**Thomasdotter, A, Shum, P, Mugnai, F, Vingiani, M, Abbiati, M and Costantini, F**

**Spineless and overlooked: DNA metabarcoding of autonomous reef monitoring structures reveals intra- and interspecific genetic diversity in Mediterranean invertebrates**

Received: 8 December 2022 | Revised: 22 June 2023 | Accepted: 4 July 2023

DOI: 10.1111/1755-0998.13836

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

# Spineless and overlooked: DNA metabarcoding of autonomous reef monitoring structures reveals intra- and interspecific genetic diversity in Mediterranean invertebrates

Anna Thomasdotter[1,2] | Peter Shum[3] | Francesco Mugnai[2] | Marina Vingiani[2,4] | Marco Abbiati[5,6,7] | Federica Costantini[2,6,7]

[1]County Administrative Board of Västerbotten, Umeå, Sweden

[2]Department of Biological, Geological and Environmental Sciences, University of Bologna, UOS Ravenna, Ravenna, Italy

[3]School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK

[4]National Research Council, Institute of Marine Sciences, CNR-ISMAR, Venice, Italy

[5]Department of Cultural Heritage, University of Bologna, Ravenna, Italy

[6]National Interuniversity Consortium for Marine Sciences (CoNISMa), Rome, Italy

[7]Interdepartmental Research Center for Environmental Sciences (CIRSA), Ravenna, Italy

**Correspondence**
Anna Thomasdotter, County Administrative Board of Västerbotten, Storgatan 71 B, Umeå 901 86, Sweden.
Email: anna.thomasdotter@gmail.com

Federica Costantini, Department of Biological, Geological and Environmental Sciences, University of Bologna, UOS Ravenna, Via Sant'Alberto 163, Ravenna 48123, Italy.
Email: federica.costantini@unibo.it

**Funding information**
ERA-NET MarTERA, Grant/Award Number: 728053-MarTERA; Ministero dell'Università e della Ricerca

**Handling Editor:** Aurélie Bonin

## Abstract

The ability to gather genetic information using DNA metabarcoding of bulk samples obtained directly from the environment is crucial to determine biodiversity baselines and understand population dynamics in the marine realm. While DNA metabarcoding is effective in evaluating biodiversity at community level, genetic patterns within species are often concealed in metabarcoding studies and overlooked for marine invertebrates. In the present study, we implement recently developed bioinformatics tools to investigate intraspecific genetic variability for invertebrate taxa in the Mediterranean Sea. Using metabarcoding samples from Autonomous Reef Monitoring Structures (ARMS) deployed in three locations, we present haplotypes and diversity estimates for 145 unique species. While overall genetic diversity was low, we identified several species with high diversity records and potential cryptic lineages. Further, we emphasize the spatial scale of genetic variability, which was observed from locations to individual sampling units (ARMS). We carried out a population genetic analysis of several important yet understudied species, which highlights the current knowledge gap concerning intraspecific genetic patterns for the target taxa in the Mediterranean basin. Our approach considerably enhances biodiversity monitoring of charismatic and understudied Mediterranean species, which can be incorporated into ARMS surveys.

**KEYWORDS**
autonomous reef monitoring structures, COI, DNA metabarcoding, haplotype diversity, invertebrate, Mediterranean Sea

## 1 | INTRODUCTION

The development and application of biodiversity monitoring tools enhance the power and pace of data generation, which is important to track ecological changes in the face of environmental shifts. From communities to populations, genetic diversity is a valuable tool to characterize these responses to surrounding pressures. Patterns of genetic differentiation are, therefore, fundamental to understanding community and species resilience (Reusch et al., 2005), evolutionary potential, adaptive ability (DuBois et al., 2022) and connectivity patterns (Darnaude et al., 2022). These aspects have implications for both marine ecosystems themselves and for the natural services they provide (Darnaude et al., 2022; Pinsky et al., 2020). Nevertheless, genetic diversity information is still low and patchy for marine invertebrates partially due to challenges with sampling (Costantini et al., 2018).

At smaller spatial scales, genetic variability affects local dynamics and thus the resulting in community composition. Biodiversity assessments are an important tool to capture this local variability at both species and population levels. In recent years, efforts have been made to standardize biodiversity data collection, allowing for studies at larger spatial and temporal scales (Obst et al., 2020). One such method is the deployment of Autonomous Reef Monitoring Structures (ARMS; Leray & Knowlton, 2015) which have been used across the world's oceans over the last few years (Obst et al., 2020; Pearman et al., 2020, see www.naturalhistory.si.edu/research/global-arms-program/publications for further updates). ARMS consist of nine stacked PVC plates arranged in a three-dimensional structure (Figure 1) that provide suitable settlement surfaces for a remarkable variety of species and can be deployed without causing a significant impact on the surrounding environment. Following retrieval, organisms collected on and within the structures can be identified using morphological and molecular techniques (Leray & Knowlton, 2015). Regarding the latter, DNA metabarcoding is a popular and efficient approach that allows for the rapid assessment of species diversity within a community (Taberlet et al., 2012). The mitochondrial cytochrome oxidase I (COI) gene has been extensively used to describe whole communities for countless metazoan surveys (e.g. Leray & Knowlton, 2015; Nichols et al., 2022), and traditional population genetic studies use the same target gene for its intraspecific variability to study phylogeographic distributions of populations (Pérez-Portela et al., 2013; Wäge et al., 2017). However, the within-species variability of the COI is typically concealed in metabarcoding studies, despite representing valuable ecological information with major implications for ecosystem dynamics (DuBois et al., 2022; Reusch et al., 2005). Unlocking this data, therefore, has enormous potential for assessing intraspecific genetic diversity for hundreds of species simultaneously.

Classical metabarcoding studies have primarily focused on generating species lists rather than assessing within-species diversity. This is primarily an error-avoidance strategy as both PCR amplification of bulk samples and next-generation sequencing inherently introduce false sequences through, e.g., chimeras and tag-jumping (Elbrecht et al., 2018; Turon et al., 2020). Clustering sequences is currently a common practice to reduce input data to molecular operational taxonomic units (MOTUs) which removes the within-species variability that is typically binned into MOTUs. Elbrecht et al. (2018) devised a method for detecting intraspecific genetic diversity by extracting haplotypes at the community level. This approach has gained traction in recent years, as evident from numerous studies (Andújar et al., 2021; Antich et al., 2021, 2022; Brandt et al., 2021; Macé et al., 2021; Porter & Hajibabaei, 2020; Shum & Palumbi, 2021; Turon et al., 2020) that have utilized this technique to reveal valuable insights into genetic diversity. The term "metaphylogeography" was coined by Turon et al. (2020) to describe the application of this method in uncovering intraspecific genetic patterns and population structures within ecological communities. With these alternative bioinformatics tools, raw metabarcoding reads can be processed to recover amplicon sequence variants (ASVs) at the species level. Once ASVs or haplotypes are generated from the metabarcoding data, they can be employed to make inferences about population-level genetic diversity. However, an issue remains in that the number of reads per haplotype does not necessarily correlate with the biomass or number of individuals exhibiting that haplotype (Elbrecht et al., 2017). Haplotype abundances are crucial for population inferences, and several approaches can be considered to quantify relative abundances. Azarian et al. (2021) proposed the use of a frequency of occurrence metric instead of read abundances when an appropriate amount of samples are collected. If reads are recorded for a haplotype in a given sample, at least one individual with that haplotype was present. This approach is advantageous for population-level inferences from community DNA collections when adequately sampled and replicated (Azarian et al., 2021), as demonstrated by Shum and Palumbi (2021).

In the present study, we show the potential of ARMS to provide new insights into intraspecific genetic patterns for several invertebrate taxa from the Mediterranean Sea. This region harbours unique

**FIGURE 1** Autonomous reef monitoring structures (ARMS) during deployment (left) and a year later (right). Photos by Gabriele Giacalone.

biodiversity hotspots with prominent levels of endemic fauna, accounting for an estimated 7% of global marine biodiversity (Coll et al., 2012). Nonetheless, Mediterranean coastal areas are subject to compounding anthropogenic impacts such as habitat disturbance, depletion and destruction (Coll et al., 2012; Micheli et al., 2013). Further, knowledge about the genetic structure and diversity of many of the local invertebrate fauna is severely lacking (Costantini et al., 2018; Mugnai et al., 2021). Information regarding these species' genetic diversity in different locations would, therefore, aid in characterizing their response to environmental pressures in a unique marine setting. Further, as demonstrated in this study, this information can be acquired through existing biodiversity monitoring efforts. Here, we use metabarcoding data gathered from ARMS deployed in the Tyrrhenian (Livorno and Palinuro, Italy) and Adriatic (Rovinj, Croatia) Seas. These marine basins constitute two distinct biogeographic environments with species-specific levels of connectivity for invertebrates (Villamor et al., 2014), which can affect both community patterns and intraspecific diversity. The strong abiotic barriers may contribute to higher genetic variability in echinoderms and polychaetes due to population isolation and ongoing speciation (Patarnello et al., 2007; Villamor et al., 2014). Conversely, lower genetic variability may be exhibited by corals and sponges due to inherently less variation within the COI region for these taxa (Calderón et al., 2006; Erpenbeck et al., 2016; Shearer et al., 2002). However, there are generally few studies on the subject in the Mediterranean region (Costantini et al., 2018).

## 2 | METHODS

### 2.1 | Study design

Standardized sampling units (Autonomous Reef Monitoring Structures [ARMS], Figure 1) were deployed in three locations in the Mediterranean Sea: two in the Tyrrhenian Sea (Palinuro and Livorno, IT) and one in the northern Adriatic Sea (Rovinj, HR; Figure 2). Each location included three sites, with three ARMS in each site (Figure S1). From each ARMS, five samples were collected corresponding to different plate positions within the structure (Figure S2). Moreover, one additional sample was collected from the natural substrate near each ARMS. Each sample was analysed in three PCR replicates. Two samples from the same ARMS in Livorno contained insufficient sample volume for genetic analysis. As such, there was a total of 480 PCR replicates (3 locations * 3 sites * 3 units * 6 samples * 3 replicates—6 unavailable replicates).

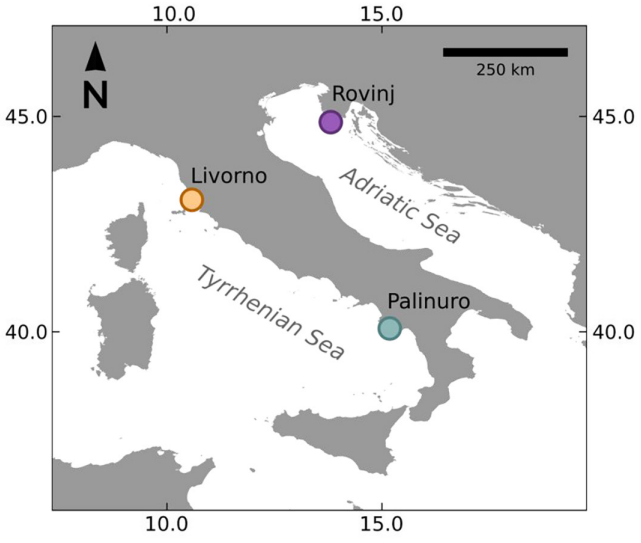



FIGURE 2 Map of sampling locations in the Mediterranean Sea.

