



## **Transcriptomes of Giant Sea Anemones from Okinawa as a Tool for Understanding Their Phylogeny and Symbiotic Relationships with Anemonefish**

Authors: Kashimoto, Rio, Tanimoto, Miyako, Miura, Saori, Satoh, Noriyuki, Laudet, Vincent, et al.

Source: Zoological Science, 39(4) : 374-387

Published By: Zoological Society of Japan

URL: <https://doi.org/10.2108/zs210111>

---

BioOne Complete ([complete.BioOne.org](https://complete.BioOne.org)) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](https://www.bioone.org/terms-of-use).

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

# Transcriptomes of Giant Sea Anemones From Okinawa as a Tool for Understanding Their Phylogeny and Symbiotic Relationships With Anemonefish

Rio Kashimoto<sup>1,2</sup>, Miyako Tanimoto<sup>3</sup>, Saori Miura<sup>1</sup>, Noriyuki Satoh<sup>2\*</sup>,  
Vincent Laudet<sup>1,4\*</sup>, and Konstantin Khalturin<sup>2\*</sup>

<sup>1</sup>Marine Eco-Evo-Devo Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan

<sup>2</sup>Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna, Okinawa 904-0495, Japan

<sup>3</sup>Okinawa Churaumi Aquarium, 424 Ishikawa, Motobu, Kunigami District, Okinawa 905-0206, Japan

<sup>4</sup>Marine Research Station, Institute of Cellular and Organismic Biology (ICOB), Academia Sinica, 23-10, Dah-Uen Rd, Jiau Shi, I-Lan 262, Taiwan

The relationship between anemonefish and sea anemones is one of the most emblematic examples of mutualistic symbiosis in coral reefs. Although this is a textbook example, the major aspects of this symbiosis are still not fully understood in mechanistic terms. Moreover, since studies of this relationship have usually been focused on anemonefish, much less is known about giant sea anemones, their similarities, their phylogenetic relationships, and their differences at the molecular level. Since both partners of the symbiotic relationship are important, we decided to explore this well-known phenomenon from the perspective of giant sea anemones. Here, we report reference transcriptomes for all seven species of giant sea anemones that inhabit fringing reefs of Okinawa (Japan) and serve as hosts for six species of local anemonefish. Transcriptomes were used to investigate their phylogenetic relations, genetic differences and repertoires of nematocyte-specific proteins. Our data support the presence of three distinct groups corresponding to three genera: *Entacmaea*, *Heteractis*, and *Stichodactyla*. The basal position among the three groups belongs to *Entacmaea*, which was the first to diverge from a common ancestor. While the magnitude of genetic difference between the representatives of *Entacmaea* and *Stichodactyla* is large, intra-specific variation within *Stichodactyla* is much smaller and seems to result from recent speciation events. Our data reconfirms that *Heteractis magnifica* belongs to the genus *Stichodactyla*, despite an overall morphological similarity with representatives of the genus *Heteractis*. The availability of reference transcriptomes will facilitate further research into the fascinating relationship between sea anemones and anemonefish.

**Key words:** sea anemones, anemonefish, transcriptome, phylogeny, nematocysts, dinoflagellate, Symbiodiniaceae, *Cladocopium*

## INTRODUCTION

The relationship between anemonefish and sea anemones is one of the most recognizable examples of mutualistic symbiosis in coral reefs (Apprill, 2020). This symbiosis had provided a unique ecological opportunity that resulted in the evolutionary radiation of anemonefish, a monophyletic group of ca. 30 species that diverged 7–12 million years ago (Litsios et al., 2012; Marcionetti et al., 2019). Indeed, after

the acquisition of this association, anemonefish have diversified into multiple ecological niches depending on the host and habitat use (Litsios et al., 2012). It is known that the fish and the sea anemone mutually protect each other: the sea anemone provides shelter since the fish is not harmed by the stinging tentacles (Nedosyko et al., 2014), and the sea anemone is protected from the predators due to the aggressive behavior of the fish defending its territory (Burke da Silva and Nedosyko, 2016). In addition, fish and the sea anemones are linked by complex metabolic exchanges with a third symbiotic partner, an intracellular photosynthetic dinoflagellate (from the family Symbiodiniaceae), which inhabits the gastric epithelium of the latter (Cleveland et al., 2011; Verde et al., 2015; LaJeunesse et al., 2018). Lastly, fish

\* Corresponding author. E-mail: norisky@oist.jp (NS);  
vincent.laudet@oist.jp (VL);  
konstantin.khalturin@oist.jp (KK)

doi:10.2108/zs210111

have been shown to saturate their host with oxygen at night when photosynthesis is not taking place, which has a beneficial effect on anemone metabolism (Alan et al., 2015).

Despite being a textbook example of mutualism, many aspects of this symbiosis are still not fully understood from a mechanistic point of view (Burke da Silva and Nedosyko, 2016). For example, it is still unclear how an anemonefish defends itself against sea anemone toxins and does not trigger the discharge of sea anemone nematocysts (Burke da Silva and Nedosyko, 2016; Roux et al., 2020). But if we look at this symbiosis from the anemone point of view, it is even more mysterious because anemones are sometimes found without fish, unlike anemonefish, which never live without their cnidarian host. Thus, the benefit of this symbiosis for sea anemones seems less obvious, although it has been shown that sea anemones grow more rapidly when anemonefish are present (Mariscal et al., 1993; Porat and Chadwick-Furman, 2004). Since both partners are important in a symbiotic relationship, we decided to investigate this well-known phenomenon from the anemone's point of view.

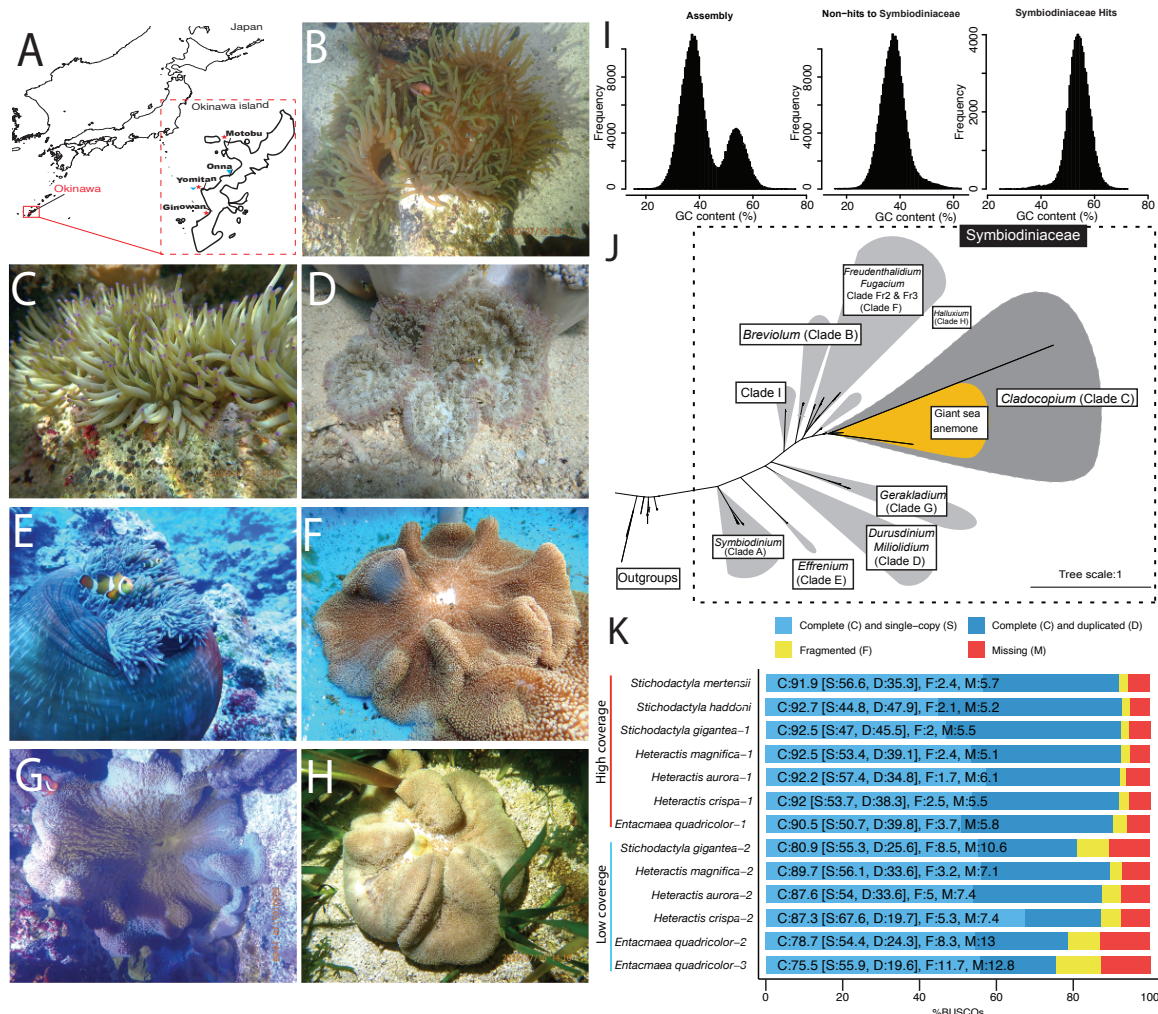
The 30 species of anemonefish can live in symbiosis with 10 species of sea anemones, but not all combinations are possible (Mariscal et al., 1993; Miyagawa-Kohshima et al., 2014; Burke da Silva and Nedosyko, 2016). There is a very clear specificity in the interaction. Some species of anemonefish (e.g., *Amphiprion clarkii*) are generalists that can live with many sea anemone species, whereas others (e.g., *A. frenatus*) live in only one sea anemone species and are therefore considered as specialists (Litsios et al., 2014; Burke da Silva and Nedosyko, 2016). Many intermediate situations exist between these two extremes (e.g., *A. ocellaris* living in three species), and there appear to be some geographical variations in association specificity, probably related to competition between species or possible imprinting for the host selection (Miyagawa-Kohshima et al., 2014; Burke da Silva and Nedosyko, 2016). It has been shown that anemonefish mucus lacks a specific N-acetylneuraminic acid, a 9-carbon monosaccharide commonly found in fish skin mucus, which triggers cnidocyte discharge (Mariscal et al., 1993; Abdullah and Saad, 2015). It is tempting to connect this observation to the fact that out of 17 genes found to be under positive selection at the base of the anemonefish radiation, two (*versican* and *N-acetylglucosaminyl hydrolase*) are associated with N-acetylated sugars (Marcionetti et al., 2019). However, although very interesting, these observations alone cannot explain the specificity of associations between certain species of anemonefish and sea anemones, nor the existence of specialists or generalist species. Therefore, additional information on the sea anemone side, for example, comparison of their genetic diversity, gene sets in different species, and the variations in the genes responsible for nematocyte development, is now essential.

