Genetic diversity of the pelagic harpacticoid copepod *Macrosetella gracilis* on colonies of the cyanobacterium *Trichodesmium* spp.

Renate Eberl[1,2,*], Sarah Cohen[1], Frank Cipriano[3], Edward J. Carpenter[1]

[1]Romberg Tiburon Center for Environmental Studies, Department of Biology, San Francisco State University, 3152 Paradise Drive, California 94920-1205, USA

[2]University of California Davis, Department of Ecology and Evolution, One Shields Avenue, Davis, California 95616-8507, USA

[3]Conservation Genetics Lab, Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, USA

ABSTRACT: The harpacticoid copepod *Macrosetella gracilis* uses colonies of the nitrogen-fixing cyanobacterium *Trichodesmium* spp. as a floating substrate and nursery. This association enables the copepod to lead a pelagic existence. Molecular and morphological data both indicate that *M. gracilis* is maintained as a single, circumglobal species across the tropical and subtropical Atlantic and Pacific. Mitochondrial DNA sequence data from cytochrome c oxidase subunit I (COI) showed moderate to high levels of diversity and limited phylogeographic structure in *M. gracilis*. Thirty-six different haplotypes were found among the 149 *M. gracilis* individuals sequenced; 4 haplotypes were shared between the Atlantic and Pacific. Samples from Japan showed the highest haplotypic and nucleotide diversity levels, suggesting that the Indo-West Pacific may be the center of diversity for *M. gracilis*. Wide distribution and large population size in *M. gracilis* may limit divergence through increased gene flow or slowing of the loss of ancestral diversity through lineage sorting. The pelagic existence of *M. gracilis* has led to less phylogenetic structuring than found in other species of harpacticoid copepods, illustrating the evolutionary significance of specialization on a floating substrate in this copepod.

KEY WORDS: *Macrosetella* · Pelagic copepod · Harpacticoidea · Holopelagic life history · *Trichodesmium* · Rafting · Cytochrome c oxidase subunit I · COI · Genetic diversity

INTRODUCTION

Many marine taxa are assumed to have high dispersal ability due to free-swimming larvae, highly motile adults, or the ability of larvae or adults to use floats.

While it remains difficult to document large-scale movement of marine taxa through direct observation, molecular methods have been successfully employed to investigate population structure, dispersal and gene flow (Avise 2000). High gene flow is expected to oppose population isolation, differentiation and speciation in marine organisms (Grosberg & Cunningham 2001). Many molecular studies of marine organisms

have confirmed a correlation between the swimming ability of adults, or the long larval duration of species with benthic adults, and the lack of population structure (reviewed in Avise 2000), but a number of studies have detected population structure in marine species over comparatively small scales (Edmands 2001).

Copepods are the most abundant fraction of marine zooplankton, yet little is known about their population structure, and molecular methods have only been used in a few studies (e.g. Bucklin et al. 2000, 2003, Edmands 2001, Rocha-Olivares et al. 2001, Goetze 2003, 2005).

The harpacticoid copepod *Macrosetella gracilis* (Miraciidae) is unique among the predominately ben-

Aquat Biol 1: 33–43, 2007

thic harpacticoid copepods due to its specialized life history and association with the nitrogen-fixing cyanobacterium *Trichodesmium* spp. (Huys & Böttger-Schnack 1994, O'Neil 1998). Like most harpacticoids, *M. gracilis* is not well adapted for swimming (Hwang & Turner 1995); it uses *Trichodesmium* spp. colonies as rafts in the open ocean, where other floating substrates are rare. This 'pseudobenthic' habitat offers the copepod an opportunity for life in the open ocean, possibly shelter from predation, and presumably allows for greater dispersal compared to benthic copepods. The areas of distribution of *M. gracilis*—tropical and subtropical marine regions—are generally limited in nutrients. The nitrogen-fixing cyanobacterium *Trichodesmium* spp. is one of the most abundant primary producers in these ocean regions and can form geographically extensive 'blooms' spread over 100s to 1000s of kilometers (Capone et al. 1997). *M. gracilis,* at various life history stages, is the most abundant metazoan regularly associated with *Trichodesmium* spp. colonies (O'Neil 1998). *M. gracilis* inhabits geographically and temporally discontinuous habitats in the form of *Trichodesmium* spp. blooms and thus may show significant genetic structure related to environmental patchiness as proposed for other marine species (Avise 2000, Goetze 2005).

Population genetic data for *Macrosetella gracilis* is of particular interest given the rarity of a pelagic lifestyle among the predominately benthic harpacticoid copepods (examples of the few known harpacticoid copepod species with a pelagic lifestyle are given in Huys & Boxshall [1991], but no population genetic data are currently available for these species). Based on morphological data, *M. gracilis* has been thought to occur pan-globally in tropical and subtropical oceans (Huys & Böttger-Schnack 1994). Recent molecular work on other zooplankton species has often revealed a greater number of unrecognized ('cryptic') and sibling species than expected (Knowlton 1993, Lee 2000, Goetze 2003). Here we use molecular data to assess support for the pan-global distribution of *M. gracilis,* previous to this analysis based only on morphology (Huys & Böttger-Schnack 1994). The goals of the present study were to: (1) use molecular data to confirm the occurrence of *M. gracilis* as a single, cohesive species in both the Atlantic and the Pacific; (2) determine levels and patterns of distribution of genetic diversity for *M. gracilis*; and (3) compare genetic diversity of the pelagic *M. gracilis* with data from benthic harpacticoid copepods.

MATERIALS AND METHODS

Sample collection. For this study, 149 individuals of *Macrosetella gracilis* from 3 collection areas were analyzed: one in the western Atlantic, north of the Amazon outflow ('A'); and 2 in the Pacific, one near the islands of Hawaii ('H'), and the other near Japan ('J') (Table 1, Fig. 1). Sampling stations in both the Atlantic (8 stations) and Hawaii (8 stations) were spread over areas of approximately 900 km^2; there was 1 sampling station in Japan (see Fig. 3). Individuals of the occasionally co-occurring copepod *Miracia efferata* (Harpacticoidea, Miraciidae) were also collected for out-group comparison. Samples of *M. gracilis* (Harpacticoidea, Miraciidae) were collected from plankton tows at 0 to 100 m depth, immediately sorted and preserved in 95% ethanol, or frozen at –20°C.

