MOLECULAR ECOLOGY RESOURCES

Supplemental Information for:

Spineless and overlooked: DNA metabarcoding of Autonomous Reef Monitoring Structures reveals intra- and interspecific genetic diversity in Mediterranean invertebrates

Anna Thomasdotter, Peter Shum, Francesco Mugnai, Marina Vingiani, Marco Abbiati, Federica Costantini

Table of Contents:

MOLECULAR ECOLOGY
RESOURCES

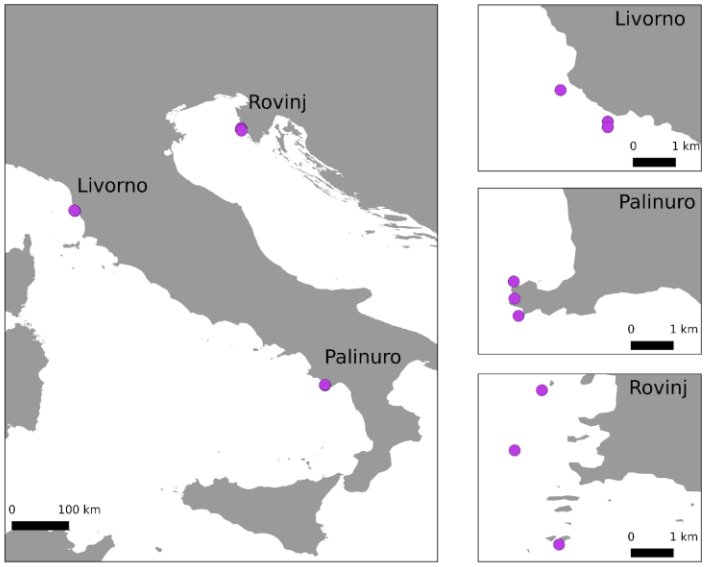


Figure S1: Overview of sampled location and sites within locations.

Table S1: Coordinates and depth for sample sites.

Region	Site	Depth (m)	Longitude	Latitude
Livorno, Italy	Calafuria (CAL)	13	10.34163	43.46233
	Sonnino (SON)	15	10.35358	43.46080
	Isola (ISO)	10	10.32751	43.47173
Palinuro, Italy	Punta Spartivento Sud (SPA)	24	15.27040	40.02233
	Cala del Ribatto (RIB)	17	15.26904	40.02709
	Costa Azzurra (COS)	20	15.26857	40.03182
Rovinj, Croatia	Piccola Figarola (FIG)	15	13.61834	45.09265
	Bagnole (BAG)	13	13.60991	45.07404
	San Giovanni (SAN)	18	13.62365	45.04493

MOLECULAR ECOLOGY
RESOURCES

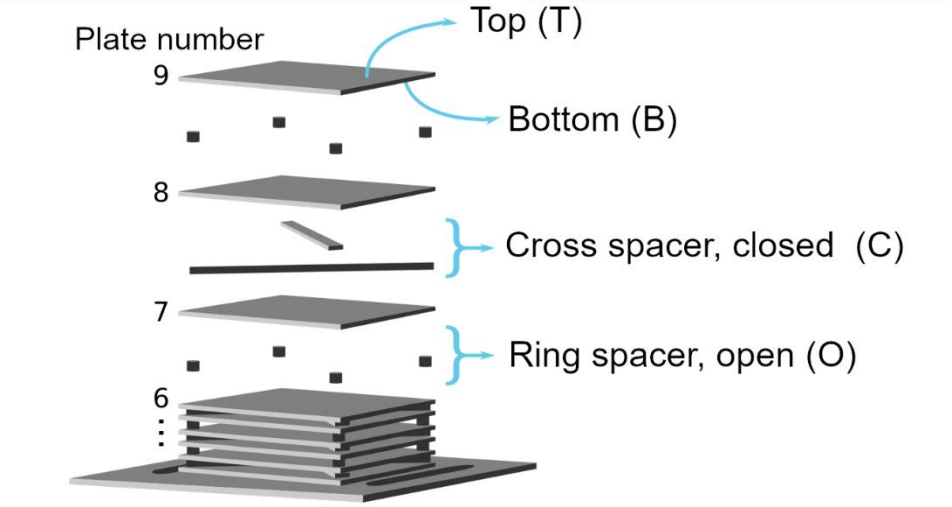


Figure S2: Internal layout of a sampling unit (ARMS). Sample pools were scraped material from top of plate 9 (T9), bottom of closed areas (BC), bottom of open areas (BO), top of open areas (TO), and top of closed areas (TC). As such, five sample pools were collected from within each ARMS. A sixth sample fraction consisted of natural substrate sampled from nearby the ARMS unit.

