University of South Bohemia in České Budějovice

Faculty of Science

Analysis of the dicyemid diversity using amplicon sequencing

Master's thesis

Bc. Tereza Flegrová

Supervisor: doc. RNDr. Jan Štefka, Ph.D.

Consultant: Tijana Cvetković, Ph.D.

České Budějovice 2022

Flegrová T., 2022: Analysis of the dicyemid diversity using amplicon sequencing. Mgr. Thesis, in English - 83 p., Faculty of Science. University of South Bohemia, České Budějovice, Czech Republic.

Annotation

Using amplicon sequencing, this study reveals dicyemid diversity around the world as well as within individual cephalopod hosts. A wide number of samples of multiple species were collected to reflect true dicyemid diversity. Heterogeneity primers were designed in order to improve the sequencing performance of the Illumina platform. A complex bioinformatic pipeline was implemented to process non-overlapping reads. The species delimitation methods were used to categorize occurring dicyemid types. Using the output of implemented methods, dicyemid diversity was assessed, proposing new trends opposing the current methodology.

DECLARATION

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

České Budějovice, 6.4.2022

Bc. Tereza Flegrová

Acknowledgment

First of all, I would like to express my endless gratitude to my dear friend and colleague Tijana Cvetković, who have helped me night and day, and without her, I would be utterly lost between all the samples in the freezer. A special thank you goes to my friend and colleague Masoud Nazarizadeh for massive help with bioinformatic analysis. I would also like to thank my supervisor Honza Štefka, who guided me through my master's studies and many thanks go to the whole lab crew for making the lab a very relaxed and fun work environment. I would also like to thank Maruška Drábková for laying a solid foundation for both my knowledge of dicyemids and this whole thesis. A big thank you also goes to all the colleagues who provided us with samples, and/or advised us regarding the bioinformatic part of the analysis. Last but not least, I also want to thank my fiancé, friends, family, and my dog for emotionally supporting me during my studies.

Table of contents

1 Introduction

Studying parasite distribution and host specificity provides us with information on phylogeographic distribution, possible migration patterns, and host population structure. A better perception of host specificity can help predict the probability of a parasite's ability to successfully establish a new population and spread to new regions and/or hosts (Poulin & Mouillot, 2003). Considering the accelerating global climate change, it is of utmost importance to understand these patterns better, and especially in the cephalopod case, we are in "uncharted waters" (Roumbedakis et al., 2018).

Due to the loss of the external shell, most coleoid cephalopods (except for *Nautilus* spp.) are more prone to contracting parasites and diseases compared to other mollusk taxa. Cephalopods' sophisticated ability to camouflage and signal makes them highly evolved, but also susceptible to lesioning, which makes them more prone to contracting disease (Kinne, 1990). Cephalopods are either intermediate or definitive hosts for many parasites, for example, cestodes, nematodes, protists, and also dicyemids (more on this subject in Hochberg, 1990). Dicyemids are poorly known parasites of benthic cephalopods. While their life cycle remains to be partly mysterious, there have been a significant number of discoveries since their first description by von Kölliker in 1849. One of those discoveries discussed possible mating strategies, and based on microsatellite markers, they established that while reproducing both sexually and clonally, dicyemid infrapopulations also form geography-bound genetic clusters (Drábková et al., 2021). In theory, marine species do not face the obvious dispersal obstacles as terrestrial species, suggesting vast spanning areas and therefore low inter-population genetic variation (Palumbi et al., 1992). In reality, many species show very high diversification (e.g., *Sepia esculenta*, see Zheng et al., 2009). This is given by both the biological traits and physical barriers, mostly sea or ocean currents (Palumbi et al., 1992).

Even though dicyemids are commonly referred to as parasites, many studies indicate that it may not be the case (Katayama et al., 1995; Furuya et al., 2004). They absorb the urine in the renal organs, possibly not causing harm to the host at all (Ridley, 1969). Nevertheless, parasites or not, it is crucial to understand dicyemids better to unveil their relationship with their cephalopod hosts, as octopuses and sepias play a major role in aquaculture and marine ecosystems. Considering dicyemids' high prevalence (see chapter 1.2: The relationship between dicyemid and its host), we might assume that every cephalopod caught by fishermen is infected. Whether this is important for human consumption is yet to be discovered. In either case, it remains

unclear whether high or low dicyemid infections play a role in cephalopods' life quality. To shed some light on dicyemids' diversity patterns, we carried out a study including cephalopod specimens from all over the world.

1.1 Dicyemids

Dicyemids (Fig. 1) are tiny, worm-like organisms that live in the renal organ of benthic cephalopods. Dicyemids are present in most octopods and sepioid decadpods, but rarely occur in teuthoid decapods (with exceptions such as *Sepioteuthis lessoniana*, *Todarodes pacificus, Sepioteuthis australis*) (Fururya et al., 2004; Catalano et al., 2014). The surface of the renal organ is usually densely covered with dicyemids, reaching up to thousands of individuals per cm^3 (Furuya & Tsuneki, 2003). Their body structure is simple, mostly consisting of only 8 to 40 cells (Furuya & Tsuneki, 2003). Dicyemids do not possess body cavities or differentiated organs (Suzuki, 2010). Their body is formed by one prolonged axial cell, surrounded by peripheral cells. In the frontal part, the peripheral cells form the "calotte" ("head"), used for attaching to the renal organ (Fig. 2). Morphologically, three different shapes of calotte are used for species determination - conical-shaped, cap-shaped, and disc-shaped (Furuya & Tsuneki 2003). The dicyemids genome is highly reduced, for example, the genome size is only 67,5 Mb for *Dicyema japonicum* (Lu et al., 2019).

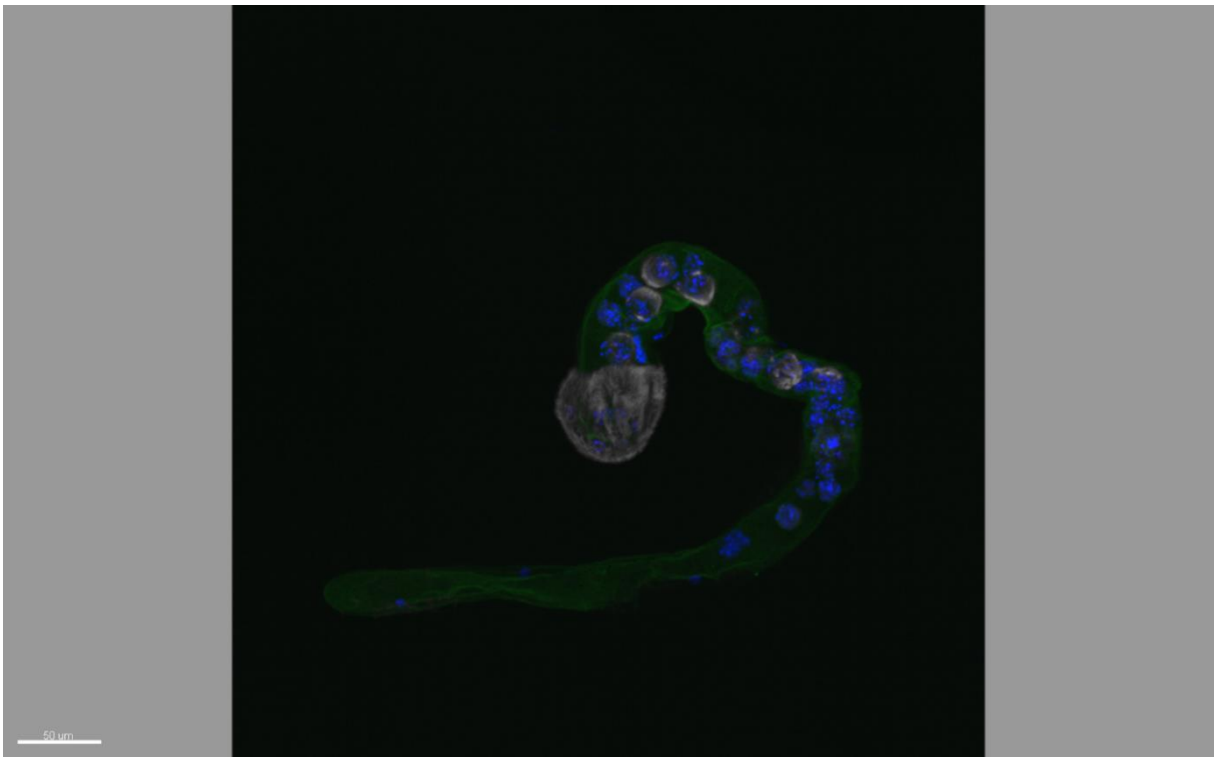


Fig. 1: Confocal microscopy of *D. moschatum*. DAPI + phalloidin (actin; green) + beta-tubulin (white). Photo by Marie Drábková.

New findings from a recent study (Drábková et al., 2021) provided more information about the life-cycle and genetic organization of dicyemids. According to microsatellite diversity results of Drábková et al. (2021) dicyemids are diploid organisms, reproducing both clonally and sexually. Moreover, the rate of heterozygosity is higher than expected, regarding their unusual habitat and life cycle. Their infrapopulation structure mostly follows the geographic pattern of the host. Multiple re-infection of the host is assumed due to several genotypes within one infrapopulation.

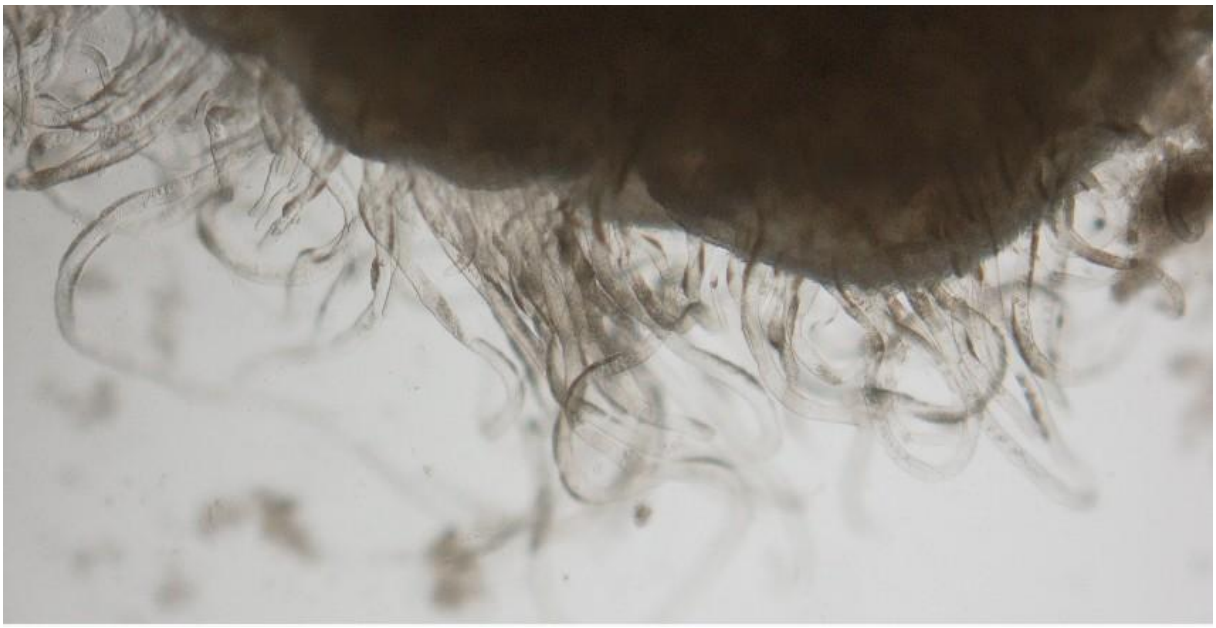


Fig. 2: Dicyemid individuals attached to the renal organ of a cephalopod. Photo by Marie Drábková.

1.1.1 Taxonomy

The most accepted phylogenetic placement of Dicyemida is among phylum Mesozoa, together with Orthonectida. However, the monophyly of the phylum has been disputed (Dunn et al., 2014; Telford et al., 2015; Bleidorn, 2019) and its placement in the tree of life remains unclear. Mesozoa were even considered as possible ancestors of Metazoa, due to their simplified body structure (Lapan & Morowitz, 1975). Stunkard (1954) lists the possibility that dicyemids could be ranked within flatworms, but this claim has been disproved based on innexin encoding cDNA revealing three distinct major protostome lineages - arthropods, nematodes, and lophotrochozoans (Suzuki et al., 2010). According to the latest phylogenomic studies, Dicyemida could be nested in Lophotrochozoa, either as a monophyletic group or with Dicyemida and Orthonectida separated (Lu et al., 2017; Schiffer et al., 2018; Zverkov et al.;

2019). Thus, a general agreement on the evolutionary origin of dicyemids has not been reached yet.

Molecular data (18S rDNA) revealed that even morphologically different species (for example *Dicyemennea brevicephala*, *Dicyemennea adscita*, *Dicyemennea adminicula*, and *Dicyema apollyoni*, all infecting *Octopus rubescens*) are genetically alike (Eshragh & Leander, 2014). That means that the current identification of morphospecies as species may be misleading; which is also supported by Catalano et al., 2012. Catalano et al. (2012) also state that out of 112 currently morphologically described species up to 20 % may be invalid. One possible answer to assessing better the taxonomy of dicyemid species is DNA barcoding. Currently, the most used barcode for eukaryotes is cytochrome c oxidase subunit 1 (COI) in mitochondria (Evans et al., 2007). However, endoparasites (such as dicyemids) tend to have highly reduced mitochondria due to the environment they occupy, making the COI unideal genetic marker (Awata et al., 2005; Tsaousis et al., 2008). A better choice for dicyemids is 18S rDNA which is known to evolve fast in parasites (Crainey et al., 2009; Bucklin et al., 2011). To answer the intra- and interspecific variation of dicyemids, the key is sequencing 18S rDNA for all morphospecies. By establishing alpha and beta diversity based on amplicon sequencing results, we can estimate how many dicyemid types are present within one host and whether there are different types among various host species and locations. Alpha diversity is a measure of species diversity within one host, whereas beta diversity measures species diversity between hosts.

1.1.2 Life cycle

The main reason for dicyemids' simple body structure is probably not their “primitive” origin, but their parasitic lifestyle (Stunkard, 1954). The life cycle includes two stages: vermiform and infusoriform stage (Furuya & Tsuneki, 2003). These two stages differ in their body structure (Fig. 3). The infusoriform stage is represented by the embryo after sexual reproduction, hatched out of a fertilized egg - the infusorigen. The embryo (nematogenic or rhombogenic individual) at vermiform stage is made asexually. The vermiform stage can only be found within the host. The infusoriform stage represents the most probable way of infecting a new host. This is due to their ability to leave the host and independently swim [for several days in vitro (McConnaughey, 1951)] in search of a new cephalopod host. Unfortunately, the precise mechanism of the transformation from infusoriform to vermiform individual is unknown, as well as the exact way of infection (Furuya & Tsuneki, 2003).

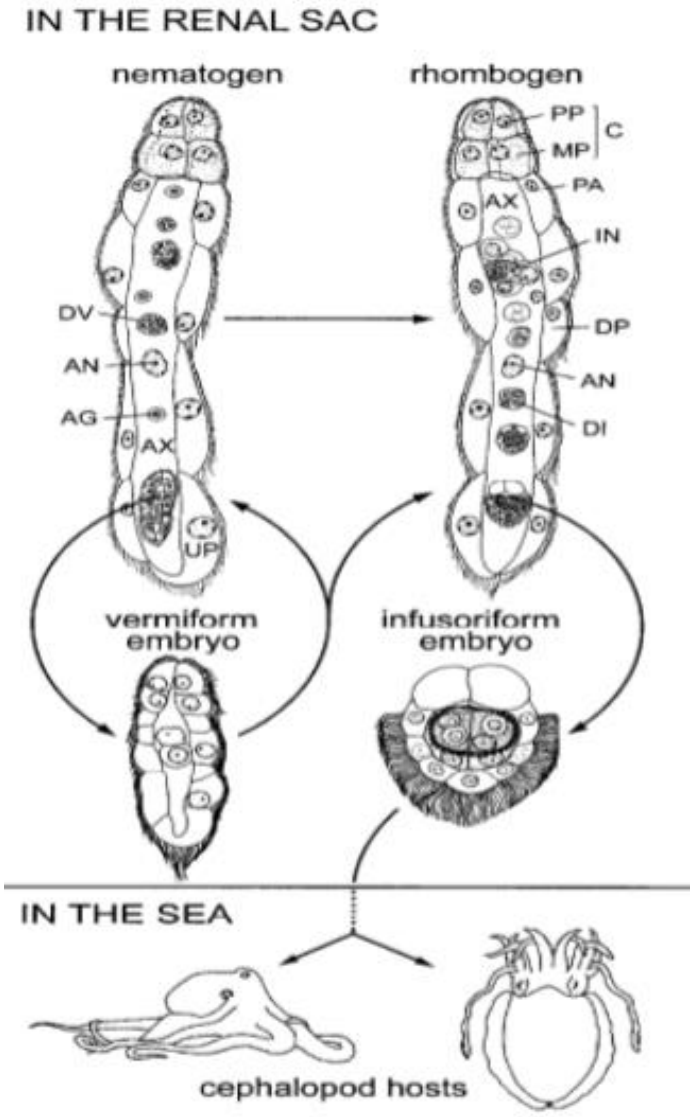


Fig. 3: Dicyemid life cycle. AG – asexual cell, AN – nucleus of axial cell, AX – axial cell, C – calotte, DI – developing infusoriform embryo, IN – infusorigen, DP – dipolar cell, DV – developing vermiform embryo, IN – infusorigen, MP – metapolar cell, PA – parapolar cell, PP – propolar cell, UP – uropolar cell (adopted from Furuya et al., 2003).

1.2 The relationship between dicyemid and its host

While exploring the relationship between dicyemids and their hosts it is vital to realize the importance of the dicyemids' way of attaching themselves to the folds of the renal organ. The type of calotte is defined by the structure of renal organ's surface, and the depth and shape of its folds. As a consequence, only morphologically compatible species of dicyemids are able to attach to specific host species.

The question is how morphologically different could be various individuals of one species. One of the proposed theories suggests that dicyemids adapt according to the renal organ's shape in which it finds itself (Eshrag & Leander, 2014). This would lead to the conclusion that only genetically closely related cephalopods can host similar dicyemids, and even the slightest change within the structure of the renal organ's folds would affect the population structure of present dicyemids (Poulin et al., 2011). Assuming that mutual relationships between host and dicyemid exist, it is probable that the change in dicyemid population structure would affect host fitness as well.

The maximum number of dicyemids present within one host is set to be three (Furuya & Soudienne, 2019), however, if this assessment was based only on morphological description of occurring dicyemids, it might be disproved by molecular data (as in Eshragh & Leander, 2014). Studies of dicyemid life cycle face many obstacles, due to simplified dicyemid morphology, and impossibility to study dicyemids in their natural habitat or to cultivate their hosts (EU legislative 2010/63/EU- animals protection for scientific purposes) (Roumbedakis et al., 2018). The simplest solution to this problem would be to infect the eggs directly, from female to its offspring, but this theory was not confirmed (Catalano, 2013). Other aspects that could take part in the way of infection, are the host's age, geographic location, or size (Catalano et al., 2014). However, despite all the obstacles to dispersal between hosts, the dicyemid prevalence in cephalopods is high – reaching up to 100% in the Mediterranean Sea (Nouvel, 1947). As dicyemids are found in renal organs, their ciliary activity seems to help maintain a constant flow of urine, which may redefine them as symbionts to cephalopods rather than their parasites (Furuya & Souidenne, 2019), supporting earlier claims about them not-harming the host at all (Katayama et al., 1995; Furuya et al., 2004).

1.3 Amplicon sequencing

Amplicon sequencing represents a cost-effective way to discover species compositions in various environments. Especially microbiologists adopted this technology to characterize microbial communities by targeting conserved regions such as 16S rRNA, or by the so-called shotgun metagenomics (randomly sheared DNA molecules) (de Muinck et al., 2017). The method is of utmost importance for discovering non-cultivable organisms which would otherwise go undetected (Fadrosh et al., 2014). The data acquired by this type of sequencing enable us to relate a huge number of samples to one another in order to reveal biological patterns. Other approaches, such as metagenomics, could be employed for the same purpose,

especially for revealing insights into genes' functions, but 16S rRNA targeting could be extremely valuable to assess biodiversity in a community of interest (Caporaso et al., 2011). The ability to relate 16S rRNA trends at the species level to the host is an especially useful and powerful tool (Hamady & Knight, 2009).

The Illumina MiSeq platform allows high-throughput sequencing, generating highly accurate single-end or pair-end reads of the maximum length of 2 x 300 bp resulting in 15 Gb of output data (Illumina website). The most popular way of employing Illumina MiSeq for amplicon sequencing is by using single or dual indexing, by attaching the barcode (the index) either from a single or from both ends of the targeted insert. This method allows for multiplexing many samples and saving time and money (Fadrosh et al., 2014).

The V4 region, a part of rRNA region, is currently the most commonly used. Nowadays, primers 515F-Y and 926R by Parada et al. (2016) are optimal for marine taxa. As described in Parada et al. (2016), previously widely used primers 515F-C and 806R (Caporaso et al., 2011, 2012) either overestimated or underestimated several marine taxa compared to the newly proposed primers. These primers are also capable of amplifying 18S eukaryotic rDNA; one more reason why they were used in this study.

1.4 OTUs and ASVs

When working with the output data from rDNA amplicon sequencing, the golden standard for years has been the use of operational taxonomic units - OTUs. OTUs are clusters of reads that do not differ by more than the fixed sequence dissimilarity threshold, which is usually set to be 3 % (Westcott & Schloss, 2015; Kopylova et al., 2016). These clusters are meant to correspond with a species of the same DNA sequence. OTU sequence table created based on this sequence similarity serves as a basis for further analysis. Clustering based on the percentage of similarity between sequences depends on the dataset that is used for *de novo* OTU clustering. This can potentially introduce error, as OTUs from two different sequencing runs of different quality can be very inconsistent. Closed reference OTUs, that are created based on a reference, are more precise. However, when comparing *de novo* OTUs to the reference ones, part of the natural biological variability may be lost due to a comparison based on the percentage of similarity (Callahan et al., 2017).

Alternatively, a method called amplicon sequence variants - ASVs - has been introduced in the last decade (Eren et al., 2013; Tikhonov et al., 2015; Eren et al., 2015; Callahan et al., 2016a;

Edgar, 2016; Amir et al., 2017). ASVs are not created based on the reads separately, but with a consideration of the whole sample, thus lowering the chance of processing possible sequencing artifacts. Unlike OTUs, ASVs reflect true biological variability, as they are not based on percentage similarity but on the truly observed sequences, making it possible to differentiate between variants differing by as little as one nucleotide. This provides higher sensitivity and specificity than OTUs, which in turn can better distinguish existing ecological patterns. ASVs reflect biological reality outside the analyzed data - real DNA sequence of studied organisms. Another important advantage of using ASVs is their reproducibility across studies and independence from reference databases (Callahan et al., 2017).

