The diversity and biogeography of planktonic copepods in the tropical and subtropical Pacific

（太平洋熱帯・亜熱帯における浮遊性カイアシ類の多様性と生物地理）

Junya Hirai

平井　惇也

The diversity and biogeography of planktonic copepods in the tropical and subtropical Pacific

（太平洋熱帯・亜熱帯における浮遊性カイアシ類の多様性と生物地理）

A dissertation submitted in partial fulfillment
of the requirements for the degree of doctoral of
philosophy in agriculture

by

Junya Hirai
平井　惇也

The University of Tokyo

2014

Contents

Chapter I. General introduction

Planktonic copepods are one of the most abundant and widely distributed taxa of metazoans in marine ecosystems, playing a significant role in marine food webs and in biogeochemical cycles (Roemmich and McGowan 1995; Mauchline 1998). Copepods diverged from other Arthropoda 388−522 million years ago, forming a highly diverse group in terms of morphology, physiology, life-strategy, and habitat preferences (Regier et al. 2005; Bron et al. 2011). More than 2,500 species are planktonic in approximately 11,500 morphological species of copepods, with the possible presence of many undescribed and cryptic species (Humes 1994; Razouls et al. 2005−2013). Copepods respond sensitively to environmental changes and have been recognized as indicators of natural and anthropogenic stressors (Beaugrand et al. 2002; Hays et al. 2005; Richardson 2008). For example, warm water species of copepods have shifted their distributions northward with temporally increasing water temperature in the Northeast Atlantic, and these biogeographic shifts led to the decrease of recruitment of Atlantic cod (Beaugrand et al. 2002, 2003). Ocean environments and ecosystems are changing with increasing temperature, acidification, overfishing, pollution, introduction of non-indigenous species etc., and it is a significant issue to describe current biogeography and diversity of marine organisms.

A study of diversity and biogeography of copepods is important for understanding marine ecosystems and food-web structures, therefore, a monitoring of global-scale community structure of copepods leads to detection of changes in marine environments and ecosystems. Under the global warming scenario, ocean with high water temperature is expanding, therefore, the biogeographical studies of copepods are especially important in subtropical and tropical areas. In the tropical and subtropical oceans, a high temperature leads to high metabolic rate and short longevity of copepods (Ikeda 1985; Mauchline 1998), and copepod community are mainly dominated by small species such as the genera *Clausocalanus* and *Paracalanus* in the epipelagic layer (Schnack-Schiel et al. 2010). Since the diversity of

oceanic taxa including copepods is strongly correlated to water temperature, large numbers of species could be observed in the tropical and subtropical oceans especially in oligotrophic subtropical gyres (Woodd-Walker et al. 2002; Rombouts et al. 2010; Tittensor et al. 2010). The oceanic copepods work as secondary and tertiary producers in the tropical and subtropical areas, and various species co-exist in the complex food web structure (McGowan and Walker 1985; Piontkovski et al. 2003). In spite of the importance of biogeographic study in the low latitude, the community structures of zooplankton taxa including copepods have been mainly investigated in the high latitude. For example, the large scale survey using continuous plankton recorder has been extensively conducted for more than 50 years in the North Atlantic and recently developed for the North Pacific and the Southern Ocean (Batten et al. 2003; Richardson et al. 2006). On the other hand, community structures of copepods have not been fully investigated in the low latitude. In addition, the large-scale copepod community in tropical and subtropical oceans has primarily been investigated in the Atlantic Ocean (Woodd-Walker et al. 2002; Schnack-Schiel et al. 2010), and few studies have been conducted in the tropical and subtropical Pacific.

The Pacific Ocean is the largest ocean basin in the world, and the majority of surface water is occupied by the tropical and subtropical regions. In the tropical and subtropical Pacific, studies of copepod communities are restricted primarily to specific taxa or local areas. The early studies by McGowan (1971, 1974) investigated distribution of major epipelagic planktonic species in the Pacific and showed six distribution patterns: (1) Subarctic and Subantarctic, (2) North and South Transition zones, (3) North and South Central, (4) Equatorial, (5) Eastern Tropical Pacific, and (6) Warm Water Cosmopolitan. These distribution patterns agreed well with water mass distributions in the surface water (Sverdrup et al. 1942). The biogeography of eupahusiids has been also studied by Brinton (1962), showing similar major distribution patterns to those of copepods (McGowan 1971), providing significant insights on biogeography of zooplankton taxa in the Pacific. However, their

studies only focused on a specific species, and there have been no study conducted for whole community of copepods covering wide range of the tropical and subtropical Pacific. As for species diversity, Rombouts et al. (2009, 2010) estimated global patterns of copepod diversity, however, data for the tropical and subtropical areas were obtained in the Atlantic Ocean. Therefore, large-scale community analysis of copepods would provide a new insight on understand diversity and biogeography of copepods in the tropical and subtropical Pacific, leading to a further understanding of marine ecosystem in the Pacific. The large-scale study of copepod community might also be helpful for constructing ocean provinces and monitoring environmental changes in the Pacific.

In addition to inaccessibility to the large tropical and subtropical Pacific, the morphological classification is limiting factor for large-scale study of copepod community structure. Morphological classification of specimens is time-consuming and requires sophisticated expertise especially in tropical and subtropical areas, where high diversity of copepods is observed. On the other hand, a DNA-based molecular technique using traditional Sanger sequencing is an effective technique for species identification without relying on subtle morphological differences. Each species shows a unique DNA sequence in molecular maker for species identification, such as cytochrome c oxidase subunit I (COI) in mitochondrial DNA, and DNA barcoding of short DNA fragments enables identification of species based on similarity to known sequences available in public databases (Hebert et al. 2003; Bucklin et al. 2011). The numbers of DNA sequences have been increasing in public database, and COI sequences have been registered in more than 2,000 zooplankton species through the products of the Census of Marine Zooplankton (Bucklin et al. 2010). In addition, even cryptic species and immature stages can be detected by molecular techniques. Therefore, species identification using molecular technique is suitable for copepods with complex morphological characters, and various studies have used molecular technique for species identification or detecting cryptic species (e.g., Bucklin et al. 1995, 2003, 2011; Kiesling et al. 2002; Goetze

2003, 2010; Nonomura et al. 2008).

In addition to species identification, the molecular techniques are also capable of revealing genetic structures within species using genetic marker with high mutation rate (e.g., mitochondrial DNA). A high genetic diversity is common in planktonic taxa in estuaries and coastal waters where there are significant geological barriers (e.g., Lee 2000; Schroth et al. 2002; Chen and Hare 2008; Ueda et al. 2011). Geological barriers to gene flow are not always clear in the open ocean, and high dispersal has been suggested in planktonic taxa (Norris 2000; Provan et al. 2009; Bortolotto et al. 2011). However, several studies have revealed distinct population subdivisions in oceanic taxa (Goetze 2005; Unal and Bucklin 2010; Blanco-Bercial et al. 2011a; Norton and Goetze 2013; Peijinenburg and Goetze 2013). The recent phylogeography of oceanic zooplankton have reported high genetic divergences in nominal species such as chaetognaths (Peijinenburg et al. 2004, 2006; Miyamoto et al. 2010, 2012), copepods (Bucklin et al. 1996; Goetze 2003, 2005, 2010), euphausiids (Bucklin et al. 2007) and molluscs (Jennings et al. 2010). These divergent genetic structures are observed both in allopatric and sympatric regions, suggesting a complex genetic structure of planktonic species in open oceans. It is considered that a study of genetic structure of planktonic species would provide significant information on how biogeography and diversity of planktonic species have been formed in the open ocean.

The traditional Sanger sequencing is an effective molecular technique, however, it remains time-consuming and cost-intensive to reveal community structure, because of need for individual sorting and analysis of specimens. On the other hand, a newly emerging metagenetic method using deep-sequencing has the potential to become a novel technique in marine biology for studying copepods and plankton communities in general. A specific genetic marker is used to reveal community structure from environmental samples in metagenetic approaches, which have been developed primarily for microbial community analysis (Margulies et al. 2005) and have primarily focused on bacteria and archaea (e.g.,

Poinar et al. 2006; Sogin et al. 2006; Huber et al. 2007). Metagenetic study of eukaryotes has mainly been applied to unicellular organisms (e.g., Medinger et al. 2010; Nolte et al. 2010), although some researchers have examined metazoans including benthic meiofauna (Creer et al. 2010; Fonseca et al. 2010), freshwater macro-invertebrates (Hajibabaei et al. 2011; Carew et al. 2013), and terrestrial arthropods (Yu et al. 2012). Metagenetic analyses are found to be useful for surveying species richness in a multicellular eukaryotic organism (Porazinska et al. 2009; Hajibabaei et al. 2011). Massive amounts of sequence data can be obtained from environmental samples containing various species, and these data are clustered into molecular operational taxonomic units (MOTUs: Floyd et al. 2002) based on sequence similarity. Therefore, the metagenetic approach is independent of morphological classification and may be an effective tool for rapidly and comprehensively revealing community structure of copepods in the tropical and subtropical Pacific.

This study focused on community structure of planktonic copepods using molecular techniques and aimed to reveal diversity and biogeography of copepods in the tropical and subtropical Pacific. A suitable genetic marker was investigated for species identification of copepods in chapter II. In chapter III, a novel metagenetic method was developed using deep-sequencing for revealing community structure of copepods. In chapter IV, the novel metagenetic method was applied to field-collected samples in the tropical and subtropical Pacific. Furthermore, chapter V focused on genetic diversity within a nominal species, and phylogeography of *Pleuromamma abdominalis* was investigated in the tropical and subtropical Pacific. All results were summarized in chapter VI, and mechanisms governing the diversity and biogeography of copepods in the tropical and subtropical Pacific were discussed.

Chapter II. Evaluation of ITS2-28S as a molecular marker for species identification of calanoid copepods

Introduction

The choice of molecular marker is an important issue for species identification of copepods using molecular techniques, requiring a high success rate of DNA amplification during polymerase chain reaction (PCR) and a sufficient genetic variability between species. The cytochrome c oxidase subunit I (COI) gene in mitochondrial DNA is a common genetic marker for molecular analysis of metazoans (Hebert et al. 2003; Bucklin et al. 2011), and widely used universal primer pair is available for recovering highly variable 5′ ends of COI fragments (Folmer et al. 1994). Mitochondrial DNA is inherited maternally without recombination, and a low effective population size and high mutation rates are suitable characteristics for molecular analysis (Avise et al. 1987; Birky et al. 1989). However, it is problematic to design universal primer pairs for COI gene for copepods, due to high mutation rates of COI and high genetic diversity of copepods (Minxiao et al. 2011; Cepeda et al. 2012; Machida and Knowlton 2012).

The nuclear genes encoding small and large rRNA subunits (18S and 28S) are relatively conserved and have been used to determine deep phylogenic relationships among families or genera of copepods (Braga et al. 1999; Bucklin et al. 2003; Blanco-Bercial et al. 2011b). 28S is more variable than 18S, suggested to be useful as a diagnostic region for species-level identification (Hassouna et al. 1984; Sonnenberg et al. 2007; Raupach et al. 2010). 28S possess a diverse region at the 5′ end (D1/D2), which has been used for species identification of copepods (e.g., Kiesling et al. 2002; Llinas 2008). In addition to 28S, the internal transcribed spacer 2 (ITS2) is as diverse as 28S and used for species identification of copepods (Goetze 2003, 2005; Machida and Tsuda 2010). ITS2 is located adjacent to the 5′ end of 28S, and it is expected that combined sequences of ITS2 and 28S (ITS2-28S) would

show genetic variability with species-level resolution in copepods.

This chapter aimed to evaluate ITS2-28S as a diagnostic region for species identification of copepods in the subtropical Pacific, using morphologically identified calanoid copepods. The taxonomy of calanoid copepods using morphological characteristics is well-established, and they are suitable for comparison of morphological and genetic analysis. The PCR success rate of ITS2-28S was compared with that of COI. In addition, the level of genetic variability of ITS2-28S was investigated within and between species.

Materials and Methods

Sampling

Calanoid copepods were collected from three stations (St. 3, St. 7, and C3100) in the subtropical area of the western North Pacific (Fig. II-1). Sampling was conducted at St. 3 (24°56.3´ N, 127°59.3´ E) and St. 7 (28°49.3´ N, 129°29.9´ E) during the KT-10-21 cruise aboard the RV "Tansei-Maru" of the Japan Agency for Marine-Earth Science and Technology, on September 28 and 30, 2010, respectively. Sampling at C3100 (31°00.0´ N, 137°59.9´ E) was carried out during the SY-11-05 cruise aboard the RV "Soyo-Maru" of the National Research Institute of Fisheries Science, Fisheries Research Agency, on May 9, 2011. Only St. 7 is located in the Kuroshio Current, while the others are outside the current. Sampling was performed by vertical tows using a vertical multiple plankton sampler (VMPS; Terazaki and Tomatsu 1997) with a 0.25 m^2 mouth-opening area and 100-μm mesh. The water column was divided into sampling layers as follows: 0–200 m, 200–500 m, 500–1000 m, and 1000–2000 m depths at St. 3 and C3100; and 0–100 m, 100–200 m, 200–400 m, and 400–600 m depths at St. 7. Sampling was carried out during the daytime at St. 3 and C3100, while sampling at St.7 was done at night. Samples were preserved in 99% ethanol and kept at 4°C. Ethanol was replaced 24 h after sampling.

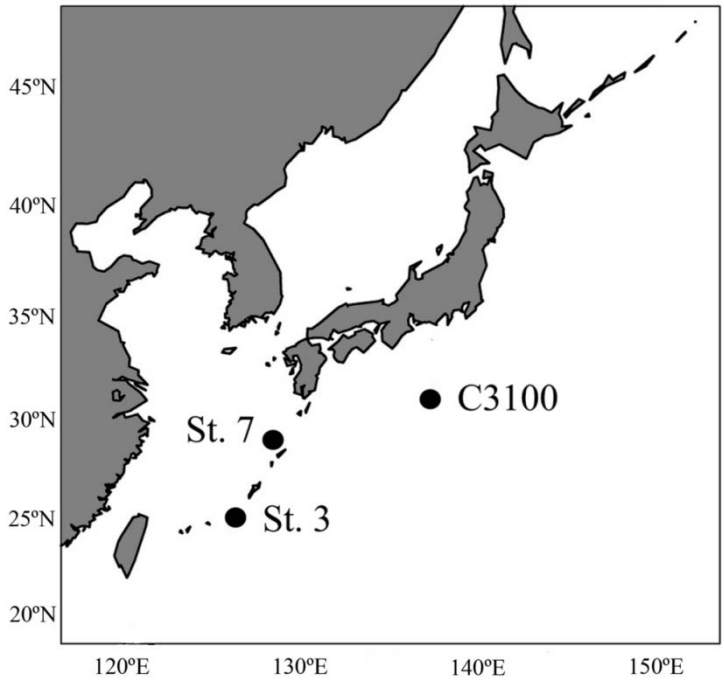


Fig. II-1. Sampling locations .
Black dots show sampling stations. C3100 and st. 3 are located outside the Kuroshio Current and st. 7 on the current.

DNA extraction and morphological classification

In this study, samples taken from St. 3 (0–200 m, 200–500 m, and 500–1000 m), St. 7 (0–100 m), and C3100 (0–200 m and 500–1000 m) were used for further molecular and morphological study. Each sample was split into 1/8 (0–100 m and 0–200 m) or 1/2 (200–500 m and 500–1000 m). Adult female calanoid copepods were selected from each aliquot. Because large numbers of taxa from certain families such as Clausocalanidae and Eucalanidae were present in the samples, not all individuals were used in this study. In total, 244 adult female calanoid copepods from various taxa were used for morphological classification and DNA analyses. Morphological classification was conducted by Dr. Shinji Shimode (Yokohama National University), according to keys by Chihara and Murano (1997). The individuals, which could not be fully identified into species using morphological characters, were classified to genus or family level by the following genetic analysis. For each individual, one of the two first antennae was used for DNA extraction. In total, 20 μL of genomic DNA was extracted using the method of Böttger-Schnack and Machida (2011), which was modified from Schizas et al. (1997), and kept at -20°C.

PCR and sequencing

For PCR amplification of ITS-28S sequence of about 1,000 bp, the primers 5.8S F1 (5´-GACACTTTGAACGCATATTGC-3´) and R635 28S (5´-GGTCCGTGTTTCAAGACGG-3´; Kiesling et al. 2002) were used (Fig. II-2). For species for which clear sequence was not obtained using this primer pair, a different primer pair (18S F1: 5´-CGTCGCTACTACCGATTG-3´ and R635 28S) was used instead. The primers 5.8S F1 and 18S F1 designed for this study were based on GenBank sequences from various species of copepods. For amplification of 710 bp of COI, the metazoan universal primer pair LCO1490 (5´-GGTCAACAAATCATAAAGATATTGG-3´) and HCO2198 (5´-TAAACTTCAGGGTGACCAAAAAATCA-3´) was used (Folmer et al. 1994).

PCR was carried out for 40 cycles in 15 µL reaction mixtures containing 9.95 µL distilled water, 1.50 µL 10× buffer (Takara), 1.20 µL dNTPs (2.5 µM each), 0.60 µL of each primer (5 µM), 0.15 µL Z Taq (Takara), and 1.00 µL template DNA. The PCR cycling was performed using a Model 9700 Thermal Cycler (Applied Biosystems). PCR cycling for ITS2-28S or 18S-28S was conducted as follows: 5 s denaturation at 94°C, 5 s annealing at 60–55°C, and 10 s extension at 72°C. Annealing temperature was decreased stepwise by 0.5°C from 60°C to 55°C for the first 10 cycles and kept at 55°C for the following 30 cycles. An annealing temperature of 45°C was used for COI amplification over 40 cycles. The PCR protocol for COI was same as that for ITS2-28S. DNA from all 244 individuals was used for both ITS2-28S and COI amplification. DNA amplifications were confirmed by electrophoresis on 2% L03 TBE agarose gels (Takara).

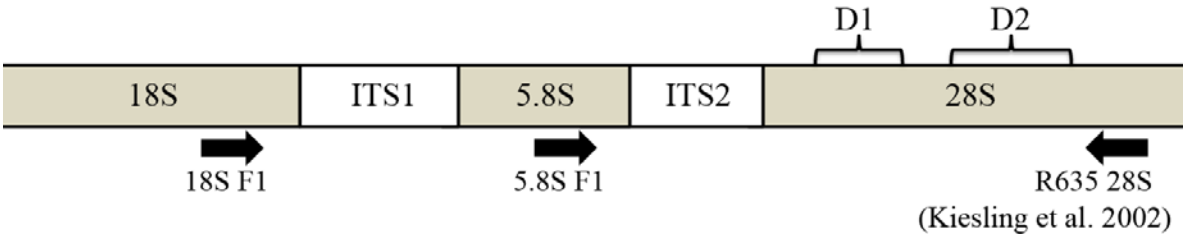


Fig. II-2. Schematic illustration of primer positions (arrows) on nuclear rRNA genes.

"PCR success" was defined as the observation of a single band around 1,000 bp (ITS2-28S) or 710 bp (COI) after staining the PCR products with ethidium bromide. Results of PCR were compared between ITS2-28S and COI amplifications. PCR products of ITS2-28S were purified with ExoSap-IT (GE Health BioScience), sequenced directly using dye-labeled terminators, and analyzed on a 3130 DNA Sequencer (Applied Biosystems). Sequencing reactions were performed according to the manufacturer's protocol using the primer pair 5.8S F1 and R635 28S.

Genetic distances

Consensus sequences of ITS2-28S were generated for each individual after alignment of forward and reverse sequences using Geneious version 5.3.4 (Biomatters Ltd; Drummond et al. 2010). Both the forward and reverse strands were sequenced to confirm the sequences. The level of genetic variability of ITS2-28S sequences was evaluated within each family. Multiple sequences in each family were aligned using Clustal X (Thompson et al. 1997), and the aligned sequences were then edited in SeaView version 4 (Gouy et al. 2010). Internal variations in ITS and 28S have been observed among rDNA gene repeats (Long and Dawid 1980). ITS and 28S are multi-copy nuclear genes, and clear double-base calls were observed in one individual. In this study, ambiguous bases were defined as equivalent double peaks and double peaks, which indicate intra-species dissimilarity, and sites containing such ambiguous bases were removed from the sequence analysis to detect genetic groups which were reproductively isolated in natural condition. The obtained ITS2-28S sequences were used for further genetic analysis. Mismatches caused by gaps (indels) and ambiguous bases were removed for calculating genetic distances and phylogenetic analyses.

Genetic distances within and between species were calculated for ITS2-28S in each family using Kimura's two-parameter (K2P) model (Kimura 1980), the model generally used to calculate genetic distance. Aligned sequence length ranged from 810 to 968 bp after removing

indels and ambiguous sites. A threshold value of 0.005 for genetic distance was used to identify genetic groups in individuals that could not be identified to species level by morphological analysis. Both morphological species and genetic groups were used in calculating genetic distances and the following phylogenetic analysis.

Phylogenetic analyses

Phylogenetic analyses were performed using maximum likelihood (ML) and Bayesian inference (BI) using GTR+G+I, which was selected as the best-fit nucleotide substitute model based on the Akaike Information Criterion in MrModeltest 2.3 (Nylander 2004). Only *Acartia negligens* was excluded from the phylogenetic analysis, owing to its excessively high variance. Only 28S sequences (587 bp after removal of indels and ambiguous sites) were used for this phylogenetic analysis because of the frequent occurrence of insertions and deletions in ITS2 when comparing sequences among various families. ML analysis was performed using 1,000 bootstrap replications for nodal support in MEGA version 5.0 (Tamura et al. 2011). The Bayesian analyses were performed in MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003). The Markov Chain Monte Carlo chains were run for 10^6 generations, and trees were sampled every 100 generations. The first 25% of the generated trees eliminated as burn-in. Three non-calanoid copepods (*Oithona* sp., *Oncaea* sp., and *Corycaeus* sp.) were used as the outgroup.

Four species (*Nannocalanus minor*, *Lucicutia flavicornis*, *Pleuromamma abdominalis*, and *Spinocalanus spinosus*) showed relatively high genetic variability within species. Phylogenetic analyses (ML and BI) were performed for these species using ITS2-28S sequences, following the procedure described above. The best-fit substitution models were GTR+G+I for Calanidae, HKY+G for Metridiniidae, and GTR+G for Spinocalanidae and Lucicutiidae. *Mecynocera clausi* was used as the outgroup for Calanoida, *Haloptilus chirchiae* for Metrididae and Lucicutiidae, and *Euchaeta indica* for Spinocalanidae. In

addition, each of the four species was further analyzed to detect morphological variation within species. Body sizes were compared among samples collected at the same site to avoid effects of temperature or food availability.

Table II-1. List of species collected in this study.
Numbers of individuals analyzed, PCR success of ITS2-28S, PCR success of COI, obtained ITS2-28S sequences, and observed ambiguous bases are shown for both morphological species and genetic groups.

Taxon	Species	Total	PCR success		ITS2-28S sequence	ambiguous base
			ITS2-28S	COI		
Augaptiloidea						
Augaptilidae	*Haloptilus chierchiae*	1	1	0	1	0
Metridinidae	*Metridia brevicauda*	3	2	0	2	0
	Metridia princeps	1	1	0	1	1
	Metridia venusta	4	4	0	1	0
	Pleuromamma abdominalis	14	14	5	9	0
	Pleuromamma gracilis	10	9	3	3	1
	Pleuromamma xiphias	6	6	4	4	1
Lucicutiidae	*Lucicutia flavicornis*	23	21	7	13	2
	Lucicutia curta	1	0	0	X	x
Centropagoidea						
Acartiidae	*Acartia danae*	1	0	0	X	x
	Acartia neligens	2	2	1	2	5
Candaciidae	*Candacia curta*	3	3	3	3	1
	Paracandacia bispinosa	1	1	1	X	x
	Paracandacia simplex	1	1	1	X	x
	Paracandacia truncate	1	1	0	X	x
Centropagidae	*Centropages calaninus*	1	0	0	X	x
	Centropages gracilis	2	1	0	1	0
Pontellidae	*Labidocera detrancata*	1	1	1	1	0
	Pontellina plumata	2	2	1	1	2
Temoridae	*Temora discaudata*	2	2	0	1	1
	Temora turbinate	2	2	1	2	0
Clausocalanoidea						
Aetideidae	*Aetideus acutus*	3	3	1	2	1
	Euchirella curticauda	2	2	2	2	1
	Euchirella messinensis	2	2	1	2	0
	Euchirella pulchra	1	1	1	1	1
	Gaetanus armiger	1	1	0	1	1
	Gaetanus minor	1	1	0	1	0
	Undeuchaeta major	2	2	1	2	1
	Undeuchaeta plumose	5	5	5	5	1
Clausocalanidae	*Clausocalanus farrani*	1	1	0	1	0
	Clausocalanus lividus	8	8	0	7	3
	Clausocalanus minor	10	10	0	8	0
	Ctenocalanus vanus	19	19	0	17	2
Euchaetidae	*Euchaeta indica*	1	1	0	1	0
	Euchaeta rimana	2	2	0	2	0
	Paraeuchaeta media	4	3	0	3	0

Table II-1. Continued.

Taxon	Species	Total	PCR success		ITS2-28S sequence	ambiguous base
			ITS2-28S	COI		
Phaeniidae	*Phaenna spinifera*	1	1	0	1	0
Scolecitrichidae	*Scaphocalanus echinatus*	1	1	0	1	0
	Scolecithrix danae	4	4	0	3	1
	Scottocalanus securifrons	1	1	0	1	0
	**Scaphocalanus* sp	1	1	0	1	0
	*Scolecitrichidae G1	1	1	1	1	0
	*Scolecitrichidae G2	1	1	1	1	0
	*Scolecitrichidae G3	1	1	1	1	0
	*Scolecitrichidae G4	1	1	1	1	0
	*Scolecitrichidae G5	1	1	1	1	0
Eucalanoidea						
Eucalanidae	*Eucalanus californicus*	7	7	6	7	1
	Eucalanus sp	1	1	0	1	1
	Pareucalanus parki	4	4	3	4	0
	Pareucalanus sp	2	2	0	2	0
	Rhincalanus rostrifrons	1	1	0	1	0
	Rhincalanus nasutus	5	5	5	5	1
	Subeucalanus subtenuis	4	4	0	4	0
Megacalanoidea						
Calanidae	*Cosmocalanus darwini*	6	5	5	5	3
	Nannocalanus minor	9	9	6	9	1
	Calanus sinicus	12	12	0	11	6
	Neocalanus gracilis	4	4	4	4	0
	Neocalanus robustior	1	1	0	1	0
	Mesocalanus lighti	1	1	0	1	0
Calocalanidae	**Calocalanus* G1	1	1	0	1	0
	**Calocalanus* G2	1	1	0	1	0
	**Calocalanus* G3	1	1	0	1	0
	**Calocalanus* G4	1	1	0	1	0
Mecynoceridae	*Mecynocera clause*	7	7	4	7	0
Paracalanidae	*Paracalanus denudatus*	2	2	0	2	0
	Paracalanus parvus	3	2	0	2	3
	**Paracalanus* sp	2	2	0	2	0
	*Paracalanidae G1	2	2	0	2	0
	*Paracalanidae G2	1	1	0	1	0
	*Paracalanidae G3	1	1	0	1	0
	**Paracalanidae	1	0	0	X	X
Spinocalanoidea						
Spinocalanidae	*Spinocalanus angusticeps*	1	1	0	1	0
	Spinocalanus longicornis	1	1	0	1	0
	Spinocalanus spinosus	2	2	0	2	0
	*Spinocalanidae G1	1	1	0	1	0
	*Spinocalanidae G2	1	1	0	1	0
	**Spinocalanidae	1	1	0	0	0

*Genetic groups separated using calculated minimum K2P genetic distance (0.005)

**Unidentified individuals, both morphologically and molecularly

x: No clear sequence obtained

Results

PCR and sequencing

Calanoid copepod of 244 individuals, representing 19 families, 35 genera, and 59 species based on morphology, were collected (Table II-1). Twenty individuals could not be identified to species level, mainly owing to the poor condition of the samples. ITS2-28S sequences (about 1,000 bp) were successfully amplified for 232 of the 244 samples (95.1% success rate). The amplification of ITS2-28S was observed in all calanoid families examined in this study. COI sequences (about 700 bp) were successfully amplified for 77 of the 244 samples (31.6% success rate). COI could not be amplified for any of the samples from some families, such as Clausocalanidae, despite the large numbers of samples analyzed.

In total, 194 ITS2-28S sequences were obtained (GenBank accession number: AB753519–AB753715), with a success rate of 79.5%. In *Lucicutia curta*, *Acartia danae*, *Paracandacia bispinosa*, *Paracandacia simplex*, *Paracandacia truncata*, and *Centropages calaninus*, no clear sequences of ITS2-28S were obtained, despite at least three sequencing trials per sample (Table II-1).

Genetic distance between species

In all families, genetic distance was observed between species (Table II-2), and the genetic divergence between species varied in each family. The least genetic distance (0.005) was observed between *Pleuromamma abdominalis* and *P. xiphias*. Therefore, a value of 0.005 was set as the threshold for discriminating between genetic groups in this study. On the basis of this criterion, 20 morphologically unidentified individuals were found to belong to 16 genetic groups. These genetic groups are shown in Table II-1 (e.g., Paracalanidae G1). Individuals of *Eucalanus* sp. and *Pareucalanus* sp. were distinguished based on morphology, using specimens in the same family, and treated as morphologically identified species. ITS2-28S sequences were obtained for 53 morphological species and 16 genetic groups.

Table II-2. Comparisons of genetic distances between/within species in each family based on ITS2-28S sequences.
Length (bp) is the number of base pairs after alignment and removal of sites containing indels and ambiguous bases in each family

Family	Species no.	Samples no.	length (bp)	Genetic distance	
				within species	between species
Acartiidae	1	2	918	0	no data
Aetideidae	8	16	860	0	0.008–0.097
Augaptilidae	1	1	890	no data	no data
Calanidae	6	31	883	0 (0.01–0.021)*1	0.016–0.114
Calocalanidae	4	4	856	no data	0.106–0.139
Candaciidae	1	3	969	0	no data
Centropagidae	1	1	882	no data	no data
Clausocalanidae	4	33	889	0–0.001	0.010–0.061
Eucalanidae	7	24	863	0	0.018–0.179
Euchaetidae	3	6	917	0	0.035–0.062
Lucicutiidae	1	13	841	(0.005–0.014)*2	no data
Mecynoceridae	1	7	903	0	no data
Metridinidae	6	20	874	0 (0.002)*3	0.005–0.049
Paracalanidae	6	10	810	0–0.001	0.005–0.118
Pheannidae	1	1	910	no data	no data
Pontellidae	2	2	903	no data	0.22
Scolecitrichidae	9	11	848	0	0.009–0.094
Spinocalanidae	5	6	835	(0.004)*4	0.009–0.146
Temoridae	2	3	864	0	0.142

*species with large genetic distances within species (1: *Nannocalanus minor*; 2: *Lucicutia flavicornis*; 3: *Pleuromamma abdominalis*; 4: *Spinocalanus spinosus*). Genetic distances within these four species are shown in parentheses.

According to the phylogenetic analysis of 28S, most species and genetic groups were grouped into clades with high bootstrap and posterior probability values that followed the morphological classification (Fig. II-3). The families Aetideidae, Calanidae, and Pontellidae were not supported with sufficient bootstrap and posterior probability values. Each family was also grouped into six superfamilies with high bootstrap and posterior probability values, although the relationship between Spinocalanoidea and Clausocalanoidea was not fully resolved. Families in the Centropagoidea, including Acartiidae, were highly divergent from other taxa. *Acartia negligens* was positioned outside the outgroups *Corycaeus* sp., *Oncaea* sp. (Poecilostomatoida), and *Oithona* sp. (Cyclopoida) and was therefore eliminated from the phylogenetic tree.

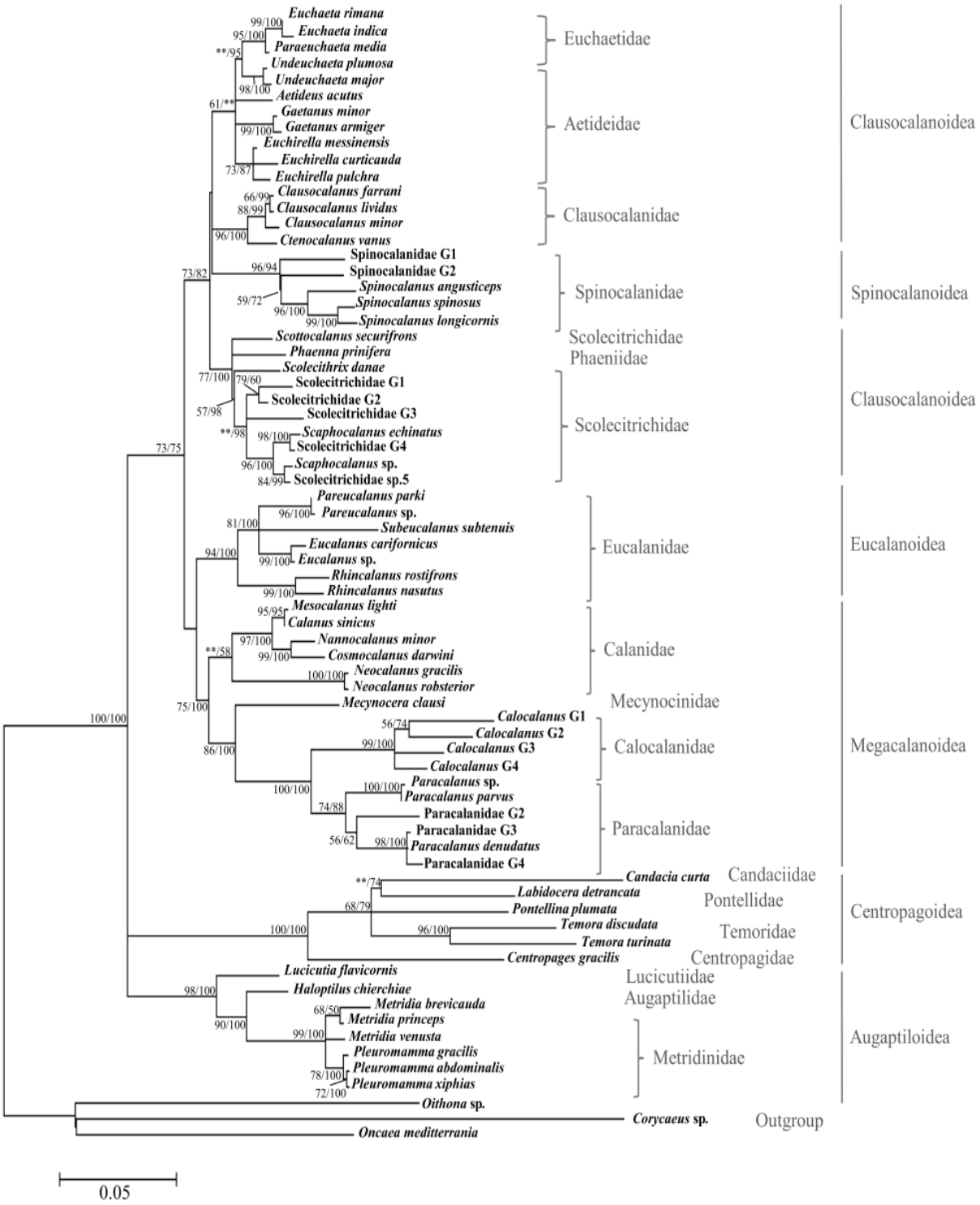


Fig. II-3. ML 50% majority-rule consensus tree based on 28S sequences of 52 species and 16 genetic groups of calanoid copepods (587 bp).