Our knowledge of giant sea anemones living in association with *Amphiprion* and *Premnas* is still imperfect. Recent molecular phylogenetic analyses have confirmed that the 10 species of giant sea anemone living in association with anemonefish do not form a monophyletic group but, in fact, three distinct ensembles within Actiniaria: (i) *Entacmaea*; (ii) *Stichodactyla* + *Cryptodendrum* and (iii) *Heteractis* + *Macrodactyla* (Titus et al., 2019; Nguyen et al., 2020). These three groups differ drastically in their morphology, with

*Stichodactyla* species being the largest in size, followed by the representatives of *Entacmaea* and *Heteractis* (Mariscal et al., 1993). However, it came as a surprise that the magnificent sea anemone *Heteractis magnifica* has been shown to cluster inside the *Stichodactyla* group and not within *Heteractis* (Titus et al., 2019; Nguyen et al., 2020). However, in the previous analyses, the placement of different species within these three groups was often poorly resolved. For example, within *Heteractis*, *Heteractis crispa* was found in three different clades, sometimes associated with other *Heteractis*, sometimes even closer to *Macrodactyla doreensis* (Titus et al., 2019; Nguyen et al., 2020). It is difficult to determine whether this result is linked to slowly evolving phylogenetic markers used in these studies or to deeper causes due to our still limited knowledge of the taxonomy of these animals. This clearly shows that more work is needed to identify phylogenetic markers that will improve our understanding of the taxonomy and phylogeny of giant sea anemones (Titus et al., 2019). Indeed, if we are not yet sure about something as basic as accurate species identification, how can we hope to understand the general rules governing the symbiotic relationship between sea anemones and anemonefish?

The three main groups of giant anemones have been characterized for their toxicity using several functional tests, including acute toxicity to brine shrimps, hemolytic activity to sheep erythrocytes, and neurotoxicity to shore crabs (Nedosyko et al., 2014). This led to the interesting observation that the anemones with intermediate toxicity are associated with the highest number of fish species, while sea anemones with very low or very high toxicity are able to host far fewer fish species (Nedosyko et al., 2014). This suggests that variation in toxicity among sea anemone hosts is important in the establishment and maintenance of the symbiosis. But, in contrast to other cnidarian models such as *Hydra* or *Nemastostella* the genomic underpinning of nematocyte specific gene sets has not been well studied in the giant sea anemones.

For all these reasons, we started a research program to improve the knowledge of giant sea anemones living in association with anemonefish. As the first step, we established a reference transcriptome dataset of the seven species (*Entacmaea quadricolor*, *Stichodactyla haddonii*, *S. gigantea*, *S. mertensii*, *Heteractis magnifica*, *H. crispa*, and *H. aurora*) that inhabit fringing reefs of Okinawa, Japan (Fig. 1 A–H) and are known to participate in symbiotic relationships with six species of anemonefish (Hayashi et al., 2018). Our goal was to estimate phylogenetic relationships among these sea anemones and to find genetic characteristics which might be responsible for differences among them. For each species, we sequenced the transcriptome of the tentacles. We selected this tissue because this is the main area where anemones and fish interact. Tentacles contain various types of nematocytes whose stings the anemonefish are able to avoid. Since the gastrodermis of sea anemones contains endosymbiotic dinoflagellates (of the family Symbiodiniaceae), the resulting transcriptomes contained both cnidarian and algal transcripts, which were separated by a computational approach. Final cleaned transcriptomes and corresponding protein sets have high BUSCO values and, thereby, capture the majority of the genes present in



**Fig. 1.** Sampling locations, specimens, removal of Symbiodiniaceae, and statistics for the cleaned transcriptome assemblies. **(A)** Samples were collected on the main island of Okinawa in Ginowan, Yomitan, Onna, and Motobu regions. Red stars: sampling locations for high coverage sequencing. Blue arrows: sampling locations for low coverage sequencing. **(B)** *Entacmaea quadricolor*, 30 cm in diameter. **(C)** *Heteractis crista*, 25 cm in diameter. **(D)** *Heteractis aurora*, 33 cm in diameter. **(E)** *Heteractis magna*, 70 cm in diameter. **(F)** *Stichodactyla mertensii*, 55 cm in diameter. **(G)** *Stichodactyla gigantea*, 30 cm in diameter. **(H)** *Stichodactyla haddoni*, 25 cm in diameter. **(I)** Separation of cnidarian and Symbiodiniaceae transcripts based on their GC content. Initial transcriptome assembly of *E. quadricolor* (left panel) and selection of cnidarian (middle) and Symbiodiniaceae (right panel) transcripts. Peak with ~38% GC corresponds to *E. quadricolor* sequences and peak with ~57% GC to the transcripts of *Cladocopium*. **(J)** Phylogenetic tree of LSU sequences from algae identified in seven species of Okinawan sea anemones. Reference sequences from nine Symbiodiniaceae clades (grey bubbles) as well as species outside Symbiodiniaceae were used as references. Location of the sequences obtained from our samples is highlighted with orange bubble. **(K)** Plot with BUSCO values for high coverage (seven samples) and low coverage (six samples) transcriptome assemblies after the removal of algal transcripts.

each of the species. Using these data as references, we show that phylotranscriptomics can be used as an efficient strategy to improve our knowledge of the diversity within each of the three clades of a giant sea anemone. Thus, our analysis is an important step towards a comprehensive genetic characterization of giant sea anemones from Japan.

## MATERIALS AND METHODS

### Sampling and sequencing

Depending on the species, from two to seven tentacles were collected per individual sea anemone. The species identification is described in Supplementary Text S2. Tentacles were cut by scissors and homogenized in 700  $\mu$ L of Trizol (Thermo Fisher Scientific, USA; Catalog #155596018). The resulting solution was immediately snap-frozen in liquid nitrogen. For high coverage transcriptomes,

samples for *E. quadricolor*, *S. haddoni* and *S. gigantea* were obtained from the marine aquarium in Yomitan Village (Umi-no-Tane Aquarium). Samples of *H. crista* and *H. aurora* were collected from the reef near Senaha Beach in Yomitan and *S. mertensii* from the Okinawa Churaumi Aquarium (see Table 1 for the coordinates of sampling locations). *Heteractis magna* from the Ginowan region was provided by a local fisherman. Animals kept in Umi-no-Tane Aquarium and Okinawa Churaumi Aquarium were originally collected from local fringing reefs within Yomitan and Motobu regions, respectively. Total RNA was extracted from the tentacles frozen in Trizol according to the manufacturer's protocol, and 4  $\mu$ g of RNA was used for library construction with a TruSeq Stranded mRNA kit (Illumina, USA; Catalog #20020594). RNA quality was checked by gel electrophoresis prior to library construction. Libraries were sequenced on the NovSeq6000 instrument with  $138 \times 10^6 - 186 \times 10^6$  reads per sample or  $34.5 \times 10^6 - 46 \times 10^6$  reads

**Table 1.** Description of the samples.

Sample No.	Species/Sample Number	Figure	Sampling Coordinates	Locality on Okinawa	Depth (m)	BioSample/TSA accession	Anemonefish species
1	<i>Entacmaea quadricolor</i> -1	1 B	26°24'31.9"N 127°42'55.9"E	Umi-no-tane Aquarium	0.3	SAMN18817341/GJFF00000000	<i>A. frenatus</i>
2	<i>Stichodactyla mertensii</i>	1 F	26°41'40.2"N 127°52'40.9"E	Okinawa Churaumi Aquarium	6	SAMN18817343/GJFE00000000	<i>A. polymnus</i>
3	<i>Stichodactyla haddoni</i>	1 H	26°24'31.9"N 127°42'55.9"E	Okinawa Churaumi Aquarium	0.3	SAMN18817345/GJFH00000000	no fish
4	<i>Stichodactyla gigantea</i> -1	1 G	26°24'31.9"N 127°42'55.9"E	Okinawa Churaumi Aquarium	0.3	SAMN18817344/GJFK00000000	<i>A. ocellaris</i>
5	<i>Heteractis crispa</i> -1	1 C	26°42'73"N 127°73'39" E	Yomitan	2.5	SAMN18817339/GJFA00000000	<i>A. clarkii</i>
6	<i>Heteractis aurora</i> -1	1 D	26°42'68"N, 127°73'51"E	Yomitan	1	SAMN18817340/GJFG00000000	<i>A. clarkii</i>
7	<i>Heteractis magnifica</i> -1	1 E	26°28'97.10"N 127°71'80.84"E	GINOWAN	8	SAMN18817342/GJFJ00000000	<i>A. ocellaris</i>
8	<i>Entacmaea quadricolor</i> -2	S2 A	26°42'96"N, 127°73'56"E	Yomitan	5	SAMN18817348	<i>A. frenatus</i>
9	<i>Entacmaea quadricolor</i> -3	S2 B	26°42'96"N, 127°73'56"E	Yomitan	7.5	SAMN18817349	<i>A. frenatus</i>
10	<i>Stichodactyla gigantea</i> -2	S2 C	26°30'25.2"N 127°52'45.3"E	Onna	4	SAMN18817351	<i>A. ocellaris</i>
11	<i>Heteractis crispa</i> -2	S2 D	26°30'25.2"N 127°52'45.3"E	Onna	2	SAMN18817346	<i>A. clarkii</i> , <i>Dascyllus trimaculatus</i>
12	<i>Heteractis aurora</i> -2	S2 E	26°42'90"N, 127°73'60"E	Yomitan	8	SAMN18817347	<i>A. clarkii</i> , <i>D. trimaculatus</i> , <i>D. reticulatus</i>
13	<i>Heteractis magnifica</i> -2	S2 F	26°42'95"N, 127°73'48"E	Yomitan	12	SAMN18817350	<i>A. ocellaris</i>