1.5 Species delimitation

Recognizing whether a group of organisms belong to a different population of a single species or whether they constitute a different species is called species delimitation. Historically, this process was based on the morphological traits of the organism only. This approach can prove problematic for species with not enough distinctive traits or highly plastic morphology (for example due to environmental factors), such as dicyemids. Additionally, morphological species delimitation depends on the expertise of the taxonomist and can be time-consuming, which is especially problematic for species undergoing extinction. The idea that the process gets semi-automated, and taxonomists only verify whether the results obtained from genomic data make sense, is therefore interesting for further investigation (Rannala & Yang, 2020).

The most known approach for species delimitation is so-called “DNA barcoding”, which focuses on sequencing one gene across the organisms (e.g., COII for most animals). However, methods based on a single locus are probably going to have low power for recently diverged species (Rannala & Yang, 2020). Nowadays, most molecular heuristic methods for species delimitation still originate from the DNA barcoding idea. Heuristic methods, even though computationally efficient, could be sometimes hard to interpret and might have poor statistical properties, depending on the specific parameters (Rannala & Yang, 2020). Various heuristic methods are available for species delimitation. Among them, the most widely used are GMYC and PTP approaches.

General Mixed Yule Coalescent (GMYC) (Pons et al., 2006) works by classifying waiting times between coalescence events in a gene tree into two categories: those within species determined by the coalescent process, and those between species determined by the Yule process. The time needed for gene-tree nodes to switch between Yule and coalescent process is estimated by

Maximum Likelihood to determine species delimitation. A likelihood ratio test is used for occurrences of single or multiple species. While this method is computationally simple, it ignores potential errors in gene tree. GMYC method is considered to work optimally for datasets with sufficient intervals between speciation events and small population sizes. bGMYC approach is a Bayesian implementation of GMYC (Reid & Carstens, 2012).

Poisson Tree Process (PTP) (Zhang et al., 2013) works with the distribution of branch lengths in gene tree, and based on them identifies species. PTP uses a rooted non-ultrametric tree, and does not rely on molecular clock. bPTP is a Bayesian implementation of standard PTP.

While convincing-enough evidence of genetic isolation might be enough for sympatric species, populations from distant geographic locations might not be so easily distinguished. The genetic isolation in those cases could be only due to isolation by distance (Rannala & Yang, 2020). In order to justify species' status, multiple sources must be combined to make a compelling case for species delimitation. In some cases, it might be the best approach to set boundaries within specific population types, rather than insisting on labeling them as species, potentially introducing errors to the taxonomy.

2 Main goals

Herein, amplicon sequencing was used for 227 cephalopod individuals to contribute to resolving still uncertain and largely unexplored dicyemid diversity and life strategy, as well as their interactions with the host species. The main goals were:

1. to test whether one or more dicyemid species per host is present;
2. to explore geography-bound diversity of dicyemids;
3. to investigate dicyemid composition among various cephalopod host species (host-specific diversity);
4. to estimate the number of dicyemid species (lineages) that occur across all the sampled hosts and locations worldwide, helping to estimate the possible true number of existing species (dicyemid diversity per se).

3 Materials and methods

3.1 Collection of samples

We obtained 264 accessions (Fig. 5) of various cephalopods' renal organs (Fig. 4). Samples were collected during multiple field trips in Europe (Fig. 6) at various localities from 2015 to 2021 (for the accession numbers and voucher information see Attachments, Table I). In addition, we included samples from China, Australia, Japan, Hawaii, and Vietnam provided by colleagues, and species downloaded from GenBank (for voucher information see Attachments, Table I).

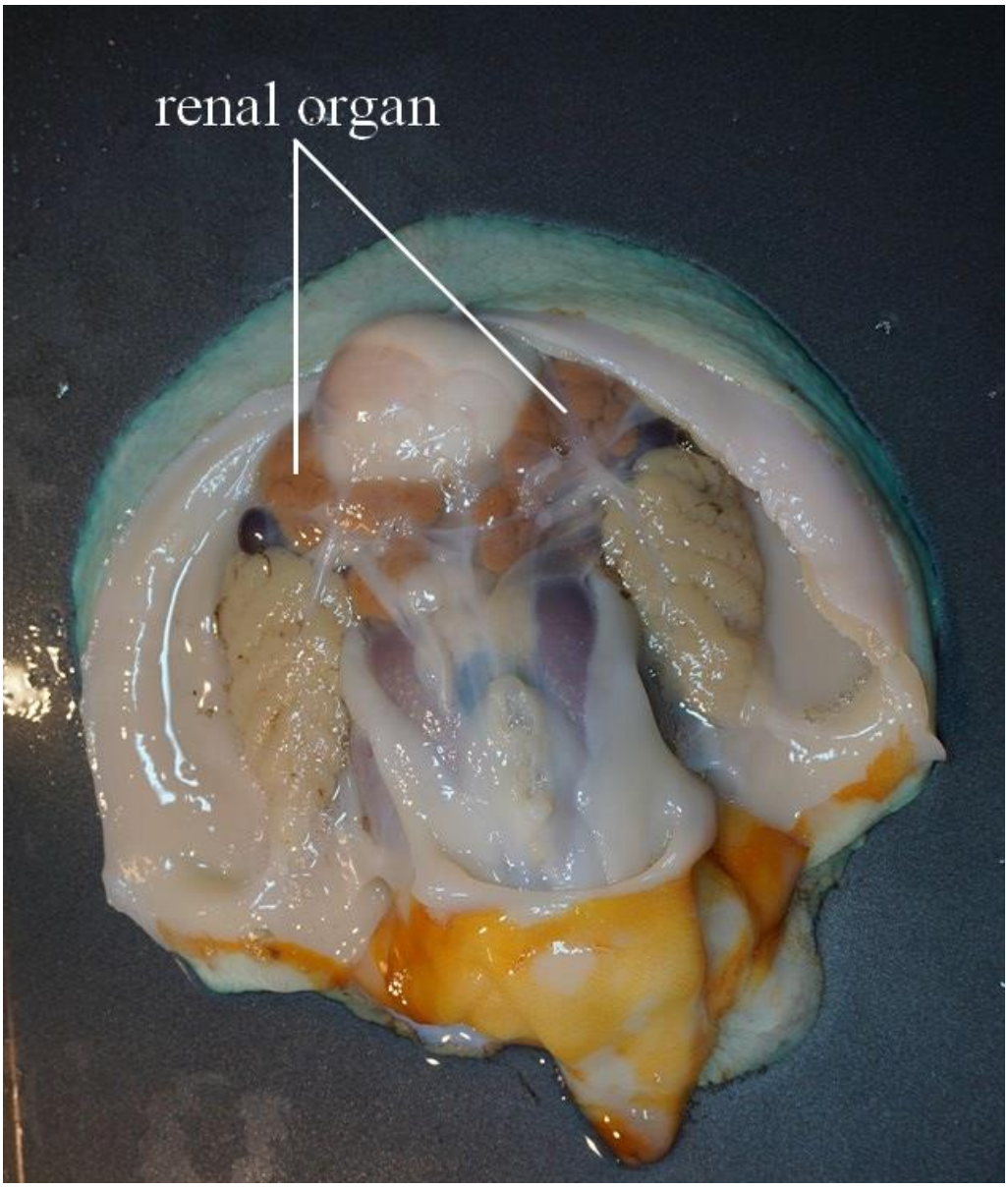


Fig. 4: Dissected octopus *Eledone moschata*. Photo by Marie Drábková.

Host species were usually purchased from the local fishermen at markets or obtained by vessel fishing. The exact collection locality was confirmed with the retailer to exclude the possibility of the uncertain geographical origin of samples. After dissection, a piece of both the renal organ and a part of a tentacle (a host tissue voucher) were preserved in pure ethanol. Some of the early-collected dicyemids samples were obtained by washing the renal organ in artificial seawater (ASW - see Lapan & Morowitz, 1975). After releasing dicyemids, the samples were carefully centrifuged to form a pellet for further DNA isolation. For a detailed sampling procedure see Drábkova et al., 2019.

Fig. 5: The map of all the accessions. Pins denote the position of sampling localities.

Fig. 6: European sampling localities. Pins denote the position of sampling localities.

3.2 DNA isolation

Before DNA extraction samples were removed from ethanol and left to dry. DNA samples from both the renal organ and tentacle were extracted using DNeasy Blood and Tissue kit (QIAGEN) according to the manufacturer's manual.

3.3 Verifying the host species

Gene marker cytochrome c oxidase I (COI) was used to determine the host species by PCR amplification and sequencing. The reaction contained: 1 µl of DNA sample, 1 µl of forward primer (5pM, F1490, Folmer et al., 1994; sequence in Tab. 1), 1 µl of reverse primer (5pM, H7005, Hafner et al., 1994; sequence in Tab. 1 or 5pM, H2198, Folmer et al., 1994; sequence in Tab. 1), 2 µl of buffer (PCR blue buffer, Top-bio), 0,5 µl of nucleotides (dNTPs 10mM, ThermoFisher), 0,2 µl of Taq polymerase (Top-bio) and 14,3 µl of H_2O. For the specific thermocycler setup, see table 2. PCR products were visualized by gel electrophoresis (1% agarose, 1KB GeneRuler ladder ThermoFisher), enzymatically cleaned (0,5 µl exo I nuclease Top-bio, 2 µl FastAP, 2,5 µl H2O), and Sanger sequenced using PCR primers in a commercial laboratory (Seqme, CZ).

Tab. 1: Primer sequences for the COI PCR amplification.

Loci	Name	Sequence of the primer 5´→ 3'	Forward/Reverse
COI of the host	F1490	GGTCAACAAATCATAAAGATATTGG	F
	H7005	CCGGATCCACANCRTARTANGTRTCRTG	R
	H2198	TAAACTTCAGGGTGACCAAAAAATCA	R

Tab. 2: Thermocycler configuration for the COI PCR amplification.

Loci	Initial denaturation	No. of cycles	Denaturation	Annealing	Elongation	Final elongation
COI of the host	94° C 5 minutes	30	92° C 1 minute	52° C 1 minute	72° C 1 minute	72° C 5 minutes

3.4 rDNA library preparation

The amplicon library was prepared following a slightly modified Illumina protocol (Illumina 16S metagenomic sequencing library preparation, 2017). A two-stage PCR protocol was chosen in order to optimize economic efficiency by allowing the sharing of barcoded oligos between different projects (opposite to one-stage PCR).

3.4.1 Designing oligonucleotides

When preparing an amplicon library it is crucial to realize that the final library will be of low complexity, as it consists of millions of the same or highly similar reads. This can cause a problem in the cluster identification and color matrix estimation for the sequencing machine, as the same signal flashes and can cause sequencing failure (de Muinck et al., 2017).

One of the most common solutions to this is adding a spike-in of PhiX DNA, which will diversify the library and therefore prevent the signal from clashing. The downside of this approach is the fact that adding a substance that is not of our interest takes up space on the sequencing lane, meaning that the amount of obtained data is lower. To avoid losing data to the PhiX spike-in, we decided to use heterogeneity primers, i.e., inserting short spacers (0-4 nucleotides long) to our construct, similarly to Fadrosh et al. (2014). These spacers are random nucleotides inserted in front of the primer itself but also after the reading site (reading primers Rd1 SP, Rd2 SP). The construct added in the first round PCR (referred to as Amplicon PCR) consists of the primer itself (primers 926R and 515F-Y by Parada et al., 2016), spacer, and the

overhang, which is necessary for binding the indices and adapters in the second-round PCR (Fig. 7). In the second round PCR (referred to as Index PCR) the construct consisting of the Illumina adapter site, also known as P5 and P7, is added together with the index and an overhang that binds to the overhang from the Amplicon PCR (see Fig. 5). In this thesis, we used dual indexing approach. For the complete list of indices see attachments, Table II. Dual indexing allows multiplexing up to 384 samples for one sequencing run, which can significantly reduce the cost of sequencing. To determine the best molarity of used primers, Bioanalyzer was run to check for primer-dimers and other potentially unwanted activities.

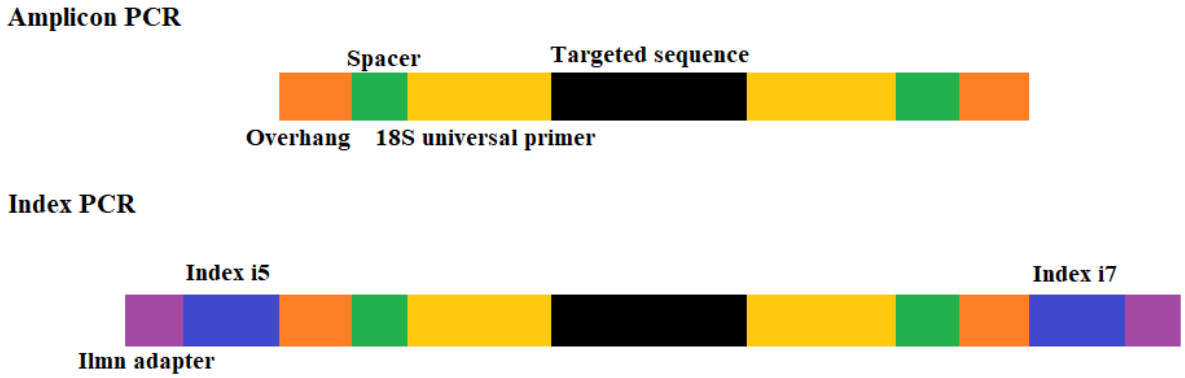


Fig. 7: Visual representation of the construct added to the Targeted sequence.

3.4.2 Amplicon PCR

Amplicon PCR was carried out for 185 samples in duplicates (to control for amplification bias) plus negative controls, i.e., 374 reactions were done in total for the first prepared library. The second library was prepared using 252 samples. The required DNA template concentration for every sample was min. 4 ng/µl, but the average was around 12 ng/µl. Every reaction consisted of 12,5 µl KAPA HotStart ReadyMix (Roche), 5 µl 1µM reverse primer (sequence as in table 3), 5 µl 1µM forward primer (sequence as in table 3), and 2,5 µl of the sample. For the thermocycler setup please see table 4.

Tab. 3: Primers used for the Amplicon PCR.

Loci	Name	Sequence of the primer 5´→ 3'	Forward/reverse

16S bacterial/18S dicyemid	Ilmn-16S-FW	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGN(0-4x)GTGCCAGCMGCCGCGGTAA	F
16S bacterial/18S dicyemid	Ilmn-16S-RV	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGN(0-4x)CCGYCAATTYMTTTRAGTT	R

Tab. 4: Thermocycler configuration for Amplicon PCR.

Loci	Initial denaturation	No. of cycles	Denaturation	Annealing	Elongation	Final elongation
16S bacterial/ 18S dicyemid	95° C 3 minutes	25	95° C 30 seconds	55° C 30 seconds	72° C 30 seconds	72° C 5 minutes

3.4.2.A Clean-up

The clean-up of the PCR reaction was carried out using AMPure XP beads (Beckman). 20 µl of magnetic beads were added to each reaction, incubated for 5 minutes outside of the magnetic stand, followed by 2-minute incubation on the magnetic stand. These steps were followed by two ethanol washes (80% EtOH) and then incubating washed beads in 52,5 µl of H_2O for 2 minutes outside the magnetic stand and for 2 minutes on the magnetic stand. Finally, 50 µl of cleaned-up PCR product was transferred to new tubes.

3.4.2.B Quantification

In order to verify the success of PCR and clean-up, all samples were measured on a Qubit fluorometer using a dsDNA High Sensitivity kit. To verify that desired product was obtained, gel electrophoresis was done as well; specifically, 1,5% agarose gel with GelRed (Biotium), ladder used was GeneRuler 100 bp Plus DNA (ThermoFisher), and to load samples the 6x

Loading Dye (ThermoFisher) was used. Only products with visible bands and/or concentrations above 0,2 ng/µl were used in further steps.

3.4.3 Index PCR

For the index PCR of the first prepared library, most of the duplicates from the first reaction were merged, and only about one-third of the duplicates were kept separate (in order to compare the sequencing results for the duplicates, to reveal whether any sequencing bias occurs or not). This merging resulted in preparing 216 reactions, instead of the original 370 (374 with negative controls). For the index PCR of the second library, we proceeded with 224 samples out of 252 original samples. The reaction was done as follows: 25 µl KAPA HotStart ReadyMix (Roche), 10 µl H_2O, 5 µl of the Amplicon PCR product, 5 µl of the 5µM forward index primer, and 5 µl of the 5µM reverse index primer (for sequences check the table 5). For detailed thermocycler setup, see table 6.

Tab. 5: Sequences of the primers used for the Index PCR. Sequences of the indices are in the Attachments, table II.

Name	Sequence of the primer 5´→ 3'	Forward/Reverse
S5XX	AATGATACGGCGACCACCGAGATCTACAC (indexS5XX)TCGTCGGCAGCGTC	F
N7XX	CAAGCAGAAGACGGCATACGAGAT (indexN7XX)GTCTCGTGGGCTCGG	R

Tab. 6: Thermocycler configuration for the Index PCR.

Loci	Initial denaturation	No. of cycles	Denaturation	Annealing	Elongation	Final elongation

Illumina index addition	95° C 3 minutes	8	95° C 30 seconds	55° C 30 seconds	72° C 30 seconds	72° C 5 minutes

3.4.3.A Cleanup

Clean-up of the indexing PCR reaction was carried out using AMPure XP beads (Beckman). 56 µl of magnetic beads were added to each reaction, incubated for 5 minutes outside of the magnetic stand, followed by 2-minute incubation on the magnetic stand. These steps were followed by two ethanol washes (80% EtOH) and then incubating washed beads in 27,5 µl of H_2O for 2 minutes outside the magnetic stand and 2 minutes on the magnetic stand. Finally, 25 µl of cleaned-up PCR product were transferred to new tubes.

3.4.3.B Quantification

As in the previous PCR, to verify the success of PCR and clean-up, all samples were measured on a Qubit fluorometer using a dsDNA High Sensitivity kit. To verify that expected product was obtained, gel electrophoresis was done as well, in the same way as described above.

3.4.4 Library quantification, normalization, and pooling

To provide optimal and unbiased sequencing results, all samples were set up to the same concentration before pooling. The first library had a 5 nM concentration, required for sequencing (Norwegian Sequencing Centre). We included 188 samples, excluding the samples of poor quality. The second library that had a 7 nM concentration included 165 samples and was sequenced by Novogene. We used the following formula for molarity calculation: (concentration in ng/µl) / (660 g/mol × average library size) × 10^6 = concentration in nM. The average library size was determined to be 660 bp long using Bioanalyzer.

3.5 Sequencing via NGS Illumina

The prepared library was sequenced on the MiSeq Illumina machine, using 250 bp paired-end sequencing (Norwegian Sequencing Centre, Oslo, Norway and Novogene, Cambridge, UK).

3.6 Bioinformatic analysis

Different software was used to analyze raw dataset. Besides the commonly-used software and pipelines for processing Illumina data (FastQC), we also used amplicon-specific software (MetReTrim), and for the most important part of the pipeline (ASV approach), we followed DADA2 Pipeline Tutorial (1.16) working with DADA2 (Callahan et al., 2016) in R (R Core team, 2021; Rstudio team, 2020) (see details in Fig. 8). In order to obtain statistical values of alpha and beta diversity, DADA2 pipeline was run within qiime2 (Bolyen et al., 2019). To validate the accuracy of our approach we ran the analysis in qiime2 using only forward reads first.

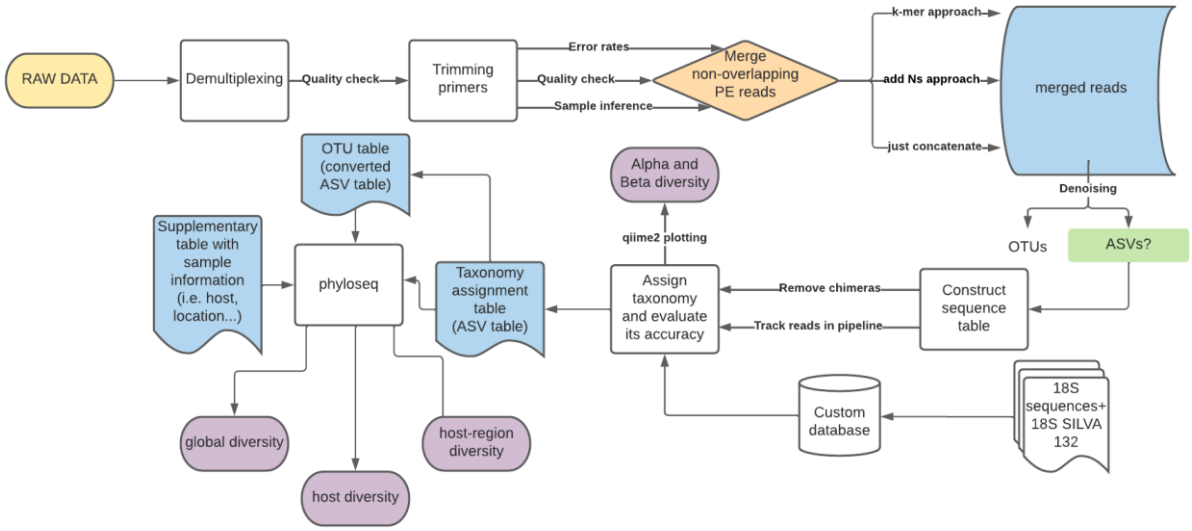


Fig. 8: Flowchart of the bioinformatic pipeline.

3.6.1 Demultiplexing

Demultiplexing of the total of 5148232 raw reads (1385654 unique reads) was done by the sequencing facility. Between 211 and 40692 reads per sample (mean = 7292, median = 4725) were obtained.

3.6.2 Quality check

FastQC (Andrews, 2010) module was run on all the fastq.gz files and then multiqc (Ewels et al., 2016) was run on the output of the FastQC.

3.6.3 Trimming primers

MetReTrim python script (Sharda, 2020) was used to trim out the primers and heterogeneity spacers.

3.6.4 Deduplication

Duplicates were both simple PCR duplicates and/or separate isolates of one sample. All 264 accessions were represented by 622 reactions and based on a wet lab work workflow, some of them were discarded or pooled together. The duplicates that remained in the dataset after the sequencing were represented by 116 sublibraries, which were in the deduplication step merged using the `cat` command. Poor quality samples (ten sublibraries) were discarded. That means that in the following analyses a total of 227 accessions was processed (see Attachments, Tab. I).

3.6.5 Quality check

FastQC module was run on all fastq.gz files, and then multiqc was run on the output of FastQC, proving a high quality of sequences (Fig. 9)

Quality check was performed by qiime2, proving the sufficient quality of sequences (Attachments, Fig. I and II).

