

RESEARCH ARTICLE

Open Access



# Shedding light: a phylotranscriptomic perspective illuminates the origin of photosymbiosis in marine bivalves

Jingchun Li<sup>1,2,3\*</sup>, Sarah Lemer<sup>3,4,5</sup>, Lisa Kirkendale<sup>6</sup>, Rüdiger Bieler<sup>7</sup>, Colleen Cavanaugh<sup>3</sup> and Gonzalo Giribet<sup>3,5</sup>

## Abstract

**Background:** Photosymbiotic associations between metazoan hosts and photosynthetic dinoflagellates are crucial to the trophic and structural integrity of many marine ecosystems, including coral reefs. Although extensive efforts have been devoted to study the short-term ecological interactions between coral hosts and their symbionts, long-term evolutionary dynamics of photosymbiosis in many marine animals are not well understood. Within Bivalvia, the second largest class of mollusks, obligate photosymbiosis is found in two marine lineages: the giant clams (subfamily Tridacninae) and the heart cockles (subfamily Fraginae), both in the family Cardiidae. Morphologically, giant clams show relatively conservative shell forms whereas photosymbiotic fragines exhibit a diverse suite of anatomical adaptations including flattened shells, leafy mantle extensions, and lens-like microstructural structures. To date, the phylogenetic relationships between these two subfamilies remain poorly resolved, and it is unclear whether photosymbiosis in cardiids originated once or twice.

**Results:** In this study, we establish a backbone phylogeny for Cardiidae utilizing RNASeq-based transcriptomic data from Tridacninae, Fraginae and other cardiids. A variety of phylogenomic approaches were used to infer the relationship between the two groups. Our analyses found conflicting gene signals and potential rapid divergence among the lineages. Overall, results support a sister group relationship between Tridacninae and Fraginae, which diverged during the Cretaceous. Although a sister group relationship is recovered, ancestral state reconstruction using maximum likelihood-based methods reveals two independent origins of photosymbiosis, one at the base of Tridacninae and the other within a symbiotic Fraginae clade.

**Conclusions:** The newly revealed common ancestry between Tridacninae and Fraginae brings a possibility that certain genetic, metabolic, and/or anatomical exaptations existed in their last common ancestor, which promoted both lineages to independently establish photosymbiosis, possibly in response to the modern expansion of reef habitats.

**Keywords:** Photosymbiosis, Tridacninae, Fraginae, Symbiodiniaceae, Reef habitat

\* Correspondence: [jingchun.li@colorado.edu](mailto:jingchun.li@colorado.edu)

<sup>1</sup>Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, USA

<sup>2</sup>Museum of Natural History, University of Colorado Boulder, Boulder, USA

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Photosymbiotic associations between marine organisms and photosynthetic algae allow each partner to thrive in nutrient-deficient environments and to constitute highly productive ecosystems [1]. This mutualistic relationship has repeatedly evolved in diverse eukaryotic lineages ranging from single-celled foraminiferans to metazoans, including corals, acoels, sacoglossans and bivalves [2]. Although the immediate threat of coral reef symbiont loss (i.e., bleaching) has generated extensive efforts in studying the short-term ecological interactions between coral hosts and symbionts, long-term evolutionary dynamics of photosymbiosis in most marine lineages are poorly understood [3, 4]. Therefore, it is crucial to start investigating the origination and extinction patterns of photosymbiosis in a diversity of organisms and identify tractable study systems, so we can better understand and predict how such systems will respond to long-term environmental change.

Among photosymbiotic host organisms, bivalve mollusks pose an evolutionary dilemma. Compared to other invertebrates that house symbionts in translucent tissues, bivalves possess opaque, protective shells that represent an inherent obstacle to symbiont exposure to sunlight. Therefore, extensive morphological or behavioral adaptations are needed to establish successful photosymbiosis. Nonetheless, at least 17 modern or extinct bivalve lineages are known/suspected to form photosymbioses, and each possess (ed) distinct morphological traits adaptive to photosymbiosis [3]. Within modern taxa, at least seven lineages have been reported to harbor photosynthetic algae [2]. However, obligate and confirmed mutualistic associations are only found in two clades within the bivalve family Cardiidae: the well-known giant clams in the subfamily Tridacninae and the less-studied heart cockles in the subfamily Fraginae [5–8]. Tridacninae has a Late Eocene to Recent fossil record with appearance of symbiont-bearing members from the Late Oligocene. Symbiont-bearing Fraginae appear slightly later in the fossil record, from the Late Miocene to Recent [3]. Comparative studies between these two lineages can reveal important insights into the evolution of photosymbiosis in complex organisms.

Giant clams (Tridacninae) were the first documented photosymbiotic bivalves [5]. A dual feeding strategy coupling photosymbiosis with filter-feeding is hypothesized to have resulted in the enormous size of *Tridacna gigas*, one of the heaviest non-colonial invertebrates known [9]. Giant clams acquire a considerable amount of organic carbon through photosymbiosis [10]. All species in the subfamily possess symbiotic algae from the family Symbiodiniaceae (also found in symbiosis with corals), mostly placed in the genera *Symbiodinium*, *Cladocopium*, and *Durusdinium* [8, 11]. The giant clam

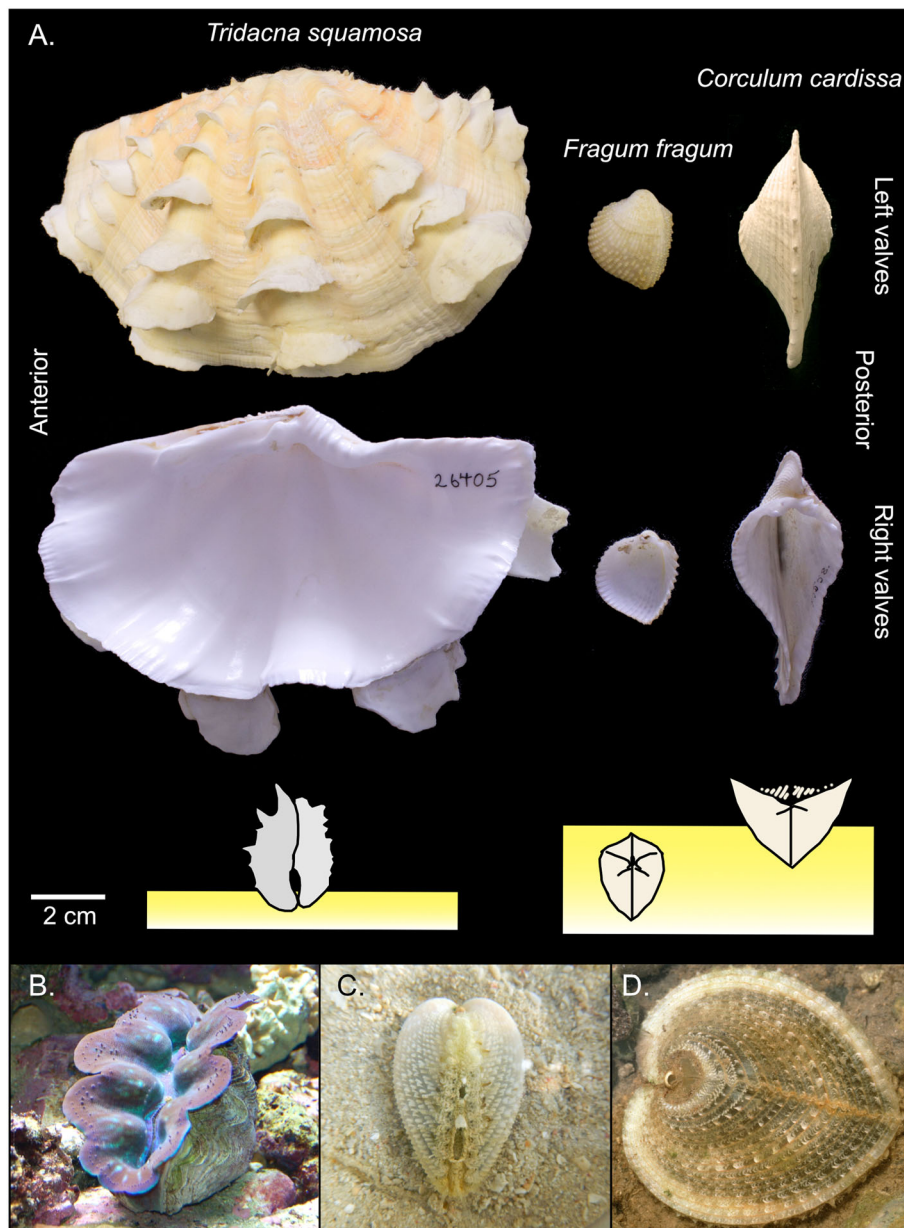
harbors extracellular symbionts within a tubular network derived from the digestive system that extends into the mantle tissue [12, 13]. The tubules only develop in the presence of the symbionts [14], indicating the existence of responsive host-symbiont interactions.

Ecological and morphological adaptations for the symbionts to obtain sufficient sunlight are apparent in Tridacninae. Unlike most species in the family Cardiidae, tridacnines do not bury in sediment. Instead, they are epibenthic (Fig. 1a) with a great expansion of the posterior body, widely gaping shells and a hypertrophied mantle that is exposed to the light (Fig. 1b). The mantle tissues also contain iridophores that scatter photosynthetically productive wavelength to symbionts distributed in the deeper regions [15].