**Table 2.** Statistics of the transcriptome assemblies.

Sample No.	Species/Sample Number	Data amount per species (million reads)	Yield (Mbases)	Total Transcripts	N50 (bases)	Longest Transcript (bases)	BUSCO (cleaned transcriptome)	BUSCO (proteins)	CG%
1	<i>Entacmaea quadricolor</i> -1	145	21,910	341,093	1,848	33,093	C:90.5%[S:50.7%,D:39.8%], F:3.7%,M:5.8%,n:954	C:86.6% [S:85.1%,D:1.5%], F:4.7%,M:8.7%,n:954	43.32
2	<i>Stichodactyla mertensii</i>	156.2	23,580	269,130	2,063	37,745	C:91.9%[S:56.6%,D:35.3%], F:2.4%,M:5.7%,n:954	C:88.1% [S:86.8%,D:1.3%], F:3.7%,M:8.2%,n:954	45.46
3	<i>Stichodactyla haddoni</i>	186.5	28,156	343,778	2,302	53,044	C:92.7%[S:44.8%,D:47.9%], F:2.1%,M:5.2%,n:954	C:87.5% [S:86.6%,D:0.9%], F:3.5%,M:9.0%,n:954	43.75
4	<i>Stichodactyla gigantea</i> -1	169.6	25,612	323,507	2,178	64,912	C:92.5%[S:47.0%,D:45.5%], F:2.0%,M:5.5%,n:954	C:88.0% [S:87.1%,D:0.9%], F:2.9%,M:9.1%,n:954	43.71
5	<i>Heteractis crispa</i> -1	138.7	20,949	252,997	1,986	32,026	C:92.0%[S:53.7%,D:38.3%], F:2.5%,M:5.5%,n:954	C:87.2% [S:86.4%,D:0.8%], F:3.6%,M:9.2%,n:954	45.57
6	<i>Heteractis aurora</i> -1	161.4	24,377	243,055	1,944	32,920	C:92.2%[S:57.4%,D:34.8%], F:1.7%,M:6.1%,n:954	C:88.5% [S:87.0%,D:1.5%], F:2.5%,M:9.0%,n:954	45.04
7	<i>Heteractis magnifica</i> -1	148.9	22,489	286,530	2,045	35,951	C:92.5%[S:53.4%,D:39.1%], F:2.4%,M:5.1%,n:954	C:88.2% [S:86.5%,D:1.7%], F:1.6%,M:10.2%,n:954	44.29
8	<i>Entacmaea quadricolor</i> -2	42.9	5,660	134,868	1,561	20,945	C:78.7%[S:54.4%,D:24.3%], F:8.3%,M:13.0%,n:954	C:78.7%[S:55.3%,D:23.4%], F:9.4%,M:11.9%,n:954	46.22
9	<i>Entacmaea quadricolor</i> -3	37.5	6,478	127,976	1,526	19,233	C:75.5%[S:55.9%,D:19.6%], F:11.7%,M:12.8%,n:954	C:75.9%[S:57.8%,D:18.1%], F:11.9%,M:12.2%,n:954	45.81
10	<i>Stichodactyla gigantea</i> -2	46	6,208	118,439	1,516	24,056	C:80.9%[S:55.3%,D:25.6%], F:8.5%,M:10.6%,n:954	C:86.2%[S:70.2%,D:16.0%], F:5.7%,M:8.1%,n:954	44.98
11	<i>Heteractis crispa</i> -2	39.7	5,991	126,553	1,594	29,259	C:87.3%[S:67.6%,D:19.7%], F:5.3%,M:7.4%,n:954	C:87.5%[S:56.4%,D:31.1%], F:5.0%,M:7.5%,n:954	45.15
12	<i>Heteractis aurora</i> -2	41.1	5,220	120,341	1,721	31,875	C:87.6%[S:54.0%,D:33.6%], F:5.0%,M:7.4%,n:954	C:87.9%[S:59.6%,D:28.3%], F:4.5%,M:7.6%,n:954	44.06
13	<i>Heteractis magnifica</i> -2	34.5	6,945	121,458	1,413	26,376	C:89.7%[S:56.1%,D:33.6%], F:3.2%,M:7.1%,n:954	C:80.1%[S:57.0%,D:23.1%], F:10.0%,M:9.9%,n:954	44.95

per sample for high coverage and low coverage transcriptomes, respectively (Table 1).

#### Assembly and annotation

Sequenced reads were processed using Trimmomatic v.0.36 (Bolger et al., 2014) in order to remove low-quality regions and any

remaining Illumina adapters. After quality trimming, the reads were assembled by using Trinity v.11 using Deigo HPC at Okinawa Institute of Science and Technology Graduate University (OIST), and redundant transcripts were removed by using CD-HIT-EST with 95% similarity cut-off (Fu et al., 2012). After the cd-hit-est step, the number of assembled transcripts per species ranged from 212,053

in *H. crispa* to 290,200 in *S. haddoni*. The quality of the resulting assemblies was assessed by using BUSCO v.4.1.2 (Table 2).

Initial assemblies contained a mixture of transcripts from sea anemones and Symbiodiniaceae (Fig. 1I). In order to detangle these meta-transcriptomes, we created a BLAST database with the collection of transcripts from *Symbiodinium* (former clade A), *Breviolum* (former clade B) and *Cladocopium* (former clade C). All the transcripts were screened against this database by using BLASTN with cut-off E-value of  $1e-20$ . It turned out that all seven species of sea anemones contained only *Cladocopium* (former clade C) as their symbiont (see Fig. 1J). Next, we calculated GC content for each transcript, analyzed the distribution of GC content frequencies in each species, and classified all transcripts into two categories based on their BLASTN matches against Symbiodiniaceae and GC content (see Fig. 1J). Since dinoflagellates and cnidarians differ drastically in their GC content (Anthozoa ~40% whereas Symbiodiniaceae ~55%), all transcripts without BLAST hits to algae, and GC content below 60% were referred to as cnidarian once. Sequences with GC content above 40% and hits to Symbiodiniaceae were considered to be from dinoflagellates. This approach allowed us to separate the transcriptomes of the sea anemones and their symbionts effectively (Fig. 1I). At the same time, the fraction of complete + fragmented BUSCOs in the resulting 'cleaned' transcriptomes remained above 94.2% (see Fig. 1K and Table 2).

After removing the algae sequences, for each transcriptome, the corresponding proteins were predicted using a combination of Transdecoder and EstScan (Lottaz et al., 2003; Haas et al., 2013). Only the transcripts encoding open reading frames (ORFs) with  $\geq 70$  amino acids were taken into further consideration. The number of predicted proteins per species varied between 53,485 for *H. aurora* and 84,403 for *S. haddoni* (see Supplementary Table S2). Redundancy in the sets of predicted proteins was reduced ~40% by using CD-HIT with a 95% similarity cut-off. For example, 32,813 peptides in *H. aurora* and 39,518 in *S. haddoni* remained after CD-HIT processing. Transcriptomes assembled by Trinity represent all detected splice variants of the genes and, therefore, multiple protein isoforms derived from the same gene are usually present. In order to reduce this redundancy, only one representative ORF from all splice variants of a gene was selected. Namely, for each group of transcripts representing a gene (in Trinity output, such transcripts differ in their last index, i.e., xxx\_i1, xxx\_i2, ... xxx\_iN), we selected a transcript with the longest open reading frame. This approach allowed us to further reduce the number of predicted proteins per species (see Supplementary Table S2). After this filtering step, the number of predicted proteins per species matched well with the number of genes in the sequenced anthozoan genomes. For example, the final number of proteins was 25,690 and 26,535 in *H. crispa* and *S. haddoni*, respectively, with the mean number of proteins per species being 25,996. In the genomes of *Nematostella*, *Acropora*, and *Exaiptasia*, 27,273, 23,700, and 29,269 genes were predicted, respectively (Putnam et al., 2007; Shinzato et al., 2011; Potter et al., 2018). BUSCO values for the proteomes of seven anemone species are shown in Table 2.

### Low-coverage transcriptomes

Generation of reference transcriptomes for seven species required one lane on a NovaSeq6000 SP flow cell. This number of raw reads is excessive for phylogenetic reconstructions, and sufficient data could be generated with ~4-fold fewer Illumina reads. With 34–45 million reads per sample, the quality of assemblies after the removal of *Cladocopium* transcripts showed only a slight reduction of BUSCO values compared to that for high-coverage sequencing. As shown in Fig. 1K, BUSCO values ranged from 75.5% in *E. quadricolor* to 89.7% in *H. magnifica* (see Table 2 for details). Six low-coverage transcriptomes were used to estimate the levels of intra-specific variability among sea anemones.

### Data availability

Raw sequence data were submitted to NCBI SRA under BioProject number PRJNA723429. Cleaned transcriptome assemblies of sea anemones were submitted to NCBI TSA archive under accessions GJFF00000000 (*E. quadricolor*), GJFG00000000 (*H. aurora*), GJFA00000000 (*H. crispa*), GJFJ00000000 (*H. magnifica*), GJFK00000000 (*S. gigantea*), GJFH00000000 (*S. haddoni*) and GJFE00000000 (*S. mertensii*). Reference protein sets (anemones.tar.gz) can be downloaded from OIST BLAST server (<http://compagen.unit.oist.jp/aurelia/datasets.html>). BLAST search against nucleotide and protein sequences can be done at <http://compagen.unit.oist.jp/aurelia/blast.html>. To access these resources, please use login "guest" and password "welcome".