MOLECULAR ECOLOGY RESOURCES

Table S2: Sequences remaining after each step of the bioinformatic pipeline. Steps are separated by sequences in real sample replicates, and sequences in MOTUs containing terrestrial positive control sequences. The latter were used to assess the level of remaining noise at each step. *Marine metazoan assignments only, excluding positive control MOTUs and with species inferred from MOTUs in cases with no conflicts. See methods for specifics on taxonomic assignment and species exclusion.

			Real sample replicates (n=480)			Terrestrial positive control MOTUs in all positive control replicates (n=42)		
		Method	MOTUs	ESVs	Reads	MOTUs	ESVs	Reads
Demultiplexed reads		-	-	-	12,999,884	-	-	-
Quality + length filtering		*Obitools*	-	-	6,870,233	-	-	-
Singleton removal + denoising + clustering		*Unoise, swarm*	10,594	61,246	3,486,238	13	782	198,434
Decontamination		*decontam*	10,547	60,602	3,347,782	13	767	198,290
Co-occurrence filter		*lulu*	24,145	9,342	3,347,782	13	106	198,731
Abundance filtering		-	2,076	4,474	2,850,076	13	21	196,392
Removal of stop codons		-	2,072	4,462	2,786,939	13	21	196,392
Taxonomic assignment*	At least class level	*RDP classifier, BOLDigger*	331	629	371,492	-	-	-
	Species level		216	455	319,385	-	-	-

MOLECULAR ECOLOGY
RESOURCES

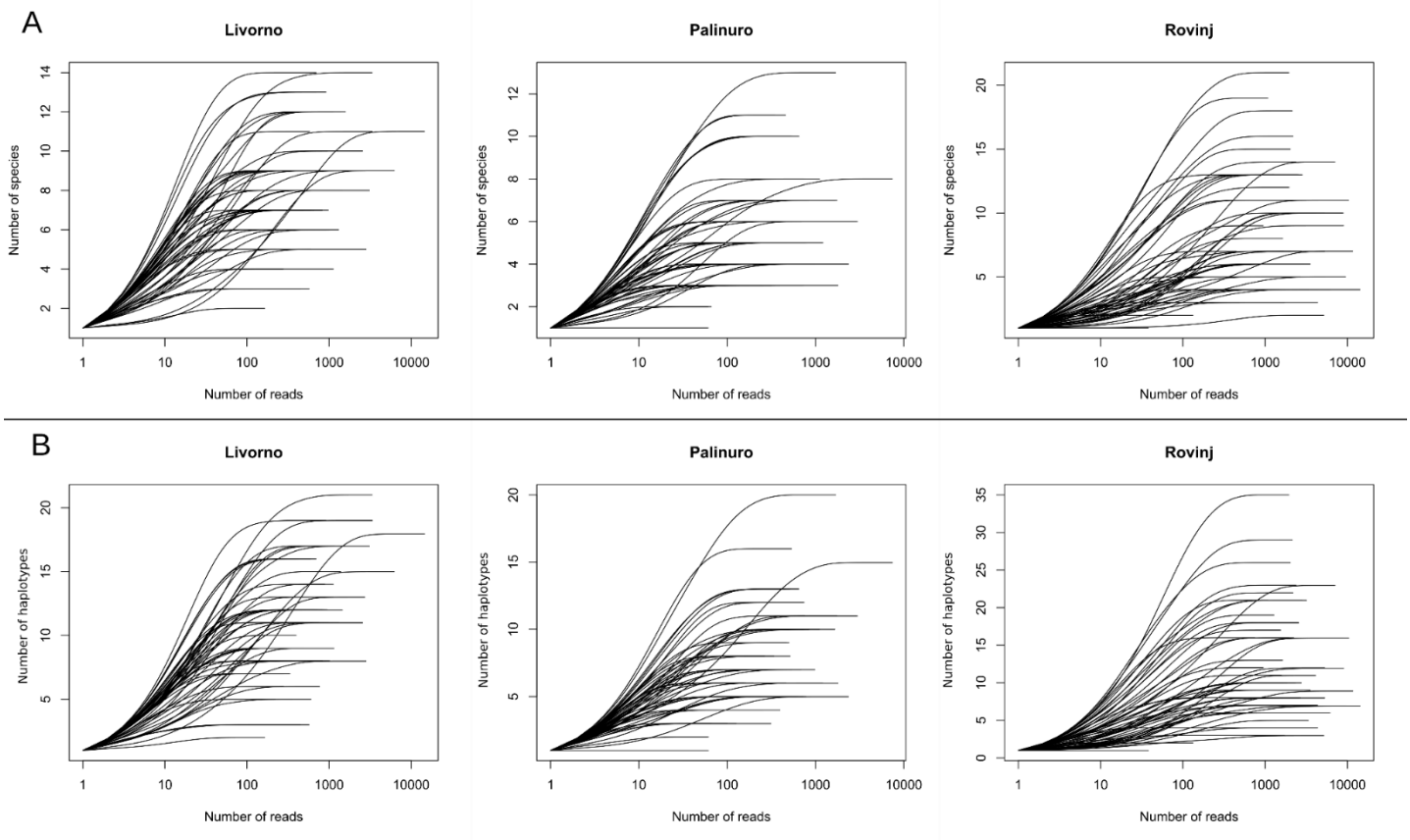


Figure S3: Rarefaction curves illustrating the number of species (panel A) or haplotypes (panel B) detected at different sequencing depths, separated by region. Each line represents one sample, which consists of three PCR replicates. Haplotypes are included only when they have a species-level taxonomic assignment.