DNA extraction. DNA extractions were initially performed on individual *Macrosetella gracilis* to compare several methods for optimal yield: 10% Chelex® (Bio-Rad) extraction, phenol chloroform extraction (Hillis et al. 1996), and extractions with the DNEasy animal kit (Qiagen). The latter method had best overall performance with the following modifications to the manufacturer's protocol: (1) the samples were left to incubate overnight with buffer and 20 µl Proteinase K (10 mg ml^{-1}); (2) columns were incubated with elution buffer at room temperature for a minimum of 20 to 30 min; (3) columns were eluted once using 200 µl of elution buffer, and the eluate was concentrated in a Speedvac® vacuum concentrator to approximately 50 µl.

Direct sequencing. Initially, 'universal' PCR primers HCO-2198 and LCO-1490 (Folmer et al. 1994) were

Table 1. Sampling locations of *Macrosetella gracilis* with ocean, cruise and station identifier, sampling coordinates and sampling dates

| Ocean	Cruise	Stn	Latitude	Longitude	Date
Pacific	MPO9	P16	19° 29.59′ N	157° 00.01′ W	Aug 2003
Pacific	MPO9	P17	19° 32.14′ N	158° 58.75′ W	Aug 2003
Pacific	MPO9	P18	20° 32.11′ N	160° 59.87′ W	Aug 2003
Pacific	MPO9	P19	21° 00.78′ N	158° 59.70′ W	Aug 2003
Pacific	MPO9	P20	20° 14.46′ N	156° 30.01′ W	Aug 2003
Pacific	MPO9	P21	20° 15.18′ N	159° 11.32′ W	Aug 2003
Pacific	MPO9	P22	19° 06.26′ N	162° 13.62′ W	Aug 2003
Pacific	St. ALOHA	Aloha	22° 45′ N	158° 00′ W	Nov 2004
Pacific	Japan	J	32° 40′ N	136° 10′ E	Oct 2004
Atlantic	MPO8	A4	10° 46.97′ N	52° 16.90′ W	Apr 2003
Atlantic	MPO8	A8	10° 20.87′ N	55° 06.07′ W	Apr 2003
Atlantic	MPO8	A18	6° 59.97′ N	49° 59.35′ W	May 2003
Atlantic	MPO8	A32	10° 42.26′ N	52° 31.65′ W	May 2003
Atlantic	MPO8	A34	10° 33.44′ N	50° 04.37′ W	May 2003
Atlantic	MPO8	A36	10° 26.42′ N	48° 18.12′ W	May 2003
Atlantic	MPO8	A38	8° 00.14′ N	48° 29.45′ W	May 2003
Atlantic	MPO8	A39	7° 51.41′ N	50° 03.96′ W	May 2003
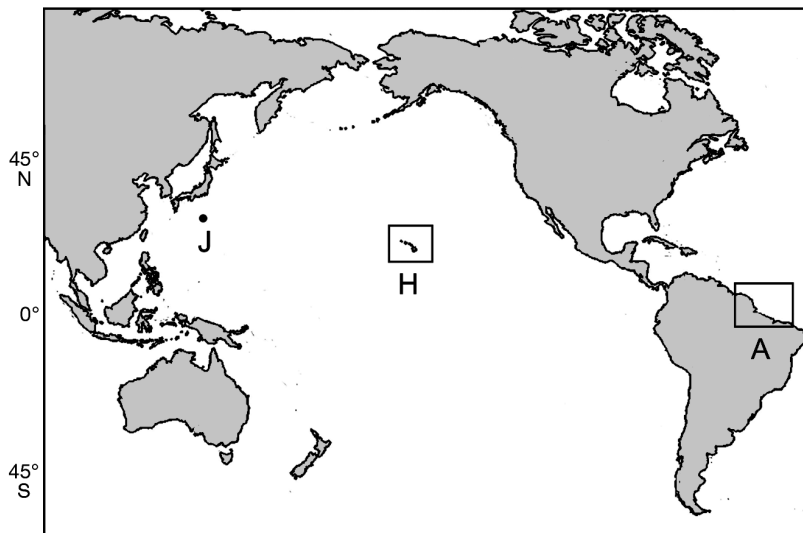
Fig. 1. *Macrosetella gracilis.* Sampling locations in the Pacific near Japan (J) and Hawaii (H) and in the Atlantic near the Amazon river outflow (A). Multiple stations were sampled at H (n = 8) and A (n = 7). For a detailed map of station locations see Fig. 3

used to amplify and sequence a 710 bp fragment of the mitochondrial cytochrome c oxidase subunit I (COI) gene from a few individuals of *Macrosetella gracilis.* New copepod-specific PCR primers MacL 5′-AAATCATAAAGATATTGGAACAC-3′ and MacH 5′-GTGGAGTTCAGATTTCGATC-3′ were designed based on our preliminary sequence results combined with sequences of other copepod species from Gen-Bank, and used to amplify and sequence 449 bp of COI from the remaining specimens. PCR reaction mixtures contained: 5 µl of 5× PCR buffer without MgCl$_2$ (Roche Diagnostics), 2.5 µl dNTP (10 mM total), 2.5 µl MgCl$_2$ (25 mM,), 1.25 µl of each primer (10 µM), 0.25 µl of Expand High Fidelity Plus taq (Roche Diagnostics), approximately 10 to 20 ng of DNA template and water to 25 µl total reaction volume. PCR reaction parameters were: initial denature at 94°C for 160 s; 35 cycles of denature at 94°C for 30 s, anneal at 50°C for 30 s, extension at 72°C for 45 s; final extension at 72°C for 300 s. PCR products were size-separated and visualized on an ethidium bromide-stained 1.8% aga-rose-Tris-borate-EDTA gel, and purified using either a QIAquick PCR purification kit (QIAGEN), or with SAP/EXO (Shrimp Alkaline Phosphatase/Exonu-cleaseI, USB) for digestion of unincorporated primers and dNTPs. Purified, amplified products were se-quenced in both directions on an ABI 3130 capillary sequencer or ABI 377 slab-gel sequencer (Applied Biosystems). Forward and reverse sequences from each individual were aligned and edited with Se-quencher 3.1.1 (Gene Codes).

Genetic diversity. Analyses of genetic diversity of *Macrosetella gracilis* were carried out with the software packages ARLEQUIN v.2.0 (Schneider et al. 2000) and DnaSP 4.0 (Rozas et al. 2003). Haplotypic diver-sity (*h*) values were calculated using Nei′s index (Nei 1987):

$$h = 1 - \Sigma f_i^2 \qquad (1)$$

where f_i is the frequency of the *i*th haplotype. Nucleotide diversity (π) was calculated as:

$$\pi = [\ \Sigma p_{ij}]/n_c \qquad (2)$$

where p_{ij} is the proportion of nucleotide differences between the *i*th and the *j*th individual, and n_c is the total number of sequence com-parisons expressed as [$n(n-1)/2$].

