Chinese Journal of Oceanology and Limnology
Vol. 32 No. 4, P. 858-870, 2014
http://dx.doi.org/10.1007/s00343-014-3251-y

Zooplankton community analysis in the Changjiang River estuary by single-gene-targeted metagenomics*

CHENG Fangping (程方平)[1, 2], WANG Minxiao (王敏晓)[1], LI Chaolun (李超伦)[1], SUN Song (孙松)[1, **]

[1] *Key Laboratory of Marine Ecology and Environmental Sciences, Institute of Oceanology, Chinese Academy of Sciences, Qingdao 266071, China*

[2] *University of Chinese Academy of Sciences, Beijing 100049, China*

Received Sep. 18, 2013; accepted in principle Dec. 2, 2013; accepted for publication Jan. 3, 2014

Abstract DNA barcoding provides accurate identification of zooplankton species through all life stages. Single-gene-targeted metagenomic analysis based on DNA barcode databases can facilitate long-term monitoring of zooplankton communities. With the help of the available zooplankton databases, the zooplankton community of the Changjiang (Yangtze) River estuary was studied using a single-gene-targeted metagenomic method to estimate the species richness of this community. A total of 856 mitochondrial cytochrome oxidase subunit 1 (cox1) gene sequences were determined. The environmental barcodes were clustered into 70 molecular operational taxonomic units (MOTUs). Forty-two MOTUs matched barcoded marine organisms with more than 90% similarity and were assigned to either the species (similarity>96%) or genus level (similarity<96%). Sibling species could also be distinguished. Many species that were overlooked by morphological methods were identified by molecular methods, especially gelatinous zooplankton and merozooplankton that were likely sampled at different life history phases. Zooplankton community structures differed significantly among all of the samples. The MOTU spatial distributions were influenced by the ecological habits of the corresponding species. In conclusion, single-gene-targeted metagenomic analysis is a useful tool for zooplankton studies, with which specimens from all life history stages can be identified quickly and effectively with a comprehensive database.

Keyword: zooplankton; DNA barcodes; cytochrome oxidase subunit 1 (cox1)

1 INTRODUCTION

Zooplankton play important roles in marine ecosystems by linking primary productivity to higher trophic levels and mediating the flux of carbon and other chemical elements essential to life on earth (Harris et al., 2000). Recent evidence has suggested that zooplankton are sensitive indicators of global climate changes (Planque and Taylor, 1998; Beaugrand, 2009). Despite the importance of zooplankton, their long-term monitoring is limited because of their fragile nature, small body size, and the large number of taxa (Bucklin et al., 2010b). Sibling species may complicate the issue by underestimating biodiversity (Knowlton, 1993). Hence, morphological identification of zooplankton is expertise-dependent and time-consuming, and even impossible for some taxa (Carvalho et al., 2010).

DNA barcoding provides an alternative approach for identifying zooplankton at the species level, regardless of the condition and life history stages of the samples (Bucklin et al., 2011a; Li et al., 2011). The validity of the approach has been affirmed in several groups including copepods (Bucklin, 2003; Wang et al., 2011a), krill (Bucklin et al., 2007), arrow

* Supported by the National Natural Science Foundation of China (No. 41230963), the National Basic Research Program of China (973 Program) (No. 2011CB403604), the "135" Fund of Institute of Oceanology, Chinese Academy of Sciences (No. 2012I0060102), the Innovative Research Group Funding of the National Natural Science Foundation of China (No. 41121064), and the Strategic Priority Research Program of Chinese Academy of Sciences (No. XDA11020305)
** Corresponding author: sunsong@qdio.ac.cn
CHENG Fangping and WANG Minxiao contributed equally to this work.

worms (Jennings et al., 2010b), medusozoans (Ortman, 2008; Ortman et al., 2010), amphipods (Browne et al., 2007), and pelagic molluscs (Jennings et al., 2010a), among others. Such approaches will make accurate species identification easier for ecologists without taxonomic expertise (Valentini et al., 2009; Li et al., 2011). DNA barcoding can be used to complete the life histories of marine animals and reveal trophic interactions (Radulovici et al., 2010). Moreover, this approach provides the prerequisite to identify zooplankton species in the local marine ecosystem by single-gene-targeted metagenomic sequencing (Machida et al., 2009; Wang et al., 2011a).

However, the low efficiency of the universal primers and the incomplete zooplankton DNA databases underestimated zooplankton richness in earlier work (Machida et al., 2009). More tests are necessary to verify the feasibility of using single genes in detecting the species diversity of zooplankton communities. Recently, a more effective primer set was developed that performed well for a wide range of zooplankton (Wang et al., 2011a; Cheng et al., 2013). The increased zooplankton DNA barcode database made it possible to assign a MOTU (molecular operational taxonomic unit) to actual species.

The Changjiang (Yangtze) River estuary is a well-known fishing ground located at the junction of the East China and the Yellow Seas. Here we test the efficiency of a single-gene (DNA barcoding locus, cox1 partial sequences) targeted metagenomic approach on zooplankton community monitoring in this subtropical estuary with a relatively complex species composition. Furthermore, we try to give a preliminary evaluation of the new molecular methodology for zooplankton by comparing the results from both new molecular and traditional morphological methods.

2 MATERIAL AND METHOD

2.1 Sample collection and identification

Two zooplankton samples were collected using a zooplankton net (160-μm mesh, 0.316-m mouth diameter) in December 2010 through the water column at each site in the Changjiang River estuary (Fig.1). Two nets were bonded in a triangle frame so that the samples could be collected simultaneously. Four sites were sampled: station A at a depth of 14.5 m, station B at 35 m, station C at 5.8 m, and station D at 14.5 m. Two samples were collected at each site. One of the samples was preserved in 5% formaldehyde seawater

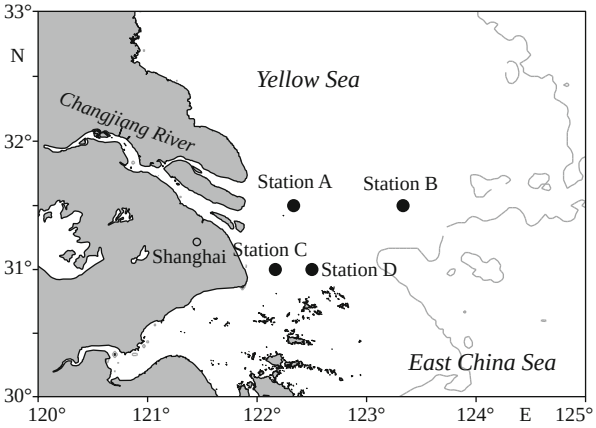

Fig.1 Sampling stations for zooplankton community studies in the Changjiang River estuary, China

Samples were collected in November 2010. Black dots indicate station locations.

for morphological identification to the lowest possible taxonomic level. The other sample was preserved in liquid nitrogen for molecular analysis. DNA extraction was performed on the entire nitrogen sample from each sampling site using an E.Z.N.A. HP Tissue DNA Maxi Kit (Omega bio-tek, USA; D5196) according to the manufacturer's instructions. A total of 600 μL of genomic DNA was generated for each station at a concentration of ~200 ng/μL. Zooplankton individuals were counted. The relative abundances (number of species individuals/sum of species individuals in a station) are given in Table 1.

