



UvA-DARE (Digital Academic Repository)

Structure at open sea : genetics of zooplankton populations

Peijnenburg, K.T.C.A.

Publication date
2005

[Link to publication](#)

Citation for published version (APA):

Peijnenburg, K. T. C. A. (2005). *Structure at open sea : genetics of zooplankton populations*.

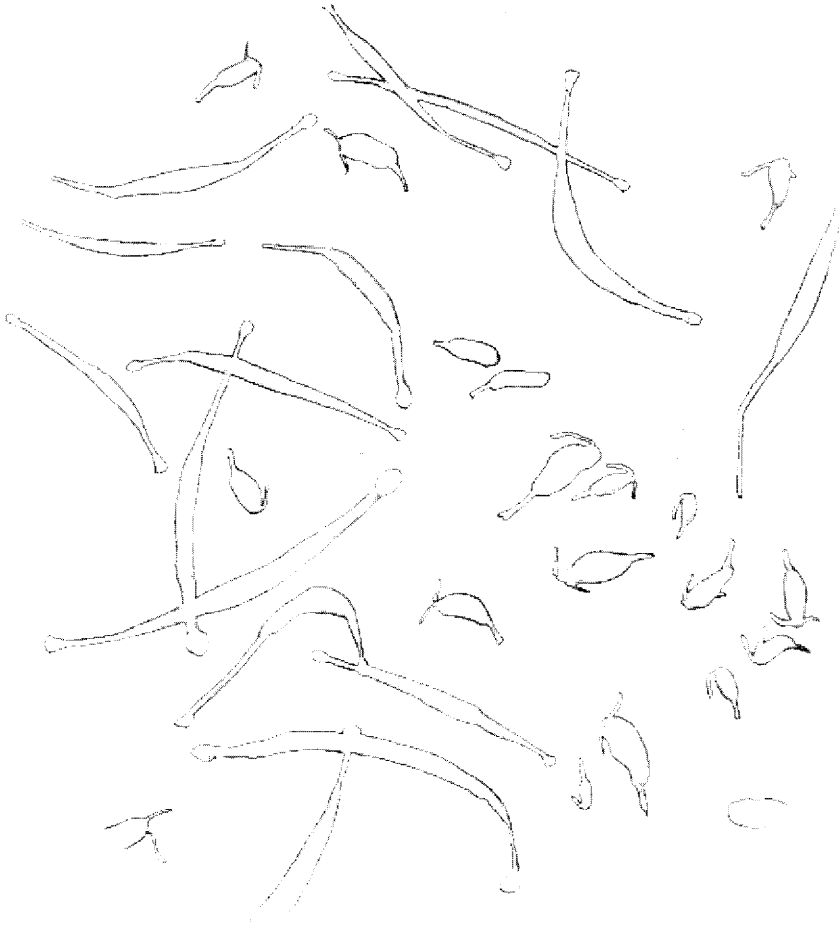
General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Chapter 4





Present-day genetic composition suggests contrasting demographic histories of two dominant chaetognaths of the North East Atlantic, *Sagitta elegans* and *S. setosa*

KATJA T.C.A. PEIJNENBURG, ELINE K. VAN HAASTRECHT, AND
CÉCILE FAUVELOT

Accepted for publication in *Marine Biology*

Abstract. — *Sagitta elegans* and *S. setosa* are the two dominant chaetognaths in the North East (NE) Atlantic. They are closely related and have a similar ecology and life history, but differ in distributional ranges. *Sagitta setosa* is a typical neritic species occurring exclusively above shelf regions, whereas *S. elegans* is a more oceanic species with a widespread distribution. We hypothesized that neritic species, because of smaller and more fragmented populations, would have been more vulnerable to population bottlenecks resulting from range contractions during Pleistocene glaciations than oceanic species. To test this hypothesis we compared mitochondrial Cytochrome Oxidase II DNA sequences of *S. elegans* and *S. setosa* from sampling locations across the NE Atlantic. Both species displayed very high levels of genetic diversity with unique haplotypes for every sequenced individual and an approximately three times higher level of nucleotide diversity in *S. elegans* (0.061) compared to *S. setosa* (0.021). Since we observed no evidence for spatial or temporal population structuring in either species, all samples were pooled to compare the genetic signatures. *Sagitta setosa* mitochondrial DNA (mtDNA) haplotypes produced a star-like phylogeny and a uni-modal mismatch distribution indicative of a bottleneck followed by population expansion. In contrast, *S. elegans* had a deeper mtDNA phylogeny and a multi-modal mismatch distribution as would be expected from a more stable population. Neutrality tests indicated that assumptions of the standard neutral model were violated for both species and results from the McDonald-Kreitman test suggested that selection may have played a role in the evolution of their mtDNA. Both species had much smaller effective population sizes estimated from genetic data when compared to census population sizes estimated from abundance data, with a factor of $\sim 10^7$ - 10^9 difference. Assuming selective effects are comparable for the two species, we conclude that the difference in genetic signature can only be explained by contrasting demographic histories. Our data are consistent with the hypothesis that in the NE Atlantic, the neritic *S. setosa* has been more severely affected by population bottlenecks resulting from Pleistocene range shifts than the more oceanic *S. elegans*.

