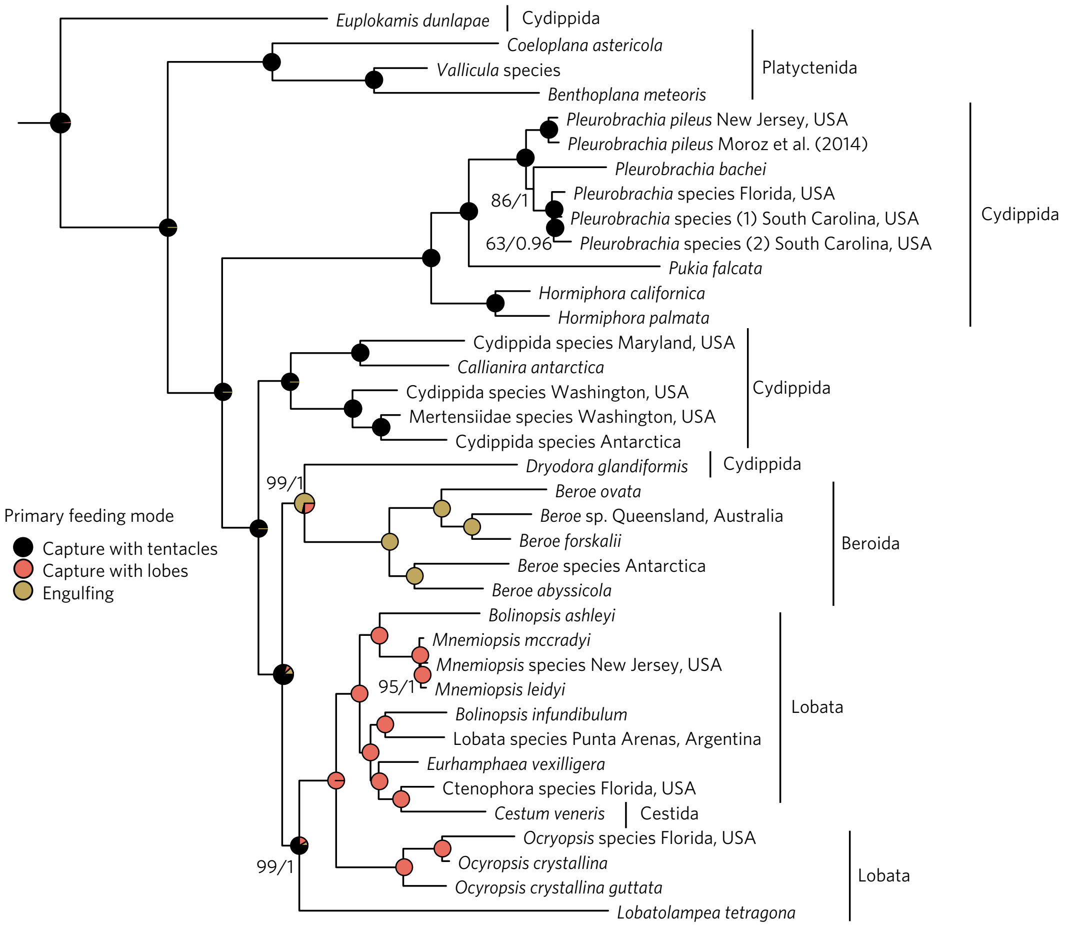

Fig. 5 | Evolutionary relationships of Ctenophora and ancestral state reconstruction of the primary feeding mode. Traditional orders are labelled. Nodes are labelled with pie charts depicting the PP of character states. Phylogeny was inferred using the dataset Ctenophore_RCFV_LB. Sponge and cnidarian outgroups that were used to root the tree have been removed for illustrative purposes. Nodes have 100% BS or 1.00 PP support unless otherwise noted (BS/PP).

Metazoa_full, Metazoa_Choano and Ctenophore_full are herein referred to as 'initial' datasets that were then filtered for OGs that had the highest potential for causing systematic error.