2.2 Generation of DNA barcodes

Primers (CO318U: CTRATTGGTGGTTTYGGN-AAHTG and CO820L: CACTTCNGGGTGACCRA-ARAAYCA) developed in our laboratory (Wang et al., 2011b) were used to generate cox1 fragments as environmental barcodes. The PCR protocol was 94°C for 4 min, 35 cycles (94°C/40 sec, 47°C/1 min, 72°C/90 sec); finally, fragments were elongated at 72°C for 5 min. PCR amplification was confirmed by electrophoresis on ethidium bromide-stained 1.5% agarose gels. After purification using the E.Z.N.A. Gel extraction Kit (Omega bio-tek, D2500), the PCR products were cloned using a PMD-18T (TaKaRa Bio, Otsu, Japan; 6011) vector. Nine-hundred and twenty-six clones were sequenced by BGI (Beijing Genomics Institute).

2.3 Data analysis

Base calling and low-quality sequence trimming were performed with PHRED, and the reads were assembled in phrap with default parameters (Ewing

Table 1 Relative abundance and species composition (determined by morphological identification) of zooplankton communities from Changjiang River estuary

Species	Relative abundance (Total zooplankton:100)				Species	Relative abundance (Total zooplankton:100)			
	Station A	Station B	Station C	Station D		Station A	Station B	Station C	Station D
Calanus sinicus	0.749	2.780	0.290	2.822	*Harpacticoida*			0.145	0.056
Labidocera euchaeta	0.204		2.032	1.580	*Clytemnestra scutellata*		0.146	0.145	
Centropages dorsispinatus	0.477		0.145	6.659	*Macrosetella gracilis*	0.068	0.439		0.226
Parvocalanus crassirostris			3.628		Bivalve larva	1.090			0.734
Paracalanus aculeatus	41.894	30.878	14.949	11.512	Gammaridea			0.435	0.056
Paracalanus parvus	6.131				*Oikopleura longicauda*		1.463		0.959
Paracalanus sp.	36.717	33.220	66.328	66.930	*Oikopleura* sp.				0.734
Acartia spinicauda				0.226	Hyperiidae				0.056
Acartia hongi	0.068				*Zonosagitta nagae*	2.861	0.146		0.508
Acartia sp.		2.634	0.145	0.169	*Zonosagitta bedoti*	1.499	1.317		0.056
Tortanus vermiculus			1.306		*Flaccisagitta enflata*	0.068	0.146		
Centropages sinensis			0.726		*Sagitta* spp.	2.725	0.146	0.290	
Euchaeta concinna	0.068	0.439			Cumacea				0.508
Euchaeta rimana		0.293		0.056	Polychaeta larva	0.613	0.878	7.837	0.847
Euchaeta plana	0.341	0.146			*Hyperacanthomysis brevirostris*				0.113
Paraeuchaeta sp.	1.499	8.780		0.564	Mysidacea				0.282
Acrocalanus gracilis		2.634		0.225	*Euphausia pacifica*		0.146		
Acrocalanus longicornis		0.293			*Pseudeuphausia sinica*		0.146		
Acrocalanus sp.				0.113	Nauplius larva (Eupdausiacea)	0.068	1.902		0.395
Centropages dorsispinatus			0.145		*Euphysora* spp.				0.056
Lucicutia flavicornis	0.068				*Diphyes chamissonis*		1.317		0.056
Scolecithricella longispinosa	0.204	2.634			*Pleurobrachia globosa*	0.068		0.145	
Pseudodiaptomus sp.			0.145		*Aequorea conica*	0.068			
Nauplius (Copepoda)	0.136			1.637	*Solmundella bitentaculata*		0.146		
Corycaeus affinis	0.749	1.756	0.581	1.242	*Aglaura hemistoma*		0.049		
Oithona spp.	0.817	0.732		0.282	*Sagitella kowalewskii*		0.439		
Oithona plumifera		0.146			*Limacina trochiformis*		0.293		
Cyclopoidea			0.581		*Creseis clava*		0.146		
Oithona similis	0.749	1.902		0.282	*Agadina stimpsoni*				0.056
Oncaea venusta		0.585			Euphausiacea eggs		0.146		
Oithona sp.		0.732							

Sampling site locations are given in Fig.1. Order of taxa chosen to correspond with that in Table 2.

and Green, 1998; Ewing et al., 1998). All assembled sequences were manually verified in CONSED (Gordon et al., 1998) to remove misassemblies. The chimeras were removed by Mothur (Schloss et al., 2009). The translated amino acids were aligned and returned to DNA sequences in Mega v.5 (Tamura et al., 2011) with default parameters. Sequences with internal termination codons were regarded as pseudogenes and were abandoned in the following analysis. All the cox1 sequences obtained were submitted to GenBank (KC731592–KC732449). The complete alignment with high-quality conserved sequences was trimmed to a length of 470 bp. Pairwise p-distance was calculated between all DNA barcodes using PAUP v.4 (Swofford, 1993). Clusters with an affinity above 95% were accepted as a MOTU for all

sequences except those belonging to Ctenodontina (94%), in which higher intraspecific divergences have been observed (Jennings et al., 2010; Miyamoto et al., 2010). Clustering was carried out in Mothur (Schloss et al., 2009) to generate MOTUs by the "cluster" command. A BLASTN Search against a dataset containing all cox1 sequences from GenBank and the zooplankton DNA barcode database (www.zooplankton.cn) with default settings was performed. Sequences giving no hits to the known cox1 sequences were removed. The results were then used to infer the taxonomic position of the queried sequences with the following criteria. If the BLASTN score was more than 350 and the BLASTN similarity was above 95%, the MOTU was assigned as the same species. If the score was more than 300 and similarity was above 90%, the MOTU was assigned as a species in the same genus. If the BLASTN score was more than 200 and the BLASTN similarity was above 80%, the sequence was assigned to the higher taxonomic group of the related species. Otherwise, the sequences were labeled unclassified. Rarefaction curves were drawn for each station by rarefaction.single in Mothur to determine if the coverage of sequences over the clone library was sufficient. Richness (Chao1) for each station and dissimilarity (Thetayc) among different sites were calculated in Mothur by summary.single and dist.shared commands separately. Definitions of Chao1 and Thetayc are given below.

$$S_{\text{chao 1}}=S_{\text{obs}}+[n_1(n_1-1)/2(n_2+1)],$$

$S_{\text{chao 1}}$=the estimated richness; S_{obs}=the observed number of species; n_1=the number of OTUs with only one sequence (i.e. singletons); n_2=the number of OTUs with two sequences (i.e. doubletons).

$$D_{\theta YC}=1-\frac{\sum_{i=1}^{S_T} a_i b_i}{\sum_{i=1}^{S_T}(a_i-b_i)^2+\sum_{i=1}^{S_T} a_i b_i},$$

where S_T=the total number of OTUs in communities A and B; a_i=the relative abundance of OTU i in community A; b_i=the relative abundance of OTU i in community B.

A parsimony-based test was carried out to check the significance of the dissimilarity among different sites by parsimony in Mothur. Indicator vector analysis was carried out following Sirovich et al. (2009) in MATLAB (2012a) to visualize similarities and relationships between haplotypes. A neighbor-joining tree was generated with default parameters in Mega v.5 (Tamura et al., 2011) to illustrate the phylogenetic positions of the MOTUs.