The resulting metabarcoding data were compared with two additional datasets: morphological identification data of organisms present on and within the ARMS, and existing sequence data derived from single-specimen samples published on GenBank.

### 2.2 | Sample collection

Autonomous reef monitoring structures (Figure 1) were deployed by SCUBA diving in June 2018 at 14–24 m depth (Table S1) and retrieved in July 2019. ARMS were placed at the same depth within a given site. During retrieval, ARMS were covered with PVC boxes to prevent loss of vagile fauna, then removed from the substrate. Once transported to the boat, each ARMS was placed into a separate PVC bin filled with 20-μm filtered seawater. Natural substrate was sampled in a 22.5 × 22.5 cm area near each ARMS using a scraper and a vacuum pump at the time of retrieval. The sample net was collected and placed into a sterile PE ziplock bag.

In the laboratory, ARMS were disassembled, and vagile fauna was collected by sieving the water from the bins. The 2000-μm fraction was placed in ethanol for morphological identification to the lowest taxonomic rank possible. Each ARMS plate (22.5 × 22.5 cm) was placed into individual sterile PVC trays and photographed on each side to estimate per cent coverage of sessile organisms. The morphological data collected were used for comparison with metabarcoding data in the present study and a summarized dataset is available in Appendix S1.

Settlement plates were then scraped and the organic material was collected according to plate position (Figure S2), resulting in five samples from within each ARMS. All samples (five from ARMS, one from the nearby natural substrate) were homogenized separately in a blender at maximum speed for 15 s. The resulting material was dried, split into sub-replicates in 15-mL falcon tubes with 96% ethanol and stored at −20°C until DNA extraction.

### 2.3 | DNA extraction, amplification and sequencing

DNA was extracted using the NucleoSpin® Soil Kit (Thermo Fisher Scientific) following the manufacturer's protocol with two modifications; the initial sample amount was doubled from 500 to 1000 mg, and the elution buffer was incubated at 70°C before use.

Samples were amplified using the versatile degenerate primer set targeting a 313-bp fragment of cytochrome c oxidase subunit 1 (COI) (forward mlCOIintF: GGWACWGGWTGAACWGTWTA YCCYCC; Leray et al., 2013, reverse jgHCO2198: TAIACYTCIGG RTGICCRAARAAYCA; Geller et al., 2013) that has been used to amplify a wide range of phyla across eukaryotes and algae. Forward and reverse primers were indexed with 12-bp tags and added in unique combinations for each sample. Each 15-μL amplification mix consisted of 7.5-μL QIAGEN® Multiplex Master Mix, 2.5-μL QIAGEN® Nuclease-Free H20, 1.5 μL of each respective tagged primer (2 μM) and 2-μL DNA. All samples, including seven positive and three negative controls, were amplified in three PCR replicates. The positive controls consisted of six identical terrestrial mock communities (containing 14 known species) and one marine mock community (containing 12 known species) with each species having a single Sanger sequenced barcode present in unimolar concentrations (Appendix S1). PCR negative controls consisted of DNA extraction blanks (1.5-mL tubes left open on the workbench during extraction procedures) and PCR controls (water).

PCR amplification was performed on an Applied Biosystems® SimpliAmp Thermal Cycler (Thermo Fisher Scientific). The PCR cycling consisted of an initial denaturation step of 15 min at 95°C, followed by five cycles of 30 s at 94°C, 40 s at 45°C and 1 min at 72°C, then 30 cycles of 30 s at 94°C, 40 s at 48°C and 1 min at 72°C, followed by a final extension step of 10 min at 72°C. PCR products, including positive and negative controls, were pooled in equal amounts for each sample (1 μL for samples and 4 μL for controls). Pooled PCR product was purified using NucleoSpin® Gel and PCR Clean-up (Thermo Fisher Scientific) following the manufacturer's instructions. The metabarcoding library was constructed by adding Illumina adapters through bridge amplification (Indexing PCR, I-PCR) on purified pooled samples using Phusion® High-Fidelity PCR Master Mix (Thermo Fisher Scientific). The indexing PCR was performed using 30-μL Phusion® High-Fidelity PCR Master Mix, 2.1 μL of each of the corresponding indexes and Illumina adapters (20 μM) and 6 μL of purified pooled DNA. PCR cycling consisted of an initial step of 30 s at 98°C, followed by 12 cycles of 40 s at 98°C, 45 s at 55°C and 1 min at 72°C, and a final elongation step of 10 min at 72°C. The Illumina library (550 bp, target region + Illumina indices/adaptors) was gel extracted and purified using the NucleoSpin® Gel and PCR Clean-up (Thermo Fisher Scientific).

The library was assessed for quantity and quality using Agilent 2100 Bioanalyzer and sequenced by the French Agricultural Research Center for International Development (CIRAD) on an Illumina MiSeq using v2 chemistry (2 × 250 bp paired-end run).

## 2.4 | Bioinformatics pipeline

The quality of the resulting raw Illumina reads was inspected using FASTQC (Andrews, 2010; www.bioinformatics.babraham.ac.uk/projects/fastqc/). OBITOOLS (Boyer et al., 2016) was used for initial quality filtering. Reads below a minimum quality threshold of 28 were removed (*obicut*), pair-end reads aligned and alignments with quality scores below 40 discarded (*illuminapairedend*). The aligned sequences were demultiplexed (*ngsfilter*), filtered strictly for length at 313 bp, reads with ambiguous bases were removed (*obigrep*) and parsed by replicate.

Quality-filtered sequences were denoised and clustered in R (R Core Team, 2022) using a modified JAMP (Elbrecht et al., 2018) denoising module excluding abundance filters. Through this module, singletons were excluded and sequences subsequently denoised within each replicate using UNOISE (alpha = 5, VSEARCH v2; Rognes et al., 2016) and clustered using SWARM v3 (d = 13; Mahé et al., 2022). Clustering was carried out on the full denoised dataset.

The final dataset was decontaminated using *decontam* (Davis et al., 2018) in R which statistically identifies contaminant sequences through comparisons with negative controls. Contaminants were removed based on the prevalence model using a 0.5 threshold after excluding eight negative control replicates with library sizes >2000.

Each step was evaluated by inspecting the number of haplotypes remaining in terrestrial positive control sequence clusters, i.e., the molecular operational taxonomic units (MOTUs) containing the terrestrial control haplotypes. The marine positive controls were not used for this purpose due to potential cross-contamination between samples. After decontamination, spurious haplotypes remained and the haplotype table was, therefore, further curated using *lulu* (Frøslev et al., 2017) in R with default settings (84% similarity threshold, 95% co-occurrence ratio). The following filters were then applied in successive order: haplotype relative abundance per replicate >0.01%, haplotype absolute abundance in replicate >5 reads and occurrence in at least two out of three PCR replicates.

The resulting sequences were inspected in MEGA11 (Tamura et al., 2021) and those containing stop codons were removed. Sequences were taxonomically assigned using the RDP classifier (v2.12; Wang et al., 2007) trained on a COI database from 2016 (Wangensteen & Turon, 2016). A second taxonomic assignment was made using BOLDigger (Buchner & Leese, 2020) to account for updated databases since the creation of the RDP classifier reference. In BOLDigger, the BOLDigger method was used to identify top hits and flag suspicious matches. Species-level uncertainties (e.g. due to incomplete taxonomic resolution in the reference match) were inspected and corrected when possible. The taxonomic tables from RDP and BOLDigger were combined so that BOLDigger assignments replaced RDP assignments in cases when BOLDigger had sequence identity match ≥98% and RDP species-level bootstrap values were lower than 98%. RDP species-level assignments were accepted for species-level bootstrap values ≥90%. If RDP bootstrap values for class were lower than 85% and no appropriate BOLDigger assignment was available to replace it, no class was assigned and the haplotype was removed. If a given MOTU contained haplotypes where some remained unassigned at species level, all haplotypes received the species assignment when no conflicts were present. MOTUs containing conflicting class assignments were removed. The final taxonomic list was filtered to exclude non-marine and non-metazoan taxa, and haplotypes clustered into MOTUs containing control sequences

(including from marine mock communities) were removed to account for cross-contamination between control and real samples.

### 2.5 | Data analysis

In the final haplotype table, it was noted that haplotypes assigned to the same species were frequently clustered into separate MOTUs. For this reason, haplotypes were grouped into species based on taxonomic assignment rather than MOTUs. Haplotypes without species-level assignments were, therefore, disregarded after initial data exploration.

To generate population-level information, the resulting dataset was converted to presence/absence per haplotype per sample (see Azarian et al., 2021; Shum & Palumbi, 2021). Note that this is different from presence per PCR replicate since a haplotype had to be present in at least two out of three PCR replicates in previous steps to be considered present in a sample. The plate position for each sample was not treated as a variable in any analyses since this was beyond the scope of the present study.

To validate and compare results, sequences from single-specimen samples were included from GenBank for species with more than three haplotypes in the metabarcoding dataset. Sequences were downloaded when more than 10 sequences from Mediterranean locations were available. Sequence data meeting these criteria were available for four groups: *Ophiothrix fragilis* (Pérez-Portela et al., 2013), *Platynereis dumerilii* and *P. massiliensis* (Calosi et al., 2013; Wäge et al., 2017), *Clytia* spp. (Cunha et al., 2017) and *Eualus* spp. (Conforti & Costantini, 2022). The GenBank sequences were combined with the species data from the present study and aligned in MEGA11 (Tamura et al., 2021) using the Muscle algorithm (Edgar, 2004). Sequences were trimmed to the longest possible overlapping fragment in MEGA11.

Haplotype networks for species with at least five haplotypes were created using *pegas* (Paradis, 2010) in R. *Pegas* was also used to compute genetic distance (K2P) and nucleotide diversity ($\pi$). The R package *ade4* (Dray & Dufour, 2007) was used for analysis of molecular variance (AMOVA) with a randomization test ($n = 999$) to test the effects of location and site (nested within location) on genetic variability in each species. Accumulation and rarefaction curves were created using the R package *vegan* (Oksanen et al., 2022).

To confirm that species found in the resulting dataset had previously been observed in the study region, species occurrence records were checked in the Global Biodiversity Information Facility (GBIF; www.gbif.org/) occurrence database.