MOLECULAR ECOLOGY
RESOURCES

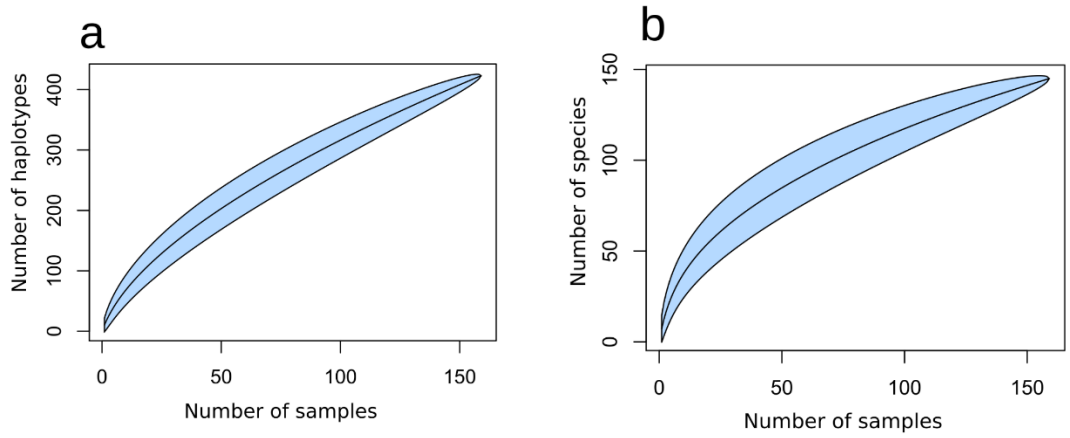


Figure S4: Accumulation curves illustrating the number of haplotypes (a) or species (b) detected with increasing number of samples across all regions. Haplotypes are included only when they have a taxonomic assignment at species-level. Samples consist of three PCR replicates.

MOLECULAR ECOLOGY RESOURCES

Table S3: Summary of data for the three regions. Means are ± standard deviation. π represents nucleotide diversity.

Region	Number of species	Number of haplotypes	Sum of haplotype occurrence	Mean haplotype diversity per species	Mean π per species	Mean π per species, excl. zeros
Livorno	79	201	600	2.54 ± 2.73	0.0073 ± 0.026	0.018 ± 0.038
Palinuro	60	139	421	2.32 ± 2.19	0.0049 ± 0.014	0.012 ± 0.020
Rovinj	89	208	656	2.33 ± 2.48	0.0043 ± 0.0088	0.0095 ± 0.011

Table S4: Analysis of Molecular Variance (AMOVA) outputs for eight species present with sufficient haplotype occurrences across the three locations.

	Variance (σ)			**P-value**		
Species	Between locations	Between samples within location	Within samples	Between location	Between samples within location	Within sample
Halisarca dujardini	8.37	-5.30	96.93	**0.001**	0.957	0.395
Strongylacidon bermudae	8.13	1.48	90.40	**0.034**	0.231	**0.002**
Amphinema dinema	1.73	0.44	97.83	0.36	0.224	0.204
Campanularia hincksii	11.89	16.49	71.62	0.072	**0.015**	**0.001**
Hesiospina aurantiaca	7.76	-5.49	97.73	0.058	0.878	0.423
Chrysopetalum debile	6.29	-6.66	100.37	0.078	0.85	0.604
Oxydromus pallidus	1.58	-2.05	100.47	0.122	0.655	0.597
Leodice cf. limosa	6.96	-2.76	95.80	0.016	0.785	0.185

MOLECULAR ECOLOGY
RESOURCES

Additional comparisons with existing data

Eualus spp.

For *Eualus* spp., a shrimp genus, 22 additional sequences were available from previous work within the SeaMoBB project (Conforti and Costantini, 2022). The additional sequences came from specimens collected from the same ARMS used in the present study but sampled in the vagile 2 mm fraction. These specimens included (potentially) three *Eualus* species: *E. cranchii, E. occultus,* and *E. pusiolus*. When combined with the seven *E. cranchii* sequences from this study and trimmed to an overlapping 313 bp fragment, the dataset collapsed into 19 distinct haplotypes. Five of the sequences found in the present study were new, and two merged with existing data. Conforti and Costantini (2022) observed four groups, likely containing the three *Eualus* species and a fourth cryptic species. Haplotypes found in the present study clustered in what Conforti and Costantini (2022) called group 4 and group 2.