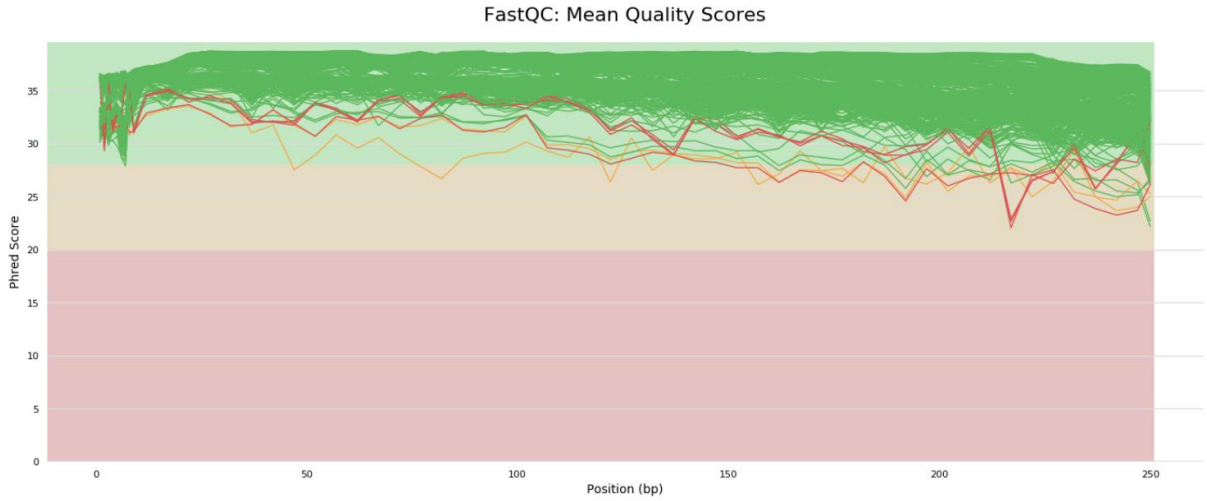


Fig. 9: multiqc report of the trimmed data showed good quality of reads.

3.6.6 Error rates and sample inference

Error rates establish how many nucleotide substitutions occur and whether it fits the expectations or not. Using the parametric error model of DADA2 we established error rates of our reads, which fitted well the expectations (forward reads in Fig. 10, reverse reads in Attachments, Fig. III).

We determined unique sequences among obtained reads using DADA2 as described in Callahan et al., 2016. Only unique sequences were used in further steps of analyses.

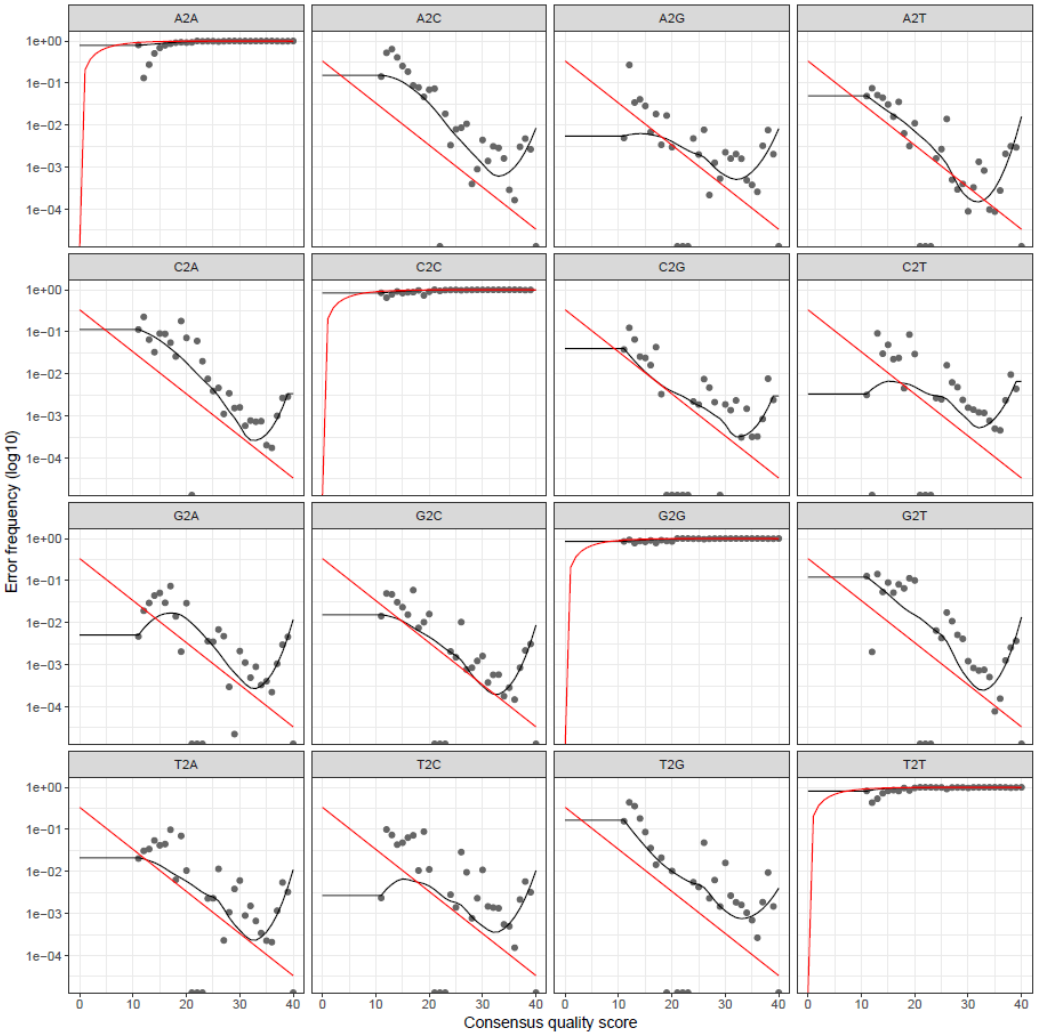


Fig. 10: Error rates of forward reads fit the expectation. Points represent the observed error rates. Black line estimates error rates after convergence of the machine/learning algorithm. Red line represents expected error rates when quality is optimal. *A2A = Adenine to adenine, A2C = Adenine to cytosine, etc.

3.6.7 Merging paired-end reads

When merging paired-end reads, most of the software depends on overlaps between forward and reverse reads. Since our insert size is around 660 bp and we obtained 250 bp paired-end reads from the sequencing facility, there is a significant gap between our reads and thus no overlap, except for the sequences that do not belong to the phylum Dicyemida (but are rather of a bacterial origin). When dealing with non-overlapping reads there are three possible approaches: k-mer based methods (such as BBMerge see Bushnell et al., 2017), adding Ns to

the gap instead (such as MeFit see Parikh et al., 2016 or DADA2 see Callahan et al., 2016) or "hard concatenation" of the reads. Multiple software was tried (e.g. PEAR, MeFit, BBMerge, vsearch, pandaseq, and DADA2), for both the overlapping and non-overlapping reads. Based on the outputs, DADA2 has been chosen, using the command `mergepairs`, as it worked the best for our dataset, discarding a minimum of reads. Choosing the "adding Ns" approach is also supported by Ansorge et al. (2021) who claim that this is an optimal approach for non-overlapping reads.

3.6.8 Denoising

Denoising was performed using `denoise-paired` command. For a more detailed output table with denoising statistics see Supplementary files (supplementary_tables.xlsx; list denoising).

3.6.9 Constructing sequence table

ASV table was constructed using the `seqtab` command by DADA2.

3.6.10 Removing chimeras

Removal of chimeric sequences was done using the `seqtab.nonchim` command by DADA2.

3.6.11 Tracking reads through the pipeline

In order to establish how many reads "survived" through the whole pipeline, the command `track` in DADA2 was used (Supplementary files, supplementary_tables.xlsx; list denoising).

3.6.12 Creating the database and assigning taxonomy

Taxonomy assignment was done using the command `assignTaxonomy` by DADA2. Reads were run against SILVA database suitable for DADA2, specifically version `silva_132.18s.99_rep_set.dada2.fa.gz` (Quast et al., 2013; Yilmaz et al., 2014; Glöckner et al., 2017). Then, a specific database was created by combining SILVA v.132 database with custom dicyemid sequences obtained by Sanger sequencing and dicyemid sequences available in Genbank.

3.6.13 Evaluating accuracy and filtering taxa

Due to possible sequencing error and contamination during lab work, assigned taxonomies were manually checked for illogical hits, which were filtered out. Out of 3,855 ASVs of 18S sequences, only 555 belonged to the phylum Dicyemida, and were used in further analyses (plotting and statistics). The rest of the ASVs represented mostly cephalopod hosts, fungi, chromidinids, and other non-targeted DNA present within the renal organ. Finally, 165 of 227 samples were used (supplementary_tables.xlsx; list TYPES_samples), the majority of discarded samples did not contain dicyemid reads. A few samples showed possible contaminant ASVs, which could have been the result of nonspecific primer annealing, cross-talk (index hopping during sequencing, MacConaill et al., 2018), or contamination during wet lab work. These contaminant reads were disregarded in further analysis.

3.6.14 Plotting in qiime2

Alpha and Beta diversity were established using a variety of plots, such as Bray-Curtis dissimilarity emperor plot (Attachments, Fig. IV), Jaccard similarity emperor plot (Attachments, Fig. VI), and rarefaction plot (sorted by locality) (Attachments, Fig. V). To test for statistical differences between groups of samples (species or localities) a variety of statistical approaches was utilized: Kruskal-Wallis, permanova, and Faith's phylogenetic diversity plot (Attachments table III, VI, IV, Fig. VII, respectively).

3.6.15 Constructing tree and species delimitation

Construction of a genetic tree was necessary to employ a variety of species delimitation methods. The tree was constructed using BEAST 2.5.2 (Bouckaert et al., 2019) with the following parameters: Markov chain Monte Carlo (MCMC) was set to 200,000,000, trees were sampled every 10,000 generations; every run was checked by Tracer v.1.7 (Rambaut et al., 2018). Posterior probability limit was set to 0,5; the maximum clade credibility tree was generated by TreeAnnotator (embed BEAST package). Based on Akaike Information Criterion, model of DNA sequence evolution was established using partionfinder v2.3 (Nylander, 2004). The tree was constructed using known 18S dicyemid sequences from Genbank (accessions in Attachments, table V; for detailed tree see Supplementary files treeDIC.pdf) and the filtered ASVs. As an outgroup, both the lophotrochozoan and bacteria were used. Based on ASVs' position in relation to Genbank dicyemid sequences, all ASVs were clustered into types/species

(Supplementary files; supplementary_tables.xlsx; lists types_asvs and types) using multiple species delimitation methods. Species delimitation results were obtained by implementing GMYC method via "splits" package (Ezard, Fujisawa & Barraclough, 2009) in R. Furthermore, bGMYC method was used via “bGMYC” package (Reid et al., 2013) in R. bPTP method was used using python3 script “bPTP.py” (Zhang et al., 2013).

3.6.16 Plotting species diversities in phyloseq

Processing of taxonomic assignment in phyloseq (McMurdie & Holmes, 2013) required the following input files: taxonomic assignment table, OTU table (ASV converted table), and a supplementary table with metadata for our samples (region, host species, sample ID). Because our qiime2-generated taxonomic assignment table was only as specific as the phylum level goes, we manually updated the ASVs’ taxonomic assignment file based on the results from the species delimitation analysis. This improved taxonomy assignment table was then used for plotting in phyloseq. Observed species, Chao, ACE, Shannon, Simpson, and inverted Simpson plots were calculated for establishing alpha diversity between octopuses, squids, and sepias, and plotted in R. PCoA (based on Bray-Curtis dissimilarity) and Network plot, clustering ASVs based on the Jaccard index were created in R using phyloseq. A filter of 20 % (at least 20 % of a certain type in at least one sample) was applied to reveal the most prevalent dicyemid types within samples.

4 Results

4.1 Global diversity

The species delimitation (Figs. 11, 12) resulted in determining 98 types (92, excluding Genbank accessions). All three species delimitation methods confirmed 37 dicyemid types (consisting of 483 ASVs); 25 ASVs clustered into 8 types that were supported by only one or two methods (these types carry a "PTP" suffix in the analyses) (Fig. 12). ASVs that did not cluster at all (47 ASVs) therefore represent their own dicyemid type in the analysis - the majority of these unclustered types are low-represented in the reads (see supplementary_tables.xlsx; list types). When 20 % filtering threshold was applied, 17 major types remained: Australia03, Catalonia01, *Dicyemennea eledones*, D_eledones_PTP, *Dicyema moschatum*, Dic01, Galicia01, China01, China02, China03, Lisbon01, Porto03, *Pseudicyema truncatum*, Tenerife01, Tyrh01, Viet01, and Viet02 type. These types were either abundant within one host or a sampling locality.

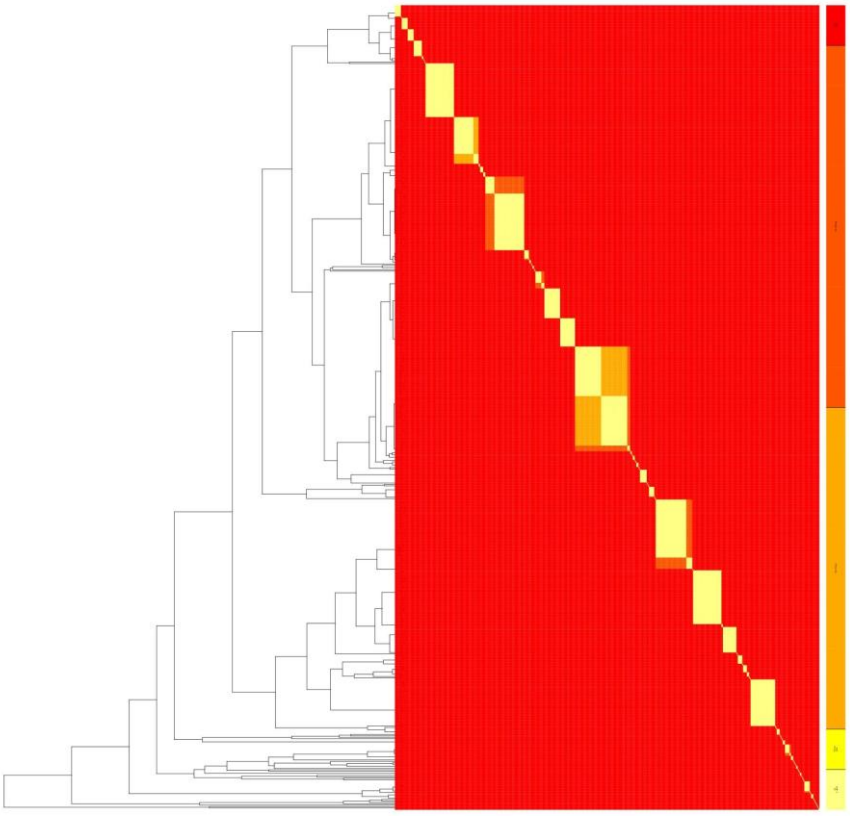


Fig. 11: The result of bGMYC species delimitation approach. P-value determines confidence of species delimitation. p=0,95-1 (pale yellow), p=0,9-0,95 (yellow), p=0,5-0,9 (orange), p=0,05-0,5 (blood orange), p=0-0,05 (red). Full figure available in Supplementary files, fig11.pdf

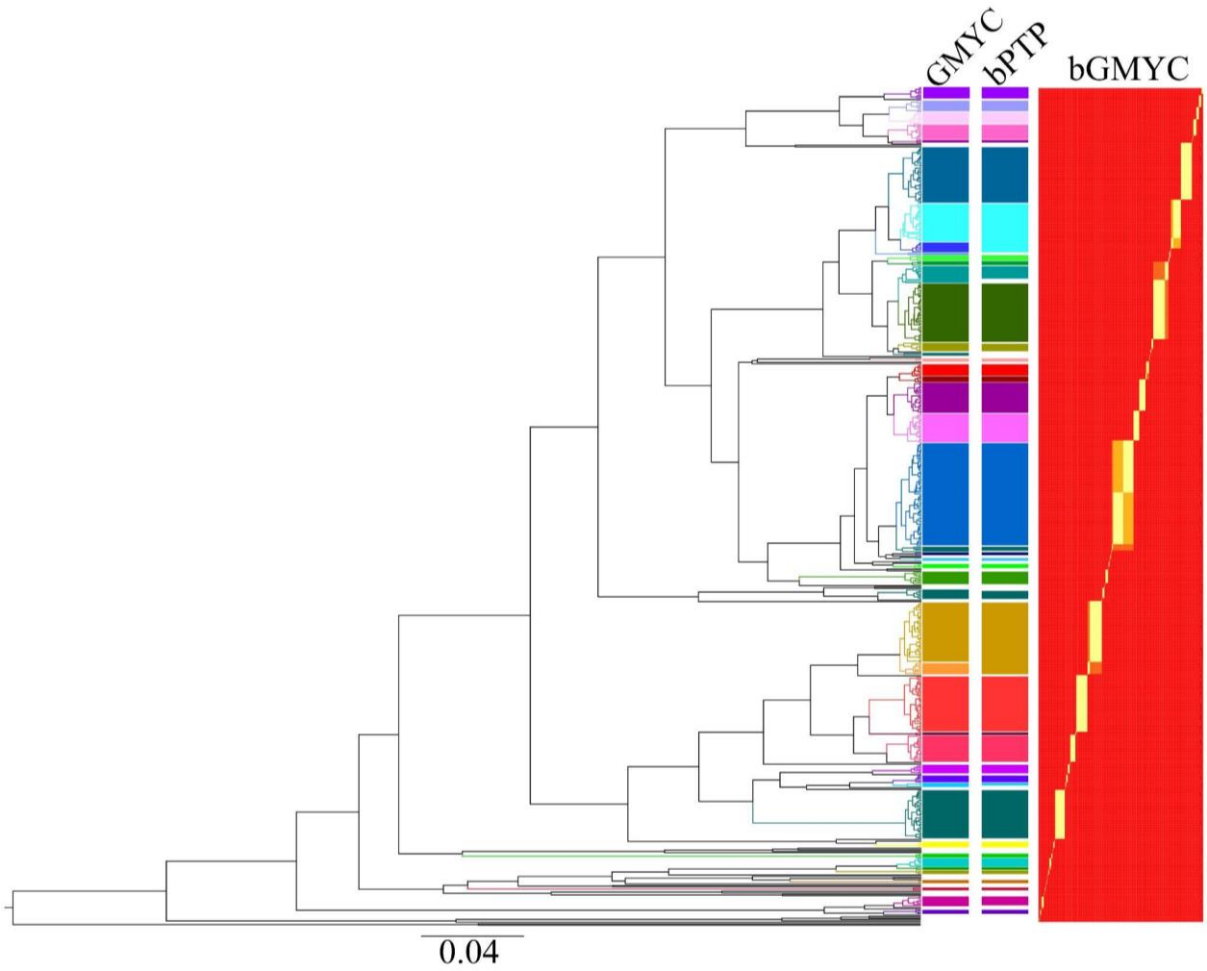


Fig. 12: Species delimitation showed 37 types of dicyemids confirmed by all three methods. Full figure available in Supplementary files, fig12.pdf.

Four Genbank accessions KJ786925, KJ786965, KJ786927, and KJ786928, belonging to four different species, *D. apollyoni*, *D. adscita*, *D. adminicula*, and *D. brevicephala*, were all clustered into one species according to all three species delimitation methods we employed (bright yellow cluster in the bottom half of Fig. 12), confirming findings of Eshragh & Leander, 2014.

All alpha diversity measure plots (Observed species, Chao1, ACE, Shannon, Simpson, and inverted Simpson) showed no significant dicyemid diversity difference between sepia and octopus samples, but a distinct difference between sepias and squids, and octopuses and squids (Fig. 13).

Besides dicyemids, chromidinids were detected within two *Abdopus aculeatus* samples (OWR36, OWR18), one *Octopus vulgaris* sample (OVIN2), and one *Sepia officinalis* sample (AZTI1). In total, 14 species carried only dicyemids (see Chapter 4.2: Diversity within specific

host), two species carried both dicyemids and chromidinids (*O. vulgaris, S. officinalis*), and one species carried only chromidinids (*Abdopus aculeatus*); seven species hosted neither dicyemids nor chromidinids (*Octopus incella*, *Octopus laqueus*, *Metasepia tullbergi*, *Sepia recurvirostra*, *Euprymna scolopes*, *Stenoteuthis oulaniensis*, *Uroteuthis duvauceli*).