Systematic error. To assess the effect of systematic error on phylogenetic inference, we generated datasets with potential sources of systematic error removed. Specifically, genes with the highest potential for causing long-branch attraction or that had the highest levels of base compositional heterogeneity were removed. By creating nested datasets with different potential causes of systematic error removed, we were able to assess whether inferred relationships were influenced by systematic error. Branch length heterogeneity scores (LB), which can be used to rank genes based on their possible contribution to long-branch attraction, were calculated using TreSpEx. This was done with individual trees for each OG in the three initial datasets; new trees for each paralog-pruned OG were inferred using RAxML as described above. Density plots of LB score heterogeneity and upper-quartile LB score for each OG and dataset were plotted using R (ref. [57]) (Supplementary Fig. 26). The two datasets designed to test relationships among metazoan phyla (that is, Metazoan_full and Metazoa_Chaono) had fewer genes than the ctenophore-centric dataset. Thus, to strike a balance between removing OGs that may cause systematic error and not having enough phylogenetic signal (that is, OGs) to accurately resolve relationships, we identified a strict and relaxed cut-off for removing genes with outlier LB scores (Supplementary Fig. 26). For the ctenophore-centric dataset, we only identified one set of genes as outliers (Supplementary Fig. 26). Using the initial datasets, nested datasets were generated by removing genes that were identified as having outlier LB scores (Supplementary Table 3). Relative composition frequency variability (RCFV)[58], which is a measure of how much base compositional heterogeneity is present in an OG, was calculated for each gene using BaCoCa[59]. A density plot of RCFV for each initial dataset was plotted in R (Supplementary Fig. 26). As with the LB scores, for the datasets Metazoa_full and Metazoa_Choano, two sets of outliers were identified and removed to create datasets with all outlier RCFV genes removed (strict) and some outlier RCFV genes removed (relaxed) (Supplementary Fig. 65 and Supplementary Table 3). Only a single set of RCFV outlier genes was identified for the dataset Cteno_full (Supplementary Fig. 26 and Supplementary Table 3). We also created datasets from the initial three datasets that had both LB and RCFV outlier genes removed (Supplementary Fig. 26 and Supplementary Table 3). For the ctenophore-centric datasets, we created corresponding datasets with outgroups removed to test whether or not the relationships among ctenophores were affected by relatively distantly related outgroups.

Phylogenetic reconstructions. Bayesian inference with the site-heterogeneous CAT-GTR substitution model was done with PhyloBayes MPI (ref. [60]). Analyses with CAT-GTR are notoriously time consuming[24] so a number of steps were taken to facilitate convergence of independent Bayesian runs. First, only two datasets were analysed with CAT-GTR: the dataset Metazoa_Choano_RCFV_strict for testing the relationships among metazoan phyla and the dataset Cteno_RCFV_LB for determining the relationships among ctenophore lineages. We removed three ctenophore taxa from the dataset Metazoa_Choano_RCFV_strict to facilitate convergence; these three ctenophores were unstable in preliminary CAT-GTR analyses that failed to converge (see Supplementary Tables 1–3). For CAT-GTR analyses on both datasets, two independent chains were sampled every generation. Trace plots of Markov chain Monte Carlo runs were visually inspected in Tracer version 1.6 (http://tree.bio.ed.ac.uk/software/tracer/) to assess stationarity and appropriate burn-in, which was determined to be 3,500 and 4,000 generations for the datasets Metazoa_Choano_RCFV_strict and Cteno_RCFV_LB, respectively. PhyloBayes runs were sampled for 18,436 generations on the dataset Metazoa_Choano_RCFV_strict and for 23,947 generations on the dataset Cteno_RCFV_LB. All parameters and tree shapes reached convergence, which was considered to have occurred when the maxdiff value was less than 0.1 as measured by bpcomp[60] and when the rel_diff value was less than 0.3 and the effective sample size was greater than 50 as measured by tracecomp[60]. Although some have advocated the use of CAT-F81 when CAT-GTR is deemed computationally prohibitive[20,61], Whelan and

Halanych[24] recently showed the CAT-F81 can result in critically inaccurate trees. Thus, tree inference was not done with the CAT-F81 model on datasets that would have been too computationally demanding for analyses with CAT-GTR.