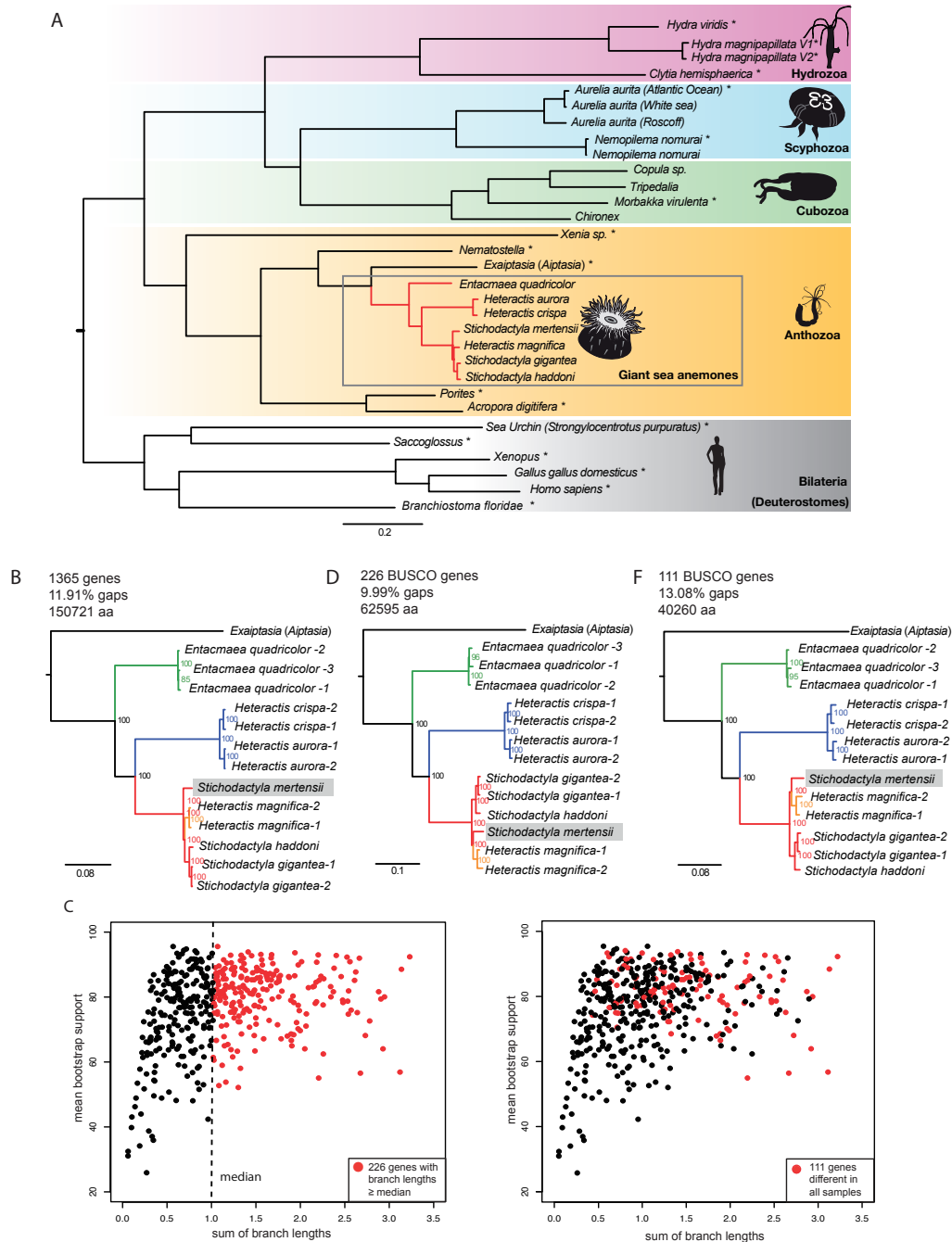
### Phylogeny reconstruction

In order to reconstruct the phylogenetic position of sea anemones, the protein sets from 16 cnidarian species and six bilaterian genomes were used as reference datasets (Fig. 2A and see Supplementary Table S3). Orthologous genes were identified by using OrthoFinder-v2.4.0 with default settings (Emms and Kelly, 2019). Phylogenetic relationships of sea anemones and their placement within Cnidaria were assessed based on concatenated multi-sequence alignments of one-to-one orthologs generated by MAFFT with '--maxiterate 1000 --localpair --leavegappyregion' settings (Kato and Standley, 2013). Low complexity regions in the alignments were removed by using TrimAL-1.2rev59 with '-gappyout' option (Capella-Gutiérrez et al., 2009). Individual alignments were concatenated by FASconCAT-G v1.02 (Kück and Longo, 2014). Phylogenetic trees were reconstructed with MPI version of RAxML 8.2.4 (raxmlHPC-MPI-AVX) with 100 rapid bootstrap inferences followed by a thorough ML search with the '-m PROTGAMMALG' option (Fig. 2A–F). Data matrices were partitioned by genes. Trees were visualized and re-rooted by using FigTree-v1.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

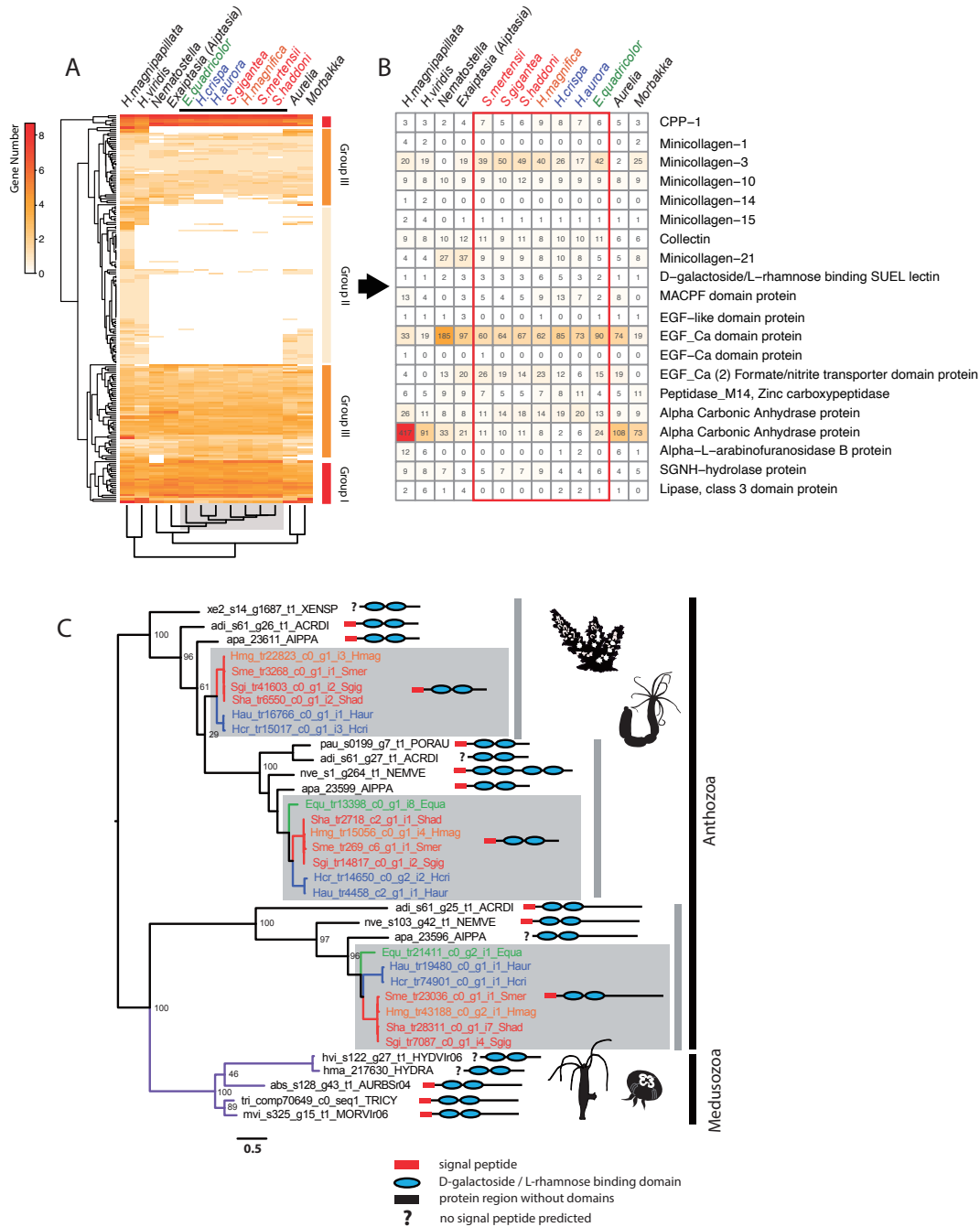
### Nematocyst-specific genes

We used the list of 410 nematocyst-specific genes of *Hydra* (Balasubramanian et al., 2012) as a reference set in order to identify putative nematocyst-specific proteins of the sea anemones. Orthologs of these nematocyst-specific proteins were identified in the proteomes of 24 cnidarian species by OrthoFinder v2.4.0. As a result of this similarity-based clustering, in each of the species, we identified groups of proteins ('orthogroups' in terms of OrthoFinder) which contained orthologs of the nematocyst-specific proteins of *Hydra*. Hence, for each species, we could estimate the number of proteins that putatively take part in the formation of nematocysts. In order to better group together paralogous sequences (for example, minicollagens) we reduced the inflation rate parameter "I" of Markov clustering algorithm (MCL) utilized by OrthoFinder from its default value of 1.5 to 1.2 (Enright et al., 2002). This setting is based on our empirical observation that the most accurate grouping of different classes of minicollagens and other structural capsule proteins is achieved with  $I = 1.2$  (Khalturin et al., 2019). Nematocyst proteins often contain repetitive regions, areas of low complexity, and multiple repeated domains. Due to these features OrthoFinder often erroneously places the paralogous genes into different orthologous groups if the inflation rate parameter is set to 1.5. The number of putative nematocyst-specific proteins in each cnidarian species is shown in Supplementary Table S6. A reduced dataset in which 13 species were clustered based on the similarity of their nematocyst-specific protein sets is shown in Fig. 3A. Several examples of variation in the number of proteins among the species are shown in Fig. 3B (a subset of Supplementary Table S6). A high-resolution heatmap representing the distribution of nematocyst-specific genes (see Fig. 3A) is shown in Supplementary Figure S1.

Sequences of D-galactoside/L-rhamnose binding SUEL lectins from the representatives of Anthozoa and Medusozoa were aligned



**Fig. 2.** Phylogenetic relationships of seven species of giant sea anemones from Okinawa. **(A)** Maximum likelihood (ML) tree based on concatenation of 354 proteins which are present in all species of the anemones and reference datasets. Protein sets from 16 species representing the major cnidarian clades as well as proteomes from six genomes of bilaterians were used as references. The data matrix consists of 97,598 distinct alignment patterns with 22.07% gaps or undetermined characters. LG + GAMMA substitution model with matrix partitioning according to the genes. All bipartitions have the maximum bootstrap support. Branches leading to seven sequenced species are shown in red and marked with a grey rectangle. **(B–F)** Phylogeny reconstructions based on gene sets with different rates of evolution. **(B)** ML tree based on a concatenated supermatrix of 1365 genes present in all samples. This shows the power of RNAseq in terms of data availability compared to single-gene approaches used previously. **(C)** The plot of mean bootstrap support versus the sum of branch lengths for 451 gene trees representing BUSCO proteins. Genes for which the sum of branch lengths  $\geq$  median value of 1.01986 are marked as red dots. Among 451 BUSCO genes, 226 genes had a sum of branch lengths  $\geq$  median. This is a subset of genes with a higher evolutionary rate, which is good for distinguishing differences among closely related species or even within the species. **(D)** Tree based on 226 BUSCO genes with an elevated rate of evolution (sum of branch lengths above the median). Here we get a higher BS support value for Equ/Eq3 branch (96% compared to 85% in **(B)**). **(E)** The plot of mean bootstrap support versus the sum of branch length for 451 gene trees. 111 BUSCO proteins for which these sequences differ in all samples of our dataset are marked with red dots. The remaining 340 genes are shown as black dots. **(F)** Phylogenetic tree based on 111 BUSCO proteins for which these sequences are different in all 13 samples of the giant sea anemones. Potentially this is the optimal marker set for population genetics studies.