## 3 | RESULTS

### 3.1 | Pipeline outputs

A total of 12,999,884 reads from 480 PCR replicates were generated after evaluation based on 42 positive control replicates with mock communities and 10 negative control replicates (Table S2). After the bioinformatics pipeline, 13 out of 14 control sequences from terrestrial mock communities were recovered, with eight spurious haplotypes remaining in their corresponding molecular operational taxonomic units (MOTUs) across the positive control replicates. Eight MOTUs with a total of 17 haplotypes were excluded from further analysis due to conflicting class assignments within the MOTU. Further, seven haplotypes were assigned to a vertebrate class (Actinopteri, ray-finned fishes) and excluded since this was outside the scope of the present study. Field samples contained a total of 349,976 reads, with an average of 2187 reads per sample (range 38–15,027, standard deviation 2592) following the pipeline. Taxonomic assignment resulted in 4498 metazoan haplotypes, of which 613 haplotypes in 322 MOTUs were taxonomically assigned to at least class level. Of these, 429 haplotypes in 205 MOTUs were taxonomically assigned to the species level. The resulting dataset was converted into presence/absence of unique haplotypes for each sample (comprising three replicates), which resulted in a total of 2243 presence counts. Rarefaction curves (Figure S3) for number of species and haplotypes observed at different sequencing depths showed that the majority of samples reached saturation. This indicated that the sequencing depth was adequate for capturing the diversity of the community. Accumulation curves for number of haplotypes and species observed with increasing number of samples are presented in Figure S4.

The pipeline generated a dataset with 145 unique species. Fifty of these species belonged to more than one MOTU, with five species clustering into four or more MOTUs. Conversely, only four MOTUs contained multiple species assignments. For this reason, analyses following initial data exploration considered species to consist of haplotypes with the same taxonomic assignment at species level, rather than individual MOTUs. As a result, haplotypes without species assignments were disregarded.

### 3.2 | Comparison with morphological observations

Photoanalysis of sessile organisms on ARMS plates identified 22 invertebrate categories to at least genus level (Figure 3). Of these categories, only one (the ascidian *Ciona edwardsi*) was also observed in the metabarcoding data. The most frequent categories identified in the photoanalysis were the Annelida Polychaeta *Salmacina* spp./*Filograna implexa* species complex (Kupriyanova & Jirkov, 1997) and the Bryozoa *Schizobrachiella sanguinea*, which both lacked haplotype occurrences in the metabarcoding data. The photoanalysis identified a larger number of bryozoan (classes Gymnolaemata and Stenolaemata) and ascidian (class Ascidiacea) species than the metabarcoding method (Figure 3).

Morphological identification of vagile fauna (≥2 mm) observed 92 categories defined to at least genus level in the ARMS. Of these, 17 categories matched with species present in the metabarcoding data (Figure 3). A larger number of mollusc species (classes Bivalvia and Gastropoda) were detected using morphological identification than metabarcoding.

Conversely, metabarcoding detected 128 species not found in either sessile or vagile data, primarily species of classes Polychaeta and Demospongiae (Figure 3).

### 3.3 | Location and site comparisons

The most abundant classes were Polychaeta, Hydrozoa and Demospongiae (Figures 4 and 5). The three classes constituted 84% of haplotype occurrences in the final dataset and were prevalent across all locations (Figure 5), although Rovinj had more classes represented overall. Rovinj had the highest numbers of unique species, haplotypes and haplotype occurrences when comparing the three locations, followed by Livorno and then Palinuro (Table S3). Only 10 species were present with more than five occurrences in all three locations. Eight of these occurred in sufficient numbers to test differences between locations. AMOVA with a randomization test ($n=999$) using locations and sites as groups revealed two species having significant variance explained by location: the Porifera *Halisarca dujardinii* (Φ 0.084, 8.4% variance explained, $p=.001$) and the Porifera *Strongylacidon bermuda* (Φ 0.081, 8.1% variance explained, $p=.034$), see Table S4. However, the overwhelming majority of variance was contained within sites, and most sites had several unique haplotypes (i.e. haplotypes occurring strictly at a given site; hashed bars in Figure 5). Polychaeta had the highest proportion of unique haplotypes per site and was the most abundant class; Hydrozoa and Demospongiae had a lower proportion of unique haplotypes despite being abundant (Figure 5). Further, within-site variance was also apparent, i.e., differences between ARMS. In total, 197 haplotypes from 82 species were only found on one ARMS, mainly belonging to the most abundant classes.

### 3.4 | Intraspecific diversity patterns

Most species (84 of 145) had only one haplotype present, and 60 of those species were only found in one sample (i.e. one settlement

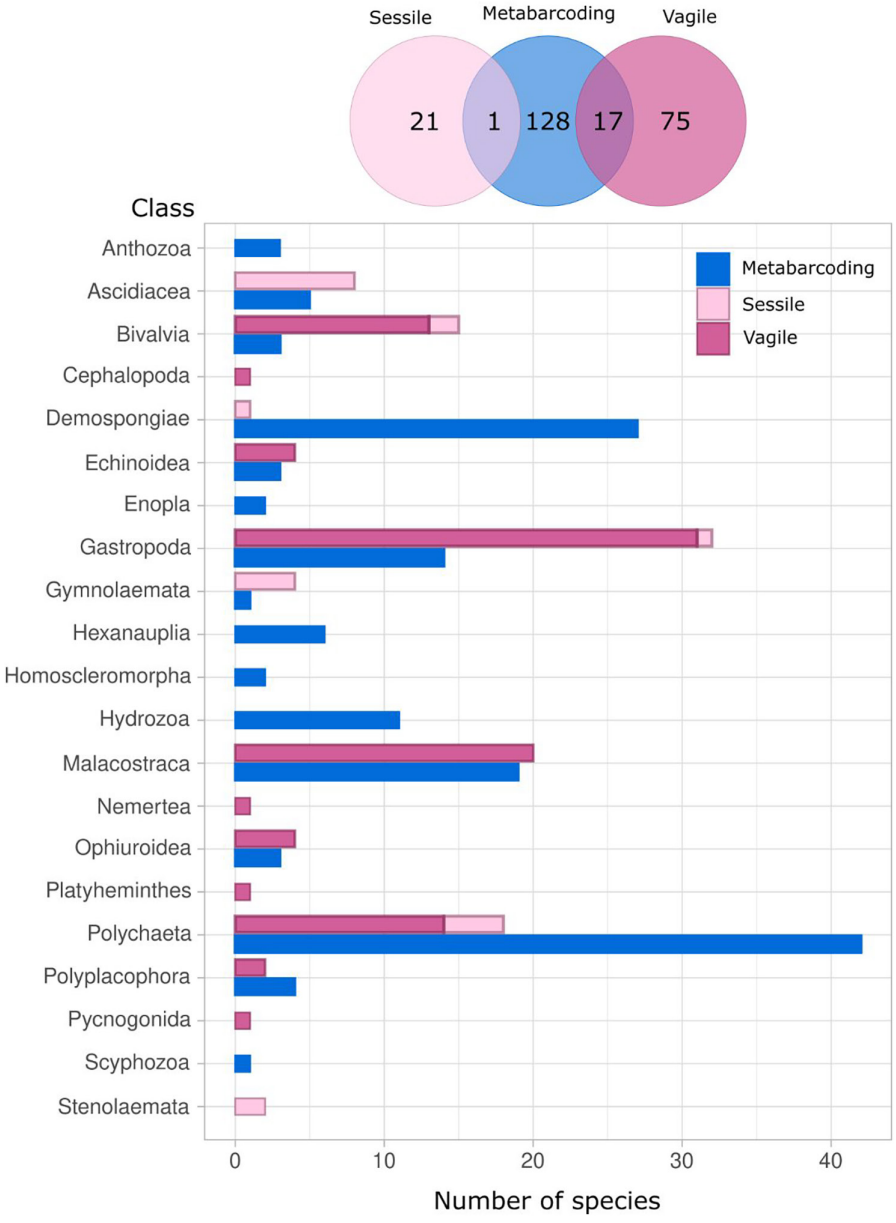



FIGURE 3 Comparison of species detected using morphological identification (sessile and vagile fractions) and metabarcoding methods at class level. Venn diagram illustrates the total number of species detected.

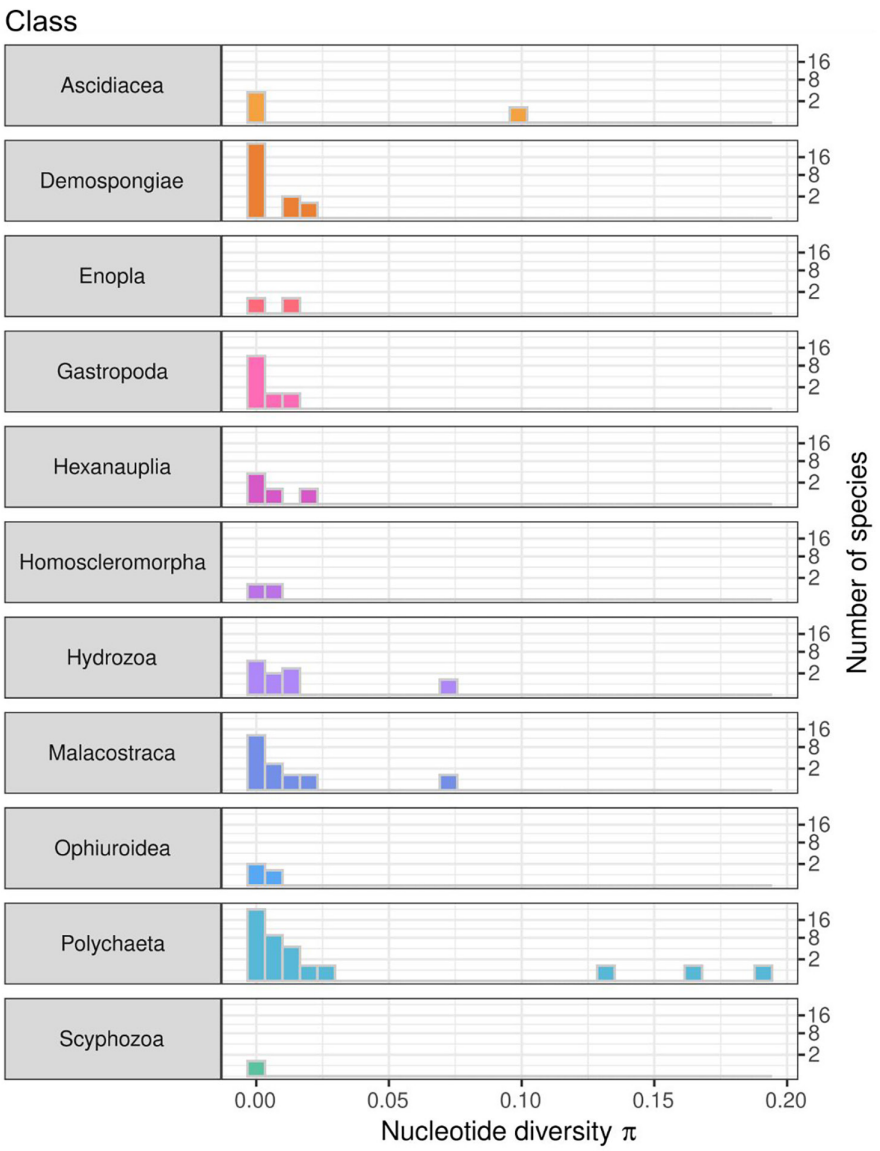



**FIGURE 4** Frequency of nucleotide diversity per species in classes with at least five haplotype occurrences.

plate or one natural substrate plot). Overall, nucleotide diversity was low across species, with some exceptions. A total of six species had nucleotide diversities >0.06 (Figure 4), Polychaeta: *Trypanosyllis zebra*, *Subadyte pellucida*, *Gyptis propinqua*; Ascidiacea: *Ciona edwardsi*; Malacostraca: *Eualus cranchii* and Hydrozoa: *Clytia gracilis*. Mean nucleotide diversity for all species was 0.0080 ± st. dev. 0.026 (0.019 ± 0.037 when excluding zeros). There was no significant difference in nucleotide diversity between classes with more than five haplotype occurrences and at least three species (Kruskal–Wallis rank sum test; $\chi^2 = 24.6$, df = 15, $p > .05$).