Maximum likelihood trees for each dataset were inferred using site-homogeneous amino acid substitution models coupled with data partitioning[62]. Best-fit partitions and amino acid substitution models for each dataset were inferred using PartitionFinder 2.0 (ref. [63]) with 20% relaxed clustering[64], the rcluster_f command and Bayesian information criteria. Maximum likelihood phylogenetic inference using best-fit partitions and amino acid substitution models was done with RAxML 8.2.4 (ref. [55]). A discrete gamma distribution with four categories was used on each partition for modelling rate heterogeneity. Nodal support was assessed with 100 fast bootstrap replicates. Files with best-fit partitions and models for each dataset have been deposited on FigShare (https://figshare.com/).

Measuring support for competing hypotheses of non-bilaterian relationships. The number of genes and sites favouring each of the two competing hypotheses—sponges as the sister group to all other extant metazoans and ctenophores as a sister to all other metazoans—was assessed under a maximum likelihood framework. For each metazoan dataset, site-wise likelihood scores were inferred for both hypotheses with RAxML 8.2.4 (option -f G). The same partitioning schemes and models used in the original tree inference were used. The two different phylogenetic hypotheses passed to RAxML (via -z) were the tree inferred with RAxML (that is, the ctenophore as the sister lineage tree) and the corresponding tree that was modified to have sponges as the sister to all other metazoans; constraints were done by modifying the original tree in Mesquite 3.2 (http://mesquiteproject.org). The numbers of genes and sites supporting each hypothesis were calculated with RAxML output and Perl scripts from Shen et al.[26]

Molecular clock analyses. Past authors have hypothesized that Ctenophora underwent a species bottleneck, possibly as recently as 65 MYA[12,13,65]. However, the bottleneck hypothesis has not been tested with molecular clock methods. BEAST 2 (ref. [66]) is a well-tested and widely used program that implements molecular clock models, but analyses with amino acids can be prohibitively slow. Thus, for molecular clock analyses, we used our smallest dataset, Metazoa_Choano_RCFV_LB_strict[67]. We also trimmed the same taxa that were deemed unstable for analyses with CAT-GTR (see above and Supplementary Table 2). The same amino acid substitution models and best-fit partitions were inferred with PartitionFinder using 20% relaxed clustering with the rcluster_f command. The best-fit number of relaxed molecular clock models for use in BEAST 2 were inferred with ClockstaR[68] using default parameters. One molecular clock was inferred to be most appropriate for this dataset. A relaxed molecular clock with a lognormal distribution[69] and a Yule tree model were used. A calibration was placed on the node representing the MRCA of Metazoa using a normal distribution with a mean of 750 MYA and a standard deviation of 35 following the findings of dos Rios et al.[70]; monophyly of Metazoa was enforced. We only used one calibration point for the molecular clock analysis, even though this may result in inaccurate absolute branching time estimates. We attempted to perform analyses with a greater number of node-age calibrations (for example, for sponges, cnidarians and bilaterian lineages; see Supplementary Table 4)[70], but Bayesian analyses failed to show evidence of convergence after over four months of run time. However, a single calibration point still allows for inference of relative timing of extant ctenophore diversification compared with better-studied lineages and lineages with better fossil records. Thus, even if the absolute timing of diversification events is imprecise in our molecular clock tree inference, we can analyse the inferred timing of ctenophore diversification relative to well-studied diversification events for which the timing of diversification is reasonably well known (for example, Bilateria, protosomes) to estimate the age of the extant ctenophore MRCA.

Molecular clock analyses with BEAST 2 consisted of two independent runs with 27,246,750 Markov chain Monte Carlo generations sampled every 250 generations. Trace plots were viewed in tracer, burn-in was visually determined (12% for run 1 and 50% for run 2). Convergence was checked and confirmed by comparing trace plots in Tracer making sure the effective sample size of each parameter was greater than 50 and that stationarity appeared to have been achieved; most parameters had effective sample sizes well in excess of 200. A maximum clade credibility tree with median heights was calculated using TreeAnnotater[66]. Bayesian phylogenetic inference using a molecular clock resulted in identical branching patterns among phyla as phylogenetic inference with RAxML and PhyloBayes (for example, Ctenophora as the sister to all other animals, $PP = 1.00$; Supplementary Figs. 1–15).