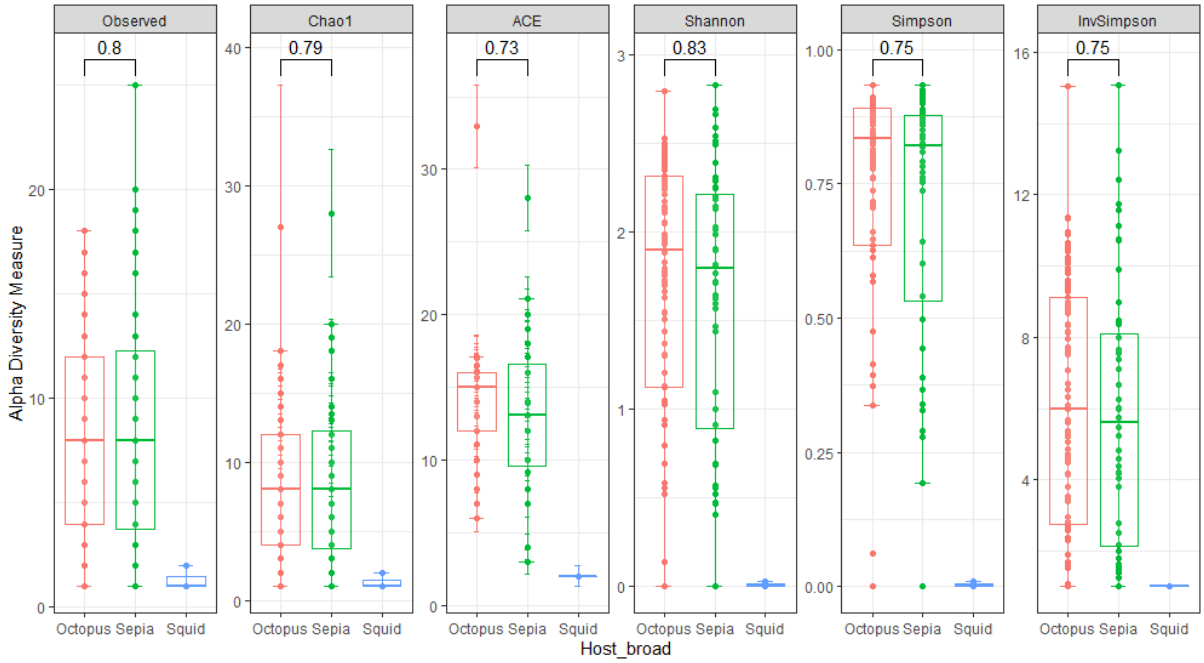
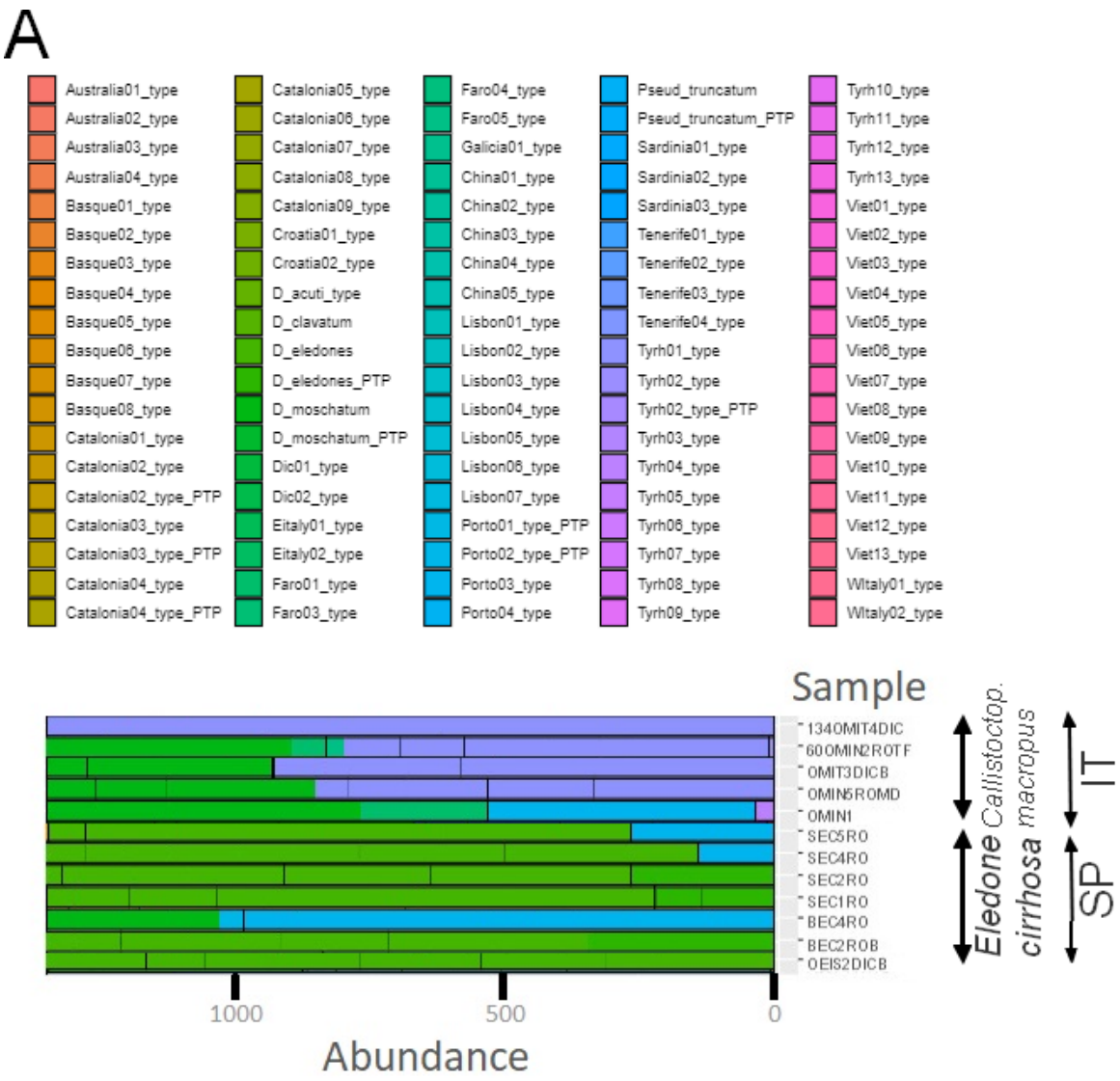
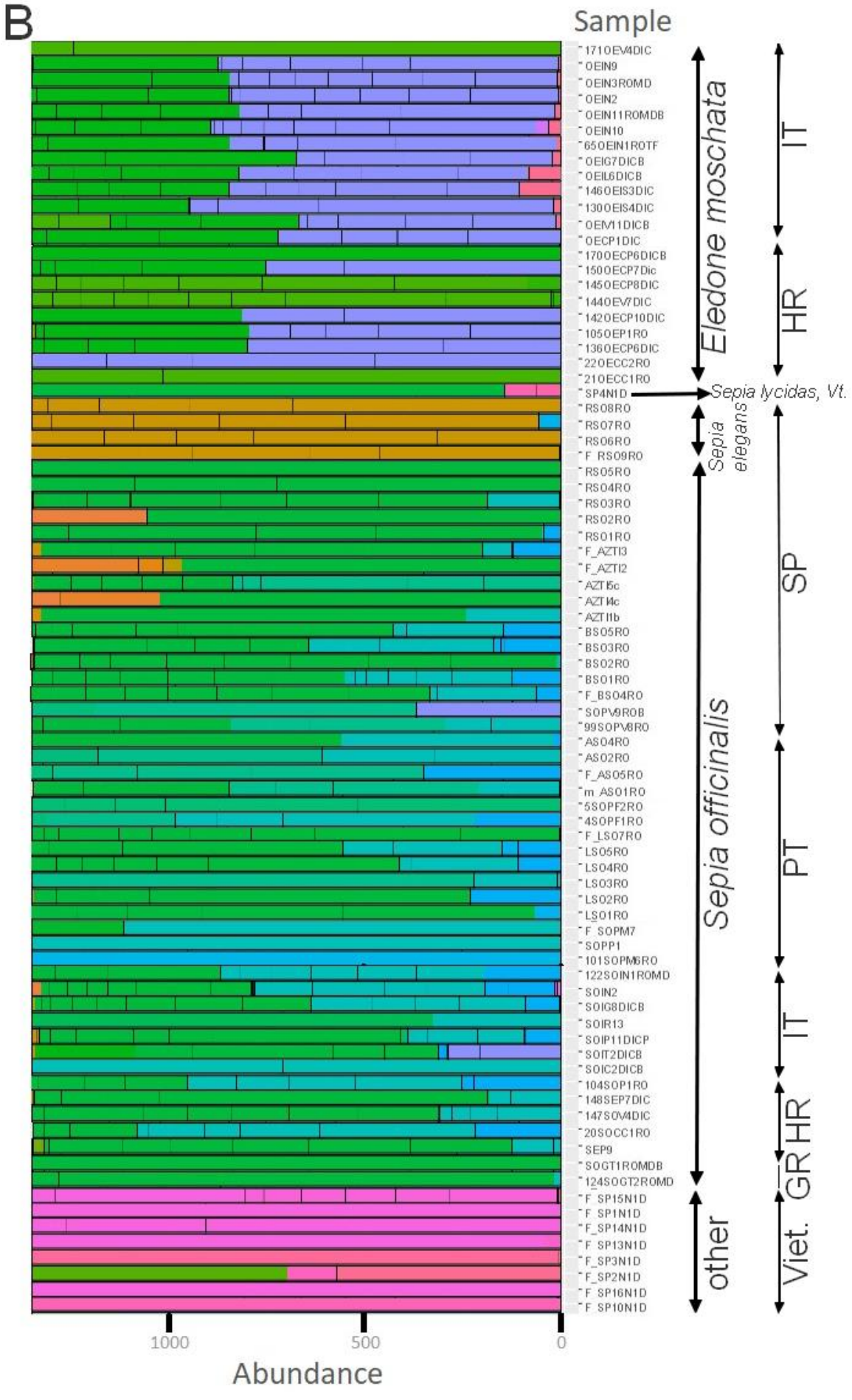


Fig. 13: Alpha diversity measures showed no significant difference between octopus' and sepia's diversity.

4.2 Diversity within specific hosts

Dicyemid types present within host individuals varied from 1 dicyemid type up to 8 dicyemid types. If we disregard types with low sequence count and considered only those with more than 100 copies of the ASV present, then the maximum number of types within one host was 4 (Fig. 14).




B
Sample
1710EV4DIC
0EIN9
0EIN3ROMD
0EIN2
0EIN11ROMDB
0EIN10
650EIN1ROTF
0EIG7DICB
0EIL6DICB
1460EIS3DIC
1300EIS4DIC
0EIV11DICB
0ECP1DIC
1700ECP6DICB
1500ECP7Dic
1450ECP8DIC
1440EV7DIC
1420ECP10DIC
1050EP1RO
1360ECP6DIC
220ECC2RO
210ECC1RO
SP4N1D
RS08RO
RS07RO
RS06RO
F RS09RO
RS05RO
RS04RO
RS03RO
RS02RO
RS01RO
F_AZTI3
F_AZTI2
AZTI5c
AZTI4c
AZTI1b
BS05RO
BS03RO
BS02RO
BS01RO
F_BS04RO
SOPV9ROB
99SOPV8RO
AS04RO
AS02RO
F_AS05RO
m AS01RO
5SOPF2RO
4SOPF1RO
F_LS07RO
LS05RO
LS04RO
LS03RO
LS02RO
LS01RO
F SOPM7
SOPP1
101SOPM6RO
122SOIN1ROMD
SOIN2
SOIG8DICB
SOIR13
SOIP11DICP
SOIT2DICB
SOIC2DICB
104SOP1RO
148SEP7DIC
147SOV4DIC
20SOCC1RO
SEP9
SOGT1ROMDB
124SOGT2ROMD
F SP15N1D
F SP1N1D
F_SP14N1D
F SP13N1D
F_SP3N1D
F_SP2N1D
F SP16N1D
F SP10N1D
1000
500
0
Abundance
Eledone moschata
Sepia lycidas, Vt.
Sepia elegans
Sepia officinalis
other
IT
HR
SP
PT
IT
HR
GR
Viet.

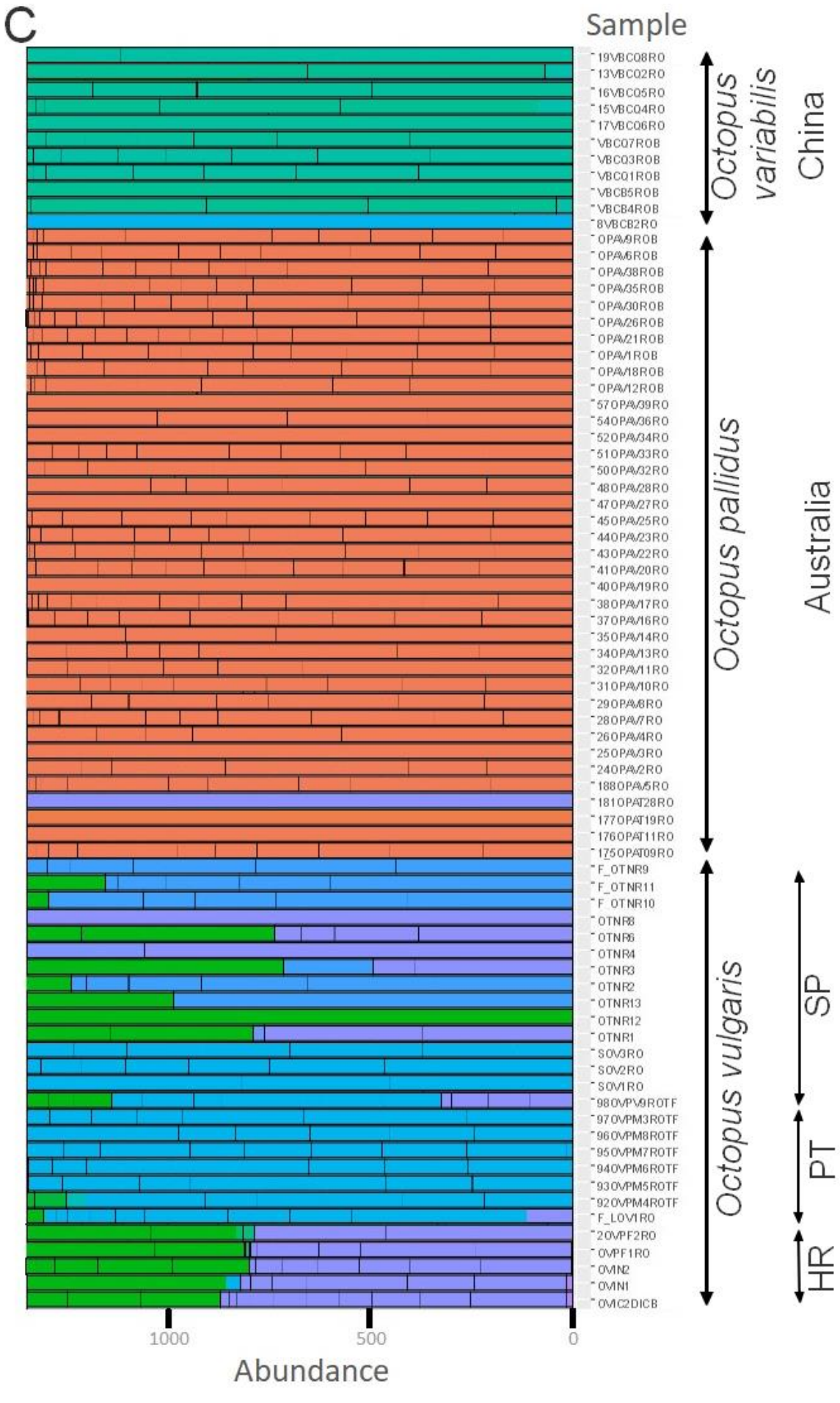

C
Sample
19VBC08RO
13VBC02RO
16VBC05RO
15VBC04RO
17VBC06RO
VBC07ROB
VBC03ROB
VBC01ROB
VBCB5ROB
VBCB4ROB
8VBCB2RO
OPAV9ROB
OPAV6ROB
OPAV38ROB
OPAV35ROB
OPAV30ROB
OPAV26ROB
OPAV21ROB
OPAV1ROB
OPAV18ROB
OPAV12ROB
57OPAV39RO
54OPAV36RO
52OPAV34RO
51OPAV33RO
50OPAV32RO
48OPAV28RO
47OPAV27RO
45OPAV25RO
44OPAV23RO
43OPAV22RO
41OPAV20RO
40OPAV19RO
38OPAV17RO
37OPAV16RO
35OPAV14RO
34OPAV13RO
32OPAV11RO
31OPAV10RO
29OPAV8RO
28OPAV7RO
26OPAV4RO
25OPAV3RO
24OPAV2RO
188OPAV5RO
181OPAT28RO
177OPAT19RO
176OPAT11RO
175OPAT09RO
F_OTNR9
F_OTNR11
F_OTNR10
OTNR8
OTNR6
OTNR4
OTNR3
OTNR2
OTNR13
OTNR12
OTNR1
SOV3RO
SOV2RO
SOV1RO
980VPV9ROTF
970VPM3ROTF
960VPM8ROTF
950VPM7ROTF
940VPM6ROTF
930VPM5ROTF
920VPM4ROTF
F_LOV1RO
20VPF2RO
OVPF1RO
OVIN2
OVIN1
OVIC2DICB
1000
500
0
Abundance
Octopus variabilis
China
Octopus pallidus
Australia
Octopus vulgaris
SP
PT
HR

Fig. 14: A - Dicyemid sequence types present through the bar plots pictured in A, B, C; diversity within *Callistoctopus macropus* and *Eledone cirrhosa*. B - Diversity within *Eledone moschata*, *Sepia lycidas*, *Sepia elegans*, *Sepia officinalis*, and Vietnamese samples. C - Diversity within *Octopus variabilis*, *Octopus pallidus*, and *Octopus vulgaris*. Each bar represents one sample.

4.2.1 *Octopus vulgaris*

Octopus vulgaris (*OV*) was sampled in the following regions: Portuguese Faro (n=2), Lisbon (n=1), and Porto district (n=6); Spanish Basque (n=3), Galicia (n=1), and Tenerife (n=11); Italian Tyrrhenian Sea (n=2), Sicily (n=1), and Sardinia (n=1).

Tab. 7: *Octopus vulgaris* and its dicyemid types: more than 100; 20-100; 1-20 present copies of ASV, marked with green (2), yellow (1), and red (0), respectively.

Country	Region	Sample ID	D_moschatum	D_moschatum_PTP	Faro01_type	Faro05_type	Lisbon01_type	Lisbon05_type	Lisbon07_type	Porto01_type_PTP	Porto03_type	Porto04_type	Tenerife01_type	Tenerife02_type	Tenerife03_type	Tenerife04_type	Tyrh01_type	Tyrh02_type	Tyrh03_type	Tyrh05_type	Tyrh08_type
PT	Faro district	OVPF 1	2		0	1											2	0			
PT	Faro district	OVPF 2	2		0	1											2				
PT	Porto district	OVPM 4		1			1				2										
PT	Porto district	OVPM 5									2										
PT	Porto district	OVPM 6								2	2										
PT	Porto district	OVPM 7									2	1									
PT	Porto district	OVPM 8									2										
PT	Porto district	OVPM 3									2										
PT	Lisbon district	LOV1	2					1	0		2						2				
SP	Galicia	OVPV 9	2								2						2				
SP	Basque country	SOV1									2										
SP	Basque country	SOV3									2										
SP	Basque country	SOV2									2										
SP	Tenerife	OTNR1	2											0	1		2				
SP	Tenerife	OTNR2	2										2								
SP	Tenerife	OTNR3	2										2				2				
SP	Tenerife	OTNR4														0	1				
SP	Tenerife	OTNR6	2														2				
SP	Tenerife	OTNR8															1				
SP	Tenerife	OTNR9											2								
SP	Tenerife	OTNR10	2										2								
SP	Tenerife	OTNR11	2										2								
SP	Tenerife	OTNR12	2																		
SP	Tenerife	OTNR13	2																		
IT	Sardinia	OVIC 2	2														2		1		
IT	Tyrrhenian sea	OVIN 1	2								2						2	0	1		0
IT	Tyrrhenian sea	OVIN 2	2														2			0	

Kruskal-Wallis statistical analysis revealed that dicyemid populations within *OV* differ significantly only from populations from Faro and Porto (p values in Attachments, table III). *OV* populations differ significantly from all the *Sepia officinalis* populations and *OV* from Faro differs significantly from *Eledone moschata* from Croatia.

There were 19 dicyemid types present (Tab. 7), including *D. moschatum*, which was previously unreported for *O. vulgaris*. A minimum of one type and a maximum of six types were present within one host.

4.2.2 *Eledone moschata*

Eledone moschata (*EM*) was sampled in the following regions: Italian Tyrrhenian Sea (n=8), Ligurian Sea (n=4), western Adriatic Sea (n=1), and Sardinia (n=2); Croatian Adriatic Sea (n=11).

Tab. 8: *Eledone moschata* and its dicyemid types: more than 100; 20-100; 1-20 present copies of ASV, marked with green (2), yellow (1), and red (0), respectively.

Country	Region	Sample ID	Croatia01_type	D_eledones	D_eledones_PTP	D_moschatum	WItaly01_type	WItaly02_type	Tyrh01_type	Tyrh06_type	Tyrh07_type	Tyrh13_type
HR	Adriatic sea east	OEP 1	0			2			2			
HR	Adriatic sea east	OECP 1				2			2			
HR	Adriatic sea east	OECP 6				2			2			
HR	Adriatic sea east	OECP 10				2			2			
HR	Adriatic sea east	OEV 7		2	0							
HR	Adriatic sea east	OECP 8		2	1							
HR	Adriatic sea east	OECP 7				2			2			
HR	Adriatic sea east	OECC 1		2					2			
IT	Adriatic sea west	OEIV 11		2		2	0		2			
IT	Tyrrhenian sea	OEIN 1				2	0		2			
IT	Tyrrhenian sea	OEIN 9				2	1		2			
IT	Tyrrhenian sea	OEIN 2				2	0		2	0		
IT	Tyrrhenian sea	OEIN 3				2	0		2			
IT	Tyrrhenian sea	OEIN 10				2		2	2		2	0
IT	Tyrrhenian sea	OEIN 11				2	0		2			
IT	Tyrrhenian sea	OEIG 7				2	1		2			
IT	Ligurian sea	OEIS 4				2	0		2			
IT	Ligurian sea	OEIL 6				2	2		2			
IT	Ligurian sea	OEIS 3				2		2	2			

Kruskal-Wallis statistical analysis revealed that dicyemid populations within *EM* did not differ significantly from each other (p values in Attachments, table III). *EM* populations differed significantly from *Sepia officinalis* populations from Greece, Galicia, and Faro; and *EM* from Croatia differs significantly from *Octopus vulgaris* from Faro.

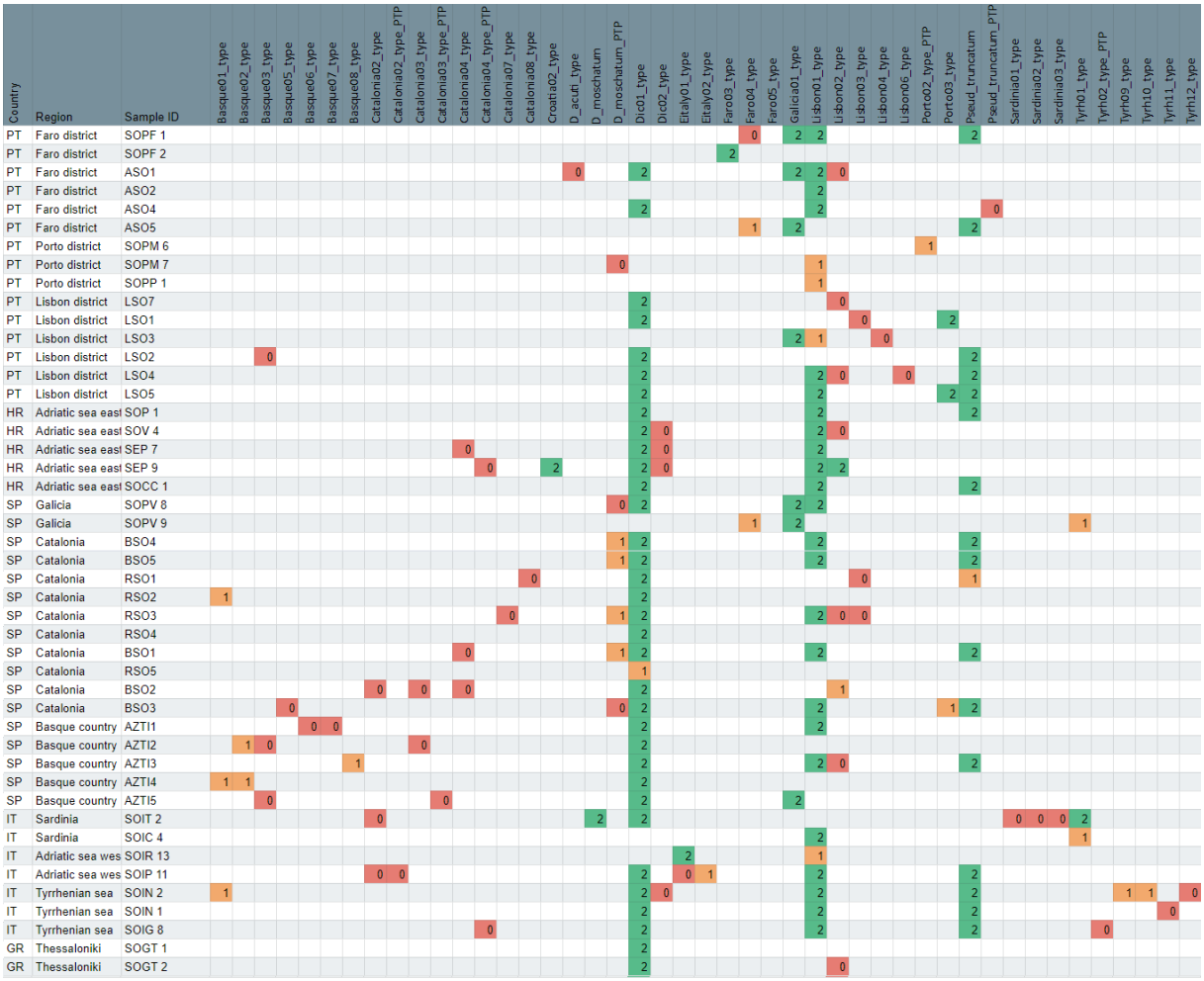
There were 10 types of dicyemids present. *Dicyema moschatum* and Tyrh01 types were dominant across the majority of samples. Within one sample, a minimum of two types and a maximum of five types were present (Tab. 8).