Phylogenetic analysis. A series of analyses were performed on complete and partial datasets in order to esti-mate phylogenetic relationships from the COI sequence data. The best-fit model of molecular evolution for this dataset was determined by hierarchical likelihood testing as imple-mented in the program MODELTEST, v.3.04 (Posada & Crandall 1998). Neighbor joining (NJ), parsimony (P) and maximum-likelihood (ML) analyses were per-formed in PAUP* 4.0b8 (Sinauer Associates, Swofford 2000). The ML analysis was performed using the suc-cessive approximations strategy (Swofford & Sullivan 2003). Tree support was assessed by bootstrapping (NJ, P = 1000 replicates, ML = 500 replicates).

In order to infer relationships between haplotypes, a minimum spanning tree was constructed using the program MINSPNET in ARLEQUIN (Schneider at al. 2000). All alternative connections were examined to determine if they resulted in fundamental changes in the topology.

Population structure and gene flow estimates. In-ference of the degree of population subdivision was based on a hierarchical analysis of molecular variance (AMOVA, Excoffier et al. 1992) implemented in ARLE-QUIN 2.0 to test for significant differences between oceans, among stations within an ocean, or within sta-tions. Pairwise Wright′s fixation index F_{ST} values be-tween different regions and average population differ-ences were calculated using ARLEQUIN to obtain an estimate of gene flow (Weir & Cockerham 1984), where the number of female migrants per generation (N_{fm}) is:

$$N_{fm} = [(1/F_{ST}) - 1]/2 \qquad (3)$$

The F_{ST} approach is based on assumptions of equilib-rium populations, symmetrical gene flow and equal populations sizes. Additional estimates of gene flow were calculated using 2 coalescent-based software pro-

grams: MDIV (Nielsen & Wakeley 2001) and MIGRATE
v.2.0.3 (Beerli & Felsenstein 2001). MDIV estimates di-
vergence time and migration rate between 2 populations
that descend from a single ancestral population using a
Markov Chain Monte Carlo (MCMC) algorithm with a
finite-sites model to account for multiple substitutions
at a site (Nielsen & Wakeley 2001). Five independent
runs with all 149 COI sequences of *Macrosetella gra-
cilis* divided between the Atlantic (n = 61) and Pacific
(n = 88) were conducted with the following parameters:
5 000 000 cycles in the Markov chain, burn-in time of
200 000 cycles, M_{max} = 10, T_{max} = 10, where M_{max} and
T_{max} are maximum values for M, the scaled migration
rate (2 × effective population size × M), and T is the
scaled divergence time (divergence time/2 × effective
population size). Since MDIV allows for only 2 popula-
tions, Hawaii and Japan were pooled as one Pacific
population. To overcome the limitation of migration es-
timates between only 2 populations, and to account for
the possibility of asymmetrical gene flow, we used
the program MIGRATE (Beerli & Felsenstein 2001) to
estimate directional gene flow between Hawaiian,
Japanese and Atlantic sampling sites using the
Metropolis-Hastings MCMC sampling, maximum like-
lihood procedure (Kuhner 2003). MIGRATE simultane-
ously estimates the effective population size (θ = $2μN_f$)
and pairwise migration rates ($M = mN_f$) between popu-
lations, where μ is the per-generation mutation rate,
and mN_f is the product of effective female population
size and mutation rate, assuming that θ remains con-
stant for each population, but can vary between popu-
lations. To decrease run time, the Atlantic and Hawai-
ian datasets were both reduced to 30 sequences each,
using a random number generator to drop individuals
from the dataset. The Japanese population included all
the original 10 samples (half of the minimum recom-
mended number of 20 individuals per population). Mul-
tiple runs were performed with 10 short chains of
length 100 000 with a sampling increment of 200, and
2 to 3 long chains of length 1 000 000 with a sampling
increment of 200, and adaptive heating with tempera-
tures of 1, 1.5, 3, 6 and 12 (Beerli & Felsenstein 2001).

To test for population expansion and selection, val-
ues of Tajima's *D* (Tajima 1989) were calculated using
DnaSP 4.0 (Rozas et al. 2003) for the whole dataset,
for different ocean basins, and for the Japanese and
Hawaiian sampling regions.

Pairs of latitude/longitude coordinates were used to
calculate geographic distances between stations using
the program 'Latitude/Longitude Distance calculations'
(Michaels 1997). The correlation between genetic and
geographic distances was evaluated with the Isolation
by Distance Web Service (IBDWS) (Jensen et al. 2005)
using Mantel test and Reduced Major Axis (RMA)
regression analysis (Sokal & Rohlf 1995).

RESULTS

Molecular characteristics

The aligned COI sequence dataset had a length of
449 bp and contained 36 unique haplotypes (GenBank
accession DQ989142–DQ989177). Sequences of 149 in-
dividual *Macrosetella gracilis* revealed 60 polymorphic
sites with a total of 67 mutations; 44 sites were parsi-
mony-informative. Of the polymorphic sites 88.4 % were
in the 3rd codon position, and 11.6 % were in the 1st po-
sition. Five non-synonymous substitutions occurred, 2 of
which were parsimony-informative. The transition-trans-
version ratio was 5.55. Average base frequencies (A =
0.30092, C = 0.12525, G = 0.17261, T = 0.38103) were sta-
tistically homogenous across samples ($χ^2$ = 3.566498, df =
105 and p = 1.00). All COI nucleotide sequences were
translated into amino acid sequences using the *Droso-
phila* mitochondrial genetic code, and no misplaced start
and stop codons or insertions or deletions were found.

Demographic statistics and genetic diversity

COI sequences of *Macrosetella gracilis* revealed
moderate to high levels of haplotype and nucleotide
diversity. Of the 36 haplotypes, 15 occurred more than
once (Table 2). The number of haplotypes per sam-
pling station (among 3 to 20 individuals sequenced)
ranged from 2 to 7 (Table 2). Japan had the overall
highest level of haplotypic and nucleotide diversity
(Table 3) and did not share any haplotypes with
Hawaii or the Atlantic (Fig. 2). Four haplotypes (31.5 %
of all sequences collected) were shared between
Atlantic and Hawaiian samples (Table 2).

The most common — haplotype A — occurred 34
times. Haplotype A and the other 3 shared haplotypes
(B, C, D) occurred in more individuals in the Atlantic
than in the Hawaiian samples (Fig. 3). Overall, the
Pacific had 21 unique haplotypes, and of these 14 were
only found once ('singletons'). Hawaii had 13 unique
haplotypes (8 singletons, Fig. 3a), and Japan had 8
unique haplotypes (6 singletons). In the Atlantic, we
found 13 unique haplotypes (10 singletons) (Fig. 3b).