Platynereis dumerilii and *P. massiliensis*

For *Platynereis* spp., a genus of annelid worms, 14 sequences from this study were combined with 123 sequences from 11 Mediterranean locations (Fig. S2) (Wäge et al., 2017, including data from Calosi et al., 2013). The combined data collapsed into 34 unique haplotypes with a length of 272 bp after removing three sequences containing likely mislabeled samples or sequences with ambiguous bases in the additional data. In the present study, seven new haplotypes were found (of which one was assigned to *P. massiliensis*) and three corresponding to haplotypes from Wäge et al. (2017). All *P. dumerilii* haplotypes from this study clustered in what Wäge et al. (2017) called clade 4 (Fig. S2). The single haplotype assigned to *P. massiliensis* from the current study clustered near clade 1 which contains other *P. massiliensis* sequences from Wäge et al. (2017).

MOLECULAR ECOLOGY
RESOURCES

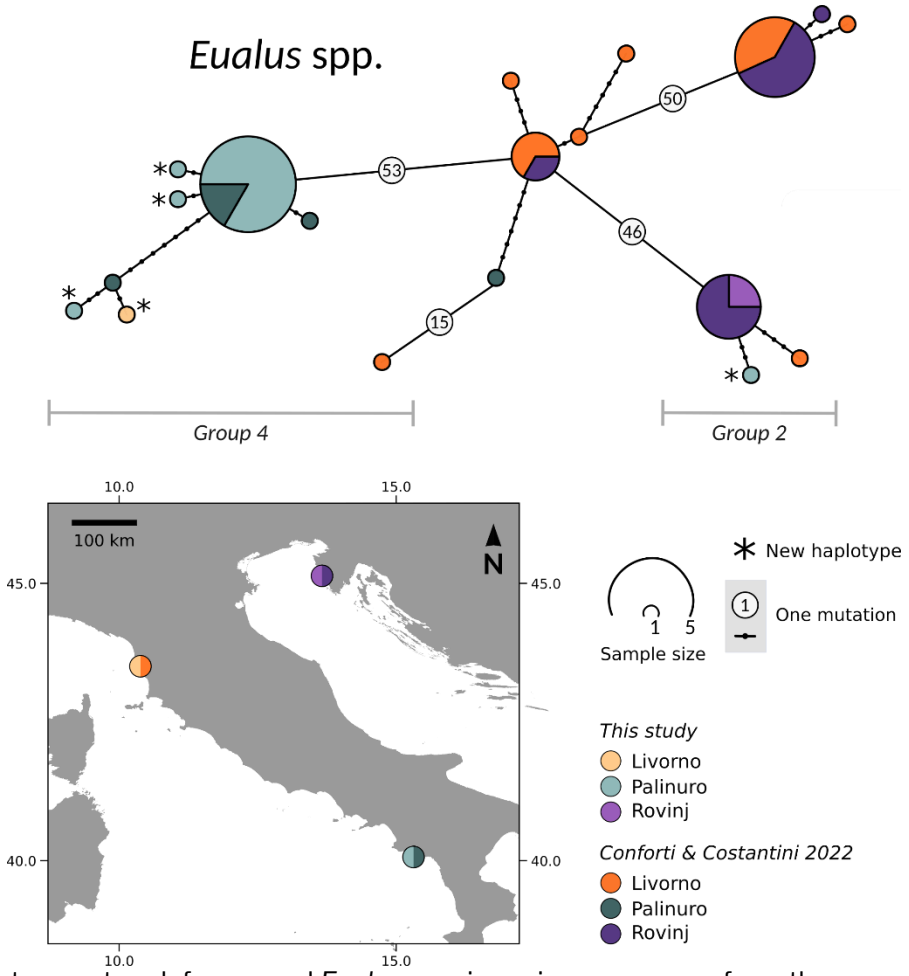


Figure S5: Haplotype network for several *Eualus* species using sequences from the present study and Conforti and Costantini (2022). Haplotypes introduced in this study are marked with an asterisk (*).