Ancestral state reconstruction. We performed ancestral state reconstruction for the following traits: (1) general body plan (that is, 'cydippid-like', 'lobata-like', Platyctenida or Cestida; Fig. 3 and Supplementary Fig. 20), (2) primary food capture mode (that is, with tentacles, with body lobes or engulfing prey with a comparatively large mouth; Supplementary Fig. 20), (3) the presence or absence of tentacles as adults (Supplementary Fig. 21), (4) the presence or absence of dioecy (Supplementary Fig. 22), (5) the presence or absence of striated muscles (Supplementary Fig. 23), (6) the presence or absence of smooth muscle (Supplementary Fig. 23), (7) a pelagic versus benthic or semi-benthic lifestyle (Supplementary Fig. 24), (8) ability to bioluminesce (Supplementary Fig. 24) and (9) the presence or absence of tentacles throughout the life cycles (Supplementary Fig. 21). Characteristics were assigned using previous descriptive work[2,6,17,19,37,38,71–77] and/or personal observations of individuals we collected (see Supplementary Table 5). Additional information about trait assignment can be found in the Supplementary Discussion. The phylogenetic signal of each trait was measured with Blomberg's K (ref. [78]) using the phytools 0.5–10 (ref. [79]) package in R (ref. [57]); each trait had a significant phylogentic signal ($P < 0.05$).

Stochastic mapping of character evolution, a Bayesian method for ancestral state reconstruction[80,81], was performed to generate character state joint probabilities on the phylogeny inferred with the dataset Cteno_RCFV_LB. This was done in R using phytools 0.5–10. Uncertainty in relationships was ignored because the only uncertain nodes were those at the tips among closely related taxa with identical character states (Fig. 3, Supplementary Figs. 16–19 and Supplementary Table 5). Analyses that incorporated uncertainty in branch lengths were effectively the same as those that ignored uncertainty (Supplementary Discussion). For ancestral state reconstruction, a Cydippida species from Friday Harbor was removed because it was labelled as *B. infundibulum* by Moroz et al.[3] and we could not confidently assign character states given the misidentification. The larval ctenophore specimen (Ctenophora species) was also removed because many character states that would be present only in adults were undetermined. These tips were removed from trees using the R package Ape[82]. Outgroups were removed from all stochastic mapping analyses except the presence or absence of striated and smooth muscle. The best-fit model of character evolution to be used for stochastic mapping was determined by fitting an equal rates model, a symmetrical model, and an all rates different model to each character state dataset using the R package Geiger[83]; corrected Akaike information criteria were used to determine the best-fit model for each respective character dataset. For each analysis, the previous probability of the root's character state was estimated directly from the data and the Bayesian Markov chain Monte Carlo method was used to generate a PP distribution for the character transition matrix. With these parameters, 1,000 stochastic maps were generated for each trait. The evolution of traits was visualized by displaying pie charts of PPs for each character state on every node.

Code availability. The code that supports the findings of this study is available from http://github.com/nathanwhelan.

Data availability. The transcriptomes sequenced as part of this study are available in the National Center for Biotechnology Information Short Read Archive under BioProject PRJNA396415. The 18S ribosomal RNA sequences have been uploaded to GenBank under accession numbers MF599304–MF599336. Datasets, model partitions and tree files have been uploaded to FigShare (https://figshare.com/).

Received: 23 January 2017; Accepted: 31 August 2017;
Published online: 09 October 2017