Uncorrected divergence ('*p*-distances') for all *Macro-
setella gracilis* sequences ranged from 0.2 to 8.4 %
(0.2 to 8.5 % for Kimura 2-parameter [K2P]-, and 0.2
to 9.3 % for Hasegawa-Kishino-Yano [HKY]-corrected
distances), and were thus smaller than distances
reported from other harpacticoid copepods and more
similar to results for differences between calanoid cope-
pods from different oceans (Table 4). Genetic distances
between *M. gracilis* and *Miracia efferata* (both in the
family Miraciidae) showed a sequence divergence
ranging from 23.3 to 24.7 % (28.7 to 31.3 % for K2P, 29.2

Table 2. *Macrosetella gracilis.* Mitochondrial cytochrome c oxidase subunit I (mtCOI) haplotypes (A–S) of the copepod at different collection locations and numbers (No.) of individuals sequenced per station, per ocean (All Atlantic, All Pacific) and for all stations combined (Total). Haplotypes A–O occurred more than once; numbers of singleton haplotypes (S) are summarized per location. Markings on the bar refer to location as used in Figs. 2 to 4

Location	No.	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	S
Stns A4+A32	7	4	1								2						
Stn A8	4	2		1													1
Stn A18	15	5	1	1	2						1		2				3
Stn A34	6	5	1														
Stn A36	3	2									1						
Stn A38	6	2	1								1					1	1
Stn A39	20	13									4					1	2
All Atlantic	61	33	4	2	2						7	2	2			2	7
Stn Aloha	5					3	1										1
Stn P16	11		1			4	2	1	1								2
Stn P17	12		1	1		4		2	2	1							1
Stn P18	5					2	1		1								1
Stn P19	9					3	2	3	1								
Stn P20	9		1			2	1	1	1	3							
Stn P21	14	1			1	5		2	4								1
Stn P22	13					3	3	3	1	1							2
All Hawaii	78	1	3	1	1	26	10	12	11	5							8
Stn J	10													2	2		6
All Pacific	88	1	3	1	1	26	10	12	11	5				2	2		14
Total	149	34	7	3	3	26	10	12	11	5	7	2	2	2	2	2	21

Table 3. *Macrosetella gracilis.* Measures of genetic diversity of the cytochrome c oxidase subunit I (COI) of *M. gracilis* separated by sampling area and total diversity: number of individuals (n_i), number of segregating sites (S), number of haplotypes (n_h), haplotype diversity (h), nucleotide diversity (π)

Location	n_i	S	n_h	h	π
All Atlantic	61	38	15	0.694	0.01305
All Hawaii	78	25	17	0.832	0.01018
Japan	10	48	8	0.956	0.0442
All Pacific	88	58	25	0.868	0.01489
Total	149	60	36	0.899	0.01677

to 32.1% for HKY), a distance value comparable to interspecific distances in other species (Table 4).

Phylogenetic analysis

The best fit to our COI data as determined with MODELTEST (Posada & Crandal 1998) was provided by the HKY85 model with unequal base frequencies, a transition transversion ratio of 5.55, a shape parameter of 0.017, and no significant amount of invariant sites. NJ, P and ML analyses all produced essentially the

Fig. 2. *Macrosetella gracilis.* Global distribution of *M. gracilis* cytochrome c oxidase subunit I (COI) haplotypes at different sampling sites with haplotypes shared between Atlantic and Hawaii (A+H), among sampling sites in the Atlantic (A), Hawaii (H) and Japan (J) and singleton haplotypes (S) where pie slices with the same shading represent different haplotypes except for singletons. Number of individuals sampled (n) is indicated for each region
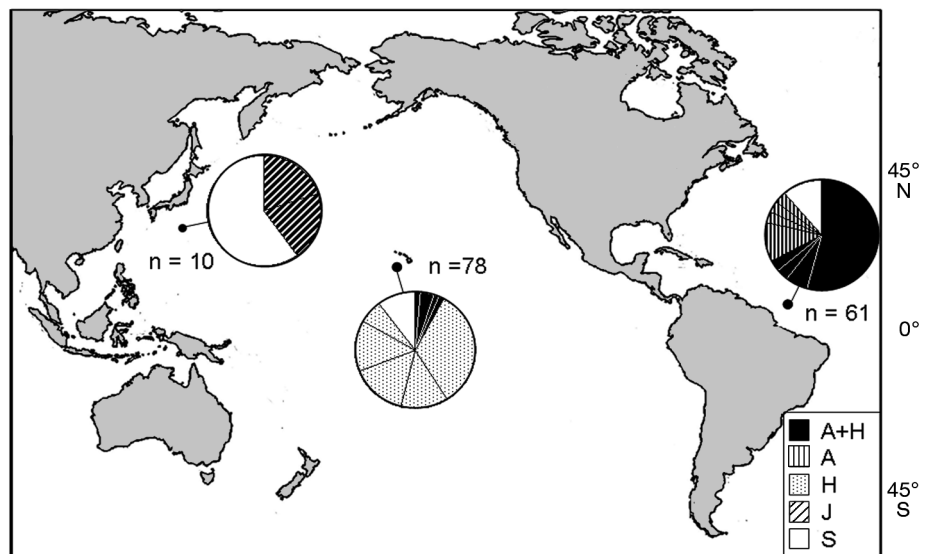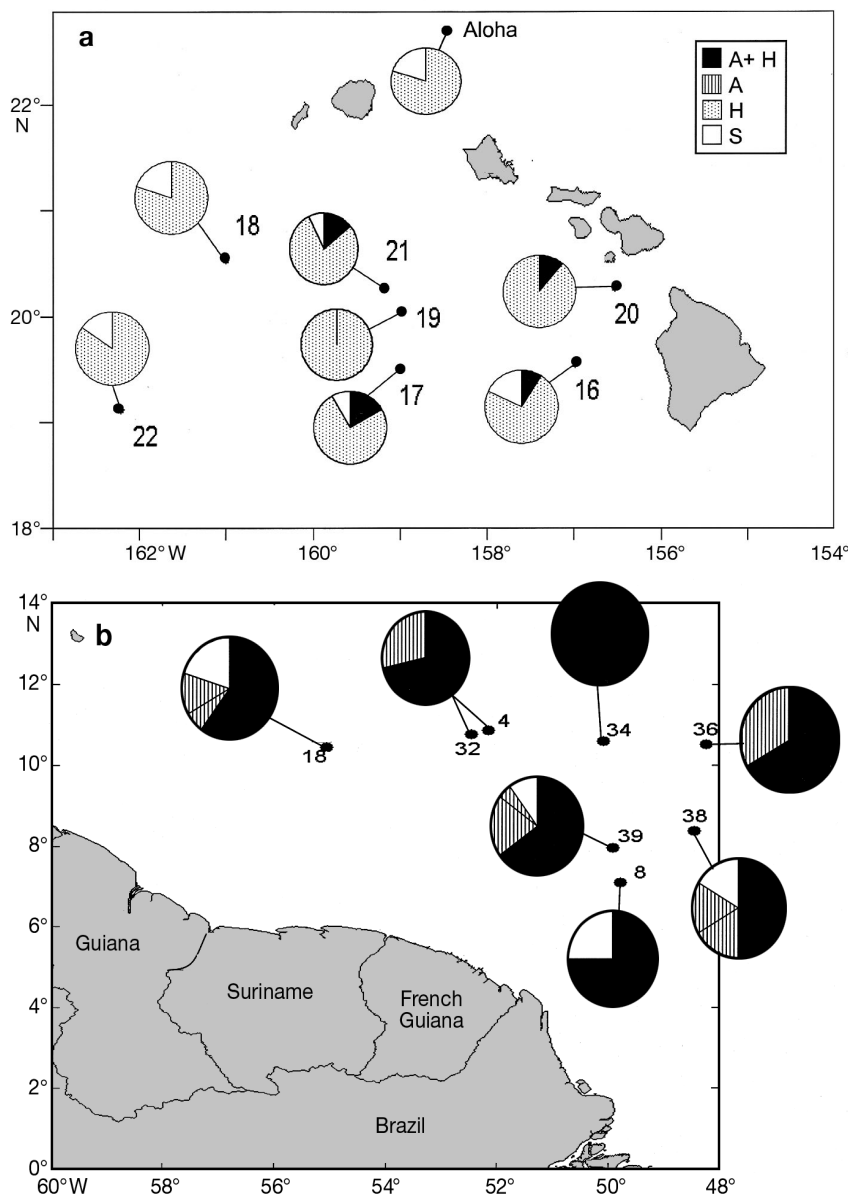
Fig. 3. *Macrosetella gracilis.* Distribution of cytochrome c oxidase subunit I (COI) haplotypes at different sampling stations at (a) Hawaiian and (b) Atlantic sampling sites with haplotypes shared between Atlantic and Hawaii (A+H), among Atlantic sampling sites (A), among Hawaiian sites (H), and singleton haplotypes (S) where pie slices with the same shading represent different haplotypes except for singletons. Numbers refer to sampling stations