The subfamily Fraginae is a more diverse but less studied group compared to the Tridacninae. It contains more than 50 species, including one symbiotic clade and one non-symbiotic clade, both reciprocally monophyletic [7]. The photosymbiotic species are exclusive to three genera: *Fragum*, *Corculum*, and *Lunulicardia*, comprising at least 23 species [7]. Part or most of their energy source is derived from algal photosynthesis [8]. Symbiont diversity within Fraginae is less explored, but species examined so far harbor algae belonging to Symbiodiniaceae genera *Symbiodinium* and *Cladocopium* [8]. Photosymbiotic fragines host algae within not only the mantle, but also the gill and part of the foot. The symbiont-containing structure is remarkably similar to that documented for giant clams – an extracellular tubular network branching into the symbiont-containing tissues [7, 8, 16].

Despite the similarities between Fraginae and Tridacninae with regards to symbiont identity and symbiont-containing structures, morphological and behavioral adaptations are otherwise quite variable. Photosymbiotic fragines are relatively small sized (~1–10 cm) and live in a diverse spectrum of habitats, ranging from shallow reef flats with clear waters to deeper lagoons [7]. Some species are entirely epifaunal (Fig. 1a, d), the rest are shallowly buried in the sediment, as in non-symbiotic cardiids (Fig. 1a). Morphologically, the shells of some species are very similar to those of non-symbiotic fragines and expose symbionts to light via shell gaping (Fig. 1a, c). In contrast, others exhibit highly flattened translucent shells with sophisticated microstructures to enhance light acquisition, referred to as “shell windows” (Fig. 1a, d [17]);). Those morphological variations suggest that fragines may depend on photosymbiosis to different degrees, and have evolved diverse solutions to maintain the associations.

Given that the two modern photosymbiotic bivalve groups both arose within the family Cardiidae but use radically different strategies, it is natural to ask whether



**Fig. 1** Morphological and ecological comparisons among Tridacninae (*Tridacna squamosa*) and Fraginae (*Fragum fragum*, *Corculum cardissa*) species. **a.** Lateral shell views of the three species and diagrams showing their typical positions in natural habitats. Yellow rectangles represent sediment. Note that *F. fragum* and *C. cardissa* both show different degrees of posterior shell compression. **b-d.** Photos of the three species in their natural habitat. *F. fragum* (**c**) was taken out of the sediment when photo was taken. Photo credits: Jingchun Li and Jeff Whitlock (the Online Zoo). Images were processed in Affinity Designer 1.8.4 (Serif Ltd.)

they inherited photosymbiosis from a common ancestor or independently evolved the associations. If photosymbiosis were an ancestral character, it would imply a single origin with subsequent losses of symbiosis in non-symbiotic fragines and possibly other cardiids. If the two subfamilies acquired photosymbiosis separately, it would suggest that their symbiont-harboring system and other metabolic adaptations are a result of convergent evolution.

In order to address the acquisition of photosymbiosis in bivalves, a phylogenetic framework is required. Currently, the phylogenetic relationships within Tridacninae and Fraginae are relatively well resolved, and both groups are monophyletic [7, 18, 19]. However, the relationship between the two subfamilies remains poorly understood. One of the first cardiid phylogenies, based on a cladistic analysis of shell microstructure and gut morphology, placed Tridacninae and Fraginae as sister

groups [20], but later work incorporating more characters resolved the two subfamilies as distantly related lineages within Cardiidae [21]. The first molecular phylogeny constructed to address this issue was based on a single genetic marker (nuclear 18S rRNA), grouping Tridacninae with non-symbiotic cardiids in the subfamilies Trachycardiinae and Laevicardiinae, leaving Fraginae as sister group to the rest [22]. The most recent multi-gene phylogenetic analyses also failed to reliably resolve the positions of Tridacninae and Fraginae [19]. The placement of Tridacninae was not consistent and was recovered as sister group to various subfamilies (e.g., Lymnocardiinae and Cardiinae) in different analyses. The same situation characterizes the placement of Fraginae as its phylogenetic position has shifted with different phylogenetic reconstruction methods. A previous thesis [23] utilizing the same multi-gene dataset, however, did recover a sister relationship between Tridacninae and Fraginae in one Bayesian analysis. Nonetheless, these conflicting results highlight the difficulties in resolving relationships based on three Sanger-based markers, thus hindering our understanding of the evolution of photosymbiosis in bivalves.

The recent development of phylogenomic approaches to bivalve phylogenies [24–26] has demonstrated the possibility of resolving difficult nodes according to previous Sanger-based approaches (e.g., [27]). Therefore, we adopted an RNAseq approach and generated transcriptomic data for Tridacninae, Fraginae and other cardiids with the aim to better understand the evolutionary of photosymbiosis in Cardiidae.

## Results

### Transcriptome assembly and data quality

The number of used reads, contigs, and other values to assess the quality of the assembled transcriptomes can be found in Table 1. Orthology assessment of this 33-taxon dataset with the OMA stand-alone algorithm recovered 244,752 orthogroups. The two super-matrices (Fig. 2 and Additional file 3: Fig. S1) respectively yielded 1108 (*Matrix 1*: occupancy of > 50%; 280,086 aa) and 313 (*Matrix 2*: occupancy of > 75%; 78,272 aa) orthologs. Levels of compositional heterogeneity were low in most orthogroups (Additional file 4: Fig. S2). The relative composition frequency variability (RCFV) per taxon and per amino acid ranged from 0.0002 to 0.001, representing overall compositional homogeneity throughout all amino acids and taxa included in *Matrix 1*.

### Phylogenetic analyses and gene evolution

All 26 analyses (Table 2) recovered strong support for the monophyly of Tridacninae, Trachycardiinae, Laevicardiinae, Cardiinae, Lymnocardiinae, photosymbiotic Fraginae, and non-symbiotic Fraginae. Monophyly of

Fraginae (including both symbiotic and non-symbiotic clades) was supported in 63% of the analyses. The clade composed of Trachycardiinae, Laevicardiinae, and Cardiinae was supported in all analyses. The placement of Lymnocardiinae was less certain, although 60% of the analyses recovered it as the sister group to all other Cardiidae.

The most supported overall topology by different analyses (Fig. 3a) is consistent with the result obtained from the maximum likelihood (ML) analysis on *Matrix 1* (Fig. 3c), which revealed a sister group relationship between Tridacninae and Fraginae, although with moderate bootstrap support (65/100). After Dayhoff recoding of *Matrix 1*, the same topology was obtained, as for the PhyloBayes analysis of *Matrix 2*, although the two independent chains did not fully converge. The conflict was mostly driven by the position of Lymnocardiinae as the sister group to all other cardiids. The posterior probability for the Tridacninae and Fraginae sister group relationship is 0.99 and does not show conflict between the two chains. Analyses of submatrices C, E, and G-I also recovered this common topology. The second most common topology is largely consistent with the first one, except that Lymnocardiinae is placed as the sister group of Tridacninae (Fig. 3b). Other matrices supporting this topology include the slowest evolving submatrix A, and three relatively fast evolving ones (J, M, O).

Besides the most supported topologies, an additional 11 topologies were supported by only one or two matrices each. In particular, ML analysis on *Matrix 2* placed Fraginae as the sister group to a clade composed of ((Trachycardiinae, Laevicardiinae), Cardiinae), Lymnocardiinae), and recovered Tridacninae as the sister group to all other Cardiidae. However, the bootstrap support values for the backbone topology are low (< 60).

The supernet analyses on the two main matrices showed relatively long branches leading to Tridacninae, photosymbiotic Fraginae, and non-photosymbiotic Fraginae, respectively (Additional file 5: Fig. S3), a result consistent with the strong support for these nodes obtained from the phylogenetic analyses. However, substantial reticulation was observed at the node leading to these three clades, indicating strong gene conflict in resolving their relationships. This is likely the reason for the low bootstrap supports obtained in the ML analyses.

BLAST results for the fastest 10% and slowest 10% genes are shown in Additional files 1 and 2: Table S1 and S2. The slow-evolving genes are composed of a higher proportion of nuclear ribosomal genes compared to the fast evolving ones (13% vs. 0%). The fast-evolving genes have a higher number of mitochondrial genes compared to the conserved genes (23% vs. 1%). This pattern is in line with what is currently known about the molecular evolution rates of these gene groups.

**Table 1** Information of all specimens used in this study. Photosymbiotic taxa are bolded. Abbreviations are as the following. UCM Museum of Natural History, University of Colorado Boulder, FMNH the Field Museum, WA Western Australian Museum, UMMZ Museum of Zoology, University of Michigan, MCZ Museum of Comparative Zoology, Harvard University, SRA NCBI Sequence Read Archive