Scale bar indicates genetic distances. Numbers at the nodes indicate the bootstrap values for ML (left) and posterior probabilities for Bayesian analyses (right) greater than 50%. Asterisks indicate values less than 50%. The sequence of *Acartia negligens* was too divergent and was removed from this analysis.

Genetic distance within species

Species in Clausocalanidae and Paracalanidae displayed genetic variability of only 0.001; no variability was observed within most species (Table II-2). However, sample sizes were not sufficient to obtain an exact measurement of the genetic variability within species. The number of ambiguous sites ranged from 0 to 6 (Table II-1). Maximum ambiguous sites (6) were observed in *Calanus sinicus*, followed by *Acartia negligens* (5 ambiguous sites). Most species and genetic groups showed 0–3 ambiguous sites, though this number depended on sample size. Genetic distances between species (≥0.005) were larger than within species, except for *Nannocalanus minor* and *Lucicutia flavicornis* (Fig. II-4 and Table II-2). Greater genetic distances were observed in *N. minor* (0.010–0.021), *L. flavicornis* (0.005–0.014), and *Spinocalanus spinosus* (0.004) than in *Pleuromamma abdominalis* (0.002), which showed morphological differences between sequence types.

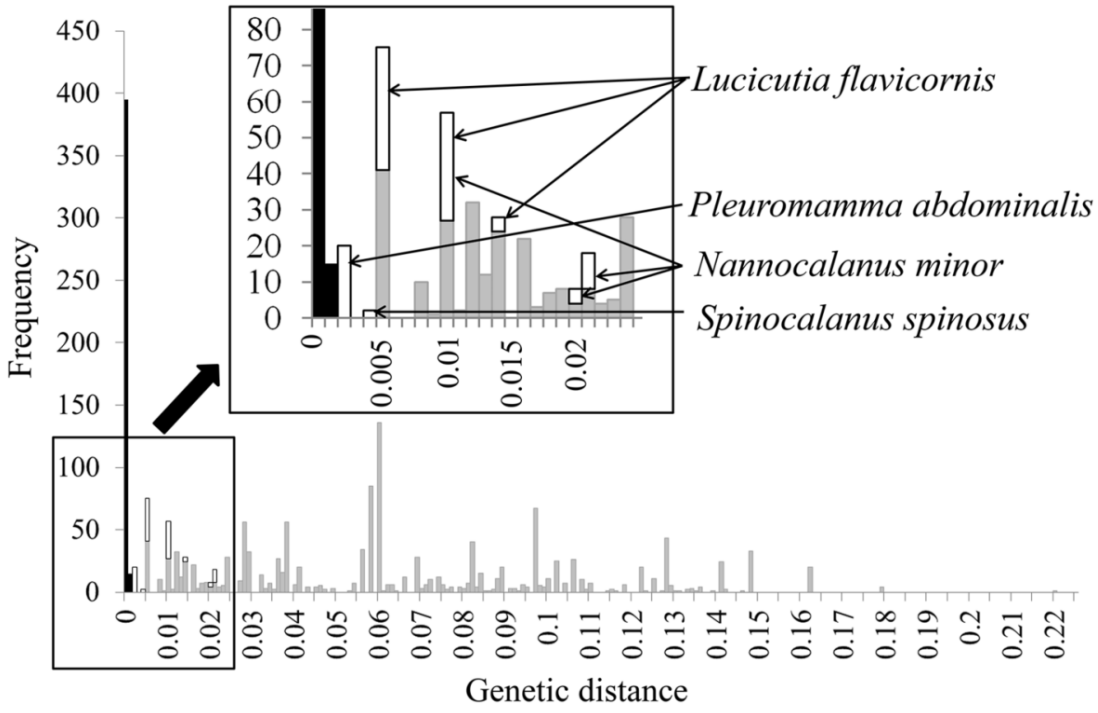


Fig. II-4. Pairwise K2P genetic distances of calanoid copepods within and between species. K2P genetic distances within species are shown by black bars and those between species by gray bars. White bars represent species with high genetic variability within species (*Nannocalanus minor*, *Lucicutia flavicornis*, *Pleuromamma abdominalis*, and *Spinocalanus spinosus*).

Samples of *N. minor* could be divided into three types according to the ITS2-28S phylogenetic tree for the family Calanidae (Fig. II-5a). Three types of *N. minor* and *Cosmocalanus darwini* formed monophyletic clade with high bootstrap values and posterior probability values (82% and 100% each), although their relationships were not fully resolved in this study. There was no genetic variability within each sequence type. Samples of *P. abdominalis* were classified into two types (Fig. II-5b). These two types were within a clade with 77% bootstrap and 99% posterior probability support. The mean prosomal lengths were 1.96 ± 0.5 mm (mean ± S.D.) in *P. abdominalis* 1 (n = 5) and 2.13 ± 1.2 mm in *P. abdominalis* 2 (n = 3).

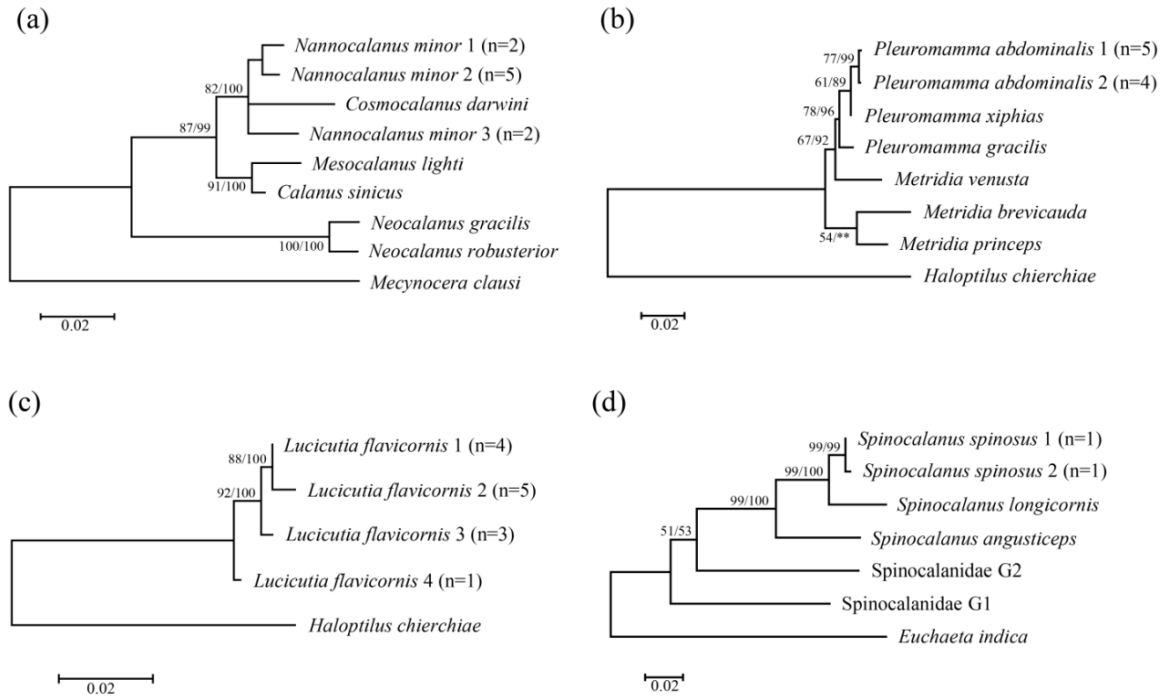


Fig. II-5. ML 50% majority-rule consensus trees of family (a) Calanidae, (b) Metridinidae, (c) Lucicutiidae, and (d) Spinocalanidae.
Numbers at the nodes indicate the bootstrap values for ML (left) and posterior probabilities for Bayesian analyses (right) which are greater than 50%. Asterisks indicate values less than 50%. Scale bars indicate genetic distances. The number of analyzed individuals (n) has been added to each sequence type. No genetic variation was observed within each group. *Mecynocera clausi* is the outgroup for Calanoidae, *Haloptilus chirchiae* for Metrididae and Lucicutiidae, and *Euchaeta indica* for Spinocalanidae.

Pleuromamma abdominalis 2 was significantly larger than *P. abdominalis* 1 at St. 3 ($p < 0.05$; *t*-test). Samples of *L. flavicornis* could also be divided into four types separated by genetic distances of 0.005–0.014 (Fig. II-5c) by ITS2-28S. A genetic distance of 0.004 was calculated between two *S. spinosus* individuals, which formed a monophyletic clade supported by 99% bootstrap and posterior probability values (Fig. II-5d).

Discussion

Suitability of ITS2-28S as a molecular marker

The method using ITS2-28S sequences achieved high rates of PCR and sequencing success. ITS2-28S was found to be sufficiently divergent in species-level and can be used as a molecular maker for DNA barcoding of calanoid copepods in the subtropical western North Pacific. ITS2-28S had a much higher PCR success rate than the COI region. The ITS2-28S primer pair used in this study was designed to amplify the highly conserved regions of 5.8S and 28S rDNA, resulting in a high PCR success rate. COI was shown to be much more divergent than ITS or 28S in the genus *Neocalanus* (Machida and Tsuda 2010). In this study, COI could only be amplified for 77 individuals (31.6% success rate), and the PCR results of amplifications were inconsistent even within single species. It may be difficult to construct universal primers for amplifying diverse mitochondrial genes (including COI), especially in the highly diverse subtropical copepods. Another study of copepods showed that COI cannot to be amplified consistently using a universal primer pair (Cepeda et al. 2012). Blanco-Bercial and Álvarez-Marqués (2007) used primers specific to *Clausocalanus* spp. to amplify COI sequences. The family Clausocalanidae could not be amplified in this study using the common universal COI primer pair by Folmer et al. (1994), suggesting it would be necessary to construct various COI primer pairs for PCR of subtropical copepods with high species diversity.

ITS2-28S sequences were not obtained for six species. According to Lindeque et al. (2006),

the amplification rate is low when samples are not well preserved in ethanol. In this study, DNA was extracted from all samples within one year, and DNA quality should not have been a significant problem. However, both PCR success and failure were observed for the same species, which might indicate that the DNA extraction procedure affected PCR results in this study. Primer mismatch in specific taxa is a more likely cause of PCR failure, because sequencing failures were observed mainly in the species belonging to the superfamily Centropagoidea. According to a 28S phylogenetic analysis of Centropagoidea, the species are highly genetically divergent within the superfamily and different from other superfamilies; this finding has also been reported in studies using multiple-gene analyses (18S, 28S, COI, and Cyt b) of the calanoid copepods (Blanco-Bercial et al. 2011b). Although the primer pair used in this study was intended to design for all calanoid copepods, the high level of divergence in the Centropagoidea may have led to the low sequencing success rate.

A total of 194 ITS2-28S sequences from the 244 individuals were used for sequencing analysis (79.5% success rate). In a similar study, 85% success rate (>1,550 specimens from 1,800 individuals) was achieved by cloning and sequencing the 28S region of copepod nauplii, including non-calanoid copepods in the Arctic Ocean (Llinas 2008). However, the number of species of copepods was much higher in this study than in the report by Llinas (2008), and success rate could be improved by designing specific primers for the Centropagoidea.

28S rRNA gene phylogeny

A 28S phylogenetic analysis of 52 species and 16 genetic groups (excluding *Acartia negligens*) of calanoid copepods from the subtropical ocean showed a high degree of similarity to classification based on morphological characters. A similar result was reported using 19 species of calanoid copepods from Florida Bay using 28S region (Kiesling et al. 2002). Therefore, 28S sequences reflected the phylogenetic relationships of calanoid copepods in this study. Blanco-Bercial et al. (2011b) conducted phylogenetic analyses of

calanoid copepods using multiple genes (18S, 28S, COI, and Cyt b) and obtained higher bootstrap support values than this study, suggesting multiple-gene analysis could be used to gain a deeper understanding of the phylogenetic relationships of the calanoid copepods in this study. Although ITS2 was highly divergent and suitable for differentiating between species, the alignment contained many indels, indicating its unsuitability for phylogenetic analysis of multiple families.

Patterns of genetic variability vary in each taxon, for example, genetic distance between species was lower in Metridinidae than in other families. Calocalanidae and Spinocalanidae showed relatively high genetic variability. There was high genetic variability, in particular, in families in the Centropagoidae, as discussed above. *Acartia negligens* was positioned outside the outgroup in the 28S phylogenetic tree. A similar analysis conducted using 18S rRNA gene sequence data from GenBank showed the same results as 28S rRNA gene, and in this analysis, some *Acartia* species were even positioned outside of *Euphausia pacifica* (Malacostraca). According to Kiesling et al. (2002), three *Acartia* species (*Acartia longiremis*, *Acartia tonsa*, and *Acartia* sp.) were included in the order Calanoida based on 28S sequences, even though high levels of genetic variability were observed and phylogenetic relationships were not resolved. The genus *Acartia* may have a more rapid mutation rate in the 28S region than other copepods (Kiesling et al. 2002), and further investigation is required. The high level of divergence of the Centropagoidae is considered to reflect the complex evolutionary history of this taxon (Adamowicz et al. 2010)

Genetic variability between species

The variability of ITS2-28S showed diagnostic sequences to identify all calanoid copepod species in this study. Calanoid copepod of 157 species were reported in the epipelagic and mesopelagic waters in Sagami Bay (Shimode et al. 2006) and 86 species in the epipelagic water in the Taiwan Strait (Lee et al. 2009), where sampling stations are close to those used in

this study. Although not all calanoid copepods were included in this study, the results show that ITS2-28S could be used for species identification of calanoid copepods in various genera or families. This finding suggests that the ITS2-28S region should be variable in other calanoid copepods. In the subarctic ocean, three closely related species of *Neocalanus* (*Neocalanus cristatus*, *N.flemingeri*, and *N.plumchrus*) were shown to be identified by ITS or 28S sequences by both Llinas (2008) and Machida and Tsuda (2010).

The 5´ end of the 28S region (D1/D2) is both divergent and suitable for species identification in various taxa (Sonnenberg et al. 2007); however, ITS2 sequences were more diverse than 28S sequences in most calanoid copepods in this study. This result corresponds with results of Machida and Tsuda (2010), which revealed higher genetic variability in ITS2 than in 28S in the genus *Neocalanus*. Goetze (2003) mentioned the utility of ITS2 as an informative site for species identification in copepods. ITS regions have recently been used for species identification or phylogenetic analysis of copepods in specific families or genera (Nonomura et al. 2008; Marszalek et al. 2009; Wyngaard et al. 2010). However, no clear phylogenetic relationship was reflected in the ITS2 phylogenetic tree when the phylogeny included various families, because of frequent occurrences of insertions and deletions in ITS2. 28S sequences can be used to infer the phylogenetic relationships in copepods, as shown in this and other studies (Kiesling et al. 2002; Song et al. 2008), and are suggested as a suitable molecular marker for the analysis of copepod covering various taxa.

Genetic variation within species

In this study, no intraspecific genetic variability was detected in most species, but other studies have shown small genetic variability in ITS2 or 28S within species (Llinas 2008; Machida and Tsuda 2010). This discrepancy in results may be due to the method of eliminating ambiguous bases derived from variation among rDNA repeats. According to Sonnenberg et al. (2007), intra-individual genetic variation in the 28S D1/D2 regions is less

than 0.1% on average in most metazoan taxa. However, Llinas (2008) reported the level of intraspecific variability in 28S D1/D2 was mostly lower than 0.5% in copepods. The method used in Llinas (2008) is different from the one used in this study; the differing results for intraspecific genetic variability might be due to the dissimilarity within rDNA genes, owing to tandem repeats in the genome. The least genetic distance between species in this study was 0.005 (between *Pleuromamma abdominalis* and *P.xiphias*), which was higher than the within-species genetic variability. In addition, the hundreds of copies of ribosomal genes in a genome typically evolve in concert (Elder and Turner 1995). The numbers of observed ambiguous sites were not large in this study, suggesting low genetic variability within a single species. The genetic distance of 0.005 was used as a threshold to discriminate between genetic groups of copepods which suggested a possibility of biological species. However, sample sizes in this study were insufficient to assess exact intraspecific and interspecific variability, and greater sampling effort is required for further understanding of genetic variability within species. In addition, some nominal species may show levels of genetic variability between 0.001 and 0.005, such as *P. abdominalis* and *Spinocalanus spinosus* in this study. For these specimens, further molecular and morphological study would be suggested.

In this study, no geographically based genetic variation within morphological species was observed between different sampling sites. Distances among sampling sites were not large, and there may be no reproductive isolation. According to Goetze (2003), both mitochondria and nuclear molecular markers showed geographically based genetic variation in the family Eucalanidae among different ocean basins, suggesting geological isolation of pelagic copepods. Since samplings were only restricted in the subtropical western North Pacific off Japan, greater sampling effort covering a wide area is required to study geological difference.

ITS2-28S revealed that the genetic difference (0.002) between samples of *P. abdominalis* was clearly coincident with body size. Two nucleotide differences were observed in 28S between the large and small forms, and there was no difference in ITS2 in this study. A similar

study conducted in *P. xiphias* using COI and ITS2 showed no genetic variability in ITS2; *P. xiphias* was grouped into four clades by COI, and body size was related to the two major clades (Goetze 2011). In *Lucicutia flavicornis*, the results showed four sequence types of ITS2-28S sequence with relatively large genetic distances (0.005–0.014), greater than the minimum genetic distance between morphological species in this study, suggesting that the groups are reproductively isolated. In *Nannocalanus minor*, three sequence types with large genetic distances were observed. In the genus *Nannocalanus*, *N. elegans* was reported from the Southeastern Pacific (Andronov 2001). The specimen of *N. minor* in this study were not clearly classified into *N. elegans*. Two forms of *N. minor*, i.e., *major* and *minor*, were reported on the basis of body size (Toda 1986). However, body size could not be successfully compared among genetic groups because of the lack of a sufficient number of samples. Similarly, Bucklin et al. (1996) reported two genetically distinct types of *N. minor* using 16S rDNA sequences corresponding to large and small forms. In addition, the large form of *N. minor* was separated into two genetic types (Bucklin et al. 1996). ITS2-28S was shown to be variable between species in this study, suggesting *N. minor* may be composed of at least three reproductively isolated populations in the subtropical western North Pacific.

The taxonomy of the family Spinocalanidae is not well established; morphological classification of this family is difficult because the individuals have fragile bodies. Genetic variation within *S. spinosus* was observed only in ITS2, not in 28S, which is a similar pattern to *Neocalanus flemingeri*, which shows genetic variation in ITS between the large and small forms (Machida and Tsuda 2010).

For four morphological species (*P. abdominalis*, *L. flavicornis*, *N. minor*, and *S. spinosus*), further morphological and molecular study using independent mitochondrial molecular markers is needed to reveal a possibility of cryptic species. As for species complexity in *P. abdominalis*, the existence of cryptic species would be discussed through the phylogeographic analysis using mitochondrial marker in chapter V.

Chapter III. A metagenetic approach to revealing community structure of marine planktonic copepods

Introduction

Chapter II showed the effectiveness of ITS2-28S for species identification of copepods with high PCR success rate and species-level genetic variations. The mutation rate of ITS is more rapid than that of 28S, however, insertion and deletion frequently occur in ITS sequences, inhibiting an accurate alignment of massive sequence data (Chapter II; Blanco-Bercial et al. 2011b; Machida and Knowlton 2012). On the other hand, 28S sequences are easily aligned and reflect phylogenetic information (Chapter II). For metagenetic analysis using deep-sequencing, genes encoding ribosomal RNA are common genetic markers due to the existence of both variable and conserved regions. In 28 sequences, the D2 region is a long (about 350 bp in copepods), hyper-variable region positioned between conserved regions that are suitable for designing a universal primer pair. The numbers of registered sequences in GenBank is relatively large for the 28S region of copepods (1,540 in November 2013), although not as high as for the COI gene (4,993 in November 2013). The 28S has been used for metagenetic analysis of Haptophyta (Bittner et al. 2013) and is expected to be a suitable molecular marker for metagenetics of copepod communities. Once a novel metagenetic method is developed for assessing copepod community, rapid and comprehensive analysis of copepod community structure is possible, which can serve as the basis for future large-scale studies that survey the diversity and distribution of planktonic copepods and other organisms.

In this chapter, a metagenetic method based on 28S-D2 sequence was developed to reveal community structure of marine planktonic copepods. Four plankton community samples were analyzed using deep-sequencing: one artificial sample containing 33 morphologically identified copepod species, and three field-collected samples. First, an approach to data analysis was developed using sequence data from the artificial community sample, and an appropriate similarity threshold was determined for molecular operational taxonomic unit

(MOTU) clustering. Next, this new method was used on field-collected samples from the Kuroshio Current region off the southern coast of Japan. Results from metagenetic and morphological analysis of these copepod communities were compared, and the accuracy of the new metagenetic method was evaluated in capturing copepod community structure in field-collected samples.

Materials and Methods

Artificial community sample

An artificial community sample was prepared from one individual of each of 33 morphologically identified copepod species (Table III-1). The 33 species were collected in the subtropical regions off Japan and belonged to 3 orders, 17 families, and 27 genera. One of the two first antennae of each individual was used for traditional Sanger sequencing to obtain reference sequences of the 28S-D2 region including the primer sites. The 33 individuals were then pooled in a single tube and used for the metagenetic analysis by pyrosequencing.

Sanger sequencing

Genomic DNA was extracted from the first antenna of each of the 33 species according to Böttger-Schnack and Machida (2011). The newly designed primer pair 28S D1F (5´-GCGGAGGAAAAGAAAACAAC-3´) and 28S D3R (5´-CGATTAGTCTTTCGCCCCT-3´) was used to amplify about a fragment (approximately 1000 bp) of the 28S nuclear rDNA gene. PCR was carried out for 40 cycles in a 15 µL volume containing 5 µL of distilled water, 7.5 µL of 2× Gflex PCR Buffer (Takara), 0.6 µL of each primer (5 µM), 0.3 µL of Tks Gflex DNA Polymerase (Takara), and 1 µL of template DNA. The PCR cycling was performed using a Model 9700 Thermal Cycler (Applied Biosystems) as follows: 10 s denaturation at 98 °C, 15 s annealing at 55 °C, and 30 s extension at 68 °C.

Table III-1. List of the 33 morphologically identified copepod species included in the artificial community analysis, and their accession numbers.
The sequence length represents the amplified region of 28S-D2 including the primer sites.

Taxon	Species	Length (bp)	Accession no.
Calanoida			
Aetideidae	*Aetideus acutus*	406	AB796399
	Euchirella curticauda	406	AB796400
	Euchirella messinensis	406	AB796401
	Gaetanus minor	406	AB796402
	Undeuchaeta major	406	AB796403
	Undeuchaeta plumosa	406	AB796404
Augaptilidae	*Haloptilus* sp.	408	AB796405
Calanidae	*Calanus sinicus*	406	AB796406
	Cosmocalanus darwini	406	AB796408
	Mesocalanus tenuicornis	406	AB796407
	Nannocalanus minor	406	AB796409
	Neocalanus gracilis	406	AB796410
Candaciidae	*Candacia curta*	429	AB796412
Centropagidae	*Centropages* sp.	407	AB796413
Eucalanidae	*Eucalanus californicus*	405	AB796414
	Pareucalanus sp.	405	AB796415
	Pareucalanus attenuatus	405	AB796416
	Subeucalanus subtenuis	405	AB796417
Euchaetidae	*Paraeuchaeta media*	406	AB796418
Lucicutiidae	*Lucicutia flavicornis*	406	AB796419
Mecynoceridae	*Mecynocera clausi*	406	AB796420
Metridinidae	*Metridia brevicauda*	406	AB796421
	Metridia venusta	406	AB796422
	Pleuromamma abdominalis	406	AB796423
	Pleuromamma gracilis	406	AB796424
Paracalanidae	*Calocalanus* sp.	407	AB796411
	Paracalanus sp.	406	AB796425
Pontellidae	*Pontellina plumata*	408	AB796426
Scolecitrichidae	*Scolecithrix danae*	406	AB796427
Temoridae	*Temora discaudata*	406	AB796428
Cyclopoida			
Oithonidae	*Oithona* sp.	408	AB796429
Poecilostomatoida			
Corycaeidae	*Corycatus* sp.	395	AB796430
Oncaeidae	*Oncaea* sp.	407	AB796431

PCR products were confirmed by electrophoresis on 1.5% TAE agarose gels (Takara), purified with ExoSap-IT (GE Health BioScience), sequenced directly using dye-labeled terminators, and analyzed on a 3130 DNA Sequencer (Applied Biosystems). Sequencing reactions were performed according to the manufacturer's protocol using the same primer pair (28S D1F and 28S D3R).

Field-collected samples

Plankton community samples were collected at three stations (Fig. III-1): Subtropical (31°00.0′ N, 137°59.9′ E), Kuroshio (32°54.7′ N, 138°00.6′ E) and Slope (33°52.0′ N, 137°44.0′ E). Sampling was conducted during the SY-11-05 cruise aboard the RV "Soyo-Maru" (National Research Institute of Fisheries Science, NRIFS, Fisheries Research Agency), May 9–18, 2011. All three stations were located on the O-line along 138°E (Sugisaki et al. 2010). The Subtropical station was located on the oceanic side of the Kuroshio

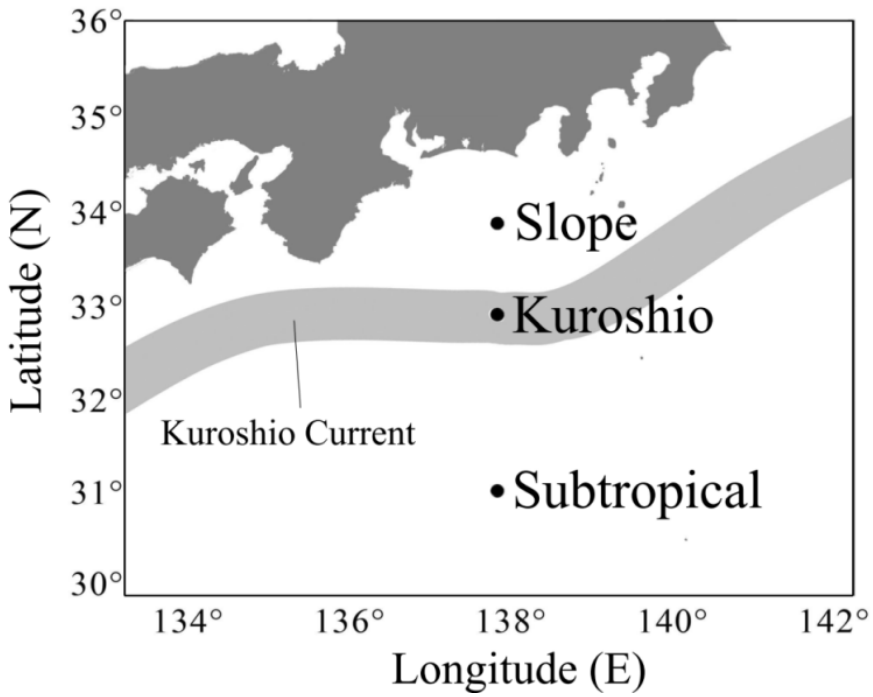


Fig. III-1. Sampling locations for field-collected samples.
Gray line indicates the surface Kuroshio Current during the sampling period, in which The Kuroshio station is located.

Current, the Kuroshio station was within the core of the current, and the Slope station was within slope waters, shoreward of the current. Sampling was performed by vertical tows using a vertical multiple plankton sampler (VMPS; Terazaki and Tomatsu 1997) with a 0.25 m^2 mouth-opening area and 100-μm mesh. All samples were collected from 0–200 m depth during the daytime and were preserved in 99% ethanol and maintained at 4 °C. Ethanol was replaced after 24 h of initial preservation. Temperature and salinity profiles were measured using a conductivity-temperature-depth (CTD) profiler (SBE-911 plus, Sea-Bird Electronics) at all stations (Fig. III-2). All samples were split, and aliquots were used for morphological classification and metagenetic analysis using 454 pyrosequencing.

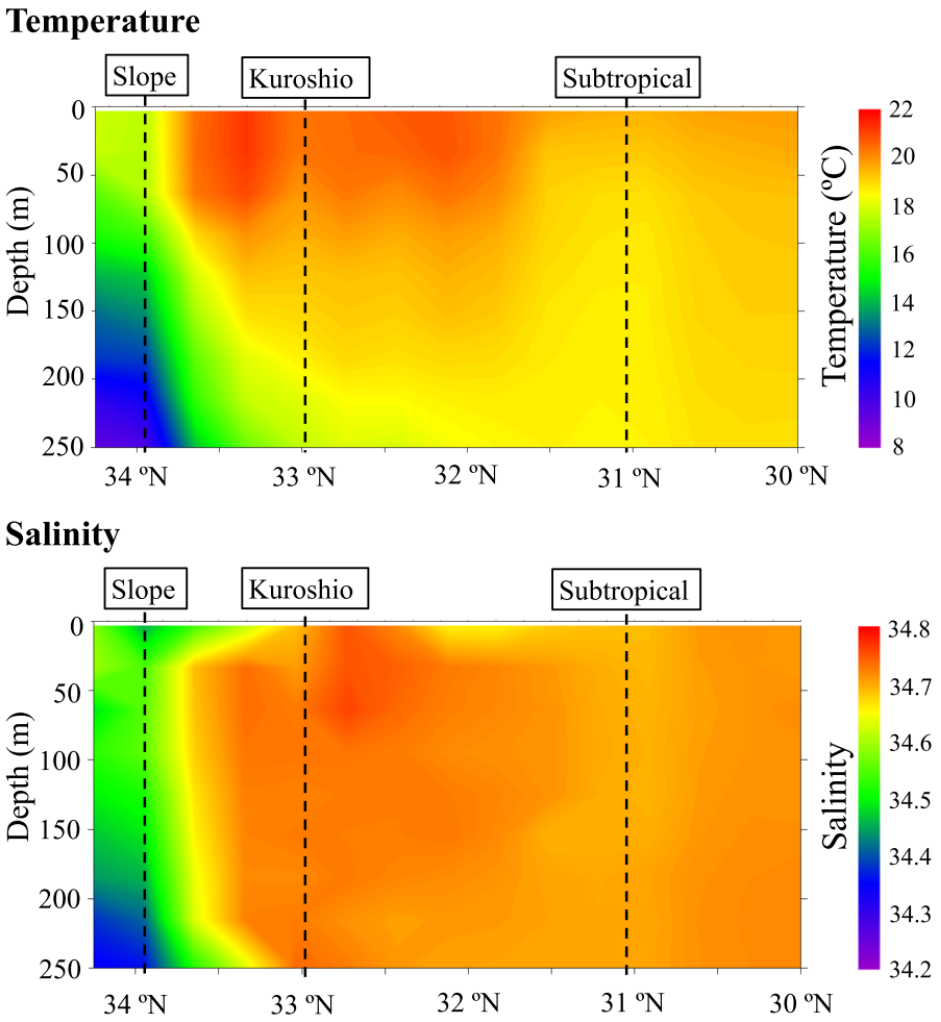


Fig. III-2. Vertical profiles of water temperature and salinity at 138°E (0–250 m depth).

Pyrosequencing

Four ethanol-preserved samples (one from the artificial community, and 1/4 aliquots of three field-collected samples) were filtered onto 100-μm Nitex mesh and transferred to 1.5 mL (artificial community) or 15 mL (field-collected samples) tubes after removing all large non-copepod animals trapped on a 2 mm mesh. Genomic DNA was extracted using the Gentra Puregene Cell and Tissue Kit (QIAGEN) according to the manufacturer's instructions, with modification of the first step to 3-h incubation at 65 °C. The reaction of Proteinase K was conducted overnight at 55 °C. The newly designed primers 28S D2F (5´-AGACCGATAGCAAACAAGTAC-3´) and 28S D2R (5´-GTCCGTGTTTCAAGACGG-3´) were used for PCR amplification of the 28S gene fragment (approximately 400 bp) containing the D2 region. These primers were designed in highly conserved regions and were shown to successfully amplify the 28S-D2 region in >100 copepod species in a preliminary experiment. PCR reactions were performed in a 25 μL volume containing 8 μL of distilled water, 12.5 μL of 2× Gflex PCR Buffer (Takara), 1 μL of each primer (5 μM), 0.5 μL of Tks Gflex DNA Polymerase (Takara), and 2 μL of template DNA (1 ng μL^{-1}). PCR cycling included initial denaturation at 94 °C for 1 min, followed by 22 cycles of 10 s denaturation at 98 °C, 15 s annealing at 58 °C, and 1 min extension at 68 °C. A final extension step was performed at 68 °C for 7 min. Recombinant DNA sequences (chimeras), which are formed during PCR and contain more than one taxon (Meyerhans et al. 1990), are a major cause of overestimation of diversity. PCR was performed with a low number of cycles, long extension time, and low concentrations of template DNA to minimize chimeras (Qiu et al. 2001; Lahr and Katz 2009; Nolte et al. 2010). PCR products of the target region were confirmed by electrophoresis on 1.5% TAE agarose gels (Takara). The concentration of each PCR product was measured with the Qubit® 2.0 Fluorometer (Life Technologies), and 1 μg of each product was purified with the MinElute PCR Purification Kit (QIAGEN). Three pyrosequencing runs were conducted using a 454 GS Junior second-generation sequencer (Roche) for each of the four plankton community samples,

according to the manufacturer's protocol. Each sample was differentiated by unique multiplex identifier (MID) adaptors. Total 184,580 sequence reads were obtained for use in bioinformatic analyses.

Quality filtering

Data analysis of raw reads was performed using MOTHUR (Schloss et al. 2009). The following quality-filtering method was applied to process the sequence data and to minimize overestimation of MOTUs during metagenetic analysis. All reads were filtered based on the following criteria: (1) contained no ambiguous bases (Ns); (2) comprised 300–420 bp not including the primer sites; (3) contained ≤ 5 homopolymers; (4) demonstrated no MID adaptor mismatch, (5) contained no more than three mismatches per primer; (6) achieved average quality score >27; and (7) contained primer sites. The minimum and maximum sequence length, maximum number of homopolymers, and primer mismatch were established based on 28S sequences of copepods in GenBank. The raw data contained both forward and reverse sequences; reverse sequences were reversed and merged with forward sequences in MOTHUR. Forward and reverse primer sites and MID adaptors were removed, and all sequence reads were classified using the naïve Bayesian classifier (Wang et al. 2007) in MOTHUR, with a threshold of >70%. Sequences classified as subclass Copepoda were selected and aligned using MAFFT (Katoh and Toh 2008). The aligned sequences were filtered using single-linkage pre-clustering (Huse et al. 2010). Possible chimeras were removed by UCHIME without a reference dataset (Edgar et al. 2011). The reference database for classification was constructed using copepod sequences containing the 28S-D2 region in GenBank. In addition, >100 28S-D2 sequences were added from various copepod species (GenBank accessions AB796399–AB796431 and AB820728–AB820803) obtained from the analyses of artificial and field-collected samples described below. The taxonomic information in the database was obtained from Boxshall and Halsey (2004). Other non-copepod metazoan taxa were selected from SILVA (Pruesse et al. 2007) and added to the database.