same 'tree' topology with few clusters receiving high bootstrap support (Fig. 4).

Phylogenetic analysis of the 36 COI haplotypes from *Macrosetella gracilis* obtained in this study showed no geographical structuring of haplotypes into separate clades. All phylogenetic reconstructions contained 2 distinct clades with moderately high bootstrap support. The smaller clade included sequences from the Atlantic and Japan only, whereas the larger clade had repre-

sentatives from all 3 sampling areas (Fig. 4). Haplotypes that were shared between ocean basins, and common haplotypes found only within one ocean, were distributed evenly across the tree. Although only 4 of the 36 haplotypes were found in both the Atlantic and Pacific, the phylogenetic relationship of the haplotypes showed no evidence for partitioning into Atlantic and Pacific clades.

The minimum spanning tree (Fig. 5) revealed several shared and centrally located haplotypes surrounded by singleton haplotypes. Examination of alternative links within the network did not significantly change the overall appearance of the minimum spanning tree. One well-supported clade was separated by 21 mutational steps from the rest of the tree and contained both Atlantic and Japanese but no Hawaiian haplotypes. Haplotypes from all 3 sampling regions occurred scattered around the tree, with the greatest spread for haplotype distribution found among Japanese samples, despite the small sample size of only 10 individuals.

Population structure and gene flow estimates

Pairwise F_{ST} values for all sampling locations ranged from 0.117 to 0.235 and were significant at $p < 0.05$ (Table 5). Estimates of migration rates derived from F_{ST} values ranged from 1.6 to 3.8 migrating females per generation. The posterior distribution of scaled migration rates (*M*) between the Atlantic and Pacific obtained with the program MDIV showed the highest probability at $M = 1.45$, suggesting exchange between regions of about 1 to 2 females per generation. Levels of interchange between Atlantic, Hawaiian and Japanese populations estimated using MIGRATE were variable between separate runs, but all population pairs showed average $M > 1$ migrant per generation. Both Atlantic and Hawaiian populations had on average >10 migrants per generation coming from Japan, whereas the migration rates in the opposite direction were much smaller.

AMOVA of Atlantic and Pacific samples showed that only 9.2% (p = 0.002) of the variation is explained

by differences between oceans, and most of the variation was found within stations (Table 6). Values of Tajima's D were not significantly different from 0 (p > 0.1) for all population ($D = -1.161$), Atlantic ($D = -0.426$), Pacific ($D = -1.487$), Hawaii ($D = -0.304$), and Japan ($D = 0.299$), and suggest large and stable populations that have not undergone a recent population expansion.

Isolation-by-distance analysis using the Mantel test and regression analyses of genetic and geographic distances between sampling stations showed no significant isolation by distance for the whole dataset (r = −0.005, p = 0.45), for the Atlantic (r = −0.191, p = 0.70), or for the Pacific (r = −0.482, p = 0.92).

Fig. 5. *Macrosetella gracilis.* Minimum spanning tree of cytochrome c oxidase subunit I (COI) haplotypes, where the size of the circle is approximately proportional to the number of haplotypes (smallest circle represents an individual), and the hatch marks represent number of mutations between haplotypes. Shading corresponds to sample location: white = Atlantic, grey = Hawaii, and black = Japan