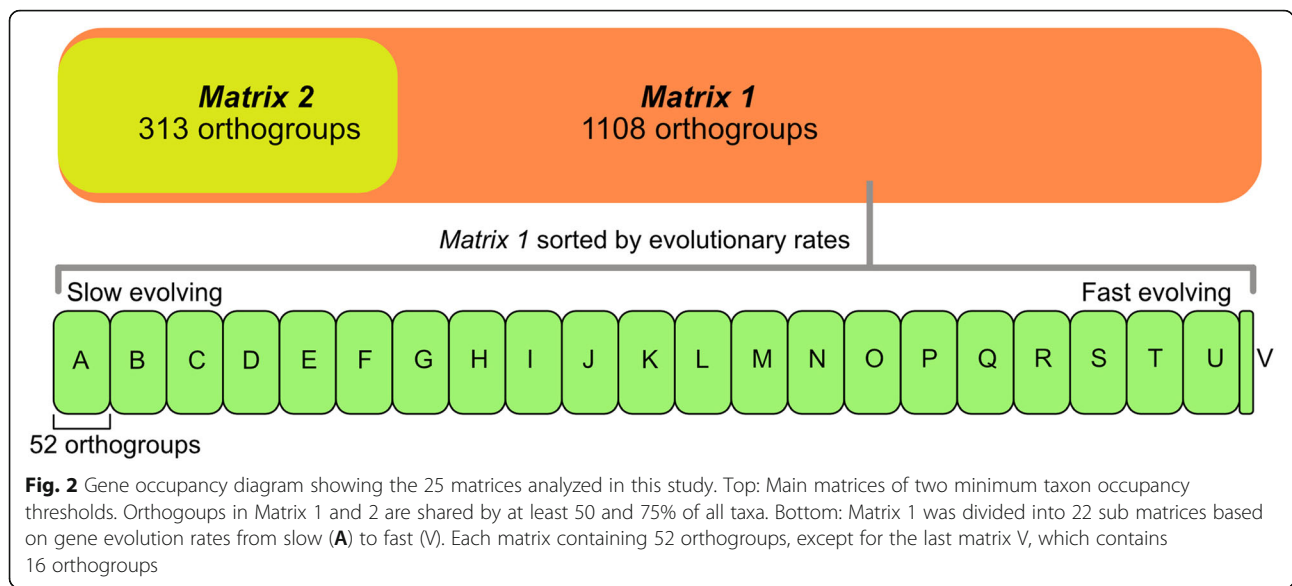
Species	Locality	Voucher	Tissue	SRA	Contigs	N50	Source
<i>Fulvia lineonotata</i>	Guam	UCM48073	Mantle	SRR11252446	243,326	688	De novo
<i>Americardia media</i>	USA	FMNH 315293	Muscle	SRR11252445	182,732	527	De novo
<b><i>Corculum cardissa</i></b>	Australia	WA263	Mantle	SRR11252439	371,151	483	De novo
<b><i>Fragum fragum</i></b>	Guam	UCM48094	Mantle	SRR11252438	608,501	460	De novo
<b><i>Fragum mundum</i></b>	Australia	WA265	Whole	SRR11252437	333,455	492	De novo
<b><i>Fragum scruposum</i></b>	Guam	UCM48090	Mantle	SRR11252436	376,388	508	De novo
<b><i>Fragum unedo</i></b>	Australia	BivAToL-75 (FMNH)	Muscle	SRR8217860	121,185	393	[26]
<b><i>Lunulicardia sp.</i></b>	Australia	FMNH 317975	Muscle	SRR8217812	244,397	404	[26]
<i>Microfragum festivum</i>	Australia	WA261	Whole	SRR11252435	174,309	548	De novo
<i>Laevicardium serratum</i>	USA	BivAtoL-57 (FMNH)	Muscle	SRR8217867	162,615	536	[26]
<i>Laevicardium sp.</i> Curaçao	Curaçao	BivAtoL-456 (FMNH)	Muscle	SRR11252434	196,243	472	De novo
<i>Acanthocardia tuberculata</i>	Italy	UMMZ 39326	Mantle	SRR11252433	241,101	554	De novo
<i>Cerastoderma edule</i>	England	BivAToL-21 (FMNH)	Muscle	SRR8217858	120,092	594	[26]
<i>Vasticardium pectiniforme</i>	Japan	UCM48108	Muscle	SRR8217813	144,372	485	[26]
<i>Vasticardium compunctum</i>	Japan	UCM48123	Muscle	SRR11252432	134,451	478	De novo
<i>Dalloccardia muricata</i>	USA	BivAtoL-454 (FMNH)	Muscle	NA	177,010	368	De novo
<i>Papyridea lata</i>	Curaçao	MCZ 383047	Muscle	SRR8217866	267,529	345	[26]
<i>Trachycardium egmontianum</i>	USA	FMNH 344567	Muscle	NA	186,788	350	De novo
<b><i>Hippopus sp.</i></b>	Japan	UCM48120	Mantle	SRR11252444	187,464	406	De novo
<b><i>Tridacna crocea</i></b>	Cultured	UMMZ 304399	Mantle	SRR11252443	251,344	1092	De novo
<b><i>Tridacna derasa</i></b>	Cultured	UMMZ 304400	Mantle	SRR11252442	245,928	1056	De novo
<b><i>Tridacna maxima 2</i></b>	Cultured	UMMZ 304398	Mantle	SRR12252441	298,624	956	De novo
<b><i>Tridacna maxima 1</i></b>	Cultured	NA	Muscle	SRR8217859	176,478	527	[26]
<b><i>Tridacna squamosa</i></b>	Cultured	UMMZ 304401	Mantle	SRR12252440	245,781	1111	De novo
Outgroup							
<i>Cardites antiquatus</i>	Spain	MCZ 379178	Muscle	SRR1560458	113,906	567	De novo
<i>Chama macerophylla</i>	Panama	MCZ 381299	Muscle	SRR8217830	165,533	686	De novo
<i>Corbicula fluminea</i>	USA	BivAToL-242 (FMNH)	Muscle	SRR1559272	176,007	763	[24]
<i>Galeomma turtoni</i>	Spain	MCZ 378975	Muscle	SRR1560274	92,358	548	[24]
<i>Hiatella arctica</i>	UK	BivAToL-195 (FMNH)	Muscle	SRR1560281	73,557	576	[24]
<i>Lyonsia floridana</i>	USA	BivAToL-248 (FMNH)	Muscle	SRR1560310	92,076	838	[24]
<i>Mya arenaria</i>	NA	MCZ 381391	Muscle	SRR1560361	98,870	873	[24]
<i>Neotrigonia margaritacea</i>	Australia	MCZ 379092	Muscle	SRR1560432	162,657	549	[24]
<i>Calyptogena gnifica</i>	USA	NA	Mantle	SRR8217831	42,445	399	[26]

The coral V-type H + -ATPase (VHA, photosymbiosis-related gene) BLAST searches recovered VHA sequences from eight Tridacninae and Fraginae species. Tridacninae and Fraginae VHA amino acid sequences are mostly identical, except for *Fragum fragum*, which has three (out of 137) amino acid substitutions. Besides *F. fragum*, there are no sequence differences between the photosymbiotic and non-symbiotic taxa (Additional file 6: Fig. S4). The coral VHA differs from bivalve sequences by 6–11 amino acid substitutions. Both coral and bivalve sequences share ~85% similarity with the

Symbiodiniaceae VHA. All animal VHA genes formed a monophyletic group with 100% bootstrap support.

#### Fossil calibration and ancestral state reconstruction

The cardiid dating estimates based on the most common topology are shown in Fig. 4. Tridacninae and Fraginae diverged in the Late Cretaceous. Soon after, the symbiotic and non-symbiotic Fraginae lineages also separated. Those long branches eventually lead to relatively recent and rapid diversification of the photosymbiotic crown groups. The diversification of photosymbiotic fragines is



dated at around 15 Ma (SD = 3.1). The diversification of crown Tridacninae is ~ 27 Ma (SD = 4.4), and the radiation of the genus *Tridacna* dates back to ~ 6 Mya (SD = 6.3). Age estimates of the other non-focal subfamilies, including Trachycardiinae, Laevicardiinae, Cardiinae, and Lymnocardiinae, are largely within the range of Herrera et al. 2015 [19].

Ancestral state reconstructions using the Binary State Speciation and Extinction (BiSSE) and mk2 models yield congruent results (Fig. 4). Both identified two independent origins of photosymbioses in Cardiidae, one within Tridacninae and the other within the photosymbiotic Fraginae. Comparison with reef abundance through time shows that the onset of Tridacninae photosymbiosis corresponds to the beginning of a rapid global reef expansion, while radiations of the genus *Tridacna* and photosymbiotic fraginae both occurred within peak reef abundance.

**Discussion**

**Cardiid phylogeny**

This is the first phylogenetic study using transcriptomic data that aims to resolve the relationships between the two cardiid subfamilies which contain photosymbiotic

taxa. The best corroborated hypothesis supports a sister relationship between Tridacninae and Fraginae, both recovered as monophyletic (Fig. 3). Other than the position of Tridacninae, the placement of all other cardiid subfamilies is consistent with the most recent multi-gene Cardiidae phylogeny comprising 110 species [19].

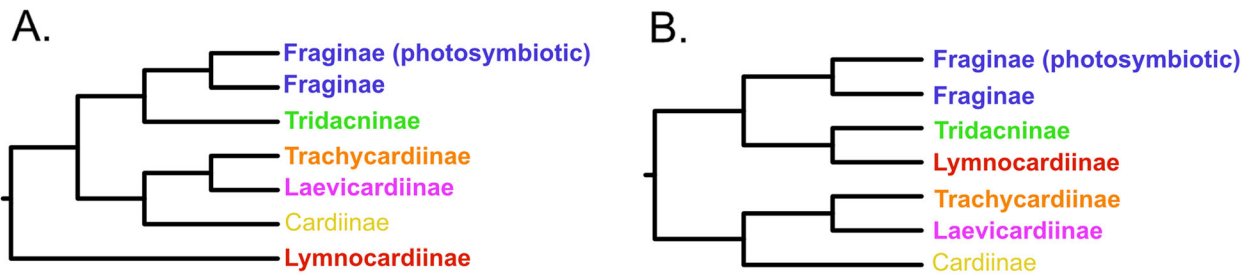
Our analyses highlighted the pervasive challenge to reconstruct a well-resolved and highly supported cardiid phylogeny. The fundamental reconstruction difficulty stems from the diversification process of Tridacninae, photosymbiotic, and non-symbiotic Fraginae. Our results show that the branches leading to the three crown groups are long and subtended by very short internodes, indicating the divergence among these groups was rapid. These ancient but fast diversification events are inherently difficult to resolve [29]. Some of the major problems include: not enough data to resolve the nodes; molecular data not variable enough at the appropriate level; strong conflicts among genes; and inadequate substitution models [29]. We discuss these concerns in the following paragraphs.