MOTU analysis of the artificial community sample

The quality-filtered sequences were clustered into MOTUs at the 95–99% similarity threshold using the average neighbor algorithm in MOTHUR. Insertions and deletions were removed from distance calculations to minimize overestimation of MOTUs resulting from homopolymers, which is the most common error of 454 pyrosequencing (Margulies et al. 2005; Huse et al. 2007). Only MOTUs containing ≥ 3 sequence reads were included in the dataset used for the following analyses. MOTU numbers were obtained from metagenetic analysis at each similarity threshold (95–99%). The representative sequences within each MOTU were obtained in MOTHUR using the default settings and blasted against the NCBI database to search for the most similar species. The clustered MOTUs were treated as "selected MOTUs" if the representative sequences showed higher similarity to at least one of the 33 reference sequences than a specific threshold (95–99%), and as "non-selected MOTUs" if they did not. MOTU numbers were also estimated from the 33 reference sequences by Sanger sequencing at each similarity threshold, and these results were compared with the numbers of MOTUs by metagenetic analysis. At the determined similarity level, the number of sequence reads for each MOTU was compared with the dry weight (DW) of the identified reference species in the artificial community. The DW (μg) of each species was calculated from prosome length (PL, μm) using the following equation (Uye and Matsuda 1988):

$$\log DW = 2.891(\log PL) - 7.467$$

The sum of the total DW was used in cases in which the MOTU contained several species. Pearson's product-moment correlation coefficients (r) were calculated for proportions of DW and sequence reads using SPSS ver. 21.0 (IBM Corporation).

MOTU analysis of field-collected samples

The quality-filtered sequence data from field-collected samples (Slope: 10,611 reads; Kuroshio: 6,221 reads; Subtropical: 12,500 reads) were clustered into MOTUs using the procedure described above at the determined similarity threshold. The numbers of useful

sequences were low, primarily because there were many short fragments in pyrosequencing data, which were removed through the quality filtering. The minimum number of sequence reads (6,221 reads) was used for MOTU clustering only for comparison of numbers of MOTUs among the sites. All MOTUs were classified into taxonomic order using a naïve Bayesian classifier, and calanoid MOTUs were further classified at the family level. MOTUs below the 70% threshold were treated as "unclassified." The taxonomic composition of MOTUs was examined using two approaches: (1) a non-biomass-based approach in which only the number of MOTUs was used, and (2) a biomass-based approach in which the number of sequence reads within each MOTU was included. For detection of biomass-dominant species, representative sequences of the top six MOTUs (those with the highest numbers of sequence reads) in each sample were obtained in MOTHUR.

Morphological analysis of field-collected samples

Morphological classification was conducted using quantitative sample aliquots for calanoid copepods by Dr. Mikiko Kuriyama (NRIFS). One-quarter or 1/8 aliquots were used for large copepods, and 1/32 or 1/64 aliquots were used for small copepods in the family Clausocalanidae, Paracalanidae, and Spinocalanidae. Adult females and males were sorted and identified to species based on morphological features. The biomass of each species was estimated from PL using the equation reported in Uye and Matsuda (1988). The 28S-D2 sequences of morphologically identified species were added to the reference database for MOTU classification. The total number of species, biomass of each species, numbers of species and total biomass per family, and biomass-dominant species determined from morphological analyses were compared with the corresponding values obtained from metagenetic analyses. Pearson's product-moment correlation coefficients (r) were calculated for correlations between sequence reads of MOTUs and biomass of corresponding species; Kendall's rank correlation coefficients (τ) were calculated for numbers of species and biomass of each family.

Results

Artificial community sample analysis

The length of the 33 reference sequences obtained by Sanger sequencing ranged from 395 to 429 bp (including primer sites), and most of the species contained 28S-D2 fragments between 405–408 bp (Table III-1). There were 0–2 mismatches in the forward primer and no mismatches in the reverse primer. The number of MOTUs determined by Sanger sequencing increased from 22 at 95% similarity to 32 at 99% similarity (Table III-2). The number of MOTUs obtained in the metagenetic analyses also increased, from 28 (95% similarity) to 59 (99%), and rarefaction curves reached the asymptote (Fig. III-3), suggesting sufficient sampling coverage. The 99% similarity threshold had the highest species-level resolution; however, metagenetic analysis of the artificial community sample overestimated the number of MOTUs, mainly because of artificial inflation of selected MOTUs at 99% similarity. The 96% and 95% similarity thresholds led to serious underestimation of the number of MOTUs. The 97% similarity threshold provided the closest match to the true MOTU number in the artificial community, with little inflation of the number. At 97% similarity, 28 selected MOTUs showed high similarity to 28 different reference species (Table III-3). The number of MOTUs estimated by Sanger sequencing was 26 at 97% similarity, and some closely related

Table III-2. The number of MOTUs obtained at each similarity value in the metagenetic analysis of the artificial community sample.
The number of MOTUs (Sanger) was estimated from 33 reference sequences by Sanger sequencing. Total MOTUs in the metagenetic analysis are separated into selected and non-selected MOTUs. Selected MOTUs are defined as those that were above a specified threshold of similarity (95–99%) to the 33 reference sequences. Non-selected MOTUs are copepod MOTUs that were below the specified similarity threshold.

Similarity	MOTU (Sanger)	MOTU	Selected	Non-selected
99%	32	59	44	15
98%	29	43	35	8
97%	26	35	28	7
96%	24	32	25	7
95%	22	28	22	5

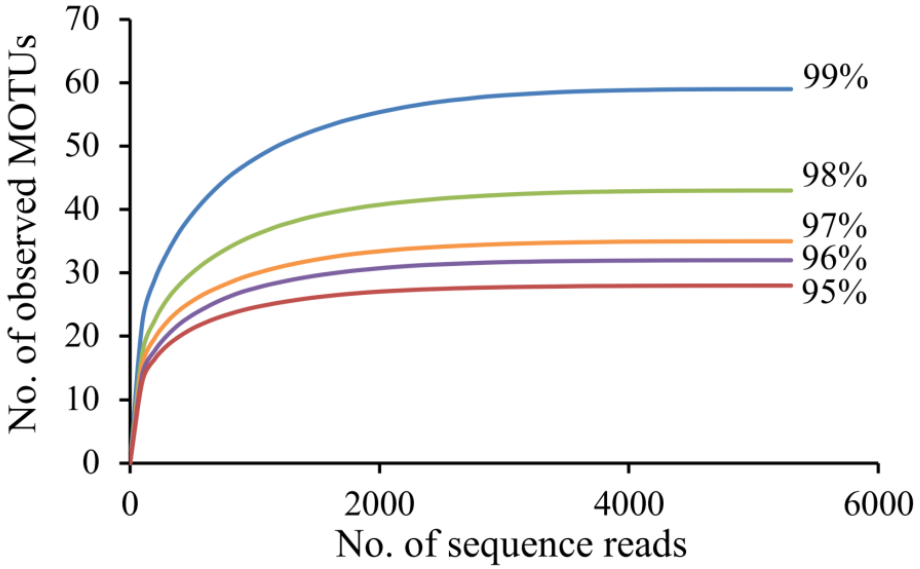


Fig. III-3. Rarefaction curves of the artificial community containing 33 morphologically identified species at 95–99% similarity thresholds.

species were considered to be clustered within a single MOTU (*Pareucalanus* sp. and *P. attenuatus*; *Pleuromamma abdominalis* and *P. gracilis*; *Undeuchaeta major* and *U. plumosa*; *Calanus sinicus, Cosmocalanus darwini, Mesocalanus tenuicornis*, and *Nannocalanus minor*). However, in the metagenetic analyses, *Pareucalanus* sp. and *P. attenuatus* were clustered into two MOTUs and *C. sinicus, C. darwini, M. tenuicornis*, and *N. minor* into three separate MOTUs. Three species (*P. gracilis, U. plumosa*, and *N. minor*) were considered to be clustered in the MOTUs of closely related species. There were no MOTUs with >97% similarity to *Metridia venusta* and *Oncaea* sp. The seven MOTUs were detected as non-selected species at 97% similarity: *Lucicutia gemina* (95% identity), *Clausocalanus furcatus* (99%), *Corycaeus* sp. (89%), *Euchirella curticauda* (93%), *Gaetanus minor* (97%), *Oithona simplex* (97%), and *Oithona helgolandica* (99%) according to a BLAST search (Table III-4). These MOTUs were probably present as gut contents of predatory copepods. The number of sequence reads was correlated with the DW of each species; MOTUs derived from species with high biomass tended to contain large numbers of sequence reads (Pearson's $R = 0.638$, $p < 0.01$; Fig. III-4), suggesting that the number of reads within each MOTU may be a proxy for biomass.

Table III-3. BLAST results of selected MOTUs in the artificial community analysis.
Species with the highest identity value are listed with GenBank accession numbers.

MOTU no.	**Reads**	**Best hit**	**Identity**	**Accession no.**
MOTU 1	1415	*Pareucalanus attenuatus*	99%	AB796416
MOTU 2	1028	*Subeucalanus subtenuis*	100%	AB796417
MOTU 3	811	*Calanus sinicus*	100%	AB796406
MOTU 4	591	*Euchirella messinensis*	99%	AB796401
MOTU 5	523	*Eucalanus californicus*	100%	AB796414
MOTU 6	262	*Undeuchaeta major*	100%	AB796403
MOTU 7	187	*Centropages* sp.	99%	AB796413
MOTU 8	93	*Paraeuchaeta media*	99%	AB796418
MOTU 9	88	*Euchirella curticauda*	99%	AB796400
MOTU 10	87	*Neocalanus gracilis*	99%	AB796410
MOTU 11	61	*Temora discaudata*	99%	AB796428
MOTU 13	53	*Mesocalanus tenuicornis*	99%	AB796407
MOTU 14	52	*Pleuromamma abdominalis*	99%	AB796423
MOTU 15	39	*Pontellina plumata*	99%	AB796426
MOTU 16	31	*Aetideus acutus*	99%	AB796399
MOTU 17	24	*Gaetanus minor*	99%	AB796402
MOTU 18	18	*Candacia curta*	100%	AB796412
MOTU 19	17	*Lucicutia flavicornis*	100%	AB796419
MOTU 20	11	*Corycaeus* sp.	99%	AB796430
MOTU 21	10	*Oithona* sp.	100%	AB796429
MOTU 23	10	*Mecynocera clausi*	100%	AB796420
MOTU 26	7	*Haloptilus* sp.	99%	AB796405
MOTU 27	7	*Paracalanus* sp.	99%	AB796425
MOTU 30	4	*Scolecithrix danae*	99%	AB796427
MOTU 31	4	*Cosmocalanus darwinii*	99%	AB796408
MOTU 33	3	*Calocalanus* sp.	99%	AB796411
MOTU 34	3	*Pareucalanus* sp.	99%	AB796415
MOTU 35	3	*Metridia brevicauda*	100%	AB796421

Table III-4. BLAST results for non-selected MOTUs at 97% threshhold in the artificial community analysis.
Species with the highest identity value are listed with GenBank accession numbers.

MOTU no.	**Reads**	**Best hit**	**Identity**	**Accession no.**
MOTU 12	57	*Lucicutia gemina*	95%	AB820776
MOTU 22	10	*Clausocalanus furcatus*	99%	AB820766
MOTU 24	8	*Gaetanus minor*	97%	AB753572
MOTU 25	8	*Euchirella curticauda*	93%	AB796400
MOTU 28	6	*Oithona helgolandica*	99%	JF419540
MOTU 29	5	*Corycaeus* sp.	91%	AB753714
MOTU 32	4	*Oithona simplex*	97%	AF385458

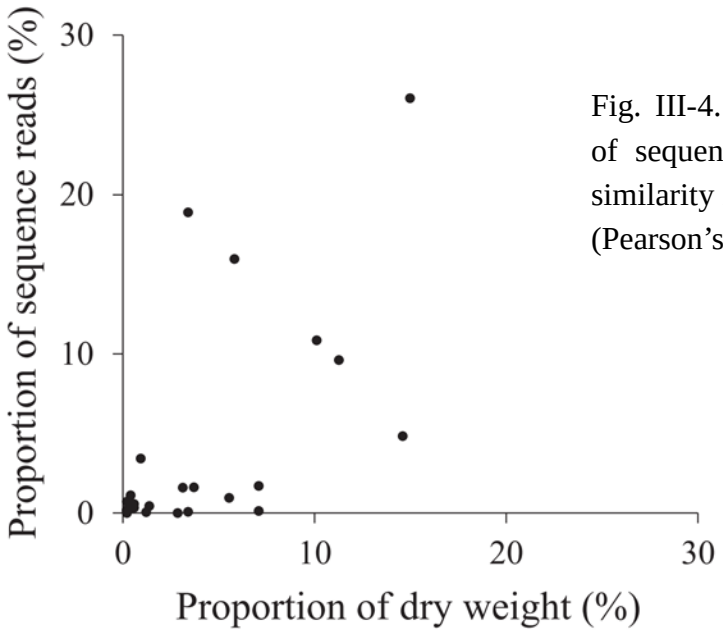


Fig. III-4. Relationship between percentages of sequence reads and dry weight at 97% similarity in the artificial community analysis (Pearson's $r = 0.638$, $p < 0.01$).

Field-collected sample analysis

A total of 110 copepod MOTUs were obtained from all field-collected samples at 97% similarity, and 70 of these were classified into calanoid copepods; 73 calanoid copepod species were identified morphologically. The number of MOTUs was lower in the Slope sample than in the other two samples, and the highest number was observed at the Kuroshio sampling site, for both calanoid and non-calanoid MOTUs (Fig. III-5). This spatial pattern agreed with the morphological analysis of calanoid copepods, supporting the utility of MOTU number as a proxy for species richness. Numbers of calanoid MOTUs exceeded numbers of observed morphological species at all stations.

All MOTUs were classified into four orders: Calanoida, Cyclopoida, Harpacticoida, and Poecilostomatoida (Fig. III-6). The proportion of MOTUs in each order was similar among the three sampling sites, and order Calanoida was dominant (Slope: 59.4% of MOTUs, Kuroshio: 65.6%, Subtropical: 63.3%). The proportion of MOTUs from the other copepod orders was small (Cyclopoida: 11.5–15.6%; Poecilostomatoida: 16.7–20.3%; Harpacticoida: 3.1–3.3%). Calanoida were also dominant in the composition of sequence reads (as a proxy

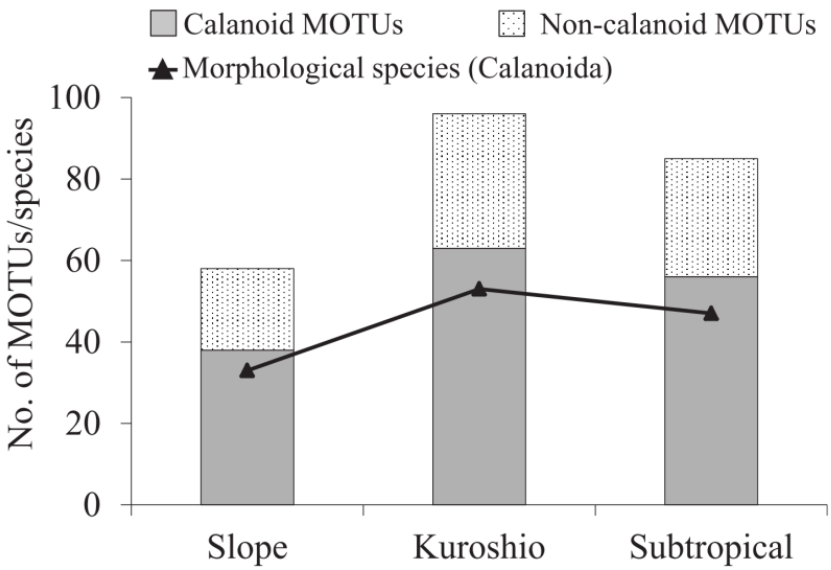


Fig. III-5. Numbers of MOTUs and morphologically identified species in field-collected samples.
MOTU numbers (clustered at 97% similarity) for calanoida and non-calanoida were obtained from the same numbers of sequence reads (6,221 in each sample). Morphological identification was performed for calanoid copepods only.

for biomass) at all stations (Slope: 49.9%; Kuroshio: 85.1%; Subtropical: 62.9%; Fig. III-6). Cyclopoida accounted for a significant fraction of the community, especially at the Slope and Subtropical stations (Slope: 49.3%; Kuroshio: 11.0%; Subtropical: 35.2%). Poecilostomatoida and Harpacticoida comprised minor proportions of sequence reads at all stations.

In comparing MOTUs and morphological species in calanoid copepods, 64 morphological species were considered to be clustered into 47 MOTUs at 97% similarity based on sequence similarities. Nine morphological species were not detected in MOTUs, while 23 MOTUs did not correspond to morphological species. The proportion of sequence reads tended to be high in MOTUs derived from species that comprised a large proportion of DW ($r = 0.620$, $p < 0.01$; Fig. III-7).

Similar patterns of family-level species richness were observed between metagenetic and morphological analyses of calanoid copepods for all three sites (Slope: $\tau = 0.694$, $p < 0.01$; Kuroshio: $\tau = 0.475$, $p < 0.01$; Subtropical: $\tau = 0.499$, $p < 0.01$; Fig. III-8). The metagenetic analysis detected species that were not found in the morphological analyses, including

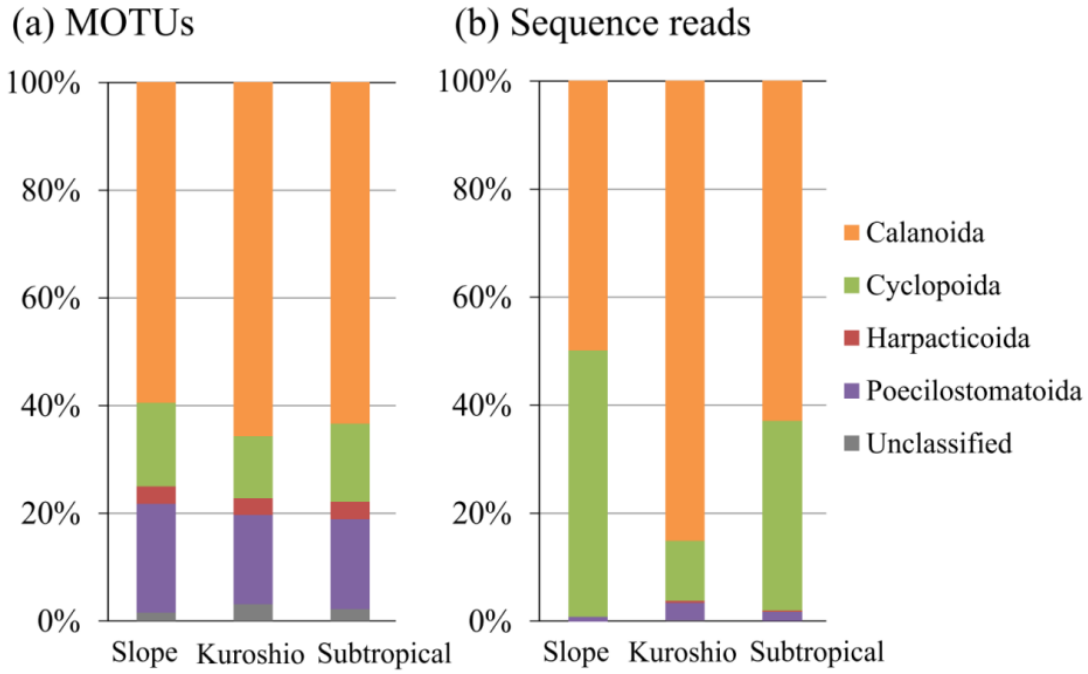


Fig. III-6. Order-level relative composition of (a) MOTUs and (b) sequence reads (all copepods). All MOTUs (clustered at 97% similarity) were assigned to taxonomic order by a naïve Bayesian classifier. The number of sequence reads was not considered in the relative composition of MOTUs.

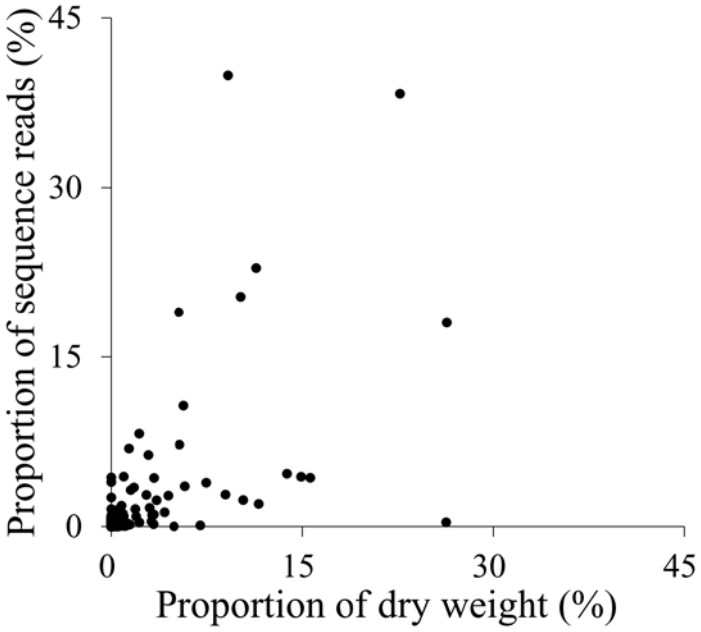


Fig. III-7. Relationship between proportion of sequence reads and dry weight (DW) in MOTUs at 97% similarity in field-collected samples (Pearson's $r = 0.620$, $p < 0.01$).

members of the Centropagidae, Phaennidae, and Pontellidae (Table III-5). On the other hand, species richness of several families (Acartiidae, Calanidae, and Clausocalanidae) was underestimated by the metagenetic analysis. In particular, only two MOTUs were observed in Clausocalanidae, although the morphological classification detected eight species. Of the 70 calanoid MOTUs, four were not assigned to a family by the MOTU classification because of insufficient sequences in the database. These unclassified MOTUs were absent in the aliquots used for morphological analysis.

Families that comprised large proportions of the total sequence reads corresponded to those with high biomass in all field-collected samples (Slope: $\tau = 0.747$, $p < 0.01$; Kuroshio: $\tau = 0.654$, $p < 0.01$; Subtropical: $\tau = 0.531$, $p < 0.01$; Fig. III-8), and included Calanidae, Clausocalanidae, Eucalanidae, and Paracalanidae. The proportions of biomass/sequence reads for each of these families (all samples combined) were as follows: Calanidae, 9.29%/16.18%; Clausocalanidae, 11.72%/5.45%; Eucalanidae, 43.11%/35.4%; and Paracalanidae, 16.81%/30.15% (Table III-5). Rhincalanidae showed relatively high biomass at the Kuroshio station, and the proportion of sequence reads in this family was higher at Kuroshio than at the other two stations. The correlation between number of sequence reads and DW in each family was supported in all data at three stations ($r = 0.836$, $p < 0.01$; Fig. III-9), supporting the usefulness of the number of reads as a proxy for biomass.

A 99% similarity threshold was used to achieve high species-level resolution for detection of dominant species. The dominant MOTUs, with large numbers of sequence reads at 99% similarity, were highly similar to species with abundant biomass in each field-collected sample (Table III-6). The MOTUs with >10% of total reads were observed in the six morphological species with the highest biomass. The identity and composition of dominant MOTUs differed among field-collected samples. MOTU 7 (99% identity to *Calanus sinicus*), MOTU 9 (99% identity to *Eucalanus californicus*), and MOTU 20 (99% identity to *Subeucalanus crassus*) were among the top six dominant MOTUs at the Slope station only, where they were also detected as morphological species with large biomass. MOTU 19 (100%

identity to *Rhincalanus nasutus*) was dominant only at the Kuroshio station and it also attained large biomass at this station. *Mesocalanus tenuicornis* was biomass-dominant at the Subtropical station and the dominant MOTU 14 registered 99% identity to this species. The most dominant MOTU (MOTU 1: 99% identity to *Subeucalanus subtenuis*) was common at both the Kuroshio and Subtropical stations.

Table III-5. Species richness and relative biomass of calanoid copepods estimated by metagenetic and morphological analyses.
MOTUs were obtained from all field-collected samples (97% similarity threshold). Reads are the average percentage of sequence reads in all field-collected samples. Dry weight is the percentage of dry weight in each family relative to the total dry weight for all field samples, estimated from prosome length of morphologically identified specimens.

	Species richness		Biomass	
	MOTUs	Morphology	Reads (%)	Dry weight (%)
Acartiidae	1	3	0.04	0.67
Aetideidae	4	2	0.32	0.49
Augaptilidae	4	2	0.11	1.81
Calanidae	4	6	16.18	9.29
Candaciidae	3	2	0.70	1.78
Centropagidae	2	0	0.06	0.00
Clausocalanidae	2	8	5.45	11.72
Eucalanidae	7	8	35.40	43.11
Euchaetidae	4	4	3.50	2.90
Fosshageniidae	1	1	0.07	0.13
Heterorhabdidae	1	1	0.11	0.48
Lucicutiidae	3	3	2.55	2.12
Mecynocenidae	1	1	0.26	0.64
Metridinidae	2	2	0.57	0.10
Paracalanidae	16	19	30.15	16.81
Phaennidae	1	0	0.02	0.00
Pontellidae	1	0	0.04	0.00
Rhincalanidae	2	2	1.98	6.45
Scolecitrichidae	5	6	0.39	1.09
Spinocalanidae	0	1	0.00	0.01
Temoridae	2	2	1.66	0.39
Unclassified	4	0	0.32	0.00

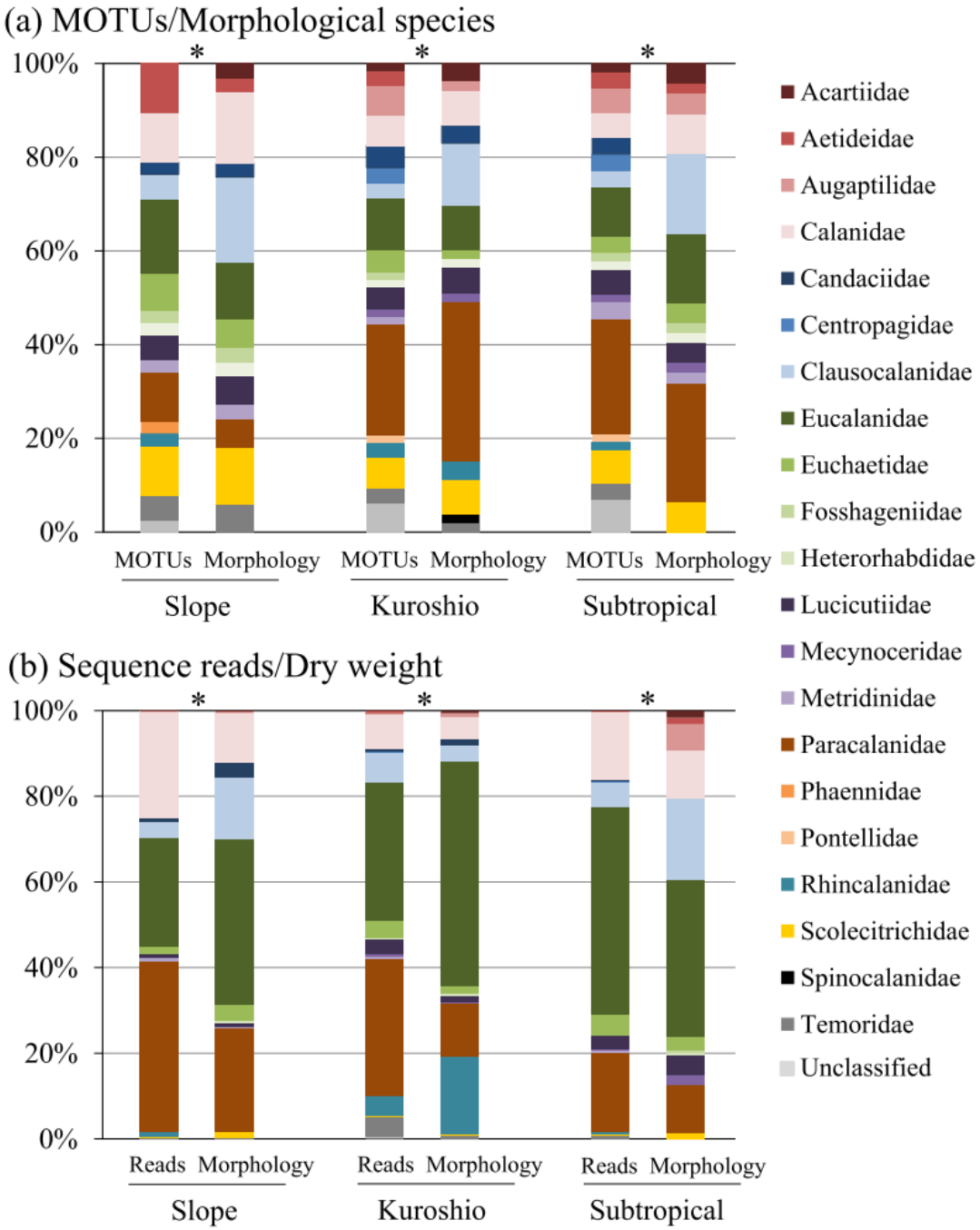


Fig. III-8. Comparison between metagenetic and morphological analyses of calanoid copepods.
(a) Family-level composition of MOTUs and morphological species.
(b) Family-level composition of sequence reads and dry weight.
Calanoid MOTUs (clustered at 97% similarity) were assigned to families by naïve Bayesian classifier. The number of sequence reads was not considered in the relative composition of MOTUs. Dry weight was estimated from prosome length of morphologically identified specimens. Significant rank correlations ($p < 0.01$) between metagenetic and morphological analyses are indicated by asterisks (*).

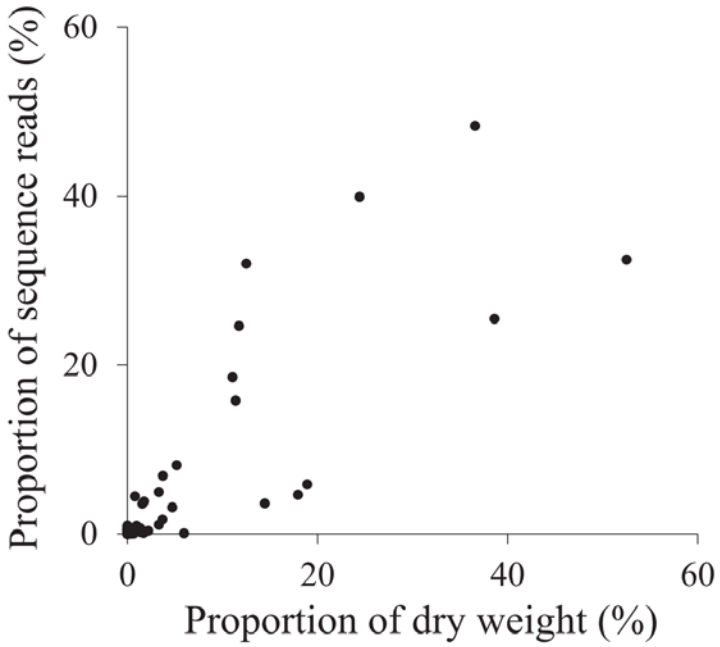


Fig. III-9. Relationship between percentages of sequence reads and dry weight in each family at 97% similarity in the field-collected samples (Pearson's $r = 0.836$, $p < 0.01$).
The data at three stations were combined in this comparison. Unclassified MOTUs were not included.

Discussion

MOTUs identified in the metagenetic analysis using the 28S-D2 gene region closely reflected the species composition of the sampling sites and may provide a proxy for species richness. Studies of Nematoda (Porazinska et al. 2009) and freshwater macro-invertebrates (Hajibabaei et al. 2011) have also suggested the utility of metagenetic analyses for detecting species richness of multicellular organisms. The 97% similarity was considered to be suitable for characterizing the community structure of pelagic copepods using 28S regions. The 28S-D2 is hyper-variable and functions as a diagnostic region for identification of copepod species (Sonnenberg et al. 2007). It is simple to design primer pairs for PCR amplification of 28S sequences (Machida and Knowlton 2012); however, the rate of evolution of nuclear 28S sequences is slow in comparison with mitochondrial DNA, and some closely related species are underestimated in metagenetic analyses because of insufficient sequence resolution (Porazinska et al. 2009). Although some closely related species were clustered into a single

Table III-6. The six MOTUs with the highest number of sequence reads at each sampling site, and morphological species with the highest biomass. Species with the highest similarity are listed for the dominant MOTUs (99% similarity), with identity values and GenBank accession numbers. Reads indicate percentage of sequence reads at each site that were calanoid copepods.

	Metagenetic analysis					Morphological analysis	
	MOTU no.	Reads (%)	BLAST-hit	Identity	Accession no.	Species	Dry weight (%)
Slope							
1	4	38.20	*Paracalanus parvus*	100%	AB820781	*Eucalanus californicus*	26.2
2	7	22.70	*Calanus sinicus*	99%	FJ967118	*Paracalanus parvus* s.l.	22.5
3	9	16.24	*Eucalanis californicus*	99%	AB753637	*Calanus sinicus*	10.7
4	1	3.93	*Subeucalanus subtenuis*	99%	AB753654	*Subeucalanus crassus*	8.9
5	20	2.61	*Subeucalanus crassus*	99%	AB820792	*Clausocalanus parapergens*	4.9
6	11	2.59	*Clausocalanus arcuicornis*	100%	HM997048	*Ctenocalanus vanus*	4.2
Kuroshio							
1	1	20.24	*Subeucalanus subtenuis*	99%	AB753654	*Eucalanus hyalinus*	25.9
2	5	18.78	*Paracalanus aculeatus*	100%	AB820778	*Rhincalanus nasutus*	14.7
3	4	5.53	*Paracalanus parvus*	100%	AB820781	*Subeucalanus subtenuis*	10
4	19	4.41	*Rhincalanus nasutus*	100%	AB753650	*Pareucalanus attenuatus*	6.8
5	12	4.31	*Subeucalanus subcrassus*	99%	AB820794	*Pareucalanus sewelli*	6.8
6	11	4.08	*Clausocalanus arcuicornis*	100%	HM997048	*Paracalanus aculeatus*	5.1
Subtropical							
1	1	39.40	*Subeucalanus subtenuis*	99%	AB753654	*Pareucalanus sewelli*	13.4
2	10	7.99	*Nannocalanus minor*	99%	AB753665	*Ctenocalanus vanus*	11.5
3	5	7.70	*Paracalanus aculeatus*	100%	AB753705	*Subeucalanus subtenuis*	9.1
4	13	7.23	*Paracalanus parvus*	99%	AB829782	*Eucalanus hyalinus*	7
5	14	3.52	*Mesocalanus tenuicornis*	99%	AB796407	*Paracalanus aculeatus*	5.3
6	4	3.49	*Paracalanus parvus*	100%	AB820781	*Mesocalanus tenuicornis*	5.3

MOTU, the 97% similarity threshold provided relatively high species-level resolution with little inflation of numbers of MOTUs. Kunin et al. (2010) also suggested a 97% similarity threshold for MOTU clustering in metagenetic analyses, using Roche 454 pyrosequencing to avoid artificial inflation of diversity estimates; the same threshold was used for metagenetic analysis of Haptophyta using the 28S region (Bittner et al. 2013). Intraspecific variation in the 28S and estimated error rates of 454 pyrosequencingare are small enough not to cause significant inflation of species richness at the 97% threshold (Sonnenberg et al. 2007; Kunin et al. 2010; Gilles et al. 2011). A threshold of 99% similarity is ideal for species identification but was inappropriate for evaluating species richness in this study because of inflation in the numbers of MOTUs, despite stringent sample preparation and quality-filtering procedures. Here, artificially inflated MOTUs generally contained small numbers of sequence reads and may not have had a significant effect on dominant MOTUs at 99% similarity.