4.2.3 *Sepia officinalis*

Sepia officinalis (*SO*) was sampled in the following regions: Italian Tyrrhenian Sea (n=3), Western Adriatic Sea (n=2), and Sardinia (n=3); Croatian Adriatic Sea (n=5); Spanish Basque (n=6), Catalonia (n=10), and Galicia (n=2); Portuguese Faro (n=6), Lisbon (n=6), and Porto district (n=3); Greek Thessaloniki (n=2).

Kruskal-Wallis statistical analysis revealed that dicyemid populations within *SO* differed significantly between Greece and Lisbon, Greece and Faro, Faro and Catalonia, Catalonia and Galicia, Greece and Basque, and Croatia and Faro (p values in Attachments, table III). *SO* populations differed significantly from *Octopus vulgaris* and *Eledone moschata* populations as described above. There were 45 dicyemid types, varying between 1 type per host to 8 types per host (Tab. 9). Three types appeared in the majority of samples - Dic01_type, Lisbon01_type, and *Pseudicyema truncatum*.

Tab. 9: *Sepia officinalis* and its dicyemid types: more than 100; 20-100; 1-20 present copies of ASV, marked with green (2), yellow (1), and red (0), respectively.



Country	Region	Sample ID	Basque01_type	Basque02_type	Basque03_type	Basque05_type	Basque06_type	Basque07_type	Basque08_type	Catalonia02_type	Catalonia02_type_PTP	Catalonia03_type	Catalonia03_type_PTP	Catalonia04_type	Catalonia04_type_PTP	Catalonia07_type	Catalonia08_type	Croatia02_type	D_acuti_type	D_moschatum	D_moschatum_PTP	Dic01_type	Dic02_type	Eitaly01_type	Eitaly02_type	Faro03_type	Faro04_type	Faro05_type	Galicia01_type	Lisbon01_type	Lisbon02_type	Lisbon03_type	Lisbon04_type	Lisbon06_type	Porto02_type_PTP	Porto03_type	Pseud_truncatum	Pseud_truncatum_PTP	Sardinia01_type	Sardinia02_type	Sardinia03_type	Tyrh01_type	Tyrh02_type_PTP	Tyrh09_type	Tyrh10_type	Tyrh11_type	Tyrh12_type
PT	Faro district	SOPF 1																								0		2	2						2										
PT	Faro district	SOPF 2																								2																				
PT	Faro district	ASO1																	0			2							2	2	0																
PT	Faro district	ASO2																												2																	
PT	Faro district	ASO4																				2								2								0									
PT	Faro district	ASO5																								1			2							2										
PT	Porto district	SOPM 6																																1												
PT	Porto district	SOPM 7																			0									1																
PT	Porto district	SOPP 1																												1																
PT	Lisbon district	LSO7																				2									0															
PT	Lisbon district	LSO1																				2										0				2										
PT	Lisbon district	LSO3																											2	1			0													
PT	Lisbon district	LSO2			0																	2															2									
PT	Lisbon district	LSO4																				2								2	0				0		2								
PT	Lisbon district	LSO5																				2								2						2	2								
HR	Adriatic sea east	SOP 1																				2								2							2								
HR	Adriatic sea east	SOV 4																				2	0							2	0													
HR	Adriatic sea east	SEP 7												0								2	0							2															
HR	Adriatic sea east	SEP 9													0			2				2	0							2	2														
HR	Adriatic sea east	SOCC 1																				2								2							2								
SP	Galicia	SOPV 8																			0	2							2	2														
SP	Galicia	SOPV 9																									1		2													1				
SP	Catalonia	BSO4																			1	2								2							2								
SP	Catalonia	BSO5																			1	2								2							2								
SP	Catalonia	RSO1															0					2									0						1								
SP	Catalonia	RSO2	1																			2																							
SP	Catalonia	RSO3													0						1	2								2	0	0												
SP	Catalonia	RSO4																				2																							
SP	Catalonia	BSO1												0							1	2								2							2								
SP	Catalonia	RSO5																				1																							
SP	Catalonia	BSO2								0		0		0								2									1														
SP	Catalonia	BSO3				0															0	2								2						1	2								
SP	Basque country	AZTI1					0	0														2								2														
SP	Basque country	AZTI2		1	0								0									2																							
SP	Basque country	AZTI3							1													2								2	0						2								
SP	Basque country	AZTI4	1	1																		2																							
SP	Basque country	AZTI5			0								0									2							2																
IT	Sardinia	SOIT 2								0									2			2																	0	0	0	2				
IT	Sardinia	SOIC 4																												2											1				
IT	Adriatic sea wes	SOIR 13																					2							1														
IT	Adriatic sea wes	SOIP 11								0	0											2	0	1						2							2								
IT	Tyrrhenian sea	SOIN 2	1																			2	0							2							2						1	1		0
IT	Tyrrhenian sea	SOIN 1																				2								2							2								0	
IT	Tyrrhenian sea	SOIG 8													0							2								2							2						0			
GR	Thessaloniki	SOGT 1																				2																							
GR	Thessaloniki	SOGT 2																				2									0														

4.2.4 Other sampled hosts

Dicyemid types within the rest of the hosts (Fig. 14) were just briefly summed up below, due to low frequency of sampling among selected localities (details in Supplementary files - supplementary_tables.xlsx; list TYPES_samples; Fig. 15).

Callistoctopus macropus (n=5) hosted seven types. Two significant lineages were Tyrh_01 type and *D. moschatum* (unreported for *C. macropus* previously). In one sample there was always at least one type of dicyemid. Maximum of five types of dicyemids were present.

Octopus pallidus (n=37) hosted four types, Australia03 type was present in all but one specimen. In one sample there was always at least one type of dicyemid. Maximum of two types of dicyemids were present.

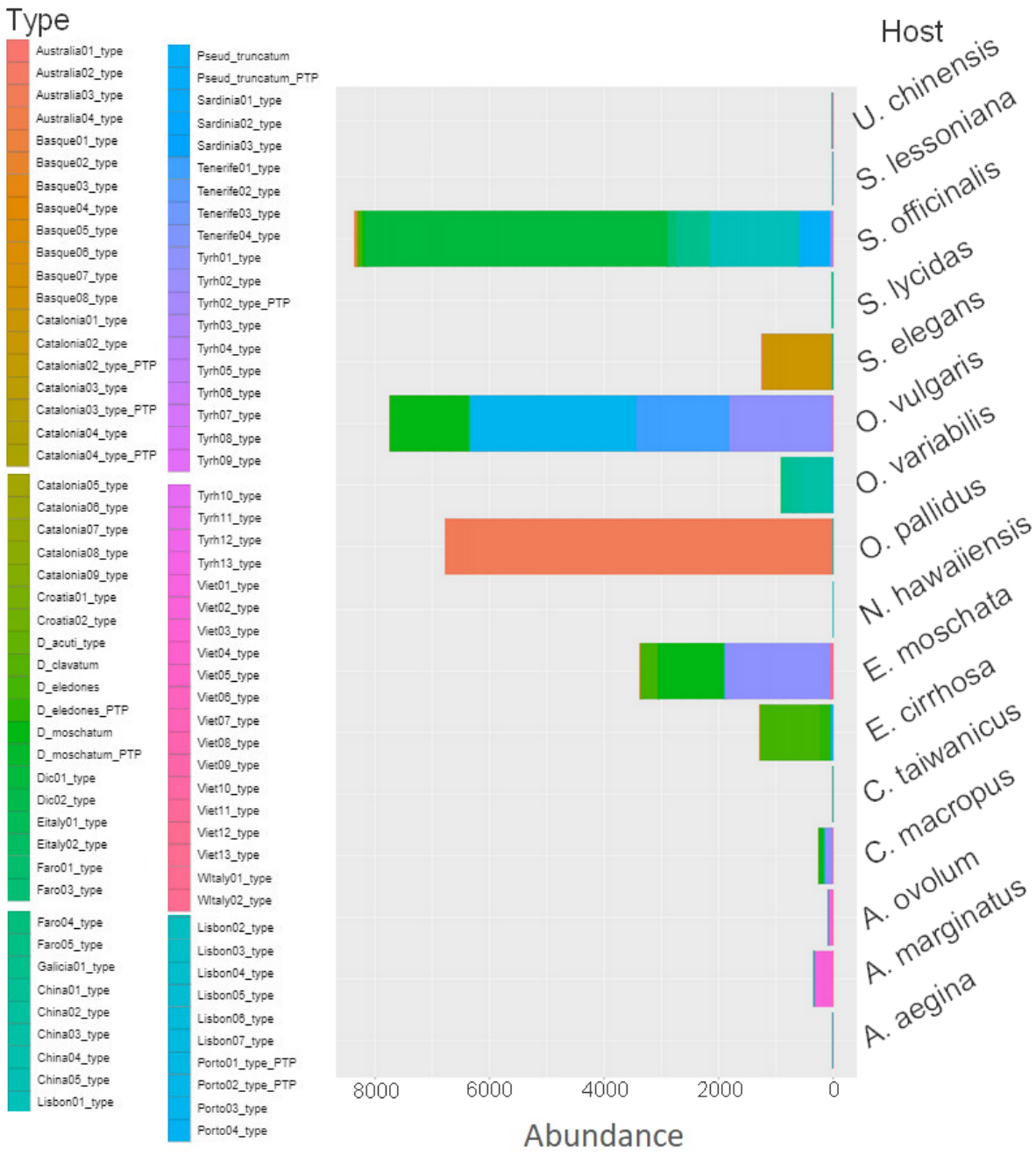


Fig. 15: Dicyemid diversity between host species.

Eledone cirrhosa (n=8) hosted seven types. *Dicyemennea eledones* was present in seven hosts in total. In one sample there was always at least one type of dicyemid. Maximum of three types of dicyemids were present.

Octopus variabilis (n=10) hosted five types. Types China01 and China02 were the most abundant ones. In one sample there was always at least one type of dicyemid. Maximum of three types of dicyemids were present.

Sepia elegans (n=4) hosted three types. Catalonia01 type was present in all *S. elegans* samples. In one sample there was always at least one type of dicyemid. Maximum of two types of dicyemids were present.

Amphioctopus marginatus (n=2) hosted seven types, only Viet01 type was present in both samples. In one sample there were always at least three types of dicyemid. Maximum of five types of dicyemids were present.

The rest of the hosts are from Vietnam and are only single-sampled (Tab. 10). *Nototodarus hawaiiensis* and *Uroteuthis chinensis* were not previously reported to have dicyemids, but both showed at least one type of dicyemid present.

Tab. 10: Individual samples from Vietnam and their dicyemid types: more than 100; 20-100; 1-20 present copies of ASV, marked with green (2), yellow (1), and red (0), respectively.

Host species	Province	Sample ID	D_clavatum	Dic01_type	Viet01_type	Viet02_type	Viet03_type	Viet04_type	Viet05_type	Viet06_type	Viet07_type	Viet08_type	Viet09_type	Viet10_type	Viet11_type	Viet12_type
Amphioctopus aegina	Phu Yen p.	SP13			1			0								
Amphioctopus ovolum	Phu Yen p.	SP14			2											
Sepia lycidas	Phu Yen p.	SP4		2								0	0			
Nototodarus hawaiiensis	Khanh Hoa p.	SP16			1											
Cistopus taiwanicus	Phu Yen p.	SP2								0						
Sepioteuthis lessoniana	Ninh Thuan p.	SP10									2					
Uroteuthis chinensis	Phu Yen p.	SP3													2	0

4.3 Diversity between sampled localities

The most significant difference between our sampled localities according to PCoA of alpha diversities was between the Australian *Octopus pallidus* and other octopuses and sepias (Fig. 16). Sepias clustered together regardless of their origin. Octopuses, except for Australians, formed a separate cluster as well. Permanova test (Attachments, table IV) showed a significant difference between octopuses in the following regions: Asia and Mediterranean, Asia and Australia, Australia and Mediterranean. All octopuses were significantly different from Mediterranean sepias. Squids were significantly different from Australian and Mediterranean octopuses, as well as Mediterranean sepias.

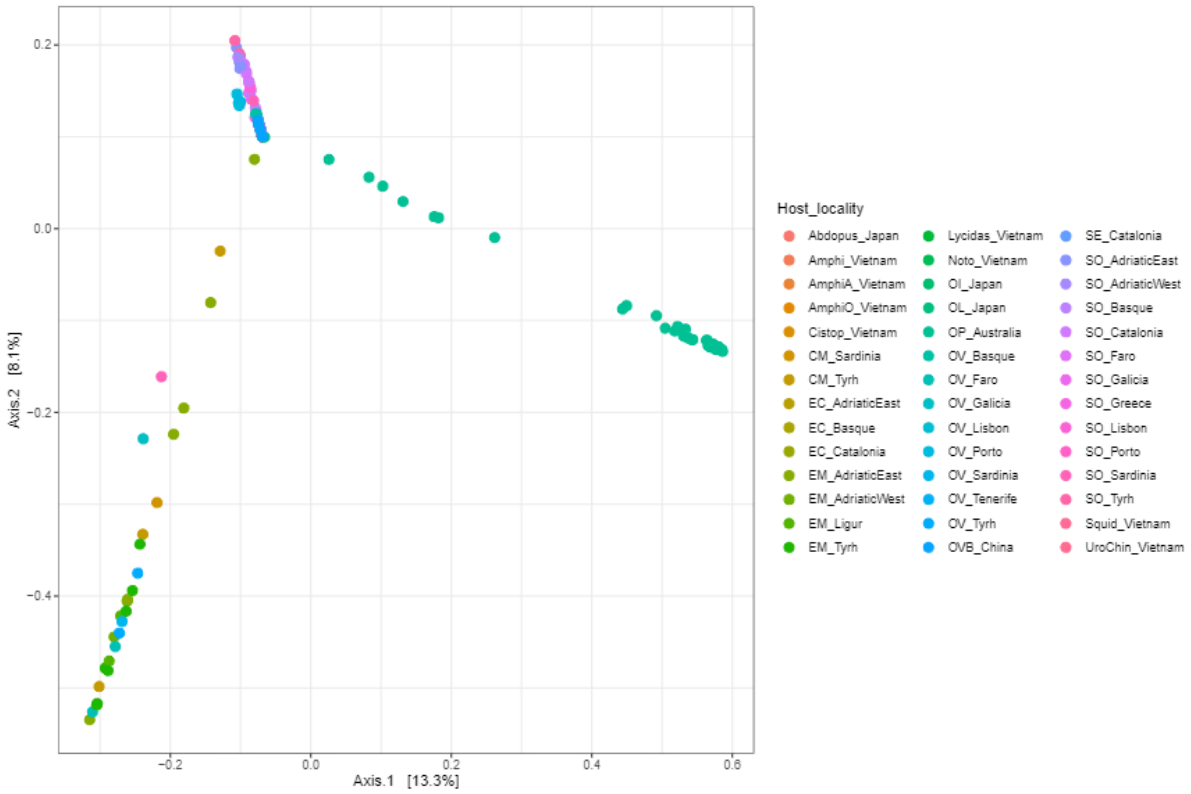


Fig. 16: Ordination plot showed a distinct difference between Sepia and Octopus samples from various localities. AmphiA= *Amphioctopus aegina*, AmphiO= *Amphioctopus ovolum*, Amphi= *Amphioctopus marginatus*, CM= *Callistoctopus macropus*, Cistop= Cistopus taiwanicus, EC= *Eledone cirrhosa*, EM= *Eledone moschata*, Lycidas= *Sepia lycidas*, Noto= *Nototodarus hawaiiensis*, OI= *Octopus incella*, OL= *Octopus laqueus*, OP= *Octopus pallidus*, OV= *Octopus vulgaris*, OVB= *Octopus variabilis*, SE= *Sepia elegans*, SO= *Sepia officinalis*, Squid= *Sepioteuthis lessoniana*, UroChin= *Uroteuthis chinensis*.

The network plot shows clustering (co-occurrence of sequence types) of dicyemid types (applied 20 % filtering threshold, Fig. 17). Opposite to the expected pattern of clustering (e.g., geographically proximate ASVs clustering together), we obtained different results of distant clustering of very closely related types of dicyemids (e.g., Chinese types).

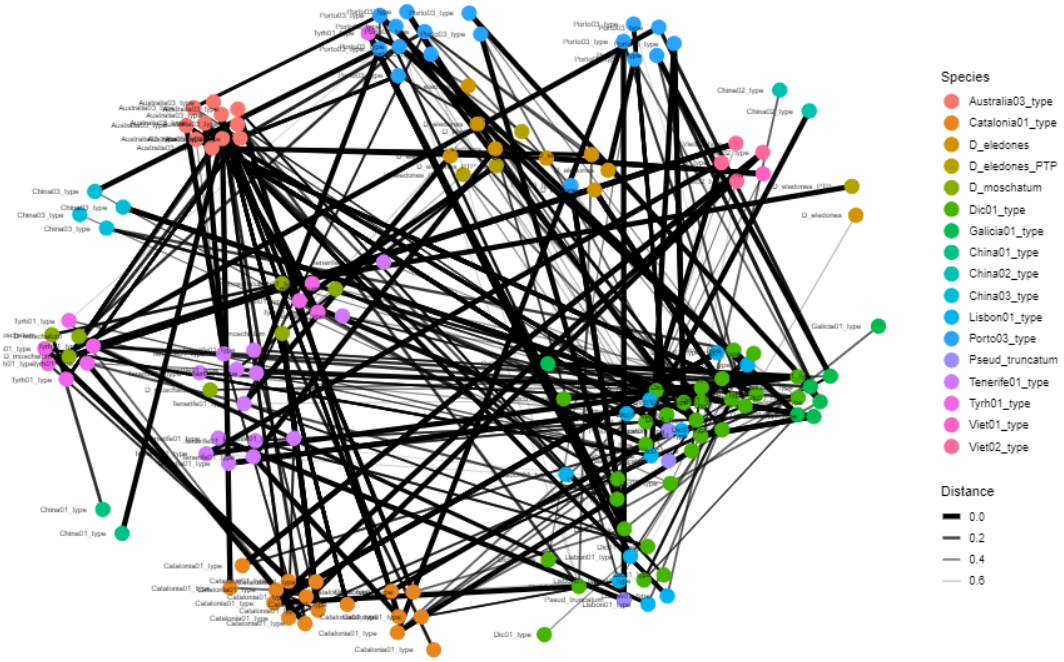


Fig. 17: Network plot showing Jaccard similarity based clustering of dicyemid types.

5 Discussion

Using amplicon sequencing approach for the first time in dicyemid parasites of cephalopods, it was possible to estimate dicyemid diversity independently of their morphological determination. Furthermore, it allowed investigation of a possible connection between several existing species and sequence-based dicyemid types established here. Lastly, the suitability of the used methodology is discussed as well as possible future direction of the research on dicyemids.

5.1 Estimate of total dicyemid diversity

Even though the available host sampling was limited to just a fraction of the global cephalopod biodiversity, the number of identified dicyemid types was relatively high. More specifically, the study included twelve octopus species out of 307 octopods reported globally, six sepia species out of 205 sepiids reported globally, and five squid species out of 300 teuthids reported globally (WoRMS, accessed March 2022); our sampling, therefore, represents 4 %, 3 %, and 1,7 % of the total octopod, sepiid, and teuthid diversity, respectively. The sampling of renal organs across selected hosts suggested that at least 37 dicyemid types exist (potentially up to 92, if all the types are to be considered credible). As the current estimate of all described dicyemid species is 121 based on examining ca 54 host species (Catalano et al., 2012; Catalano, 2013; Catalano & Furuya, 2013; Castellanos-Martinez et al., 2016), we believe that the number of known dicyemid species significantly under-represents their true diversity.

Dicyemids rarely occur within squid hosts, but in the presented study three out of five sampled hosts hosted an abundant number of dicyemids; two of them were not previously reported to host dicyemids (*Uroteuthis chinensis* and *Nototodarus hawaiiensis*). That might suggest dicyemid infection among teuthid hosts might be more common than previously reported.

In Europe, 112 specimens were sampled (mostly of commercial importance), eight of them produced no dicyemid reads, suggesting a 93 % prevalence (opposed to Nouvel's 1947, who claimed 100% prevalence). According to the study, the suggested prevalence for China (two out of 12) was 83 % (prevalence previously unreported). The obtained prevalence for Australia (11 out of 48) was 77 % (previously reported prevalence 24 - 100 % according to Catalano et al., 2014; 60 % for *O. pallidus* reported by Finn et al., 2005), however, most (eleven) of the Australian dicyemid-free specimens are from one sampling locality, which proved problematic

during PCR. If those samples were excluded, the prevalence rate would increase to 89 % (33 out of 37).