**Fig. 3.** Identification of putative nematocyst-specific proteins in the giant sea anemones. **(A)** Heatmap shows the variation in the number of putative nematocyst-specific proteins in seven species of sea anemones and six reference cnidarians. Three groups of proteins with varying levels of evolutionary conservation are highlighted. **(B)** Selection of several examples where variation is especially evident or might indicate functional significance for symbiotic relations between sea anemones and anemonefish. **(C)** Phylogenetic tree of D-galactoside/L-rhamnose-binding SUEL lectin proteins that are expanded in the anthozoan lineage. ML tree, LG-GAMMA substitution model. The grey rectangles indicate the proteins from Okinawan sea anemones. The color code of branches and species names is similar to that in Fig. 2 B–F. Proteins from the reference species have the following identifiers: xe2 - *Xenia*, adi - *Acropora digitifera*, apa - *Exaiptasia (Aiptasia)*, pau - *Porites*, nve - *Nematostella*, hvi - *Hydra viridis*, hma - *Hydra magnipapillata*, abs - *Aurelia* (Baltic sea), tri - *Tripedalia*, mvi - *Morbakka virulenta*.

by MAFFT, areas of low complexity in the alignment were trimmed by TrimAL, and the gene tree was reconstructed by using RAxML with the LG + GAMMA substitution model (see Fig. 3C).

## RESULTS

### Reference transcriptomes of Okinawan sea anemones

We sequenced the transcriptomes of seven sea anemone species from Okinawa that establish symbiotic relations



with anemonefish (Fig. 1; Tables 1, 2). These were *E. quadricolor*, *H. aurora*, *H. crispa*, *H. magnifica*, *S. gigantea*, *S. haddoni* and *S. mertensii*. Since it was difficult to estimate the amount of raw sequence data required to obtain high-quality transcriptomes, two successive rounds of sequencing were performed (see Table 2). In the first round, one representative individual per species was selected for high coverage sequencing, and ca. 138–186 million reads were sequenced per species (which corresponds to 21–28 Gbp of 2x150bp NovaSeq6000 reads). For the second round with lower coverage, we selected two additional individuals of *E. quadricolor*, and one additional individual each of *H. aurora*, *H. crispa*, *H. magnifica*, and *S. gigantea*. We were not able to get other samples of *S. haddoni* or *S. mertensii* from our sampling location. In the second sequencing round, data was reduced to 34–46 million reads per specimen (5.2–8.9 Gbp per library). The first round of sequencing provided high-quality reference transcriptomes for seven anemone species, and additional data from the second round were used to identify markers that can resolve not only inter-specific but also intra-specific diversity (Tables 1, 2).

Giant sea anemones harbor symbiotic dinoflagellates in their gastrodermis and in the endodermal epithelial cells of their tentacles. Therefore, our initial transcriptome assemblies contained a mixture of cnidarian and algal transcripts, which should be separated to obtain pure sequence data from the giant sea anemones (Fig. 1I). Since no prior knowledge about the species identity of the dinoflagellates that inhabit giant sea anemones on Okinawa was available, we used BLASTN to screen our data sets against the transcriptomes of *Symbiodinium* (former clade A), *Breviolum* (former clade B), and *Cladocopium* (former clade C), which are the most frequent symbionts of corals and other marine invertebrates (LaJeunesse et al., 2018; Pochon and LaJeunesse, 2021). It is known that dinoflagellates and cnidarians differ drastically in their GC content (Shoguchi et al., 2013). While the GC content in Anthozoa is ~40%, the GC content in Symbiodiniaceae is ~55%. This feature, therefore, can be used as an additional criterion to separate transcripts belonging to sea anemones from those belonging to Symbiodiniaceae. Combined filtering based on BLASTN search (no hit to algae) and GC content (below 60%) allowed us to effectively separate the transcriptomes of the sea anemones and their symbionts (Fig. 1I). Based on the above-mentioned screening parameters, 70.5% of all assembled transcripts belonged to sea anemones, and only 29.5% were derived from *Cladocopium* (Fig. 1I). This ratio indicates that tentacles are the optimal tissue source for transcriptomic analysis, since the number of cnidarian transcripts was ~2-fold higher than that of algal ones. Thus, by using tentacles as an RNA source, the price overhead for unavoidable sequencing of algal transcripts remains within the acceptable range.

Since, in addition to the transcriptomes of the giant sea anemones, we obtained seven high-quality transcriptomes of their symbionts (see Supplementary Table S1 for BUSCO values), it was interesting to estimate the genetic diversity of the symbionts derived from different hosts. In order to address this question, we used the LSU rDNA reference data set (see Supplementary Text S1) that includes Symbiodiniaceae and other subclades as query sequences (LaJeunesse

et al., 2018). Phylogenetic analysis of LSU sequences showed that symbiotic algae from all seven species of giant sea anemones belong to *Cladocopium* (former Clade C) (Fig. 1J). There was little previous information on the *Durusdinium* (former Clade D) that inhabit Okinawan giant sea anemones (Santos et al., 2003). In contrast, there were several reports from Japan and outside of Japan about giant sea anemones' symbiotic relationship with *Cladocopium* in previous research (Rodriguez-Lanetty et al., 2003; Ono et al., 2010; Hill et al., 2012; Behrouz et al., 2021).

After the removal of *Cladocopium* (former clade C) transcripts, the combined values for complete (C) and fragmented (F) BUSCOs in all transcriptomes were above 94.2% and 87.2% for high- and low-coverage datasets, respectively (Fig. 1K, Table 2). N50 values of the transcriptomes ranged from 1.8–2.2 Kb, and the longest assembled transcript length ranged between 32–64 Kb (Table 2). Thus, by using on average 162 million reads derived from tentacle tissue, we were able to obtain high-quality reference transcriptomes suitable for phylogenetic reconstructions. Detailed quality statistics for the cleaned transcriptomic assemblies (without algae sequences) are shown in Table 2.

### Prediction and annotation of protein sequences

After the removal of algae-derived sequences, the 'cleaned' transcriptomes were used to predict peptides by using the combination of Transdecoder and EstScan (Haas et al., 2013; Lottaz et al., 2003). Only the transcripts with open reading frames (ORFs) encoding  $\geq 70$  amino acids were taken into consideration for further analyses. For each transcript, only the best ORF was selected. Redundancy in the sets of predicted proteins was reduced ~40% by using CD-HIT with a 95% similarity cut-off (see Supplementary Table S2). The resulting protein sets were annotated by BLASTP search with E-value  $1e-5$  against the human proteome, HMMER search against Pfam-33.1 release (Potter et al., 2018; Mistry et al., 2021), and InterProScan pipeline (Jones et al., 2014). BUSCO values for the final non-redundant proteomes of seven anemone species are shown in Table 2. Protein sequences for seven anemone species in FASTA format are available at the OIST BLAST server (<http://compagen.unit.oist.jp/aurelia/datasets.html>).

### Phylogeny and genetic variability among giant sea anemones

The major goal of our transcriptomic analyses was to clarify phylogenetic relationships of the anemonefish-hosting sea anemones and to estimate the levels of divergence among them. In addition, we sought to identify a set of markers that would be optimal for population genetics studies in the future.

In order to reconstruct the giant sea anemone phylogeny, we used protein sets from 14 cnidarian and six bilaterian species as references (see Supplementary Table S3). All bilaterian proteomes were derived from genomic data. For cnidarian species, we utilized data sets of either genomes or transcriptomes. In addition, two sets of proteins from two versions of *Hydra magnipapillata* genome annotations were used to assess the level of variability that might be caused by differences in protein model predictions within the same species. For the same reason, we selected two

protein sets of *Nemopilema nomurai* (Nomura jellyfish) – one derived from the genome (Kim et al., 2019) and another one from the transcriptome (Khalturin et al., 2019). Our dataset includes several reference species for which genetic distances and divergence times are well known. Such information is available for *Hydra viridis* and *H. magnipapillata* within Hydrozoa (Chapman et al., 2010), for *Aurelia* within Scyphozoa (Khalturin et al., 2019), and for the corals *Acropora* spp (Shinzato et al., 2011, 2021). Divergence times within Bilateria have reliable estimations based on fossil records (Benton et al., 2015).

Orthologous genes were identified by using OrthoFinder (see Materials and Methods) with 737,660 proteins (90.5% of total number) assigned to 46,993 orthologous groups (orthogroups). There were 2405 orthogroups with all species present. Six of these consisted entirely of single-copy genes. 354 orthogroups were represented by single-copy genes in a minimum 71% of species. Phylogenetic relationships of the sea anemones and their placement within cnidaria were assessed based on a concatenated alignment of these 354 genes that were present in all 27 species. The phylogenetic tree was reconstructed by a maximum-likelihood method in RAxML using the LG + GAMMA substitution model (see Materials and Methods for details).

As shown in Fig. 2A, the general placement of the giant sea anemones (*H. crispera*, *H. aurora*, *E. quadricolor*, *H. magnifica*, *S. mertensii*, *S. gigantean*, and *S. haddoni*) within Cnidaria was consistent with that in previous reports (Titus et al., 2019; Nguyen et al., 2020). As expected from the combined usage of hundreds of genes, all branches have maximum bootstrap support, which had not been achieved with the limited number of markers used previously. *Exaiptasia* (referred to as *Aiptasia* in Baumgarten et al., 2015) is the most closely related species among the outgroups used in our phylogeny reconstruction.

Giant sea anemones make up three distinct groups corresponding to three genera: *Entacmaea*, *Heteractis*, and *Stichodactyla*, with the genus *Entacmaea* branching off prior to the divergence of the *Heteractis* and *Stichodactyla* genera. Our results support previous observations that *H. magnifica* actually belongs to the genus *Stichodactyla* (Titus et al., 2019; Nguyen et al., 2020). Anemones belonging to each of these three genera are quite similar genetically. The range of divergence among them is less than that within the extreme representatives of the genus *Hydra*. At the same time, if compared with bilaterians, the genetic distance between *Entacmaea* and *Stichodactyla* is equivalent to that between humans and chickens. Although differences among the genera are clear, genetic diversity within genera is low (Fig. 2A, B, D, F). When compared to the variability observed within the *Aurelia* species complex, it is more in line with differences between strains than between species. Relationships within the genus *Stichodactyla* are formally resolved with our set of markers, but the level of variability is extremely low, making it uncertain whether *S. mertensii* occupies the basal position within the genus. Thus, several additional samples of *S. mertensii* would be of high importance.