References

1. Hyman, L. *The Invertebrates* Vol. 1 (McGraw-Hill, New York, 1940).
2. Mackie, G. O., Mills, C. E. & Singla, C. L. Structure and function of the prehensile tentialla of *Euplokamis* (Ctenophora, Cydippida). *Zoomorphology* **107**, 319–337 (1988).
3. Moroz, L. L. et al. The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).
4. Roohi, A. et al. Changes in biodiversity of phytoplanton, zooplankton, fishes and macrobenthos in the Southern Caspian Sea after the invasion of the ctenophore *Mnemiopsis leidyi*. *Biol. Invasions* **12**, 2342–2361 (2010).
5. Ryan, J. F. et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 1242592 (2013).
6. Harbison, G. R. in *The Origins and Relationships of Lower Invertebrates* (eds Morris, S. C., George, J. D., Gibson, R. & Platt, H. M.) 78–100 (Oxford Univ. Press, Oxford, 1985).
7. Dunn, C. W., Leys, S. & Haddock, S. H. D. The hidden biology of sponges and ctenophores. *Trends Ecol. Evol.* **30**, 282–291 (2015).
8. Whelan, N. V., Kocot, K. M. & Halanych, K. M. Employing phylogenomics to resolve the relationships among cnidarians, ctenophores, sponges, placozoans and bilaterians. *Integr. Comp. Biol.* **55**, 1084–1095 (2015).
9. Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).
10. Hejnol, A. et al. Assessing the root of bilaterian animals with scalable phylogenomic models. *Proc. Biol. Sci.* **276**, 4261–4270 (2009).
11. Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl. Acad. Sci. USA* **112**, 5773–2778 (2015).
12. Podar, M., Haddock, S. H. D., Sogin, M. L. & Harbison, G. R. A molecular phylogenetic framework for the phylum Ctenophora using 18S rRNA genes. *Mol. Phylogen. Evol.* **21**, 218–230 (2001).