All 33 reference species showed amplification of 28S-D2 using a primer pair developed in preliminary experiment, but two species (*Metridia venusta* and *Oncaea* sp.) from the artificial community sample were not detected in the MOTUs. Undetectable species were also noted in a previous metagenetic analysis of Nematoda using 18S and 28S regions (Porazinska et al. 2009). An insufficient quantity of template DNA can result in omission of species (Hajibabaei et al. 2011). Although *Oncaea* sp. was a small specimen, *Metridia venusta* was a medium-sized copepod in the artificial community sample. Other factors, such as PCR biases, may be partly responsible for undetected species. The metagenetic approach is known to be effective for assessing diet (Pompanon et al. 2012); the appearance of species in the artificial community that were not included in the initial sample is attributed to the gut contents of carnivorous copepods.

There are known discrepancies between biomass and numbers of sequence reads, because of primer mismatches, length of amplicons, and copy numbers of rDNA genes (Porazinska et al. 2009; Machida and Knowlton 2012). In the artificial community analysis, there were up to two nucleotide mismatches to a single primer, which might have affected PCR efficiency. To

minimize mismatch, the 3′ region of each primer, which is important for PCR amplification, was designed to be conserved. Sequence length is susceptible to bias that occurs during the PCR emulsion step, however, 28S-D2 sequence length does not differ significantly among copepods. The number of sequence reads within MOTUs increased with dry weight in the artificial community (Fig. III-4), and the number of reads was suggested to be a proxy for biomass, though there were some discrepancies possibly because of biases described above.

The utility of metagenetic analysis was supported by analysis of field-collected samples. Both metagenetic and morphological analyses showed high species richness in the warm, western-boundary Kuroshio Current. In the epipelagic zone, copepod diversity is strongly correlated with temperature, and higher diversity is observed in warm oligotrophic oceans (Woodd-Walker et al. 2002; Rombouts et al. 2010). The highest species richness was observed at sampling sites that were strongly affected by the Kuroshio Current, as reflected in high temperature and salinity (Fig. III-2; Fig. III-5). The Kuroshio Current also transports plankton from lower latitudes and causes increased species diversity in the western North Pacific (Shih and Chiu 1998). Low temperature and salinity were observed at the Slope station and corresponded to low species richness and a community structure that differed from that of the other two sites.

The numbers of MOTUs were slightly higher than those of morphological species at all stations (Fig. III-5). Morphological identifications were made only for adult copepods, but the metagenetic analysis could detect immature stages, possible cryptic species, and gut contents, leading to higher estimates of species richness. Rarefaction curves of field-collected samples did not fully stabilize at maximum MOTU numbers, and greater numbers of sequence reads from field-collected samples might have led to larger numbers of MOTUs. The proportions of sequence reads were reasonably well correlated to those of DW in the corresponding species. Discrepancies between sequence reads and DW might occur as a result of the methodological biases of the metagenetic and morphological analyses.

In addition to correct MOTU classification at the family level, metagenetic analysis

revealed species richness of taxa that are difficult to identify morphologically because of their small size and subtle morphological differences but that have high species richness (e.g., Paracalanidae and Scolecitrichidae; Table III-5). The characteristic species composition of taxa in each hydrographic area was also detected in the MOTUs, although the species richness of some families was underestimated. The family Acartiidae is known to have distinct 28S sequences from other copepods (Chapter II; Kiesling et al. 2002); the divergent sequences within this family do not form a single phylogenetic group. The Acartiidae showed primer mismatches and short sequence length of the 28S-D2 region. Species richness of the Clausocalanidae as detected by MOTUs was low compared to that revealed by morphological analysis (Table III-5). The database of 28S-D2 regions showed small genetic distance among species in Clausocalanidae and thus the 97% similarity threshold may have been inadequate for this taxon. The lower species richness shown by MOTUs for the Calanidae may also be the result of small genetic variability in this family. The metagenetic method in this study was optimized to reflect a wide range of copepod taxa; however, different methods should be used for data analysis, or a different molecular marker should be selected, to detect species richness of taxa including the Acartiidae, Calanidae, and Clausocalanidae.

The relative composition of sequence reads agreed reasonably well with biomass composition at the family-level (Fig. III-8, III-9). Morphological analysis requires time-consuming sorting and dissection of individuals and fragmentation of some samples means that not all individuals can be analyzed. In contrast, metagenetic analysis includes all individuals, including immature stages, and rapidly detects biomass-dominant taxa. Discrepancies between sequence reads and dry weight in dominant families may have derived from methodological biases in the metagenetic analysis or from individuals that were not identified in the morphological classification. Despite known methodological biases, dominant taxa determined by metagenetic analysis provide valuable insights into the composition of the copepod community.

The identification of dominant species is important to understanding copepod community

structure and environmental conditions. For example, *Calanus sinicus* and *Paracalanus parvus* s.l., which were dominant at the Slope station (Table III-6), are known to be key species and important prey of forage fish in this region (Nakata and Hidaka 2003). In addition, metagenetic analysis detected species richness and biomass of small copepods (e.g., *Oithona, Paracalanus,* and *Clausocalanus*), the importance of which has been underestimated (Turner 2004). Although significant overestimation of MOTUs may occur at 99% similarity, this threshold may be appropriate for detecting dominant species because artificially inflated MOTUs normally contain small numbers of sequence reads.

The metagenetic analysis in this study can provide a rapid means of obtaining valuable information on copepod community structure, including immature stages and cryptic species. However, this is the first study to use metagenetic analysis and deep sequencing to focus on copepod diversity, and several aspects of the method (e.g., PCR bias and bioinformatics procedures) must be improved in future analyses. The lack of copepod 28S data is the primary problem in distilling massive quantities of sequence data into fine-scale taxonomy. Deficiencies in the reference database hinder accurate classification (Bik et al. 2012). The ecology of copepods has been well studied compared to other zooplankton. Each family of calanoid copepods has specific ecological characteristics, such as feeding habitat, patterns of vertical migration and energy storage, and reproductive strategy. However, classification to the genus-level would facilitate understanding of the functional roles of MOTUs from sequence data. Improving the robustness of the database is also necessary for identification of dominant species. Total 109 sequence data of 28S-D2 region were added to GenBank in this study. The 28S is commonly used for phylogenetic analysis or species identification; classification of MOTUs at finer scales might be achievable as the 28S database is expanded. The method in this study can be easily adapted to field-collected samples from other areas and can contribute to improving understanding of copepod community structure on the global scale.

Chapter IV. Spatial patterns of copepod community structure using metagenetic approach in the tropical and subtropical Pacific

Introduction

The novel metagenetic method was developed in chapter III, enabling a rapid and comprehensive analysis of copepod communities. This method is especially effective in the tropical and subtropical Pacific, where morphological classification of copepods is difficult because of occurrence of large numbers of species, small body size, and limited morphological information. In the tropical and subtropical Pacific, community analyses of copepods are restricted to local areas close to Japan, Hawaii and eastern tropical Pacific (e.g., Longhurst 1985a; Landry et al. 2001; Sugisaki et al. 2010). The subtropical area in South Pacific is the least-well-known region of the oceans (Longhurst 2007), and only limited numbers of data are available for copepod community (Williamson and McGowan 2010). There is currently no study comparing community structure of copepods across large areas in the tropical and subtropical Pacific. Therefore, this study could serve as a basis for the understanding of diversity and biogeography of copepods in the tropical and subtropical Pacific.

The purpose of this chapter is to reveal spatial patterns of copepod community structure in the tropical and subtropical Pacific. The metagenetic method was applied to field-collected samples in four ocean regions: Kuroshio, North Pacific subtropical gyre (NPSG), eastern tropical Pacific (ETP), and South Pacific subtropical gyre (SPSG). The community structures of copepods were compared among and within ocean regions, and mechanisms governing copepod diversity and biogeography were discussed in relation to environmental variables.

Materials and Methods

Sampling

Total 20 plankton samples were collected at 19 stations in four ocean regions in the tropical

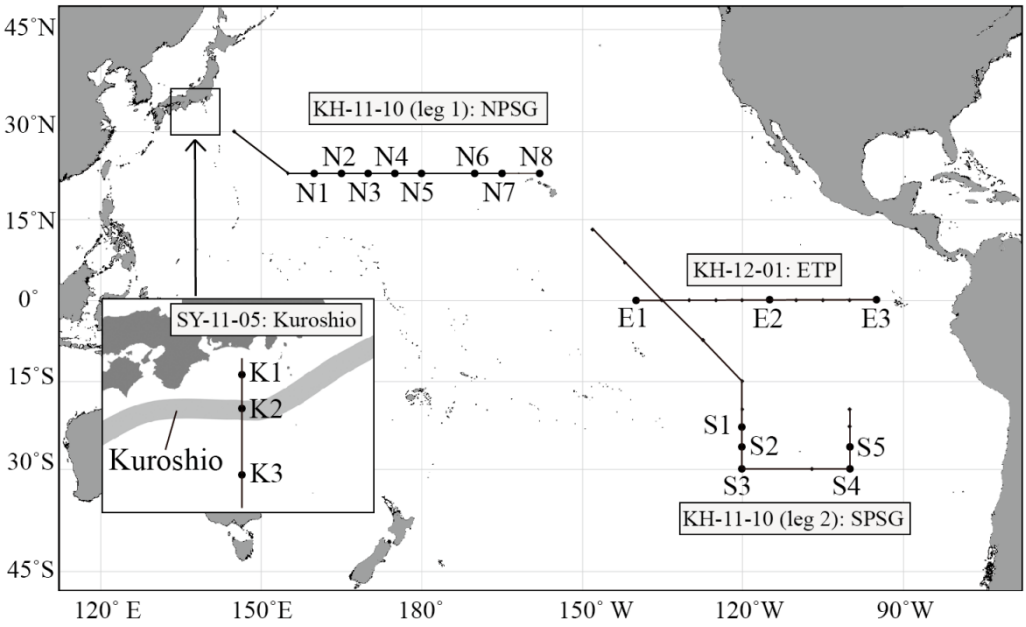


Fig. IV-1 Sampling locations and cruise tracks in this study.
Zooplankton samples were collected at 19 stations during three cruises of SY-11-05, KH-11-10 and KH-12-01 in four ocean regions: Kuroshio, NPSG, ETP and SPSG. Samplings were performed both during the day and night at the station E1. Chl-*a*, salinity and temperature were measured throughout cruises.

and subtropical Pacific (Fig. IV-1): Kuroshio (K1-K3), NPSG (N1-N8), ETP (E1-E3) and SPSG (S1-S5). According to Longhurst's biogeographical provinces (Longhurst 2007), the region Kuroshio was located in Kuroshio Current Province, NPSG in North Pacific Tropical Gyre Province, ETP in Pacific Equatorial Divergence Province, and SPSG in South Pacific Subtropical Gyre Province. Samples in chapter III were used for the station K1-K3 in Kuroshio, which were collected at the daytime during the SY-11-05 cruise aboard the RV "Soyo-Maru" (National Research Institute of Fisheries Science, Fisheries Research Agency). The station K1 was located shoreward of the Kuroshio Current and K2 at the core of the current. Other samples were newly collected in this study at 16 stations during the KH-11-10 and KH-12-01 cruises aboard the RV "Hakuho-Maru" (Japan Agency for Marine-Earth Science and Technology). Samples were collected at 0–200 m during the night using a vertical multiple plankton sampler (VMPS; Terazaki and Tomatsu 1997) with a 0.5 m^2 mouth opening area and 100 μm mesh size. At the station E1, samplings were performed both during the day

and night. All samples were preserved in 99% ethanol and kept at 4ºC. Ethanol was replaced after 24 h from initial preservation. Parameters of environmental variables were obtained throughout cruises (Fig. IV-1). The profiles of temperature and salinity were obtained using CTD systems (SBE-911 plus, Sea-Bird Electronics), and dissolved oxygen was measured using SBE-43 dissolved oxygen sensor (Sea-Bird Electronics). Water samples were collected with Niskin bottles attached to CTD system and were filtered using Whatman GF/F filters for chl-*a* analysis. Chl-*a* was extracted with N, N-dimethylformamide, and chl-*a* concentration was analyzed using a Turner fluorometer according to Welschmeyer (1994). Mixed layer depth (MLD) was calculated by the depth at temperature of ΔT = 0.2 ºC from temperature at 10 m depth (de Boyer Montégut et al. 2004).

DNA extraction, PCR amplification and 454 pyrosequencing

DNA extraction and PCR amplification were conducted for newly collected samples using the method in chapter III. The method of DNA extraction using the Gentra Puregene Cell and Tissue Kit (QIAGEN) was slightly modified, and 50 mL tube with 12 mL of Cell Lysis Solution was used in the first step. Three ethanol-preserved samples at the station E1-E3 were split, and 1/2 aliquots were used for DNA extraction due to large biomass. Two pyrosequencing runs were carried out on a 454 GS FLX system (Roche). Each sample was differentiated by unique multiplex identifier (MID) adaptors. All procedures for 454 pyroseqencing were performed according to the manufacturer's protocol.

The newly obtained 454 data were combined with previously obtained data of K1-K3 (Chapter III). All sequence reads were quality-filtered and clustered into MOTUs at 97% similarity threshold using the method in chapter III which was mainly performed in MOTHUR (Schloss et al. 2009). In the step of alignment of sequence reads, I used add-fragments option in MAFFT with default setting (Katoh and Standley 2013). To standardize sequence data, 6,229 sequence reads (minimum number of sequence reads among all samples) was randomly selected in each sample for the following community analysis.

Fig. IV-2 Vertical profiles of temperature, (2) salinity, and (3) chlorophyll *a* at 0-250 m depth during cruises of (a) SY-11-05 in Kuroshio, (b) KH-11-10 leg 1 in NPSG, (c) KH-12-01 in ETP, and (d) KH-11-10 leg 2 in SPSG.
Each cruise track is shown in Fig. IV-1. Sampling stations for plankton community are indicated by black triangles.

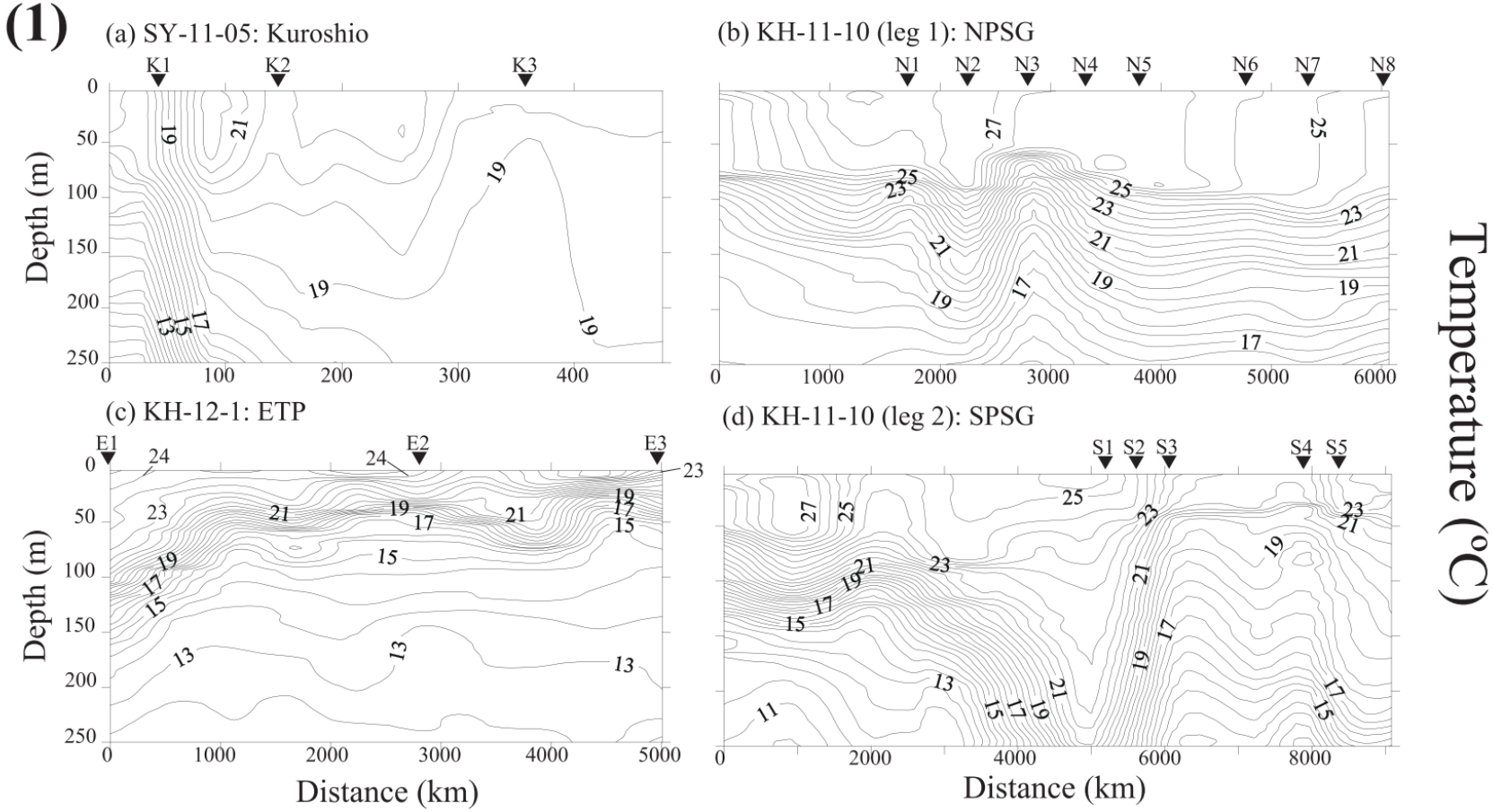


Fig. IV-2 Continued.

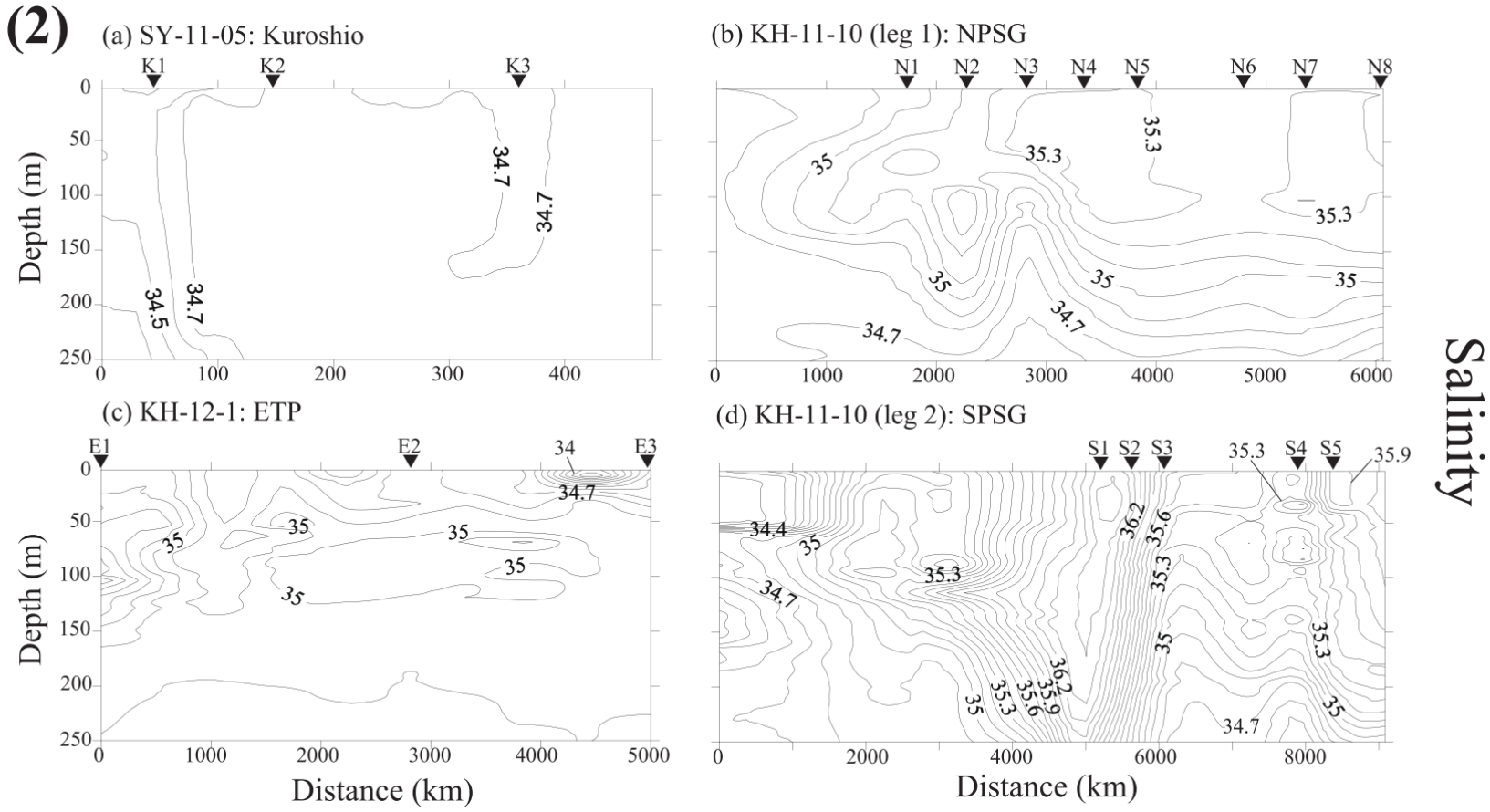
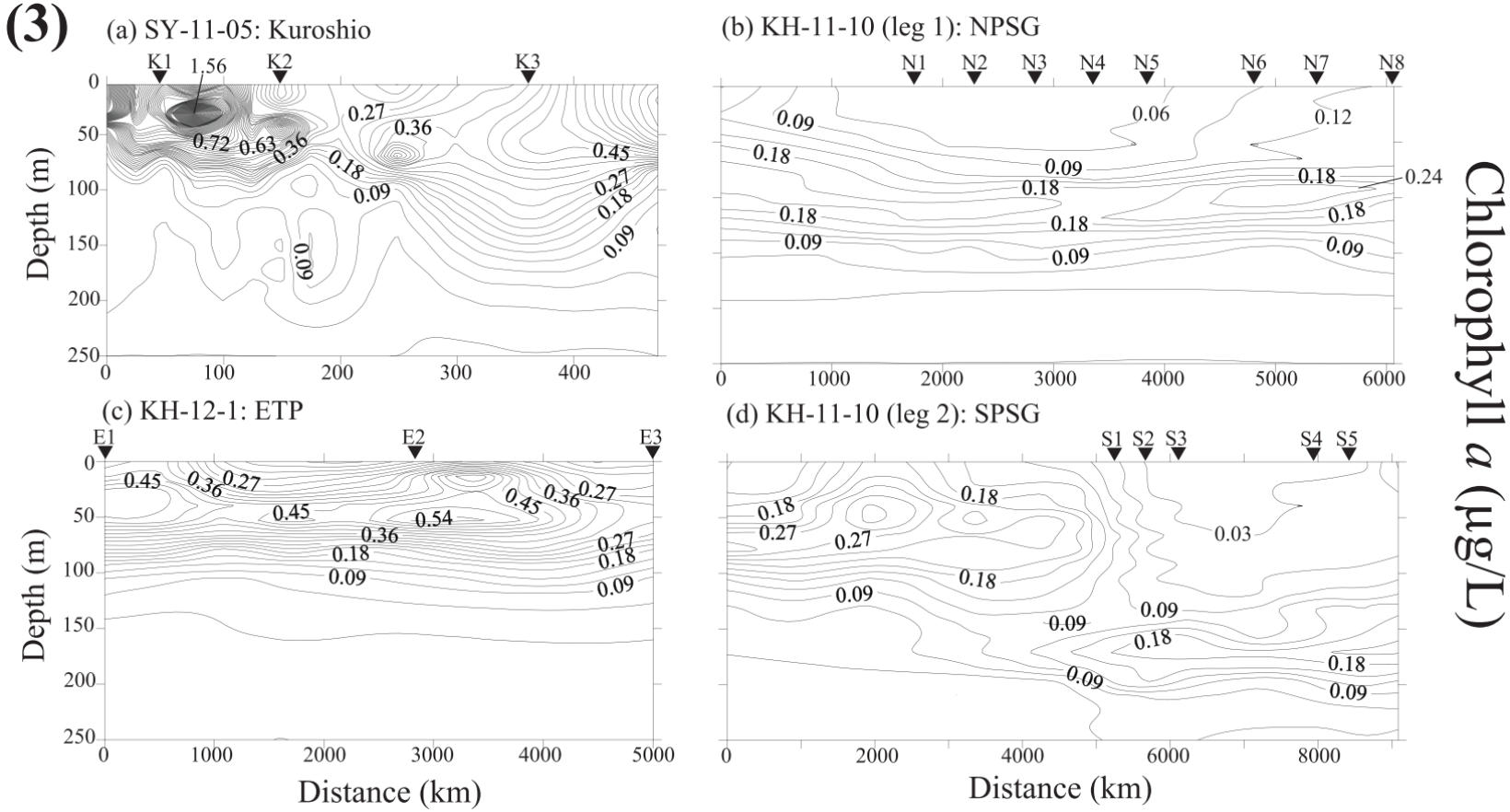


Fig. IV-2 Continued.



Community analysis of copepods

A number of MOTUs was used as a proxy of number of species and a number of sequence reads as a proxy of biomass. It is noted that studies of copepods have been conducted based on abundance (number of individuals), however, metagenetic data used in this study were based on proportions of biomass for community analysis of copepods. All MOTUs were classified into taxonomic order using the naïve Bayesian classifier (Wang et al. 2007) in MOTHUR, with 70% threshold. MOTUs below the 70% threshold were treated as 'unclassified' MOTUs. Calanoid MOTUs were classified into family, and the genus *Calocalanus* was further distinguished from the family Paracalanidae. The composition of MOTUs and sequence reads in each taxon were obtained in each sample. A similarity of copepod community was investigated based on the Bray-Curtis similarity by cluster analysis using group average and by multidimensional scaling (MDS). Both biomass-based and non-biomass-based were conducted. Biomass-based analysis contained numbers of sequence reads in each MOTU, and the Bray-Curtis similarity was calculated based on number of sequence reads. Non-biomass-based only used data of presence or absence of MOTUs. To detect major distribution patterns of copepods, cluster analysis based on the Bray-Curtis similarity (Bray & Curtis 1957) was also performed for MOTUs with at least 30 sequence reads in a single sample. To standardize sequence data of each MOTU, sequence reads were divided by total sequence reads. Therefore, this cluster analysis was based on relative proportions of sequence reads in each MOTU. Cluster and MDS analysis were performed in PRIMER version 6 (Plymouth Marine Laboratory, UK; Clarke and Warwick 2001).

The relationship between copepod community and environmental variable was assessed using redundancy analysis (RDA) in CANOCO version 5 (Microcomputer Power). RDA was selected based on the result of detrended canonical correspondence analysis. I used environmental variables of temperature, salinity, dissolved oxygen, chl-*a*, and MLD. Both surface and average at 0–200 m were used for temperature. Salinity and oxygen were averages of 0–200 m, and chl-*a* was integrated amount at 0–200 m. The forward selection

method (ter Braak and Verdonschot 1995) was used to determine and order statistically significant environmental variables for RDA ($p < 0.05$). A Monte Carlo permutation test was performed using 499 unrestricted permutations to test significant correlation between environmental variables and copepod community. The significance of the first axis and all combined axes in RDA ordination were also tested by 499 permutations.

Diversity of copepods

MOTU richness (proxy of species richness) was evaluated by number of MOTUs, non-parametric abundance-based coverage estimator Chao 1 and ACE, which were calculated using MOTHUR. In addition, diversity index of Shannon and Simpson index (Gini-Simpson index) were calculated to investigate species evenness (Shannon 1948; Simpson 1949). These values were calculated based on sequence reads (proxy of biomass) and should be differentiated from previously published data based on number of individuals. The non-parametric Kruskal-Wallis test was used to identify significant differences of observed MOTU numbers among ocean regions. A significant difference between two ocean regions was tested by the non-parametric Mann-Whitney *U* test. Total number of observed MOTUs was obtained in each ocean region. The number of shared MOTUs was also calculated between ocean regions. The correlations of environmental variables with MOTU richness and diversity index were investigated by Pearson's correlation coefficient to detect effective environmental variables on copepod diversity. Only night sample was used at the station E1 for statistical analyses. Statistical analyses were performed using SPSS 21.0 (IBM Corporation).

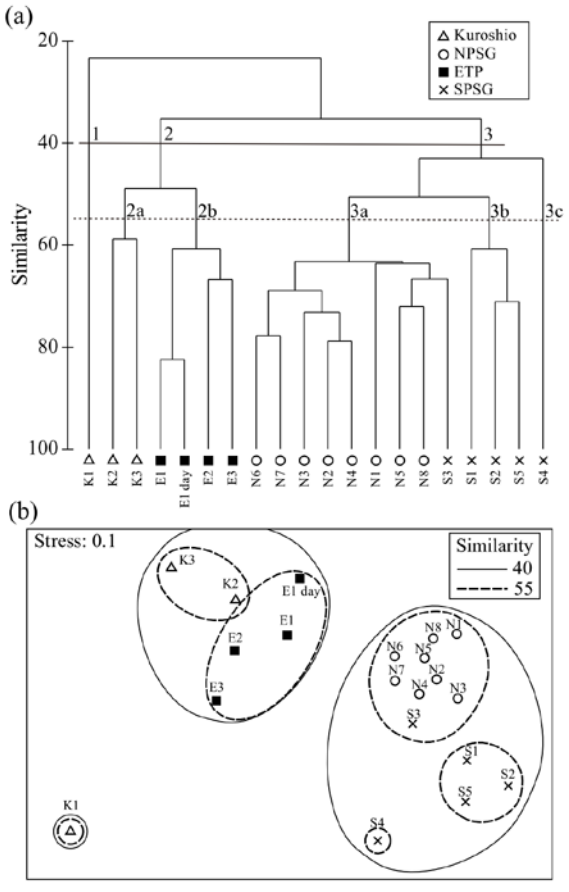
Results

Temperature, salinity and chl-a

There were significant differences of water temperature and chl-*a* concentration among ocean regions (Fig. IV-2). A significantly low total chl-*a* was observed in NPSG (14.2–26.7

mg m^{-2}) and SPSG (10.3–14.7 mg m^{-2}), compared with those in Kuroshio (52.6–82.6 mg m^{-2}) and ETP (33.2–56.7 mg m^{-2}). Sea surface temperature (SST) was the highest in NPSG (24.5–27.8 ºC), followed by ETP (23.9–25.4 ºC) and SPSG (22.4–25.4 ºC). SST was relatively low in Kuroshio (17.3–21.9 ºC). SPSG were characterized by the highest salinity (35.2–36.4), however, salinity was not significantly different among ocean regions.

Variations of environmental factors were also observed within each ocean region. In Kuroshio region, a large hydrographic difference was observed at the boundary of the Kuroshio Current. The lowest SST and the highest chl-*a* among all 19 stations were observed at the station K1. The higher SST and lower chl-*a* were observed at the station K2 than K3. In ETP, there were decreases of SST, salinity and chl-*a* from E1 to E3. In NPSG, relatively high chl-*a* and low temperature were observed at the station N6-N8. In SPSG, high temperature and salinity were observed through the epipelagic layer at the station S1-S2. The station S3-S5 was affected by the intrusion of cold and low salinity water, which was the most evident at the station S4.



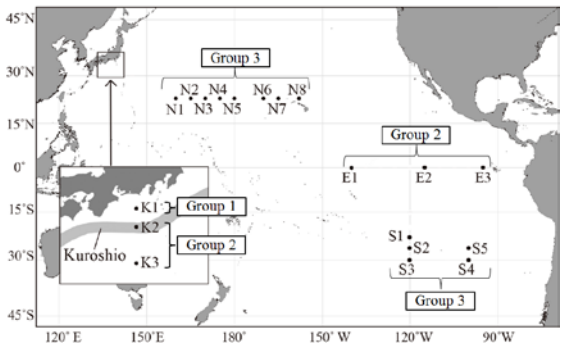


Fig. IV-3 Biomass-based Community analysis of copepods: (a) dendrogram by cluster analysis and (b) MDS ordination based on Bray-Curtis similarity. Copepod communities are clustered into three groups at the 40% similarity and six groups at the 55% similarity.

Biomass-based community structure of copepods

Total 434,304 quality-filtered sequence reads were obtained, and they were clustered into 404 MOTUs at 97% similarity threshold. To standardize sequence reads, 6,229 sequence reads were randomly selected from each sample. In this standardized dataset, 383 of 404 MOTUs were selected.

In biomass-based cluster analysis, copepod communities were clustered into three groups at the 40% similarity: group 1: station K1, group 2: Equator-Kuroshio group containing the station K2-K3 and E1-E3, and group 3: subtropical gyre group containing N1-N8 and S1-S5 (Fig. IV-3a). At the 55% similarity, group 2 was further divided into two subgroups of Kuroshio (group 2a) and ETP (group 2b). Samples during the day and night at the station E1 showed the highest similarity in all samples. Three subgroups were obtained from group 3 at the 55% similarity. The community structure at the station S4 (group 3c) was distinct from other samples in subtropical gyres. All stations in NPSG and the station S3 in the SPSG were in a single cluster (group 3a). Three stations in SPSG of S1, S2 and S5 formed a single cluster (group 3b). MDS showed consistent results with cluster analysis (Fig. IV-3b). The community structure at the station S3 in group 3a showed close relationship to other stations in SPGS. NPSG showed more similar community structures to Equator-Kuroshio group than did SPSG.

In the taxonomic composition of sequence reads (proxy of biomass), the group 1 in the cluster analysis (station K1) was characterized by large proportions of Cyclopoida, followed by Paracalanidae, Calanidae and Eucalanidae (Fig. IV-4). Other taxa only accounted for small proportions of sequence reads. Group 2 (Equator-Kuroshio group) was characterized by the low proportions of Cyclopoida, compared with group 1, although those of Calanidae, Eucalanidae and Paracalanidae remained high with proportions. Within group 2, proportion of Eucalanidae was higher in Kurohio region (group 2a) than in ETP (group 2b), whereas those of Calanidae and Paracalanidae were more dominant in ETP than in Kuroshio. The family Clausocalanidae and Euchaetidae also showed relatively high proportions in ETP. In group 3 (subtropical gyre group), Calanidae, Clausocalanidae and Paracalanidae (including

Calocalanus) were dominant. There were also high proportions of sequence reads in various taxa, such as Aetideidae, Augaptilidae, Candaciidae, Centropagidae, Clausocalanidae Lucicutiidae, Metridinidae and Poecilostomatoida, compared with group 1 and 2. The absence of Rhincalanidae and small proportion of Eucalanidae were also characteristics in group 3. Eucalanidae were almost restricted to east side of NPSG (N6-N8). In SPSG, station S4 was distinct from other group 3, and high proportions of Mecynoceriidae and Cyclopoida were observed.