INTRODUCTION

Comparison of the population genetic composition of closely related species can be a powerful approach to investigate the importance of contrasting life history and ecological characteristics on a species' evolutionary history. Several studies have used this approach with marine fish (*e.g.* Arnaud *et al.* 1999; Dudgeon *et al.* 2000; Fauvelot *et al.* 2003), but to our knowledge only one study has used this approach with zooplankton (Bucklin and Wiebe 1998). In that study the genetic composition of two planktonic copepods with different distributions, and different ecological and life history traits was compared, and results indicated that different levels of genetic diversity may be a legacy of contrasting effects of Pleistocene glaciations. Planktonic taxa are ideal for such comparative approaches because, by definition, they are unable to swim against major ocean currents (Van der Spoel and Heyman 1983) and hence, do not exhibit confounding differences in behavior that often hamper comparisons of the impact of climatic shifts on a species' evolutionary history.

Chaetognaths are found in many marine habitats from coastal waters to the open seas. They are often highly abundant and are functionally very important in marine food webs (Bone *et al.* 1991). *Sagitta elegans* Verrill 1873 and *S. setosa* Müller 1847 are the two dominant planktonic chaetognaths in the North East (NE) Atlantic, and based on morphological and genetic evidence, appear to be very closely related. A recent revision of the phylum Chaetognatha placed the two species in the same genus '*Parasagitta*' based on morphological similarities (Bieri 1991), and a molecular phylogeny based on 28S ribosomal DNA sequences indicated a sister species relationship (Telford and Holland 1997).

Sagitta elegans and *S. setosa* have similar ecological characteristics. They are both active predators on other zooplankton (mainly copepods) and demonstrate little dietary differentiation. Differences that have been observed mainly reflect differences in size (*S. setosa* is smaller and takes smaller prey) and prey availability associated with vertical distribution (*S. setosa* occurs higher in the water column, Øresland 1987; Feigenbaum 1991). The two species also have similar life histories. Like all chaetognaths they are protandrous hermaphrodites (testes mature before the ovaries), and reproduction is with internal fertilisation occurring while eggs are in the ovary (Pearre 1991). Cross-fertilisation is believed to be the rule in nature, although self-fertilisation has been observed in the laboratory (Alvariño 1965; Dallot 1968). Eggs of *S. elegans* and *S. setosa* are released in the sea and remain in the open water column (Pearre 1991) and both species

can lay several batches, sometimes over prolonged periods (Dallot 1968; Jakobsen 1971). This, together with evidence that different cohorts may be transported from elsewhere into a particular region, has made it difficult to estimate the number of breeding cycles per year. It has been suggested that *S. elegans* has between one and five, and *S. setosa* has between one and six breeding cycles in the NE Atlantic, with the number generally increasing southwards (reviewed in Pearre 1991). However, Jakobsen (1971) and Øresland (1983, 1986) criticized Russell's (1932a, 1932b) findings of multiple (five or six) breeding cycles in the English Channel and concluded that there is no conclusive evidence for more than a single generation per year, and that the number of cycles of *S. elegans* and *S. setosa* in different regions of the NE Atlantic should be regarded as an open question.

The main difference between *S. elegans* and *S. setosa* is related to their habitat preferences. Traditionally, the two species have been regarded as useful indicators of different water masses in the NE Atlantic, with *S. elegans* characterising the inflow of oceanic water and *S. setosa* characterising coastal water (e.g. Meek 1928; Russell 1939; Fraser 1952; Bainbridge 1963). This difference is also reflected in their distributions. *Sagitta elegans* has an extensive cold-water distribution, occurring in the upper 100–150 m in Arctic and Subarctic regions, and extending into the northern parts of the Atlantic and Pacific oceans (Fig. 1A; Alvariño 1965; Pierrot-Bults and Nair 1991). The species is sometimes referred to as 'distant neritic' because it usually has denser populations in nearshore areas (Tokioka 1979; Pierrot-Bults and Van der Spoel 2003). *Sagitta setosa* has a more restricted distribution with disjunct populations above continental shelf areas in the NE Atlantic, Mediterranean Sea, and Black Sea (Fig. 1A; Alvariño 1965; Peijnenburg *et al.* 2004). Thus, *S. setosa* can be regarded as a typically neritic species and *S. elegans* as a more oceanic one.