Haplotype networks for species with at least five haplotypes are available in Appendix S2.

## 3.5 | Species with high intraspecific variation

Five species clustered into four or more MOTUs; *Halisarca dujardinii* (Demospongiae), *Myrianida quindecimdentata* (Polychaeta), *Sabellaria spinulosa* (Polychaeta), *Eualus cranchii* (Malacostraca) and *Clytia gracilis* (Hydrozoa). Nucleotide diversity within these species ranged from 0.01 to 0.07, and maximum within-species distances were 7.4%–24.5% (K2P distances). Besides these species, the polychaetes *Trypanosyllis zebra*, *Subadyte pellucida*, *Eusyllis lamelligera*, *Hesiospina aurantiaca*, the hydroids *Campanularia hincksii* and *Bougainvillia muscus*, and the gastropod *Bittium reticulatum*, were the species with the highest nucleotide diversity (Figure 6) that have documented presence in the study regions according to Global Biodiversity Information Facility (GBIF; gbif.org).

## 3.6 | Comparison with GenBank sequences

Additional data from population genetic studies employing single-species methods in Mediterranean locations were available for four groups: *Ophiothrix fragilis* (Pérez-Portela et al., 2013), *Clytia* spp. (Cunha et al., 2017), *Platynereis dumerilii* and *P. massiliensis*

(Calosi et al., 2013; Wäge et al., 2017) and *Eualus* spp. (Conforti & Costantini, 2022). Results for *O. fragilis* and *Clytia* spp. are presented and discussed in more detail for comparative purposes; haplotype networks for *Platynereis* spp. and *Eualus* spp. are available in Appendix S2 (Figures S5 and S6).

### 3.6.1 | *Ophiothrix fragilis*

For the brittle star *Ophiothrix fragilis*, five haplotype sequences were found in the present study across the three study locations. These were combined with 124 COI sequences from nine Mediterranean locations collected between 2006 and 2011 (Pérez-Portela et al., 2013). After alignment and trimming, the combined data collapsed into 89 unique haplotypes with a length of 313 bp (Figure S7). In the present study, four haplotypes were new, and one had previously been found by Pérez-Portela et al. (2013) (Figure 7). Two distinct clusters emerged in the haplotype network, where all sequences found in this study belonged to the larger cluster (Figure 7). Pérez-Portela et al. (2013) described two distinct lineages in *Ophiothrix fragilis* where all haplotypes except BLA1 in lineage I originated from Atlantic samples. However, once trimmed to the overlapping 313-bp target amplicon, BLA1 grouped with the largest Mediterranean cluster (Figure 7).

### 3.6.2 | *Clytia gracilis* and *C. hemisphaerica*

For hydrozoans in the genus *Clytia*, *C. gracilis* and *C. hemisphaerica* were selected for further analysis due to previous indications that these species may present population structure or cryptic species (Cunha et al., 2017; Govindarajan et al., 2006). Seven *C. gracilis* and five *C. hemisphaerica* sequences from this study were combined with 17 sequences from Cunha et al. (2017) and Govindarajan et al. (2006). Sequences from Govindarajan et al. (2006) were removed due to the presence of ambiguous bases. The combined data

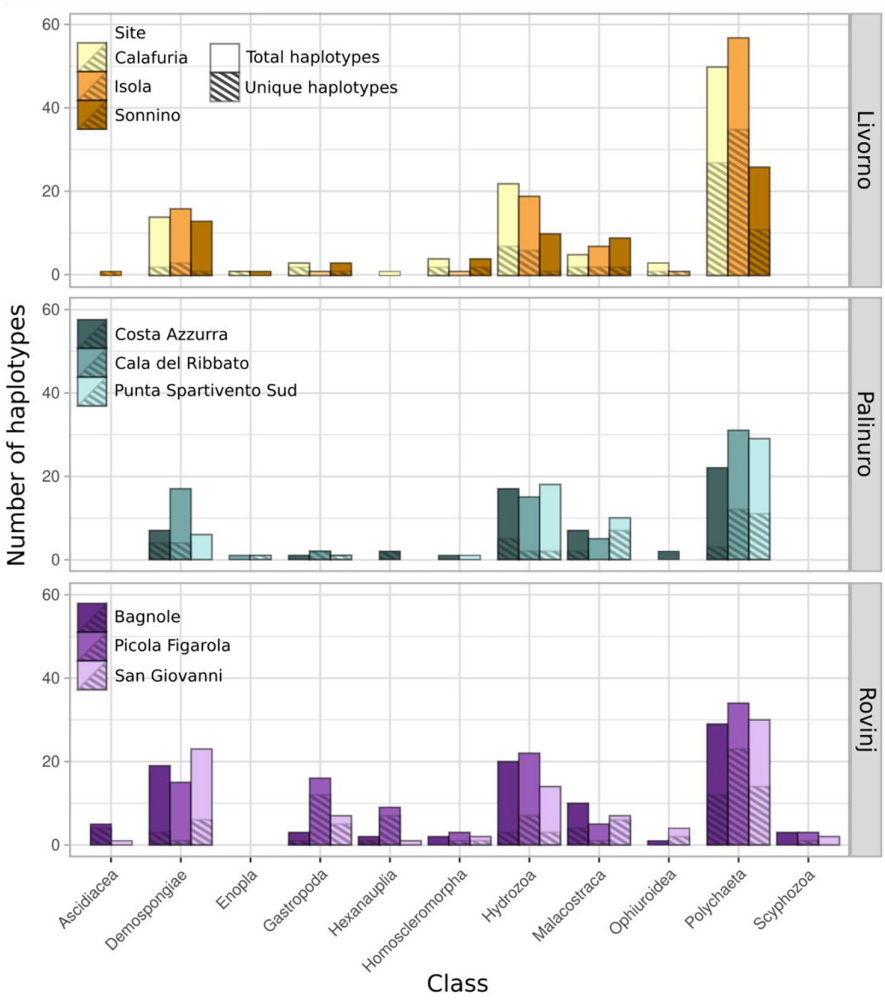
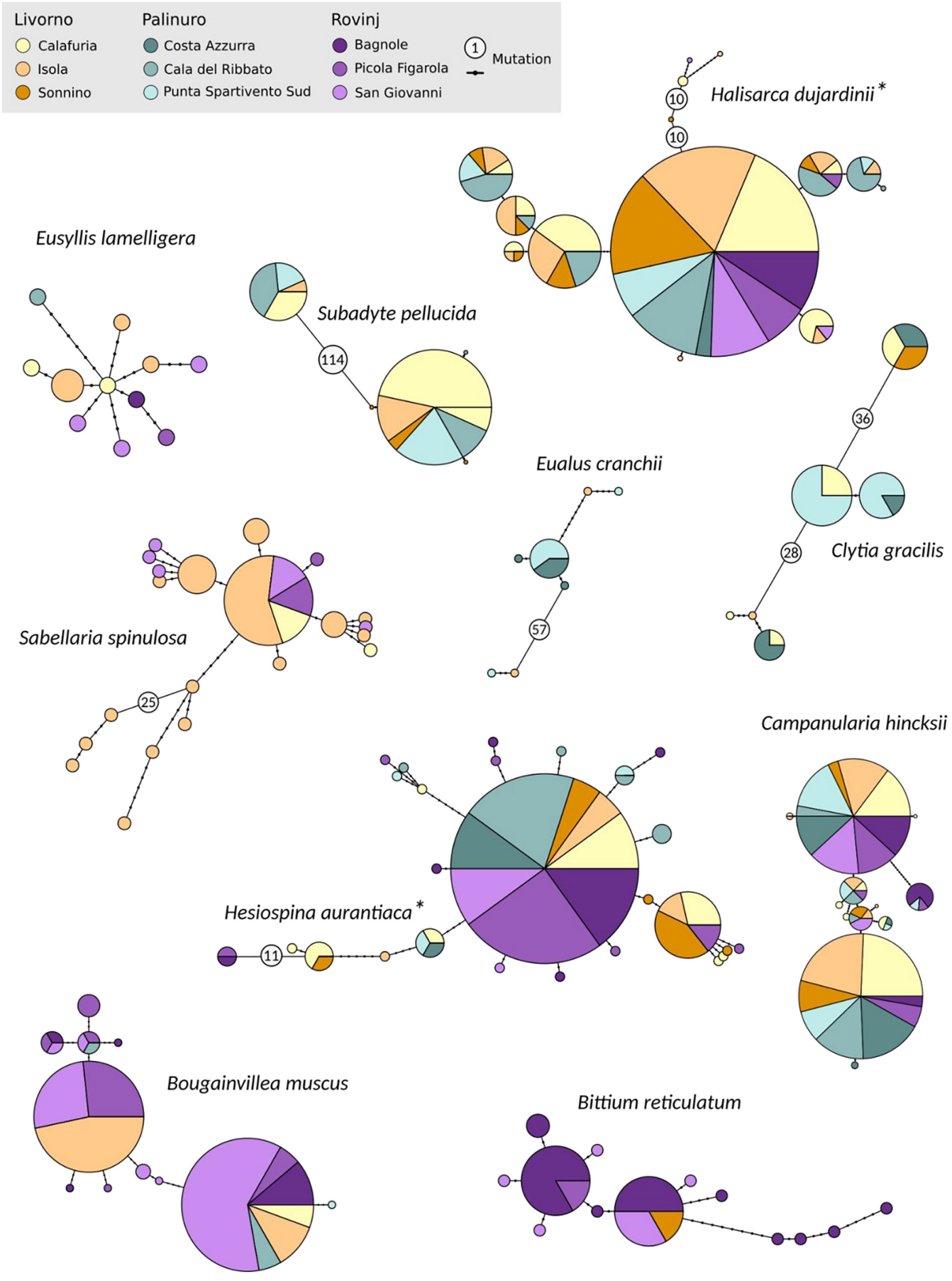



**FIGURE 5** Number of haplotypes per class, region and site (one site = 3 ARMS). Hashed bars represent site-unique haplotypes (i.e. do not occur in any other site but may occur on multiple ARMS within the site). Plain bars represent the total number of haplotypes occurring within the site. Haplotype occurrences are shown for classes with more than four occurrences across the dataset.

**FIGURE 6** Haplotype networks for the most diverse species in the dataset. Colours correspond to sites within regions. Species are shown if they have at least five haplotypes and confirmed distribution in the study region. Haplotype circle sizes are not to scale due to differences in abundance between species. *Region explained a significant amount of genetic variance.