MOLECULAR ECOLOGY
RESOURCES

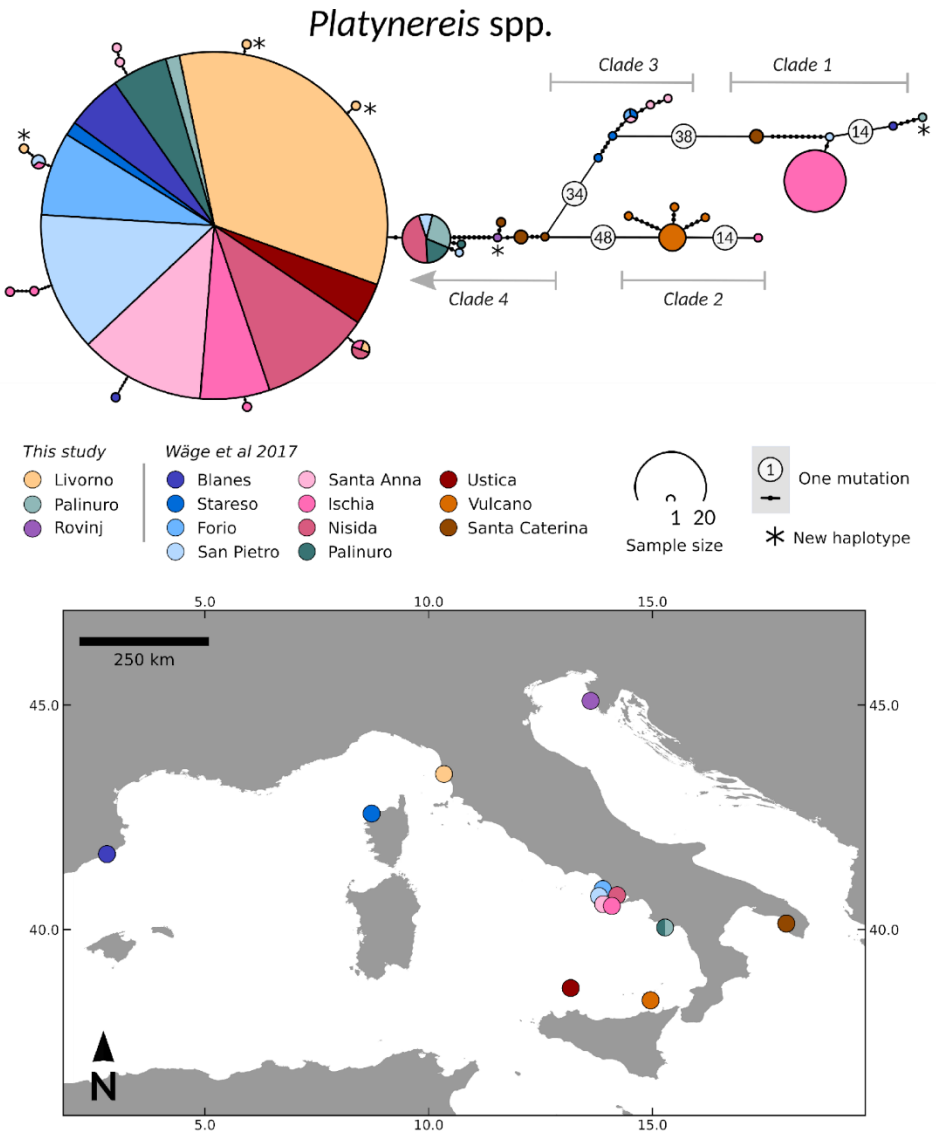


Figure S6: Haplotype network and map of sample sites for *Platynereis dumerilii* and *P. massiliensis,* including sequences from the present study and Mediterranean locations from Wäge et al. (2017) who incorporate data from Calosi et al. (2013). Haplotypes introduced in this study are marked with an asterisk (*).