Neritic species, because of smaller and more fragmented populations are potentially more vulnerable to population bottlenecks resulting from climate induced range shifts and compressions than more widely distributed oceanic species (see also Grant and Bowen 1998; Arnaud *et al.* 1999). In a recent study of mitochondrial diversity of *S. setosa* populations from the NE Atlantic, Mediterranean, and Black Sea (Peijnenburg *et al.* 2004), it was shown that although levels of diversity were very high, there was significant evidence of population bottlenecks, particularly in the NE Atlantic and Black Sea. During Pleistocene glacials, the NE Atlantic was covered with land ice and polar desert (Andersen and Borns 1994) and was therefore not habitable for marine plankton. Since *S. elegans* and *S. setosa* present-

ly co-occur in the NE Atlantic, they provide an opportunity to compare the effect of differences in distributional range on the genetic impact of Pleistocene glaciations. Population size changes leave detectable patterns in the population genetic composition of species, which can be revealed by DNA sequencing (Slatkin and Hudson 1991; Rogers and Harpending 1992). This is called a genetic signature and it will be different if species experi-

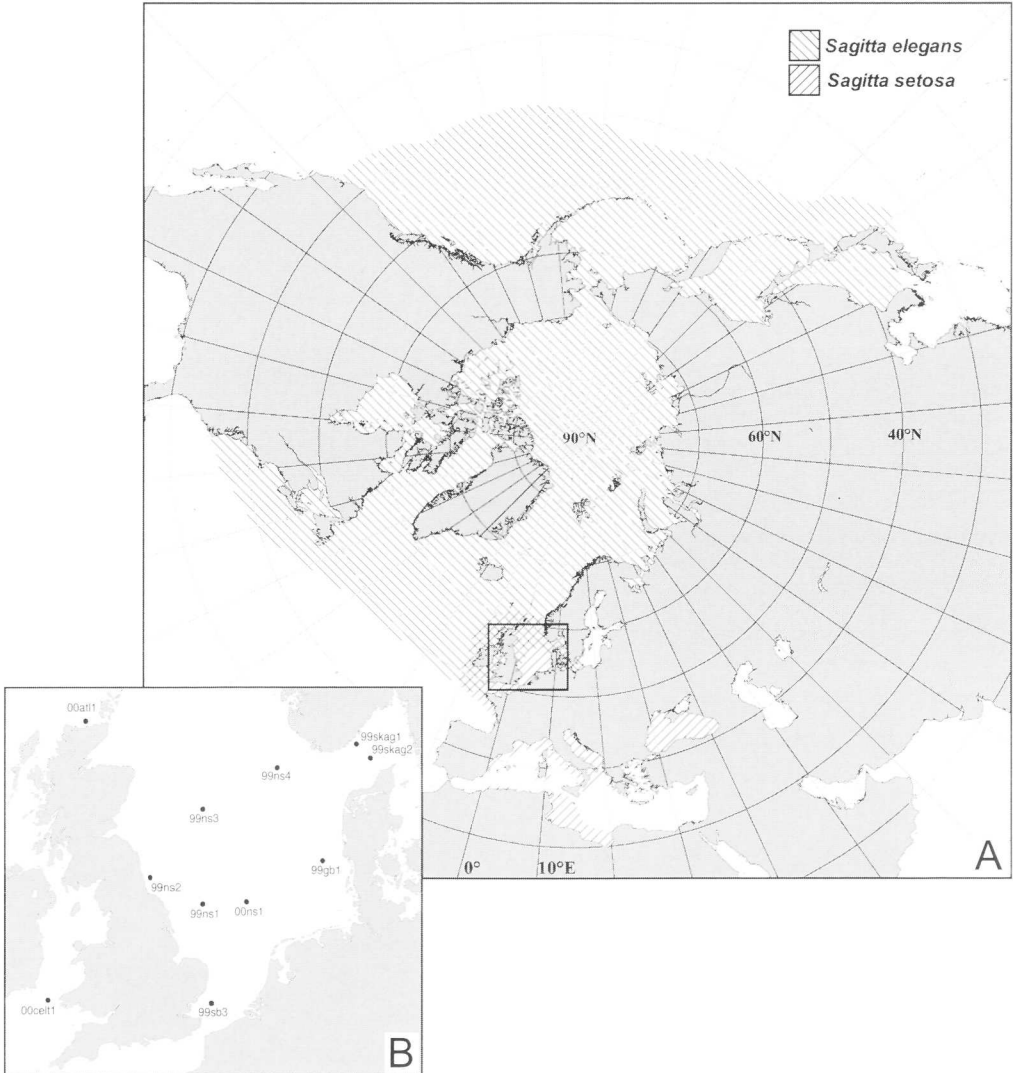


FIGURE 1. — (A) Distribution of *Sagitta elegans* and *S. setosa* shown by different hatching patterns (based on maps in Alvariño 1965; Pierrot-Bults and Nair 1991; Peijnenburg *et al.* 2004). (B) Sampling localities of *S. elegans* and *S. setosa* from the North East Atlantic (see also Table 1). Sample abbreviations used in a previous study by Peijnenburg *et al.* (2004) were N1=99skag1, N3=99gb1, N9=99sb3, N10=00at1, and N11=00cel1.

GENETIC SIGNATURES OF *SAGITTA ELEGANS* AND *S. SETOSA*

TABLE 1. — *Sagitta elegans* and *S. setosa* sampling information, with sample abbreviations and regions (see also Fig. 1B). Approximate abundance per haul (–: absent, +: present (<20 individuals), ++: abundant (20–100 individuals), +++: highly abundant (>100 individuals); n: the number of sequenced individuals for the mitochondrial Cytochrome Oxidase II fragment. Genbank Accession Numbers are AY585585–AY585616 (*S. setosa*), AY585563, AY585564, and AY942657–AY942691 (*S. elegans*).