Fig. 4. *Macrosetella gracilis.* Maximum likelihood tree, where numbers indicate bootstrap support for neighbor-joining (above branch) and maximum likelihood analysis (below branch). Letters are haplotype names as used for submission to GenBank
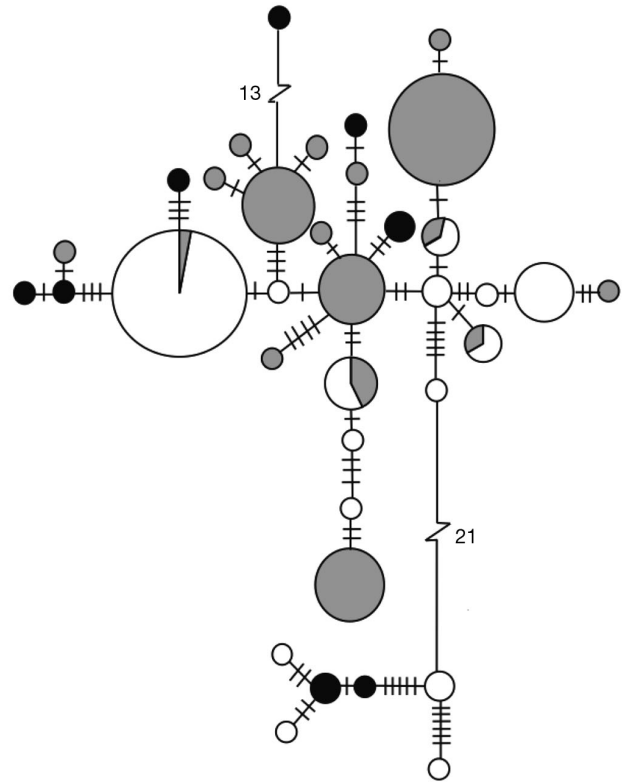
DISCUSSION

Our study represents the first population genetic dataset available for harpacticoid copepods in the family Miraciidae, and one of just several available for oceanic copepods, despite their abundance, diversity and importance for pelagic ecosystems. The results of our phylogenetic analyses of *Macrosetella gracilis* COI sequences were in agreement with previous studies on the distribution of *M. gracilis* based solely on morphological characteristics (Huys & Böttger-Schnack 1994), and confirmed the occurrence of genetically similar *M. gracilis* in both the Atlantic and the Pacific. The hypothesis of a global distribution of *M. gracilis* was supported by low genetic distances between most COI sequences across the sampling range, and by shared haplotypes between the Atlantic and the Pacific. Surprisingly, we did not find shared haplotypes between Hawaiian and Japanese samples. Genetic diversity of COI sequences in *M. gracilis* was highest in Japanese samples.

Table 4. Copepod spp. Genetic distances for cytochrome c oxidase subunit I (COI) for *Macrosetella gracilis* and *Miracia efferata* using different substitution models (K2P: Kimura 2-parameter; HKY: Hasegawa-Kishino-Yano) compared to distance values for other copepods obtained from the literature. *p*: uncorrected divergence ('*p*-distances')

Taxa	Range	Metric	Source
M. gracilis (Atlantic and Pacific)	0.2–8.4	*p*	Present study
M. gracilis (Atlantic and Pacific)	0.2–8.5	K2P	Present study
M. gracilis (Atlantic and Pacific)	02.–9.3	HKY	Present study
M. gracilis and *M. efferata*	23.3–24.7	*p*	Present study
M. gracilis and *M. efferata*	28.7–32.3	K2P	Present study
M. gracilis and *M. efferata*	29.2–32.1	HKY	Present study
Other harpacticoid copepods			
Cleptocamptus deitseri (among major lineages)	20.8–26.8	*p*	Rocha-Olivares (2001)
Cleptocamptus deitseri (among major lineages)	24.7–34.1	K2P	Rocha-Olivares (2001)
Cleptocamptus deitseri (within one cryptic species)	8.7	*p*	Rocha-Olivares (2001)
Tigriopus californicus (along the California coast)	18.2–21.1	*p*	Burton & Lee (1994)
Calanoid copepods			
Eurytemora affinis (worldwide estuarine)	17	*p*	Lee (2000)
Eucalanidae (different species Atlantic and Pacific)	0.5–10.1	*p*	Goetze (2003)
Rhinocalanus nasutus (cryptic species)	20–24.3	*p*	Goetze (2003)
Eucalanus hyalanus (cryptic species)	7.9–8.7	*p*	Goetze (2003)
Acartia clausi (within Norway)	1.0–2	*p*	Bucklin et al. (2000)
Clausocalanus furcatus (Atlantic–Aegean)	4	*p*	Bucklin et al. (2003)
Clausocalanus acruicornis (New Zealand–Aegean)	4	*p*	Bucklin et al. (2003)
Clausocalanus jobei (New Zealand–Aegean)	4	*p*	Bucklin et al. (2003)
Neocalanus robustior (Atlantic–Pacific)	3	*p*	Bucklin et al. (2003)
Clausocalanus pergens (New Zealand–Aegean)	1	*p*	Bucklin et al. (2003)

Phylogenetic analysis placed samples from the Atlantic and the Pacific within the same clades, suggesting some level of exchange between the 2 oceans or retention of ancestral polymorphisms across a huge geographic area. Genetic distances observed for *Macrosetella gracilis* average less than 2% (K2P dis-

Table 5. *Macrosetella gracilis.* Pairwise measures of genetic differentiation of *M. gracilis* between sampling regions. F_{ST} values above, Φ_{ST} below diagonal. Significance values were obtained from 1000 permutations (**p < 0.01, *p < 0.05)

	Atlantic	Hawaii	Japan
Atlantic	–	0.235**	0.204**
Hawaii	0.173**	–	0.117**
Japan	0.143*	0.328**	–

Table 6. *Macrosetella gracilis.* Analysis of molecular variance (AMOVA) of 149 individuals from sampling stations in the Atlantic and Pacific

Source of variation	df	SS	Variance components	% variation	p-value
Among oceans	1	31.6	0.367	9.27	0.002
Among stations within oceans	14	66.6	0.145	3.67	0.051
Within stations	133	458.5	3.447	87.06	<0.001
Total	148	556.7	3.959	100.00	

tances) within (Atlantic 1.8%, Pacific 1.5%) and between oceans (1.9%), which is comparable to average intraspecific distances for metazoans, but lower than COI intraspecific diversity in calanoid copepod species (Bucklin et al. 2003, Goetze 2003). Genetic diversity data are available for only 2 benthic harpacticoid copepods, the intertidal *Tigriopus californicus* (Edmands 2001) and the meiobenthic *Cleptocampus deitseri* (Rocha-Olivares et al. 2001). Extremely high levels of genetic differentiation at mitochondrial loci (20.8 to 26.8% for COI) led to the suggestion of a species complex with 4 cryptic species within *C. deitseri* (Rocha-Olivares et al. 2001). Intraspecific genetic distances for *T. californicus*—18.2 to 21.1% sampled across the species range (Edmands 2001)—were also much higher than we found for *M. gracilis* (Table 4). The evolution of a pelagic lifestyle appears to have strongly influenced the population genetics of *M. gracilis* as evidenced by comparatively low levels of population differentiation inferred from our genetic results. The harpacticoid copepods previously studied inhabit distinctly different habitats—tide-pools (*T. californicus*), estuaries and inland salt ponds (*C. deitseri*)—compared to the tropical and subtropical open ocean areas inhabited by *M. gracilis*. Furthermore, both *T. californicus* and *C. deitseri* lack a pelagic naupliar stage, whereas *M. gracilis*