3 RESULT

3.1 Zooplankton composition based on morphological analysis

The zooplankton species composition inferred from environmental barcodes was compared with those from traditional morphological examinations at the same stations by zooplankton taxonomists (Table 1). Sixty-one species were identified at four stations, with 33 at station A, 25 at station B, 33 at station C, and 19 at station D. The abundance of small copepods, including *Paracalanus* species and *Oithona similis*, dominated all samples. High abundances of *Calanus sinicus* and Ctenodontina species were also observed. Many species, such as *Scolecithricella longispinosa*, occurred only once in all of the samples.

3.2 Clustering by MOTUs

After removing problematic sequences including pseudogenes (Bensasson et al., 2001) and chimera sequences, the final alignments comprised 856 environmental barcodes ~470 bp in length. Pairwise divergences (PWD, represented by *p*-distance) of the combined dataset were calculated. The mismatch distribution of the PWD revealed a high frequency of very small (<0.05) genetic distance sequence pairs (Fig.2) that were separated from the larger genetic distance (>0.1) pairs by gaps. The species richness index Chao1 and the number of species estimated decreased continuously from 0 to 0.05, and were consistent from 0.05 to 0.065 (Fig.3). The level of 0.05 was considered the threshold for coalescence. This value is similar to the intraspecific divergence found from barcoding studies based on environmental samples in Jiaozhou Bay (Wang et al., 2011a).

3.3 Species composition analysis based on environmental barcodes

We clustered 856 environmental barcodes into 70 MOTUs according to the specified criteria (Table 2). Most of the MOTUs were rare or narrowly distributed. More than half of them occurred in fewer than two sites. Fifty species were only identified in one site. In accordance with the morphological results, copepods were the predominant zooplankton, represented by 593 clones belonging to 18 MOTUs. The top four MOTUs contained 207, 139, 130, and 61 sequences. None of the rarefaction curves (Fig.4) reached an asymptote, indicating insufficient sequencing for a full representation of diversity (e.g. station A and C

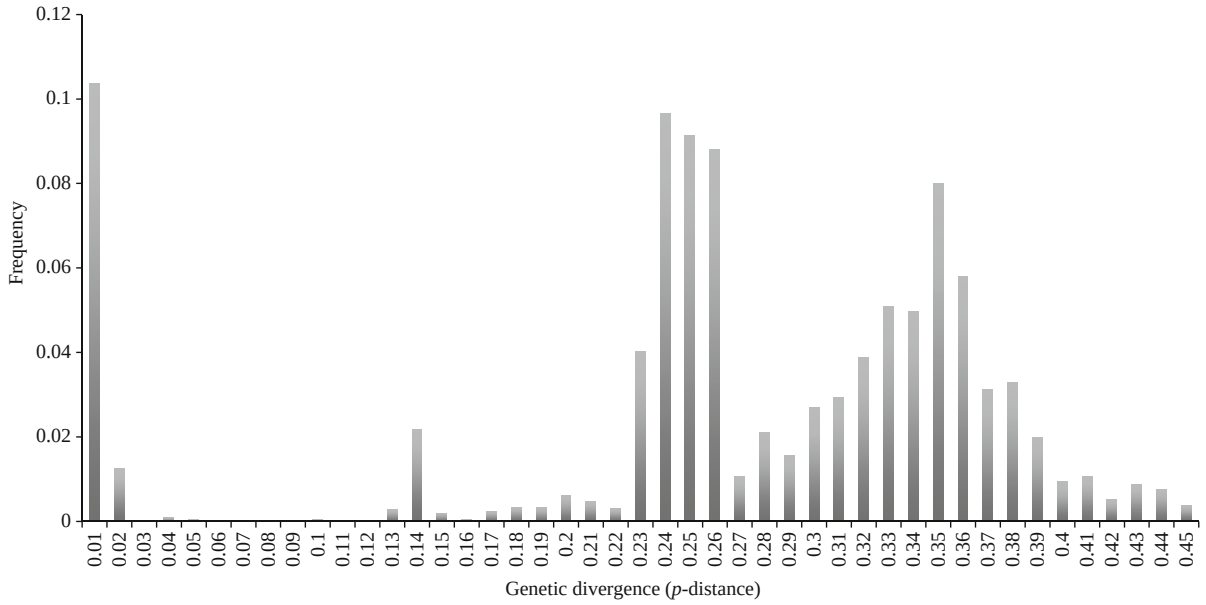

Fig.2 Mismatch distribution of the pairwise genetic distances (*p*-distance) for the 856 environmental barcodes from zooplankton communities in the Changjiang River estuary

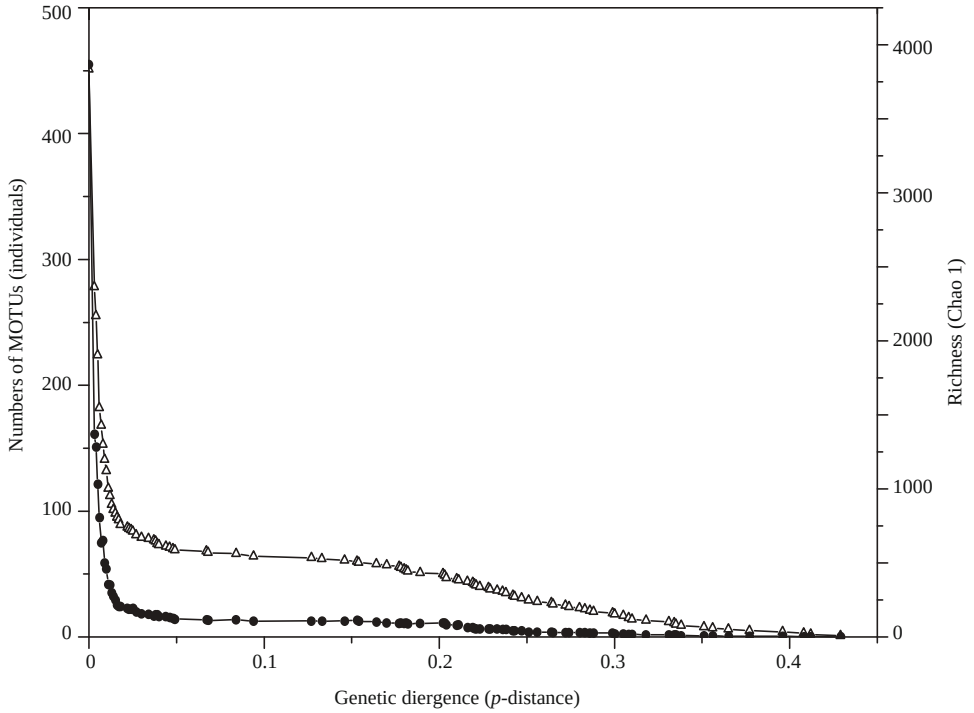

Fig.3 The numbers of MOTUs (triangles) and species richness (Chao 1, black dots) estimated at different genetic divergences (*p*-distance) for zooplankton communities in the Changjiang River estuary

Numbers on the left *y*-axis indicate the number of MOTUs estimated. Numbers on the right *y*-axis are the Chao 1 values.

for example, coverage<0.95). However, more taxonomic units were identified by this method than by morphological analysis.