5.2 Known dicyemid species versus dicyemid sequence types

Current literature (Hochberg 1990; Furuya 1999; Fururya & Souidenne, 2019) describes co-occurrence of several dicyemid species at least in several cephalopod hosts (relevant ones for the analysis see Tab. 11). In other cases (*O. incella, O pallidus, O. variabilis* etc.), there are only mentions of the presence of unspecified dicyemids. However, in the case of previously described dicyemid species, it might be possible to try matching herein presented results to the respective types in certain cephalopod hosts. For example, the type Porto 03 occurs only within *E. cirrhosa* and *O. vulgaris*, just like the previously described *Dicyemennea lameerei*. Similarly, Tyrh 01 type occurs in *C. macropus*, *O. vulgaris*, and *E. moschata* whilst *Dicyema paradoxum* can be found in the first two mentioned above. Type Catalonia 01 is found only within *Sepia elegans*, which leads us to believe it might be either *D. schulzianum* or *D. macrocephalum*. Also, type Lisbon 01 can be found only in *S. officinalis*, suggesting possible ties to one of the *S. officinalis*' described dicyemids. Contrary to the cases discussed above, Vietnam 07 type found in *Sepioteuthis lessoniana* does not correspond to *Dicyema orientale* previously reported for this cephalopod, as Genbank accession of *D. orientale* was included in the analysis and it did not form a cluster together with Vietnam 07 type. However, these are only assumptions, and morphological assessment would be crucial to determine whether there is any connection between described species and presented types.

Tab. 11: Dicyemid species and their hosts described in current literature, and the respective most prevalent sequence types.

Cephalopod host	Dicyemids described	Prevalent types
Sepia elegans	*D. schulzianum*	Catalonia 01
	D. macrocephalum	
Eledone moschata	*D. moschatum*	Tyrh 01, *D. moschatum*, *D. eledones*
	D. eledones	
Sepia officinalis	*D. gracile*	Dic 01, Lisbon 01, *P. truncatum*
	P. truncatum	
	Microcyema vespa	
	D. whitmani	

Eledone cirrhosa	*D. lameerei*	*D. eledones*, Porto 03
	D. eledones	
Sepia lycidas	*Pseudicyema nakaoi*	Dic 01, Vietnam 08, 09
	D. lycidoeceum	
Sepioteuthis lessoniana	*D. orientale*	Vietnam 07
Callistoctopus macropus	*D. paradoxum*	*D. moschatum*, Tyrh 01
Octopus vulgaris	*D. paradoxum*	Porto 03, Tyrh 01, *D. moschatum*, Tenerife 01
	D. lameerei	
	Conocyema polymorpha	
	**D. megalocephalum*	
	**D. typus*	
	**D. misakiense*	
	**D. japonicum*	
	**D. acuticephalum*	
	**D. monodi*	
	**D. bilobum*	
	**D. aegira*	

*found in non-European waters.

Here we found up to eight dicyemid types (see chapter 4.2: Diversity within specific hosts) within one host, opposite to previous studies based on morphology (Furuya et al., 2004; Furuya & Souidenne 2019), presenting only up to three dicyemid types within one host. The high prevalence of *P. truncatum* sequence type within most sepia samples in the study confirmed Furuya & Souidenne (2019) findings. Several species of cephalopods hosted no dicyemids (*Octopus incella*, *Octopus laqueus*, *Sepia recurvirostra*, *Stenoteuthis oulaniensis*, *Uroteuthis duvauceli*, *Euprymna scolopes*, and *Metasepia tullbergi*), however, due to the insufficient sampling depth of most of them (between one and seven indviduals), it can not be confirmed whether it was caused by the actual absence of dicyemids or if it was only caused by sampling error.

5.3 Employed methodology

Generally, the accuracy of the results of an amplicon study depends on choosing both the optimal pipeline and database (Pauvert et al., 2019; Smith et al., 2020). For example, the same

software and parameters can not be applied to fungal ITS, bacterial 16S, or eukaryotic 18S communities. The choice of the applied pipeline depends on many variables (e.g. sequencing platform, read-assembly method and overlap length, selected output of either OTUs or ASVs). Selecting a suitable database could face a lot of obstacles, as many of them are not curated, not updated, or simply do not include sufficient coverage of the organisms of interest. However, the herein presented approach was validated by multiple studies using DNA extraction from the host with the intention to analyze its parasite infection (Cooper et al., 2018; Chaudhry et al., 2019; Hammoud et al, 2021). By implementing heterogeneity primers, a sequencing failure due to low PhiX spike-in was avoided (similarly to Chaudhry et al., 2019), and following Hammoud et al. (2021) suggestions our own reads were implemented in order to create a relevant database for our data.

The main disadvantage of the employed methodology was the absence of morphological description of sequenced species, due to the fact that samples were generally either taken under field conditions not allowing fixation of dicyemid individuals for slide mounting or shipped to us as frozen host tissues. Furthermore, relying only on genetic information can create false-negative samples due to improper DNA extraction (due to possible DNA inhibitors not being removed properly, etc.) or due to primer site mutations. The combination of a morphological description and presented species delimitation results could have matched previously described species to the dicyemid types. Also, the sampling methodology remains inconsistent within various research groups. Standardizing the sampling process (for example guidelines on reporting cephalopod cause of death, complete health assessment, including notion of parasitic infections and if possible, complementing morphological assessment of the present parasites with molecular data) as proposed by Roumbedakis et al. (2018) would improve and facilitate research of both the dicyemids and cephalopods in general.

5.4 Future research areas

Furuya & Souidenne (2019) suggested that a mutualist-symbiotic way of life might have been adopted by dicyemids, transferring from a previous strategy more harmful to the host. Considering the advantage of the host having symbiotic dicyemids for aiding with urine excretion (as suggested by Lapan, 1975), the high prevalence - reported elsewhere and confirmed here - seems reasonable. Furthermore, varying dicyemid prevalence rates between geographical localities were also recorded here, which could be tied to the occurrence and/or rareness of the host, or its population size. Locally common hosts might allow for larger

population sizes (and higher prevalences) of their dicyemids. On the contrary, hosts with fragmented distribution and/or less dense populations provide fewer opportunities for their parasites to propagate. Dicyemids' dispersal depends on both the dispersal and migration of their cephalopod hosts. This fact is the starting point for a future study to compare the dicyemid amplicon data with the host COI data defining the depth of population structure (similar to a trematode study by Blasco-Costa & Poulin, 2013).

In addition to the presented dicyemid dataset, data about cephalopod microbiomes were also obtained, using the same universal 16S primers. The bacterial microbiome can provide information about feeding patterns, fat storage, and host metabolism (Gill et al., 2006). It could be an indicator of general health and activity of the host (Heitlinger et al., 2017). Employing the obtained bacterial sequence data could provide additional insight into cephalopod microbiome studied before (Pernice et al., 2006; Farto et al., 2019; Lutz et al., 2019).

Possible future studies may focus on a more thorough comparison between morphological traits and dicyemid genotypes, as the current publicly-available molecular data proved to be insufficient. Establishing a connection between host COI and dicyemid 18S rDNA could also resolve many questions regarding dicyemid diversity and infection pattern. Moreover, bacterial microbiome in the renal organ could also affect the diversity or infection patterns of dicyemid populations. Close ties between bacteriome and commensal or mutualistic eukaryome are known from other systems (Lukeš et al., 2015).

6 Conclusion

The results of our amplicon study suggest that colonization of a host by multiple species is common with up to eight dicyemid types coexisting within the renal organ. The species delimitation results showed 37 (up to 92) dicyemid types present across 227 cephalopod accessions sampled globally, suggesting a possible under-estimation of the true dicyemid diversity by the currently described dicyemid species. Measures of dicyemid diversity between various localities, as well as between various hosts, showed distinct patterns, suggesting that dicyemid composition varies significantly. Prevalences in the Mediterranean, China, and Australia were estimated, showing 77 - 93% prevalence. The high prevalence rate might suggest a symbiotic rather than parasitic life strategy of dicyemids. Based on this thesis, future studies might focus on a microbiome composition within cephalopod hosts and on a thorough analysis of hosts' and dicyemids' genetic markers. We have provided molecular resources which could be used for further research on the connection between currently known species and genetic types established in this study.

7 Bibliography

Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., Kightley, E. P., Thompson, L. R., Hyde, E. R., Gonzalez, A., & Knight, R. (2017). Deblur rapidly resolves single-nucleotide community sequence patterns. *MSystems*, *2*(2). https://doi.org/10.1128/msystems.00191-16

Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ansorge, R., Birolo, G., James, S. A., & Telatin, A. (2021). Dadaist2: A Toolkit to automate and simplify statistical analysis and plotting of metabarcoding experiments. *International Journal of Molecular Sciences*, *22*(10), 5309. https://doi.org/10.3390/ijms22105309

Awata, H., Noto, T., & Endoh, H. (2005). Differentiation of somatic mitochondria and the structural changes in mtDNA during development of the dicyemid *Dicyema japonicum* (Mesozoa). *Molecular Genetics and Genomics*, *273*(6), 441–449. https://doi.org/10.1007/s00438-005-1157-2

Blasco-Costa, I., & Poulin, R. (2013). Host traits explain the genetic structure of parasites: A meta-analysis. *Parasitology*, *140*(10), 1316–1322. https://doi.org/10.1017/s0031182013000784

Bleidorn, C. (2019). Recent progress in reconstructing lophotrochozoan (Spiralian) phylogeny. *Organisms Diversity & Evolution*, *19*(4), 557–566. https://doi.org/10.1007/s13127-019-00412-4

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., . . . Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., …

Drummond, A. J. (2019). Beast 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology*, *15*(4). https://doi.org/10.1371/journal.pcbi.1006650

Bucklin, A., Steinke, D., & Blanco-Bercial, L. (2011). DNA barcoding of Marine Metazoa. *Annual Review of Marine Science*, *3*(1), 471–508. https://doi.org/10.1146/annurev-marine-120308-080950

Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge – accurate paired shotgun read merging via overlap. *PLOS ONE*, *12*(10). https://doi.org/10.1371/journal.pone.0185056

Callahan, B. J., McMurdie, P. J., & Holmes, S. P. (2017). Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal*, *11*(12), 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., & Holmes, S. P. (2016). Dada2: High-resolution sample inference from Illumina Amplicon Data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina Hi-seq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, *108*(Supplement_1), 4516–4522. https://doi.org/10.1073/pnas.1000080107

Castellanos-Martinez, S., Aguirre-Macedo, M. L., & Furuya, H. (2016). Two new species of dicyemid mesozoans (Dicyemida: Dicyemidae) from *Octopus maya* Voss & Solis-Ramirez (Octopodidae) off Yucatan, Mexico. *Systematic Parasitology*, *93*(6), 551–564. https://doi.org/10.1007/s11230-016-9644-x

Catalano, S. R. (2012). A review of the families, genera, and species of Dicyemida van Beneden, 1876. *Zootaxa*, *3479*(1), 1. https://doi.org/10.11646/zootaxa.3479.1.1

Catalano, S. R. (2013). Five new species of dicyemid mesozoans (Dicyemida: Dicyemidae) from two Australian cuttlefish species, with comments on dicyemid fauna composition. *Systematic Parasitology*, *86*(2), 125–151. https://doi.org/10.1007/s11230-013-9443-6

Catalano, S. R., & Furuya, H. (2013). Two new species of Dicyemid (Dicyemida: Dicyemidae) from two Australian cephalopod species: *Sepioteuthis australis* (Mollusca: Cephalopoda: Loliginidae) and *Sepioloidea lineolata* (Mollusca: Cephalopoda: Sepiadariidae). *Journal of Parasitology*, *99*(2), 203–211. https://doi.org/10.1645/ge-3252.1

Catalano, S. R., Whittington, I. D., Donnellan, S. C., & Gillanders, B. M. (2013). Using the Giant Australian Cuttlefish (*Sepia apama*) mass breeding aggregation to explore the life cycle of dicyemid parasites. *Acta Parasitologica*, *58*(4). https://doi.org/10.2478/s11686-013-0186-y

Catalano, S. R., Whittington, I. D., Donnellan, S. C., & Gillanders, B. M. (2014). Dicyemid fauna composition and infection patterns in relation to Cephalopod host biology and ecology. *Folia Parasitologica*, *61*(4), 301–310. https://doi.org/10.14411/fp.2014.034

Chaudhry, U., Ali, Q., Rashid, I., Shabbir, M. Z., Ijaz, M., Abbas, M., Evans, M., Ashraf, K., Morrison, I., Morrison, L., & Sargison, N. D. (2019). Development of a deep amplicon sequencing method to determine the species composition of Piroplasm Haemoprotozoa. *Ticks and Tick-Borne Diseases*, *10*(6), 101276. https://doi.org/10.1016/j.ttbdis.2019.101276

Cooper, C., Keatley, S., Northover, A., Gofton, A. W., Brigg, F., Lymbery, A. J., Pallant, L., Clode, P. L., & Thompson, R. C. A. (2018). Next generation sequencing reveals widespread trypanosome diversity and polyparasitism in marsupials from Western Australia. *International Journal for Parasitology: Parasites and Wildlife*, *7*(1), 58–67. https://doi.org/10.1016/j.ijppaw.2018.01.005

Crainey, J. L., Wilson, M. D., & Post, R. J. (2009). An 18S ribosomal DNA barcode for the study *Ofisomermis lairdi*, a parasite of the *Blackflysimulium damnosum* S.L. *Medical and Veterinary Entomology*, *23*(3), 238–244. https://doi.org/10.1111/j.1365-2915.2009.00814.x

DADA2 Pipeline Tutorial (1.16). Retrieved September 20, 2021, from https://benjjneb.github.io/dada2/tutorial.html

de Muinck, E. J., Trosvik, P., Gilfillan, G. D., Hov, J. R., & Sundaram, A. Y. (2017). A novel ultra high-throughput 16S rRNA gene amplicon sequencing library preparation method for the Illumina HiSeq platform. *Microbiome*, *5*(1). https://doi.org/10.1186/s40168-017-0279-1

Drábková, M., Flegrová, T., Myšková, E., Hypša, V., & Štefka, J. (2021). Genetic analysis of dicyemid infrapopulations suggests sexual reproduction and host colonization by multiple individuals is common. *Organisms Diversity & Evolution*. https://doi.org/10.1007/s13127-021-00493-0

Drábková, M., Jachníková, N., Tyml, T., Sehadová, H., Ditrich, O., Myšková, E., Hypša, V., & Štefka, J. (2019). Population co-divergence in common cuttlefish (*Sepia officinalis*) and its dicyemid parasite in the Mediterranean Sea. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-50555-9

Dunn, C. W., Giribet, G., Edgecombe, G. D., Hejnol, A. (2014). Animal phylogeny and its evolutionary implications. *Annual Review of Ecology, Evolution, and Systematics*, *45*(1), 371–395. https://doi.org/10.1146/annurev-ecolsys-120213-091627

Edgar, R. C. (2016). UNOISE2: Improved error correction for Illumina 16S and its amplicon sequencing. *BioRxiv*. https://doi.org/10.1101/081257

Eren, A. M., Maignien, L., Sul, W. J., Murphy, L. G., Grim, S. L., Morrison, H. G., & Sogin, M. L. (2013). Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods in Ecology and Evolution*, *4*(12), 1111–1119. https://doi.org/10.1111/2041-210x.12114

Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2014). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, *9*(4), 968–979. https://doi.org/10.1038/ismej.2014.195

Eshragh, R., & Leander, B. S. (2014). Molecular contributions to species boundaries in dicyemid parasites from eastern Pacific cephalopods. *Marine Biology Research*, *11*(4), 414–422. https://doi.org/10.1080/17451000.2014.943241

Evans, K. M., Wortley, A. H., & Mann, D. G. (2007). An assessment of potential diatom “barcode” genes (COX1, RBCL, 18s and its rDNA) and their effectiveness in determining

relationships in Sellaphora (Bacillariophyta). *Protist*, *158*(3), 349–364. https://doi.org/10.1016/j.protis.2007.04.001

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ezard T., Fujisawa T., & Barraclough T.G. (2009). SPLITS: SPecies' LImits by Threshold Statistics. R package version 1.0-20/r45. Available from: URL http://R-Forge.R-project.org/projects/splits/, last accessed February 2022

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq Platform. *Microbiome*, *2*(1). https://doi.org/10.1186/2049-2618-2-6

Farto, R., Fichi, G., Gestal, C., Pascual, S., & Nieto, T. P. (2019). Bacteria-affecting cephalopods. *Handbook of Pathogens and Diseases in Cephalopods*, 127–142. https://doi.org/10.1007/978-3-030-11330-8_8

Finn, J. K., Hochberg, F. G., & Norman, M. D. (2005). Phylum Dicyemida in Australian waters: first record and distribution across diverse cephalopod hosts. *Phuket Mar. Biol. Cent. Res. Bull*, (66), 83–96.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology*, *3*(5), 294–299.

Furuya, H. (1999). Fourteen new species of Dicyemid mesozoans from six Japanese cephalopods, with comments on host specificity. *Species Diversity*, *4*(2), 257–319. https://doi.org/10.12782/specdiv.4.257

Furuya, H. (2002). The biology of dicyemid mesozoans. *ZOOLOGICAL SCIENCE*, *19*(12), 1398–1399. https://doi.org/10.2108/zsj.19.1398

Furuya, H., Ota, M., Kimura, R., & Tsuneki, K. (2004). Renal organs of cephalopods: A habitat for dicyemids and chromidinids. *Journal of Morphology*, *262*(2), 629–643. https://doi.org/10.1002/jmor.10265

Furuya, H., & Souidenne, D. (2019). Dicyemids. *Handbook of Pathogens and Diseases in Cephalopods*, 159–168. https://doi.org/10.1007/978-3-030-11330-8_11

Gill, S. R., Pop, M., DeBoy, R. T., Eckburg, P. B., Turnbaugh, P. J., Samuel, B. S., Gordon, J. I., Relman, D. A., Fraser-Liggett, C. M., & Nelson, K. E. (2006). Metagenomic analysis of the human distal gut microbiome. *Science*, *312*(5778), 1355–1359. https://doi.org/10.1126/science.1124234

Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., & Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and Tools. *Journal of Biotechnology*, *261*, 169–176. https://doi.org/10.1016/j.jbiotec.2017.06.1198

Hafner, M. S., Sudman, P. D., Villablanca, F. X., Spradling, T. A., Demastes, J. W., & Nadler, S. A. (1994). Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science*, *265*(5175), 1087–1090. https://doi.org/10.1126/science.8066445

Hamady, M., & Knight, R. (2009). Microbial community profiling for human microbiome projects: Tools, techniques, and challenges. *Genome Research*, *19*(7), 1141–1152. https://doi.org/10.1101/gr.085464.108

Hammoud, C., Mulero, S., Van Bocxlaer, B., Boissier, J., Verschuren, D., Albrecht, C., & Huyse, T. (2021). Simultaneous genotyping of snails and infecting trematode parasites using high-throughput amplicon sequencing. *Molecular Ecology Resources*, *22*(2), 567–586. https://doi.org/10.1111/1755-0998.13492

Heitlinger, E., Ferreira, S. C., Thierer, D., Hofer, H., & East, M. L. (2017). The intestinal eukaryotic and bacterial biome of spotted hyenas: The impact of social status and age on diversity and composition. *Frontiers in Cellular and Infection Microbiology*, *7*. https://doi.org/10.3389/fcimb.2017.00262

Hochberg, F. G. (1990). “Diseases of Mollusca: Cephalopoda: diseases caused by protistans and metazoans,” in Diseases of Marine Animals, Vol. III, Cephalopoda to Urochordata, ed. O. Kinne (Hamburg: Biologisches Anstalt Helgoland), 47–227

Katayama, T., Wada, H., Furuya, H., Satoh, N., & Yamamoto, M. (1995). Phylogenetic position of the DICYEMID Mesozoa inferred from 18S rDNA sequences. *The Biological Bulletin*, *189*(2), 81–90. https://doi.org/10.2307/1542458

Kopylova, E., Navas-Molina, J. A., Mercier, C., Xu, Z. Z., Mahé, F., He, Y., Zhou, H.-W., Rognes, T., Caporaso, J. G., & Knight, R. (2016). Open-source sequence clustering methods improve the state of the art. *MSystems*, *1*(1). https://doi.org/10.1128/msystems.00003-15

Lapan, E. A., & Morowitz, H. J. (1975). The DICYEMID mesozoa as an integrated system for morphogenetic studies. I. Description Isolation and maintenance. *Journal of Experimental Zoology*, *193*(2), 147–159. https://doi.org/10.1002/jez.1401930204

Lu, T.-M., Kanda, M., Furuya, H., & Satoh, N. (2019). Dicyemid mesozoans: A unique parasitic lifestyle and a reduced genome. *Genome Biology and Evolution*, *11*(8), 2232–2243. https://doi.org/10.1093/gbe/evz157

Lu, T.-M., Kanda, M., Satoh, N., & Furuya, H. (2017). The phylogenetic position of dicyemid mesozoans offers insights into Spiralian Evolution. *Zoological Letters*, *3*(1). https://doi.org/10.1186/s40851-017-0068-5

Lukeš, J., Stensvold, C. R., Jirků-Pomajbíková, K., & Wegener Parfrey, L. (2015). Are human intestinal eukaryotes beneficial or commensals? *PLOS Pathogens*, *11*(8). https://doi.org/10.1371/journal.ppat.1005039

Lutz, H. L., Ramírez-Puebla, S. T., Abbo, L., Durand, A., Schlundt, C., Gottel, N. R., Sjaarda, A. K., Hanlon, R. T., Gilbert, J. A., & Mark Welch, J. L. (2019). A simple microbiome in the European common cuttlefish, *Sepia officinalis*. *MSystems*, *4*(4). https://doi.org/10.1128/msystems.00177-19

MacConaill, L. E., Burns, R. T., Nag, A., Coleman, H. A., Slevin, M. K., Giorda, K., Light, M., Lai, K., Jarosz, M., McNeill, M. S., Ducar, M. D., Meyerson, M., & Thorner, A. R. (2018). Unique, dual-indexed sequencing adapters with UMIS effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-017-4428-5

McConnaughey, B.H. (1951) The life cycle of the dicyemid Mesozoa. *Univ. Calif. Publ. Zool.*, 55: 295–336

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Metagenomic Library Prep Guide - Illumina, Inc.. (n.d.). Retrieved February 18, 2021, from https://support.illumina.com/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Miseq system. Focused power for targeted gene and small genome sequencing. Retrieved October 12, 2021, from https://www.illumina.com/systems/sequencing-platforms/miseq.html

Nylander, J.A.A. (2004). MrAIC.pl v1.4.6. Evolutionary Biology Centre, Uppsala University

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution*, *7*(4), 114–118. https://doi.org/10.1016/0169-5347(92)90144-z

Parada, A. E., Needham, D. M., & Fuhrman, J. A. (2015). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series, and Global Field samples. *Environmental Microbiology*, *18*(5), 1403–1414. https://doi.org/10.1111/1462-2920.13023

Parikh, H. I., Koparde, V. N., Bradley, S. P., Buck, G. A., & Sheth, N. U. (2016). MeFiT: Merging and filtering tool for Illumina paired-end reads for 16S rRNA amplicon sequencing. *BMC Bioinformatics*, *17*(1). https://doi.org/10.1186/s12859-016-1358-1