In order to explore the relationships of Okinawan giant sea anemones in more detail, we included data of the low coverage transcriptomes so that we had several (e.g., two or three) individuals per species to compare. We also excluded

distant outgroups (all deuterostomes, Hydrozoa, Scyphozoa, and Cubozoa) and used only *Nematostella* and *Exaiptasia* protein sets to root the phylogenetic trees. As a result, the number of markers present in all 15 datasets (which correspond to nine species) increased to 6889 genes. From them, we selected 1365 orthogroups with a minimum of 86.7% of species having single-copy genes in any orthogroup. As shown in Fig. 2B, the ML tree based on concatenated alignments of 1365 proteins fully supports the topology shown in Fig. 2A obtained with 354 markers. This large set of 1365 markers allows distinguishing between individuals in all species except *S. mertensii* and *S. haddoni*, for which only one individual was available for sequencing so far. However, with this dataset, the bootstrap support for two clades within *Entacmaea* was 85%. Of note, the topology within the *Stichodactyla* clade was identical to that with the 354 markers set, with *S. mertensii* having the basal position within the genus *Stichodactyla*.

The selection of markers for phylogenetic reconstruction may influence the topology of the resulting trees considerably. Taking this into consideration, we reconstructed the phylogeny of the sea anemones by using the controlled set of markers that all belong to basic universal single-copy orthologs (BUSCO genes) (Seppey et al., 2019). These genes, by definition, are convenient markers for phylogenetic reconstruction since, in the majority of metazoan species, they are present in just one copy. In most cases, this feature removes the problem of marker selection from several alternative paralogs. In the proteomes of *Nematostella*, *Exaiptasia*, and seven species of sea anemones, we identified 451 orthologous groups (OGs) that correspond to the proteins from the BUSCO metazoa\_odb09 database. Next, we sought to identify the subset of proteins with the strongest phylogenetic signal. For each BUSCO orthogroup, individual ML gene trees were built by RAxML with the LG + GAMMA substitution model. The mean branch lengths were calculated for each tree and 226 genes with the mean branch lengths larger or equal to the median distribution were selected (shown as red dots in Fig. 2C). We thereby selected the subset of BUSCO proteins with an elevated rate of evolution in the sea anemones. After that, we constructed a supermatrix by concatenation of 226 proteins from 15 datasets. As shown in Fig. 2D, the general topology of the tree reconstructed with these 226 markers did not change except for the position of *S. mertensii*, which now grouped together with *H. magnifica*. At the same time, the bootstrap support within *Entacmaea* increased from 85% to 96%.

The analysis of individual gene trees of the BUSCO set showed that the cases in which the protein sequences were identical in two or more species were relatively frequent. Thus, from 451 BUSCO proteins, we selected a subset where all sequences were always different among the species as well as among all the individuals within each species. In total, as shown in Fig. 2E, this subset of BUSCO proteins with the highest degree of variability comprised 111 genes. The ML tree based on this subset is shown in Fig. 2F. Bootstrap support was maximal for all branches except those in *E. quadricolor*, and the topology was identical to the tree obtained previously with 226 markers (Fig. 2D). For population genetics studies, this may be the most informative and convenient set of markers since it allows the capture

of intra-specific and inter-specific variation simultaneously. The list of 111 genes used in Fig. 2E, F with their annotations is shown in Supplementary Table S4.

### Nematocyst-specific genes in the giant sea anemones

Transcriptomic data from the tentacles potentially allow us to compare the repertoires of genes expressed in the nematocytes of the different sea anemone species. Hence, the molecular basis of interactions between anemonefish and sea anemones can be better understood. However, this type of analysis is only possible if nematocyte development occurs in the tentacles and not in the body column, as in several model cnidarians such as *Hydra*, *Clytia* and *Aurelia* (Chapman et al., 2010). There is evidence that in the sea anemones, the areas of nematocyte development differ from those in the representatives of Medusozoa. Indeed, in *Nematostella*, in strong contrast to that in Hydrozoa and Scyphozoa, three minicollagen genes (*NvNcol-1*, *NvNcol-3*, and *NvNcol-4*) known to be the major structural constituents of the nematocysts were shown to be expressed predominantly in the tentacles (Zenkert et al., 2011). Similar indication that the nematoblasts originate in the tentacles was also obtained by the transcriptomic analysis of *NvNcol-3* reporter line in *Nematostella* (Sunagar et al., 2018). Given these results, it is reasonable to expect that nematocyte development also occurs in the tentacles in the giant sea anemones, as in *Nematostella*.

In order to verify whether the nematocyte genes are indeed expressed in the tentacles, we utilized the set of 90 proteins from *Nematostella*, the most widely studied sea anemone model organism. The single-cell transcriptomic approach confirmed that these proteins were cnidocyte-specific (Sebé-Pedrós et al., 2018). We screened for the presence of these proteins in the tentacle-derived transcriptomes of seven species of Okinawan sea anemones. Out of 90 *Nematostella* proteins, 68–70 proteins were present in all species (see BLAST results in Supplementary Table S5). It is important to note that from the 90 nematocyte-specific proteins reported in Sebé-Pedrós et al. (2018), only 69 could be identified among the filtered protein models of the *Nematostella* genome by BLASTP with *e*-value  $1e-5$  (Putnam et al., 2007). Thus, the actual detection rate of cnidocyte-specific genes in our case was not ~76.6% (69/90) but close to 100%. Our result demonstrates that active nematocyte proliferation occurs in the tentacles of giant sea anemones. Therefore, our transcriptomes representatively capture the repertoire of genes associated with nematocyte development.

Currently, there are no data comprehensively describing a proteome of anthozoan nematocysts. Transcriptomic approaches were used to characterize cnidocyte genes in *Nematostella* (Sebé-Pedrós et al., 2018; Sunagar et al., 2018), but that type of data does not enable one to separate structural proteins that constitute the nematocyst wall and its contents from the proteins localized elsewhere in the cytoplasm of this cell type. The only comprehensive proteome of nematocysts known to date is from *Hydra*, a more distant species, where 410 proteins constituting the capsule wall or localized in the capsule lumen were identified by a proteomic approach followed by MALDI/TOF (Balasubramanian et al., 2012). Despite the large evolutionary distance between

*Hydra* and Anthozoa, we decided to use this comprehensive proteomic dataset to identify putative nematocyst-specific proteins in seven sea anemone species.

Orthologs of these *Hydra* proteins were identified in 22 cnidarian species by OrthoFinder (Emms and Kelly, 2019) (see Supplementary Table S6). Since functional characteristics of the stinging cells are dependent on the repertoire of capsule proteins, our main interest was to identify differences in nematocyst composition among the giant sea anemones as well as between sea anemones and other representatives of Hexacorallia such as *Nematostella*, *Exaiptasia*, *Acropora*, and *Porites*. This comparison should be considered as a preliminary analysis because a homology-based search with a distant evolutionary species like *Hydra* may miss important anemone-specific genes. Moreover, not all types of nematocytes in anemones are localized to the tentacles (England, 1991). For example, nematocysts of a particular type, such as mastigopore nematocysts, are usually found only in mesenterial filaments. Our tentacle transcriptomes, therefore, might have missed the genes associated with the capsule types that are not present in the tentacles. However, this is the first step to identifying genes that might be responsible for the differences in stinging capacity among sea anemones and, therefore, important for the interactions between anemonefishes and sea anemones. From the set of 410 *Hydra* proteins, we were able to identify 197 orthologs in at least one species other than *Hydra*. Variation in a number of putative nematocyst-specific proteins in 13 cnidarian species is shown as a heatmap in Fig. 3A. Detailed information about the phylogenetic distribution of 20 genes which we found the most interesting, is given in Fig. 3B, and the complete dataset is represented in Supplementary Table S6 (see also Supplementary\_Text\_S1).

As shown in the cladogram at the bottom of Fig. 3A, a grouping of the species based on the repertoires of their nematocyst-specific proteins perfectly recapitulates the phylogenetic relationships obtained by maximum likelihood analysis described previously in Fig. 2A, B, D, F. *Nematostella* and *Exaiptasia* (*Aiptasia*) form outgroups, and *Entacmaea* is the basal species among the sea anemones. Both *H. crispata* and *H. aurora* form a distinct clade, and all representatives of *Stichodactyla* as well as *H. magnifica* group together.

Our analysis shows that among putative nematocyst-specific genes, there are several groups with varying degrees of evolutionary conservation. First, there is a group of 27 genes that are highly conserved and present in all cnidarians in large copy numbers (group I in Fig. 3A). This group contains modular proteins with repeated vWFA, EGF, thrombospondin domains, zinc-dependent metalloproteinases and peptidases. One representative of the minicollagen family, *minicollagen-3*, also belongs to this category. Second, a large group contains genes that are present only in *H. magnipapillata* and *H. viridis* and, thereby, represent taxonomically restricted genes at the genus level (group II in Fig. 3A). Several genes from this category are also sporadically present in Medusozoa (Hydrozoa, Scyphozoa, and Cubozoa) but absent in Anthozoa. Except for ATPases, one Zn-finger transcription factor, several minicollagens, and toxin precursors, these are mostly transmembrane and secreted proteins lacking any functionally characterized

domains. The most interesting category (group III in Fig. 3A) includes proteins that are present in the majority of species but with variation in the number of orthologs among the species.

The number of putative nematocyst-specific genes is relatively uniform across all giant sea anemone species. For example, as shown in Fig. 3B, there are from five to nine copies of cnidarian proline-rich protein-1 (CPP-1), which is one of the structural components of the nematocyst wall. Minicollagens are present in large numbers without any clear trend within sea anemones, except for the absence of several types of Hydrozoa-specific minicollagens in Anthozoa. This trend has been observed earlier (David et al., 2008) and is fully supported by our data. The presence of only five copies of *minicollagen-21* in *E. quadricolor*, compared to 8–10 copies in *Stichodactyla* and *Heteractis*, is also interesting. Such a difference may have a functional significance, since the mechanical strength of nematocysts, and hence their stinging power, depends on the repertoire of minicollagens that reinforce the capsule structure.