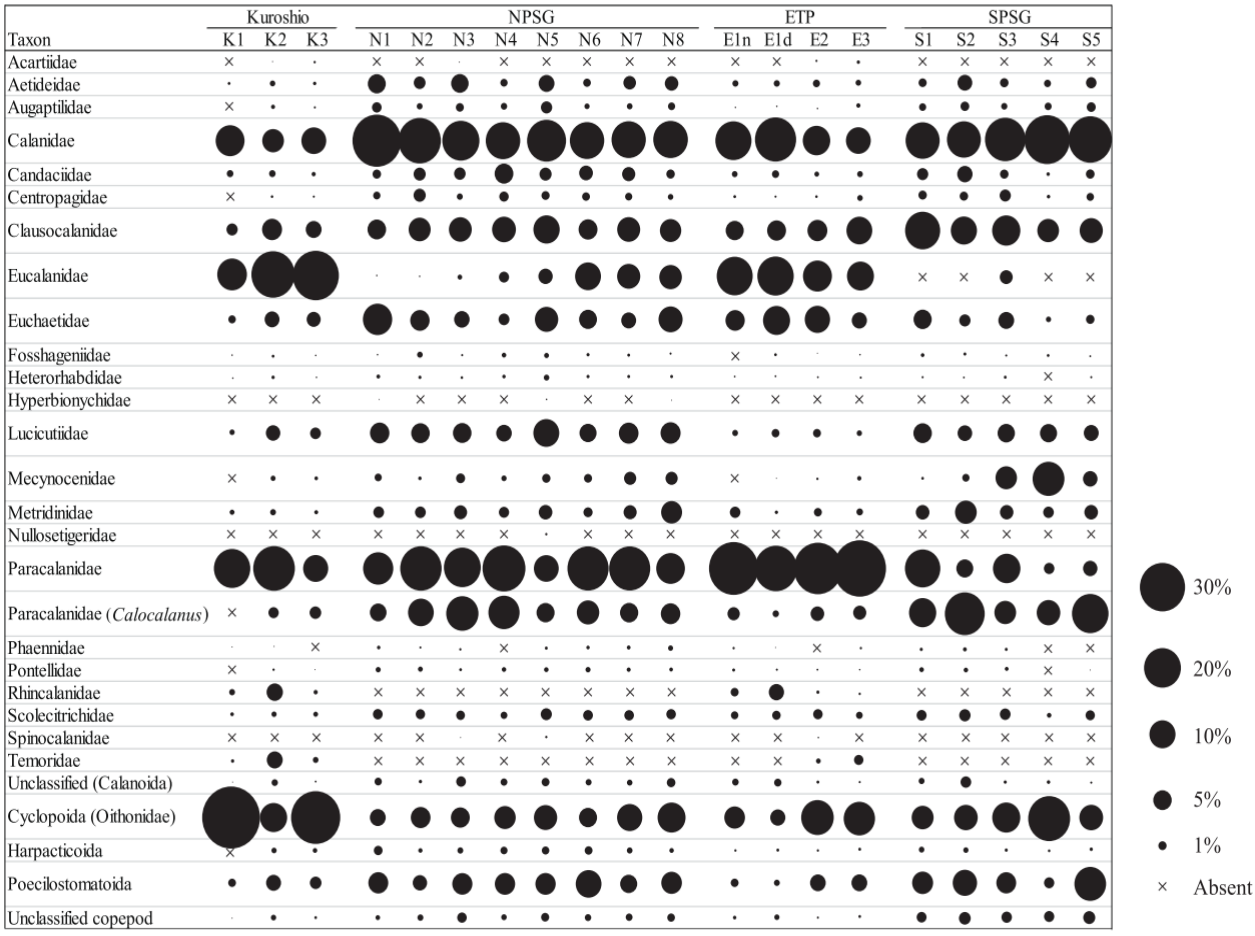


Fig. IV-4 Taxonomic compositions of copepod sequence read at each station.
The normalized sequence data (6,229 reads) is used for each station. Black circle in indicates proportion sequence reads in each taxon. The order calanoid were classified into family, and the genus *Calocalanus* was further classified from the family Paracalanidae.

Distribution patterns of major MOTUs

Total 83 MOTUs with at least 30 sequence reads in a single site were clustered into seven groups (group A-G) at the 40% similarity, based on the relative distribution peaks of sequence reads (Fig. IV-5). Group A-D was characterized by distributions in subtropical gyres, whereas MOTUs in Group E-G were almost absent in subtropical gyres (Fig. IV-6). Group A showed distribution peak at the station N6-N8 in NPSG, whereas group B at the station S4 in SPSG. Major part of MOTUs was clustered into group C (47/83 MOTUs) and group D (17/83 MOTUs). Both groups were widely distributed; however, distribution peaks of group C were in subtropical gyres, whereas those of group D were in Equator-Kuroshio region. The distributions of group E-G were restricted to specific area, and the distribution peak was at the station K1 in group E, at the station E1 in group F, and at the station K2 in group G.

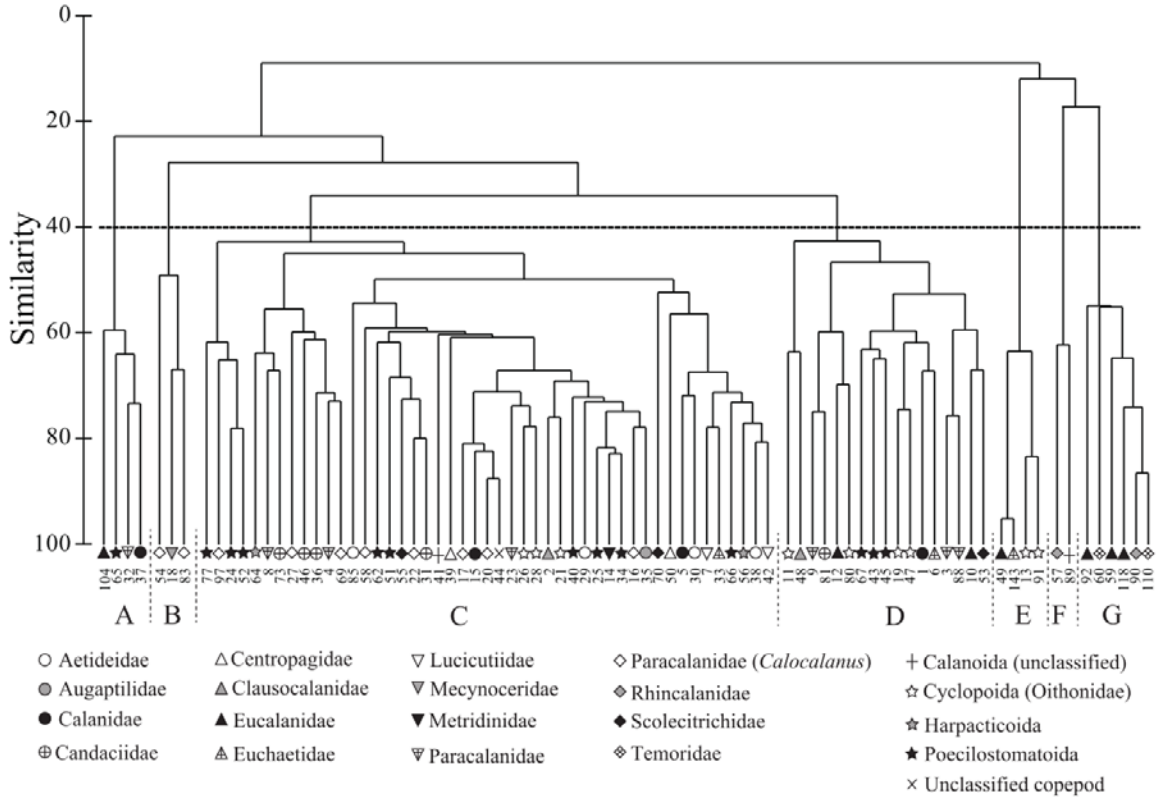


Fig. IV-5 Cluster analysis of major MOTUs based on Bray-Curtis similarity.
Total 83 MOTUs with at least 30 sequence reads in a single site are clustered by the relative peak of distribution. Sequence reads in a single sample is divided by total number of sequence reads in each MOTU. MOTUs are clustered into seven groups (A-H) at the 40% similarity. Taxonomic information is determined using the naïve Bayesian classifier.

The cluster analysis of MOTUs based on distribution patterns were related to taxonomy. For example, MOTUs in genus *Calocalanus* (8/83 MOTUs) were only observed in group B and C, whereas MOTUs in Eucalanidae (7/83 MOTUs) were absent in group B and C. Temoriidae MOTUs were restricted to group G and Rhincalanidae MOTUs to group F and G. A large MOTU number with various distribution patterns was observed in Paracalanidae except for *Calocalanus* (7 MOTUs in group A, C and D), Oithonidae in Cyclopoida (9 MOTUs in group C-E), and Poecilostomatoida (13 MOTUs in group A, C and D). In group C, there were various taxa of 4 orders and 11 families in calanoid copepods.

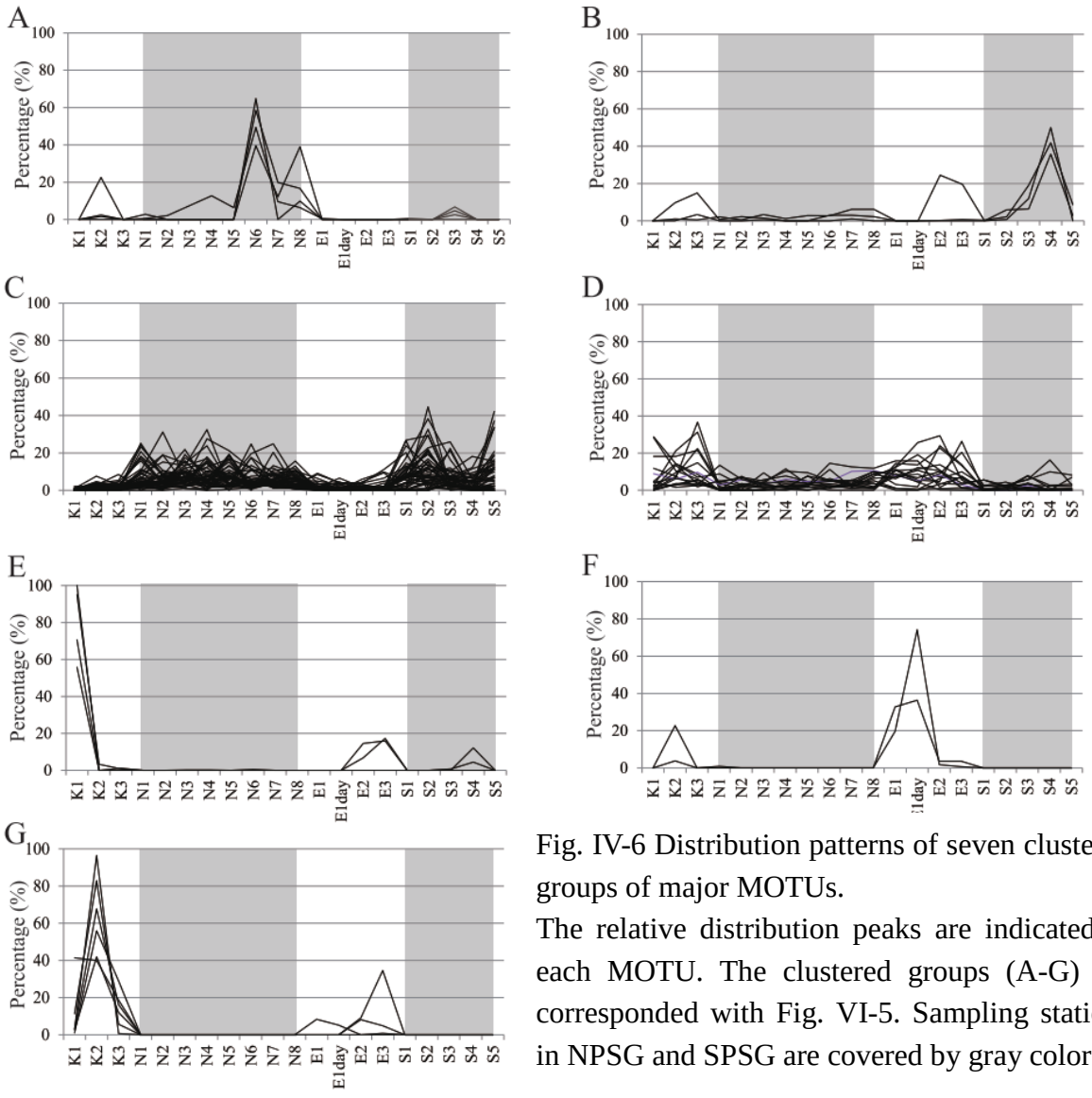


Fig. IV-6 Distribution patterns of seven clustered groups of major MOTUs.
The relative distribution peaks are indicated in each MOTU. The clustered groups (A-G) are corresponded with Fig. VI-5. Sampling stations in NPSG and SPSG are covered by gray colors.

RDA ordination of biomass-based copepod community

According to the forward selection, the most significant environmental variable was chl-*a* (F = 6.0, p = 0.002) to explain copepod community structure in first axis and all combined axes, followed by SST (F = 4.0, p = 0.002), dissolved oxygen (F = 2.9, p = 0.024) and MLD (F = 2.3, p = 0.036). These four environmental variables explained 30.9% of the variance of MOTU data by the first axis and 13.6% by the second axis in RDA ordination (Fig. IV-7). The variability of copepod community structure was explained significantly by both the first axis and the all axes combined ($p < 0.01$). The clustered groups of Equator-Kuroshio (Group 2 in Fig. IV-3) and subtropical gyre (Group 3) were clearly separated by the environmental variable of chl-*a*. The station K1 (Group 1) was discriminated from other samples mainly by the high chl-*a* and low surface temperature.

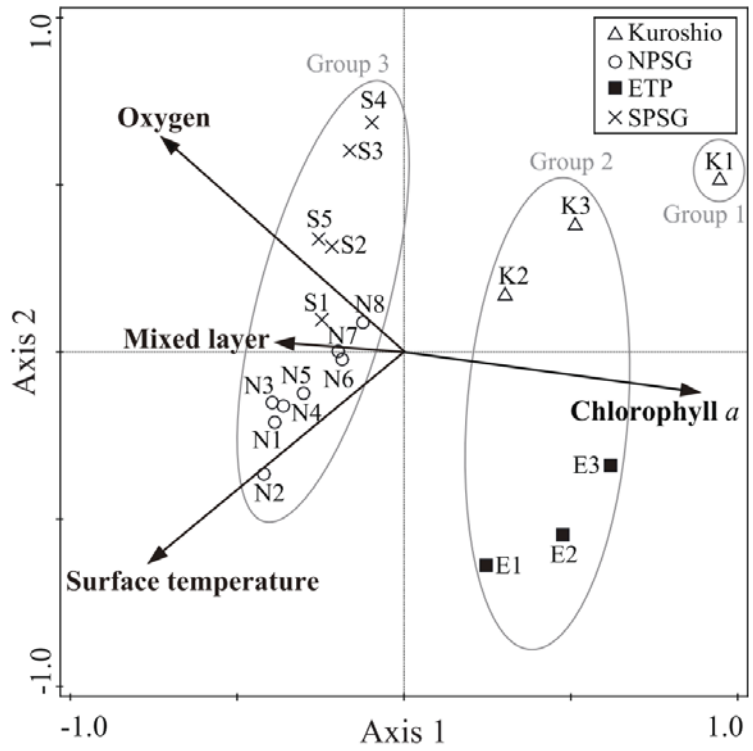


Fig. IV-7 RDA ordination for copepod community and environmental variable. Black arrows indicate significant environmental variables on MOTU data, which are selected by forward selection. 30.9% of the variance of MOTU data is explained by the first axis and 13.6% by the second axis. The arrow length of environmental variable indicates the strength of relationship with community structures. The cluster groups at 40% similarity (Fig. IV-3) are represented by gray circles.

Non-biomass- based community structure of copepods

Non-biomass-based community analysis showed same clustered groups as those in biomass-based analysis at 55% similarity (Fig. IV-8). An overall similarity among samples in non-biomass-based analysis was high, compared with biomass-based analysis. As well as biomass-based analysis, the station K1 was significantly different from other stations, and the station S4 showed different MOTU composition from other stations in subtropical gyres. In contrast to biomass-based analysis, the group 2a (oceanic side in Kuroshio region) was more similar to stations in subtropical gyres (group 3a-3c) than to stations in ETP (group 2b). As well as biomass-based analysis, NPSG showed higher similarity of MOTU composition to ETP and Kuroshio regions than did SPSG.

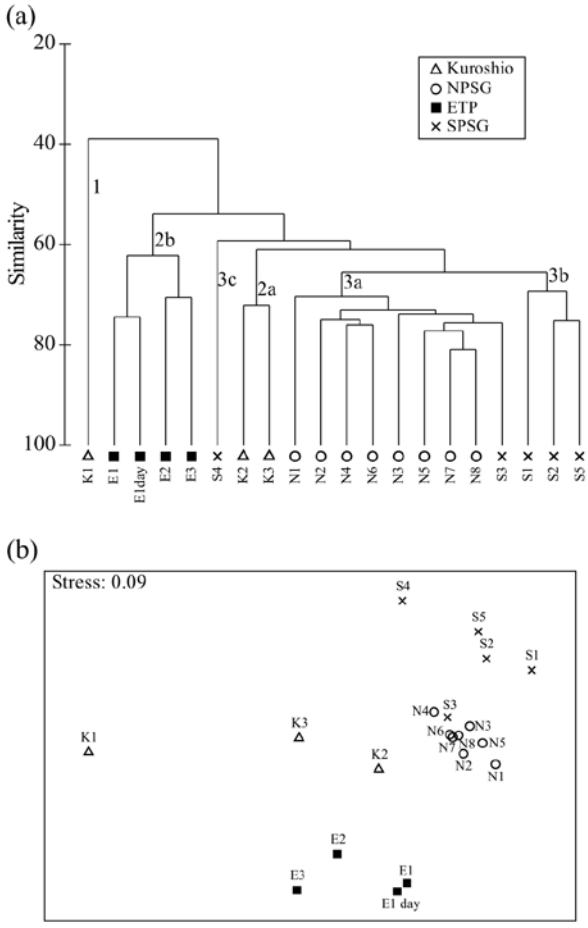


Fig. IV-8 Non-biomass-based (presence/absence of MOTUs) community analysis of copepods: (a) dendrogram by cluster analysis and (b) MDS ordination of all copepod community samples based on Bray-Curtis similarity.
The number of sequence reads was not taken into account. The clustered groups at a 55% similarity in biomass-based cluster analysis (Fig. IV-3) are represented in dendrogram.

As represented in relatively high similarity among stations in non-biomass-based cluster analysis, the difference of taxonomic compositions of MOTUs (MOTU number in each taxon) were less evident among samples (Fig. IV-9), compared with taxonomic compositions of sequence reads (Fig. IV-4). Most of families had relatively large MOTU numbers in the subtropical gyre especially in NPSG. However, some taxa, such as Eucalanidae, Rhincalanidae, Temoriidae and Cyclopoida showed large MOTU numbers in ETP and Kuroshio regions. At the station K1, there was absence (less than threshold value) of specific taxa, such as family Augaptilidae, Centropagidae, Mecynoceridae, Pontellidae and the genus *Calocalanus*.

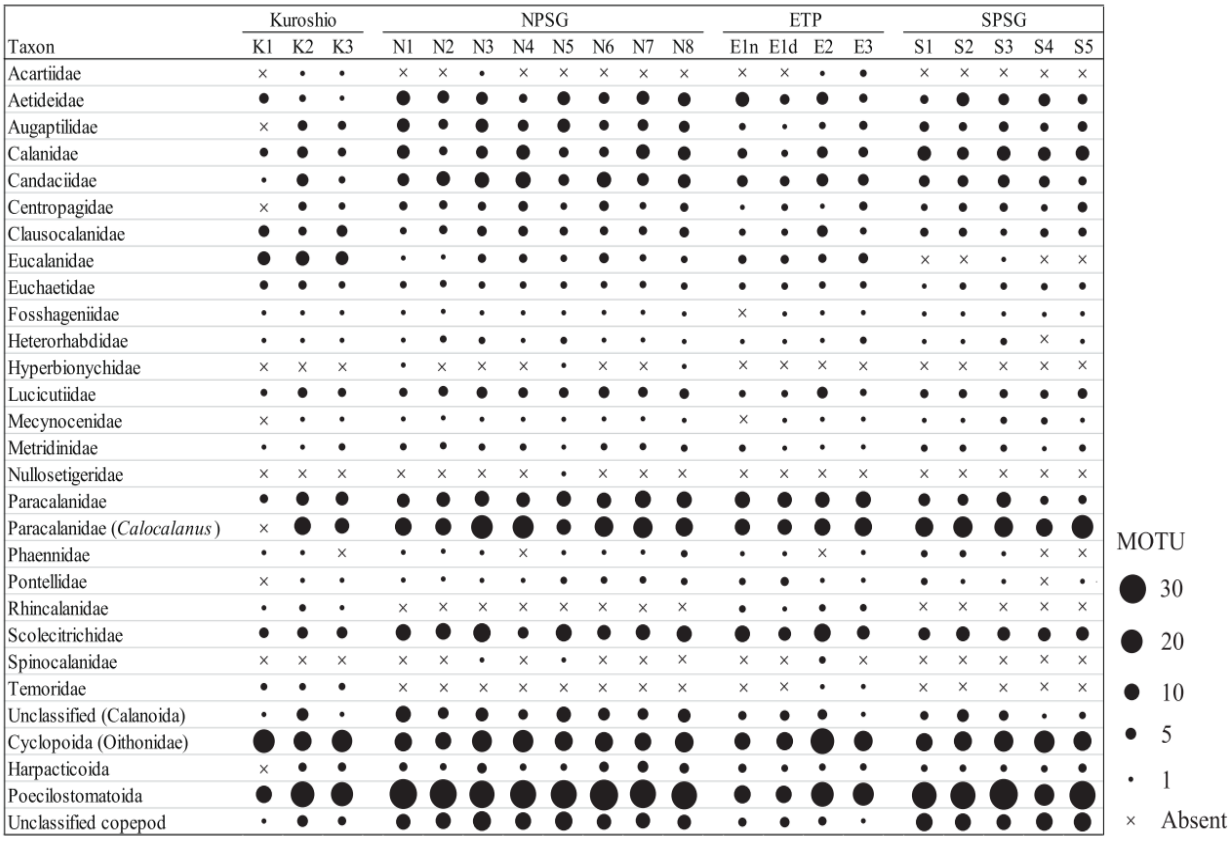


Fig. IV-9 Taxonomic compositions of copepod MOTUs at each station.
The normalized sequence data (6,229 reads) is used for each station. Black circle in indicates numbers of MOTUs in each taxon. MOTUs in the order calanoid were classified into family, and the genus *Calocalanus* was further classified from the family Paracalanidae.

Spatial patterns of copepod diversity

The MOTU richness (observed MOTU number, Chao 1 and ACE) and diversity index (Shannon and Simpson) were higher in NPGS than in other ocean regions, with the highest numbers of 166 observed MOTUs at the station N3 (Table IV-1). Both Chao 1 and ACE richness estimators were also high at the station N3 (Chao 1: 213.1 MOTUs; ACE: 216.7 MOTUs). MOTU richness and diversity index were relatively low at the station N2 within NPSG. MOTU richness and diversity index were also high in SPSG, though MOTU richness

Table IV-1 MOTU richness and diversity index in each sampling site.
The standardized sequence data (6,229 sequence reads) is used for a single sample. MOTU richness is estimate from the observed number of MOTUs at 97% similarity, and non-parametric abundance-based coverage estimator Chao 1 and ACE. The diversity index is estimated from Shannon and Simpson (Gini-Simpson) index.

		Richness			Diversity index	
Area	Station	MOTUs	Chao 1	ACE	Shannon	Simpson
Kuroshio	K1	72	91.1	97.5	2.11	0.78
	K2	125	160.1	147.4	3.25	0.92
	K3	105	116.7	124.4	2.62	0.84
NPSG	N1	143	182.0	217.5	3.18	0.91
	N2	130	142.7	147.1	3.19	0.92
	N3	166	213.1	216.7	3.60	0.95
	N4	145	190.3	227.4	3.44	0.93
	N5	143	178.4	174.0	3.60	0.95
	N6	147	174.1	169.5	3.56	0.95
	N7	143	167.5	171.1	3.48	0.94
	N8	146	175.3	176.9	3.62	0.96
ETP	E1 (night)	99	128.0	156.7	2.80	0.90
	E1 (day)	89	112.1	115.7	2.65	0.89
	E2	133	199.2	179.9	2.82	0.88
	E3	108	119.9	122.2	2.84	0.90
SPSG	S1	119	181.0	152.1	3.19	0.92
	S2	129	152.0	148.0	3.58	0.95
	S3	139	181.5	174.9	3.52	0.95
	S4	107	126.5	131.3	2.73	0.89
	S5	129	188.5	164.8	3.36	0.93

in SPSG was lower than in NPSG. Within SPSG, MOTU richness and diversity index tended to increase from S1 to S3 but significantly decreased at the station S4. The diversity index in ETP was low, compared with that in NPGS and SPGS (except for the station S4). MOTU richness was also low at the station E1 and E3, however, the station E2 showed high MOTU richness with almost equivalent or superior to those in SPSG. At the station E1, MOTU number increased 11% at the night, and diversity index were higher at the night than the day. All the lowest values of MOTU richness and diversity index were observed at the station K1 in Kuroshio. Within Kuroshio, there was the peak of MOTU richness and diversity index at the station K2, with higher diversity index than that in ETP.

The Kruskal-Wallis test identified a significant difference of observed numbers of MOTUs among ocean regions ($p < 0.005$). The Mann-Whitney *U* test detected that the number of observed MOTUs in NPSG was significantly higher than that in other areas ($p < 0.05$). All MOTU richness and diversity index were correlated positively to SST ($p < 0.01$) and negatively to chl-*a* ($p < 0.01$; Table IV-2). The correlation of MOTU richness was the highest to SST (observed MOTUs: R=0.77, Chao 1: R = 0.66, ACE: R = 0.717). In terms of diversity index, the correlation to total chl-*a* was the highest (Shannon: R = -0.762, Simpson: R = -0.792).

Table IV-2 Pearson's correlation between environmental variables and copepod diversity (MOTU richness and diversity index).
*p < 0.05, **p < 0.01

	Observed MOTUs	Cho 1	ACE	Shannon	Simpson
Surface temperature	0.77 **	0.66 **	0.717 **	0.676 **	0.746 **
Average temperature	0.44	0.428	0.481 *	0.699 **	0.641 **
Average salinity	0.308	0.358	0.142	0.465 *	0.524 *
Total chlorophyll *a*	-0.684 **	-0.623 **	-0.541 **	-0.762 **	-0.792 **
Dissolved Oxygen	0.171	0.324	0.278	0.519 *	0.412
Mixed layer	0.77	0.58	0.167	0.325	0.169

Total numbers of observed MOTUs was the highest in NPSG of 274 MOTUs, followed by 232 MOTUs in SPSG (Fig. IV-10). Total 163 MOTUs was observed in Kuroshio and 187 MOTUs in ETP. Total 90 MOTUs were distributed in all ocean regions. NPSG highly shared MOTUs with other ocean region, and 130/163 (79.8%) MOTUs in Kuroshio and 136/187 (72.7%) MOTUs in ETP were also observed in NPSG. The highest numbers of 184 MOTUs were shared between NPGS and SPSG, occupying 67.2% of MOTUs in NPSG and 79.3% of MOTUs in SPSG.

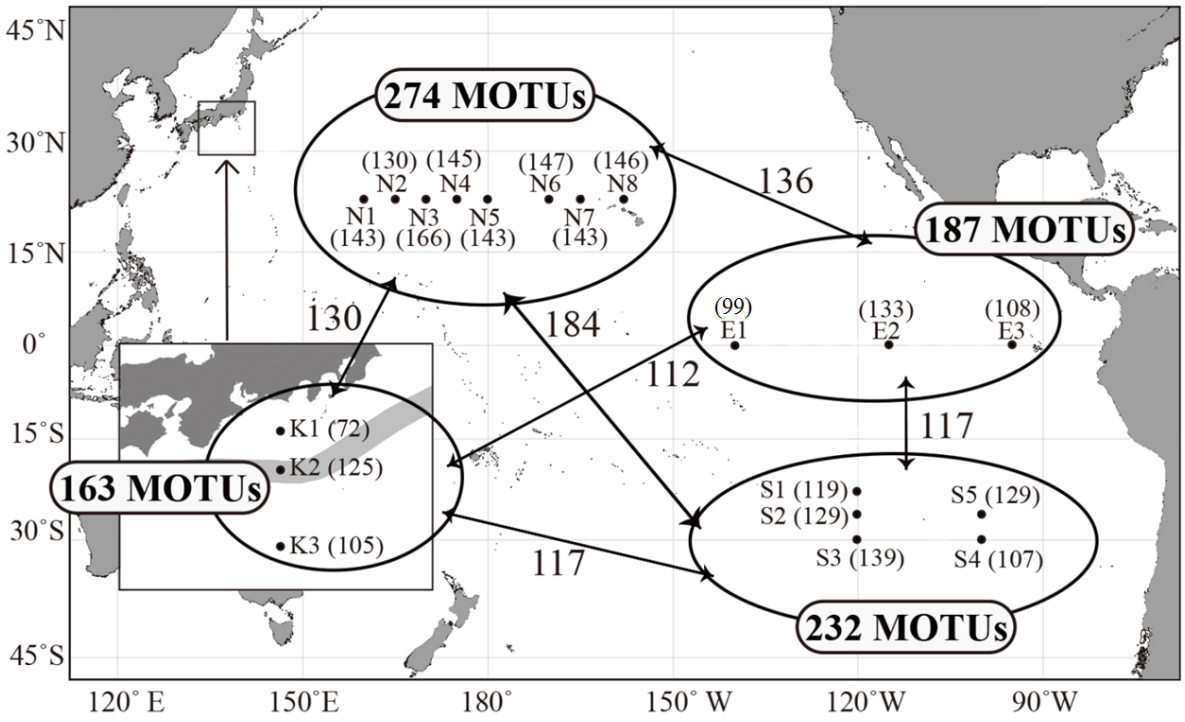


Fig. IV-10. Number of total and shared MOTUs in each ocean region.
The standardized sequence data (6,229 reads in a single sample) is used to calculate observed numbers of MOTUs. Total numbers of MOTUs were represented in each ocean regions of Kuroshio, NPSG, ETP and SPSG. The number next to arrows indicate numbers of shared MOTUs between ocean regions. Values in parentheses indicate observed numbers of MOTUs in each station. The sample collected at the night is used for the station E1.

Discussion

A newly developed metagenetic technique revealed the community structure of copepods, which were largely affected by ocean regions with different water masses and environmental conditions in the tropical and subtropical Pacific. The cluster analysis showed the highest similarity between day and night samples at the station E1, suggesting an insignificant effect of sampling time on community structure in this study. Woodd-Walker et al. (2002) also mention that the sampling time is not a major source of difference in large-scale community structures of copepods. Seasonal changes are not evident in lower latitude than in higher latitude (Longhurst 2007), and seasonality is not considered in this study. A boundary of biogeographic province is mainly determined by ocean current and water mass with large hydrographic differences (Longhurst 2007), agreeing with cluster analysis and MDS ordination in this study (Fig. IV-3). In this study, hydrographic condition was significantly different at the boundary of the Kuroshio Current, and copepod community was distinct at the shoreward side of the current where influence of coastal water is dominated (Fig. IV-3, 8). The other copepod communities were clearly divided into the subtropical gyre and the Equator-Kuroshio groups. These groups agreed with major distribution patterns with peaks in subtropical gyres or Equator-Kuroshio, which are common distribution patterns of planktonic taxa in the Pacific (McGowan 1971). The subtropical gyre were characterized by lower chl-*a* compared with Equator-Kuroshio region, and copepod communities of these two groups are clearly correlated with chl-*a* difference (Fig. IV-5). Woodd-Walker et al. (2002) has also proposed that temporal and spatial patterns of primary production determine the large-scale copepod communityies.

MOTU compositions (present/absent) in the oceanic side of the Kuroshio region (station K2 and K3) are more similar to those in subtropical gyres. However, biomass-based community analysis showed similar community structure between Kuroshio and Equator, indicating that there are common dominant species between Equator and Kuroshio region. The Equator-Kuroshio region, including the station K1, was characterized by dominant

MOTUs belonging to specific family, such as Calanidae, Eucalanidae, Paracalanidae and Oithonidae. These families were main taxa of MOTUs with distribution peaks in Equator-Kuroshio region (Fig. VI-5, VI-6), suggested adaptation to high food availability. For example, MOTUs in the family Eucalanidae were clearly dominant under high chl-*a* condition with MOTU specific distribution peaks. In oligotrophic subtropical gyres, Eucalanidae were almost absent or restricted to relatively high chl-*a* area at the station N6-N8 (Fig. IV-2, IV-4). One of the dominant MOTUs (MOTU 49) at the station K1 was identified as *Eucalanus carifornicus* (100 % similarity) in the family Eucalanidae by blast search (data not shown). This species is considered to maintain population with ontogenetic vertical migration in high primary production areas in the western North Pacific (Shimode et al. 2012). In ETP, other Eucalanidae species of *Subeucalanus subtenuis* and *S. subcrassus* are dominant (Chen et al. 1986). There were also restricted distribution patterns to specific locations in Kuroshio or ETP in some taxa (e.g., Temoriidae and Rhincalanidae), and these MOTUs might be adapted to specific area such as coastal waters or the core of the Kuroshio Current.

The large numbers of MOTUs belonging to various taxa are hypothesized to be an adaptation to oligotrophic areas where food source is poor, as suggested by McGowan and Walker (1979). The various functions including body size, feeding habits or diel vertical distribution and migration (DVM) are important for niche partition. For example, the proportions of sequence reads were high in carnivorous copepods (e.g. family Candaciidae) and in copepods with DVM behavior (e.g. family Aetideidae and Metridinidae) in the subtropical gyres, and most of MOTUs in these taxa had distribution peaks in the subtropical gyres (Fig. IV-5, IV-6). According to Sameoto (1986), carnivorous copepods are more common in non-upwelling region with high diversity than in high productive upwelling region where copepod community is dominated by herbivores. An importance of copepods with DVM behavior for export flux is also known in oligotrophic subtropical gyre (Al-Mutairi and Landry 2001). Therefore, various functions are considered as an important factor for maintenance of complex food web structure in subtropical gyres. Some taxa clearly adapted to

the oligotrophic oceans, for example, the genus *Calocalanus* was especially diverse and increased proportion of biomass toward low chl-*a* areas in this study (Fig. IV-4, IV-9). In previous study in the Atlantic, the genus *Calocalanus* was abundant in the subtropical areas where low chl-*a* was observed (Longhurst 2007; Schnack-Schiel et al. 2010).

The community structures in NPSG and SPSG were highly similar each other both in biomass and non-biomass-based community analysis. Williamson and McGowan (2010) have also reported high similarity of species compositions between NPSG and SPSG. However, the community structure at the station S4 was relatively different from other samples in subtropical gyres. In SPSG, intrusion of water with low temperature and salinity was evident at the station S4 (Fig. IV-2), and some species might be adapted to low temperature or advection from the coastal and high-latitude areas. The station S4 is close to the distribution boundary between central species (distribution in subtropical gyre) and transition zone species (distribution in between subtropical gyre and subantarctic; McGowan 1971). There is a possibility of existence of transition zone species at the station S4. An additional sampling in the transition water would be helpful to understand community structure at the station S4.