The environmental barcodes were searched against the database using BLASTN. Their phylogenetic affiliation with the published DNA barcodes was determined by construction of a phylogenetic tree (Fig.5) and indicator vector analysis (Fig.6). In total, 60% of the MOTUs were affiliated with the barcoded marine organisms with more than 90% similarity. These MOTUs could be linked with known marine invertebrates at either the species (similarity>96%) or

Table 2 BLASTN search results for environmental barcodes of MOTUs (molecular operational taxonomic units) against the sequences in the GenBank zooplankton DNA barcode database

OTU ID	Clone access number	Expect	Score	Similarity	Lowest taxonomy	Acc. for closely related sp.	Vector ID	Station A	Station B	Station C	Station D
OTU-01	KC731661	0	463	1.00	*Calanus sinicus*	HQ619228	16	22	151	15	19
OTU-02	KC731943	0	463	1.00	*Centropages dorsispinatus*	EU599519	56	14	0	15	110
OTU-03	KC732094	0	463	1.00	*Paracalanus aculeatus*	EU856807	47	3	4	38	85
OTU-04	KC731979	0	463	1	*Bestiolina* sp.	KC784343	25	1	0	50	10
OTU-05	KC732271	0	454	0.99	*Creseis acicula*	HM045333	11	0	25	0	0
OTU-06	KC732206	0	463	1.00	*Hyperacanthomysis longirostris*	HM045290	68	0	0	18	5
OTU-08	KC732293	0	454	0.99	*Salanx ariakensis*	HM151583	64	0	0	18	0
OTU-07	KC732364	0	369	0.96	*Sagitta bipunctata*	JN258007	15	3	15	0	0
OTU-09	KC732356	0	442	0.99	*Zonosagitta nagae*	AP011545	10	10	3	2	0
OTU-11	KC732411	4.36E-120	238	0.84	Bacillariophyta species	AB706233	21	0	1	8	5
OTU-10	KC732443	0	373	0.94	*Metridium* sp.	U36783	60	6	0	0	8
OTU-14	KC732311	0	460	1	*Sus scrofa*	EF545593	14	0	0	12	0
OTU-12	KC732241	2.76E-82	170	0.8	Stomatopoda species	HM138780	4	0	0	12	0
OTU-13	KC731953	3.37E-121	240	0.84	Euchaetae species	JQ819825	41	0	0	12	0
OTU-15	KC732389	0	466	1	*Muggiaea atlantica*	JQ353741	42	8	4	0	0
OTU-18	KC732230	0	463	1	*Pseudeuphausia sinica*	AY947487	37	5	3	1	0
OTU-16	KC732179	0	454	0.99	*Subeucalanus crassus*	HM045347	26	0	9	0	0
OTU-17	KC731805	0	445	0.99	*Labidocera euchaeta*	HM045392	6	0	0	7	2
OTU-19	KC732425	0	463	1	*Aequorea conica*	JQ353765	28	8	0	0	0
OTU-20	KC732043	0	395	0.96	*Bestiolina* sp.	KC784343	9	0	0	5	3
OTU-21	KC732336	0	380	0.97	*Zonosagitta bedoti*	JN258003	13	4	0	0	2
OTU-22	KC732339	0	360	1	*Zonosagitta bedoti*	DQ862800	63	6	0	0	0
OTU-23	KC732214	3.23E-151	294	0.88	Mysidae species	HM045290	1	0	0	4	1
OTU-24	KC732246	1.69E-64	138	0.78	Unclassified	DQ230111	69	0	0	3	1
OTU-25	KC731974	0	463	1	*Paracalanus parvus*	EU856802	7	1	0	0	3
OTU-26	KC732251	2.03E-118	235	0.84	Mollusca species	FJ876888	20	0	4	0	0
OTU-27	KC731963	0	463	1	*Euchaeta plana*	HM045309	62	2	2	0	0
OTU-28	KC731968	0	448	0.99	*Tortanus vermiculus*	JN605791	17	0	0	4	0
OTU-29	KC732396	9.11E-142	277	0.86	Diphyidae species	GQ119973	32	1	3	0	0
OTU-30	KC732301	0	349	0.92	*Benthosema* sp.	AP012260	55	0	4	0	0
OTU-31	KC731948	0	460	1	*Scolecithricella longispinosa*	HM045346	58	0	3	0	0
OTU-32	KC732223	0	445	0.99	*Iiella pelagica*	HM045339	57	0	3	0	0
OTU-33	KC732321	0	444	0.99	*Cypridina nana*	HM045340	50	3	0	0	0
OTU-34	KC732187	0	463	1	*Corycaeus affinis*	HQ848872	5	0	1	1	1
OTU-35	KC732380	1.35E-45	104	0.75	Unclassified	HQ024438	39	0	2	0	0
OTU-36	KC732399	2.16E-73	154	0.78	Unclassified	FJ949002	61	0	0	2	0
OTU-37	KC732433	0	463	1	*Corymorpha bigelowi*	JQ353733	23	2	0	0	0
OTU-38	KC732431	0	415	1	*Nanomia bijuga*	JQ716071	66	0	2	0	0
OTU-39	KC732418	9.04E-147	286	0.87	Naviculaceae species	HQ317076	35	0	0	2	0

Spatial distributions are also given for each MOTU. Corresponding vector IDs for each MOTU are listed. Sampling site locations are given in Fig.1. Acc: GenBank accession number. Hit scores (expect, score, similarity) are also given.