Livorno
Calafuria
Isola
Sonnino
Palinuro
Costa Azzurra
Cala del Ribbato
Punta Spartivento Sud
Rovinj
Bagnole
Picola Figarola
San Giovanni
1 Mutation
Halisarca dujardinii *
10
10
Eusyllis lamelligera
Subadyte pellucida
114
Eualus cranchii
57
36
28
Clytia gracilis
Sabellaria spinulosa
25
Campanularia hincksii
Hesiospina aurantiaca *
11
Bougainvillea muscus
Bittium reticulatum

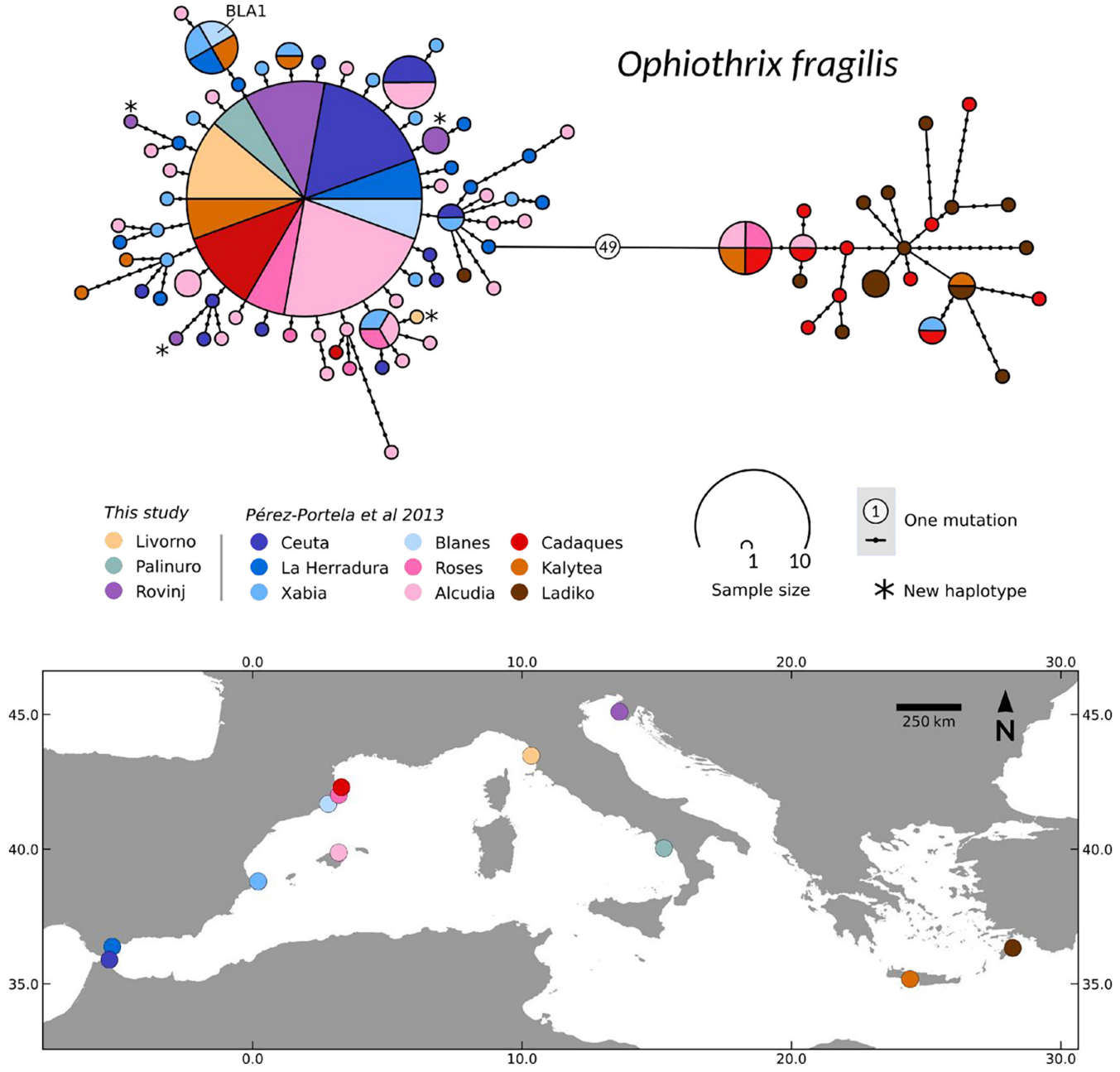



FIGURE 7 Haplotype network for *Ophiothrix fragilis* including sequences from the present study and Mediterranean locations in Pérez-Portela et al. (2013). The haplotype from lineage I in Pérez-Portela et al. (2013) has been marked, BLA1. Haplotypes introduced in this study are marked with an asterisk (*).

collapsed into 22 haplotypes of 313 bp. In this study, seven new and four previously described haplotypes were found (Figure 8). *Clytia gracilis* was polyphyletic in the combined data (Figure S8).

## 4 | DISCUSSION

Autonomous reef monitoring structures have been deployed throughout the world's oceans to monitor biodiversity and evaluate the influence of environmental and anthropogenic impacts on species richness (Obst et al., 2020; Pearman et al., 2020). Here, we use DNA metabarcoding data from ARMS deployed in highly diverse locations around Italian coastal waters in the Tyrrhenian and Adriatic Seas. With a combination of existing bioinformatics tools, we disentangle haplotypic variation at an intraspecific level across biogeographic regions. We reveal trends of genetic variability between and within both locations and sites that would otherwise be missed using traditional single-species methods. Our approach documents valuable genetic information for 145 benthic invertebrate species simultaneously and enables the first investigation of the genetic patterns of diversity for several species. This study highlights the potential of ARMS biodiversity surveys coupled with DNA

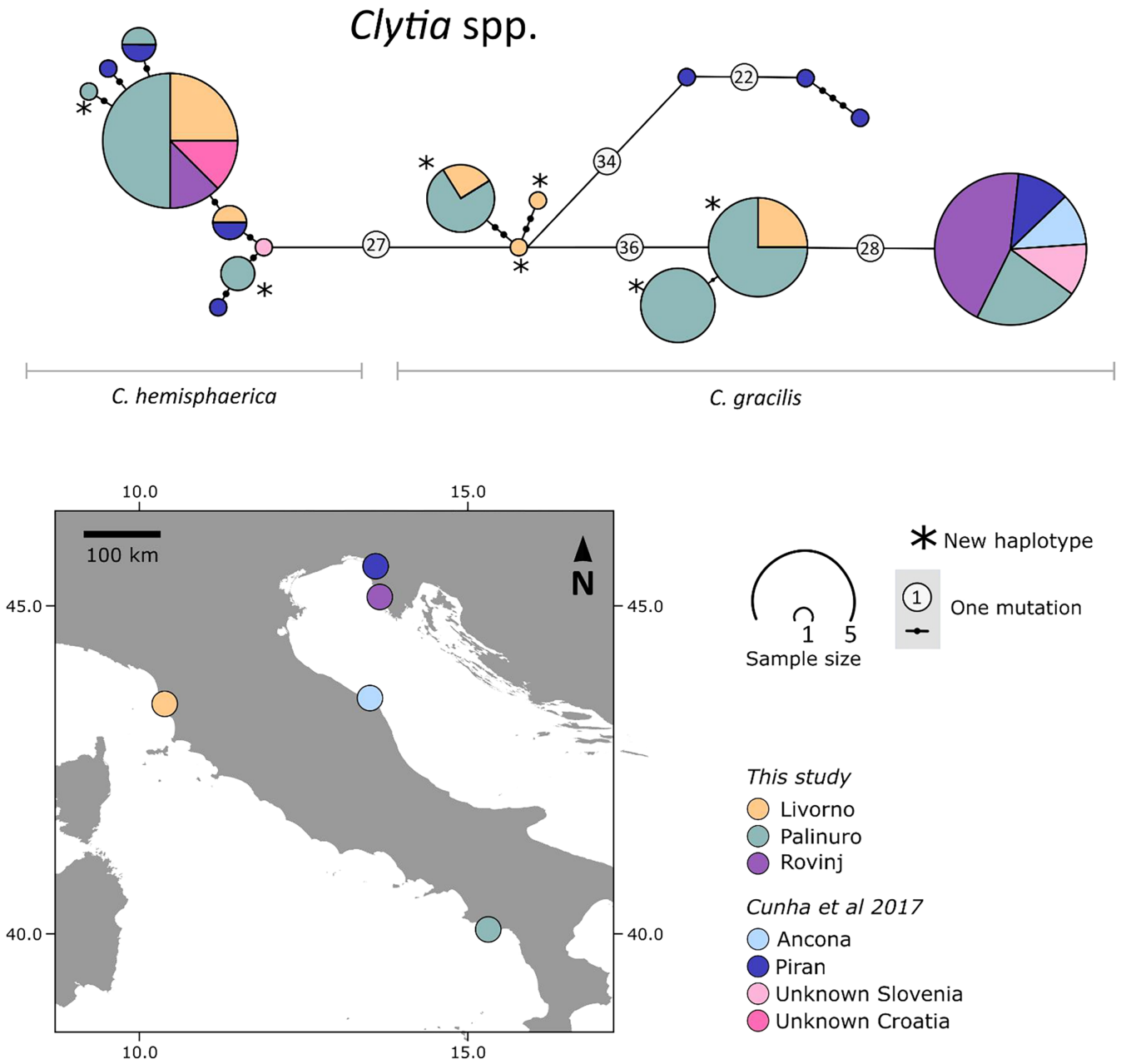



FIGURE 8 Haplotype network and map of sample sites for *Clytia gracilis* and *C. hemisphaerica* including sequences from the present study and Cunha et al. (2017). Haplotypes introduced in this study are marked with an asterisk (*).

metabarcoding-based haplotype analysis to reveal hidden genetic patterns among Mediterranean invertebrates.

### 4.1 | Comparison with morphological data

In addition to the molecular information, morphological data were collected from the ARMS. This included estimates of percentage cover of various sessile organism categories and morphological identification to at least genus level of vagile fauna ≥2 mm. Overall, species occurrences from the DNA metabarcoding data show slight overlap with morphological observations (Figure 3), a pattern which has been observed in similar studies (Cahill et al., 2018; Obst et al., 2020). In the present study, metabarcoding detected 128 species not found through the morphological identification methods, or alternatively, found but not identified at genus level. Conversely, metabarcoding methods failed to detect the most abundant genera in morphological observations. The fact that morphological identifications did not match well with the DNA metabarcoding data, despite the rarefaction curves being mostly saturated (Figure S3), highlights the potential limitations and discrepancies between the two approaches. The saturation of rarefaction curves suggests that the sequencing depth was sufficient to capture the majority of the species diversity in the samples. However, the mismatch between

the morphological and molecular data could be attributed to various factors.