MOLECULAR ECOLOGY
RESOURCES

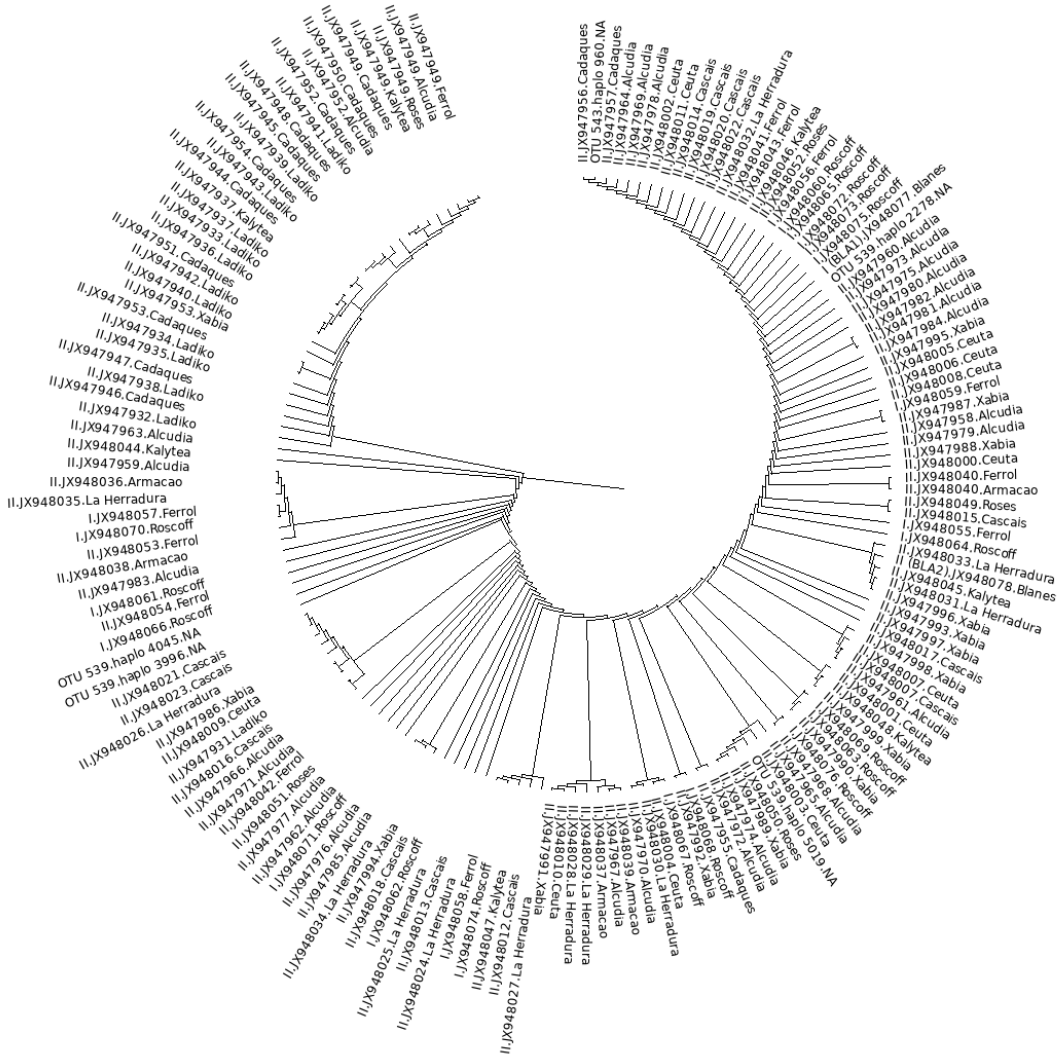


Figure S7: Combined UPGMA phylogenetic tree for *Ophiothrix fragilis* sequences from the present study and Pérez-Portela et al. (2013). Sequences from the present study are labeled with MOTU ID – haplotype ID. Sequences from Pérez-Portela et al. (2013) are labeled with lineage ID – accession number– sample location.