To be continued

Table 2 Continued

OTU ID	Clone access number	Expect	Score	Similarity	Lowest taxonomy	Acc. for closely related sp.	Vector ID	Station A	Station B	Station C	Station D
OTU-40	KC732330	0	419	0.97	*Flaccisagitta enflata*	KC784346	30	0	2	0	0
OTU-41	KC732318	0	395	1	*Crassostrea* sp.	HM003526	36	0	0	0	2
OTU-42	KC732219	0	445	0.99	*Paradorippe granulata*	EU636974	31	0	2	0	0
OTU-43	KC732314	2.35E-13	46	0.85	Unclassified	AY376998	33	0	0	0	2
OTU-44	KC732325	4.39E-115	229	0.84	*Nemertea species*	HQ848621	34	0	2	0	0
OTU-45	KC732279	0	376	0.94	*Creseis* sp.	FJ876888	29	0	1	0	0
OTU-46	KC732382	1.73E-49	111	0.76	Unclassified	JQ711382	24	0	0	0	1
OTU-47	KC732329	5.76E-104	209	0.82	Polychaeta species	GU014062	46	0	0	1	0
OTU-48	KC732327	9.50E-112	223	0.83	Brachiopoda species	AB621915	22	1	0	0	0
OTU-49	KC732328	2.64E-112	224	0.84	Mollusca species	DQ207350	27	0	1	0	0
OTU-50	KC732324	2.90E-47	107	0.75	Unclassified	JN009913	48	0	0	1	0
OTU-51	KC732363	2.44E-167	323	0.94	*Sagitta bipunctata*	JN258007	49	1	0	0	0
OTU-52	KC732233	0	451	0.99	*Euphausia pacifica*	HQ700929	51	0	1	0	0
OTU-53	KC732323	1.60E-104	210	0.82	Mollusca species	HQ380202	52	0	0	1	0
OTU-54	KC732415	4.24E-140	274	0.86	Bacillariophyta species	AB020223	53	1	0	0	0
OTU-55	KC732417	4.45E-105	211	0.83	Bacillariophyta species	FN557039	54	0	0	1	0
OTU-56	KC732416	9.84E-87	178	0.79	Unclassified	AB020223	38	0	0	1	0
OTU-57	KC732420	2.06E-108	217	0.83	Bacillariophyta species	AB706216	18	0	0	0	1
OTU-58	KC732421	9.30E-127	250	0.85	Bacillariophyta species	AB020223	59	0	0	1	0
OTU-59	KC732422	2.07E-103	208	0.82	Oomycetes species	EF408874	40	0	0	1	0
OTU-60	KC732189	0	460	1	*Oithona similis*	JN230869	67	0	1	0	0
OTU-61	KC732435	0	397	0.99	*Nemopsis bachei*	JQ716072	19	0	0	1	0
OTU-62	KC732332	8.49E-93	169	0.84	Ctenodontina species	KC784346	12	0	1	0	0
OTU-63	KC732316	8.98E-152	295	0.88	Bacteria	CP000157	44	0	0	1	0
OTU-64	KC732190	6.31E-39	92	0.75	Unclassified	JQ390574	65	0	0	0	1
OTU-65	KC731971	0	424	0.97	*Pseudodiaptomus poplesia*	AF536521	8	0	0	1	0
OTU-66	KC732379				Unclassified	Unclassified	3	0	1	0	0
OTU-67	KC732317	0	460	1	*Temnopleurus reevesii*	JN128630	45	0	1	0	0
OTU-68	KC731950	0	460	1	*Euchaeta concinna*	HM045350	2	0	1	0	0
OTU-69	KC732045	1.45E-174	336	0.91	*Paracalanus* sp.	EU856801	70	0	0	1	0
OTU-70	KC732176	1.88E-173	334	0.91	*Paracalanus* sp.	EU856801	43	0	1	0	0

genus level (similarity<96%). With the exception of eight MOTUs that exhibited extremely low similarities (<80%) with known barcodes, the others were assigned to higher taxonomic levels. Seventeen taxonomic groups were identified, including Copepoda (18 species), Medusae (seven species), and Mollusca (six species), among others. The taxonomic distributions revealed by the environmental barcodes were similar to those found by morphological analysis. Species that occurred with high frequency in the morphological analysis were all recovered. Some swimming (e.g. *Salanx ariakensis*) and benthic species (e.g. *Metridium* sp.) that were usually absent with the morphological identification method were identified by the molecular method.

All MOTUs were grouped as short branches (Fig.5) or red squares (indicator vector) (Fig.6). The affinities between the MOTUs and the zooplankton barcodes were confirmed. The correct rate for the assignment was confirmed to be 100%. In addition, intraspecific divergences for chaetognath species were much larger than for other taxa.

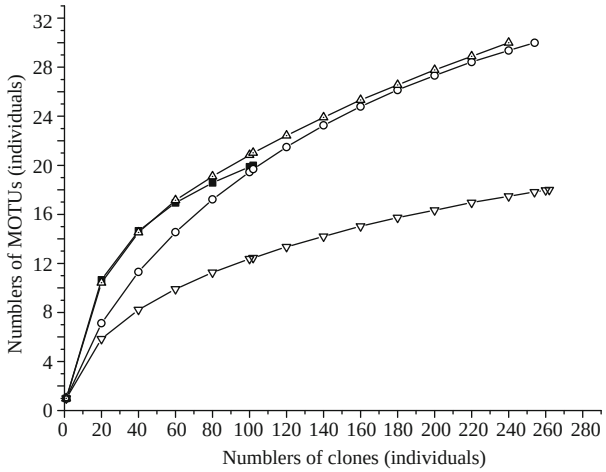

Fig.4 Rarefaction curves generated for environmental barcodes for zooplankton communities in the Changjiang River estuary

Station A, B, C, and D are represented by black squares, hollow circles, hollow triangles, and inverted triangles, respectively.

3.4 Community structure and similarity

Coverage, richness, and diversity were estimated for the four stations (Table 3). High values of coverage were calculated for all stations, indicating that an adequate estimate of species composition could be made. In accordance with the morphological results, richness was highest at station C (with a depth of only 5.8 m) and lowest at station D. More species occurred in the pelagic zones.

Differences in community composition and structure were evaluated among the four stations using an abundance-based approach (Thetayc). The stations clustered into two groups based on the dissimilarity matrix (Fig.7). Sites from the same latitudes assembled together. Use of the parsimony method to test dissimilarity among stations revealed significantly different community structures (P<0.000 1).

4 DISCUSSION

4.1 Performance of new primer sets in zooplankton community studies

Previous work by Machida et al. (2009) has confirmed that single-gene-targeted metagenomic sequencing can be a powerful tool for estimating zooplankton species richness. However, the low binding efficiency of the universal primers developed by Folmer et al. (1994) prohibited either successful or efficient amplification of barcodes for some taxonomic groups (Crandall, 2009; Bucklin et al., 2010b). The priming heterogeneity among taxonomic groups leads to a biased estimation of zooplankton community richness (Machida et al., 2009). During laboratory work to accumulate zooplankton DNA barcodes, amplification success rates were low (less than 70%), especially for Tunicata and Ctenophora (Cheng et al., 2013 and unpublished data). The newly developed primers exhibited excellent taxonomic compatibility by generating barcodes for 17 taxonomic groups, and showing excellent amplification rates (95%) for all taxa (unpublished data). Hence, the primers used in this study should reduce the risk of underestimation of species richness. However, richness for *Oikopleura* sp. seemed underestimated, which probably resulted from low binding efficiency for the group. Other regions such as the internal transcribed spacer region (ITS) should be tested.

Table 3 Richness sample coverage at four stations in the Changjiang River estuary based on MOTUs (molecular operational taxonomic units)

Group	Sobs	Richness (Chao 1)	Coverage
Station A	20	25.00	0.94
Station B	30	37.86	0.96
Station C	30	49.50	0.94
Station D	18	21.00	0.98

Sobs: Species observed.

4.2 Performance of the molecular method to determine zooplankton richness

Fewer species were identified by morphological analysis than by single-gene-targeted metagenomic analysis (Table 1). The taxonomic distributions revealed by environmental barcodes were similar to those found by morphological methods, indicating that the molecular method provided accurate profiling of this zooplankton community. Species occurring at high frequencies in the morphological analysis were all recovered by DNA analysis. Minor differences were found in the species composition of low-abundance zooplankton. These differences can be explained by the limited number of clones that were sequenced, systematic error during sample collection, underestimated larva/egg diversity, and unknown contaminants such as zooplankton gut contents and organic debris. Surprisingly, we found DNA barcodes for the pig (*Sus scrofa*) at station C, which likely came from food waste discarded into the sea by the many fishing boats there.