The high MOTU richness (proxy of species richness) was the most correlated to SST, and high species richness was observed in subtropical gyres especially in NPSG where the highest water temperature was observed. Rombouts et al. (2010) also reported SST as the most explanatory environmental factor for species richness of copepods, and high species richness is predicted under high temperature, salinity and low chl-*a*, which are characteristics in subtropical gyres. A high temperature leads to a high evolution rate, and extinction rate is also low in the low latitude (reviewed by Dowle et al. 2013). A zooplankton metabolic rate is high, and longevity of copepods is short in the low latitude with a high temperature (Ikeda 1985, Mauchline 1998), providing a chance of rapid molecular evolution under high temperature.

In addition to possibility of high evolution rate, historically stable environments might facilitate the co-existence of various species in the subtropical gyres. The diversity index was the most correlated to chl-*a*, due to niche partitioning under low food availability as described

above. This high diversity index indicates co-existence of various species with few dominant species, therefore, large numbers of species are maintained with few dominant species in subtropical gyres. In addition, the emergence of the central water species with distribution peaks in subtropical gyres are hypothesized to be the earliest in planktonic taxa, and they are usually ancestral to cold water fauna (van der Spoel and Heyman 1983). The subtropical gyres are old, and high stability of water temperature is suggested during the glacial-interglacial cycles in subtropical gyres (McGowan and Walker 1979; Pflaumann et al. 2003). Therefore, it is hypothesized that copepods have evolved under high temperature, and both historically and contemporary stable environment have helped to maintain species diversity with low extinction rate under co-existing with various species in subtropical gyres.

The highest species richness in NPSG agreed with the previous studies of oceanic taxa including copepods in the Atlantic (Angel 1993; Rombouts et al. 2009). Williamson and McGowan (2010) also reported larger number of copepod species in NPSG than in SPSG. In this study, the high species richness in NPSG was explained by large numbers of shared MOTUs with other ocean regions. In addition to large number of shared MOTUs (184 MOTUs) between NPSG and SPSG, NPSG contained more MOTUs with Equator-Kuroshio distribution than did SPSG (Fig. IV-10). One possible explanation of high diversity in NPSG is the strong Kuroshio Current and its recirculation, which would contribute to transporting species with Equator-Kuroshio distribution to NPSG. In population genetics of copepods, haplotype are highly shared between Kuroshio and NPSG (Norton and Goetze 2013), suggesting connectivity between the NPSG and Kuroshio Current region. In addition, higher chl-*a* concentration in NPSG than in SPSG might be preferable for species in Equatorial-Kuroshio group favoring high food availability. In comparison between NPSG and SPSG, the proportion of MOTUs with distribution peaks in Equator-Kuroshio region (Group D in Fig. IV-6) was higher in NPSG (average: 3.73%) than in SPSG (average: 1.76%), suggesting Equator-Kuroshio species are more adapted to NPSG than SPSG. The similarity of community structure was also higher both in biomass and non-biomass-based analysis

between Equator-Kuroshio and NPSG than Equator-Kroshio and SPSG. It is known that a zooplankton diversity pattern shows peak in intermediate level of zooplankton biomass (Irigoien et al. 2004). SPSG is the most oligotrophic area in the world, which might lead to low biomass and diversity of copepods.

This is the first study to investigate large-scale community structure of copepods in the tropical and subtropical Pacific using metagenetic method, and the sampling areas were restricted to specific areas. In addition, community structures of copepods would be strongly related to vertical niche partitioning, and many species are known to be distributed with species-specific vertical depth (Longhurst 1985b). The closely related planktonic species including copepods are known to show different vertical distributions (Mackas et al. 1993; Rutherford et al. 1999; Fragopoulu et al. 2001). The method used in this study is easily adapted to other field-collected zooplankton samples. Additional samplings covering horizontal and vertical areas would reveal fine-scale community structure, diversity and distribution patterns of copepods, leading to further understandings of marine ecosystem and ocean provinces.

Chapter V. Phylogeography of oceanic copepod *Pleuromamma abdominalis* (Lubbock, 1856) in the tropical and subtropical Pacific

Introduction

Pleuromamma abdominalis (Lubbock, 1856) is a medium-sized oceanic copepod in the genus *Pleuromamma* in the family Metridinidae and is characterized by the presence of a 'pigment knob' on the right or left side of the prosome (Blades-Eckelbarger and Youngbluth 1988). *Pleuromamma abdominalis* has a worldwide distribution mainly in the tropical and subtropical oceans with high abundance in the upper mesopelagic zone, therefore, *P. abdominalis* serves as important prey to larger predatory crustaceans and fish (e.g., Foxton and Roe 1974; Lancraft et al. 1988; Bacha and Amara 2009). *Pleuromamma abdominalis* is also known to migrate to the surface layer during the night (Roe 1972). Due to its high abundance and extensive diel vertical migration (DVM), *P. abdominalis* acts as a transporter of organic matters to the mesopelagic zone with other vertically migrating zooplankton (Hays et al. 1997; Steinberg et al. 2000).

The intraspecific morphological variations of *P. abdominlis* have been reported, and there are three forms (Fornshell and Ferrari 2010): *P. abdominlis* forma *abdominalis* Lubbock, 1856 (forma *typica*), *P. abdominlis* forma *abyssalis* Giesbrecht, 1893, and *P. abdominlis* forma *edentata* Steuer, 1932. The forma *edentata* is only known in females and the forma *abyssalis* only in males. The forma *edentata* is discriminated from forma *typica* by the absence of a large curved attenuation on proximal segment of first antenna in female (Steuer 1932). In males, the forma *typica* show strongly asymmetrical urosome and simple lateral seta on the left caudal ramus compared with the forma *abyssalis* (Steuer 1932; Fornshell and Ferrari 2010). The difference of vertical and horizontal distribution is not obvious among each forma (Steuer 1932; Fornshell and Ferrari 2010). On the other hand, the different distribution patterns have been also reported between the forma *typica* and *edentata* in the eastern North

Pacific Ocean, suggesting different adaptation ability to the environment (Haury 1988).

The genetic variability has been observed in *P. abdominalis*, and the forma *typica* in the western subtropical Pacific is composed of at least two genetic groups with different body size (Chapter II). A high species complexity has been also reported in other species in the genus *Pleuromamma*. The phylogeography of *P. xiphias* showed four distinct genetic lineages in mitochondrial DNA in global area (Goetze 2011). A global study of *P. gracilis* and *P. piseki* revealed that both morphological species do not form monophyletic group respectively, and twelve genetically-divergent genetic lineages in mitochondrial DNA were observed in these two nominal species (Halbert et al. 2013). There is no phylogeographic study of *P. abdominalis*, and the relationship between morphological and genetic variations in this nominal species is still unclear. Therefore, the understanding of genetic variations in *P. abdominalis* might be important issue to understand complex genetic structure of both the genus *Pleuromamma* and oceanic planktonic species. In addition, an understanding of genetic structure of *P. abdominalis* would provide a significant information on how the current biogeography and diversity of copepods have been formed.

This chapter studied phylogeography of *P. abdominalis* in the tropical and subtropical Pacific and aimed to reveal a mechanism of biogeography and diversity of copepods in the Pacific. Both mitochondrial and nuclear genetic markers were used to detect genetic variation within *P. abdominalis*. Results of genetic analyses were compared with morphological variation in the adult female. The phylogeography of *P. abdominalis* was also discussed in relation to hydrographic structures in the tropical and subtropical Pacific.

Materials and Methods

Sampling

Sampling was performed at 25 stations in the tropical and subtropical Pacific (Fig. V-1 and Table V-1). All stations are categorized into three ocean regions: subtropical North Pacific,

equatorial Pacific and subtropical South Pacific. In this study, equatorial Pacific was defined as the area between 10°N–10°S, according to the approximate boundaries of the subtropical gyres (Longhurst 2007). A variety of sampling gear was used in this study. Sampling using NORPAC and IKMT nets was conducted during the night. Sampling using a VMPS net at 200–500 m depth and an ORI net were performed during the day. All samples were preserved in 99% ethanol and kept at 4ºC. Ethanol was replaced after 24 h from initial preservation.

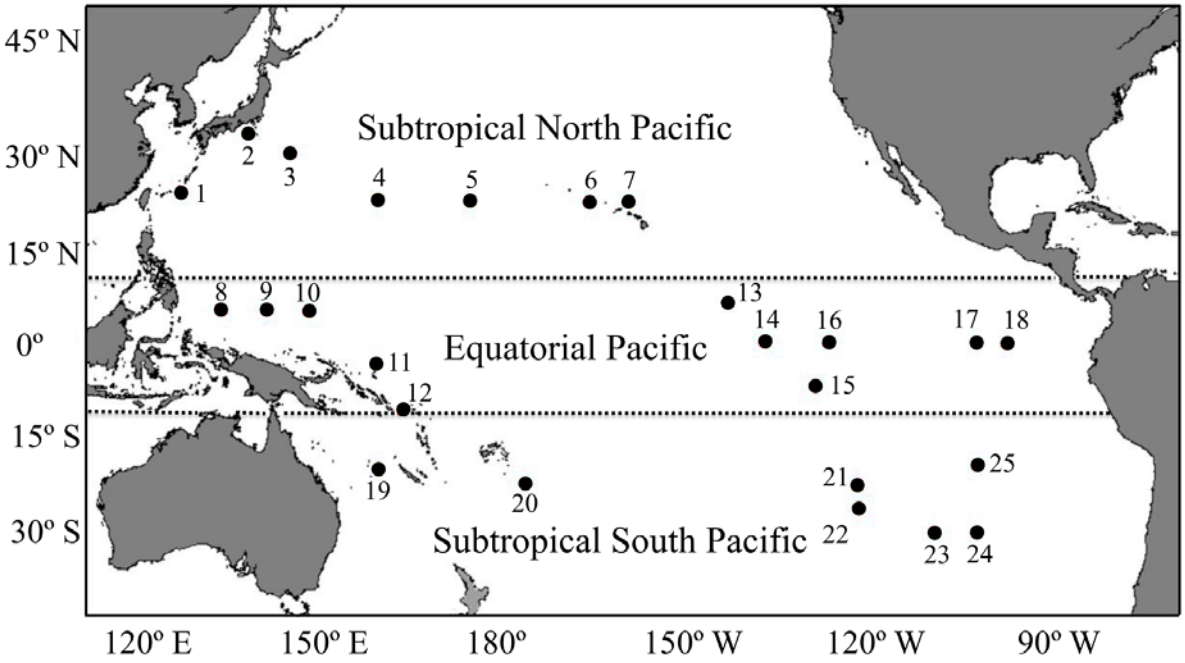


Fig. V-1 Sampling locations in this study.
Sampling stations 1−25 are categorized into three ocean regions: subtropical North Pacific, equatorial Pacific and subtropical South Pacific. A dotted line indicates the boundaries of ocean regions.

Table V-1 Sampling information including location, date, cruise, depth, sampling gear and sample size (N).
The number in brackets in the sample size column indicates the number of adult females of *Pleuromamma abdominalis*.

Area	Station	Latitude	Longitude	Date	Cruise	Depth	Sampling gear	N
Subtropical North	1	24° 56.3′N	127° 59.3′E	09/28/2010	KT-10-21	200-500 m	VMPS	27(25)
	2	33° 52.0′N	137° 44.0′E	05/05/2011	SY-11-05	200-500 m	VMPS	24(24)
	3	30° 00.2′N	144° 59.9′E	12/02/2011	KH-11-10	0-200 m	NORPAC	29(14)
	4	23° 00.5′N	159° 59.0′E	12/06/2011	KH-11-10	0-200 m	NORPAC	13(5)
	5	23° 00.2′N	174° 54.8′E	12/12/2011	KH-11-10	0-200 m	NORPAC	27(15)
	6	23° 00.6′N	165° 01.2′W	12/16/2011	KH-11-10	0-200 m	NORPAC	14(8)
	7	22° 46.8′N	158° 05.7′W	12/18/2011	KH-11-10	0-200 m	NORPAC	28(23)
Equatorial	8	05° 00.1′N	134° 00.1′E	01/13/2013	KH-13-01	0-180 m	IKMT	20(19)
	9	05° 00.1′N	140° 00.7′E	01/15/2013	KH-13-01	0-200 m	IKMT	17(17)
	10	05° 00.1′N	146° 00.2′E	01/19/2013	KH-13-01	0-224 m	IKMT	14(13)
	11	02° 00.3′S	160° 00.0′E	01/23/2013	KH-13-01	0-202 m	IKMT	18(18)
	12	10° 00.1′S	163° 00.2′W	24/01/2013	KH-13-01	0-193 m	IKMT	11(11)
	13	07° 00.2′N	141° 59.7′W	12/29/2011	KH-11-10	0-200 m	NORPAC	16(9)
	14	00° 00.0′N	134° 59.9′W	01/01/2012	KH-11-10	200-500 m	VMPS	15(4)
	15	07° 29.9′S	127° 30.4′W	01/03/2012	KH-11-10	0-200 m	NORPAC	11(11)
	16	00° 00.1′N	125° 00.1′W	02/09/2012	KH-12-01	200-500 m	VMPS	25(7)
	17	00° 00.4′N	100° 04.5′W	02/03/2012	KH-12-01	200-500 m	VMPS	16(7)
	18	00° 00.2′N	95° 30.0′W	02/02/2012	KH-12-01	0-200 m	VMPS	7(5)
Subtropical South	19	19° 59.7′S	159° 57.8′E	01/27/2013	KH-13-01	0-210 m	IKMT	18(18)
	20	21° 59.9′S	179° 59.4′W	02/13/2013	KH-13-02	0-220 m	IKMT	19(19)
	21	22° 59.7′S	119° 15.8′W	01/08/2012	KH-11-10	0-396 m	ORI	24(23)
	22	26° 29 9′S	120° 00.0′W	01/10/2012	KH-11-10	0-200 m	NORPAC	20(10)
	23	29° 59.5′S	107° 00.2′W	01/13/2012	KH-11-10	0-459 m	ORI	32(16)
	24	30° 00.0′S	100° 00.1′W	01/15/2012	KH-11-10	0-200 m	NORPAC	12(7)
	25	20° 00.0′S	100° 00.0′W	01/20/2012	KH-11-10	200-500 m	VMPS	14(9)

Morphological analysis

A total of 495 individuals of *Pleuromamma abdominalis* were sorted from bulk samples under the microscope. Adult females were mainly selected, however, males and immature stage were also used in cases of insufficient numbers of adult females. The genus *Pleuromamma* was mainly composed of *P. abdominalis*, *P. gracialis*, *P. xiphias* and *P. quadrungulata* in this area. According to Chihara and Murano (1997), *P. abdomimalis* is medium in body size and easily distinguished from the smaller-sized *P. gracialis*, *P. piseki*, and *P. borealis*, as well as from the larger-sized *P. xiphias*. *Pleuromamma quadrungulata* has two curved attenuations on the proximal segment of the first antenna that distinguish it from *P. abdominalis* with one or no curved attenuation. The location of the pigment knob is on the left side in *P. abdominalis* males and on the right side in *P. quadrungulata* males. Only adult females of *P. abdominalis* were used for morphological analyses, and the following morphological characters were recorded: (1) prosome length, (2) location of the pigment knob on the lateral prosome (side), and (3) the shape of the antennule spines on the proximal segment of the first antenna. The shape of the antennule spines on first antenna was classified into two categories: one large curved attenuation or an absence of a curved attenuation on the proximal segment of the first antenna, according to Fornshell and Ferrari (2010). In this study, I described the individuals with curved attenuation as “forma *typica*” and those without curved attenuation as “forma *edentata*” (Fig. V-2).

DNA extraction, PCR amplification and sequencing

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (QIAGEN) and kept at -20 °C. Mitochondrial cytochrome oxidase subunit I (COI) and nuclear large subunit ribosomal DNA (28S) were used to detect genetic variations within *P. abdominalis*. Another nuclear marker, internal transcribed spacer (ITS), was also used to support COI and 28S sequence variations. Polymerase chain reaction (PCR) was carried out in 15 μL reaction

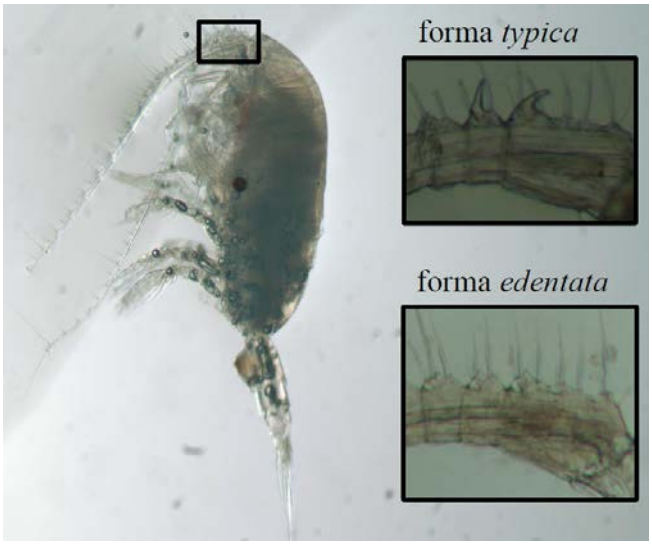


Fig. V-2 Digital images of *Pleuromamma abdominalis.*
The shapes of the antennule spines on the proximal segment of the first antenna are shown for forma *typica* and *edentate*.

mixtures containing 7.62 µL distilled water, 1.5 µL 10× buffer (Takara), 1.2 µL dNTPs (2.5 µM each), 1.8 µL of each primer (5 µM), 0.08 µL Z-Taq (Takara), and 1.00 µL template DNA. The primer pair L1384-COI and HCO PabR was used for COI, 28S D1F and 28S D3R for 28S, and 18S F1 and R63 28S for ITS amplifications (Table V-2). The PCR cycling was performed using a Model 9700 Thermal Cycler (Applied Biosystems). PCR cycling was denaturation at 94 °C for 2 min, followed by 35 cycles of 5 s denaturation at 94 °C, 5 s annealing at 45 °C (COI) or 55°C (28S and ITS), and 10 s extension at 72 °C. A final extension was set at 72 °C for 2 min. In case of poor amplification of COI, a second PCR was conducted using primer pair COI F1 and HCO2198ab. PCR products were purified with ExoSap-IT (GE Health BioScience), prepared for sequencing using the BigDye Terminator v3.1 (Applied Biosystems), and analyzed on a 3130 DNA Sequencer (Applied Biosystems). Sequencing reactions were performed according to the manufacturer's protocol. The primer COI F1 and HCO2198ab were used for COI sequencing. The same primer pairs for PCR were used for sequencing of 28S and ITS.

Table V-2 Primers used for PCR and sequencing in this study.

Gene	Primer name	Sequence (5'→3')	Reference
COI	L1384-COI	GGTCATGTAATCATAAAGATATTGG	Machida et al. (2004)
	HCO PabR	GTKGTAAAATATGCYCGTGTGTC	This study
	COI F1	AATGTTGTAGTAGACAGCGC	This study
	HCO2198ab	TAVACTTCAGGRTGVCCAAAAAAYCA	Modified from Folmer et al. (1994)
28S	28S D1F	GCGGAGGAAAAGAAAACAAC	Chapter III
	28S D3R	CGATTAGTCTTTCGCCCCT	Chapter III
ITS	18S F1	CGTCGCTACTACCGATTG	Chapter II
	R63 28S	CTTTTCCTCCGCTTATTAATATGC	Modified from Kiesling et al. (2002)

Data analysis

The obtained sequences were manually confirmed and edited by comparing aligned sequences for both forward and reverse strands using Geneious version 5.3.4 (Drummond et al. 2010). Multiple sequences of each genetic marker were aligned using MUSCLE (Edgar 2004). Phylogenetic analyses were performed for COI using Bayesian inference and maximum likelihood. The GTR+I+G model was selected as the best-fit nucleotide substitute model based on the Akaike Information Criterion in MrModeltest 2.3 (Nylander 2004). The Bayesian analysis was performed in MrBayes v.3.1.2 (Ronquist and Huelsenbeck 2003). The Markov Chain Monte Carlo chains were run for 2.0×10^7 generations, and trees were sampled every 1000 generations. The first 25% of the generated trees were eliminated as burn-in. Maximum likelihood analysis was performed using 1,000 bootstrap replications for nodal support in MEGA version 5.0 (Tamura et al. 2011). *Pleuromamma xiphias* was used as outgroup. The obtained COI clades were compared with sequence variations of 28S and ITS to detect reproductively isolated cryptic species.

Genetic distances between and within COI clades were calculated using Kimura's two-parameter (K2P) model (Kimura 1980) in ARLEQUIN 3.5 (Excoffier and Lischer 2010). Numbers of haplotype, haplotype diversity (h, Nei 1987), nucleotide diversity (π, Nei and Li 1979), and Tajima's D (Tajima 1989) at each sampling site were calculated in ARLEQUIN.

Results

Genetic variations

A total of 441 bp of COI sequence was obtained from 471 individuals of *P. abdominalis*. The COI phylogenetic tree showed 13 distinct COI clades that were highly supported by posterior probability and bootstrap values (Fig. V-3). Two genetically distinct lineages (*Pleuromamma abdominalis* 1 and 2) were observed as monophyletic groups of COI 1a-1b (1.00 posterior probability) and COI 2a–2k (0.98 posterior probability), though COI 2i–2j were not supported as a monophyletic group within *P. abdominalis* 2 in maximum likelihood analysis. There were 0.167 to 0.224 inter-clade K2P distances of COI sequences between *P. abdominalis* 1 and 2, with the average of 0.198 (Table V-3). K2P distance within COI clade was 0–0.015. The inter-clade K2P distance was 0.093 in *P. abdominalis* 1 and 0.058–0.187 (average: 0.138) in *P. abdominalis* 2.

28S sequences of 406 bp were obtained in 443/471 individuals. 28S sequences were clearly classified into two sequence types by three diagnostic nucleotide sites. 28S sequence types agreed with two distinct lineages of *P. abdominalis* 1 and 2 in COI phylogenetic tree (Fig. V-3). ITS sequences were obtained from randomly selected two individuals in each COI clade to investigate further cryptic speciation. There were no clear sequence differences in total 26 ITS sequences of 745 bp. An ambiguous site of equivalent peaks of two bases (e.g., C and T registered as Y according to the standard IUPAC code) was observed in 28S and ITS sequences. These ambiguous bases were not treated as diagnostic sites. The following analyses were conducted for *P. abdominalis* 1 and 2 respectively.

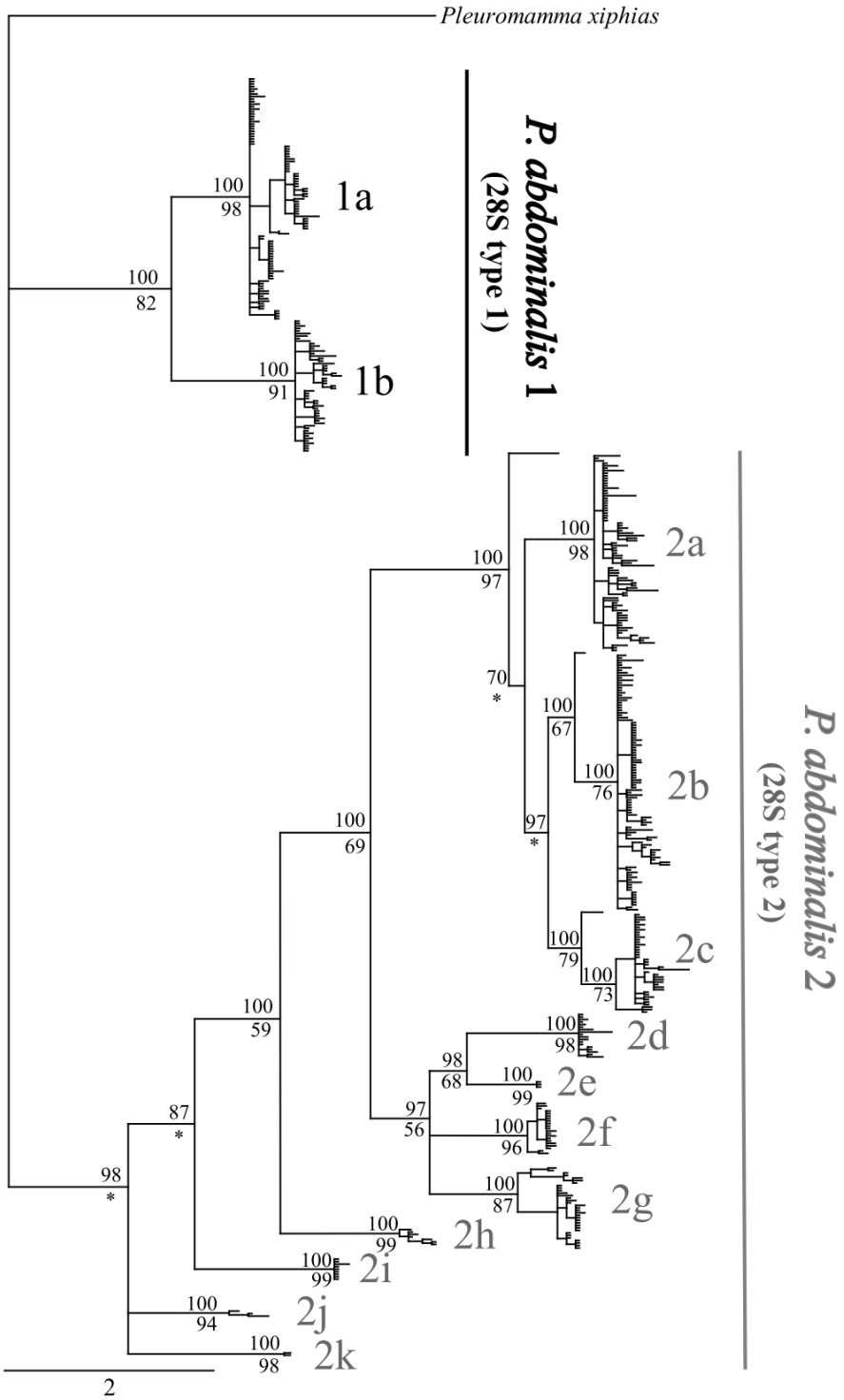


Fig. V-3 The 50% majority-rule consensus tree from Bayesian analysis of 471 COI sequences. Numbers at the nodes indicate the posterior probabilities for Bayesian analyses (upper) and bootstrap values for maximum likelihood (lower) greater than 50%. Asterisks indicate values less than 50%. Two major clades (*Pleuromamma abdominalis* 1 and 2) correspond to sequence types of 28S. Letters indicate COI clades for *P. abdominalis* 1 (1a-1b) and *P. abdominalis* 2 (2a-2k).

Table V-3 Average K2P genetic distance between and within COI clades.
Bold values indicate average K2P genetic distance within each COI clade.

	1a	1b	2a	2b	2c	2d	2e	2f	2g	2h	2i	2j	2k
COI 1a	**0.011**												
COI 1b	0.093	**0.010**											
COI 2a	0.204	0.214	**0.013**										
COI 2b	0.196	0.211	0.069	**0.009**									
COI 2c	0.207	0.212	0.059	0.058	**0.007**								
COI 2d	0.199	0.184	0.154	0.145	0.143	**0.005**							
COI 2e	0.198	0.196	0.127	0.125	0.119	0.075	**0.000**						
COI 2f	0.210	0.206	0.140	0.124	0.140	0.099	0.085	**0.005**					
COI 2g	0.218	0.224	0.134	0.127	0.140	0.113	0.089	0.089	**0.015**				
COI 2h	0.184	0.192	0.166	0.145	0.157	0.147	0.145	0.111	0.123	**0.006**			
COI 2i	0.175	0.193	0.179	0.163	0.161	0.148	0.141	0.152	0.158	0.130	**0.001**		
COI 2j	0.167	0.187	0.157	0.173	0.170	0.155	0.158	0.143	0.152	0.129	0.132	**0.011**	
COI 2k	0.187	0.201	0.169	0.187	0.172	0.185	0.183	0.175	0.182	0.148	0.137	0.108	**0.002**

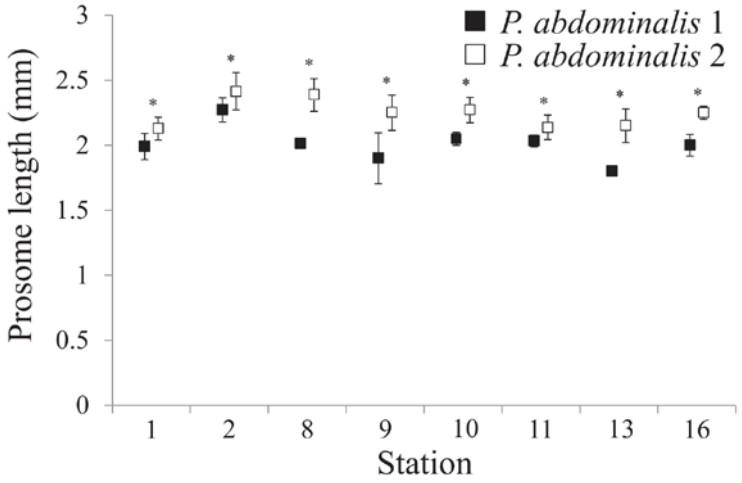


Fig. V-4 Adult female prosome length between *Pleuromamma abdominalis* 1 and 2.
Asterisk (*) indicates a significant difference ($p < 0.05$). Error bar represents the standard deviation. The prosome length is compared at stations with more than three females sampled for both *P. abdominalis* 1 and 2

Morphological variations of female Pleuromamma abdominalis

In 471 individuals of *P. abdominalis*, 337 individuals were adult female and used for morphological analysis. Prosome length was significantly larger in *P. abdominalis* 2 than in *P. abdominalis* 1 ($p < 0.05$), at stations where more than three female individuals were obtained for both cryptic species (Fig. V-4, other sites not examined).

Table V-4. Morphological characters on the first antenna in the adult female.
Only adult female is used (N=337), and the number of individuals with curved attenuation (forma *typica*) and no curved attenuation (forma *edentata*) is listed for each COI clade in each ocean region. ND: no data.

	Subtropical North		Equatorial		Subtropical South	
	typica	*edentata*	*typica*	*Edentata*	*typica*	*edentata*
COI 1a	10	0	1	50	ND	ND
COI 1b	33	2	ND	ND	3	0
COI 2a	ND	ND	0	2	0	54
COI 2b	54	0	2	0	16	1
COI 2c	4	0	ND	ND	27	1
COI 2d	ND	ND	9	1	ND	ND
COI 2e	2	0	ND	ND	ND	ND
COI 2f	ND	ND	15	1	ND	ND
COI 2g	2	0	31	0	ND	ND
COI 2h	ND	ND	0	4	ND	ND
COI 2i	4	0	1	0	ND	ND
COI 2j	2	0	0	1	ND	ND
COI 2k	ND	ND	2	0	ND	ND

Table V-5. Location of the pigment knob on adult females (side).
The number of specimens examined for pigment knob location (left/right) is listed for each COI clade in each ocean region. ND: no data.

	Subtropical North		Equatorial		Subtropical South	
	Right	Left	Right	Left	Right	Left
COI 1a	4	6	21	30	ND	ND
COI 1b	16	19	ND	ND	0	3
COI 2a	ND	ND	1	1	45	9
COI 2b	43	11	0	2	14	3
COI 2c	4	0	ND	ND	26	2
COI 2d	ND	ND	8	2	ND	ND
COI 2e	1	1	ND	ND	ND	ND
COI 2f	ND	ND	11	5	ND	ND
COI 2g	1	1	22	9	ND	ND
COI 2h	ND	ND	1	3	ND	ND
COI 2i	1	3	0	1	ND	ND
COI 2j	2	0	1	0	ND	ND
COI 2k	ND	ND	0	2	ND	ND

Both forma *typica* (curved attenuation) and *edentata* (no curved attenuation) were observed in *P. abdominalis* 1 and 2 (Table V-4). In *P. abdominalis* 1, the occurrence of individuals with a specific shape of the antennule spines on first antenna was corresponded with ocean regions. In subtropical North Pacific, there were 43 individuals of forma *typica* and only 2 individuals of forma *edentata*. In equatorial Pacific, there were 50 individuals of forma *typica* and only one individual of forma *edentata* in analyzed samples. In *P. abdominalis* 2, all individuals were forma *typica* in subtropical North Pacific. Both forma *typica* and *edentata* were observed in equatorial and subtropical South Pacific. In Equatorial Pacific, COI 2a, 2h and 2j were forma *edentata*, whereas clade 2d, 2f and 2g, 2i and 2k were mostly forma *typica*. In subtropical South Pacific, COI 2b and 2c were forma *typica* except for one individual in each clade. All individuals in COI 2a were forma *edentata*.

The pigment knob was observed on both sides of the lateral prosome in *P. abdominalis* 1 and 2 (Table V-5). In all COI clades of *P. abdominalis* 1, there were more individuals with the

pigment knob on left side than those on the right side. On the other hand, the numbers of individuals with the pigment knob on the right side were greater than those on left side in most COI clades of *P. abdominalis* 2 (except for COI 2h, 2j and 2k, which had small sample sizes). However, the side of the pigment knob on the prosome showed no clear relationship with ocean regions or COI clades in both *P. abdominalis* 1 and 2.

Distribution of Pleuromamma abdominalis 1 and 2

Pleuromamma abdominalis 1 exhibited high proportions along the equator and Kuroshio Current (Fig. V-5). The proportions of *P. abdominalis* 1 decreased toward the center of North Pacific subtropical gyre. In subtropical South Pacific, *P. abdominalis* 1 was absent in the eastern part (station 21–25), and only small proportion was observed in the western part (station 19–20). *Pleuromamma abdominalis* 2 were widely distributed, however, they were absent in the eastern equatorial Pacific (station 17 and 18). The proportion of *P. abdominalis* 2 was high in the center of North and South Pacific subtropical gyre.

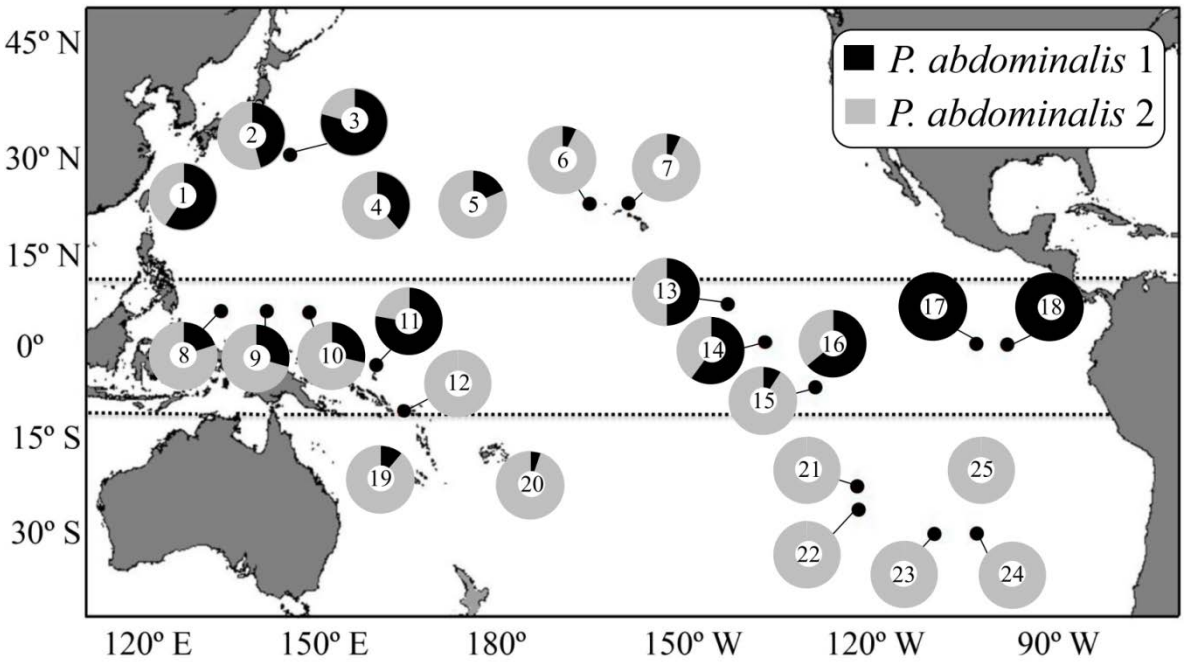


Fig. V-5 Relative proportions of *Pleuromamam abdominalis* 1 and 2 at 25 sampling sites. The number inside each circle indicates the station number (see Table V-1).