nauplii hold onto *Trichodesmium* spp. colonies with modified appendages and use the cyanobacterial colonies as a floating substrate during development. The availability of floating substrate has been shown by Thiel & Gutow (2005a,b) to be important for connectivity between areas. While these authors acknowledge the importance of rafting for harpacticoid copepods, they do not include *M. gracilis* in their long list of rafting animals, nor discuss *Trichodesmium* spp. as rafting substrate. The role of *Trichodesmium* as an important floating habitat was previously recognized elsewhere (Sheridan et al. 2002). Our results show that habitat and life history of *M. gracilis* have a strong influence on the genetic structure of these harpacticoid copepods. The effect of habitat preference and life history on population structure of copepods has previously been shown by a comparison of drifting and resident calanoid copepod species (Bucklin et al. 2000). Resident *Acartia clausii* showed significant genetic differences between different fjords, whereas drifting *Calanus* spp. did not differ significantly. Genetic distances between *M. gracilis* in different regions found in this study were closer to the distances found in studies of some open ocean calanoid copepods within and across ocean basins (0.5 to 10.1%) (Bucklin et al. 2003, Goetze 2003). Genetic distances between *M. gracilis* and *Miracia efferata* (28.7 to 31.3%, K2P) in the present study were similar to distances found between different species and genera in calanoid copepods (Bucklin et al. 2000, 2003).

One clade in our dataset was separated from the rest by 21 mutational steps (up to 8.5% sequence divergence). This is a high intraspecific divergence for a pelagic copepod (Goetze 2005), but within the divergence range of the tidepool copepod *Tigriopus californianus* (Edmands 2001). While the possibility of cryptic speciation within *Macrosetella gracilis* exists, future studies using multiple loci are needed to address this possibility. Studies of marine invertebrates have found frequent evidence for cryptic speciation in the marine environment (Knowlton 1993). Several species of copepods appear to have diverged at the molecular level without becoming morphologically distinct, for example, the meiobenthic harpacticoid copepod *Cletocamptus deitseri* (Rocha-Olivares et al. 2001), the estuarine copepod *Eurytemora affinis* (Lee 2000), and the pelagic copepod *Rhinocalanus nasutus* (Goetze 2003) — all at levels of divergence similar to those found between *M. gracilis* clades. Data from other genes are needed to determine whether the observed divergences between *M. gracilis* COI sequences are signs of cryptic speciation.

While the existence of cryptic speciation in our dataset is unresolved, the possibility of sympatric speciation exists. Sibling species often have distinct habitat preferences, e.g. differences in salinity or depth distribution (Knowlton 1993, Lee 2000). The observed depth distribution of *Macrosetella gracilis* in the Red Sea with a shallow and a deep 'population' (Böttger-Schnack & Schnack 1989) could suggest the presence of sibling species and requires further study. Behavioral differences in feeding ecology have been shown in sympatric species of the harpacticoid copepod *Tisbe* (Marcotte 1984). The cyanobacterium *Trichodesmium* spp. is capable of producing toxins (Cox et al. 2005). Adaptation and various levels of tolerance to these toxins by the copepod could also lead to speciation in *M. gracilis*. It has been shown that *M. gracilis* is resistant to these toxins to a certain extent (O'Neil et al. 1996), although the differences in utilization of *Trichodesmium* spp. as a food source (O'Neil 1998, Eberl & Carpenter 2007) may be attributed to variations in toxin concentration.

The largest genetic diversity in our study was found among Japanese haplotypes, despite a low sample size there. It is possible that this area harbors remnants of the oldest and most diverse population. Our sampling area was north of the presumed Center of Origin in the Indo-West Pacific, where the highest levels of diversity have been observed for many marine species (Briggs 1999). Briggs (1999) suggested that within the Pacific, radiation of diversity occurred eastward. This hypothesis is in accordance with our genetic data, where levels of diversity are higher in Japan than in Hawaii. Due to the availability of only 10 samples, Japan was likely under-sampled. If the hypothesis of an older population with higher genetic diversity near Japan holds true, increased sampling should reveal more Japanese haplotypes and perhaps the presence of common haplotypes A to E in Japanese waters. Increased sampling may also reveal more rare haplotypes that are separated from these common haplotypes by only one mutational step and have been missed due to under-sampling.

Shared haplotypes between oceans, as well as lack of geographical partitioning of haplotypes, suggest gene flow between the Atlantic and the Pacific (either historical or current) or long-term maintenance of widespread 'ancestral' haplotypes. Geographical lineage sorting between Atlantic and Pacific populations might be expected if the populations of *Macrosetella gracilis* in the different oceans had been separated since the rise of the Isthmus of Panama (3.5 million years ago) and no gene flow exists. This expectation might be in error given the very large population sizes possible for small plankton species. One factor likely affecting the population structure of many copepods is their widespread distribution and large population size. Despite widespread distribution, however, populations may experience local bottlenecks that could lead to local

lineage extinctions. Variance in reproductive success also influences the persistence of different haplotypes in a population (Avise 2000) as is the case for *M. gracilis*, where reproduction and larval survival are linked to the presence of *Trichodesmium* spp. colonies (Björnberg 1964).

Anthropogenic factors have been implicated in changing the distribution of many marine species, especially dispersal through ballast water exchange (Carlton & Geller 1993). In addition to rafting on *Trichodesmium* spp., dispersal of *Macrosetella gracilis* could be facilitated by shipping activity. *M. gracilis* was found in ballast water after mid-ocean exchange in the Pacific (Choi et al. 2005). Dispersal of *M. gracilis* by means of ballast water is feasible, although to date there is not sufficient data available to determine the role of humans in dispersal of *M. gracilis*.

In this study of *Macrosetella gracilis*, 'migration' (interchange between regions) appears to counteract genetic differentiation. Estimates of gene flow from F_{ST} values for all possible migration patterns between the 3 sampling regions and between different oceans indicate the general magnitude of interchange as >1 female per generation. It is also possible that due to the large population sizes of *M. gracilis*, remnants of old haplotypic lineages are still present in the now separated ocean basins, and it will take a long time before Atlantic and Pacific lineages will be completely separated. Further study using multiple molecular markers and an increased sampling range will help to distinguish between current and past gene flow in *M. gracilis*.