The absence of some rare species (e.g. *Oikopleura*

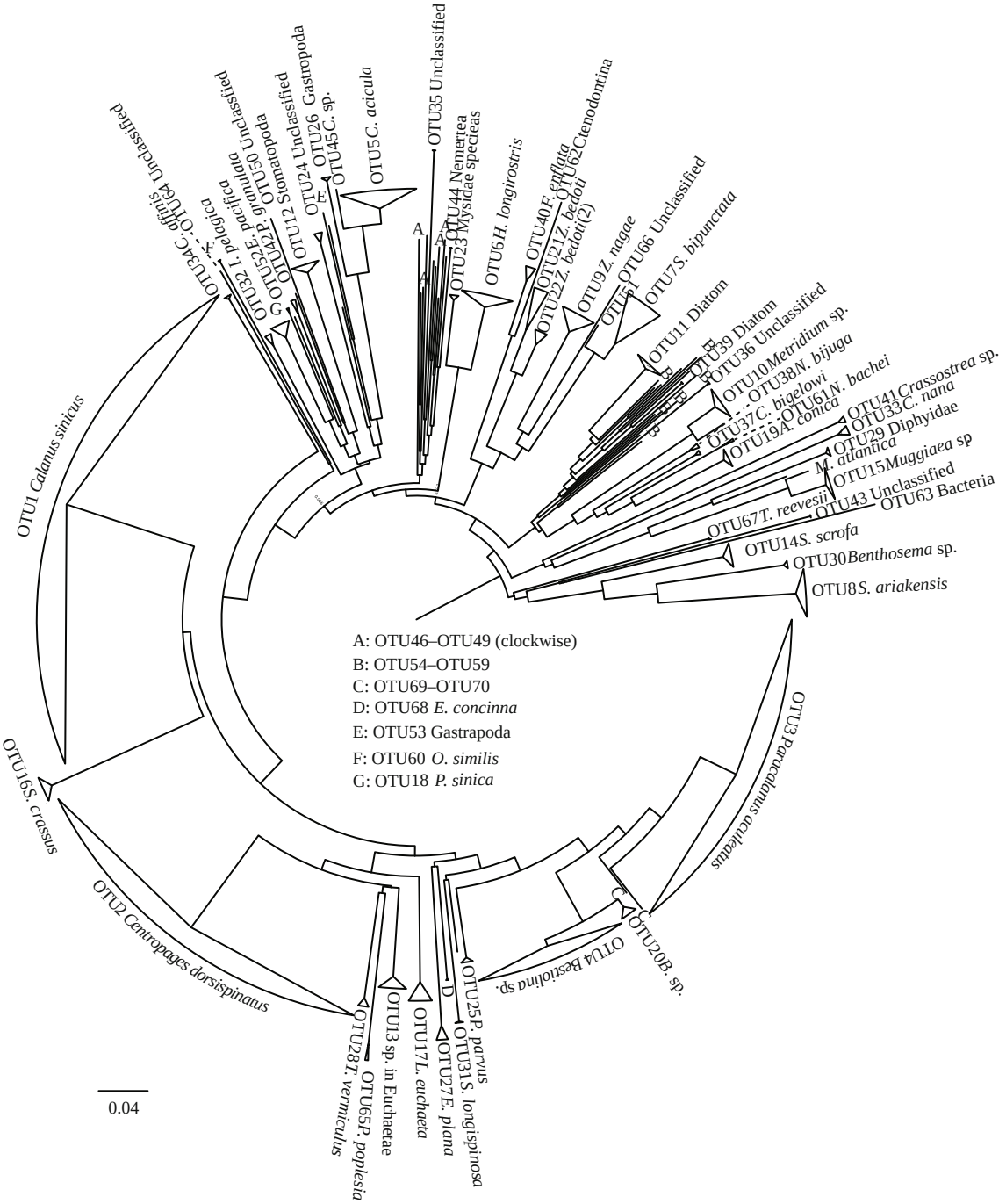

Fig.5 Neighbor-joining phylogenetic tree of the cox1 sequences recovered from both the marine environmental barcodes of zooplankton in the Changjiang River estuary and their close relatives as revealed by BLASTN analysis

The scale bar corresponds to a 4% difference. Numbers of MOTUs (molecular operational taxonomic units) are represented by the width of the tips. The depth of the tips indicates intraspecific divergences. Species names associated with OUT identifications are given in Table 2.

spp.) in the molecular methods may be due to technical deficiencies such as low priming efficiency (Machida et al., 2009) for some taxa. Because the molecular and morphological analyses were carried out on different samples, zooplankton patchiness may also have contributed to the absence of rare species. Although new primer sets have been developed to fit diverse taxonomic groups, bias for certain species seems unavoidable (Machida et al., 2009). Simulated experiments are needed to assess the performance of the current system in diversity analysis. Many species absent in the morphological results were identified by molecular methods, especially for gelatinous zooplankton and merozooplankton (animals that

spend only part of their life cycle in the plankton). *Salanx ariakensis*, which was recorded as a fish larva by morphology, was recognized by molecular analysis at station C where the species has been reported previously (Sun et al., 1994; Hua et al., 2009). Sibling species were successfully distinguished by molecular methods. *Bestiolina* sp., which is morphologically similar to *Paraclanus parvus*, was mistakenly identified by microscopy as a *Paracalanus* sp., but correctly identified by molecular methods.

The planktonic life history stages of merozooplankton have been assigned to species using molecular methods. These observations highlight the fact that this type of analysis enables estimation of larval dynamics, which is almost impossible by morphology alone (Ko et al., 2013). Ecological studies on marine larvae have repeatedly emphasized their pivotal role in elucidating the patterns and processes that influence marine populations, communities, and ecosystems (Uye et al., 2002; Cowen et al., 2006). Almost all larvae could be distinguished by the molecular method and a comprehensive database of DNA barcodes. Based on our comparison of the two methods, the single-gene-targeted metagenomic method was confirmed reliable for zooplankton species composition studies. It can also provide higher resolution for zooplanktonic larval studies.

4.3 Community structure

Copepods dominated all of the samples in terms of species richness and abundance, as found in other zooplankton community studies using morphological methods (Liu, 2012). MOTU spatial distributions were related to the ecological habits of the corresponding species. MOTUs representing high-salinity pelagic species (*Subeucalanus crassus*, *Sagitta bipunctata*, *Creseis clava*, and *Scolecithricella longispinosa*) appeared at station B; MOTUs representing estuarine low-salinity species (*Pseudodiaptomus poplesia*, *Tortanus vermiculus*, and *Sinocalanus sinensis*) occurred at stations C and D. MOTUs of euryhaline species such as *Calanus sinicus*, *Paracalanus aculeatus*, and *P. parvus* were present at all stations. In the station with the lowest chlorophyll *a* (station C, unpublished data), species richness was also lowest as inferred by both morphological and molecular methods.

MOTUs representing *Sagitta bipunctata* in the northern transect (31.5°N) were observed in our study. *S. bipunctata* is considered a warm species

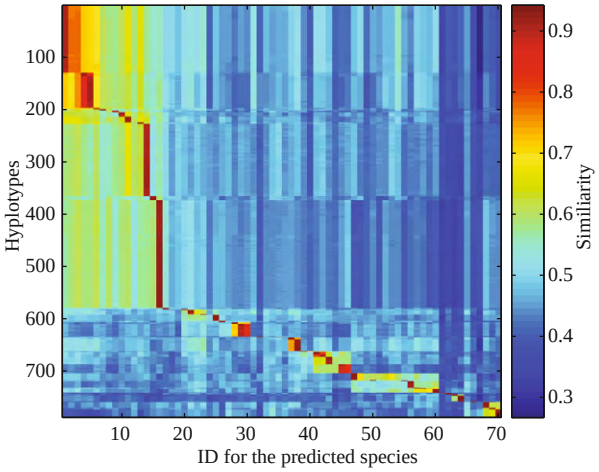

Fig.6 Vector analysis of 856 barcodes belonging to 70 MOTUs (molecular operational taxonomic units)

Results are shown as a Klee diagram. Different colored bar scales emphasize off diagonal resemblance between matrices. Similarity increases from blue to red. MOTU IDs represented by the numbers on the *x*-axis are given in Table 1.