Firstly, current studies on cardiid phylogeny may suffer from some level of data limitation. For example, although Herrera et al. 2015 [19] had excellent taxon sampling through Cardiidae, the gene coverage was low. In the present study, taxon sampling was limited due to the need to obtain high quality transcriptomes, which requires freshly collected specimens. These constraints could be overcome by incorporating DNA-based phylogenomic approaches (e.g., RADseq, exome capture, etc.) using well-preserved museum specimens.

In addition, some of the data may lack the appropriate level of resolution. For example, the ML Analysis on *Matrix 2* resulted in poorly resolved topologies with low

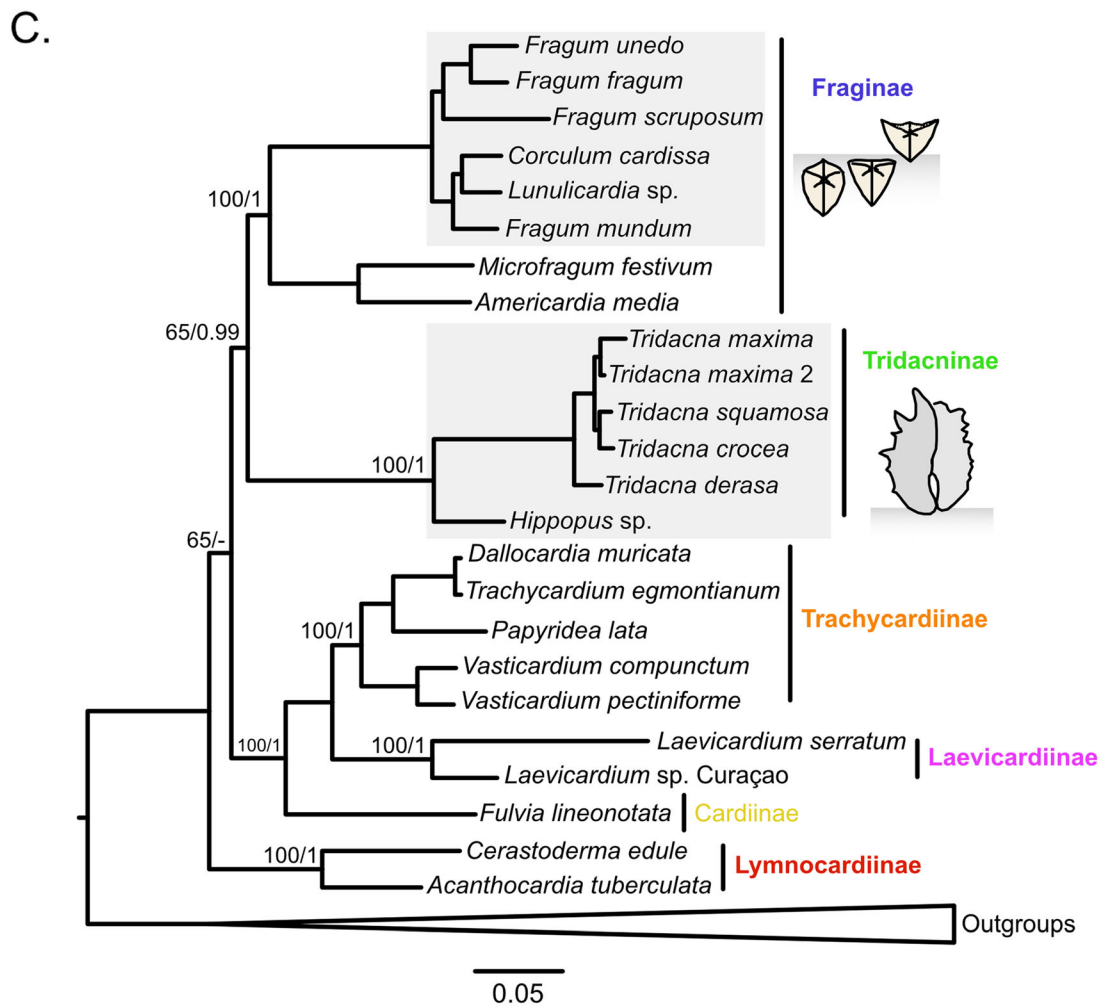
**Table 2** A summary of phylogenetic analyses conducted in this study

Gene Matrix	Analysis	Model
1,2, A-V	PhyML	PCMA
1	RAXML-Dayhoff	Multigamma GTR
2	PhyloBayes	CAT-GTR
Individual genes	RAXML	PROTGAMMALG4X

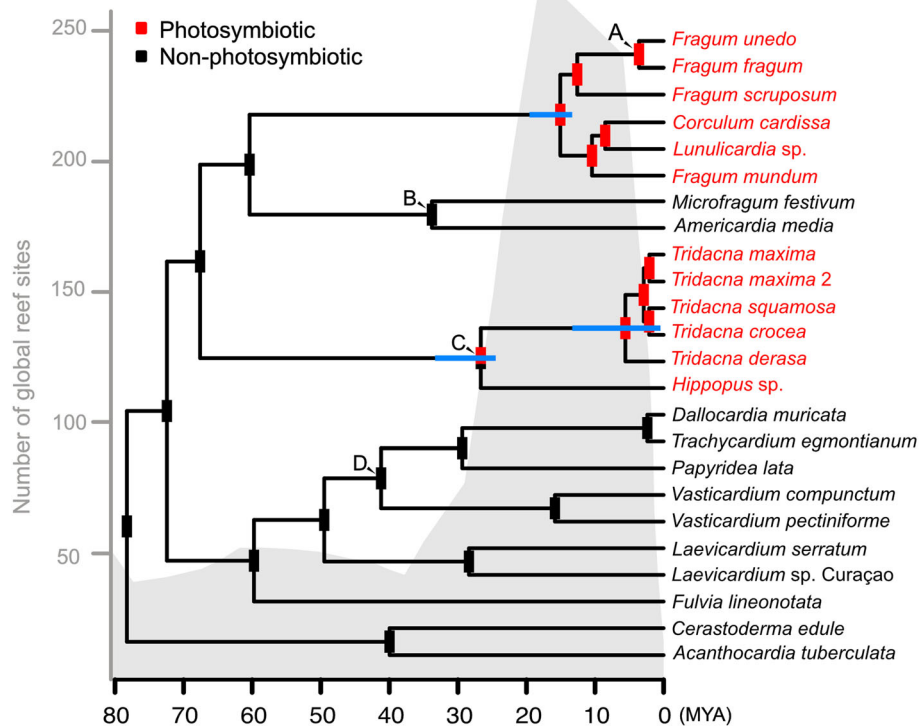


Supporting analyses:  
 Matrix 1 (1108 orthogroups): PhyML, RAxML-Dayhoff  
 Matrix 2 (313 orthogroups): PhyloBayes (unrooted)  
 Evolutionary rate matrices C, E, G, H, I: PhyML

Supporting analyses:  
 Evolutionary rate matrices A, J, M, O: PhyML



**Fig. 3 a-b.** The two best supported topologies obtained from the analyses of the two main matrices and 22 submatrices. Cardiid subfamilies are indicated by different colors. Supporting matrices and corresponding analytical methods are listed under each topology. **c.** Phylogenetic results based on maximum likelihood analysis (PhyML-PCMA) of Matrix 1 and Bayesian analysis (PhyloBayes) of Matrix 2. This is also the topology supported by the most analyses. Node labels represent bootstrap supports / posterior probabilities of each subfamily and the backbone. Photosymbiotic clades are shaded in grey. Shell position of Fraginae and Tridacninae species in their natural habitats is shown in the two diagrams



**Fig. 4** Fossil calibrated phylogeny and ancestral state reconstruction based on the most supported topology. Letters (a–d) at nodes indicate calibration points. Blue bars at nodes represent standard deviation of age estimation. Red and black bars at nodes represent probabilities of the common ancestor being photosymbiotic (red) or non-photosymbiotic (black). The grey shading in the background corresponds to number of global reef sites through time (following [28])

bootstrap support values. This observation where large, albeit more incomplete matrices provided better resolution than small and more complete matrices is not unique to our dataset (e.g., [24, 30]). It is likely influenced by gene choice in the more complete matrix [30]. If a gene is present in more taxa, it is likely to be relatively conserved; hence lacking enough genetic variation to resolve rapid divergence, as seen in some of our individual gene trees. These very conserved/slow-evolving genes might be over-represented in a smaller matrix, but contribute minimally to any phylogenetic resolution.

Gene conflicts are also apparent in our dataset based on the supernetwork and submatrix analyses, and in part may be explained by incomplete lineage sorting at these rapid-divergence nodes. In addition, genes with different rates of evolution gave rise to divergent topologies. Some faster evolving genes (e.g., Matrices Q–V) produced clearly problematic topologies, such as placing photosymbiotic Fraginae as the sister clade to all other cardids. These rapidly evolving genes might be subject to more genetic saturation and could be under the influence of strong selection, all of which hinder their ability to resolve deep phylogenetic nodes and conflict with other more informative genes.

Lastly, sequence compositional heterogeneity is known to be especially problematic for inferring short internal nodes [31]. However, based on the low level of composition heterogeneity in our dataset and the Dayoff recoding analysis, it is unlikely to produce significant bias in the results.