13. Simion, P., Bekkouche, N., Jager, M., Quéinnec, E. & Manuel, M. Exploring the potential of small RNA subunit and ITS sequences for resolving the phylogenetic relationships within the phylum Ctenophora. *Zoology* **118**, 102–114 (2015).
14. Tang, F., Bengtson, S., Wang, Y., Wang, X.-l. & Yin, C.-Y. *Eoandromeda* and the origin of Ctenophora. *Evol. Dev.* **13**, 408–414 (2011).
15. Morris, S. C. & Collins, D. H. Middle Cambiran ctenophores from Stephen Formation British Columbia, Canada. *Phil. Trans. R. Soc. Lond. B* **351**, 279–308 (1996).
16. Chen, J.-Y. et al. Raman spectra of a Lower Cambrian ctenophore embryo from southwestern Shaanxi, China. *Proc. Natl Acad. Sci USA* **104**, 6289–6292 (1997).
17. Mills, C. E. Revised classification of the genus *Euplokamis* Chun, 1880 (Ctenophora: Cydippida: Euplokamidae n. fam.) with a description of the new species *Euplokamis dunlapae*. *Can. J. Zool.* **65**, 2661–2668 (1987).
18. Harbison, G. R. & Miller, R. L. Not all ctenophores are hermaphrodites. Studies on the sytematics, distribution, sexuality and development of two species of *Ocyropsis*. *Mar. Biol.* **90**, 413–424 (1986).
19. Uyeno, D., Lasley, R. M., Moore, J. M. & Berumen, M. L. New records of *Lobatolampea tetragona* (Ctenophora: Lobata: Lobatolampeidae) from the Red Sea. *Mar. Biodivers. Rec.* **8**, e33 (2015).
20. Pisani, D. et al. Genomic data do not support comb jellies as the sister group to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 15402–15407 (2015).
21. Simion, P. et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).
22. Halanych, K. M., Whelan, N. V., Kocot, K. M., Kohn, A. B. & Moroz, L. L. Miscues misplace sponges. *Proc. Natl Acad. Sci. USA* **113**, E946–E949 (2016).
23. Moroz, L. L. & Halanych, K. M. A sisterly dispute: methodological misconceptions. *Nature* **529**, 286–287 (2016).
24. Whelan, N. V. & Halanych, K. M. Who let the CAT out of the bag? Accurately dealing with subtitutional heterogeneity in phylogenomics analyses. *Syst. Biol.* **66**, 232–255 (2017).
25. Arcila, D. et al. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nat. Ecol. Evol.* **1**, 0020 (2017).
26. Shen, X.-X., Hittinger, C. T. & Rokas, A. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
27. Heath, T. A., Hedtke, S. M. & Hillis, D. M. Taxon sampling and the accuracy of phylogenetic analyses. *J. Syst. Evol.* **46**, 239–257 (2008).
28. Hedtke, S. M., Townsend, T. M. & Hillis, D. M. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Syst. Biol.* **55**, 522–529 (2006).
29. Zwickl, D. J. & Hillis, D. M. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* **51**, 588–598 (2002).
30. Benton, M. J. & Twitchett, R. J. How to kill (almost all life): the end-Permian extinctoin event. *Trends Ecol. Evol.* **18**, 358–365 (2003).
31. Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol. Biol. Evol.* **21**, 1095–1109 (2004).
32. Steinmetz, P. R. H. et al. Independent evolution of striated muscles in cnidarians and bilaterians. *Nature* **487**, 231–234 (2012).
33. Martindale, M. Q. in *Atlas of Marine Invertebrate Larvae* (eds Young, C. M., Sewell, M. A. & Rice, M. E.) 109–122 (Academic Press, Cambridge, MA, 2002).
34. Rigby, S. & Milsom, C. Benthic origins of zooplankton: an environmentaly determined macroevolutionary effect. *Geology* **24**, 52–54 (1996).
35. Haddock, S. H. D. & Case, J. F. Not all ctenophores are bioluminescent: Pleurobrachia. *Biol. Bull.* **189**, 356–362 (1995).
36. Widder, E. A. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. *Science* **328**, 704–708 (2010).
37. Gershwin, L.-A., Zeidler, W. & Davie, P. J. F. Ctenophora of Australia. *Mem. Queensl. Mus.* **54**, 1–45 (2010).
38. Haddock, S. H. D. Comparative feeding behaviour of planktonic ctenophores. *Integr. Comp. Biol.* **47**, 847–853 (2007).
39. Moroz, L. L. Covergent evolution of neural systems in ctenophores. *J. Exp. Biol.* **218**, 598–611 (2015).
40. Moroz, L. L. & Kohn, A. B. Independent origins of neurons and synapses: insights from ctenophores. *Phil. Trans. R. Soc. Lond. B* **371**, 20150041 (2016).
41. Borowiec, M. L., Lee, E. K., Chiu, J. C. & Plachetzki, D. C. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* **16**, 987 (2015).
42. Cannon, J. T. et al. Xenacoelomorpha is the sister group to Nephrozoa. *Nature* **530**, 89–93 (2016).
43. Chang, E. S. et al. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proc. Natl Acad. Sci. USA* **112**, 14912–14917 (2015).
44. Nabhan, A. R. & Sarkar, I. N. The impact of taxon sampling on phylogenetic inference: a review of two decades of controversy. *Brief. Bioinformatics* **13**, 122–134 (2011).
45. Brown, T., Howe, C., Zhang, A., Pyrkosz, Q. & Brom, A. B. A reference-free algorithm for computational normalization of shotgun sequencing data. Preprint at https://arxiv.org/abs/1203.4802 (2012).
46. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
47. Ebersberger, I., Strauss, S. & von Haeseler, A. HaMStR: profile hidden Markov model based search for orthologs in ESTs. *BMC Evol. Biol.* **9**, 157 (2009).
48. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
49. Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
50. Kocot, K. M. et al. Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst. Biol.* **66**, 256–282 (2017).
51. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
52. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
53. Kocot, K. M., Citarella, M. R., Moroz, L. L. & Halanych, K. M. PhyloTreePruner: a phylogenetic tree-based approach for selection of orthologous sequences for phylogenomics. *Evol. Bioinform.* **9**, 429–435 (2013).
54. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
55. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
56. Struck, T. H. TreSpEx—detection of misleading signal in phylogenetic reconstructions based on tree information. *Evol. Bioinform.* **10**, 51–67 (2014).
57. R Development Core Team. *R: A Language and Environment for Statistical Computing* http://www.R-project.org/ (R Foundation for Statistical Computing, Vienna, 2015).
58. Zhong, M. et al. Detecting the symplesiomorphy trap: a multigene phylogenetic analysis of terebelliform annelids. *BMC Evol. Biol.* **11**, 369 (2011).
59. Kück, P. & Struck, T. H. BaCoCa—a heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Mol. Phylogen. Evol.* **70**, 94–98 (2014).
60. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
61. Nosenko, T. et al. Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogen. Evol.* **67**, 223–233 (2013).
62. Brown, J. M. & Lemmon, A. R. The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst. Biol.* **56**, 643–655 (2007).
63. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.*, **34**, 772–773 (2016).
64. Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82 (2014).
65. Jékely, G., Paps, J. & Nielsen, C. The phylogenetic position of ctenophores and the origin(s) of nervous systems. *EvoDevo* **6**, 1 (2015).
66. Bouckaert, R. et al. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
67. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
68. Duchêne, S., Molak, M. & Ho, S. Y. W. ClockstaR: choosing the number of relaxed-clock models in molecular phylogenetic analysis. *Bioinformatics* **30**, 1017–1019 (2014).
69. Drummond, A. J., Ho, S. Y. W., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
70. Dos Reis, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 29392950 (2015).
71. Brusca, R. C., Moore, W. & Shuster, S. M. in *Invertebrates* 3rd edn 1104 (Sinauer Associates, Sunderland, MA, 2016).
72. Eechkaut, I., Flammang, P., Bue, C. L. & Jangoux, M. Functional morphology of the tentacles and tentilla of *Coeloplana bannworthi* (Ctenophora, Platyctenida), and ectosymbiont of *Diadema setosum* (Echinodermata, Echinoida). *Zoomorphology* **117**, 165–174 (1997).
73. Harbison, G. R. & Madin, L. P. in *Synopsis and Classification of Living Organisms* (ed. Parker, S. P.) 707–715 (McGraw-Hill, New York, 1982).