Station label	Region	Depth of haul	Date of collection	Abundance		Analysed (n)	
				<i>S. elegans</i>	<i>S. setosa</i>	<i>S. elegans</i>	<i>S. setosa</i>
99ns1	Central North Sea	57	31.08.99	++	–	4	
99ns2	Central North Sea	75	01.09.99	++	–	5	
99ns3	Central North Sea	92	02.09.99	+++	–	5	
99ns4	Central North Sea	80	04.09.99	++	+	4	
00ns1	Central North Sea	45	18.10.00	++	+	5	
00atl1	Atlantic inflow	113	16.10.00	++	++	5	7
00celt1	Celtic Sea	100	29.09.00	+++	++	4	5
99skag1	Skagerrak	100	05.09.99	++	++		7
99skag2	Skagerrak	100	05.09.99	+++	++	5	
99gb1	German Bight	29	08.09.99	–	+++		8
99sb3	Southern Bight	44	15.09.99	–	+++		6
Total						37	32

enced different levels of population size reductions and expansions resulting from climatic fluctuations. Thus, using DNA sequences from the same mitochondrial fragment for *S. elegans* as for *S. setosa* (Peijnenburg *et al.* 2004), we compare levels of genetic diversity and signatures between the two species sampled across the NE Atlantic (Fig. 1B). We specifically test whether differences in genetic diversity are related to differences in population size of the two species and whether genetic signatures are different, as would be expected if neritic species are more prone to bottlenecks than oceanic taxa.

MATERIALS AND METHODS

Sampling and molecular analysis

Individuals of *S. elegans* and *S. setosa* were collected from the NE Atlantic during two research expeditions on the Research Vessel 'Pelagia' of the Royal Netherlands Institute for Sea Research (NIOZ) in August–September 1999 (Cruise 64PE144) and September–October 2000 (Cruise 64PE169, Table 1). Samples from 11 locations were selected for analyses, eight included *S. elegans* and five included *S. setosa* (Fig. 1B, Table 1). Vertical hauls were

taken from close to the bottom (except for station '99skag1' that was 450 m deep) to the surface using a ring net of 1 m in diameter and a 500 μm mesh size. Approximate relative abundances per haul are indicated in Table 1. These results are congruent with earlier reports, that although both species co-occur in the NE Atlantic, *S. elegans* can be regarded as a species characteristic of the inflow of oceanic water and *S. setosa* as characteristic of coastal waters (Russell 1939; Bainbridge 1963). All live chaetognaths were examined under a binocular microscope aboard ship. Only mature individuals in good condition that were clearly identified to species, and without visible gut contents or parasites were individually stored in $\sim 300 \mu\text{l}$ BLB buffer (250 mM EDTA, 5% SDS, 50 mM Tris-HCl, pH=8.0; Holland 1993). Molecular protocols are described in Peijnenburg *et al.* (2004). Thirty-seven sequences of the mitochondrial Cytochrome Oxidase II (COII) region of *S. elegans* were added to a previous dataset of 32 sequences of *S. setosa* from the NE Atlantic (Peijnenburg *et al.* 2004; Table 1).

Data analysis

General sequence characteristics. — A 504 bp fragment was unambiguously aligned for all individuals of *S. elegans* and *S. setosa* (69 sequences in total) in DAMBE (Xia and Xie 2001). DNA sequences were translated into amino-acids using the invertebrate mitochondrial code to verify that these were protein coding genes, and not nuclear copies of mitochondrial DNA (Bensasson *et al.* 2001). We encountered two stop-codons and some gaps in *S. elegans* sequences, so we double-checked seven individuals by direct sequencing PCR products on an ABI 3100 Genetic Analyser (Applied Biosystems). We confirmed that the two substitutions resulting in stop-codons and all gaps were PCR/cloning artefacts, and estimated a PCR/cloning error rate of 0.004. This estimate was slightly higher than our previous estimate of 0.003 for *S. setosa*, which was based on comparisons of sequences from different clones (Peijnenburg *et al.* 2004). PCR/cloning errors are random with respect to evolutionary history and do not contribute phylogenetic signal. For analyses, we used only sequences from a single clone for each individual and treated all gaps as missing data.

Because of high sequence variability, we examined the extent of mutational saturation (multiple substitutions at single nucleotide positions) in DAMBE by plotting the total number of transitions and transversions against Jukes-Cantor (1969) genetic distances for every pairwise sequence comparison and for all three codon positions. Sequence comparisons within

S. setosa were not saturated, and slight saturation within *S. elegans* comparisons was observed for transitions at third codon positions only. Interspecific comparisons were saturated at all three codon positions.

Population structure. — Evolutionary relationships among all haplotypes were examined using a distance matrix approach in PAUP* (Swofford 1998). Uncorrected P-distances and corrected Maximum Likelihood distances (using the General Time Reversible Model (Rodríguez *et al.* 1990) with gamma-distributed rates with shape parameter $\alpha=0.6647$, as estimated in MODELTEST; Posada and Crandall 1998) were calculated for within and between species comparisons, respectively. The neighbor-joining (NJ) algorithm (Saitou and Nei 1987) was used to construct unrooted phylogenies from distance matrices and the robustness of nodes was assessed with 1,000 NJ-bootstrap replicates.