Interestingly, there is considerable variation among species in the number of proteins with MACPF (membrane attack complex/perforin), EGF-like, and EGF\_Ca domains. The same trend is true for Zinc carboxypeptidase (Peptidase M14) and two types of alpha carbonic anhydrase proteins. Potentially interesting is the loss of Alpha-L-arabinofuranosidase in all representatives of *Stichodactyla*, *H. magnifica*, and *E. quadricolor*, while one and two copies of this enzyme involved in the hydrolysis of polysaccharides have been retained in *H. crispa* and *H. aurora*, respectively. A similar case is the loss of Lipase-3 in all representatives of *Stichodactyla* as well as in *H. magnifica*. This observation may indicate differences in lipid metabolism within the group. Interestingly, the same type of lipase also has been lost in all representatives of Cubozoa.

D-galactoside/L-rhamnose binding SUEL lectins demonstrate another trend—expanding a gene family in the anthozoan lineage. While in Hydrozoa, Scyphozoa, and Cubozoa, this is a single-copy gene, in Anthozoa, several paralogs with slightly different protein structures are present. As shown in Fig. 3C, each species of giant sea anemones, except for *E. quadricolor*, has at least three copies of the D-galactoside/L-rhamnose binding SUEL lectin gene. These genes encode secreted proteins with two adjacent D-galactoside/L-rhamnose binding domains followed by a relatively long C-terminal region without recognizable homology to any other proteins. In *H. magnifica*, in addition to proteins shown in the phylogenetic tree (Fig. 3C), there are three additional proteins that show high similarity to the C-terminal part but do not contain D-galactoside/L-rhamnose binding domains. In the giant sea anemones, we do not yet have information about the location of these genes in the genome, but in *Acropora digitifera*, all three D-galactoside/L-rhamnose binding lectin genes are clustered together in scaffold 61, which might indicate their origin as a result of tandem gene duplication (Shinzato et al., 2011).

## DISCUSSION

Giant sea anemones are species with a broad geographical distribution that exhibit a wide variety of morphologies, rendering species identification very difficult (Dunn,

1981). Therefore, the taxonomy of these species is still confusing. To date, species-level differences in the taxonomy of giant sea anemones are still not fully resolved since, in recent analyses, no clear delineation of the various species was obtained based on several concatenated markers such as CO3, 12S, 16S, 18S, and 28S (Titus et al., 2019; Nguyen et al., 2020).

In this paper, we performed transcriptomic analyses of seven giant anemone species native to Okinawa, Japan, to obtain key information on these important partners of the mutualistic symbiosis with anemonefish and to gain insight into their genetic makeup and relationships. As a result, we obtained high-quality transcriptomes for all seven species (over 90% complete based on BUSCO values), which can serve as references for future research on giant sea anemones.

Phylogenetic trees based on several sets of protein-coding genes from the transcriptomes (ranging from 111 to 1365 concatenated markers) showed that these giant sea anemones cluster in three groups: the first divergence happened between the *Entacmaea* group and a group containing *Heteractis* and *Stichodactyla* species. Furthermore, we confirmed that *H. magnifica* should be included in the *Stichodactyla* clade and not in the genus *Heteractis*. This confirmed previous reports and therefore suggests that *H. magnifica* is indeed more closely related to *Stichodactyla* than to *Heteractis* (Titus et al., 2019; Nguyen et al., 2020). This result corresponds well to the observation that *Amphiprion ocellaris* is able to establish mutualistic relationships with both *S. gigantea* and *H. magnifica* (Miyagawa-Kohshima et al., 2014). The same is true for *Amphiprion polymnus*, which uses the anemones of the *Stichodactyla* group as well as *H. magnifica* as its host. Thus, clownfish are sometimes better at identifying the genus of anemones based on their morphology than humans.

Due to sampling limitations (only one sample was available), it is still difficult to identify the place of *S. mertensii* within the respective genus unambiguously. While two sets of markers favor the most basal position of *S. mertensii* within *Stichodactyla* with the maximum support (Fig. 2A, B), one alternative topology is highly supported with two subsets of BUSCO markers (Fig. 2D, F). Thus, an increased number of samples from multiple locations around Okinawa Island is needed for *S. mertensii*. This also applies to *S. haddoni*, for which a single individual was also sampled. Morphological comparisons with type specimens and specimens from type localities may allow us in the future to better define the species composition within the *Heteractis*, *Stichodactyla*, and *Entacmaea* genera, as well as to detect eventual cryptic species previously unrecognized. The main difficulty is that the ethanol-preserved type specimens are usually not suitable for transcriptomic analysis due to RNA degradation. That greatly hinders direct comparison of the type specimens with newly collected samples at the molecular level.

The high BUSCO values, as well as the total number of non-redundant proteins, clearly indicate that the tentacle-derived transcriptomes provide a surprisingly complete representation of gene sets typical of Anthozoa. Such a high proportion of recovered genes is rather unexpected since only a small subset of cell types is typically present in the

tentacles of other cnidarians. For example, all types of gland cells, as well as all stages of developing nematocytes, would be missing in the tentacles of *Hydra* or *Aurelia*. Hence, a transcriptome derived from the tentacles of the aforementioned species would be missing thousands of genes. Interestingly, we do not observe this tendency in the tentacle transcriptomes of *Entacmaea*, *Heteractis*, and *Stichodactyla*. In contrast, about 200 genes that are known to be involved in nematocyst formation were detected, indicating that stinging cells also develop within the tentacles and not only in the body column. This unexpected feature made it possible to compare the repertoires of nematocyst-specific genes among the representatives of *Entacmaea*, *Heteractis*, and *Stichodactyla*, which differ considerably in their stinging capacity.

We used a set of 410 nematocyst-specific genes from *Hydra* (Balasubramanian et al., 2012) as a reference and made three interesting observations: (i) First, we observed the grouping of the species based on the repertoires of their nematocyst-specific proteins, which perfectly recapitulates the phylogenetic relationships that were obtained through transcriptomic data (bottom of Fig. 3A). In particular, we found that the set of putative *H. magnifica* nematocyte proteins is more closely related to the *Stichodactyla* set of proteins than to those from *Heteractis*. This solves the paradox previously highlighted by Nedosyko et al. (2014), who observed the strong toxicity of *H. magnifica* in contrast to other members of this genus. Our data clearly suggest that *H. magnifica* is indeed a *Stichodactyla* and shares similar toxicity with these species. However, this conclusion needs to be confirmed by more direct functional experiments comparing the toxicity of these various anemones. This observation also suggests that the divergence of nematocyte genes among the three groups of giant anemones can be explained, at least in part, by phylogenetic divergence. (ii) Second, we noticed that the number of putative nematocyst-specific genes is relatively uniform across the species of giant sea anemones. Therefore, differences in toxicity measured between various sea anemone species (Nedosyko et al., 2014) and known to be important for the establishment and maintenance of anemonefish–anemone symbiosis is likely related to variations in a few genes and/or changes in their expression levels, not in any huge variations in the copy number of those genes (Marcionetti et al., 2019). Analysis of the differences in the expression levels of nematocyte-specific genes among the species is, therefore, the next logical step for future research. (iii) We, however, observed some interesting differences in gene copy numbers between giant anemones and other anthozoans, as well as among the three types of giant anemones. Among those, we think it is worth noting the case of alpha-L-arabinofuranosidase, an enzyme important for the hydrolysis of polysaccharides and, in particular, cellulose (Numan and Bhosle, 2006). This gene is present in *H. crispata* and *H. aurora*, but not in the other giant sea anemones (*Stichodactyla*, *H. magnifica*, and *E. quadricolor*). Similarly, D-galactoside/L-rhamnose-binding SUEL lectins are present in three copies in all giant sea anemones except *E. quadricolor*. These observations are particularly interesting in the context of the symbiosis with the anemonefish and the Symbiodiniaceae. SUEL lectins have been implicated in the coral-zooxanthellae symbiosis

mechanism, suggesting that these genes may be important for the interactions between sea anemones and *Cladocopium* (Zhou et al., 2017). It has also been shown previously that specific sugars such as N-acetylneuraminic acid are able to stimulate nematocyte discharge (Ozacmak et al., 2001; Anderson and Bouchard, 2009) and are present in low levels in clownfish skin (Abdullah and Saad, 2015). In addition, recent genomic analysis has found sugar genes (namely *versican core protein* and *protein OGLcNAse*, whose functions are associated with N-acetylated sugars) under positive selection at the base of the anemonefish radiation (Marcionetti et al., 2019). It is not yet known whether sugar biology is also involved in the mechanisms that allow tentacles to avoid nematocyte discharge upon contact with neighboring ones, but it is clear that more information is needed about the role of sugars in giant sea anemones and their symbiosis with anemonefishes.

Taken together, transcriptomic analyses of seven species of giant sea anemones have improved our understanding of their taxonomic relationships within the group. We obtained comprehensive transcriptomic data for each of the species, allowing us to perform the first phylotranscriptomic analysis with sufficient resolution at all evolutionary scales with high reliability. Our results indicate that the first divergence happened between the *Entacmaea* group and a clade containing *Heteractis* and *Stichodactyla* species. We hope that our data will be useful as a platform for future experiments on population structure and symbiotic relationships. High-quality reference transcriptomes will also serve as an important dataset for future giant sea anemone genome projects. Our data open up promising avenues of research to better understand this fascinating example of mutualistic symbiosis.

## ACKNOWLEDGMENTS

We thank Eiichi Shoguchi for help with analyzing and interpreting sequence data from Symbiodiniaceae, Polina Piliava and Michael Izumiyama for help with sampling, and Lilian Carlu for culturing sea anemones. Gratitude is due to Marcela Sarrias-Herrera, Yann Guerardel, Michael Izumiyama, and Natacha Roux for critical reading, and Nicolas Salamin and Benjamin Titus for helpful discussion. We thank OIST for funding and in particular the OIST Kicks program for funding. We are grateful for the help and support provided by the Scientific Computing and Data Analysis section of Research Support Division at OIST. The authors would also like to thank OIST SQC for sequencing services.

## COMPETING INTERESTS

The authors declare no competing interest.

## AUTHOR CONTRIBUTIONS

RK and KK performed experiments and analyzed the data. MT provided samples and expert assistance on giant sea anemone. SM provided technical assistance on RNA extraction. VL, NS, and KK designed the study. RK, VL, and KK wrote the manuscript.