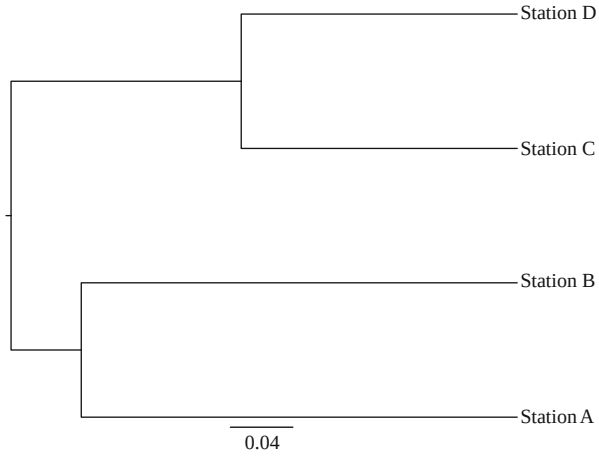

Fig.7 Cluster analysis for the zooplankton community at four stations in the Changjiang River estuary

Tree lengths represent differences between stations. Positions of sampling stations are given in Fig.1.

transported by the high-salinity Kuroshio currents. This species had not been recorded in these locations before (Lin, 1985; Xiao, 2004). More environmental parameters should be measured to elucidate the possible reasons, such as global warming or invasive species, for the appearance of this species.

4.4 Insights from high intraspecific variations

Closer examination of intraspecific variation will lead to a better understanding of cryptic species and geographic distribution of lineages and phylogeography (Dawson et al., 2001; Baird et al.,

2011). Chaetognath species exhibited unusually high intraspecific divergence, as in previous studies (Miyamoto et al., 2010; Wang et al., 2011b; Miyamoto et al., 2012). The disjunctive distribution of genetic distances indicated species for further morphological examination. Although collected at adjacent sites, the MOTUs representing *Zonosagitta bedoti* clustered into two clades, which suggests discrete lineages. Large genetic divergence was found in *Sagitta bipunctata* for the corresponding MOTUs, which overlapped, similar to the genetic structure of *Aidanosagitta crassa* (Wang et al., 2011b). The presence of significant genetic diversity without geographic structure could indicate reproductive mixing of different haplotypes or insufficient time for lineage sorting in isolated populations (Jennings et al., 2010b).

4.5 Prospect

When a comprehensive database is available, high-throughput techniques like next-generation sequencing (Creer et al., 2010) and microarrays (Kochzius et al., 2008; Lee et al., 2011) can be used for species identification with high accuracy and efficiency. Furthermore, real-time quantitative PCR using specific primers can be used for the accurate quantification of species abundance (Bucklin et al., 2011). These tools will aid marine ecologists to uncover trophic relationships, invasive species, and historical range expansion, and will facilitate population genetic and biogeographic analyses (Valentini et al., 2009). The application of DNA barcoding will provide more information; it will also improve our understanding of zooplankton biodiversity and their functions in marine ecosystems (Li et al., 2011).

5 CONCLUSION

Owing to the boosting of the DNA barcoding project in China, a zooplankton DNA barcode database has been constructed. The database and single-gene-targeted metagenomic sequencing were applied in combination to environmental zooplankton net samples in the Changjiang River estuary. It was possible to determine the zooplankton species composition regardless of the condition or developmental stages of the target species. Compared with the molecular approach, species richness tended to be underestimated by microscopic analysis, especially for gelatinous zooplankton and planktonic larvae. Our results confirm that the molecular approach is a reliable method for zooplankton species composition determination. The zooplankton community structure differed significantly among all stations. MOTU spatial distributions corresponded to the ecological habits of the corresponding species.

6 ACKNOWLEDGEMENT

We thank the crew of the R/V *Science 3* for their assistance in sample collection. We thank DAI Luping, GONG Han, and WANG Rencheng for help in the laboratory.

References

Baird H P, Miller K J, Stark J S. 2011. Evidence of hidden biodiversity, ongoing speciation diverse patterns of genetic structure in giant Antarctic amphipods. *Molecular Ecology*, **20**(16): 3 439-3 454.

Beaugrand G. 2009. Decadal changes in climate and ecosystems in the North Atlantic Ocean and adjacent seas. *Deep-Sea Research, Part II*, **56**(8-10): 656-673.

Bensasson D, Zhang D X, Hartl D L, Hewitt G M. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, **16**(6): 314-321.

Browne W E, Haddock S H D, Martindale M Q. 2007. Phylogenetic analysis of lineage relationships among hyperiid amphipods as revealed by examination of the mitochondrial gene, cytochrome oxidase I (COI). *Integrative and Comparative Biology*, **47**(6): 815-830.

Bucklin A, Frost B W, Bradford-Grieve J, Allen L D, Copley N J. 2003. Molecular systematic phylogenetic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. *Marine Biology*, **142**(2): 333-343.

Bucklin A, Hopcroft R R, Kosobokova K N, Nigro L M, Ortman B D, Jennings R M, Sweetman C J. 2010a. DNA barcoding of Arctic Ocean holozooplankton for species identification andrecognition. *Deep-Sea Research, Part II*, **57**(1-2): 40-48.

Bucklin A, Ortman B, Jennings R, Nigro L, Sweetman C, Copley N, Sutton T, Wiebe P. 2010b. A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep-Sea Research, Part. II*, **57**(24-26): 2 234-2 247.

Bucklin A, Steinke D, Blanco-Bercial L. 2011. DNA Barcoding of marine metazoa. *Annual Review of Marine Science*, **3**: 471-508.

Bucklin A, Wiebe P H, Smolenack S B, Copley N J, Beaudet J G, Bonner K G, Farber-Lorda J, Pierson J J. 2007. DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). *Journal of Plankton Research*, **29**(6): 483-493.

Carvalho G, Creer S, Allen M J, Costa F, Tsigenopoulos C, Le Goff-Vitry M, Magoulas A, Medlin L, Metfies K. 2010.

Genomics in the discovery and monitoring of marine biodiversity. Introduction to Marine Genomics. p.1-32.

Cheng F P, Wang M X, Sun S, Li C L, Zhang Y S. 2013. DNA barcoding of Antarctic marine zooplankton for species indetification and recognition. *Advances in Polar Science*, **24**(2): 119-127.

Cowen R K, Paris C B, Srinivasan A. 2006. Scaling of connectivity in marine populations. *Science*, **311**(5760): 522-527.

Crandall K. 2009. Workshops Report and Recommendations, DNA barcoding of Marine Biodiversity, http://www.cmarz.org/barcode/pdf/MarBOL_Workshops_Report_06jul09.pdf. Accessed on 2009-07-06.