We tested for population subdivision between different regions in the NE Atlantic (see Table 1) and between different years (1999 and 2000) for each species, using analyses of molecular variance (AMOVA; Excoffier *et al.* 1992) as implemented in ARLEQUIN (Schneider *et al.* 2000). The null hypothesis of genetic homogeneity among samples was tested using Φ -statistics (using uncorrected P-distances) and significance was assessed using 10,000 permutations.

Mitochondrial diversity and population size. — To investigate levels of genetic diversity and the genetic signatures of *S. elegans* and *S. setosa* populations from the NE Atlantic, all sequences from different areas were pooled. Using DNASP version 4.0 (Rozas *et al.* 2003), we calculated different estimators of genetic diversity, namely haplotype diversity (h ; Nei 1987), nucleotide diversity (π ; Nei 1987), and the average number of nucleotide differences between two sequences (k ; Tajima 1983). Additionally, the number of segregating sites (S), the number of mutations (η), and Watterson's theta (θ_w , based on S ; Watterson 1975) were computed.

Under a standard neutral model (Kimura 1983) of a random-mating population of constant size without selection, both π and θ_w are estimates of the population mutation rate; $\theta=2N_e\mu$ (in the case of a mitochondrial gene in a diploid organism, where N_e is the female effective population size (the number of breeding females in an idealized population; Nei 1987), and μ the mutation rate per generation). Hence, effective population size can be estimated if the mutation rate per generation is known. Note that in our

case of hermaphroditic species, the effective female population size would be (approximately) equal to the effective total population size if variance in reproductive success of males and females is similar. We used the conventional mutation rate of 2% divergence per million years for animal mitochondrial DNA (Avisé 2000) and assumed a generation time of one year for both species (see Introduction).

We wished to compare our estimates of effective population size of *S. elegans* and *S. setosa* (based on genetic data) to estimated actual total population sizes. However, to our knowledge such estimations are unavailable for *S. elegans* and *S. setosa*. To roughly generate estimates we multiplied published data on average densities by our own estimates of the surface area and volume of known distributions. In the case of *S. elegans*, this included all marine habitats north of latitude 40° N, comprising the northern parts of the Atlantic and Pacific oceans and Arctic and Subarctic regions, to a maximum depth of 150 m (Fig. 1A). In the case of *S. setosa*, only the distribution area of the NE Atlantic population was considered because it is now isolated from populations in the Mediterranean and Black Sea (Peijnenburg *et al.* 2004). Hence, this distribution comprised all neritic marine habitats between latitudes 45° N and 61° 30' N, and longitudes 12° W and 15° E, roughly ranging from the western Baltic to the continental seas of the NE Atlantic (Fig. 1A). Volumes were adjusted for varying water depths according to a grid of 0.5×0.5 degrees (H. van Aken, pers. comm. 2004).

Demographic history and neutrality. — Demographic history of the two species was examined using mismatch distributions, *i.e.* the distribution of the observed number of differences among all pairs of haplotypes. Population expansions (within the useable time frame of the genetic marker) produce a wave in the mismatch distribution creating a uni-modal distribution. In contrast, multi-modal distributions are expected for samples drawn from populations with constant population sizes (Slatkin and Hudson 1991; Rogers and Harpending 1992). For each species, we also computed two theoretical distributions using DNASP, the first one as expected for an equilibrium population of constant population size, based on θ_w as estimated from the data, and the second one as expected under a sudden expansion model (Rogers and Harpending 1992). This model considers a population that was formerly at equilibrium, but τ generations ago suddenly expanded (assuming θ_1 (after the expansion) to be infinite according to recommendations in DNASP). To test whether one model fitted the observed data better than the other model, we calculated square deviations

(SD) between frequencies of observed pairwise differences and their expectations under the two different models. We used a paired t-test to test for differences in SD's. If significant, we assumed that the model with the smaller sum of all SD's (SSD) best explained our data.

To test for neutrality we computed the commonly used Tajima's D statistic (Tajima 1989a) as well as two additional ones, the Fu and Li's F* statistic (Fu and Li 1993), and the R₂ statistic (Ramos-Onsins and Rozas 2002). These tests use information from mutation frequencies and were chosen because they were found to be the most powerful to reject neutrality in the case of population growth (Ramos-Onsins and Rozas 2002). One of the most powerful tests according to these authors, the F_s test (Fu 1997), could not be performed because all haplotypes were unique (see Results). Using DNASP, test statistics for each species were compared to their empirical distributions expected under the neutral model as generated by 10,000 coalescent simulations based on S. Additionally, the use of the same protein coding gene in two related species allowed us to perform a McDonald-Kreitman test (McDonald and Kreitman 1991). This test is based on a prediction from neutral theory that the ratio of replacement (non-synonymous) to silent (synonymous) fixed differences observed between species should equal the ratio of replacement to silent polymorphisms within species. For instance, a significant excess of replacement fixations indicates positive divergent selection.