74. Horita, T. An undescribed lobate ctenophore, *Lobatolampea tetragona* gen. nov. & spec. nov., representing a new family, from Japan. *Zool. Meded.* **73**, 457–464 (2000).
75. Matsumoto, G. I. & Harbison, G. R. In situ observations of foraging, feeding, and escape behavior in three orders of oceanic ctenophores: Lobata, Cestida, and Beroida. *Mar. Biol.* **117**, 279–287 (1993).
76. Purcell, J. E., Sturdevant, M. V. & Galt, C. P. *Response of Marine Ecosystems to Global Change: Ecological Impact of Appendicularians* (eds Gorsky, G., Yongbluth, M. J. & Deibel, D.) 359–435 (GB Science Publishers-Editions Scientifiques, Paris, 2005).
77. Stretch, J. J. Observations on the abundance and feeding behavior of the cestid ctenophore. *Velamen Parallelum. Bull. Mar. Sci.* **32**, 796–799 (1982).
78. Blomberg, S. P., Garland, T. Jr & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
79. Revell, L. J. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
80. Huelsenbeck, J. P., Nielsen, R. & Bollback, J. P. Stochastic mapping of morphological characters. *Syst. Biol.* **52**, 131–158 (2003).
81. Nielsen, R. Mapping mutations on phylogenies. *Syst. Biol.* **51**, 729–739 (2002).
82. Paradis, E., Claude, J. & Strimmer, K. APE: analysis of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
83. Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24, 129–131 (2008).

Acknowledgements

This work was made possible in part by a grant of high-performance computing resources and technical support from the Alabama Supercomputer Authority and was supported by the United States National Aeronautics and Space Administration (grant NASA-NNX13AJ31G), the National Science Foundation (grants ANT-1043670, ANT-1043745, 1557923, 1548121 and 1645219) and the Ministry of Education of the Russian Federation (#14W03.31.0015). We thank the International SeaKeepers Society, captains and crew of RV *Laurence M Gold*, RV *Capasetic*, RV *Penny Mae*, RV *Defiance*, RV *Basic Explorer*, RV *Harle* of Fleet Miami and RV *Miss Phebe II*, as well as J. Jacoby for their help with the collection of ctenophores around the globe. A. Kohn helped with laboratory work and data curation. C. Mills helped with species identifications. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the United States Fish and Wildlife Service. This is Molette Biology Laboratory contribution 70 and Auburn University Marine Biology Program contribution 166.

Author contributions

N.V.W., K.M.K., L.L.M. and K.M.H. designed the study. K.M.K. collected and identified the Australian speices. L.L.M. and G.P. collected and identified all other ctenophores. P.W., K.M., T.P.M. and L.L.M. sequenced and assembled the ctenophore transcriptomes. N.V.W. and K.M.K. performed phylogenetic analyses and ancestral state reconstruction. N.V.W., K.M.K., L.L.M. and K.M.H. wrote the manuscript. All authors edited manuscript versions and approved the final version.

Competing interests

The authors declare no competing financial interests

Additional information

Supplementary information is available for this paper at doi:10.1038/s41559-017-0331-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.V.W. or L.L.M. or K.M.H.

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