In sum, we have found that overly conservative genes and fast evolving genes do not provide informative resolution to the nodes of interest, a phenomenon observed in other taxonomic groups [32]. These genes generate conflicting topologies and could be responsible for the observed low bootstrap values. On the contrary, ML analyses on the large matrix, moderately evolved genes, as well as the Bayesian analysis, all support a sister relationship between Tridacninae and Fraginae.

#### Implications on photosymbiosis evolution

Our analyses support two independent origins of photosymbiosis in Cardiidae. This result should be robust to phylogenetic uncertainties, because a non-sister relationship between Tridacninae and Fraginae will only reinforce the independent evolution scenario. Our ancestral state reconstruction was model-based; an alternative parsimony-based analyses would deem “two



independent origins” or “one origin and one loss (in the non-symbiotic *Fraginae* clade)” equally likely. However, the latter scenario is not well supported because fossil record indicates that visible shell adaptations in both *Tridacninae* and *Fraginae* appeared after the Late Oligocene [3, 33]. If the common ancestor of the two subfamilies is photosymbiotic, it would imply that photosymbiosis was established in late Cretaceous but persisted more than 30 Ma without any apparent shell adaptations. It is much more probable that there are two separate origins of photosymbiosis with shell adaptations evolved shortly after.

The repeated evolution of bivalve photosymbiosis suggests its adaptive advantage. The relatively rapid crown group diversification, coupled with morphological responses [3], is consistent with criteria for adaptive radiation - the generation of new species exhibiting pronounced morphological divergence over relatively short timeframes, typically in response to new environmental conditions [34]. In photosymbiotic cardiid, the species number is modest compared to other documented examples. However, this might be due to the lack of systematics research in these groups, as more studies have started to reveal their hidden diversities (e.g., [35]).

As for most cardiid lineages, *Tridacninae* and *Fraginae* have inferred ancestral distributions that span the Indo-Pacific; *Tridacninae* has a wider ancestral range, reaching the western temperate northern Pacific [19]. Given that the origin of photosymbiosis in both subfamilies overlaps with the expansion of modern reefs, it is likely that the formation of shallow marine habitat in the Indo-Pacific is a key environmental driver for bivalve photosymbiotic adaptation. This is further supported by the fact that photosymbiotic fragines have a sister non-symbiotic lineage that still inhabits deep sandy bottoms [7], possibly representing the ancestral ecology of these bivalves before they shifted to shallower habitat. Kiesling 2009 [28] proposed a fundamental question regarding the formation of reef biodiversity – Have reef taxa originated within reefs or have they evolved somewhere else then moved into reef habitats? Our results indicate that at least for photosymbiotic bivalves, they have likely originated and diversified within the reef habitats. More biogeographical, palaeontological and phylogenetic data are certainly needed to further corroborate this point.

Both *Tridacninae* and photosymbiotic *Fraginae* are thought to exhibit phenotypic adaptations to benefit the symbionts, making some species’ shell and mantle morphologies drastically different from typical cardiid (Fig. 1). What is striking, and little mentioned until now, is the divergent morphological trends in photosymbiotic fragines contrasted with the uniform morphological trends in *Tridacninae*, essentially two different responses

to expose symbionts to optimal irradiance in sister lineages. Previous studies of adaptive radiations have well highlighted examples of divergent morphological responses to newly available niches (e.g., [36]), as seen in fragines. But uniform morphological responses, as seen in tridacnines, are less documented. Both strategies have advantages. While divergent morphologies provide highly specialized adaptations to different fine-scale niches, an uniform/static morphology may enable the lineages to become broadly adapted generalists to mediate environmental fluctuations [37]. The different morphological evolution patterns in photosymbiotic cardiid suggests that in the acquisition of a key innovation, historical morphological contingencies (e.g., opaque heavy shell, unexposed tissues, infaunal habit) and common ancestry do not predict the directionality of morphological evolution, not even for sister lineages that use the same ecological strategy to adapt to similar habitats.

Despite the versatile shell and mantle morphologies in *Fraginae* and *Tridacninae*, their symbiont-containing tubular system share striking similarities. Both stem from the digestive system of the hosts and form tertiary tubular networks [13, 16]. The only other similar molluscan structures are found in some marine gastropods who temporarily maintain algae or chloroplasts in their tissues [38]. The development of giant clam tubules are only triggered by the presence of symbionts [14], indicating the acquisition of photosymbiosis is a highly interactive process between hosts and symbionts.

It has long been hypothesized that different photosymbiotic bivalves express homologous genes to build the tubular system, in response to similar symbiont signals [16]. The newly-found sister relationship between *Tridacninae* and *Fraginae* lend further support to this theory, suggesting that genes homologous to the tubular-formation ones are ancestral to the two lineages. It is even possible that the genes are ancestral to bivalve and gastropods, as the latter can form similar anatomical structures. Given that the different mollusk lineages evolved photosymbiosis independently, it is likely that the gene/anatomical level similarities are generated from parallel evolution. That is, each lineage independently coopted similar genetic mechanisms for generating symbiont-containing structures. Molecular level parallel evolution has been shown in photosymbiotic systems. For example, both corals and giant clams repurpose the expression of the vacuolar H<sup>+</sup>-ATPase gene (VHA) to facilitate their carbon concentrating process and promote algal photosynthesis, even though the two lineages are very distantly related [39, 40]. Our analyses further support that VHA is a conserved gene, ancestral to bivalves, corals, and other animals. Photosymbiotic bivalves do not possess any “special” version of VHA; the amino acid sequence is the same as the ones found in

non-symbiotic taxa. It is more likely that the adaptation to photosymbiosis is realized at the regulatory/expression level, where VHA is highly expressed in tissues that harbor symbionts. In contrast with the relatively labile shell and mantle adaptations, it is possible that the evolution of host metabolomic and developmental adaptations are more constrained genetically, resulting in similar mechanisms in diverse animal groups. More in-depth studies on the molecular mechanisms behind photosymbiosis are needed to gain better insights.

To further investigate macroevolution of bivalve photosymbiosis, a better understanding of potential photosymbiotic fossil taxa is essential. However, they are challenging to identify in the fossil record based on shell morphology alone [2]. Although photosymbiotic bivalves possess a suite of morphological traits to enhance light exposure, almost all characters used to support photosymbiotic condition are found in modern non-photosymbiotic bivalves. Examples include permanent shell gaping (Family Limidae, Galeommatidae), enlarged mantle (Galeommatidae), or transparent shells (Placunidae, Pinnidae). The only photosymbiotic-exclusive shell character is the shell “window” microstructure [17], but many photosymbiotic bivalve species do not possess it, and similar features have not been found in fossil taxa. Similarly, other identifiable photosymbiotic ecological traits (shallow water distribution, fast growth, reclining habits) are not unique to photosymbiotic bivalves either [2]. Therefore, alternative methods need to be developed (e.g. better isotopic metrics, metabolite signatures) if we wish to fully understand the evolutionary dynamics of bivalve photosymbiosis.

Lastly, further characterization of host-symbiont diversity and interactions are essential for developing a full picture of animal-algal photosymbiosis. A comprehensive documentation of host-symbiont biodiversity can provide insights into photosymbiotic mechanisms in diverse habitats and with different organismal complexities. For example, Li et al. 2018 [8] demonstrated that frangible species at varying water depths have different dependencies on algal photosynthesis, with deeper host species utilizing less photosynthetically derived carbon and exhibiting fewer shell modifications. Therefore, it is highly likely that host-symbiont interactions and reliance vary greatly among host-symbiont pairs, depending on symbiont types, as well as the degree of host adaptations. Li et al. 2018 [8] summarized known symbiont diversity in giant clams and showed that several frangible species harbor symbionts from the genus *Cladocopium*, which also occupies cnidarians, giant clams and foraminifera. However, this by no means captured the full spectrum of potential symbiont diversity in photosymbiotic bivalves, especially when new host species are discovered regularly (e.g., [35]). In most photosymbiotic bivalve

systems, the roles symbionts play in host ontogeny, reproduction, response to environmental fluctuation, etc. are unexplored, all of which require our continued effort to identify and characterize existing photosymbiotic diversity.

## Conclusions

This study revealed that two closely related bivalve subfamilies independently evolved photosymbiosis, further demonstrating the prevalence of photosymbiosis in marine ecosystems. Selection pressure for photosymbioses is high in oligotrophic environments, because the associations provide immediate metabolic benefits to the partners [41]. In some parts of the ocean, the diversity and abundance of photosymbiotic plankton hosts are substantially higher than that of phototrophic protists [42]. However, despite the ecological and evolutionary importance of photosymbioses, our knowledge of their origin, diversity, environmental impact, and genetic mechanisms remains rudimentary [41]. The evolution of photosymbioses has been hypothesized to drive rapid diversification, have significant impacts on the community composition, and influence productivity and biogeochemical cycling of the ecosystem [43]. Most of these hypotheses have not been formally tested or have only been tested on a small number of lineages. Therefore, it is now time to move beyond a few model groups and start to comparatively address evolutionary patterns, genomic adaptations, and geological impacts of diverse photosymbiotic systems.

## Methods

### RNA extraction, Transcriptome sequencing, and assembly

We obtained transcriptomes of 24 Cardiid specimens and 9 bivalve outgroups. Information about the sampled specimens can be found in Table 1 and in the MCZ online collections database (<http://mczbase.mcz.harvard.edu>).