RESULTS

General sequence characteristics

All 69 sequenced individuals of *S. elegans* and *S. setosa* represented unique haplotypes resulting in maximum haplotypic diversity. Nevertheless, we are confident that our sequences represent mitochondrial COII coding sequences, for the following three reasons: (1) we found a close match with the COII gene of the recently sequenced entire mitochondrial genome of the first representative of the phylum Chaetognatha, *Paraspadella gotoi*, by BLASTP searching Genbank (Accession Number YP054597; Helfenbein *et al.* 2004), (2) we did not encounter any double peaks, stopcodons, or gaps in direct sequences of the same fragment, and (3) restriction digests of PCR products did not show multiple banding patterns (data not shown). DNA sequences are deposited in Genbank under Accession Numbers AY585585-AY585616 (*S. setosa* from the NE Atlantic), AY585563, AY585564, and AY942657-AY942691 (*S. elegans*).

Most mutations were at third codon positions in both species (77.26% and 67.28% for *S. elegans* and *S. setosa*, respectively), but the relative proportions of transitions and transversions differed between the two species (57.52% and 91.36% transitions at third codon positions for *S. elegans* and *S. setosa*, respectively), reflecting saturation at third codon transitions observed for *S. elegans* only (see Materials and Methods). Translation into amino acid sequences revealed that replacement mutations were common. From a total of 167 aligned amino acid positions, 17 were polymorphic in both species, 44 were polymorphic in *S. elegans* only, 17 in *S. setosa* only, 71 were fixed and shared between the two species, and 18 were fixed but different between the two species. These results suggest that similar regions of the COII fragment were conserved between the species, but sequences of *S. elegans* were much more variable compared to *S. setosa*, both at the nucleotide as well as the protein level.

Population structure

Uncorrected pairwise distances between haplotypes within *S. elegans* were larger than for *S. setosa*, ranging from 0.99 to 12.95% (average \pm standard

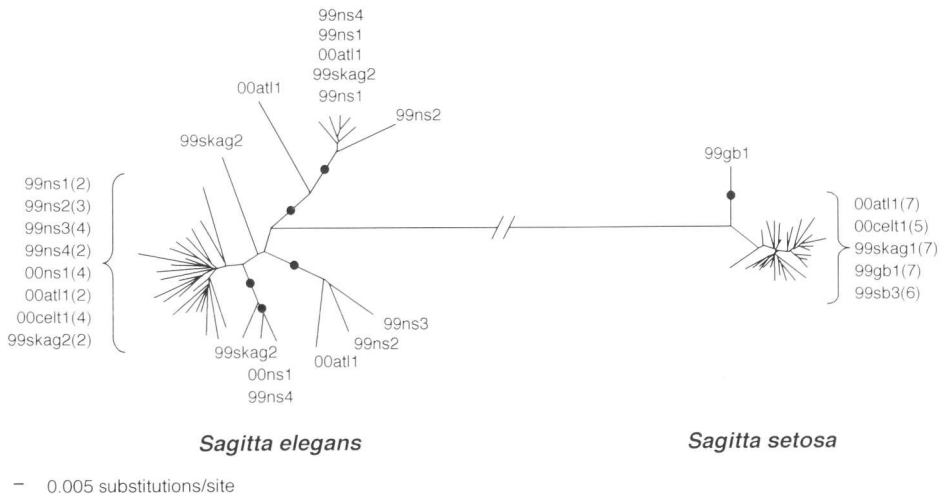


FIGURE 2. — Unrooted Neighbor-Joining phylogram of all *Sagittia elegans* (shown on the left) and *S. setosa* (shown on the right) mitochondrial Cytochrome Oxidase II haplotypes (based on P-distances). Nodes with more than 80% bootstrap support are indicated with black dots. Numbers in brackets denote the number of individuals per population sample found for a particular clade, if more than one. Average maximum likelihood divergence between *S. elegans* and *S. setosa* was $77.70 \pm 3.45\%$ (standard deviation).

deviation, $6.30 \pm 2.74\%$) compared to a range of $0.20\text{--}5.56\%$ ($2.08 \pm 0.95\%$) for *S. setosa*. The average maximum-likelihood-corrected divergence between the two species was $77.70 \pm 3.45\%$ (range $68.18\text{--}89.68\%$). Reconstructions of evolutionary relationships amongst haplotypes revealed no within-species spatial or temporal structuring of samples (Fig. 2). Similarly, AMOVA analyses revealed no significant within-species structuring (Φ_{st} -values ranged from -0.02 to 0.004 between regions or between years for both species, all comparisons $P > 0.26$). The unrooted NJ tree (Fig. 2) showed a star-like structure for *S. setosa* haplotypes (with the exception of a single deeper lineage from the German Bight, population sample '99gb1'). By contrast, the phylogeny of *S. elegans* haplotypes consisted of several well-supported subclades indicating a deeper genealogical structure (Fig. 2).