MOLECULAR ECOLOGY
RESOURCES

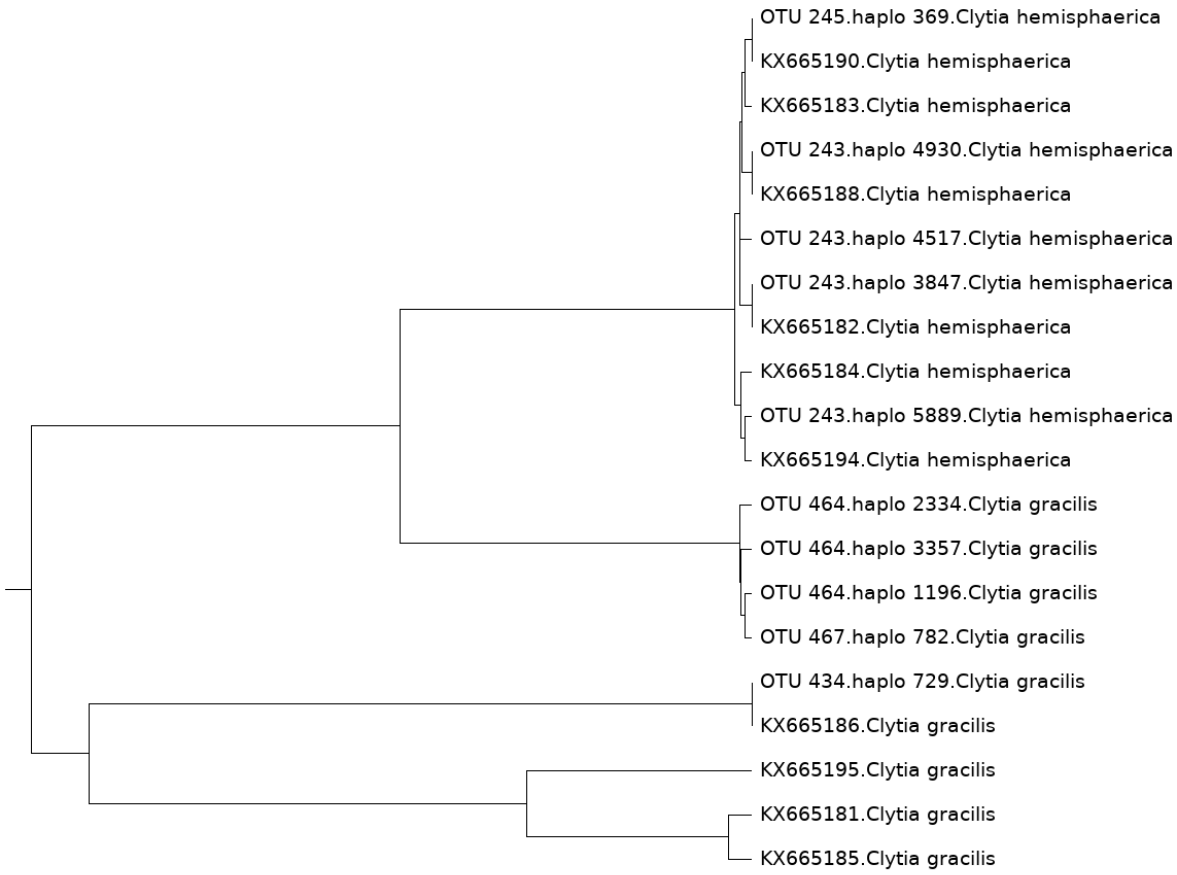


Figure S8: Combined UPGMA phylogenetic tree for *Clytia gracilis* and *C. hemisphaerica* sequences from the present study and Cunha et al. (2017). Sequences from the present study are labeled with MOTU ID – haplotype ID - species. Sequences from Cunha et al. (2017) are labeled with accession number– species.

MOLECULAR ECOLOGY
RESOURCES

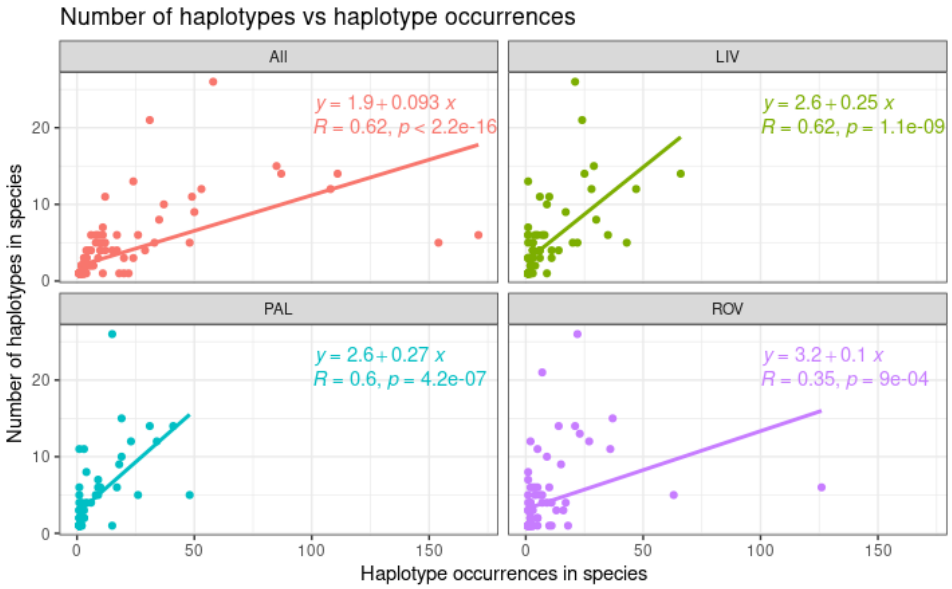


Figure S9: Plots illustrating correlation between haplotype diversity (number of haplotypes observed) and number of haplotype occurrences per species, separated by location. The correlation was significant ($p<0.05$) for all locations.