Pauvert, C., Buee, M., Laval, V., Edel-Hermann, V., Fauchery, L., Gautier, A., Lesur, I., Vallance, J. & Vacher, C. (2019). Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecology*, 23 - 33. ff10.1016/j.funeco.2019.03.005ff. ffhal-02627344f

Pernice, M., Pichon, D., Domart-Coulon, I., Favet, J., & Boucher-Rodoni, R. (2006). Primary co-culture as a complementary approach to explore the diversity of bacterial associations in marine invertebrates: The example of *Nautilus macromphalus* (Cephalopoda: Nautiloidea). *Marine Biology*, *150*(5), 749–757. https://doi.org/10.1007/s00227-006-0413-2

Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA

taxonomy of undescribed insects. *Systematic Biology*, *55*(4), 595–609. https://doi.org/10.1080/10635150600852011

Poulin, R., & Mouillot, D. (2003). Parasite specialization from a phylogenetic perspective: A new index of host specificity. *Parasitology*, *126*(5), 473–480. https://doi.org/10.1017/s0031182003002993

Poulin, R., Krasnov, B. R., & Mouillot, D. (2011). Host specificity in phylogenetic and geographic space. *Trends in Parasitology*, *27*(8), 355–361. https://doi.org/10.1016/j.pt.2011.05.003

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2012). The Silva ribosomal RNA gene database project: Improved Data Processing and web-based tools. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1219

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Rannala, B., & Yang, Z. (2020). Species Delimitation. In Scornavacca, C., Delsuc, F., and Galtier, N., editors, Phylogenetics in the Genomic Era, chapter No. 5.5, pp. 5.5:1– 5.5:18. No commercial publisher | Authors open access book. The book is freely available at https://hal.inria.fr/PGE

Reid, N. M., & Carstens, B. C. (2012). Phylogenetic estimation error can decrease the accuracy of species delimitation: A Bayesian implementation of the general mixed yule-coalescent model. *BMC Evolutionary Biology*, *12*(1), 196. https://doi.org/10.1186/1471-2148-12-196

Reid, N. M., Hird, S. M., Brown, J. M., Pelletier, T. A., McVay, J. D., Satler, J. D., & Carstens, B. C. (2013). Poor fit to the multispecies coalescent is widely detectable in empirical data. *Systematic Biology*, *63*(3), 322–333. https://doi.org/10.1093/sysbio/syt057

Ridley, R. K. (1969). Electron microscopic studies on Dicyemid Mesozoa. II. Infusorigen and infusoriform stages. *The Journal of Parasitology*, *55*(4), 779. https://doi.org/10.2307/3277218

Roumbedakis, K., Drábková, M., Tyml, T., & di Cristo, C. (2018). A perspective around cephalopods and their parasites, and suggestions on how to increase knowledge in the field. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.01573

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/

Schiffer, P. H., Robertson, H. E., & Telford, M. J. (2018). Orthonectids are highly degenerate annelid worms. *Current Biology*, *28*(12). https://doi.org/10.1016/j.cub.2018.04.088

Sharda, M. (2020, January 30). *MetReTrim: This pipeline trims Heterogeneity 'n' spacers from the pre-processed reads given the primer sequences.* Mohak91/MetReTrim. Retrieved July 10, 2021, from https://github.com/Mohak91/MetReTrim

Smith, P. E., Waters, S. M., Gómez Expósito, R., Smidt, H., Carberry, C. A., & McCabe, M. S. (2020). Synthetic sequencing standards: A guide to database choice for rumen microbiota amplicon sequencing analysis. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.606825

Stunkard, H. W. (1954). The life-history and systematic relations of the Mesozoa. *The Quarterly Review of Biology*, *29*(3), 230–244. https://doi.org/10.1086/400243

Suzuki, T. G., Ogino, K., Tsuneki, K., & Furuya, H. (2010). Phylogenetic analysis of Dicyemid mesozoans (phylum Dicyemida) from innexin amino acid sequences: Dicyemids are not related to Platyhelminthes. *Journal of Parasitology*, *96*(3), 614–625. https://doi.org/10.1645/ge-2305.1

Telford, M. J., Budd, G. E., & Philippe, H. (2015). Phylogenomic insights into Animal Evolution. *Current Biology*, *25*(19). https://doi.org/10.1016/j.cub.2015.07.060

Tikhonov, M., Leach, R. W., & Wingreen, N. S. (2014). Interpreting 16S metagenomic data without clustering to achieve sub-OTU resolution. *The ISME Journal*, *9*(1), 68–80. https://doi.org/10.1038/ismej.2014.117

Tsaousis, A. D., Kunji, E. R., Goldberg, A. V., Lucocq, J. M., Hirt, R. P., & Embley, T. M. (2008). A novel route for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature*, *453*(7194), 553–556. https://doi.org/10.1038/nature06903

Westcott, S. L., & Schloss, P. D. (2015). *De novo* clustering methods outperform reference-based methods for assigning 16S rRNA gene sequences to operational taxonomic units. *PeerJ*, *3*. https://doi.org/10.7717/peerj.1487

WoRMS - world register of marine species. (n.d.). Retrieved March 4, 2022, from https://www.marinespecies.org/aphia.php?p=browser&id%5B%5D=325343#focus

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2013). The Silva and "all-species living tree project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1209

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, *29*(22), 2869–2876. https://doi.org/10.1093/bioinformatics/btt499

Zheng, X., Ikeda, M., Kong, L., Lin, X., Li, Q., & Taniguchi, N. (2009). Genetic diversity and population structure of the Golden Cuttlefish, *Sepia esculenta* (Cephalopoda: Sepiidae) indicated by microsatellite DNA variations. *Marine Ecology*, *30*(4), 448–454. https://doi.org/10.1111/j.1439-0485.2009.00294.x

Zverkov, O. A., Mikhailov, K. V., Isaev, S. V., Rusin, L. Y., Popova, O. V., Logacheva, M. D., Penin, A. A., Moroz, L. L., Panchin, Y. V., Lyubetsky, V. A., & Aleoshin, V. V. (2019). Dicyemida and Orthonectida: Two stories of body plan simplification. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/fgene.2019.00443

8 Attachments

Tab. I: List of accessions used in this thesis.

Sample code	Host species	Country	Locality	Group	Sampled by	Year
OVPF 1	*Octopus vulgaris*	Portugal	Faro	Faro district	B. Heroutová	2020
OVPF 2	*Octopus vulgaris*	Portugal	Olhao	Faro district	B. Heroutová	2020
SOPF 1	*Sepia officinalis*	Portugal	Faro	Faro district	B. Heroutová	2020
SOPF 2	*Sepia officinalis*	Portugal	Faro	Faro district	B. Heroutová	2020
ASO1	*Sepia officinalis*	Portugal	Albufeira	Faro district	J. Štefka	2021
ASO2	*Sepia officinalis*	Portugal	Albufeira	Faro district	J. Štefka	2021
ASO4	*Sepia officinalis*	Portugal	Albufeira	Faro district	J. Štefka	2021
ASO5	*Sepia officinalis*	Portugal	Albufeira	Faro district	J. Štefka	2021
OVPM 4	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
OVPM 5	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
OVPM 6	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
OVPM 7	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
OVPM 8	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
OVPM 3	*Octopus vulgaris*	Portugal	Matosinhos	Porto district	E. Myšková	2015
SOPM 6	*Sepia officinalis*	Portugal	Matosinhos	Porto district	E. Myšková	2015
SOPM 7	*Sepia officinalis*	Portugal	Matosinhos	Porto district	E. Myšková	2015

SOPP 1	*Sepia officinalis*	Portugal	Porto	Porto district	E. Myšková	2015
LSO7	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LOV1	*Octopus vulgaris*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LSO1	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LSO3	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LSO2	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LSO4	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
LSO5	*Sepia officinalis*	Portugal	Lisbon	Lisbon district	J. Štefka	2021
SOP 1	*Sepia officinalis*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015
OEP 1	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015
OECP 1	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
OECP 6	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
OECP 10	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
OECP 9	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
OEV 7	*Eledone moschata*	Croatia	Veruda	Adriatic Sea east	M. Drábková	2015
OECP 8	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
SOV 4	*Sepia officinalis*	Croatia	Veruda	Adriatic Sea east	M. Drábková	2015
SEP 7	*Sepia officinalis*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015

OECP 7	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2016
SEP 9	*Sepia officinalis*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015
OEV 4	*Eledone cirrhosa*	Croatia	Veruda	Adriatic Sea east	M. Drábková	2015
SOCC 1	*Sepia officinalis*	Croatia	Cres	Adriatic Sea east	D. Míšek	2018
OECC 1	*Eledone moschata*	Croatia	Cres	Adriatic Sea east	D. Míšek	2018
OECC 2	*Eledone moschata*	Croatia	Cres	Adriatic Sea east	D. Míšek	2018
SOP 1	*Sepia officinalis*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015
OEP 1	*Eledone moschata*	Croatia	Pula	Adriatic Sea east	T. Tyml	2015
OPAV 1	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 2	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 3	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 4	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 6	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 7	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 8	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020

OPAV 9	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 10	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 11	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 12	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 13	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 14	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 15	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 16	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 17	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 18	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 19	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 20	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 21	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 22	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 23	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020

OPAV 25	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 26	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 27	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 28	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 30	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 32	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 33	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 34	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 35	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 36	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 37	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 38	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 39	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAV 41	*Octopus pallidus*	Australia	Victoria	Bass strait mainland	Q. Hua	2020
OPAT 01	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020

OPAT 02	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 05	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 09	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 11	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 19	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 21	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 22	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 27	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 28	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAT 34	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OPAV 5	*Octopus pallidus*	Australia	Stanley	Bass strait Tasmania	Q. Hua	2020
OVPV 9	*Octopus vulgaris*	Spain	Vigo	Galicia	E. Myšková	2015
SOPV 8	*Sepia officinalis*	Spain	Vigo	Galicia	E. Myšková	2015
SOPV 9	*Sepia officinalis*	Spain	Vigo	Galicia	E. Myšková	2015
BSO4	*Sepia officinalis*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO8	*Sepia elegans*	Spain	Roses	Catalonia	J. Štefka	2021
BEC2	*Eledone cirrhosa*	Spain	Barcelona	Catalonia	J. Štefka	2021

BSO5	*Sepia officinalis*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO9	*Sepia elegans*	Spain	Roses	Catalonia	J. Štefka	2021
RSO1	*Sepia officinalis*	Spain	Roses	Catalonia	J. Štefka	2021
RSO2	*Sepia officinalis*	Spain	Roses	Catalonia	J. Štefka	2021
BEC4	*Eledone cirrhosa*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO3	*Sepia officinalis*	Spain	Roses	Catalonia	J. Štefka	2021
RSO4	*Sepia officinalis*	Spain	Roses	Catalonia	J. Štefka	2021
BSO1	*Sepia officinalis*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO5	*Sepia officinalis*	Spain	Roses	Catalonia	J. Štefka	2021
BSO2	*Sepia officinalis*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO6	*Sepia elegans*	Spain	Roses	Catalonia	J. Štefka	2021
BSO3	*Sepia officinalis*	Spain	Barcelona	Catalonia	J. Štefka	2021
RSO7	*Sepia elegans*	Spain	Roses	Catalonia	J. Štefka	2021
SEC1	*Eledone cirrhosa*	Spain	Sukarrieta	Basque country	J. Štefka	2021
SEC2	*Eledone cirrhosa*	Spain	Sukarrieta	Basque country	J. Štefka	2021
SOV1	*Octopus vulgaris*	Spain	Bermeo	Basque country	J. Štefka	2021
SOV3	*Octopus vulgaris*	Spain	Bermeo	Basque country	J. Štefka	2021
SOV2	*Octopus vulgaris*	Spain	Bermeo	Basque country	J. Štefka	2021
SEC4	*Eledone cirrhosa*	Spain	Sukarrieta	Basque country	J. Štefka	2021
SEC5	*Eledone cirrhosa*	Spain	Sukarrieta	Basque country	J. Štefka	2021

AZTI1	*Sepia officinalis*	Spain	Sukarrieta	Basque country	I. Mendibil	2021
AZTI2	*Sepia officinalis*	Spain	Sukarrieta	Basque country	I. Mendibil	2021
AZTI3	*Sepia officinalis*	Spain	Sukarrieta	Basque country	I. Mendibil	2021
AZTI4	*Sepia officinalis*	Spain	Sukarrieta	Basque country	I. Mendibil	2021
AZTI5	*Sepia officinalis*	Spain	Sukarrieta	Basque country	I. Mendibil	2021
OTNR1	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR2	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR3	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR4	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR6	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR8	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR9	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR10	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR11	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR12	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OTNR13	*Octopus vulgaris*	Spain	Canary islands	Tenerife	E. Almansa Berro	2021
OMIT 3	*Callistoctopus macropus*	Italy	Oristano	Sardinia	M. Drábková	2016

OEIC 1	*Eledone moschata*	Italy	Cagliari	Sardinia	M. Drábková	2016
OMIT 4	*Callistoctopus macropus*	Italy	Oristano	Sardinia	M. Drábková	2016
OEIC 5	*Eledone moschata*	Italy	Cagliari	Sardinia	M. Drábková	2016
SOIT 2	*Sepia officinalis*	Italy	Oristano	Sardinia	M. Drábková	2016
OVIC 2	*Octopus vulgaris*	Italy	Cagliari	Sardinia	M. Drábková	2016
SOIC 4	*Sepia officinalis*	Italy	Cagliari	Sardinia	M. Drábková	2016
OEIV 11	*Eledone moschata*	Italy	Vieste	Adriatic Sea west	M. Drábková	2015
SOIR 13	*Sepia officinalis*	Italy	Rimini	Adriatic Sea west	M. Drábková	2015
SOIP 11	*Sepia officinalis*	Italy	Pescara	Adriatic Sea west	M. Drábková	2015
OMIN 2	*Callistoctopus macropus*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OMIN 5	*Callistoctopus macropus*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OMIN 1	*Callistoctopus macropus*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OVIN 1	*Octopus vulgaris*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OVIN 2	*Octopus vulgaris*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIN 1	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIN 9	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017

OEIN 2	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIN 3	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIN 10	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
SOIN 2	*Sepia officinalis*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIN 11	*Eledone moschata*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
SOIN 1	*Sepia officinalis*	Italy	Naples	Tyrrhenian Sea	M. Drábková	2017
OEIG 7	*Eledone moschata*	Italy	Gaeta	Tyrrhenian Sea	M. Drábková	2015
OEIG 8	*Eledone moschata*	Italy	Gaeta	Tyrrhenian Sea	M. Drábková	2015
SOIG 8	*Sepia officinalis*	Italy	Gaeta	Tyrrhenian Sea	M. Drábková	2015
OEIM 2	*Octopus vulgaris*	Italy	Marsala	Sicily	O. Ditrich	2015
OEIS 4	*Eledone moschata*	Italy	La Spezia	Ligurian Sea	M. Drábková	2016
OEIL 6	*Eledone moschata*	Italy	Livorno	Ligurian Sea	M. Drábková	2015
OEIS 2	*Eledone cirrhosa*	Italy	La Spezia	Ligurian Sea	M. Drábková	2016
OEIS 3	*Eledone moschata*	Italy	La Spezia	Ligurian Sea	M. Drábková	2016
OWR1	*Octopus incella*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR2	*Octopus incella*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR3	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021

OWR5	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR6	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR7	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR8	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR9	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR10	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR11	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR13	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR14	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR15	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR16	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR17	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR18	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR19	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR20	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021

OWR21	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR22	*Octopus laqueus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR23	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR24	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR25	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR26	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR27	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR28	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR29	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR30	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR31	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR32	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR33	*Metasepia tullbergi*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR34	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
OWR12	*Sepioteuthis lessoniana*	Japan	Okinawa	Okinawa	Z. Lajbner	2021

OWR36	*Abdopus aculeatus*	Japan	Okinawa	Okinawa	Z. Lajbner	2021
SP1	*Amphioctopus marginatus*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP2	*Cistopus taiwanicus*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP3	*Uroteuthis chinensis*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP4	*Sepia lycidas*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP5	*Uroteuthis duvauceli*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP7	*Stenoteuthis oulaniensis*	Vietnam	Ninh Thuan p.	Vietnam	F. Lisenko	2021
SP8	*Uroteuthis chinensis*	Vietnam	Ninh Thuan p.	Vietnam	F. Lisenko	2021
SP9	*Sepia recurvirostra*	Vietnam	Ninh Thuan p.	Vietnam	F. Lisenko	2021
SP10	*Sepioteuthis lessoniana*	Vietnam	Ninh Thuan p.	Vietnam	F. Lisenko	2021
SP11	*Uroteuthis duvauceli*	Vietnam	Ninh Thuan p.	Vietnam	F. Lisenko	2021
SP13	*Amphioctopus aegina*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP14	*Amphioctopus ovolum*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP15	*Amphioctopus marginatus*	Vietnam	Phu Yen p.	Vietnam	F. Lisenko	2021
SP16	*Nototodarus hawaiiensis*	Vietnam	Khanh Hoa p.	Vietnam	F. Lisenko	2021
HW2.2	*Euprymna scolopes*	USA	Paiko beach	Hawaii	H. Osland	2020

HW3.2	*Euprymna scolopes*	USA	Paiko beach	Hawaii	H. Osland	2020
VBCB 2	*Octopus variabilis*	China	Bohai Sea	Bohai Sea	T. Scholz	2019
VBCB 3	*Octopus variabilis*	China	Bohai Sea	Bohai Sea	T. Scholz	2019
VBCB 4	*Octopus variabilis*	China	Bohai Sea	Bohai Sea	T. Scholz	2019
VBCB 5	*Octopus variabilis*	China	Bohai Sea	Bohai Sea	T. Scholz	2019
VBCQ 1	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 2	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 3	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 4	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 5	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 6	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 7	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
VBCQ 8	*Octopus variabilis*	China	Qingdao	Yellow Sea	T. Scholz	2019
SOGT 1	*Sepia officinalis*	Greece	Thessaloniki	Thessaloniki	A. Bartoňová	2017
SOGT 2	*Sepia officinalis*	Greece	Thessaloniki	Thessaloniki	A. Bartoňová	2017

Tab. II: List of indices used in this thesis.

Index name	Sequence 5'-3'	Index name	Sequence 5'-3'
N701	TCGCCTTA	N723	GAGCGCTA
N702	CTAGTACG	N724	CGCTCAGT
N703	TTCTGCCT	N726	GTCTTAGG
N704	GCTCAGGA	N727	ACTGATCG
N705	AGGAGTCC	N728	TAGCTGCA
N706	CATGCCTA	N729	GACGTCGA
N707	GTAGAGAG	S502	CTCTCTAT
N710	CAGCCTCG	S503	TATCCTCT
N711	TGCCTCTT	S505	GTAAGGAG
N712	TCCTCTAC	S506	ACTGCATA
N714	TCATGAGC	S507	AAGGAGTA
N715	CCTGAGAT	S508	CTAAGCCT
N716	TAGCGAGT	S510	CGTCTAAT
N718	GTAGCTCC	S511	TCTCTCCG
N719	TACTACGC	S513	TCGACTAG
N720	AGGCTCCG	S515	TTCTAGCT
N721	GCAGCGTA	S516	CCTAGAGT
N722	CTGCGCAT		

Tab. III: Output of Kruskal-Wallis (only significant p-values).

Group 1	Group 2	p-value
SO_Greece (n=2)	SO_Lisbon (n=6)	0.0455
SO_Faro (n=6)	SO_Greece (n=2)	0.0455
SO_Catalonia (n=10)	SO_Faro (n=6)	0.0126
SO_Catalonia (n=10)	SO_Galicia (n=2)	0.0317
SO_Basque (n=6)	SO_Greece (n=2)	0.0455
SO_AdriaticEast (n=5)	SO_Faro (n=6)	0.0446
OV_Tenerife (n=11)	SO_Greece (n=2)	0.0299

OV_Porto (n=6)	SO_Sardinia (n=2)	0.0455
OV_Porto (n=6)	SO_Tyrh (n=3)	0.0201
OV_Porto (n=6)	SO_Greece (n=2)	0.0455
OV_Porto (n=6)	SO_Catalonia (n=10)	0.0011
OV_Porto (n=6)	SO_Faro (n=6)	0.0039
OV_Porto (n=6)	SO_AdriaticEast (n=5)	0.0062
OV_Faro (n=2)	OV_Porto (n=6)	0.0455
OV_Basque (n=3)	SO_Tyrh (n=3)	0.0495
OV_Basque (n=3)	SO_Catalonia (n=10)	0.0112
OV_Basque (n=3)	SO_AdriaticEast (n=5)	0.0253
EM_Tyrh (n=7)	SO_Greece (n=2)	0.0404
EM_AdriaticEast (n=10)	SO_Faro (n=6)	0.0393
EM_AdriaticEast (n=10)	SO_Galicia (n=2)	0.0317
EM_AdriaticEast (n=10)	SO_Greece (n=2)	0.0317
EM_AdriaticEast (n=10)	OV_Porto (n=6)	0.0011

Tab. IV: Output of permanova.

Group 1	Group 2	Sample size	Permutations	p-value
Octopus_asia	Octopus_australia	57	999	0.001
Octopus_asia	Octopus_medit	79	999	0.001
Octopus_asia	Sepia_asia	20	999	0.058
Octopus_asia	Sepia_medit	70	999	0.001
Octopus_asia	Teuthida	22	999	0.063
Octopus_australia	Octopus_medit	98	999	0.001
Octopus_australia	Sepia_asia	39	999	0.033
Octopus_australia	Sepia_medit	89	999	0.001
Octopus_australia	Teuthida	41	999	0.001
Octopus_medit	Sepia_asia	61	999	0.014
Octopus_medit	Sepia_medit	111	999	0.001
Octopus_medit	Teuthida	63	999	0.007
Sepia_asia	Sepia_medit	52	999	0.144

Sepia_asia	Teuthida	4	999	0.253
Sepia_medit	Teuthida	54	999	0.014

Tab. V: Accessions downloaded from Genbank.

Species	Accession no
Dicyema acuticephalum	D26530.1
Dicyema orientale	D26529.1
Dicyema moschatum	MT703900.1
Dicyema clavatum	LC571905.1
Pseudicyema truncatum	MN066367.1
Dicyemennea eledones	LT669912.1
Dicyemennea brevicephala	KJ786928.1
Dicyemennea adminicula	KJ786927.1
Dicyemennea adscita	KJ786926.1
Dicyema apollyoni	KJ786925.1
Dicyemennea rossiae	KJ786921.1
Dicyema sphyrocephalum	LC571906.1
Dugesia japonica	D83382.1
Neisseria gonorrhoeae	ROU58369.1

Tab. VI: Output of Kruskal-Wallis analysis based on general host and location.