Firstly, morphological identification may not be able to detect cryptic species, which are genetically distinct but morphologically indistinguishable, leading to an underestimation of true species diversity. Secondly, the DNA metabarcoding approach may detect species that are difficult to identify morphologically, such as sponges, or specimens otherwise overlooked due to their small size or rarity. Lastly, molecular methods may fail to identify—or misidentify—certain species due to, e.g., primer bias or mismatch (Cahill et al., 2018), insufficient sample amounts and incomplete or inaccurate reference databases (Mugnai et al., 2021).

Overall, these patterns underscore the importance of integrating both morphological and molecular methods to achieve a more comprehensive understanding of biodiversity in ecological studies.

### 4.2 | Community patterns across locations

Polychaeta, Demospongiae and Hydrozoa dominated observations in this study, accounting for 84% of metazoan haplotype occurrences. Many pioneer species, such as the sedentary polychaete *Sabellaria spinulosa* and the hydroid *Obelia dichotoma*, were found in the samples and sometimes with high haplotype occurrences. This was expected since ARMS are artificial substrates and were deployed for only 1 year, mainly allowing the growth of an early-stage community. Previous studies in the Adriatic and Tyrrhenian Seas have found similar classes and species as those observed in this study on artificial structures within the first year of deployment (Ardizzone et al., 1989; Ponti et al., 2015; Spagnolo et al., 2014).

### 4.3 | Location and site comparisons

The three locations had similar class compositions, with Polychaeta, Demospongiae and Hydrozoa constituting the majority of haplotypes identified. Rovinj, in the northern Adriatic Sea, had the greatest number of haplotypes and represented classes. In a similar study by Pearman et al. (2020), differences in community composition between Western Mediterranean and Adriatic locations were more apparent, with Arthropoda dominating in the former and Polychaeta in the latter subregions.

Many haplotypes were unique for the different sites within each location, indicating potential within-region variation in community composition and intraspecific genetic variability (Figure 5). This was particularly the case in Rovinj sites. Polychaetes had a larger number of site-unique haplotypes across the three locations, while most hydrozoan and demosponge haplotypes were shared between two or more sites. For example, the polychaete *Sabellaria spinulosa* had a large number of haplotypes in the Isola site (Livorno) but few in other sites and none in the Palinuro location (Figure 6). Other species, such as the hydrozoan *Campanularia hincksii*, had haplotypes represented in and shared between almost all sites (Figure 6). This highlights species-specific patterns in genetic diversity across the study area. Further, a comparatively large number of haplotypes were also only found on one ARMS unit. As such, variation at all spatial scales, from location to individual ARMS unit, was observed at species level. This brings an additional perspective to the within-site community differences presented by Pearman et al. (2020).

Assessment of location differences in species-specific genetic patterns was made difficult by the small number of haplotype occurrences remaining after stringent filtering, the overall low diversity and the fact that most species did not occur across the three locations. Despite this, eight abundant species were analysed during location comparison and two of these (both demosponges) had significant genetic variance explained by location, *Halisarca dujardinii* and *Strongylacidon bermuda*. Several invertebrate species show population genetic structure in the Mediterranean, sometimes relating to geographic barriers or isolation by distance (e.g. Costantini et al., 2018; Villamor et al., 2014). For some species, Villamor et al. (2014) indicated connectivity between the Tyrrhenian and Adriatic Seas, despite significant environmental differences between these areas. In the present study, we found no clear patterns of regional variation but highlight the potential of metabarcoding methods for this purpose (see Turon et al., 2020).

### 4.4 | Intraspecific diversity patterns

Most species had low COI diversity, with 58% of recorded species having only one haplotype. Moreover, 41% of described species were only found in one sample. Across all species, including those with a single occurrence, the average nucleotide diversity was 0.8% ± 2.6 SD. When excluding species with zero diversity, the corresponding number was 1.9% ± 3.7 SD. Similarly low intraspecific diversity was found by both Shum and Palumbi (2021) and Turon et al. (2020) using metabarcoding methods in the Pacific Ocean and Atlantic/Mediterranean Seas respectively. The early-stage communities on artificial structures can be expected to have lower diversity than established communities on natural substrates. In addition, diversity estimates can be influenced by the size of the gene fragment since the 313 bp used in this study is shorter than the 650 bp in conventional COI-based population genetics. In the present study, this was demonstrated in the case of *Ophiothrix fragilis*, where trimming the additional data from Pérez-Portela et al. (2013) collapsed 125 sequences to 90 haplotypes and removed the distinction between two known lineages (Figure S7). Further, COI inherently harbours low variability in several taxa, which can also explain the observed patterns. These taxa are typically at the base of the Metazoan tree (Huang et al., 2008) and include, for example, sea anemones (Shearer et al., 2002), corals (Calderón et al., 2006; Shearer et al., 2002) and demosponges (Erpenbeck et al., 2016). Apart from the class Demospongiae, this aligns with the observed patterns in the present study (Figures 4 and 5).

At the other end of the spectrum, some instances of high intraspecific diversity were found. High nucleotide and haplotype

diversities were observed in Polychaeta, Malacostraca, Hydrozoa and Demospongiae, which also were the most abundant classes. Haplotype diversity per species significantly increased with increasing haplotype occurrences (Figure S9), while nucleotide diversity unexpectedly did not (Figure S10). For example, some abundant species had only one haplotype or low nucleotide diversity (e.g. the Porifera *Oscarella lobularis* and the Arthropoda Decapoda *Synalpheus gambarelloides*). Conversely, the Polychaeta *Trypanosyllis zebra*, with only five haplotype occurrences, had a nucleotide diversity of 0.19. This was the highest nucleotide diversity observed in our samples, and an outlier compared to other species. Álvarez-Campos et al. (2017) concluded that *T. zebra* in the Mediterranean may consist of a species complex with genetic distances ranging between 10.5 and 27.4% among *Trypanosyllis* pseudocryptic lineages, which was lower than the maximum "intraspecific" distance found in this study (42.6%; Appendix S1).

Another species with unexpected genetic patterns was the Ross worm, *Sabellaria spinulosa*, an abundant reef-building polychaete in the Mediterranean Sea (Schimmenti et al., 2015) which had many haplotype occurrences in our samples. The abundance was expected as it is a pioneer species, and its presence was confirmed by the morphology-based records. More unexpectedly, however, this study presents relatively high nucleotide diversity (0.025) and very high intraspecific distances (up to 10.7%) for this species. For comparison, a previous study of *S. spinulosa* in the Mediterranean found the intraspecific distances to be only 0.9% (Schimmenti et al., 2015) employing a longer COI fragment (610 bp). This highlights potential population structure or crypticity within an important yet understudied species in this region. Similarly, the polychaete *Subadyte pellucida* had extremely high nucleotide diversity (16.1%) and has been found living on or attached to other organisms such as coral colonies (Mastrototaro et al., 2010), sponges (Goren et al., 2021), anemones (Mangano et al., 2010) and sea stars/brittle stars (e.g. *Ophiothrix fragilis*; Pettibone, 1993). However, this species has not been studied at a population genetic level in the Mediterranean and requires further investigation related to their host specialization and diversity. These results demonstrate the benefit of metabarcoding-based methods to identify potential cryptic lineages and highlight prospective avenues of research in species discovery and distribution.

On a species level, we were able to compare our findings with existing studies of four groups: the brittle star *Ophiothrix fragilis*, the annelid polychaetes *Platynereis dumerilii* and *P. massiliensis*, the hydrozoans *Clytia gracilis* and *C. hemisphaerica* and the shrimp *Eualus* spp. We found both new and previously described haplotypes in our dataset for all species groups, and further provide records from previously unstudied locations. The findings in the present study aligned with results from studies applying traditional methods, for which two examples are highlighted here. In the case of *C. gracilis* and *C. hemisphaerica*, we find that *C. gracilis* forms a polyphyletic clade (Figure 8; Figure S10) and therefore contend this species consists of a species complex in accordance with Cunha et al. (2017) and Govindarajan et al. (2006). For *O. fragilis*, Pérez-Portela et al. (2013) suggested the existence of two distinct lineages separated by genetic distances of at least 15.8% (*p* distances). Lineage I was observed almost exclusively in the Atlantic Ocean and lineage II in the Mediterranean Sea, except for one sequence from lineage I collected in Blanes, Spain (BLA1; Figure 7). However, when sequences were trimmed to 313 bp and collapsed, the lineage I sequence BLA1 clustered with lineage II haplotypes. Upon further inspection, any distinction between the two lineages disappeared when using the shorter sequence (Figure S7). As such, this COI fragment failed to detect significant population structure within *O. fragilis* and highlights target gene limitations. Haplotype networks combining data from the current and comparable studies for *Eualus* spp. and *Platynereis* spp. are available in Appendix S2 (Figures S5 and S6).

### 4.5 | Method limitations

The mismatch between morphological and metabarcoding records and the presence of extreme outliers in nucleotide diversity suggest that the metabarcoding methods may fail to detect, or mislabel, certain species. For example, our data contain some unlikely taxonomic assignments, such as the demosponge *Strongylacidon bermuda*. This species was the most abundant demosponge across samples but has previously only been recorded in North America (de Voogd et al., 2022). Unless this is a recently introduced and highly successful non-indigenous species, which we consider unlikely, its presence in our dataset is almost certainly some form of error. In this case, the error is likely due to incomplete public databases, which is a common issue in DNA metabarcoding (Mugnai et al., 2021). In our study, it is exemplified by the small proportion of metazoan haplotypes that received taxonomic assignment at species level (less than 10%) despite using the BOLD database. In addition, for the species we identified, we find very limited sequence data from the Mediterranean region for comparisons in GenBank, which further highlights the regional discrepancy in database coverage.

Intraspecific studies involving data from ARMS will have some inherent differences from traditional approaches. ARMS units provide a snapshot of diversity, capturing only those species that settle onto the plates during the deployment period. This design could potentially lead to undersampling of species diversity, which may affect the overall assessment of community composition. On the other hand, haplotype diversity refers to the genetic variation within species or populations. The observed haplotype patterns in our study may be influenced by factors such as natural population structure, selection pressures and study design, rather than solely undersampling. The single haplotypes observed for many species could reflect the true genetic structure of these populations. By recognizing the difference between these two aspects of diversity, we can better interpret our study results and understand the potential limitations associated with the ARMS units and the chosen molecular techniques. Future studies should consider incorporating additional sampling methods and strategies to capture a more comprehensive representation of both species and haplotype diversity.

Furthermore, metabarcoding approaches include bulk samples, in this case, material scraped from plates, and the sample composition can affect the outcome. In the present study, we excluded macrophyte taxa during analysis following our research objective, despite rhodophytes constituting a significant portion of the biomass. Biomass-dominant taxa can affect the read abundances of both rarer and smaller organisms, and therefore limit the probability of their detection (Elbrecht et al., 2017; Leray & Knowlton, 2015). Subdividing scraped material by size fraction may, therefore, be useful for future studies when not all target organisms constitute a considerable proportion of the biomass (Wangensteen et al., 2018). Lastly, any bioinformatic pipeline will have trade-offs when attempting to remove erroneous haplotypes from the dataset. For the present study, we chose a stringent filtering approach which almost certainly excludes rare species or haplotypes and therefore prevents more in-depth analysis such as demographic inferences (Shum & Palumbi, 2021). Incorporating positive controls with intraspecific variation in the experimental design would also be beneficial for pipeline development, which was not the case with the current dataset. As such, both bioinformatics methods and study design should be chosen carefully depending on the research objectives moving forward.