Bivalve tissues were collected fresh and flash frozen in liquid nitrogen or preserved in RNAlater® (Life Technologies) and stored at  $-80^{\circ}\text{C}$ . Total RNA from bivalve mantle and foot tissues was extracted and purified as described in [26]. mRNA quality was assessed with a picoRNA assay in an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantity was measured in a Qubit fluorometer (Life Technologies). 28S rRNA in mollusk samples breaks down into two segments of comparable sizes to 18S rRNA during Bioanalyzer quality assessment, resulting in non-meaningful RIN scores [44–46]. Therefore, we used the RNA electropherogram itself to validate the RNA integrity. When a significant single peak (composed of 18S and the broken 28S) was observed, it indicated good RNA integrity.

The cDNA libraries were constructed with an Apollo 324 NGS Library Prep System (TakaraBio) as described in [26]. cDNA library concentrations, quality and sizes were assessed as in [26] before being sequenced on an Illumina HiSeq 2500 platform with paired-end reads of 150 bp at the FAS Center for Systems Biology at Harvard University. Tridacninae samples were sequenced at the Advanced Genomics Core, University of Michigan.

Demultiplexed sequencing results were retrieved in FASTQ format. Each sample was processed in the following steps. Trimalore 0.3.3 [47] was used to quality filter the data and trim adapters (all reads with an average Phred score lower than 30 and shorter than 25 bp, were discarded). Ribosomal RNA (rRNA) was filtered out using Bowtie 2.0.0 [48] by building a bowtie index using all known mollusk rRNA sequences downloaded from GenBank. Some of the bivalve tissues (mantle) used for RNA extraction contain symbionts, whose expressed genes could interfere with the transcriptome assembly. Therefore, Symbiodiniaceae reads were filtered out of photosymbiotic bivalve sequences (Tridacninae and Fraginae) by aligning reads to the transcriptomes of Symbiodiniaceae strains from multiple genera [49–51]. All reads that did not align with the rRNA or Symbiodiniaceae indices were stored in FASTA format as single files and used in our downstream analyses.

Detailed methods for de novo transcriptome assembly, filtration and translation can be found in [26]. In brief, assemblies were conducted with Trinity r2014-04-13 [52, 53], filtration of redundant reads with CD-HIT version 4.6 [54], translation with TransDecoder [53] and isoform filtration with a custom Python script. All reads generated for this study are deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI-SRA; Table 1) and all assembled transcriptomes can be found in the online Harvard database: [https://dataverse.harvard.edu/dataverse/cardiid\\_transcriptome](https://dataverse.harvard.edu/dataverse/cardiid_transcriptome).

#### Orthology assignment, gene matrix construction

Orthologous genes across all 33 taxa were identified for downstream phylogenetic analyses using OMA standalone v0.99z.3 [55, 56]. Contrary to standard bidirectional best-hit approaches, the OMA algorithm uses evolutionary distances instead of scores, considers distance inference uncertainty and differential gene losses, and includes many-to-many orthologous relations; making it more powerful in identifying true orthologs [57]. The ortholog matrix is constructed from all-against-all Smith–Waterman protein alignments. The algorithm then identifies stable ortholog pairs, verifies them, and checks against potential paralogous genes before clustering cliques of stable pairs as groups of orthologs.

For each of the 244,752 unique orthogroups predicted by OMA, amino acid sequences from all available taxa were aligned using MUSCLE version 3.6 [58]. Positions showing poor alignment scores were identified and discarded using a probabilistic character masking approach with ZORRO [59] (selected parameters are detailed in [26]).

Two initial gene matrices were generated by selecting orthogroups based on minimum taxon occupancy thresholds ([60]; see Fig. 2 and Additional file 3: Fig. S1). *Matrix 1*, with a minimum gene occupancy of 50% (i.e., each gene is present in at least 50% of the taxa), is composed of OMA orthogroups shared by 17 or more taxa; and *Matrix 2* includes orthogroups shared by 25 or more taxa (gene occupancy > 75%). Orthogroups for the two matrices were then concatenated using Phyutility 2.6 [61]. All data were treated as amino acids and the two matrices do not contain any contaminant sequences from the Symbiodiniaceae symbionts. In addition to filtering out any symbiont genes during the transcriptome assembly process, no Tridacninae or Fraginae exclusive genes (i.e., possibly from algal symbionts) were used for downstream analyses. All genes used in downstream analyses were strictly shared among symbiotic and non-symbiotic cardiid taxa.

#### Phylogenetic analyses, compositional heterogeneity assessment, and gene function assessment

A summary of all analyses is shown in Table 2. The two main matrices, *Matrices 1* and 2, were analyzed using maximum likelihood (ML) inferences conducted by PhyML-PCMA with 10 principal components [62] using Subtree Pruning and Regrafting (SPR) and three initial random starting trees with a random seed. One hundred bootstrap replicates were conducted for *Matrices 2* using PhyML-PCMA. Bootstrap replicates for *Matrix 1* were computed with RAxML 7.7.5 with the fast bootstrap algorithm [63], because of its large size and the intense computation required by PhyML-PCMA.

*Matrix 2* (75% minimum gene occupancy, 313 orthogroups) was also analyzed using a Bayesian inference with PhyloBayes MPI v.1.4e [64]. The heterogeneous CAT-GTR model of evolution [65] was used, accounting for potential site-specific amino acid preferences. Other settings were kept as default. Three independent Markov chain Monte Carlo (MCMC) runs were conducted for 7496–11,852 cycles and their convergence was evaluated with the bpcomp and tracecomp commands.

BaCoCa v.1.1r was used to estimate relative compositional frequency variability (RCFV) in *Matrix 1*. RCFV is a measure of the absolute deviation from the mean for each amino acid and for each taxon summed up over all amino acids and all taxa [66]. High values of

compositional heterogeneity may negatively impact phylogenetic inferences [67]. RCFV values were plotted in a heatmap using the R package gplots. To assess the impact of compositional heterogeneity, Dayhoff recoding was employed on *Matrix 1*, as described in [26]. This recoding should minimize effects of compositional heterogeneity as well as of long-branch attraction [68]. The recoded matrix was analyzed with RAxML 7.7.5 using multigamma GTR (-m MULTIGAMMA -K GTR).

*Matrices 1* and *2* showed some inconsistent tree topologies across analyses, we thus tested them for putative gene incongruence by generating individual gene trees for each orthogroup included in each of the three main matrices with RAxML 7.7.5 and the PROTGAMMALG4X substitution model [69]. The choice of the PROTGAMMALG4X model was made based on past experience with bivalve transcriptomic data analyses [24–26], where this model of amino acid substitution revealed to consistently be the best fitted for this kind of data. All individual best-scoring trees were concatenated per matrix, intergene conflict was analyzed with SuperQ v1.1 [70], as described in [26] and the super-networks were visualized with SplitsTree v.4.13.1 [71].

Because the individual gene tree super-networks revealed gene conflict at specific nodes (see results for details), we further explored the effects of molecular evolution rate heterotachy on tree topology. All 1108 orthogroups from *Matrix 1* were sorted based on their evolutionary rate, as detailed in [26]. The sorted orthogroups were then divided into 22 matrices (Matrices A to V, Fig. 2), each containing 52 sorted loci, except for the last one which contains 16 loci. The first matrix contains the 52 slowest evolving genes (most conserved) and the last contains the 16 fastest evolving genes (least conserved). Each matrix was then analyzed using PhyML-PCMA as described above. In addition, we investigated gene functions of the 10% slowest (matrices A and B) and 10% fastest (matrices T-V) evolving genes. The amino acid sequences were queried against the SWISS-PROT protein sequence database [72] using the NCBI BLAST server [73].

We also assessed the evolution of one particular gene that is known to promote photosymbiosis in *Tridacna* [39] and corals [40], the V-type H<sup>+</sup>-ATPase (VHA). We identified VHA genes from our *Tridacninae* and *Fraginae* species by running a BLAST search using the scleractinian coral (*Acropora yongei*) VHA amino acid sequence reported in [40]. Any bivalve sequences that share a more than 90% identity with the coral VHA were considered bivalve VHA. The bivalve sequences were then aligned with coral and Symbiodiniaceae [40] VHA sequences, as well as Genbank sequences from a mosquito (AAR13789), a fish (BAC75967) and an alveolate protist (XP\_002782779). A maximum likelihood

phylogeny with 100 bootstraps was constructed using PhyML as described above.

#### Fossil calibration and ancestral state reconstruction

A fossil-calibrated phylogeny was generated based on the best-supported topology from multiple analyses (see results and discussion for more details). We applied a penalized likelihood method [74] using a truncated Newton optimization algorithm implemented in r8s 1.81 [75]. This method was chosen to avoid computational limitations imposed by analyzing big datasets in a Bayesian framework, especially given that the two approaches tend to yield similar divergence time estimates (e.g., [76]). The *Matrix 1* ML phylogeny (outgroup removed) was used as the input tree for r8s. Fossil dates representing the earliest occurrences of well-defined taxa/genera were used to constrain the minimal node ages of four groups: *Fragum fragum*, *Americardia media*, *Tridacna*, and *Vasticardium* (Table 3, [77–80]). A cross-validation analysis [62] was performed on the ML tree with smoothing parameters varying from 1 to  $1 \times 10^5$  to determine the optimal parameter. To account for branch length and topological uncertainties, we repeated the analyses using 100 RAxML bootstrap trees obtained from previous phylogenetic analyses on *Matrix 1*. Node age statistics from the 100 analyses were summarized using the *profile* command in r8s.