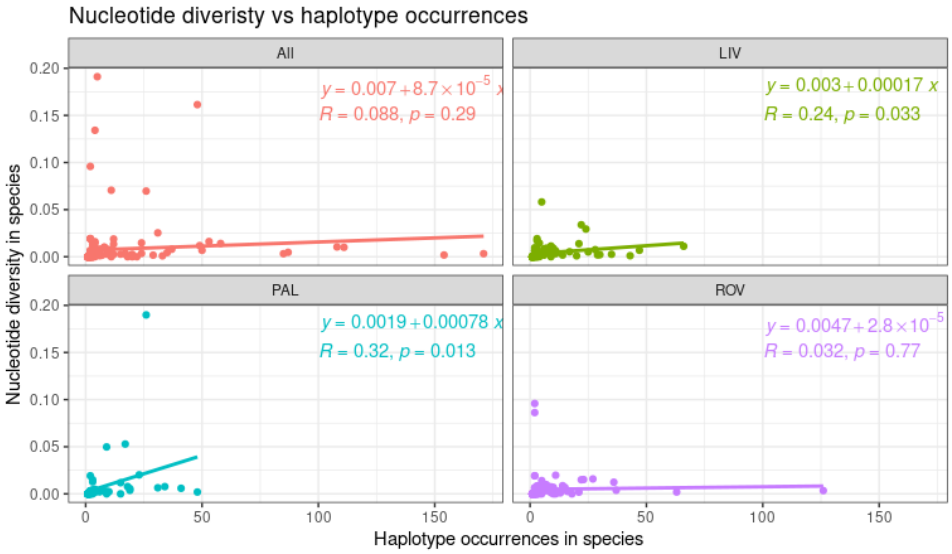
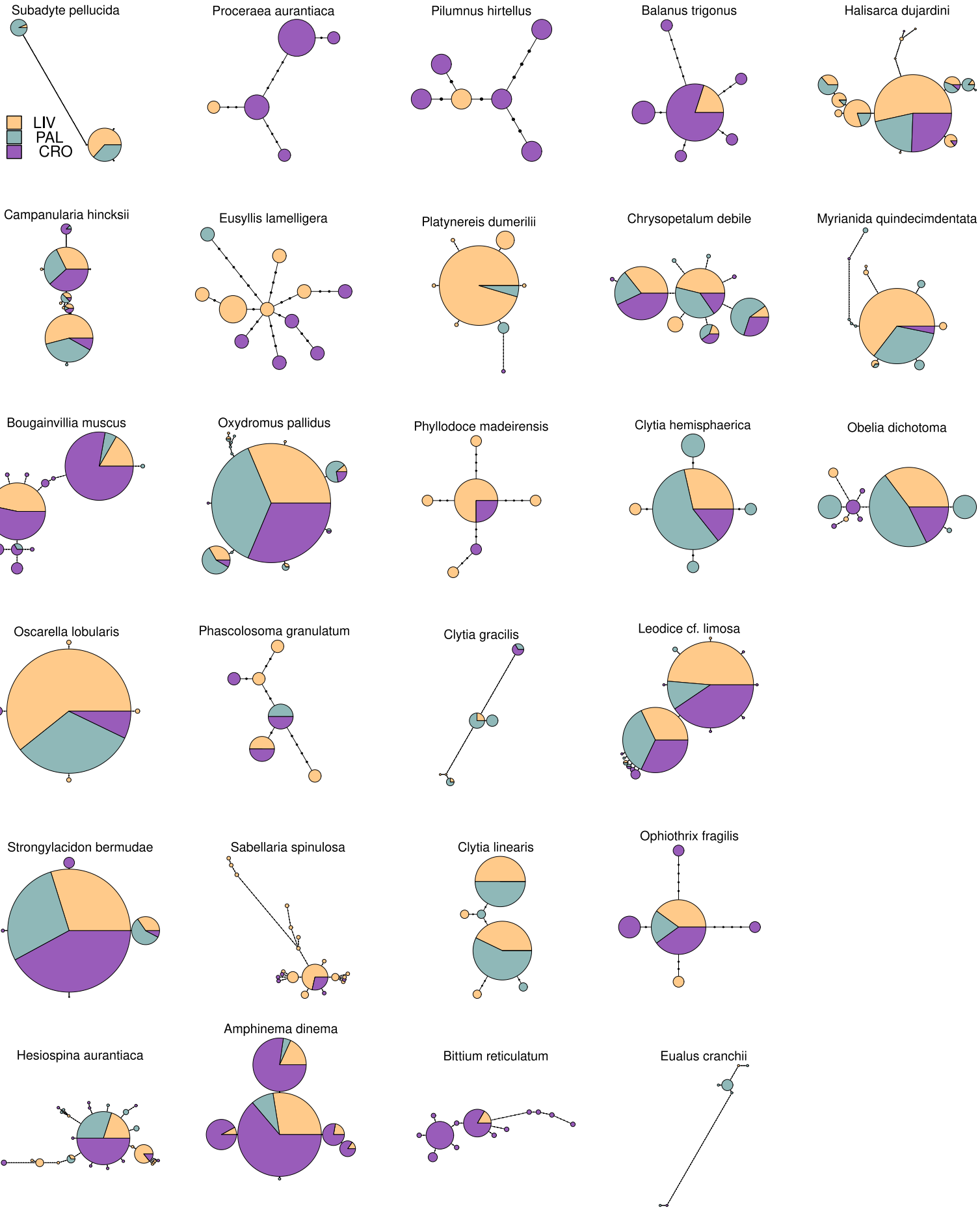


Figure S10: Plots illustrating correlation between nucleotide diversity and number of haplotype occurrences per species, separated by location. The correlation was not significant overall, although Livorno and Palinuro have p-values < 0.05.


Subadyte pellucida
LIV
PAL
CRO
Proceraea aurantiaca
Pilumnus hirtellus
Balanus trigonus
Halisarca dujardini
Campanularia hincksii
Eusyllis lamelligera
Platynereis dumerilii
Chrysopetalum debile
Myrianida quindecimdentata
Bougainvillia muscus
Oxydromus pallidus
Phyllodoce madeirensis
Clytia hemisphaerica
Obelia dichotoma
Oscarella lobularis
Phascolosoma granulatum
Clytia gracilis
Leodice cf. limosa
Strongylacidon bermudae
Sabellaria spinulosa
Clytia linearis
Ophiothrix fragilis
Amphinema dinema
Hesiospina aurantiaca
Bittium reticulatum
Eualus cranchii