Distributions of COI clades

All individuals of *P. abdominalis* 1 were COI 1a in the equatorial Pacific (Fig. V-6). Both COI 1a and 1b were observed in subtropical North Pacific, however, clade 1b was dominant.

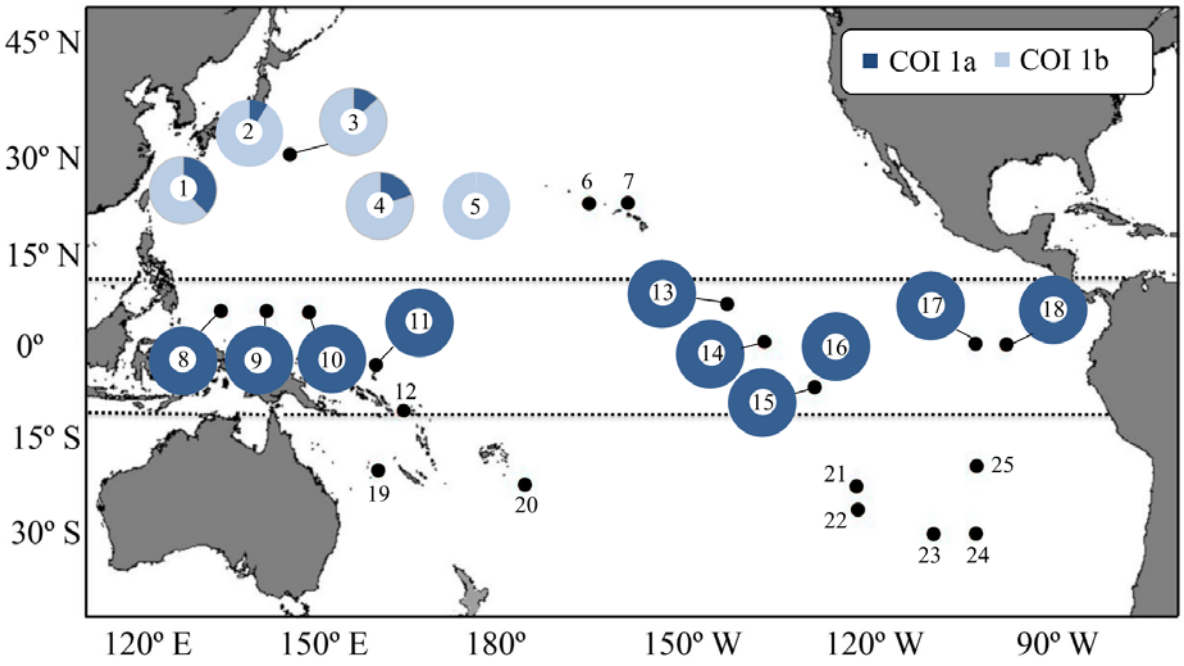


Fig.V- 6 Distribution of COI clades in *Pleuromamma abdominalis* 1.
Proportion of clades COI 1a and COI 1b is plotted in each station where, at least 4 individuals of *P. abdominalis* 1 are plotted. The number inside each circle indicates the station number (see Table V-1).

In contrast to *P. abdominalis* 1, complex genetic structure was observed in *P. abdominalis* 2. A different composition of COI clade was observed among subtropical North Pacific, equatorial Pacific and subtropical South Pacific (Fig. V-7). The basal distinct clades (COI 2i–2k) were mainly distributed in the west part of subtropical North Pacific and equatorial Pacific. COI 2b was dominant clade in subtropical and accounted for majorities of COI clades in the station 5–7. Various COI clades were observed in western North Pacific especially at the station 1 with six COI clades. In subtropical South Pacific, the composition of COI 2a was high in the station 21–23 and 25. The station 19–20 and 24 were mainly composed of COI 2b and 2c. The COI 2a was domestic in subtropical South Pacific, whereas COI 2a and 2c were shared with other areas. In equatorial Pacific, compositions of COI clade were different between west part (station 8–12) and east part (station 13–15). The COI 2d and 2h were mainly found in the east part and the clade 2g in west part.

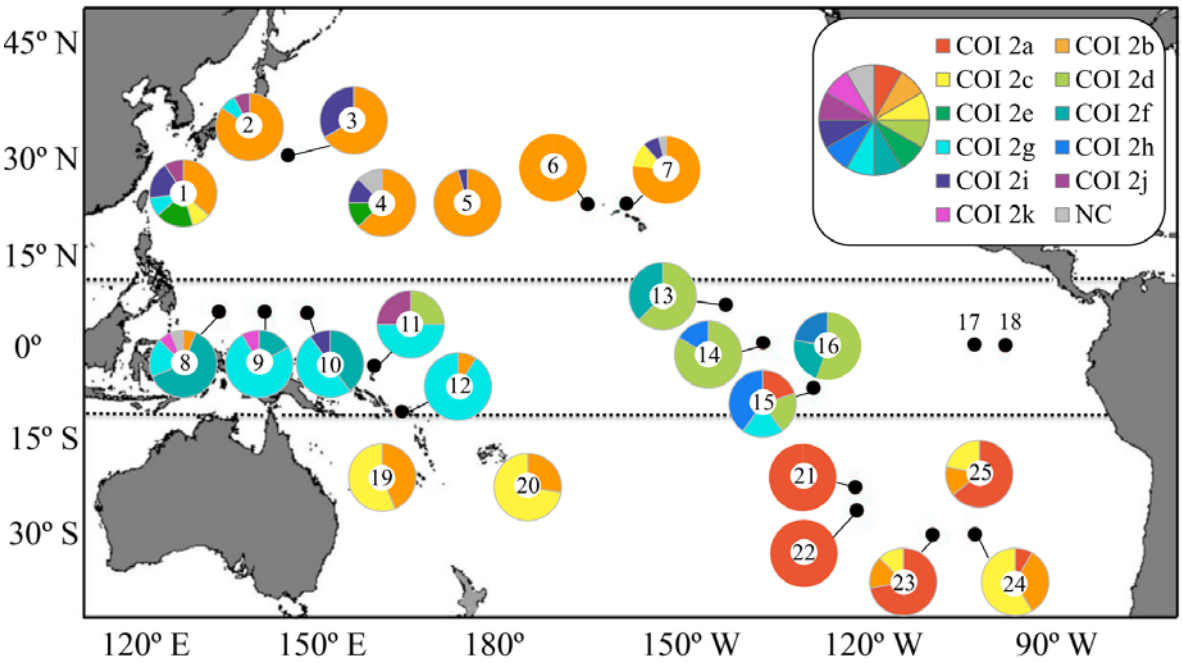


Fig. V-7 Distribution of COI clades in *Pleuromamma abdominalis* 2.
Proportion of COI clades (2a–2k) is plotted in each station where at least 4 individuals of *P. abdominalis* 2 are plotted. The number inside each circle indicates the station number (see Table V-1).

Genetic diversity of Pleuromamma abdominalis 1 and 2

The high genetic diversity was observed in *P. abdominalis*, and haplotype diversity was 0.50–1.00 in *P. abdominalis* 1 and 0.83–1.00 in *P. abdominalis* 2. The nucleotide diversity of all COI sequences was higher in *P. abdominalis* 2 (0.082) than *P. abdominalis* 1 (0.047; Table V-6). The highest nucleotide diversity was observed at the station 1 (western subtropical North Pacific), with 0.045 in *P. abdominalis* 1 and 0.105 in *P. abdominalis* 2. In *P. abdominalis* 1, nucleotide diversity was relatively high in North Pacific (0.027–0.045) except for the station 5 (0.01), compared with that in equatorial Pacific (0.002–0.013). In *P. abdominalis* 2, nucleotide diversity was low (0.011–0.037) in subtropical South Pacific. The value of nucleotide diversity tended to decrease from the west to the east in subtropical North Pacific (0.008–0.105). The nucleotide diversity was also high in equatorial Pacific (0.03–0.10). The significant value of Tajima's D was only observed at the station 2 in *P. abdominalis* 1 and at the station 5 and 12 in *P. abdominalis* 2, suggesting neutral evolution of COI.

Discussion

Cryptic species of Pleuromamma abdominalis

The phylogeography of *Pleuromamma abdominalis* revealed high genetic variations in this nominal species in the tropical and subtropical Pacific, and at least two cryptic species were detected. Two cryptic species co-occur in a same station, suggesting sympatric reproductive isolation. The K2P genetic distance of COI sequences between *P. abdominalis* 1 and 2 was large (0.167–0.224) and equivalent to inter-species variation of other copepods (e.g., Bucklin et al. 2003). Two cryptic species of *P. abdominalis* were corresponded with large and small forms of adult female, and *P. abdominalis* 1 were smaller than *P. abdominalis* 2 in all stations where at least three individuals of *P. abdominalis* 1 and 2 were collected respectively. The body size varied among sampling station, and overlaps of prosome length were observed between *P. abdominalis* 1 and 2, if all adult females were compared.

Table V-6 Summary of the number of individuals (N), number of haplotypes (H), haplotype diversity (h), nucleotide diversity (π) and Tajima's D (D) for *Pleuromamma abdominalis* 1 (left) and 2 (right).
Bold values of Tajima's D test represent significant p-value (<0.05). ND: no data.

		Pleuromamma abdominalis 1 / *Pleuromamma abdominalis* 2																	
Area	Station	N			H			Hd			π			D					
Subtropical North	1	16	/	11	13	/	9	0.97	/	0.96	0.045	/	0.105	1.39	/	1.10			
	2	11	/	13	11	/	10	1.00	/	0.96	0.027	/	0.046	**-1.55**	/	-1.30			
	3	23	/	6	18	/	5	0.98	/	0.93	0.028	/	0.080	-0.50	/	1.28			
	4	5	/	8	4	/	8	0.90	/	1.00	0.038	/	0.071	-0.94	/	-0.44			
	5	5	/	22	5	/	13	1.00	/	0.88	0.010	/	0.018	-0.93	/	**-2.33**			
	6	1	/	13	1	/	7	ND	/	0.83	ND	/	0.008	ND	/	-0.61			
	7	2	/	26	2	/	17	ND	/	0.93	ND	/	0.038	ND	/	-1.02			
Equator	8	4	/	16	2	/	13	0.50	/	0.97	0.011	/	0.068	-0.83	/	-0.37			
	9	5	/	12	4	/	10	0.90	/	0.97	0.008	/	0.057	-0.44	/	-0.70			
	10	4	/	10	2	/	8	0.50	/	0.96	0.002	/	0.065	-0.71	/	0.32			
	11	14	/	4	8	/	4	0.89	/	1.00	0.013	/	0.103	0.74	/	0.14			
	12	0	/	11	ND	/	7	ND	/	0.87	ND	/	0.026	ND	/	**-1.86**			
	13	8	/	8	3	/	5	0.68	/	0.86	0.005	/	0.049	-0.67	/	1.93			
	14	9	/	6	8	/	5	0.97	/	0.93	0.010	/	0.050	-1.41	/	-1.24			
	15	1	/	10	1	/	10	ND	/	1.00	ND	/	0.100	ND	/	1.51			
	16	16	/	9	10	/	9	0.90	/	1.00	0.009	/	0.074	-1.00	/	1.15			
	17	16	/	0	9	/	ND	0.91	/	ND	0.005	/	ND	-1.26	/	ND			
	18	7	/	0	6	/	ND	0.95	/	ND	0.005	/	ND	-0.73	/	ND			
Subtropical South	19	2	/	16	2	/	13	ND	/	0.97	ND	/	0.031	ND	/	1.11			
	20	1	/	18	1	/	10	ND	/	0.85	ND	/	0.024	ND	/	0.80			
	21	0	/	24	ND	/	15	ND	/	0.89	ND	/	0.011	ND	/	-1.27			
	22	0	/	20	ND	/	18	ND	/	0.98	ND	/	0.014	ND	/	-1.03			
	23	0	/	32	ND	/	25	ND	/	0.97	ND	/	0.034	ND	/	-0.10			
	24	0	/	12	ND	/	11	ND	/	0.98	ND	/	0.035	ND	/	0.09			
	25	0	/	14	ND	/	11	ND	/	0.96	ND	/	0.037	ND	/	0.39			

The difference of body size corresponding to genetic variation has been reported in other oceanic copepods of *Nannocalanus minor* (Bucklin et al. 1996) and *Neocalanus flemingeri* (Machida and Tsuda 2010). Both *P. abdominalis* 1 and 2 contained the forma *typica* and *edentata*, and variations of first antenna were mainly explained by differences of ocean regions and COI clades. For example, COI 1a in *P. abdominalis* 1 was forma *typica* in subtropical North Pacific but forma *edentata* in equatorial Pacific (Table V-4), indicating a possibility of current geological isolations in *P. abdominalis* 1. However in *P. abdominalis* 2, COI 2a was forma *edentata* in *P. abdominalis* 2 in subtropical South Pacific, whereas COI 2b and COI 2c was mostly forma *typica*. The side of pigment knob of adult females showed no clear relationship with genetic variations and locality, agreeing with the result in *P. xiphias* (Goetze 2011). Only adult females were analyzed morphologically in this study, and morphological variations in males are still unknown. In addition, population genetics after determining species boundaries among COI clades would be necessary to understand morphological differences.

The main distribution of *P. abdominlais* 1 was along the equator and the Kuroshio Current. They were almost absent in South Pacific subtropical gyre. On the other hand, *P. abdominalis* 2 were widely distributed including center of the subtropical gyres and equatorial Pacific. However, they were absent in the eastern equatorial Pacific where there is the equatorial cold tongue with relatively cool water from the Peru Current and by equatorial upwelling along the equator (Wyrtki 1966, 1981). According to McGowan (1971, 1974), distribution patterns of *P. abdominalis* 1 and 2 are common in plankton taxa in the Pacific. The similar distribution pattern was observed between closely related species of *Clausocalanus farrani* and *C. minor* (Frost 1969) and *Pontellina morii* and *P. plumata* (Fleminger and Hulsemann 1974). In each genus, there are also closely related species of *C. jobei* and *P. sobrina* which are restricted to the eastern equatorial Pacific, where *P. abdominalis* 2 were absent. A different distribution pattern of two cryptic species suggests that *P. abdominalis* 1 are adapted to areas with low temperature and high food availability compared with *P. abdominalis* 2

which are dominant in the subtropical gyres with high temperature and low food availability.

Pleuromamma abdominalis is considered to follow a general evolutionary pattern of planktonic taxa in the Pacific. The central water fauna with distribution peaks in subtropical gyres were ancestral, then the equatorial fauna with distribution peaks in Equator-Kuroshio region were subsequently emerged (van der Spoel and Heyman 1983). According to this pattern, *P. abdominalis* 2 is hypothesized to be ancestral due to distribution pattern of central water fauna. *Pleuromamma abdominalis* 1 and 2 were estimated to share most recent common ancestors approximately during the middle Miocene (11.6─16.0 million years ago), if a common crustacean COI mutation rate of 1.4% per one million years is applied (Knowlton and Weigt 1998). A global-scale cooling event began in the middle Miocene, intensifying latitudinal gradient of temperature (Zachos et al. 2001). An enhancement of upwelling and dominance of diatom is proposed in the central equatorial Pacific during the cooling period (Barron 1985), and the historical increases of diatom abundance are proposed to promote diversification of various oceanic taxa (Suto et al. 2012). Therefore, it is hypothesized that speciation into *P. abdominalis* 1 might have occurred in concordance with adaption to low temperature and high food availability during the cooling event. A further study of morphology and ecology of *P. abdominalis* are necessary to understand different adaptation abilities within *P. abdominalis*.

Genetic structure of Pleuromamma abdominalis 1

The lower genetic diversity was observed in *P. abdominalis* 1 than in *P. abdominalis* 2, supporting that *P. abdominalis* 1 was newly emerged species. The distribution of *P. abdominalis* 1 along strong ocean currents in the Kuroshio and equatorial regions might also contribute to maintaining high gene flow and low genetic diversity. In *P. abdominalis* 1, COI clade in equatorial Pacific (COI 1a) was also observed in subtropical North Pacific, however, COI 1b was restricted to the subtropical North Pacific. This genetic structure indicates population isolation in the past, and recent secondary contact

with equatorial populations might have occurred in the subtropical North Pacific. There is the strong western boundary current of the Kuroshio Current in the subtropical North Pacific, and the Kuroshio Current is considered to transport equatorial population to the subtropical North Pacific. The historical isolations and secondary contacts of divergent genetic lineages are common in the coastal waters in the western subtropical North Pacific, and the Kuroshio Current contributes to gene flows and secondary contacts of several genetic clades (Shen et al. 2011; Hirase et al. 2012). Though a mechanism of past population interactions are not obvious in the open ocean, the part of Equator-Kuroshio populations of *P. abdominalis* 1 might be the result of past transportation of equatorial species to the Kuroshio Current region. A fossil evidence of planktonic foraminifera also showed several historical intrusions of warm water fauna in the Kuroshio Current region due to past climate changes (Tsuchi 1997). The individuals of COI 1a were forma *typica* in subtropical North Pacific and forma *edentata* in equatorial Pacific (Table V-4), suggesting limited contemporary gene flows between these two regions. The Equator-Kuroshio distribution is common in planktonic species as shown in chapter IV. This distribution pattern is considered to have been formed through the same mechanism in *P. abdominalis*

Genetic structure of Pleuromamma abdominalis 2

The genetic structure within *P. abdominalis* 2 is more complex than that of *P. abdominalis* 1 due to high genetic variations which suggest a long evolutionary history. In addition, wide distribution *P. abdominalis* 2 including historically and spatially stable subtropical gyres might lead to low dispersal and high genetic variability among and within ocean regions. Distinct compositions of COI clades were observed mainly among the subtropical North Pacific, equatorial Pacific and subtropical South Pacific. The previous population studies of oceanic copepods have reported genetic break between North and South hemisphere (Goetze 2005, 2011), and equatorial regions are suggested as a strong dispersal barrier for copepods with distribution peaks in subtropical gyres (Norton and Goetze 2013).

In addition to north and south subtropical gyre systems, equatorial Pacific was distinct subdivisions of *P. abdominalis* and worked as genetic barrier between the North and South Pacific subtropical gyres. In addition, *P. abdominalis* 2 also showed different composition of COI between west and east part in each ocean region. Therefore, COI clades of *P. abdominalis* are largely related to locality among and within ocean regions.

The co-existence of genetically divergent COI clades was common in *P. abdominalis* 2, indicating several historical isolations and secondary contacts. The various distinct COI clades of *P. abdominalis* 2 with high nucleotide diversity were found in the subtropical North Pacific (0.058-0.163 K2P distance) and Equatorial Pacific (0.069-0.189 K2P distance), whereas the low nucleotide diversity was observed with relatively newly divergent COI in the subtropical South Pacific (0.058-0.069 K2P distance). The main COI clade in equatorial Pacific (COI 2g) was observed in subtropical North Pacific, whereas equatorial COI clades were not found in the subtropical South Pacific. In the western equatorial Pacific, the New Guinea Coastal Undercurrent flows northward from southern hemisphere to northern hemisphere, in addition to the strong Kuroshio Current in the North Pacific (Fine et al. 1994; Qu and Lindstrom 2002). These ocean currents possibly lead to higher population interactions between subtropical North Pacific and equatorial Pacific, as suggested in *P. abdominalis* 1. On the other hand, there may be limited contemporary and historical gene flow between subtropical South Pacific and equatorial Pacific. In particular, the highest nucleotide diversity and various COI clades were found in station 1 (western part of the Kuroshio Current), supporting the introgression and accumulation of COI clades to the subtropical North Pacific especially to the Kuroshio Current region. As discussed in *P. abdominalis* 1, it is hypothesized that historical climate changes, such as glacial and inter-glacial cycles, have caused hydrographic changes which led to isolation, introgression and admixture of populations, forming current complex genetic structure of *P. abdominalis* in the Pacific. For the current population connectivity, population genetics and physical modeling data would be helpful for the understanding of biogeography of *P.*

abdominalis.

High genetic variation with a possibility of further cryptic species

The high COI genetic divergence in *P. abdominalis* suggests two possibilities: (1) divergent COI clades are result of geological isolations and capable of interbreeding between clades, and (2) some of COI clades are reproductively isolated species. COI variations within both *P. abdominalis* 1 and 2 are higher than those between some sibling species of crustaceans (Bucklin et al. 2003; Costa et al. 2007), suggesting the existence of an additional cryptic species. A value of nucleotide diversity in a single station is also high in *P. abdominalis* in comparison with other population genetic studies of zooplankton taxa (Peijnenburg and Goetze 2013). Each COI clade might be adapted to local environment, for example, COI 2b in *P. abdominalis* 2 was frequently observed in the subtropical North and South Pacific but not in Equatorial Pacific. If population in COI 2b is adapted to oligotrophic subtropical gyre, the environment in Equatorial Pacific might act as a barrier of population connectivity. As different salinity preferences have been reported among genetic lineages of a neritic copepod *Acartia tonsa* (Chen and Hare 2011), each COI clade in *P. abdominalis* may have different adaptation to environments, affecting distribution patterns of COI clades.

A high species complexity is also suggested with four genetic clades in *P. xiphias* (Goetze 2011) and twelve genetic clades in *P. gracilis* and *P. piseki* (Halbert et al. 2013). Therefore, high genetic diversity is characteristics of the genus *Pleuromamma*. A low genetic connectivity due to population retention through DVM behavior is suggested in oceanic copepod of *Metridia lucence* (Blanco-Bercial et al. 2010; Bucklin et al. 2011). A high genetic diversity in the genus *Pleuromamma* is probably related to DVM behavior. In this study of *P. abdominalis*, no clear genetic variability was observed in 28S and ITS sequences between COI clades within each cryptic species. As nuclear microsatellite marker revealed reproductive isolations of mesopelagic copepods between North and south subtropical gyres (Goetze, personal communication), multi-locus analysis are necessary for

further understandings of species complexity and diversity of *P. abdominalis*. If there are more cryptic species in *P. abdominalis*, each distribution pattern would provide a new insight on biogeography of copepods in the open oceans.

Pleuromamma abdominalis is distributed in global oceans, however, sampling stations were restricted to the tropical and subtropical Pacific in this study. The basal COI clades (COI 2h-k) of *P. abdominalis* 2 were distinct from other clades and not supported as a monophyletic group of *P. abdominalis* 2 in the maximum likelihood analysis (Fig. V-3). They were mainly restricted to the west part of the Pacific. This area is adjacent to the coral triangle and the Indian Ocean, and these clades might be related to populations in the Indian Ocean. In the eastern Indian Ocean and western Pacific, a specific distinct genetic clade has been reported in other oceanic copepods of *Pleuromamma xiphias* (Goetze 2011) and *Pareucalanus* sp. (Goetze and Ohman 2010). In addition to horizontal distribution, a vertical niche partitioning is also important factor to separate sibling species in planktonic taxa (Mackas et al. 1993; Fragopoulu et al. 2001). A difference of vertical distribution by genetic types has been also reported in planktonic foraminifera (Weiner et al. 2012). *Pleuromamma abdominalis* perform extensive DVM, and the vertical distribution of *P. abdominalis* is considered to be affected by environmental factors such as thermocline depth, and a different vertical distribution is also expected among cryptic species. Therefore, more sampling efforts covering horizontal and vertical areas are necessary for further understanding of speciation, distribution and genetic structure of *P. abdominalis.*

Chapter VI. General discussion

Methodology in this paper

DNA-based molecular approaches worked as effective tools to reveal the diversity and biogeography of planktonic copepods in the tropical and subtropical Pacific. The different molecular markers of nuclear DNA (28S and ITS) and mitochondrial DNA (COI) were used in this study. In the analysis of biological diversity using molecular techniques, choice of molecular marker is strongly related to diversity levels. Chapter II-IV aimed to focus on 'species-level' distribution and biogeography of copepods, whereas 'genetic-level' diversity was discussed within a nominal species in chapter V.

Though there are many concepts and definitions of species (de Queiroz 2007), the classical concept of species is morphological species, which is morphologically distinguished. The majority of copepod studies have been conducted based on species concept of morphological species. The common species-level nuclear markers of ITS and 28S are relatively slowly evolving gene compared with mitochondrial DNA sequences, however, sequence differences have been observed between morphological species in marine environments (Chapter II; Kiesling et al. 2002; Goetze 2003; Llinas 2008). These markers also could detect cryptic species in a single morphological species, which are not fully identified by morphological characters (Chapter II; Goetze 2003, 2005; Chen and Hare 2008; Ueda et al. 2011; Miyamoto et al. 2012). Though there are existences of some cryptic species, molecular analysis using ITS and 28S show good concordance to morphological analysis due to the sophisticate works by taxonomic experts of copepods (e.g., Bucklin et al. 2003; Blanco-Bercial et al. 2011b). Therefore, ITS and 28S work as a species-level molecular marker, and they are easily applied to various species of copepods.

The MOTU-based metagenetic analysis using 28S-D2 region also reasonably agreed with morphological analysis. The 28S-D2 was useful molecular marker for metagenetic analysis because this marker showed high sequence variations, high PCR success rate, phylogenetic information and

appropriate sequence length for deep-sequencing. Though some closely related species might be lumped in a single MOTU due to lack of enough sequence variations, the characteristics of community structure were successfully reflected in metagenetic data in comparison with morphological analysis. Therefore, chapter IV using metagenetic analysis offered 'species diversity' in the tropical and subtropical Pacific, which is almost comparable to classical studies based on morphological identification. This metagenetic method could be easily applied to other field-collected samples and might contribute to reveal global-scale community structure of copepods. In particular, there are advantages to use this metagenetic method for samples containing many immature individuals, and temporally and spatially fine-scale samples also could be also analyzed easily. In addition, once metagenetic data would be accumulated, we can easily compare data in various places and times. It is also possible to document current biogeography of copepods and detect changes of distribution in the future. The limitation of method, such as biases and taxonomic classification, would be improved in the future with a development of technology of sequencer and bioinformatics method. Therefore, the metagenetic method would be a powerful tool for study of diversity and biogeography of copepods.

In chapter V, COI revealed 'genetic diversity' in a nominal species of *Pleuromamma abdominalis* in which species complexity was suggested in chapter II. A mutation rate is much higher in mitochondrial DNA than in nuclear DNA (Brown et al. 1979), and COI is frequently used for intra-species genetic analysis in planktonic taxa (Peijnenburg and Goetze 2013). At least two cryptic species were detected in *P. abdominalis* based on COI and 28S sequence variations, providing a good example of comparing genetic structures between closely related species in the tropical and subtropical Pacific. The high COI genetic divergences were also observed within two cryptic species of *P. abdominalis*. In addition to common species-level nuclear maker of 28S and ITS at the present, an additional nuclear marker (e.g., microsatellite marker) would help further understanding of species complexity of planktonic taxa, as discussed in chapter V. The study of genetic diversity in a

nominal species, such as *P. abdominalis*, is an important to understand current species-level diversity and biogeography of copepods. Therefore, results of phylogeography of *P. abdominalis* in chapter V would provide significant insights on metanetenic data in chapter IV, as discussed in the following parts.

Copepod biogeography and diversity in the tropical and subtropical Pacific

The metagenetic analysis revealed copepod community structures, which are based on distribution pattern of MOTUs. The distribution pattern of MOTU was based on relative proportions of sequence reads (proxy of biomass), indicating percentage of biomass in copepod community. The copepod community at the shoreward area of the Kuroshio Current was unique, and this area was considered to be affected by coastal waters mass and environment. Other oceanic stations were clustered into two major groups: Equator-Kuroshio and subtropical gyre group. These groups agreed with major distribution patterns of planktonic taxa in the Pacific (McGowan 1971). Two cryptic species in *P. abdominalis* also followed these distribution patterns. *Pleuromamma abdominalis* 1 was suggested to be adapted to relatively low temperature and high food availability in Equator-Kuroshio region, whereas dominance of *P. abdominalis* 2 were observed in oligotrophic subtropical gyres. Therefore, genetic analysis within *P. abdominalis* is a good example to understand how major patterns of current biogeography of copepods have been formed between closely related species.

As described in chapter V, it is hypothesized that speciation of *P. abdominalis* might have occurred in concordance with adaption to low temperature and high food availability during the cooling event after the middle Miocene (11.6─16.0 million years ago). It is proposed that diversification occurred in many marine taxa after the middle Miocene, in relation to eutrophication indicated by increase of diatom (Suto et al. 2012). Other sibling species of pelagic copepods *Eucalanus hyalinus* and *E. spinifer* are also suggested to share the most recent common ancestors in the Miocene (Goetze and Bradford-Grieve 2005). Therefore, many oceanic species containing

copepods are considered to have experienced same divergence procedure as *P. abdominalis*, and speciation between closely related species is one of the explanations of current major biogeography of Equator-Kuroshio and subtropical gyre distributions.

It is difficult to understand the mechanism of biogeography from the morphological or species-level molecular analysis. On the other hand, genetic-level analysis using COI gene provided some insight on how the current biogeography of copepods have been formed. The genetic structure of *P. abdominalis* indicated limited gene flows at the present among subtropical North Pacific, equatorial Pacific and subtropical South Pacific. The connectivity is especially low in the subtropical South Pacific with other ocean areas, whereas relatively high historical population interactions were suggested in the subtropical North Pacific and Equatorial Pacific. This historical interaction is evident especially in the Kuroshio Current region, suggesting that the Kuroshio Current has transported equatorial species to the subtropical North Pacific. Therefore, the Equator-Kuroshio distribution patterns, which were detected by both metagenetic and *P. abdominalis* study, might be the result of past transportation of equatorial species into the Kuroshio Current region, as suggested by McGowan (1971). In the biomass-based metagenetic analysis, community structure in Equator-Kuroshio region was correlated to high chl-*a*. Therefore, relatively high food availability in Kuroshio region works to facilitate their survival and maintenance of high biomass of equatorial species. On the other hand, small proportions of Equator-Kuroshio species in subtropical gyres might be related to low population interactions and/or poor adaptation to low food availability.

The spatial pattern of MOTU richness in this study was well agreed with estimated global patterns of copepod diversity by Rombouts et al. (2010), which are mainly based on data in the Atlantic Ocean, and high species richness was observed in the subtropical gyres. Both North Pacific Subtropical gyre (NPSG) and South Pacific Subtropical Gyre (SPSG) showed high values of diversity index, which indicate high evenness with few dominant species. However, the number of MOTUs (proxy of species number) was higher in NPSG than in SPSG. The high species number of

oceanic taxa including copepods has been also reported in the subtropical areas in northern hemisphere (Woodd-Walker et al. 2002; Rombouts et al. 2009, 2010; Tittensor et al. 2010). In addition to large numbers of MOTUs of subtropical gyre group, MOTUs with Kuroshio-Equator group were more frequently observed in NPSG than in SPSG. This metagenetic result agreed with the study of *P. abdominalis*, *Pleuromamma abdominalis* 1 with Equator-Kurohio distribution was distributed in NPSG with small frequencies but almost absent in SPSG. As discussed above, higher connectivity and historical population interactions are suggested in the subtropical North Pacific than in the subtropical South Pacific. The population interactions and accumulation of species are known as main reasons of diversity hot spot at the coastal coral areas in the Caribbean and the Indo-West Pacific oceans (e.g., Rocha et al. 2008; Gaither et al. 2011). In addition, chl-*a* concentration is higher in NPSG than SPSG, and species with adaptation to high food availability might be maintained in NPSG. The food web structure is complex in NPSG (McGowan and Walker 1985), and various species are capable of co-existing. Therefore, high connectivity and food availability and co-existence of various species might be possible factors of high species richness in NPSG, compared with SPSG in sampling areas in this study.

Sampling stations in this paper have been restricted to specific areas in the tropical and subtropical Pacific. In metagenetic analysis, there is no data available in the western subtropical North Pacific where the highest genetic diversity of *P. abdominalis* was observed. Therefore, an additional sampling is necessary to reveal detailed diversity and biogeography of copepods in the Pacific. In addition, the Pacific Ocean is contemporary and historically related to other ocean basins. Populations between Pacific and Atlantic were connected until the isolation after the closure of the Isthmus of Panama (Knowlton and Weigt 1988; Blanco-Bercial et al. 2011a). The Coral Triangle is located between the Pacific and the Indian Ocean, however, the Coral Triangle do not act as strong barrios for oceanic copepods (Goetze 2011). A global sampling must provide a significant insight on global-scale diversity and biogeography of copepods. Once global data of metagenetic analysis is

established, it is possible to compare diversity pattern among ocean basins or to investigate distribution pattern of specific species.

Future perspective

The technique of DNA-barcoding has been developed to identify species easily based on similarity to known sequences available in public databases (Hebert et al. 2003). In zooplankton taxa, the DNA database of various species have been accounted through the Census of Marine Zooplankton (Bucklin et al. 2010), and a new method has been expected for revealing community structures efficiently from the environmental bulk samples. This paper is the first metagenetic study focusing on community structures of copepods using deep-sequencing. A metagenetic approach is currently not common method for community analysis of zooplankton taxa, however, it is expected that this rapid and comprehensive method will be widely used in the future. A choice of molecular marker is important for sensitivity of identification and universality of primer pair, and nuclear 28S-D2 region was selected for metagenetic analysis of copepods in this study. Other studies also proposed molecular markers and universal primer pairs for metagenetic analysis of metazoans, such as COI (Leray et al. 2013), nuclear 28S and 18S (Machida and Knowlton 2012) and mitochondrial 12S (Machida et al. 2012). The method without PCR step using deep-sequencing has been also suggested for diversity study of metazoans to eliminate biases during PCR procedure (Dettai et al. 2012; Zhou et al. 2013). The technologies of deep-sequencing have been developing, and sequences data and length are increasing with improvement of sequencing accuracy, in addition, cost of sequence run is getting low (Glenn et al. 2011). The current methodological problem of metagenetic analysis (e.g., PCR bias and inflation of MOTUs) will be improved with the development of sequencing and bioinformatics technologies.

In addition to analysis of community structure, deep-sequencing works in many aspects of ecological studies of zooplankton. For example, metagenetic analysis is an effective method for

feeding habits, contributing to understanding of food-web structures in complex marine ecosystems (Pompanon et al. 2012). Fernandez-Silva et al. (2013) introduced a new workflow to construct microsatellite markers, which would help understanding species complexity of *P. abdominalis* in this paper. In addition, deep-sequencing has been used for obtaining whole mitochondrial genome in euphausiids (Johansson et al. 2012) or transcriptome analysis in copepods (Barreto et al. 2011; Ning et al. 2013; Christie et al. 2013a, b). There are not many studies using deep-sequencing for copepods, however, deep-sequencing is expected to be used for various aspects of ecological studies of copepods in the future. It is expected that those new technologies contribute a further understanding of global-level diversity and biogeography of copepods in the future.