To explore the origin of photosymbiosis in *Cardiidae*, ancestral state reconstructions were performed on the fossil calibrated phylogeny. A maximum likelihood reconstruction was conducted using the Binary State Speciation and Extinction (BiSSE) model [81] on the calibrated *Matrix 1* ML phylogeny. This method was chosen because it is the only method that accommodates incomplete taxon sampling, not because we pre-assumed photosymbiotic and non-symbiotic lineages exhibit different diversification rates. Photosymbiotic and non-photosymbiotic ecologies were coded as discrete characters for each taxon included in the ingroup. Sampling fraction for photosymbiotic (33%) and all non-symbiotic *cardiids* (6%) were calculated based on the most recent taxon records from the World Register of Marine Species (WoRMS). Given the known caveats of the BiSSE model [82], we also conducted a reconstruction using the mk2 model [83], which does not assume trait-based

**Table 3** Fossil records used for calibrating the *cardiid* phylogeny

Node	Taxa	Min. node age (MYA)	Source
A	<i>Fragum fragum</i>	3.6	[77]
B	<i>Americardia media</i>	7.2	[78]
C	<i>Tridacna</i>	23	[79]
D	<i>Vasticardium</i>	41.2	[80]

diversification rates, although the analysis does not take missing taxa into consideration. Both analyses were conducted using the “diversetree” package [84] in R 3.4.3.

Lastly, we compared the timing of bivalve photosymbiotic evolution to the abundance of global reef habitats through time, to assess the importance of habitat availability. The total number of global reef sites for the past 90 million years (following [28]) were overlaid on the fossil calibrated cardiid phylogeny for visual comparison.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s12862-020-01614-7>.

**Additional file 1 : Table S1.** BLAST search results of the 10% fastest-evolving genes (matrices A, B) used in the phylogenetic reconstruction.

**Additional file 2 : Table S2.** BLAST search results of the 10% slowest-evolving genes (matrices T-V) used in the phylogenetic reconstruction.

**Additional file 3 : Figure S1.** Gene occupancy representation per species for matrices 1 and 2. A white cell indicates a gene that was not sampled. Taxa are sorted from the highest (top) to lowest (bottom) gene representation.

**Additional file 4 : Figure S2.** A heat map of relative composition frequency variability (RCFV) values for all orthogroups in Matrix 1. Species name are coded as in Additional file 3: Figure S1.

**Additional file 5 : Figure S3.** Gene super-network for matrices 1 and 2. Species name are coded as in Additional file 3: Figure S1.

**Additional file 6 : Figure S4.** Maximum likelihood phylogeny of the VHA genes from Fraginae, Tridacninae, the coral *Acropora yongei*, Symbiodiniaceae, and other eukaryotes. Non-photosymbiotic bivalve taxa are labeled red. The bootstrap value for the animal VHA clade is shown above the branch.

## Abbreviations

BiSSE: Binary State Speciation and Extinction; MCMC: Markov chain Monte Carlo; ML: Maximum likelihood; RCFV: Relative composition frequency variability; SPR: Subtree pruning and regrafting; VHA: V-type H + -ATPase; WoRMS: World Register of Marine Species

## Acknowledgements

The authors thank Jan Johan ter Poorten for providing taxonomical expertise.

## Authors' contributions

J.L., L.K., S.L., C.C., and G.G. designed the study. G.G., R.B., and L.K. collected specimens. S.L. and J.L. carried out the experiments, generated data, and conducted analyses. All authors wrote the manuscript. The authors read and approved the final manuscript.

## Funding

This work is supported by NSF 1420967, National Geographic W367–15, and Packard Fellowship 2019-69653 to J.L. for the fieldwork, lab work, data analysis and manuscript writing; NSF 0732854, 0732903, and 0732860 to R.B., G.G. and Paula M. Mikkelsen for fieldwork and manuscript writing; and NSF 1457769 to S.L. for data analysis and manuscript writing. Transcriptome generation and sequencing are supported by internal funds from the Faculty of Arts and Sciences and the Museum of Comparative Zoology (Harvard) to G.G.

## Availability of data and materials

All reads generated for this study are deposited in the NCBI-SRA repository. All transcriptome accession and museum catalogue numbers are listed in Table 1. Fully assembled transcriptomes and phylogenies are deposited in Harvard dataverse ([https://dataverse.harvard.edu/dataverse/cardiid\\_transcriptome](https://dataverse.harvard.edu/dataverse/cardiid_transcriptome)).

## Ethics approval

The use of invertebrate animals does not require protocol approval by the Institutional Animal Care and Use Committee (IACUC), University of Colorado Boulder.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, USA. <sup>2</sup>Museum of Natural History, University of Colorado Boulder, Boulder, USA. <sup>3</sup>Organismic and Evolutionary Biology, Harvard University, Cambridge, USA. <sup>4</sup>University of Guam Marine Laboratory, Mangilao, Guam. <sup>5</sup>Museum of Comparative Zoology, Harvard University, Cambridge, USA. <sup>6</sup>Department of Aquatic Zoology, Western Australian Museum, Welshpool, Australia. <sup>7</sup>Negaunee Integrative Research Center, Field Museum of Natural History, Chicago, USA.

Received: 25 September 2019 Accepted: 15 April 2020

Published online: 01 May 2020

## References

- Stanley GD. Photosymbiosis and the evolution of modern coral reefs. *Science*. 2006;312:857–8.
- Kirkendale L, Paulay G. Part N, Volume 1, chapter 9: Photosymbiosis in Bivalvia. *Treatise Online*. 2017;89:1–39.
- Vermeij GJ. The evolution of molluscan photosymbioses: a critical appraisal. *Biol J Linn Soc*. 2013;109:497–511.
- Clavijo JM, Donath A, Serôdio J, Christa G. Polymorphic adaptations in metazoans to establish and maintain photosymbioses. *Biol Rev*. 2018;93:2006–20.
- Yonge CM. Mode of life, feeding, digestion and symbiosis with zooxanthellae in the Tridacnidae. *Great Barrier Reef Expedition, 1928–1929, scientific reports*. British Museum. 1936;1:283–321.
- Kawaguti S. Heart shell *Corculum cardissa* (L.) and its zooxanthella. *Kagaku Nanyo*. 1941;3:45–6.
- Kirkendale L. Their day in the sun: molecular phylogenetics and origin of photosymbiosis in the ‘other’ group of photosymbiotic marine bivalves (Cardiidae: Fraginae). *Biol J Linn Soc*. 2009;97:448–65.
- Li J, Volstead M, Kirkendale L, Cavanaugh C. Characterizing photosymbiosis between Fraginae bivalves and *Symbiodinium* using phylogenetics and stable isotopes. *Front Ecol Evol*. 2018;6:45.
- Hawkins AJS, Klumpp DW. Nutrition of the giant clam *Tridacna gigas* (L.). II. Relative contributions of filter-feeding and the ammonium-nitrogen acquired and recycled by symbiotic alga towards total nitrogen requirements for tissue growth and metabolism. *J Exp Mar Biol Ecol*. 1995; 190:263–90.
- Klumpp DW, Griffiths CL. Contributions of phototrophic and heterotrophic nutrition to the metabolic and growth requirements of four species of giant clam (Tridacnidae). *Mar Ecol Prog Ser*. 1994;115:103–15.
- LaJeunesse TC, Parkinson JE, Gabrielson PW, Jeong HJ, Reimer JD, Voolstra CR, Santos SR. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Curr Biol*. 2018;28:2570–80.
- Mansour K. Communication between the dorsal edge of the mantle and the stomach of *Tridacna*. *Nature*. 1946;157:844.
- Norton JH, Shepherd MA, Long HM, Fitt WK. The zooxanthellal tubular system in the giant clam. *Biol Bull*. 1992;183:503–6.
- Fitt WK, Fisher CR, Trench RK. Contribution of the symbiotic dinoflagellate *Symbiodinium microadriaticum* to the nutrition, growth and survival of larval and juvenile tridacnid clams. *Aquaculture*. 1986;55:5–22.
- Holt AL, Vahidinia S, Gagnon YL, Morse DE, Sweeney AM. Photosymbiotic giant clams are transformers of solar flux. *J R Soc Interface*. 2014;11:20140678.
- Farmer MA, Fitt WK, Trench RK. Morphology of the symbiosis between *Corculum cardissa* (Mollusca: Bivalvia) and *Symbiodinium corculorum* (Dinophyceae). *Biol Bull*. 2001;200:336–43.
- Carter JG, Schneider JA. Condensing lenses and shell microstructure in *Corculum* (Mollusca: Bivalvia). *J Paleontol*. 1997;71:56–61.