Mitochondrial diversity and population size

Since we found no evidence of genetic differentiation in either species, all population samples within species could be grouped to estimate effective population sizes and conduct tests of selective neutrality and demographic history. The number of observed segregating sites (S) of *S. elegans* sequences was approximately twice that of *S. setosa* sequences (192 and 94, respectively), whereas nucleotide diversity (π) for *S. elegans* was about three times that for *S. setosa* (0.0612 and 0.0208 , respectively; Table 2A). Likewise, estimates of θ_w suggested a ~ 2.6 times deeper genealogical structure in *S. elegans* compared to *S. setosa* (Table 2A). Converting estimates of π and θ_w into effective population sizes of the two species resulted in a ~ 2.6 to 3 times larger estimate for *S. elegans* than for *S. setosa*, assuming similar mutation rates for both species. Applying a standard mutation rate of 10^{-8} mutations per generation, we estimated effective population sizes of $4.7\text{--}6.1 \times 10^6$ for *S. elegans*, and $2.1\text{--}2.3 \times 10^6$ for *S. setosa* (Table 2B).

Estimates of effective population sizes were substantially smaller than our estimates of total population sizes based on the distribution and average abundances of the two species. Published estimates of the average density of *S. elegans* are 460 and 560 individuals per m^2 in the Celtic Sea (Conway and Williams 1986) and Arctic Ocean (Choe and Deibel 2000), respectively, and 1.4 and 6.8 individuals per m^3 in the SW North Sea (Wimpenny 1936) and North Pacific (Fager and McGowan 1963), respectively. Assuming an average abundance for *S. elegans* of $500/\text{m}^2$ or $5/\text{m}^3$ over its entire range, and an estimated distribution of 4.2×10^{13} m^2 or

TABLE 2. — Genetic diversity estimators (A), estimated census and effective population sizes (B), and results from neutrality tests (C) for *Sagitta elegans* and *S. setosa*. The order of neutrality tests is by decreasing power of rejecting the stable population size model according to Ramos-Onsins and Rozas (2002). Significance of the tests is indicated to the right of test statistics as NS (not significant), *: $P < 0.05$, **: $P < 0.01$, and ***: $P < 0.001$.

	<i>S. elegans</i>	<i>S. setosa</i>
A: Diversity estimators		
Number of haplotypes	37	32
Number of segregating sites (S)	192	94
Number of mutations (η)	251	96
Average number of nucleotide differences (k)	29.87	10.41
Nucleotide diversity (π)	0.0612	0.0208
Theta per site from S (θ_w ; Watterson 1975)	0.0943	0.0467
B: Estimation of population size		
Average density per m ²	500	1000
Average density per m ³	5	50
Distribution (area) in m ²	4.2×10^{13}	1.0×10^{12}
Distribution (volume) in m ³	5.6×10^{15}	6.6×10^{13}
Census population size (based on area)	2.1×10^{16}	1.0×10^{15}
Census population size from (based on volume)	2.8×10^{16}	3.3×10^{15}
Effective population size from π	6.1×10^6	2.1×10^6
Effective population size from θ_w	4.7×10^6	2.3×10^6
C: Neutrality tests		
R ₂ (Ramos-Onsins and Rozas 2002)	0.0583 **	0.0400 ***
F* (Fu and Li 1993)	-3.2440 *	-3.5940 **
D (based on S; Tajima 1989a)	-1.3149 NS	-2.0914 *

5.6×10^{15} m³, we obtained an overall population size of 2.1×10^{16} (based on area) or 2.8×10^{16} (based on volume, Table 2B). Compared to *S. elegans*, observed densities of *S. setosa* in the NE Atlantic were larger, $\sim 1100/\text{m}^2$ (Øresland 1983) and $\sim 55/\text{m}^3$ (Wimpenny 1936). Using an overall average density of $1000/\text{m}^2$ and $50/\text{m}^3$ and an estimated distribution in the NE Atlantic of 1.0×10^{12} m² or 6.6×10^{13} m³, the overall population size of *S. setosa* was 1.0×10^{15} (based on area) or 3.3×10^{15} (based on volume, Table 2B). Thus, both surface area and volume estimates of census population sizes were similar for both species, and *S. elegans* had an approximately 8–20 times larger census population size than *S. setosa*. Compared with effective population size estimates, our calculations of census population sizes were much larger, by a factor of $3.4\text{--}6.0 \times 10^9$ for *S. elegans* and $4.4\text{--}15.7 \times 10^8$ for *S. setosa*.