Acknowledgements

I would like to thank my supervisor, Professor Atsushi Tsuda (The University of Tokyo) for providing me an opportunity of doctoral study, thoughtful guidance and active support throughout this dissertation.

I would also like to thank my thesis committee members, Professor Shuhei Nishida, Associate Professors Koji Hamasaki, Kazutaka Takahashi, and Sachihiko Ito (The University of Tokyo) for their critical reading of this dissertation.

The morphological classifications were kindly performed by Associate Professor Shinji Shimode (Yokohama National University) in chapter II and by Dr. Mikiko Kuriyama (National Research Institute of Fisheries Science) in chapter III. Without their technical supports, this dissertation would not have been possible.

I am deeply grateful to Dr. Hiroomi Miyamoto (Tohoku National Fisheries Research Institute) for teaching me methods of genetic experiment and data analysis.

I also wish to express appreciation to Dr. Erica Goetze (University of Hawaii) for giving insightful comments and suggestions on an earlier version of this dissertation.

For the samplings in Kuroshio Current regions, Drs. Hiroya Sugisaki, Tadafumi Ichikawa and Kiyotaka Hidaka (National Research Institute of Fisheries Science) gave me opportunities of sample collections on “Soyo-maru”. I also indebted to Dr. Jun Aoyama (The University of Tokyo), who kindly provided samples collected in the western equatorial and south Pacific for chapter V.

I also thank the captains, crew, and researchers aboard the “Hakuho-Maru”, “Soyo-Maru” and “Tansei-maru” for assistance with field collections.

The metagenetic works were carried out using 454 second-generation sequencers (Roche) owned by the Center for Earth Surface System Dynamics and the Center for Omics and Bioinformatics (The University of Tokyo). I appreciate staffs in both centers for technical supports.

Finally, I would like to offer special thanks to all members of the Plankton Laboratory and

Laboratory of International Research Cooperation (The University of Tokyo) for supporting me in various ways. Without their help and encouragement, this dissertation would not have materialized.

References

Adamowicz SJ, Menu-Marque S, Halse SA, Topan JC, Zemlak TS, Hebert PD, Witt JD (2010) The evolutionary diversification of the Centropagidae (Crustacea, Calanoida): A history of habitat shifts. *Molecular Phylogenetics and Evolution* **55**, 418–430.

Al-Mutairi H, Landry MR (2001) Active export of carbon and nitrogen at Station ALOHA by diel migrant zooplankton. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**, 208–2103.

Andronov VN (2001) On the taxonomy of the genus *Nannocalanus* Stars, 1925 (Crustacea, Copepoda: Calanidae). *Zoosystematica Rossica* **9**, 277–283.

Angel MV (1993) Biodiversity of the Pelagic Ocean. *Conservation Biology* **7**, 760–772.

Avise J, Arnold J, Ball R, Bermingham E, Lamb T, Neigel J, Reeb C, Saunders N (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review in Ecology and Systematic*, **18**, 489–522.

Bacha M, Amara R (2009) Spatial, temporal and ontogenetic variation in diet of anchovy (*Engraulis encrasicolus*) on the Algerian coast (SW Mediterranean). Estuarine. *Coastal and Shelf Science* **85**, 257–264.

Barreto FS, Moy GW, Burton RS (2011) Interpopulation patterns of divergence and selection across the transcriptome of the copepod *Tigriopus californicus*. *Molecular Ecology* **20**, 560–572.

Barron JA (1985) Diatom paleoceanography and paleoclimatology of the central and eastern equatorial Pacific between 18 and 6.2 Ma. *Initial Reports, Deep Sea Drilling Project* **85**, 935–945.

Barton AD, Pershing AJ, Litchman E, Record NR, Edwards KF, Finkel ZV, Kiørboe T, Ward BA (2013) The biogeography of marine plankton traits. *Ecology Letters* **16**, 522–534.

Batten SD, Clark R, Flinkman J, Hays G, John E, John AWG, Jonas T, Lindley JA, Stevens DP, Walne A (2003) CPR sampling: the technical background, materials and methods, consistency and comparability. *Progress in Oceanography* **58**, 193–215.

Beaugrand G, Reid PC, Ibañez F, Lindley JA, Edwards M (2002) Reorganization of North Atlantic

Marine Copepod Biodiversity and Climate. *Science* **296**, 1692–1694.

Beaugrand G, Brander KM, Alistair Lindley J, Souissi S, Reid PC (2003) Plankton effect on cod recruitment in the North Sea. *Nature* **426**, 661–664.

Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution* **27**, 233–243.

Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**, 613–627.

Bittner L, Gobet A, Audic S, Romac S, Egge ES, Santini S, Ogata H, Probert I, Edvardsen B, de Vargas C (2013) Diversity patterns of uncultured Haptophytes unravelled by pyrosequencing in Naples Bay. *Molecular Ecology* **22**, 87–101.

Blades-Eckelbarger PI, Youngbluth MJ (1988) Ultrastructure of the “pigment knob” of *Pleuromamma* spp. (Copepoda: Calanoida). *Journal of Morphology* **197**, 315–326.

Blanco-Bercial L, Álvarez-Marqués F (2007) RFLP procedure to discriminate between *Clausocalanus* Giesbrecht, 1888 (Copepoda, Calanoida) species in the Central Cantabrian Sea. *Journal of Experimental Marine Biology and Ecology* **344**, 73–77.

Blanco-Bercial L, Bucklin A, Viñas MD, Cornils A. (2010) Zooplankton phylogeography as a measure of Atlantic Ocean ecosystem connectivity. Paper presented at Ocean Sciences Meeting, Portland.

Blanco-Bercial L, Álvarez-Marqués F, Bucklin A (2011a) Comparative phylogeography and connectivity of sibling species of the marine copepod *Clausocalanus* (Calanoida). *Journal of Experimental Marine Biology and Ecology* **404**, 108–115.

Blanco-Bercial L, Bradford-Grieve J, Bucklin A (2011b) Molecular phylogeny of the Calanoida (Crustacea: Copepoda). *Molecular Phylogenetics and Evolution* **59**, 103–113.

Bortolotto E, Bucklin A, Mezzavilla M, Zane L, Patarnello T (2011) Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill *Euphausia superba*. *BMC genetics* **12**, 32.

Böttger-Schnack R, Machida RJ (2011) Comparison of morphological and molecular traits for species identification and taxonomic grouping of oncaeid copepods. *Hydrobiologia* **666**, 111–125.

Boxshall GA, Halsey SH (2004) *An Introduction to Copepod Diversity*. The Ray Society, London.

Braga E, Zardoya R, Meyer A, Yen J (1999) Mitochondrial and nuclear rRNA based copepod phylogeny with emphasis on the Euchaetidae (Calanoida). *Marine Biology* **133**, 79–90.

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. . *Ecological Monographs* **27**, 325–349.

Brinton E (1962) The distribution of Pacific euphausiids. *Bulletin of the Scripps Institute of Oceanography* **8**, 51–270.

Bron JE, Frisch D, Goetze E, Johnson SC, Lee CE, Wyngaard GA (2011) Observing copepods through a genomic lens. *Frontiers in Zoology* **8**, 22.

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences* **76**, 1967–1971.

Bucklin A, Frost BW, Kocher TD (1995) Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). *Marine Biology* **121**, 655–664.

Bucklin A, LaJeunesse TC, Curry E, Wallinga J, Garrison K (1996) Molecular diversity of the copepod, *Nannocalanus minor*: Genetic evidence of species and population structure in the North Atlantic Ocean. *Journal of Marine Research* **54**, 285–310.

Bucklin A, Frost, BW, Bradford-Grieve J, Allen LD, Copley NJ (2003) Molecular systematics and phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Marine Biology* **142**, 333–343.

Bucklin A, Wiebe PH, Smolenack SB, Copley NJ, Beaudet JG, Bonner KG, Färber-Lorda J, Pierson JJ (2007) DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). *Journal of Plankton Research* **29**, 483–493.

Bucklin A, Nishida S, Schnack-Schiel S, Wiebe PH, Lindsay D, Machida RJ, Copley NJ (2010) A

Census of Zooplankton of the Global Ocean. In: *Marine Life: Diversity, Distribution, and Abundance*. (ed McIntyre A). pp. 247–265, Wiley-Blackwell, Oxford.

Bucklin A, Steinke D, Blanco-Bercial L (2011) DNA Barcoding of Marine Metazoa. *Annual Review of Marine Science* **3**, 471–508.

Carew ME, Pettigrove VJ, Metzeling L, Hoffmann AA (2013) Environmental monitoring using next generation sequencing: rapid identification of macroinvertebrate bioindicator species. *Frontiers in Zoology* **10**, 45.

Cepeda GD, Blanco-Bercial L, Bucklin A, Beron CM, Vinas MD (2012) Molecular systematic of three species of *Oithona* (Copepoda, Cyclopoida) from the Atlantic Ocean: comparative analysis using 28S rDNA. *Plos One* **7**, e35861.

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, *Acartia tonsa*. *Molecular Ecology* **17**, 1451–1468.

Chen G, Hare MP (2011) Cryptic diversity and comparative phylogeography of the estuarine copepod *Acartia tonsa* on the US Atlantic coast. *Molecular ecology* **20**, 2425–2441.

Chen YQ (1986) The vertical distribution of some pelagic copepods in the eastern tropical Pacific. *California Cooperative Oceanic Fisheries Investigations Reports* **27**, 205–227.

Chihara M. Murano M (1997) *An illustrated guide to marine plankton in Japan*. Tokai university Press, Tokyo.

Christie AE, Fontanilla TM, Nesbit KT, Lenz PH (2013a) Prediction of the protein components of a putative *Calanus finmarchicus* (Crustacea, Copepoda) circadian signaling system using a de novo assembled transcriptome. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **8**, 165–193.

Christie AE, Roncalli V, Wu L-S, Ganote CL, Doak T, Lenz PH (2013b) Peptidergic signaling in *Calanus finmarchicus* (Crustacea, Copepoda): In silico identification of putative peptide hormones and their receptors using a de novo assembled transcriptome. *General and Comparative Endocrinology* **187**, 117–135.

Clarke KR, Warwick RM (2001) Change in marine communities: an approach to statistical analysis and interpretation, 2nd edition. PRIMER-E, Plymouth.

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M, Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 272–295.

Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML, Lambshead PJ, Thomas WK (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology* **19 Suppl 1**, 4–20.

de Boyer Montégut C, Madec G, Fischer AS, Lazar A, Iudicone D (2004) Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans* **109**, C12003.

de Queiroz K (2007) Species Concepts and Species Delimitation. *Systematic Biology* **56**, 879–886.

Dettai A, Gallut C, Brouillet S, Pothier J, Lecointre G, Debruyne R (2012) Conveniently pre-tagged and pre-packaged: extended molecular identification and metagenomics using complete metazoan mitochondrial genomes. *Plos One* **7**, e51263.

Dowle EJ, Morgan-Richards M, Trewick SA (2013) Molecular evolution and the latitudinal biodiversity gradient. *Heredity* **110**, 501–510.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.3.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200.

Elder JF, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review in Biology* **70**, 297–320

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Fernandez-Silva I, Whitney J, Wainwright B, Andrews KR, Ylitalo-Ward H, Bowen BW, Toonen RJ, Goetze E, Karl SA (2013) Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *Plos One* **8**, e55990.

Fine RA, Lukas R, Bingham FM, Warner MJ, Gammon RH (1994) The western equatorial Pacific: A water mass crossroads. *Journal of Geophysical Research: Oceans* **99**, 25063–25080.

Fleminger A, Hulsemann K (1974) Systematics and distribution of the four sibling species comprising the genus *Pontellina* Dana (Copepoda, Calanoida). *Fishery Bulletin* **72**, 63–120.

Floyd R, Abebe E, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology* **11**, 839–850.

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**, 294–299.

Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, Neill SP, Packer M, Blaxter ML, Lambshead PJD, Thomas WK, Creer S (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature Communications* **1**, 98.

Fornshell J, Ferrari F (2010) Morphological variability of *Pleuromamma abdominalis* (Copepoda, Calanoida, Metridinidae) along two latitudinal transects in the eastern North Pacific Ocean. *Crustaceana* **83**, 753–765.

Foxton P, Roe HSJ (1974) Observations on the nocturnal feeding of some mesopelagic decapod crustacea. *Marine Biology* **28**, 37–49.

Fragopoulu N, Siokou-Frangou I, Christou ED, Mazzocchi MG (2001) Patterns of vertical distribution of Pseudocalanidae and Paracalanidae (Copepoda) in pelagic waters (0 to 300 m) of the eastern mediterranean sea. *Crustaceana* **74**, 49–68.

Frost BW (1969) *Distribution of the ocean, epipelagic copepod genus Clausocalanus with an analysis of sympatry of North Pacific species*. PhD Thesis, Scripps Institute of Oceanography, University of California, San Diego.

Gaither M, Bowen B, Bordenave T-R, Rocha L, Newman S, Gomez J, van Herwerden L, Craig M (2011) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates

Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. *BMC Evolutionary Biology* **11**, 189.

Gilles A, Meglecz E, Pech N, Ferreira S, Malausa T, Martin J-F (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. *BMC Genomics*, **12**, 245.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* **11**, 759–769.

Goetze E (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. *Proceedings of the Royal Society B* **270**, 2321–2331.

Goetze E (2005) Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer*. *Evolution* **59**, 2378–2398.

Goetze E (2010) Species discovery in marine planktonic invertebrates through global molecular screening. *Molecular Ecology* **19**, 952–967.

Goetze E (2011) Population differentiation in the open sea: Insights from the Pelagic Copepod *Pleuromamma xiphias*. *Integrative and Comparative Biology* **51**, 580–597.

Goetze E, Bradford-Grieve J (2005) Genetic and morphological description of *Eucalanus spinifer* T. Scott, 1894 (Calanoida: Eucalanidae), a circumglobal sister species of the copepod *E. hyalinus* s.s. (Claus, 1866). *Progress in Oceanography* **65**, 55–87.

Goetze E, Ohman MD (2010) Integrated molecular and morphological biogeography of the calanoid copepod family Eucalanidae. *Deep Sea Research Part II: Topical Studies in Oceanography* **57**, 2110–2129.

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* **27**, 221–224.

Hajibabaei M, Shokralla S, Zhou X, Singer GA, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS One* **6**, e17497.

Halbert KMK, Goetze E, Carlon DB (2013) High cryptic diversity across the global range of the

migratory planktonic copepods *Pleuromamma piseki* and *P. gracilis*. *Plos One* **8**, e77011.

Hassouna N, Michot B, Bachelleire JP (1984) The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Research* **12**, 3563–3583.

Haury L (1988) Vertical distribution of *Pleuromamma* (Copepoda: Metridinidae) across the eastern North Pacific Ocean. In: *Biology of Copepods* (eds Boxshall G, Schminke HK). Springer, Netherlands.

Hays GC, Harris RP, Head RN (1997) The vertical nitrogen flux caused by zooplankton diel vertical migration. *Marine Ecology Progress Series* **160**, 57–62.

Hays GC, Richardson AJ, Robinson C (2005) Climate change and marine plankton. *Trends in Ecology & Evolution* **20**, 337–344.

Hebert P, Cywinska A, Ball S, deWaard J (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B* **270**, 313–321.

Hirase S, Ikeda M, Kanno M, Kijima A (2012) Phylogeography of the intertidal goby *Chaenogobius annularis* associated with paleoenvironmental changes around the Japanese Archipelago. *Marine Ecology Progress Series* **450**, 167–179.

Huber JA, Mark Welch DB, Morrison HG, Huse SM, Neal PR, Butterfield DA, Sogin ML (2007) Microbial population structures in the deep marine biosphere. *Science* **318**, 97–100.

Humes AG (1994) How many copepods? *Hydrobiologia* **292–293**, 1–7.

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology* **8**, R143.

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology* **12**, 1889–1898.

Ikeda T (1985) Metabolic rates of epipelagic marine zooplankton as a function of body mass and temperature. *Marine Biology* **85**, 1–11.

Irigoien X, Huisman J, Harris RP (2004) Global biodiversity patterns of marine phytoplankton and

zooplankton. *Nature* **429**, 863–867

Jennings RM, Bucklin A, Ossenbrügger H, Hopcroft RR (2010) Species diversity of planktonic gastropods (Pteropoda and Heteropoda) from six ocean regions based on DNA barcode analysis. *Deep Sea Research Part II: Topical Studies in Oceanography* **57**, 2199–2210.

Johansson ML, Sremba AL, Feinberg LR, Banks MA, Peterson WT (2012) The mitochondrial genomes of *Euphausia pacifica* and *Thysanoessa raschii* sequenced using 454 next-generation sequencing, with a phylogenetic analysis of their position in the Malacostracan family tree. *Molecular Biology Reports* **39**, 9009-9021.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics* **9**, 286–298.

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780.

Kiesling TL, Wilkinson E, Rabalais J, Ortner PB, McCabe MM, Fell JW (2002) Rapid identification of adult and naupliar stages of copepods using DNA hybridization methodology. *Marine Biotechnology* **4**, 30–39.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111–120.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society B* **265**, 2257–2263.

Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**, 118–123.

Lahr DJ, Katz LA (2009) Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques* **47**, 857–866.

Lancraft TM, Hopkins TL, Torres JJ (1988) Aspects of the ecology of the mesopelagic fish *Gonostoma elongatum* (Gonostomatidae, Stomnformes) in the eastern Gulf of Mexico, *Marine Ecology Progress Series* **49**, 27–40.

Landry MR, Al-Mutairi H, Selph KE, Christensen S, Nunnery S (2001) Seasonal patterns of mesozooplankton abundance and biomass at Station ALOHA. *Deep Sea Research Part II: Topical Studies in Oceanography* **48**, 2037–2061.

Lee CE (2000) Global phylogeography of a cryptic copepod species complex and reproductive isolation between genetically proximate "populations". *Evolution* **54**, 2014–2027.

Lee CY, Liu DC, Su WC (2009) Seasonal and Spatial Variations in the Planktonic Copepod Community of Ilan Bay and Adjacent Kuroshio Waters off Northeastern Taiwan. *Zoological Studies* **48**, 151–161.

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (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**, 34.

Lindeque PK, Hay SJ, Heath MR, Ingvarsdottir A, Rasmussen J, Smerdon GR, Waniek JJ (2006) Integrating conventional microscopy and molecular analysis to analyse the abundance and distribution of four *Calanus* congeners in the North Atlantic. *Journal of Plankton Research* **28**, 221–238.

Llinas L (2008) *Distribution, reproduction, and transport of zooplankton in the western Arctic*. Ph.D. Dissertation. University of Miami, Coral Gables, Florida.

Long EO, Dawid B (1980) Repeated genes in eukaryotes. *Annual Review of Biochemistry* **49**, 726–727.

Longhurst AR (1985a) The structure and evolution of plankton communities. *Progress in Oceanography* **15**, 1–35.

Longhurst AR (1985b) Relationship between diversity and the vertical structure of the upper ocean. *Deep Sea Research Part A: Oceanographic Research Papers* **32**, 1535–1570.

Longhurst AR (2007) *Ecological Geography of the Sea*. Academic Press, London.

Machida RJ, Miya MU, Nishida M, Nishida S (2004) Large-scale gene rearrangements in the mitochondrial genomes of two calanoid copepods *Eucalanus bungii* and *Neocalanus cristatus*

(Crustacea), with notes on new versatile primers for the srRNA and COI genes. *Gene* **332**, 71–78.

Machida RJ, Tsuda A (2010) Dissimilarity of species and forms of planktonic *Neocalanus* copepods using mitochondrial *COI*, *12S*, nuclear *ITS*, and *28S* gene sequences. *Plos One* **5**, e10278.

Machida RJ, Knowlton N (2012) PCR primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. *Plos One* **7**, e46180.

Machida RJ, Kweskin M, Knowlton N (2012) PCR primers for metazoan mitochondrial 12S ribosomal DNA sequences. *Plos One* **7**, e35887.

Mackas DL, Sefton H, Miller CB, Raich A (1993) Vertical habitat partitioning by large calanoid copepods in the oceanic subarctic Pacific during Spring. *Progress in Oceanography* **32**, 259–294.

Marszalek M, Dayanandan S, Maly E (2009) Phylogeny of the genus *Hesperodiaptomus* (Copepoda) based on nucleotide sequence data of the nuclear ribosomal gene. *Hydrobiologia* **624**, 61–69.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

Mauchline J (1998) *The biology of calanoid copepods*. Academic Press, San Diego.

McGowan JA (1971) Oceanic biogeography of the Pacific. In: *The micropaleontology of the oceans*. (eds Funnell BM, Riedel WR). pp. 3–74, Cambridge University Press, Cambridge.

McGowan JA (1974) The nature of oceanic ecosystems. In: *The biology of the Pacific Ocean*. (ed Miller CB). pp. 9–28, Oregon State University Press, Corvallis.

McGowan JA, Walker PW (1979) Structure in the Copepod Community of the North Pacific Central Gyre. *Ecological Monographs* **49**, 195–226.

McGowan JA, Walker PW (1985) Dominance and Diversity Maintenance in an Oceanic Ecosystem. *Ecological Monographs* **55**, 103–118.

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwalder B, Schlotterer C, Boenigk J (2010) Diversity in a hidden world: potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Molecular Ecology* **19 Suppl 1**, 32–40.

Meyerhans A, Vartanian J-P, Wain-Hobson S (1990) DNA recombination during PCR. *Nucleic Acids Research* **18**, 1687–1691.

Minxiao W, Song S, Chaolun L, Xin S (2011) Distinctive mitochondrial genome of Calanoid copepod *Calanus sinicus* with multiple large non-coding regions and reshuffled gene order: useful molecular markers for phylogenetic and population studies. *BMC genomics* **12**, 73.

Miyamoto H, Machida RJ, Nishida S (2010) Genetic diversity and cryptic speciation of the deep sea chaetognath *Caecosagitta macrocephala* (Fowler, 1904). *Deep Sea Research Part II: Topical Studies in Oceanography* **57**, 2211–2219.

Miyamoto H, Machida RJ, Nishida S (2012) Global phylogeography of the deep-sea pelagic chaetognath *Eukrohnia hamata*. *Progress in Oceanography* **104**, 99–109.

Nakata K, Hidaka K (2003) Decadal-scale variability in the Kuroshio marine ecosystem in winter. *Fisheries Oceanography* **12**, 234–244.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* **76**, 5269–5273.

Ning J, Wang M, Li C, Sun S (2013) Transcriptome sequencing and *De Novo* analysis of the copepod *Calanus sinicus* using 454 GS FLX. *Plos One* **8**, e63741.

Nolte V, Pandey RV, Jost S, Medinger R, Ottenwalder B, Boenigk J, Schlotterer C (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. *Molecular Ecology* **19**, 2908–2915.

Nonomura T, Machida RJ, Nishida S (2008) Stage-V copepodites of *Calanus sinicus* and *Calanus*

jashnovi (Copepoda: Calanoida) in mesopelagic zone of Sagami Bay as identified with genetic markers, with special reference to their vertical distribution. *Progress in Oceanography* **77**, 45–55.

Norris RD (2000) Pelagic species diversity, biogeography, and evolution. *Paleobiology* **26**, 236–258

Norton EL, Goetze E (2013) Equatorial dispersal barriers and limited population connectivity among oceans in a planktonic copepod. *Limnology and Oceanography* **58**, 1581–1596.

Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey J (2004) Bayesian Phylogenetic Analysis of Combined Data. *Systematic Biology* **53**, 47–67.

Peijnenburg KT, Breeuwer JA, Pierrot-Bults AC, Menken SB (2004) Phylogeography of the planktonic chaetognath *Sagitta setosa* reveals isolation in European seas. *Evolution* **58**, 1472–1487.

Peijnenburg KT, Fauvelot C, Breeuwer JA, Menken SB (2006) Spatial and temporal genetic structure of the planktonic *Sagitta setosa* (Chaetognatha) in European seas as revealed by mitochondrial and nuclear DNA markers. *Molecular Ecology* **15**, 3319–3338.

Peijnenburg KTCA, Goetze E (2013) High evolutionary potential of marine zooplankton. *Ecology and Evolution* **3**, 2765–2781.

Pflaumann U, Sarnthein M, Chapman M, d'Abreu L, Funnell B, Huels M, Kiefer T, Maslin M, Schulz H, Swallow J, van Kreveld S, Vautravers M, Vogelsang E, Weinelt M (2003) Glacial North Atlantic: Sea-surface conditions reconstructed by GLAMAP 2000. *Paleoceanography* **18**, 1065.

Piontkovski SA, Landry MR, Finenko ZZ, Kovalev AV, Williams R, Gallienne CP, Mishonov AV, Skryabin VA, Tokarev YN, Nikolsky VN (2003) Plankton communities of the South Atlantic anticyclonic gyre. *Oceanologica Acta* **26**, 255–268.

Poinar HN, Schwarz C, Qi J, Shapiro B, Macphee RD, Buigues B, Tikhonov A, Huson DH, Tomsho LP, Auch A, Rampp M, Miller W, Schuster SC (2006) Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. *Science* **311**, 392–394.

Pompanon F, Deagle BE, Symondson WO, Brown DS, Jarman SN, Taberlet P (2012) Who is eating what: diet assessment using next generation sequencing. *Molecular Ecology* **21**, 1931–1950.

Porazinska DL, Giblin-Davis RM, Faller L, Farmerie W, Kanzaki N, Morris K, Powers TO, Tucker AE, Sung W, Thomas WK (2009) Evaluating high-throughput sequencing as a method for metagenomic analysis of nematode diversity. *Molecular Ecology Resources* **9**, 1439–1450.

Provan J, Beatty GE, Keating SL, Maggs CA, Savidge G (2009) High dispersal potential has maintained long-term population stability in the North Atlantic copepod *Calanus finmarchicus*. *Proceedings of the Royal Society B* **276**, 301–307.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35**, 7188–7196.

Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. *Applied and Environmental Microbiology* **67**, 880–887.

Qu T, Lindstrom EJ (2002) A Climatological Interpretation of the Circulation in the Western South Pacific. *Journal of Physical Oceanography* **32**, 2492–2508.

Raupach M, Astrin J, Hannig K, Peters M, Stoeckle M, Wagele J-W (2010) Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in Zoology* **7**, 26.

Razouls C, de Bovée F, Kouwenberg J, et Desreumaux N (2005–2013) Diversity and Geographic Distribution of Marine Planktonic Copepods. Available at http://copepodes.obs-banyuls.fr/en [Accessed September 04, 2013].

Regier JC, Shultz JW, Kambic RE (2005) Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. *Proceedings of the Royal Society B* **272**, 395–401.

Richardson AJ (2008) In hot water: zooplankton and climate change. *ICES Journal of Marine Science* **65**, 279–295.

Richardson AJ, Walne AW, John AWG, Jonas TD, Lindley JA, Sims DW, Stevens D, Witt M (2006) Using continuous plankton recorder data. *Progress in Oceanography* **68**, 27–74.

Rocha L, Rocha C, Robertson DR, Bowen B (2008) Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean. *BMC Evolutionary Biology* **8**, 157.

Roe HSJ (1972) The Vertical Distributions and Diurnal Migrations of Calanoid Copepods Collected on the Sond Cruise, 1965 III. Systematic Account: Families Euchaetidae up to and Including the Metridiidae. *Journal of the Marine Biological Association of the United Kingdom* **52**, 525–552.

Roemmich D, McGowan J (1995) Climatic warming and the decline of zooplankton in the California current. *Science* **267**, 1324–1326.

Rombouts I, Beaugrand G, Ibanez F, Gasparini S, Chiba S, Legendre L (2009) Global latitudinal variations in marine copepod diversity and environmental factors. *Proceedings of the Royal Society B* **276**, 3053–3062.

Rombouts I, Beaugrand G, Ibañez F, Gasparini S, Chiba S, Legendre L (2010) A multivariate approach to large-scale variation in marine planktonic copepod diversity and its environmental correlates. *Limnology and Oceanography* **55**, 2219–2229.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Rutherford S, D'Hondt S, Prell W (1999) Environmental controls on the geographic distribution of zooplankton diversity. *Nature* **400**, 749–753.

Sameoto DD (1986) Influence of the biological and physical environment on the vertical distribution of mesozooplankton and micronekton in the eastern tropical Pacific. *Marine Biology* **93**, 263–279.

Schizas NV, Street GT, Coull BC, Chandler GT, Quattro JM (1997) An efficient DNA extraction method for small metazoans. *Molecular Marine Biology and Biotechnology* **6**, 381–383.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing MOTHUR: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* **75**, 7537–7541.

Schnack-Schiel SB, Mizdalski E, Cornils A (2010) Copepod abundance and species composition in the Eastern subtropical/tropical Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography* **57**, 2064–2075.

Schroth W, Jarms G, Streit B, Schierwater B (2002) Speciation and phylogeography in the cosmopolitan marine moon jelly, *Aurelia* sp. *BMC Evolutionary Biology* **2**, 1.

Shannon CE (1948) A mathematical theory of communication. T*he Bell System Technical Journal* **27**, 379–423 and 623–656.

Shen K-N, Jamandre BW, Hsu C-C, Tzeng W-N, Durand J-D (2011) Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus*. *BMC Evolutionary Biology* **11**, 83.

Shih CT, Chiu TS (1998) Copepod diversity in the water masses of the southern East China Sea north of Taiwan. *Journal of Marine Systems* **15**, 533–542.

Shimode S, Toda T, Kikuchi T (2006) Spatio-temporal changes in diversity and community structure of planktonic copepods in Sagami Bay, Japan. *Marine Biol*ogy **148**, 581–597.

Shimode S, Takahashi K, Shimizu Y, Nonomura T, Tsuda A (2012) Distribution and life history of the planktonic copepod, *Eucalanus californicus*, in the northwestern Pacific: Mechanisms for population maintenance within a high primary production area. *Progress in Oceanography* **96**, 1–13.

Simpson EH (1949) Measurement of diversity. *Nature* **163**, 688.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences* **103**, 12115–12120.

Song Y, Wang GT, Yao WJ, Gao Q, Nie P (2008) Phylogeny of freshwater parasitic copepods in the Ergasilidae (Copepoda : Poecilostomatoida) based on 18S and 28S rDNA sequences. *Parasitology Research* **102**, 299–306.

Sonnenberg R, Nolte A, Tautz D (2007) An evaluation of LSU rDNA D1-D2 sequences for their use in species identification. *Frontiers in Zoology* **4**, 6.

Steinberg DK, Carlson CA, Bates NR, Goldthwait SA, Madin LP, Michaels AF (2000) Zooplankton vertical migration and the active transport of dissolved organic and inorganic carbon in the Sargasso Sea. *Deep Sea Research Part I: Oceanographic Research Papers* **47**, 137–158.

Steuer A (1932) Copepoda 6: *Pleuromamma* Giesbr. 1898 der Deutschen Tiefsee-Expedition. *Wissenschaftliche Ergebnisse der Deutschen Tifsee-Expedition 1898–1899* **24**, 1–128.

Sugisaki H, Nonaka M, Ishizaki S, Hidaka K, Kameda T, Hirota Y, Oozeki Y, Kubota H, Takasuka A (2010) Status and trends of the Kuroshio region, 2003-2008. In: *Marine ecosystems of the North Pacific Ocean, 2003–2008. PICES Special Publication 4* (eds McKinnell SM, Dagg MJ). pp. 330–359, North Pacific Marine Science Organization, Sidney.

Suto I, Kawamura K, Hagimoto S, Teraishi A, Tanaka Y (2012) Changes in upwelling mechanisms drove the evolution of marine organisms. *Palaeogeography, Palaeoclimatology, Palaeoecology* **339–341**, 39–51.

Sverdrup HU, Johnson MV, Fleming RH (1942) *The Oceans, Their Physics, Chemistry and Biology*, Prentice Hall, Englewood Cliffs.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

Tamura K, Peterson D, Peterson N, Stecher G, Neo M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731–2739.

ter Braak CF, Verdonschot PM (1995) Canonical correspondence analysis and related multivariate methods in aquatic ecology. *Aquatic Science* **57**, 255–289.

Terazaki M, Tomatsu C (1997) A vertical multiple opening and closing plankton sampler. *Journal of Advanced Marine Science and Technology Society* **3**, 127–132.

Thompson J, Gibson T, Plewniak F, Jeanmougin F, Higgins D (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.

Nucleic Acids Research **24**, 4876–4882.

Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe EV, Worm B (2010) Global patterns and predictors of marine biodiversity across taxa. *Nature* **466**, 1098–1101.

Toda T (1986) Some problems in the classification of calanoid copepods. *Aquabiology*, **8**, 182–192.

Tsuchi R (1997) Marine climatic responses to Neogene tectonics of the Pacific Ocean seaways. *Tectonophysics* **281**, 113–124.

Turner JT (2004) The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zoological Studies* **43**, 255–266.

Ueda H, Yamaguchi A, Saitoh S-i, Sakaguchi SO, Tachihara K (2011) Speciation of two salinity-associated size forms of *Oithona dissimilis* (Copepoda: Cyclopoida) in estuaries. *Journal of Natural History* **45**, 2069–2079.

Unal E, Bucklin A (2010) Basin-scale population genetic structure of the planktonic copepod *Calanus finmarchicus* in the North Atlantic Ocean. *Progress in Oceanography* **87**, 175–185.

Uye S, Matsuda O (1988) Phosphorus content of zooplankton from the Inland Sea of Japan. *Journal of the Oceanographical Society of Japan* **44**, 280–286.

van der Spoel S, Heyman RP (1983) *A comparative atlas of zooplankton*. Springer-Verlag, Berlin.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**, 5261–5267.

Weiner A, Aurahs R, Kurasawa A, Kitazato H, Kucera M (2012) Vertical niche partitioning between cryptic sibling species of a cosmopolitan marine planktonic protist. *Molecular Ecology* **21**, 4063–4073.

Welschmeyer (1994) Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. *Limnology and Oceanography* **39**, 1985–1992.

Williamson M, McGowan JA (2010) The copepod communities of the north and south Pacific central gyres and the form of species-abundance distributions. *Journal of Plankton Research* **32**,

273–283.

Woodd-Walker RS, Ward P, Clarke A (2002) Large-scale patterns in diversity and community structure of surface water copepods from the Atlantic Ocean. *Marine Ecology Progress Series* **236**, 189–203.

Wyngaard GA, Holynska M, Schulte II JA (2010) Phylogeny of the freshwater copepod *Mesocyclops* (Crustacea: Cyclopidae) based on combined molecular and morphological data, with notes on biogeography. *Molecular Phylogenetics and Evolution* **55**, 753–764.

Wyrtki K (1966) Oceanography of the eastern equatorial Pacific Ocean. *Oceanography and Marine Biology, An Annual Review* **4**, 33–68.

Wyrtki K (1981) An Estimate of Equatorial Upwelling in the Pacific. *Journal of Physical Oceanography* **11**, 1205–1214.

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, Ding Z (2012) Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* **3**, 613–623.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* **292**, 686–693.

Zhou X, Li Y, Liu S, Yang Q, Su X, Zhou L, Tang M, Fu R, Li J, Huang Q (2013) Ultra-deep sequencing enables high-fidelity recovery of biodiversity for bulk arthropod samples without PCR amplification. *GigaScience* **2**, 4.