Creer S, Fonseca V G, Porazinska D L, Giblin-Davis R M, Sung W, Power D M, Packer M, Carvalho G R, Blaxter M L, Lambshead P J D, Thomas W K. 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. *Molecular Ecology*, **19**(s1): 4-20.

Dawson M N, Jacobs D K. 2001. Molecular evidence for cryptic species of *Aurelia aurita* (Cnidaria, Scyphozoa). *Biological Bulletin*, **200**(1): 92-96.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Research*, **8**(3): 186-194.

Ewing B, Hillier L, Wendl M C, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research*, **8**(3): 175-185.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**(5): 294-299.

Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Research*, **8**(3): 195-202.

Harris R, Wiebe P, Lenz J, Skjoldal H, Huntley M. 2000. ICES Zooplankton Methodology Manual. Academic Press, San Diego, US.

Hua X, Wang W, Yin W, He Q, Jin B, Li J, Chen J, Fu C. 2009. Phylogeographical analysis of an estuarine fish, *Salanx ariakensis* (Osmeridae: Salanginae) in the north-western Pacific. *Journal of Fish Biology*, **75**(2): 354-367. (in Chinese with English abstract)

Jennings R M, Bucklin A, Ossenbrügger H, Hopcroft R R. 2010a. Species diversity of planktonic gastropods (Pteropoda and Heteropoda) from six ocean regions based on DNA barcode analysis, *Deep-Sea Research, Part II*, **57**(24-26): 2 199-2 210.

Jennings R M, Bucklin A, Pierrot-Bults A. 2010b. Barcoding of arrow worms (Phylum Chaetognatha) from three oceans: genetic diversity and evolution within an enigmatic phylum. *PloS One*, **5**(3): e9949.

Knowlton N. 1993. Sibling species in the sea. *Annual Review of Ecology Evolution and Systematics*, **24**: 189-216.

Ko H L, Wang Y T, Chiu T S, Lee M A, Leu M Y, Chang K Z, Chen W Y, Shao K T. 2013. Evaluating the accuracy of morphological identification of larval fishes by applying DNA barcoding. *PloS One*, **8**(1): e53451.

Kochzius M, Nolte M, Weber H, Silkenbeumer N, Hjorleifsdottir S, Hreggvidsson G O, Marteinsson V, Kappel K, Planes S, Tinti F, Magoulas A, Vazquez E G, Turan C, Hervet C, Falgueras D C, Antoniou A, Landi M, Blohm D. 2008. DNA microarrays for identifying fishes. *Marine Biotechnology*, **10**(2): 207-217.

Lee G, Park S Y, Hwang J, Lee Y H, Hwang S Y, Lee S, Lee T K. 2011. Development of DNA chip for jellyfish verification from South Korea. *Biochip Journal*, **5**(4): 375-382.

Li C L, Wang M X, Cheng F P, Sun S. 2011. DNA barcoding and its application to marine zooplankton ecology. *Biodiversity Science*, **19**(06): 805-814. (in Chinese with English abstract)

Lin Y. 1985. Vertical distribution of chaetognaths in the East China Sea in summer of 1978. *Acta Ecologica Sinica*, **5**(2): 175-186. (in Chinese with English abstract)

Liu Z S. 2012. Community Structure and Biodiversity of Zooplankton in the Changjiang Estuary and Its Adjacent Waters. PhD dissertation, Ocean University of China. 154p. (in Chinese with English abstract)

Machida R J, Hashiguchi Y, Nishida M, Nishida S. 2009. Zooplankton diversity analysis through single-gene sequencing of a community sample. *BMC Genomics*, **10**(1): 438.

Miyamoto H, Machida R J, Nishida S. 2010. Genetic diversity and cryptic speciation of the deep sea chaetognath *Caecosagitta macrocephala* (Fowler, 1904). *Deep-Sea Research, Part II*, **57**(24): 2 211-2 219.

Miyamoto H, Machida R J, Nishida S. 2012. Global phylogeography of the deep-sea pelagic chaetognath *Eukrohnia hamata*. *Progress in Oceanography*, **104**: 99-109.

Ortman B. 2008. DNA Barcoding the Medusozoa and Ctenophora. PhD dissertation, University of Connecticut. 134p.

Ortman B, Bucklin A, Pages F, Youngbluth M. 2010. DNA Barcoding the Medusozoa using mtCOI. *Deep-Sea Research, Part II*, **57**(24): 2 148-2 156.

Planque B, Taylor A H. 1998. Long-term changes in zooplankton and the climate of the North Atlantic. *ICES Journal of Marine Science*, **55**(4): 644-654.

Radulovici A E, Archambault P, Dufresne F. 2010. DNA barcodes for marine biodiversity: moving fast forward? *Diversity*, **2**(4): 450-472.

Schloss P D, Westcott S L, Ryabin T, Hall J R, Hartmann M, Hollister E B, Lesniewski R A, Oakley B B, Parks D H, Robinson C J, Sahl J W, Stres B, Thallinger G G, Van Horn D J, Weber C F. 2009. Introducing Mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**(23): 7 537-7 541.

Sirovich L, Stoeckle M Y, Zhang Y. 2009. A scalable method for analysis and display of DNA sequences. *PLoS One*, **4**(10): e7051.

Sun G, Zhu Y, Zhou Z, Chen J. 1994. The reproductive biology

of *Lateolabrax japonicus* in the Yangzi River estuary and Zhejiang offshore waters. *Journal of Fishery Science of China*, **18**(1): 18-23. (in Chinese with English abstract)

Swofford D L. 1993. Paup—a computer-program for phylogenetic inference using maximum parsimony. *Journal of General Physiology*, **102**(6): A9.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011, MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**(10): 2 731-2 739.

Uye S, Aoto I, Onbe T. 2002. Seasonal population dynamics and production of *Microsetella norvegica*, a widely distributed but little-studied marine planktonic harpacticoid copepod. *Journal of Plankton Research*, **24**(2): 143-153.

Valentini A, Pompanon F, Taberlet P. 2009. DNA barcoding for ecologists. *Trends in Ecology & Evolution*, **24**(2): 110-117.

Wang M X. 2010. Application of Molecular Markers to the Researches on Pelagic Copepods in the Chinese Coastal Regions. PhD dissertation. Institute of Oceanography, Chinese Academy of Sciences. 164p. (in Chinese with English abstract)

Wang M X, Cheng F P, Li C L, Sun S. 2011a. DNA barcoding of zooplankton of zooplankton in the Jiaozhou Bay for species identification. *Oceanologia et Limnologia Sinica*, **42**(5): 702-710. (in Chinese with English abstract)

Wang R C, Wang M X, Cheng F P, Li C L, Sun S. 2011b. species composition analysis of zooplankton in the Jiaozhou Bay using environmental barcoding of the mitochondrial cox1 gene. *Oceanologia et Limnologia Sinica*, **42**(5): 711-721. (in Chinese with English abstract)

Xiao Y C. 2004. Fauna Sinica Invertebrata (Vol.38) Chaetognatha: Sagittoidea. Science Press, Beijing. 201p. (in Chinese with English abstract)

Xu Z L, Shen X Q, Yuan Q, Chen Y Q. 2003. Distribution characteristics of zooplankton in waters around Yangshan Islands in Hangzhou Bay. *Journal of Fishery Science of China*, **27**(suppl.): 69-75. (in Chinese with English abstract)