## SUPPLEMENTARY MATERIALS

Supplementary materials for this article are available online. (URL: <https://doi.org/10.2108/zs210111>)

**Supplementary Text S1.** Symbiodiniaceae LSU rDNA sequences.

**Supplementary Text S2.** Description of sea anemones.

**Supplementary Figure S1.** Enlarged version of a heatmap in

Fig. 5A.

**Supplementary Figure S2.** Images of sea anemones used for low coverage transcriptome sequencing. **(A)** *Entacmaea quadricolor*, **(B)** *Entacmaea quadricolor*, **(C)** *Stichodactyla gigantea*, **(D)** *Heteractis crispa*, **(E)** *Heteractis aurora*, and **(F)** *Heteractis magnifica*.

**Supplementary Table S1.** Algae transcriptome assembly quality.

**Supplementary Table S2.** Peptide prediction for cleaned sequences.

**Supplementary Table S3.** Information about the data sources.

**Supplementary Table S4.** List of 111 BUSCO genes used for phylogeny reconstruction (Fig. 4E) with their Gene Ontology (GO) and InterProScan annotations.

**Supplementary Table S5.** Orthologs of nematocyte-specific genes of *Nematostella vectensis* in sea anemones from Okinawa.

**Supplementary Table S6.** Number of putative nematocyst-specific genes in the sea anemones and other cnidarian species (homology search based on *Hydra* dataset).

## REFERENCES

- Abdullah NS, Saad S (2015) Rapid detection of N-Acetylneuraminic acid from false clownfish using HPLC-FLD for symbiosis to host sea anemone. *Asian J Appl Sci* 3: 858–864
- Alan Verde E, Cleveland A, Lee RW (2015) Nutritional exchange in a tropical tripartite symbiosis II: direct evidence for the transfer of nutrients from host anemone and zooxanthellae to anemonefish. *Mar Biol* 162: 2409–2429
- Anderson PAV, Bouchard C (2009) The regulation of cnidocyte discharge. *Toxicon* 54: 1046–1053
- Aprill A (2020) The role of symbioses in the adaptation and stress responses of marine organisms. *Ann Rev Mar Sci* 12: 291–314
- Balasubramanian PG, Beckmann A, Warnken U, Schnölzer M, Schüler A, Bornberg-Bauer E, et al. (2012) Proteome of *Hydra* nematocyst. *J Biol Chem* 287: 9672–9681
- Benton M, Donoghue PCJ, Asher RJ, Friedman M, Near TJ, Vinther J (2015) Constraints on the timescale of animal evolutionary history. *Palaeontol Electron* 18: 1–107
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120
- Burke da Silva K, Nedosyko A (2016) Sea anemones and anemonefish: A match made in heaven. In: “The Cnidaria, Past, Present and Future: The World of Medusa and Her Sisters” Ed by S Goffredo, Z Dubinsky, Springer International Publishing, pp 425–438
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25: 1972–1973
- Chapman JA, Kirkness EF, Simakov O, et al. (2010) The dynamic genome of *Hydra*. *Nature* 464: 592–596
- Cleveland A, Verde EA, Lee RW (2011) Nutritional exchange in a tropical tripartite symbiosis: direct evidence for the transfer of nutrients from anemonefish to host anemone and zooxanthellae. *Mar Biol* 158: 589–602
- David CN, Özbek S, Adamczyk P, Meier S, Pauly B, Chapman J, et al. (2008) Evolution of complex structures: minicollagens shape the cnidarian nematocyst. *Trends Genet* 24: 431–438
- Dunn DF (1981) The clownfish sea anemones: Stichodactylidae (Coelenterata: Actiniaria) and other sea anemones symbiotic with pomacentrid fishes. *Trans Am Philos Soc* 71: 3–115
- Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20: 238
- England KW (1991) Nematocysts of sea anemones (Actiniaria, Ceriantharia and Corallimorpharia: Cnidaria): nomenclature. *Hydrobiologia* 216/217: 691–697
- Enright AJ, Van Dongen S, Ouzounis CA (2002) An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30: 1575–1584
- Fu L, Niu B, Zhengwei Z, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28: 3150–3152
- Haas BJ, Papanicolaou A, Yassour M, et al. (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc* 8: 1494–1512
- Hayashi K, Tachihara K, Reimer JD (2018) Patterns of coexistence of six anemonefish species around subtropical Okinawa-jima Island, Japan. *Coral Reefs* 37: 1027–1038
- Hill R, Scott A (2012) The influence of irradiance on the severity of thermal bleaching in sea anemones that host anemonefish. *Coral Reefs* 31: 273–284
- Jones P, Binns D, Chang H-Y, et al. (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30: 1236–1240
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30: 772–80
- Khalturin K, Shinzato C, Khalturina M, et al. (2019) Medusozoan genomes inform the evolution of the jellyfish body plan. *Nature Ecology & Evolution* 3: 811–822
- Kim H-M, Weber JA, Lee N, et al. (2019) The genome of the giant Nomura’s jellyfish sheds light on the early evolution of active predation. *BMC Biology* 17: 28
- Kück P, Longo GC (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. *Frontiers in Zoology* 11: 81
- LaJeunesse TC, Parkinson JE, Gabrielson PW, et al. (2018) Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology* 28: 2570–2580.e6
- Litsios G, Kostikova A, Salamin N (2014) Host specialist clownfishes are environmental niche generalists. *Proceedings of the Royal Society B: Biological Sciences* 281: 20133220
- Litsios G, Sims CA, Wüest RO, et al. (2012) Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology* 12: 212
- Lottaz C, Iseli C, Jongeneel CV, Bucher P (2003) Modeling sequencing errors by combining hidden Markov models. *Bioinformatics* 19 Suppl 2: 103–112
- Marcionetti A, Rossier V, Roux N, et al. (2019) Insights into the genomics of clownfish adaptive radiation: Genetic Basis of the Mutualism with Sea Anemones. *Genome Biol Evol* 11: 869–882
- Mariscal RN, Fautin DG, Allen GR (1993) Field guide to anemonefishes and their host sea anemones. *Copeia* 1993: 899
- Mistry J, Chuguransky S, Williams L, et al. (2021) Pfam: The protein families database in 2021. *Nucleic Acids Research* 49: D412–D419
- Miyagawa-Kohshima K, Odoriba S, Okabe D, et al. (2014) Embryonic learning of chemical cues via the parents’ host in anemonefish (*Amphiprion ocellaris*). *Journal of Experimental Marine Biology and Ecology* 457: 160–172
- Nedosyko AM, Young JE, Edwards JW, Burke da Silva K (2014) Searching for a toxic key to unlock the mystery of anemonefish and anemone symbiosis. *PLoS ONE* 9: e98449
- Nguyen H-TT, Dang BT, Glenner H, Geffen AJ (2020) Cophylogenetic analysis of the relationship between anemonefish *Amphiprion* (Perciformes: Pomacentridae) and their symbiotic host anemones (Anthozoa: Actiniaria). *Marine Biology Research* 16: 117–133
- Numan MT, Bhosle NB (2006) Alpha-L-arabinofuranosidases: the potential applications in biotechnology. *J Ind Microbiol Biotechnol* 33: 247–260
- Ozacmak VH, Thorington GU, Fletcher WH, Hessinger DA (2001) N-acetylneuraminic acid (NANA) stimulates in situ cyclic AMP

- production in tentacles of sea anemone (*Aiptasia pallida*): possible role in chemosensitization of nematocyst discharge. *Journal of Experimental Biology* 204: 2011–2020
- Pochon X, LaJeunesse TC (2021) *Miliolidium* n. gen, a new Symbiodiniacean genus whose members associate with soritid foraminifera or are free-living. *Journal of Eukaryotic Microbiology* 68: e12856
- Porat D, Chadwick-Furman NE (2004) Effects of anemonefish on giant sea anemones: expansion behavior, growth, and survival. *Hydrobiologia* 530: 513–520
- Potter SC, Luciani A, Eddy SR, et al. (2018) HMMER web server: 2018 update. *Nucleic Acids Research* 46: W200–W204
- Putnam NH, Srivastava M, Hellsten U, et al. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. *Science* 317: 86–94
- Rodriguez-Lanetty M, Chang S-J, Song J-I (2003) Specificity of two temperate dinoflagellate–anthozoan associations from the north-western Pacific Ocean. *Marine Biology* 143: 1193–1199
- Roux N, Salis P, Lee S-H, et al. (2020) Anemonefish, a model for Eco-Evo-Devo. *EvoDevo* 11: 20
- Seppely M, Manni M, Zdobnov EM (2019) BUSCO: Assessing genome assembly and annotation completeness. In: Kollmar M (ed) *gene prediction*. Springer New York, New York, NY, pp 227–245
- Sebé-Pedrós A, Saudemont B, Chomsky E, et al. (2018) Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell* 173: 1520–1534.e20
- Shinzato C, Khalturin K, Inoue J, Zayasu Y, Kanda M, Kawamitsu M, et al. (2021) Eighteen coral genomes reveal the evolutionary origin of acropora strategies to accommodate environmental changes. *Mol Biol Evol.* 38(1): 16–30
- Shinzato C, Shoguchi E, Kawashima T, et al. (2011) Using the *Acropora digitifera* genome to understand coral responses to environmental change. *Nature* 476: 320–323
- Shoguchi E, Shinzato C, Kawashima T, et al. (2013) Draft assembly of the symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Current Biology* 23: 1399–1408
- Sunagar K, Columbus-Shenkar YY, Fridrich A, et al. (2018) Cell type-specific expression profiling unravels the development and evolution of stinging cells in sea anemone. *BMC Biology* 16: 108
- Titus BM, Benedict C, Laroche R, et al. (2019) Phylogenetic relationships among the clownfish-hosting sea anemones. *Molecular Phylogenetics and Evolution* 139: 106526
- Zarei Darki Behrouz, Fahimeh E, Neda S (2021) Dinoflagellate endosymbiont of the soft coral *stichodactyla haddoni* savillekent collected from the coast of hormuz island, strait of hormuz. *Russ J Mar Biol* 47: 126–133
- Zenkert C, Takahashi T, Diesner M-O, Özbek S (2011) Morphological and molecular analysis of the *Nematostella vectensis* Cnidom. *PLOS ONE* 6: e22725
- Zhou Z, Yu X, Tang J, et al. (2017) Dual recognition activity of a rhamnose-binding lectin to pathogenic bacteria and zooxanthellae in stony coral *Pocillopora damicornis*. *Dev Comp Immunol* 70: 88–93

(Received December 1, 2021 / Accepted March 8, 2022 /  
Published online May 2, 2022)