## 5 | CONCLUSION

In this study, we recover haplotype information for hundreds of Mediterranean marine invertebrate species simultaneously. This was achieved by implementing COI-based metabarcoding on ARMS deployed in coastal marine ecosystems. Our haplotype findings overlap with comparable published studies using single-species population genetic methods. In general, we find low intraspecific diversity across several invertebrate classes in the three study locations in the central Mediterranean Sea. However, we also identify species with high diversity that warrant further investigation regarding potential crypticity and population structure. In addition, we highlight the presence of unique haplotypes at all spatial scales for almost all classes studied, indicating small-scale genetic variability within the invertebrate species presented here. We demonstrate DNA metabarcoding as an important tool to generate intraspecific genetic information, yet emphasize the limitations of these methods, as they rely on taxonomic accuracy in barcode reference databases. The approach presented here greatly enhances DNA metabarcoding methods to reveal interspecific COI variation both within and between ARMS units. This has the potential to strengthen species monitoring across vast scales that help track geographical range shifts and climate-related impacts on biodiversity.

### AUTHOR CONTRIBUTIONS

Anna Thomasdotter developed the metabarcoding protocol, abstracted and analysed data and wrote the manuscript. Peter Shum contributed significantly to the development of the metabarcoding protocol and revised the manuscript. Francesco Mugnai performed the molecular analysis in the laboratory and edited the manuscript. Marina Vingiani performed some laboratory analysis and reviewed the manuscript. Marco Abbiati helped with the field activities and edited the manuscript. Federica Costantini conceived the experiment, performed the field activities and the sample processing and wrote, reviewed and edited the manuscript.

### ACKNOWLEDGEMENTS

This work was supported by the European project SEAMoBB: "Solutions for Semi Automated Monitoring of Benthic Biodiversity" funded by ERA-NET MarT-ERA and the Ministry of University and Research (MUR). Anna Thomasdotter was supported by the IMBRSea master program (International Master of Science in Marine Biological Resources), as an Erasmus Mundus Joint Master's Degree (EMJMD) scholarship Awardee. Special thanks to Dr. Kenan O. Matterson for the field sampling and the molecular-related contributions, bioinformatics included. We are grateful to Francesco Paolo Mancuso for his help during the field activity during the first deployment of ARMS. We would like to thank all of the Marine Biology M.Sc. students involved in SEAMoBB project field- and lab-related activities during the 2019 and 2020 campaigns: Giuditta Carbone, Cinzia Cozzulla, Gabriele Giacalone, Alessandro Giordano, Massimiliano Lippi, Giorgia Olivieri, Alberto Orengo, India Ronchi and Alessandra Silvestri. Many thanks to Barbara Mikac for the identification of the Polychaetes in the vagile fauna fraction. Finally, we would like to express our thanks to the diving centres that supported us during the field activities (Rovinj Sub, Rovinj, Croatia; Palinuro Sub, Palinuro, Italy; Accademia Blu Diving Center, Livorno, Italy). We would also like to thank the CIBM, Centro Interuniversitario di Biologia Marina ed Ecologia Applicata "G.Bacci", Livorno, Italy, the Ruđer Bošković Institute, Rovinj, Croatia and the Hotel Le Tre Caravelle, Palinuro (SA), Italy for their support during the sample processing in the field.

### CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

### DATA AVAILABILITY STATEMENT

Sequence data and pipeline step outputs have been deposited to Zenodo and are available at https://zenodo.org/record/7781906. Scripts used for the bioinformatics pipeline, data analysis and visualization are available at https://github.com/thomasdotter/spineless-haplotypes.

### BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

### ORCID

*Anna Thomasdotter* https://orcid.org/0000-0003-4501-9656

*Peter Shum* https://orcid.org/0000-0001-8154-9828
*Marina Vingiani* https://orcid.org/0000-0003-4154-8673
*Marco Abbiati* https://orcid.org/0000-0002-2426-4524
*Federica Costantini* https://orcid.org/0000-0002-8813-1923

## REFERENCES

Álvarez-Campos, P., Giribet, G., San Martín, G., Rouse, G. W., & Riesgo, A. (2017). Straightening the striped chaos: Systematics and evolution of *Trypanosyllis* and the case of its pseudocryptic type species *Trypanosyllis krohnii* (Annelida, Syllidae). *Zoological Journal of the Linnean Society*, *179*(3), 492–540. https://doi.org/10.1111/zoj.12443

Andújar, C., Creedy, T. J., Arribas, P., López, H., Salces-Castellano, A., Pérez-Delgado, A. J., Vogler, A. P., & Emerson, B. C. (2021). Validated removal of nuclear pseudogenes and sequencing artefacts from mitochondrial metabarcode data. *Molecular Ecology Resources*, *21*(6), 1772–1787. https://doi.org/10.1111/1755-0998.13337

Antich, A., Palacín, C., Turon, X., & Wangensteen, O. S. (2022). DnoisE: Distance denoising by entropy. An open-source parallelizable alternative for denoising sequence datasets. *PeerJ*, *10*, e12758. https://doi.org/10.7717/peerj.12758

Antich, A., Palacín, C., Wangensteen, O. S., & Turon, X. (2021). To denoise or to cluster that is not the question: Optimizing pipelines for COI metabarcoding and metaphylogeography. *BMC Bioinformatics*, *22*, 177. https://doi.org/10.1186/s12859-021-04115-6

Ardizzone, G. D., Gravina, M. F., & Belluscio, A. (1989). Temporal development of epibenthic communities on artificial reefs in the Central Mediterranean Sea. *Bulletin of Marine Science*, *44*(2), 592–608.

Azarian, C., Foster, S., Devloo-Delva, F., & Feutry, P. (2021). Population differentiation from environmental DNA: Investigating the potential of haplotype presence/absence-based analysis of molecular variance. *Environmental DNA*, *3*(3), 541–552. https://doi.org/10.1002/edn3.143

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). Obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182.

Brandt, M. I., Trouche, B., Quintric, L., Günther, B., Wincker, P., Poulain, J., & Arnaud-Haond, S. (2021). Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. *Molecular Ecology Resources*, *21*(6), 1904–1921. https://doi.org/10.1111/1755-0998.13398

Buchner, D., & Leese, F. (2020). BOLDigger – A python package to identify and organise sequences with the barcode of life data systems. *Metabarcoding and Metagenomics*, *4*, e53535. https://doi.org/10.3897/mbmg.4.53535

Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J.-P., Olenin, S., Šiaulys, A., Somerfield, P. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, *8*(17), 8908–8920. https://doi.org/10.1002/ece3.4283

Calderón, I., Garrabou, J., & Aurelle, D. (2006). Evaluation of the utility of COI and ITS markers as tools for population genetic studies of temperate gorgonians. *Journal of Experimental Marine Biology and Ecology*, *336*(2), 184–197. https://doi.org/10.1016/j.jembe.2006.05.006

Calosi, P., Rastrick, S. P. S., Lombardi, C., de Guzman, H. J., Davidson, L., Jahnke, M., Giangrande, A., Hardege, J. D., Schulze, A., Spicer, J. I., & Gambi, M. (2013). Adaptation and acclimatization to ocean acidification in marine ectotherms: An in situ transplant experiment with polychaetes at a shallow $CO_2$ vent system. *Philosophical Transactions of the Royal Society. Series B, Biological Sciences*, *368*(1627), 1–15. https://doi.org/10.1098/rstb.2012.0444

Coll, M., Piroddi, C., Albouy, C., Ben Rais Lasram, F., Cheung, W. W. L., Christensen, V., Karpouzi, V. S., Guilhaumon, F., Mouillot, D., Paleczny, M., Palomares, M. L., Steenbeek, J., Trujillo, P., Watson, R., & Pauly, D. (2012). The mediterranean sea under siege: Spatial overlap between marine biodiversity, cumulative threats and marine reserves. *Global Ecology and Biogeography*, *21*(4), 465–480. https://doi.org/10.1111/j.1466-8238.2011.00697.x

Conforti, S., & Costantini, F. (2022). Potential crypticity within two decapod (crustacea) genera: *Galathea* Fabricius, 1793 and *Eualus* Thallwitz, 1891 suggested by integrative taxonomic approach. *Mediterranean Marine Science*, *23*(3), 499–524. https://doi.org/10.12681/mms.27733

Costantini, F., Ferrario, F., & Abbiati, M. (2018). Chasing genetic structure in coralligenous reef invertebrates: Patterns, criticalities and conservation issues. *Scientific Reports*, *8*, 5844. https://doi.org/10.1038/s41598-018-24247-9

Cunha, A. F., Collins, A. G., & Marques, A. C. (2017). Phylogenetic relationships of *Proboscoida* broch, 1910 (cnidaria, hydrozoa): Are traditional morphological diagnostic characters relevant for the delimitation of lineages at the species, genus, and family levels? *Molecular Phylogenetics and Evolution*, *106*, 118–135. https://doi.org/10.1016/j.ympev.2016.09.012

Darnaude, A. M., Arnaud-Haond, S., Hunter, E., Gaggiotti, O., Sturrock, A., Beger, M., Volckaert, F., Pérez-Ruzafa, A., López-López, L., Tanner, S. E., Turan, C., Doğdu, S. A., Katsanevakis, S., & Costantini, F. (2022). Unifying approaches to functional marine connectivity for improved marine resource management: The European SEA-UNICORN COST action. *Research Ideas and Outcome Journal*, *8*(2), 339–321. https://doi.org/10.3897/rio.8.e80223

Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, *6*, 226–230.

de Voogd, N. J., Alvarez, B., Boury-Esnault, N., Carballo, J. L., Cárdenas, P., Díaz, M., Dohrmann, M., Downey, R., Goodwin, C., Hajdu, E., Hooper, J. N. A., Kelly, M., Klautau, M., Lim, S. C., Manconi, R., Morrow, C., Pinheiro, U., Pisera, A. B., Ríos, P., … Xavier, J. (2022). *Strongylacidon bermuda* (de Laubenfels, 1950). *World porifera database*. https://www.marinespecies.org/porifera/porifera.php?p=taxdetails&id=168796

Dray, S., & Dufour, A. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, *22*(4), 1–20. https://doi.org/10.18637/jss.v022.i04

DuBois, K., Pollard, K. N., Kauffman, B. J., Williams, S. L., & Stachowicz, J. J. (2022). Local adaptation in a marine foundation species: Implications for resilience to future global change. *Global Change Biology*, *28*(8), 2596–2610. https://doi.org/10.1111/gcb.16080

Edgar, R. C. (2004). MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, *5*(1), 113. https://doi.org/10.1186/1471-2105-5-113

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., Leese, F., & Yu, D. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), 1265–1275. https://doi.org/10.1111/2041-210X.12789

Elbrecht, V., Vamos, E. E., Steinke, D., & Leese, F. (2018). Estimating intraspecific genetic diversity from community DNA metabarcoding data. *PeerJ*, *6*, e4644. https://doi.org/10.7717/peerj.4644

Erpenbeck, D., Voigt, O., Al-Aidaroos, A. M., Berumen, M. L., Büttner, G., Catania, D., Guirguis, A. N., Paulay, G., Schätzle, S., & Wörheide, G. (2016). Molecular biodiversity of Red Sea demosponges. *Marine Pollution Bulletin*, *105*(2), 507–514. https://doi.org/10.1016/j.marpolbul.2015.12.004

Frøslev, T. G., Kjøller, R., Bruun, H. H., Ejrnæs, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*, 1188. https://doi.org/10.1038/s41467-017-01312-x

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular Ecology Resources*, *13*(5), 851–861.