Group 1	Group 2	p-value	q-value
Octopus_asia (n=19)	Octopus_australia (n=38)	0.318	0.367
Octopus_asia (n=19)	Octopus_medit (n=60)	0.010	0.036
Octopus_asia (n=19)	Sepia_asia (n=1)	0.099	0.166
Octopus_asia (n=19)	Sepia_medit (n=51)	0.000	0.000
Octopus_asia (n=19)	Teuthida (n=3)	0.315	0.367
Octopus_australia (n=38)	Octopus_medit (n=60)	0.000	0.000
Octopus_australia (n=38)	Sepia_asia (n=1)	0.091	0.166

Octopus_australia (n=38)	Sepia_medit (n=51)	0.000	0.000
Octopus_australia (n=38)	Teuthida (n=3)	0.040	0.100
Octopus_medit (n=60)	Sepia_asia (n=1)	0.088	0.166
Octopus_medit (n=60)	Sepia_medit (n=51)	0.023	0.070
Octopus_medit (n=60)	Teuthida (n=3)	0.628	0.628
Sepia_asia (n=1)	Sepia_medit (n=51)	0.405	0.434
Sepia_asia (n=1)	Teuthida (n=3)	0.180	0.263
Sepia_medit (n=51)	Teuthida (n=3)	0.193	0.263

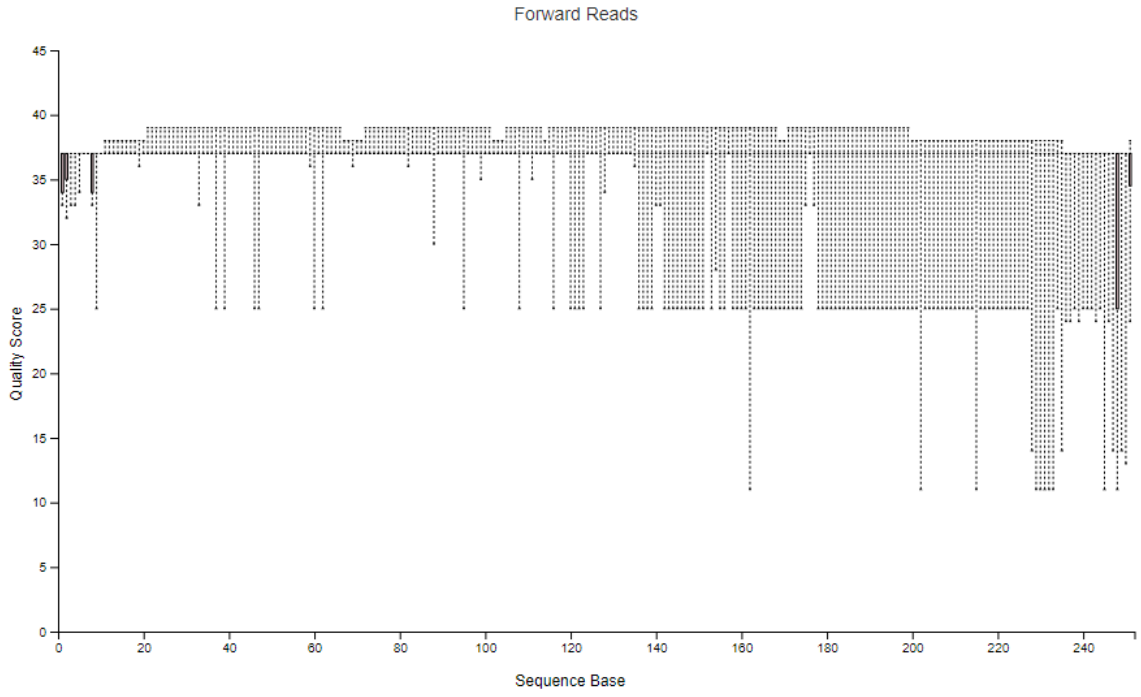


Fig. I: Quality report for forward reads (qiime2).

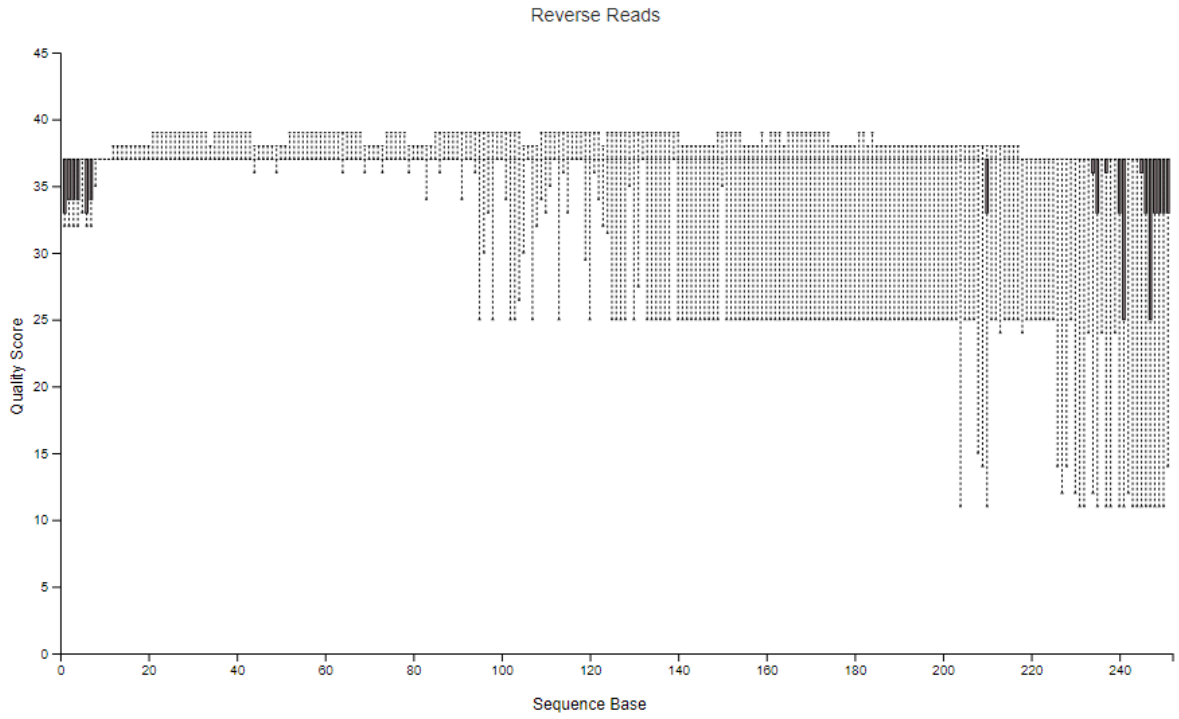


Fig. II: Quality report for reverse reads (qiime2).

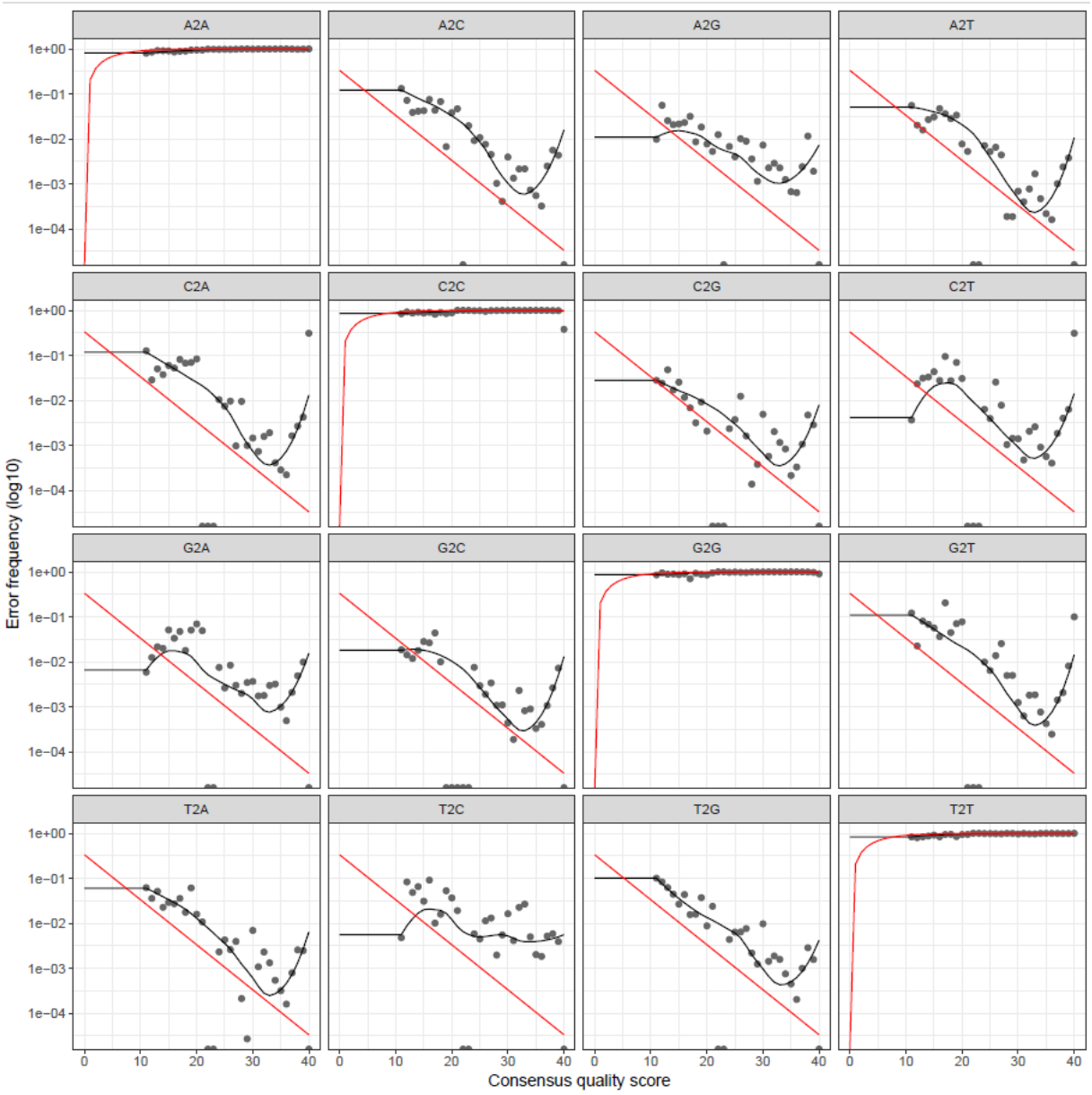


Fig. **III:** Error rates of reverse reads. Points represent the observed error rates. Black line estimates error rates after convergence of the machine/learning algorithm. Red line represents expected error rates when quality is optimal.

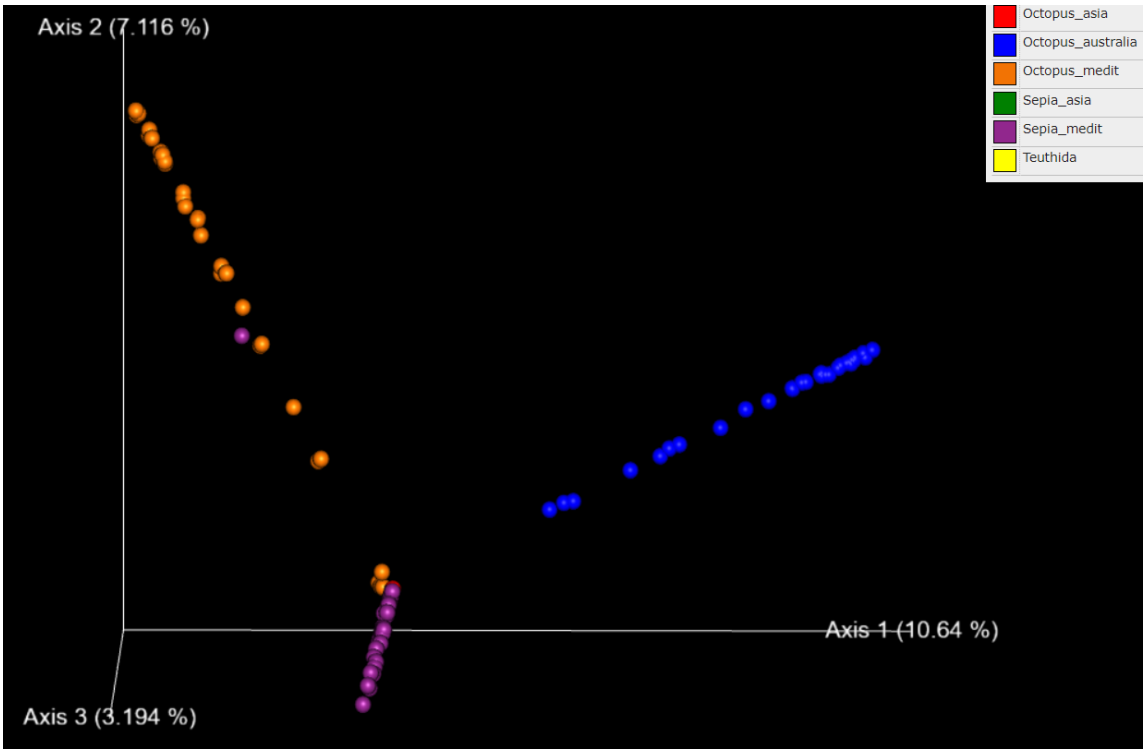


Fig. IV: PCoA of Bray-Curtis dissimilarity emperor plot. Measures dissimilarity between datasets.

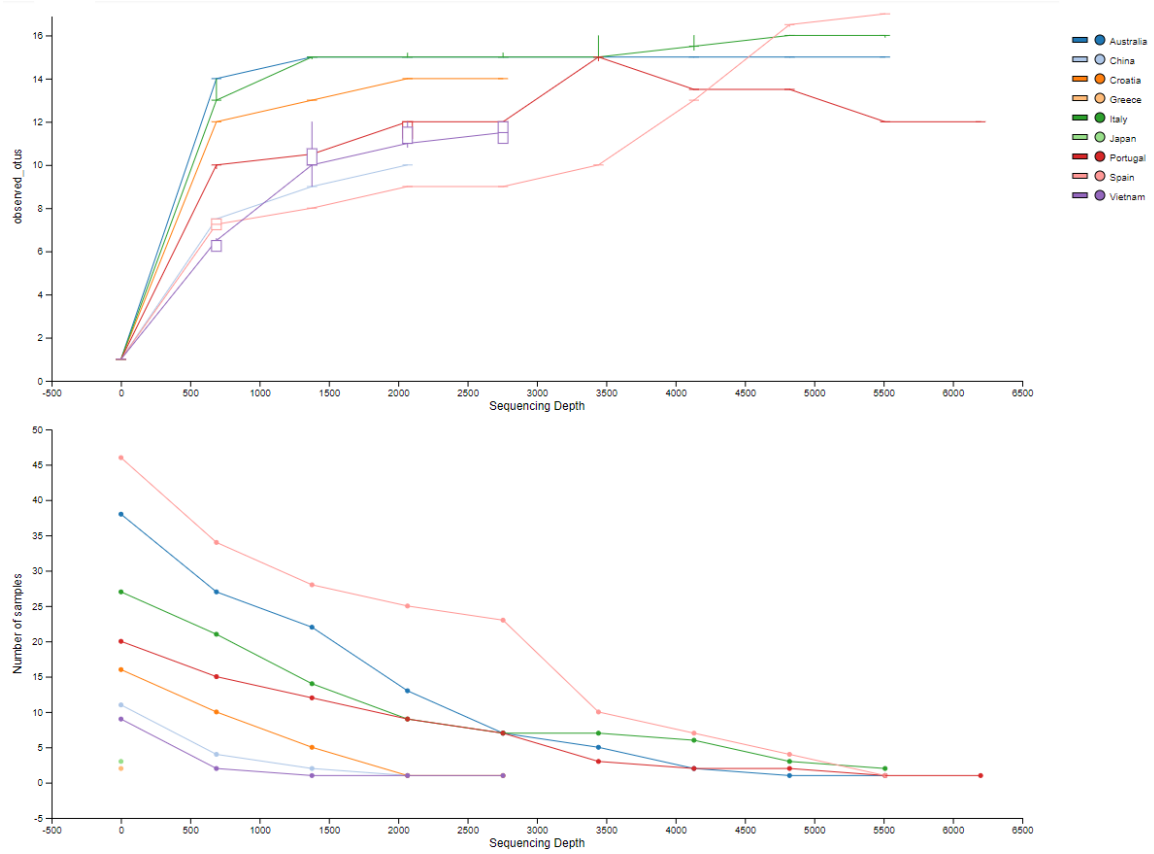


Fig. V: Rarefaction plot (by locality). Shows ASV richness per locality based on sequencing depth.

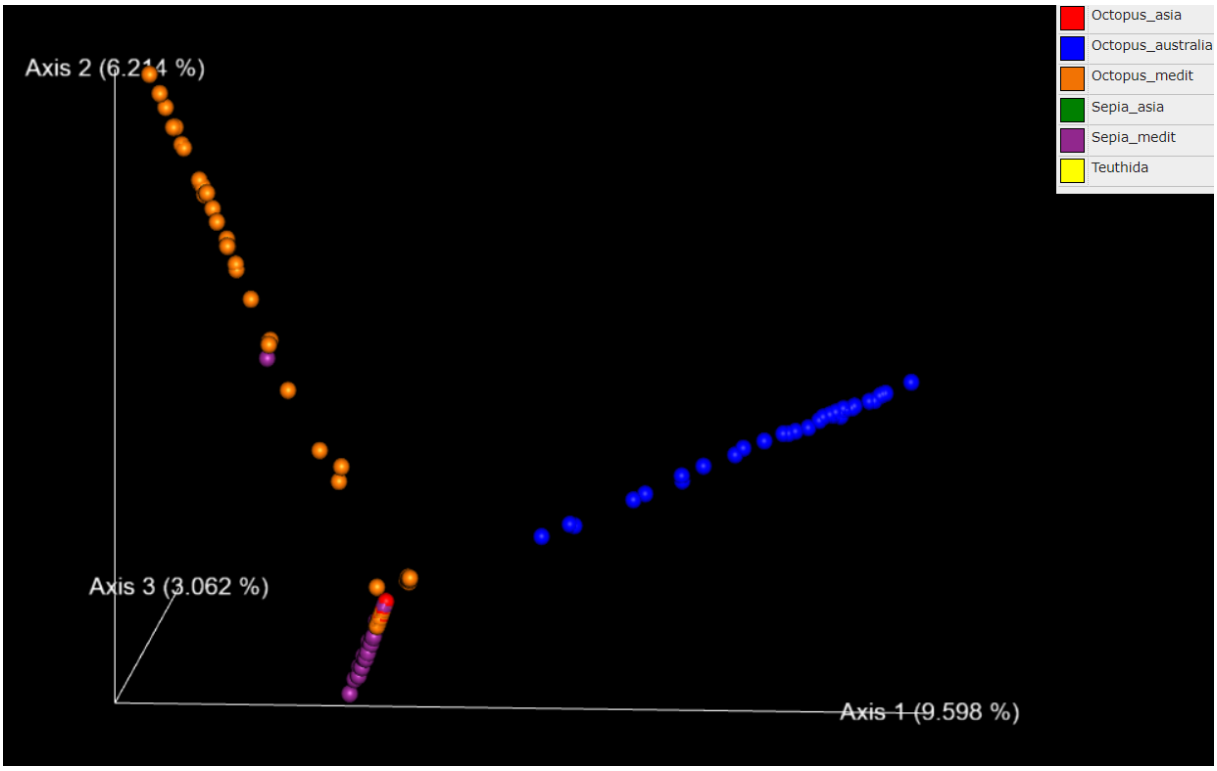


Fig. VI: Jaccard emperor plot calculates Jaccard similarity index. Shows fraction of unique features, regardless of abundance.

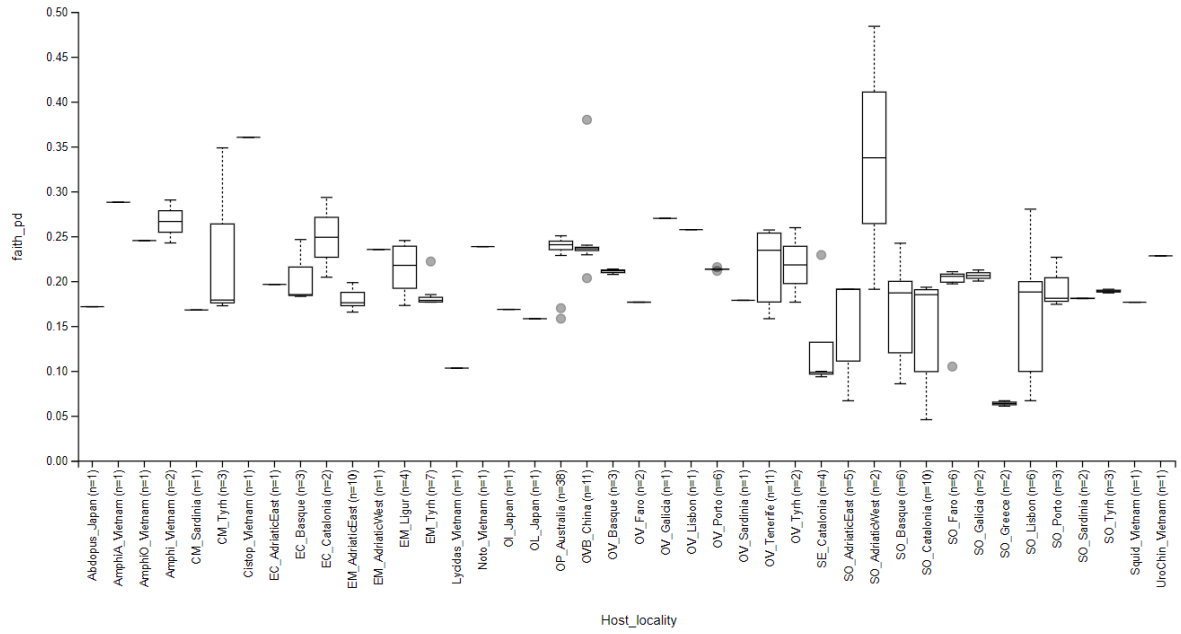


Fig. VII: Faith's phylogenetic distance plot measuring biodiversity by incorporating phylogenetic differences between species. AmphiA= *Amphioctopus aegina*, AmphiO=

Amphioctopus ovolum, Amphi= *Amphioctopus marginatus*, CM= *Callistoctopus macropus*, Cistop= Cistopus taiwanicus, EC= *Eledone cirrhosa*, EM= *Eledone moschata*, Lycidas= *Sepia lycidas*, Noto= *Nototodarus hawaiiensis*, OI= *Octopus incella*, OL= *Octopus laqueus*, OP= *Octopus pallidus*, OV= *Octopus vulgaris*, OVB= *Octopus variabilis*, SE= *Sepia elegans*, SO= *Sepia officinalis*, Squid= *Sepioteuthis lessoniana*, UroChin= *Uroteuthis chinensis*.