18. Schneider JA, Ó Foighil D. Phylogeny of giant clams (Cardiidae: Tridacninae) based on partial mitochondrial 16S rDNA gene sequences. *Mol Phylogenet Evol.* 1999;13:59–66.
19. Herrera ND, ter Poorten JJ, Bieler R, Mikkelsen PM, Strong EE, Jablonski D, Stepan SJ. Molecular phylogenetics and historical biogeography amid shifting continents in the cockles and giant clams (Bivalvia: Cardiidae). *Mol Phylogenet Evol.* 2015;93:94–106.
20. Schneider JA. Preliminary cladistic analysis of the bivalve family Cardiidae. *Am Malacol Bull.* 1992;9:145–55.
21. Schneider JA. Phylogeny of the Cardiidae (Bivalvia): phylogenetic relationships and morphological evolution within the subfamilies Clinocardiinae, Lymnocardiinae, Fraginae and Tridacninae. *Malacologia.* 1998;40:321–73.
22. Maruyama T, Ishikura M, Yamazaki S, Kanai S. Molecular phylogeny of zooxanthellate bivalves. *Biol Bull.* 1998;195:70–7.
23. Herrera ND. Molecular phylogenetics and historical biogeography of cockles and giant clams (Bivalvia: Cardiidae). M.Sc. thesis. Tallahassee: Florida State University; 2013.
24. González VL, Andrade SC, Bieler R, Collins TM, Dunn CW, Mikkelsen PM, Taylor JD, Giribet G. A phylogenetic backbone for Bivalvia: an RNA-seq approach. *Proc Royal Soc B.* 2015;282:20142332.
25. Lemer S, González VL, Bieler R, Giribet G. Cementing mussels to oysters in the pteriomorphian tree: a phylogenomic approach. *Proc Royal Soc B.* 2016;283:20160857.
26. Lemer S, Bieler R, Giribet G. Resolving the relationships of clams and cockles: dense transcriptome sampling drastically improves the bivalve tree of life. *Proc Royal Soc B.* 2019;286:20182684.
27. Combosch DJ, Collins TM, Glover EA, Graf DL, Harper EM, Healy JM, Kawauchi GY, Lemer S, McIntyre E, Strong EE, Taylor JD, Zardus JD, Mikkelsen PM, Giribet G, Bieler R. A family-level tree of life for bivalves based on a Sanger-sequencing approach. *Mol Phylogenet Evol.* 2017;107:191–208.
28. Kiessling W. Geologic and biologic controls on the evolution of reefs. *Annu Rev Ecol Evol Syst.* 2009;40:173–92.
29. Whitfield JB, Lockhart PJ. Deciphering ancient rapid radiations. *Trends Ecol Evol.* 2007;22:258–65.
30. Fernández R, Edgecombe GD, Giribet G. Exploring phylogenetic relationships within Myriapoda and the effects of matrix composition and occupancy on phylogenomic reconstruction. *Syst Biol.* 2016;65:871–89.
31. Jermiin LS, Ho SY, Ababneh F, Robinson J, Larkum AW. The biasing effect of compositional heterogeneity on phylogenetic estimates may be underestimated. *Syst Biol.* 2004;53:638–43.
32. Rasmussen MD, Kellis M. Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. *Genome Res.* 2007;17:1932–42.
33. Harzhauser M, Mandic O, Piller WE, Reuter M, Kroh A. Tracing back the origin of the Indo-Pacific mollusc fauna: basal Tridacninae from the Oligocene and Miocene of the Sultanate of Oman. *Palaeontology.* 2008;51:199–213.
34. Schluter D. *The ecology of adaptive radiation.* Oxford: Oxford University Press; 2000.
35. ter Poorten JJ. The Cardiidae of the Panglao marine biodiversity project 2004 and the Panglao 2005 Deep-Sea cruise with descriptions of four new species (Bivalvia). *Vita Malacologica.* 2009;8:9–96.
36. Losos JB, Glor RE, Kolbe JJ, Nicholson K. Adaptation, speciation, and convergence: a hierarchical analysis of adaptive radiation in Caribbean *Anolis* lizards. *Ann Missouri Bot Gard.* 2006;93:24–34.
37. Simpson GG. *Tempo and mode in evolution* (no. 15). New York: Columbia University Press; 1994.
38. Trench RK. Of 'leaves that crawl': functional chloroplasts in animal cells. In: *Symposia of the Society for Experimental Biology*, vol. 29; 1975. p. 229.
39. Armstrong EJ, Roa JN, Stillman JH, Tresguerres M. Symbiotic photosynthesis in giant clams is promoted by V-type H<sup>+</sup>-ATPase from host cells. *J Exp Biol.* 2018;221:jeb177220.
40. Barott KL, Venn AA, Perez SO, Tambutté S, Tresguerres M. Coral host cells acidify symbiotic algal microenvironment to promote photosynthesis. *Proc Natl Acad Sci U S A.* 2015;112:607–12.
41. Archibald JM. Endosymbiosis and eukaryotic cell evolution. *Curr Biol.* 2015;25:R911–21.
42. De Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J, Bittner L, Chaffron S, Dunthorn M, Engelen S, Karsenti E, et al. Eukaryotic plankton diversity in the sunlit ocean. *Science.* 2015;348:1261605. <https://science.sciencemag.org/content/348/6237/1261605>.
43. Cowen R. The role of algal symbiosis in reefs through time. *Palaios.* 1988;3:221–7.
44. Barcia R, Lopez-García JM, Ramos-Martínez JI. The 28S fraction of rRNA in molluscs displays electrophoretic behaviour different from that of mammal cells. *IUBMB Life.* 1997;42:1089–92.
45. McCarthy SD, Dugon MM, Power AM. 'Degraded' RNA profiles in Arthropoda and beyond. *PeerJ.* 2015;3:e1436.
46. Natsidis P, Schiffer PH, Salvador-Martínez I, Telford MJ. Computational discovery of hidden breaks in 28S ribosomal RNAs across eukaryotes and consequences for RNA integrity numbers. *Sci Rep.* 2019;9:1–10.
47. Krueger F. Trim galore. A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files; 2015.
48. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9:357.
49. Bayer T, Aranda M, Sunagawa S, Yum LK, DeSalvo MK, Lindquist E, Coffroth MA, Voolstra CR, Medina M. *Symbiodinium* transcriptomes: genome insights into the dinoflagellate symbionts of reef-building corals. *PLoS One.* 2012;7:e35269.
50. Ladner JT, Barshis DJ, Palumbi SR. Protein evolution in two co-occurring types of *Symbiodinium*: an exploration into the genetic basis of thermal tolerance in *Symbiodinium* clade D. *BMC Evol Biol.* 2012;12:217.
51. Pinzón JH, Kamel B, Burge CA, Harvell CD, Medina M, Weil E, Mydlarz LD. Whole transcriptome analysis reveals changes in expression of immune-related genes during and after bleaching in a reef-building coral. *R Soc Open Sci.* 2015;2:140214.
52. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, Di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644.
53. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nat Protoc.* 2013;8:1494.
54. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics.* 2012;28:3150–2.
55. Altenhoff AM, Schneider A, Gonnet GH, Dessimoz C. OMA 2011: orthology inference among 1000 complete genomes. *Nucleic Acids Res.* 2010;39:D289–94.
56. Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One.* 2013;8:e53786.
57. Roth AC, Gonnet GH, Dessimoz C. Algorithm of OMA for large-scale orthology inference. *BMC Bioinform.* 2008;9:518.
58. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–7.
59. Wu M, Chatterji S, Eisen JA. Accounting for alignment uncertainty in phylogenomics. *PLoS One.* 2012;7:e30288.
60. Hejnol A, Obst M, Stamatakis A, Ott M, Rouse GW, Edgecombe GD, Martínez P, Baguña J, Bailly X, Jondelius U, Wiens M, Müller WEG, Seaver E, Wheeler WC, Martindale MQ, Giribet G, Dunn CW. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc Royal Soc B.* 2009;276:4261–70.
61. Smith SA, Dunn CW. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics.* 2008;24:715–6.
62. Zoller S, Schneider A. Improving phylogenetic inference with a semiempirical amino acid substitution model. *Mol Biol Evol.* 2012;30:469–79.
63. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–3.
64. Lartillot N, Rodrigue N, Stubbs D, Richer J. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol.* 2013;62:611–5.
65. Lartillot N, Philippe H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol.* 2004;21:1095–109.
66. Kück P, Struck TH. BaCoCa-A heuristic software tool for the parallel assessment of sequence biases in hundreds of gene and taxon partitions. *Mol Phylogenet Evol.* 2014;70:94–8.

67. Foster PG. Modeling compositional heterogeneity. *Syst Biol*. 2004;53:485–95.
68. Susko E, Roger AJ. On reduced amino acid alphabets for phylogenetic inference. *Mol Biol Evol*. 2007;24:2139–50.
69. Berger SA, Krompass D, Stamatakis A. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Syst Biol*. 2011;60:291–302.
70. Grunewald S, Spillner A, Bastkowski S, Bogershausen A, Moulton V. SuperQ: computing supernetworks from quartets. *IEEE/ACM Trans Comput Biol Bioinform*. 2013;10:151–60.
71. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol*. 2005;23:254–67.
72. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res*. 2000;28:45–8.
73. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
74. Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol*. 2002;19:101–9.
75. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 2003;19:301–2.
76. Li HL, Wang W, Mortimer PE, Li RQ, Li DZ, Hyde KD, Xu JC, Soltis DE, Chen ZD. Large-scale phylogenetic analyses reveal multiple gains of actinorhizal nitrogen-fixing symbioses in angiosperms associated with climate change. *Sci Rep*. 2015;5:14023.
77. Nomura S, Zinbo N. Fossil and recent Mollusca from the island of Kita-Daito-Zima. *Sci Rep Tohoku Univ Geol*. 1935;18:41–51.
78. Vokes HE. Neogene paleontology in the northern Dominican Republic 9. The family Cardiidae (Mollusca: Bivalvia). *Bull Am Paleontol*. 1989;97:95–160.
79. Ladd HS, Schlanger SO. Drilling operations on Eniwetok atoll. *U S Geol Surv Prof Pap*. 1960;260:863–903.
80. Oppenheim P. Zur Kenntnis alttertiärer Faunen in Aegypten. *Palaeontographica*. 1903;30:1–348.
81. Maddison WP, Midford PE, Otto SP. Estimating a binary character's effect on speciation and extinction. *Syst Biol*. 2007;56:701–10.
82. Rabosky DL, Goldberg EE. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst Biol*. 2015;64:340–55.
83. Pagel M. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc Royal Soc B*. 1994;255:37–45.
84. FitzJohn RG. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods Ecol Evol*. 2012;3:1084–92.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