LITERATURE CITED

Avise JC (2000) Phylogeography. Harvard University Press, London

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proc Natl Acad Sci USA 98:4563–4568

Björnberg TKS (1964) Observations on the development and the biology of the Miraciidae Dana (Copepoda: Crustacea). Bull Mar Sci 15:512–520

Böttger-Schnack R, Schnack D (1989) Vertical distribution and population structure of *Macrosetella gracilis* (Copepoda: Harpacticoida) in the Red Sea in relation to the occurrence of Oscillatoria (*Trichodesmium*) spp. (Cyanobacteria). Mar Ecol Prog Ser 52:27–31

Briggs JC (1999) Coincident biogeographic patterns: Indo-West Pacific Ocean. Evolution 53:326–335

Bucklin A, Kaartvedt S, Guarnieri M, Goswami U (2000) Population genetics of drifting (*Calanus* spp.) and resident (*Arcatia clausi*) plankton in Norwegian fjords. J Plankton Res 22:1237–1251

Bucklin A, Frost BW, Bradford-Grieve J, Allen LD, Copley NJ (2003) Molecular systematic and phylogentic assessment of 34 calanoid copepod species of the Calanidae and Clausocalanidae. Mar Biol 142:333–343

Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ (1997) *Trichodesmium,* a globally significant marine cyanobacterium. Science 276:1221–1230

Carlton JT, Geller JB (1993) Ecological roulette: the global transport of nonindigenous marine organisms. Science 261:78–82

Choi K, Kimmerer W, Smith G, Ruiz GM, Lion K (2005) Post-exchange zooplankton in ballast water of ships entering the San Fransisco Estuary. J Plankton Res 27:707–714

Cox PA, Banack SA, Murch SJ, Rasmussen U and 6 others (2005) Diverse taxa of cyanobacteria produce β-N-methyl-amino-ʟ-alanine, a neurotoxic aminoacid. Proc Natl Acad Sci USA 102:5074–5078

Eberl R, Carpenter EJ (2007) Association of the copepod *Macrosetella gracilis* with the cyanobacterium *Trichodesmium* spp. in the North Pacific Gyre. Mar Ecol Prog Ser 333:205–212

Edmands S (2001) Phylogeography of the intertidal copepod *Tigriopus californicus* reveals substantially reduced population differentiation at northern latitudes. Mol Ecol 10:1743–1750

Excoffier L, Smouse P, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479–491

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. Mol Mar Biol Biotechnol 3:294–299

Goetze E (2003) Cryptic speciation on the high seas; global phylogenetics of the copepod family Eucalanidae. Proc R Soc Lond B 270:2321–2331

Goetze E (2005) Global population genetic structure and biogeography of the oceanic copepods *Eucalanus hyalinus* and *E. spinifer.* Evolution 59(11):2378–2398

Grosberg R, Cunningham CW (2001) Genetic structure in the sea from populations to communities. In: Bertram MD, Gaines SD, Hey ME (eds) Marine community ecology. Sinauer Associates, Sunderland, MA

Hillis DME, Moritz CE, Mable BKE (1996) Molecular systematics, 2nd edn. Sinauer Associates, Sunderland, MA

Huys R, Böttger-Schnack R (1994) Taxonomy, biology and phylogeny of Miracidae (Copepoda: Harpacticoida). Sarsia 79:207–283

Huys R, Boxshall GA (1991) Copepod evolution. Ray Society, London

Hwang J, Turner J (1995) Behaviour of cyclopoid, harpacticoid and calanoid copepods from coastal waters of Taiwan. PSZN I: Mar Ecol 16:207–216

Jensen JL, Bohonak AJ, Kelly ST (2005) Software isolation by distance, web service. BMC Genet 6:13. Published online March 11, 2005. doi: 10.1186/1471-2156-6-13

Knowlton N (1993) Sibling species in the sea. Annu Rev Ecol Syst 24:189–216

Kuhner MK (2003) LAMARC: estimating population genetic parameters from molecular data. In: Salemi M, Vandamme AM (eds) The phylogenetic handbook: a practical approach to DNA and protein phylogeny. Cambridge University Press, Cambridge, p 378–398

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

Marcotte BM (1984) Behaviorally defined ecological resources and speciation in *Tisbe* (Copepoda: Harpacticoida). J Crustac Biol 4(3):404–416

Michaels C (1997) Latitude/longitude distance calculation. Avalable at: http://jan.ucc.nau.edu/~cvm/latlongdist.html

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov Chain Monte Carlo approach. Genetics 158:885–896

O'Neil JM (1998) The colonial cyanobacterium *Trichodesmium* as a physical and nutritional substrate for the harpacticoid copepod *Macrosetella gracilis*. J Plankton Res 20:43–59

O'Neil JM, Metzeler P, Gilbert PM (1996) Ingestion of $^{15}N_2$-labelled *Trichodesmium,* and ammonium regeneration by the pelagic harpacticoid copepod *Macrosetella gracilis.* Mar Biol 125:89–96

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818

Rocha-Olivares A, Fleeger JW, Foltz W (2001) Decoupling of molecular and morphological evolution in deep lineages of a meiobenthic harpacticoid copepod. Mol Biol Evol 18: 1088–1102

Rozas J, Sachez-del Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN. A software for population genetic analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Sheridan CC, Steinberg DK, Kling GW (2002) The microbial and metazoan community associated with colonies of *Trichodesmium* spp.: a quantitative survey. J Plankton Res 24:913–922

Sokal RR, Rohlf FJ (1995) Biometry. WH Freeman, New York

Swofford DL (2000) PAUP*: phylogenetic analysis using parsimony and other methods. Sinauer Associates, Sunderland, MA

Swofford DL, Sullivan J (2003) Phylogeny inference based on parsimony and other methods. In: Salemi M, Vandamme AM (eds) The phylogenetic handbook. Cambridge University Press, Cambridge, p 160–206

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

Thiel M, Gutow L (2005a) The ecology of rafting in the marine environment. I. The floating substrata. In: Gibbson RN, Atkinson RJA, Gordon JDM (eds) Oceanography and marine biology: an annual review, Vol 42. Aberdeen University Press, Aberdeen, p 181–263

Thiel M, Gutow L (2005b) The ecology of rafting in the marine environment. II. The rafting organisms and community. In: Gibbson RN, Atkinson RJA, Gordon JDM (eds) Oceanography and marine biology: an annual review, Vol 43. Aberdeen University Press, Aberdeen, p 279–418

Weir BS, Cockerham CC (1984) Estimating *F* statistics for the analysis of population structure. Evolution 38:1358–1370