Goren, L., Idan, T., Shefer, S., & Ilan, M. (2021). Macrofauna inhabiting massive demosponges from shallow and mesophotic habitats along the Israeli Mediterranean coast. *Frontiers in Marine Science*, *7*, 1–14. https://doi.org/10.3389/fmars.2020.612779

Govindarajan, A. F., Boero, F., & Halanych, K. M. (2006). Phylogenetic analysis with multiple markers indicates repeated loss of the adult medusa stage in Campanulariidae (hydrozoa, cnidaria). *Molecular Phylogenetics and Evolution*, *38*(3), 820–834. https://doi.org/10.1016/j.ympev.2005.11.012

Huang, D., Meier, R., Todd, P. A., & Chou, L. M. (2008). Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *Journal of Molecular Evolution*, *66*(2), 167–174. https://doi.org/10.1007/s00239-008-9069-5

Kupriyanova, E. K., & Jirkov, I. A. (1997). Serpulidae (Annelida, polychaeta) of the Arctic Ocean. *Sarsia*, *82*(3), 203–236. https://doi.org/10.1080/00364827.1997.10413651

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., Boehm, J. T., & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, *10*(1), 34.

Macé, B., Hocdé, R., Marques, V., Guerin, P., Valentini, A., Arnal, V., Pellissier, L., & Manel, S. (2021). Evaluating bioinformatics pipelines for population-level inference using environmental DNA. *Environmental DNA*, *4*, 674–686. https://doi.org/10.1002/edn3.269

Mahé, F., Czech, L., Stamatakis, A., Quince, C., de Vargas, C., Dunthorn, M., & Rognes, T. (2022). Swarm v3: Towards tera-scale amplicon clustering. *Bioinformatics*, *38*(1), 267–269. https://doi.org/10.1093/bioinformatics/btab493

Mangano, M. C., Cosentino, A., De Domenico, F., & Spanò, N. (2010). *Antedon mediterranea* (Echinodermata: Crinoidea) as host of commensal *Subadyte pellucida* (Annelida: Polychaeta). *Biologia Marina Mediterranea*, *17*(1), 302–303.

Mastrototaro, F., D'Onghia, G., Corriero, G., Matarrese, A., Maiorano, P., Panetta, P., Gherardi, M., Longo, C., Rosso, A., Sciuto, F., Sanfilippo, R., Gravili, C., Boero, F., Taviani, M., & Tursi, A. (2010). Biodiversity of the white coral bank off Cape Santa Maria di Leuca (Mediterranean Sea): An update. *Deep Sea Research Part II: Topical Studies in Oceanography*, *57*(5), 412–430. https://doi.org/10.1016/j.dsr2.2009.08.021

Micheli, F., Halpern, B. S., Walbridge, S., Ciriaco, S., Ferretti, F., Fraschetti, S., Lewison, R., Nykjaer, L., & Rosenberg, A. A. (2013). Cumulative human impacts on Mediterranean and Black Sea marine ecosystems: Assessing current pressures and opportunities. *PLoS One*, *8*(12), e79889. https://doi.org/10.1371/journal.pone.0079889

Mugnai, F., Meglécz, E., Abbiati, M., Bavestrello, G., Bertasi, F., Bo, M., Capa, M., Chenuil, A., Colangelo, M. A., De Clerck, O., Gutiérrez, J. M., Lattanzi, L., Leduc, M., Martin, D., Matterson, K. O., Mikac, B., Plaisance, L., Ponti, M., Riesgo, A., … Costantini, F. (2021). Are well-studied marine biodiversity hotspots still blackspots for animal barcoding? *Global Ecology and Conservation*, *32*, e01909. https://doi.org/10.1016/j.gecco.2021.e01909

Nichols, P. K., Timmers, M., & Marko, P. B. (2022). Hide 'n seq: Direct versus indirect metabarcoding of coral reef cryptic communities. *Environmental DNA*, *4*(1), 93–107. https://doi.org/10.1002/edn3.203

Obst, M., Exter, K., Allcock, A. L., Arvanitidis, C., Axberg, A., Bustamante, M., Cancio, I., Carreira-Flores, D., Chatzinikolaou, E., Chatzigeorgiou, G., Chrismas, N., Clark, M. S., Comtet, T., Dailianis, T., Davies, N., Deneudt, K., de Cerio, O. D., Fortič, A., Gerovasileiou, V., … Pavloudi, C. (2020). A marine biodiversity observation network for genetic monitoring of hard-bottom communities (ARMS-MBON). *Frontiers in Marine Science*, *7*, 572680. https://doi.org/10.3389/fmars.2020.572680

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Michin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2022). *vegan: Community ecology package. R package version 2.6–4*. https://CRAN.R-project.org/package=vegan

Paradis, E. (2010). Pegas: An R package for population genetics with an integrated–modular approach. *Bioinformatics*, *26*(3), 419–420. https://doi.org/10.1093/bioinformatics/btp696

Patarnello, T., Volckaert, F. A., & Castilho, R. (2007). Pillars of Hercules: Is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, *16*(21), 4426–4444. https://doi.org/10.1111/j.1365-294X.2007.03477.x

Pearman, J. K., Chust, G., Aylagas, E., Villarino, E., Watson, J. R., Chenuil, A., Borja, A., Cahill, A. E., Carugati, L., Danovaro, R., David, R., Irigoien, X., Mendibil, I., Moncheva, S., Rodríguez-Ezpeleta, N., Uyarra, M. C., & Carvalho, S. (2020). Pan-regional marine benthic cryptobiome biodiversity patterns revealed by metabarcoding autonomous reef monitoring structures. *Molecular Ecology*, *29*(24), 4882–4897. https://doi.org/10.1111/mec.15692

Pérez-Portela, R., Almada, V., & Turon, X. (2013). Cryptic speciation and genetic structure of widely distributed brittle stars (Ophiuroidea) in Europe. *Zoologica Scripta*, *42*(2), 151–169. https://doi.org/10.1111/j.1463-6409.2012.00573.x

Pettibone, M. H. (1993). Scaled polychaetes (Polynoidae) associated with ophiuroids and other invertebrates and review of species referred to *Malmgrenia* McIntosh and replaced by *Malmgreniella* Hartman, with descriptions of new taxa. *Smithsonian Contributions to Zoology*, *538*, 1–92. https://doi.org/10.5479/si.00810282.538

Pinsky, M. L., Selden, R. L., & Kitchel, Z. J. (2020). Climate-driven shifts in marine species ranges: Scaling from organisms to communities extinction. *The Global Disappearance of a Species*, *12*, 153–179. https://doi.org/10.1146/annurev-marine-010419-010916

Ponti, M., Fava, F., Perlini, R. A., Giovanardi, O., & Abbiati, M. (2015). Benthic assemblages on artificial reefs in the northwestern Adriatic Sea: Does structure type and age matter? *Marine Environmental Research*, *104*, 10–19. https://doi.org/10.1016/j.marenvres.2014.12.004

Porter, T. M., & Hajibabaei, M. (2020). Putting COI metabarcoding in context: The utility of exact sequence variants (ESVs) in biodiversity analysis. *Frontiers in Ecology and Evolution*, *8*, 248. https://doi.org/10.3389/fevo.2020.00248

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Reusch, T. B. H., Ehlers, A., Hämmerli, A., & Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(8), 2826–2831. https://doi.org/10.1073/pnas.0500008102

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584.

Schimmenti, E., Lo Brutto, S., Badalamenti, F., Giangrande, A., Mikac, B., & Musco, L. (2015). DNA-barcoding to solve the tricky case of co-occurring *Sabellaria* (Annelida) species in the Mediterranean Sea. *Biologia Marina Mediterranea*, *22*(1), 109–110.

Shearer, T. L., van Oppen, M. J. H., Romano, S. L., & Wörheide, G. (2002). Slow mitochondrial DNA sequence evolution in the Anthozoa (cnidaria). *Molecular Ecology*, *11*(12), 2475–2487. https://doi.org/10.1046/j.1365-294X.2002.01652.x

Shum, P., & Palumbi, S. R. (2021). Testing small-scale ecological gradients and intraspecific differentiation for hundreds of kelp forest species using haplotypes from metabarcoding. *Molecular Ecology*, *30*(13), 3355–3373. https://doi.org/10.1111/mec.15851

Spagnolo, A., Cuicchi, C., Punzo, E., Santelli, A., Scarcella, G., & Fabi, G. (2014). Patterns of colonization and succession of benthic assemblages in two artificial substrates. *Journal of Sea Research*, *88*, 78–86. https://doi.org/10.1016/j.seares.2014.01.007

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*(7), 3022–3027. https://doi.org/10.1093/molbev/msab120

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From metabarcoding to metaphylogeography: Separating the wheat from the chaff. *Ecological Applications*, *30*(2), e02036. https://doi.org/10.1002/eap.2036

Villamor, A., Costantini, F., & Abbiati, M. (2014). Genetic structuring across marine biogeographic boundaries in rocky shore invertebrates. *PLoS One*, *9*(7), e101135. https://doi.org/10.1371/journal.pone.0101135

Wäge, J., Valvassori, G., Hardege, J. D., Schulze, A., & Gambi, M. C. (2017). The sibling polychaetes *Platynereis dumerilii* and *Platynereis massiliensis* in the Mediterranean Sea: Are phylogeographic patterns related to exposure to ocean acidification? *Marine Biology*, *164*(10), 1–12. https://doi.org/10.1007/s00227-017-3222-x

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wangensteen, O.S., & Turon, X. (2016). Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests. In: Rossi, S., Bramanti, L., Gori, A., Orejas Saco del Valle, C. (Eds.), *Marine Animal Forests*.Springer.https://doi.org/10.1007/978-3-319-17001-5_53-1

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, *6*, e4705. https://doi.org/10.7717/peerj.4705

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Thomasdotter, A., Shum, P., Mugnai, F., Vingiani, M., Abbiati, M., & Costantini, F. (2023). Spineless and overlooked: DNA metabarcoding of autonomous reef monitoring structures reveals intra- and interspecific genetic diversity in Mediterranean invertebrates. *Molecular Ecology Resources*, *00*, 1–17. https://doi.org/10.1111/1755-0998.13836