Demographic history and neutrality

Mismatch distributions differed strikingly between the two species (Fig. 3). The distribution of *S. elegans* was multi-modal with a mean of 29.87 pairwise differences and a distribution ranging from five to 62 differences. This is among the patterns that might be expected in a stable size population (Slatkin and Hudson 1991; Rogers and Harpending 1992). Conversely, the mismatch distribution of *S. setosa* was less broad (between one and 28 pair-

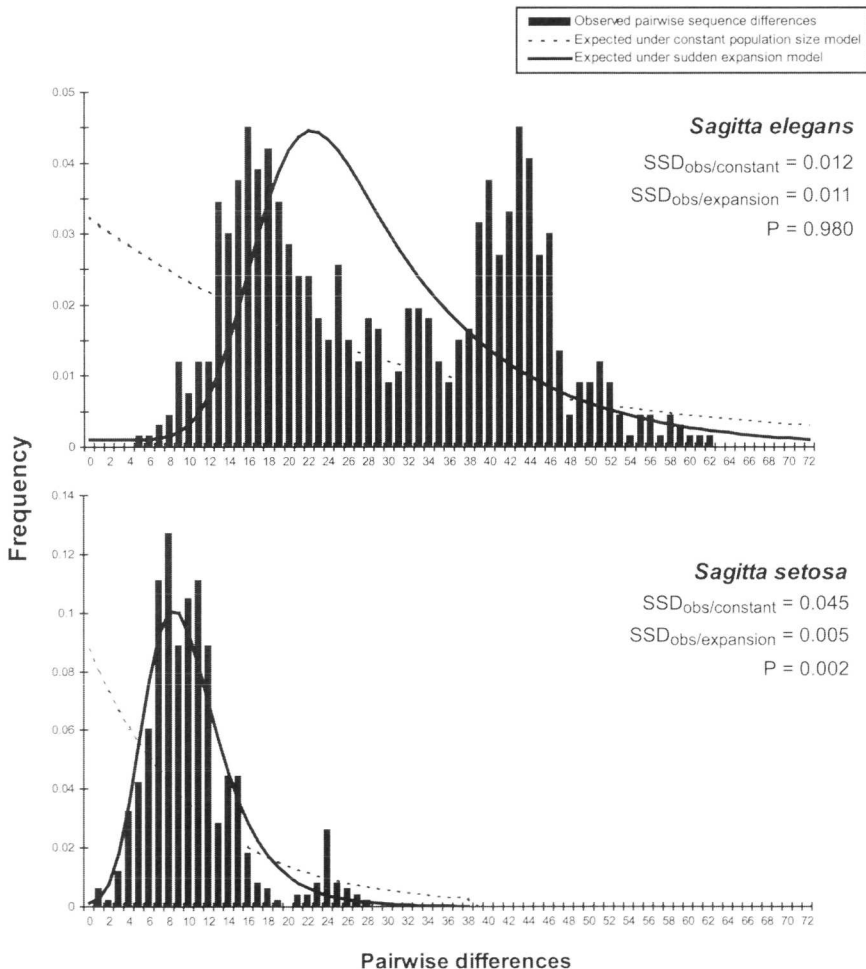


FIGURE 3. — Mismatch distributions of pairwise sequence differences for *Sagitta elegans* (top) and *S. setosa* (bottom). Expected distributions under a model of constant population size and under a sudden population expansion model are indicated as computed in DNASP (Rozas *et al.* 2003). Sum of squared deviations (SSD) between the observed data and the two theoretical models are shown on the right as well as the P values of the paired t-test testing for a difference in fit between the two models and the observed data.

wise differences and a mean of 10.41 differences) and had a uni-modal shape with the main peak at eight pairwise differences reflecting the star phylogeny found for this species (Fig. 2). This pattern is suggestive of a bottleneck followed by a sudden population expansion (Slatkin and Hudson 1991; Rogers and Harpending 1992). An additional small peak was apparent at 24 differences, which resulted from the inclusion of a single representative of a deeper clade (individual N3-3 in Peijnenburg *et al.* 2004). For *S. elegans*, no significant difference was found between the fit of the observed data to that expected under a stable population size model, or the sudden expansion model (paired t-test, $df=58$, $P=0.98$), whereas for *S. setosa*, the fit of the two models was significantly different (paired t-test, $df=28$, $P=0.002$). The lower SSD for the sudden expansion model indicated that this model provided the best fit to our data (Fig. 3). Based on this model, with parameters $\theta_0=3.51$, $\theta_1=\infty$, and $\tau=6.9$, we calculated that an expansion of the NE Atlantic population of *S. setosa* started approximately 685,000 generations ago (using the formula $\tau=2ut$, where u is the mutation rate of the entire sequence and t is the number of generations; Rogers and Harpending 1992).

Results of neutrality tests are presented in Table 2C and are ordered by decreasing power to reject the null hypothesis of neutrality in the case of a demographic expansion according to a simulation study by Ramos-Onsins and Rozas (2002). The R_2 test resulted in a significant rejection of the neutral model for both species, with a lower R_2 value for *S. setosa*, which is typical of a recently expanded population. Likewise, Fu and Li's F^* test rejected the neutral model for both species. The commonly used Tajima's D test, however, did not reject neutrality for *S. elegans*, but did reject this for *S. setosa* (Table 2C). The McDonald-Kreitman test revealed a significantly higher ratio of replacement to silent fixed differences between species ($35:38=0.921$) compared to the ratio of replacement to silent polymorphisms within species ($54:139=0.388$) indicating positive selection of the COII gene during the evolution of these species (Fisher's exact test, $P=0.003$).

DISCUSSION

Our study of mitochondrial COII sequences of two chaetognaths *S. elegans* and *S. setosa* from the NE Atlantic is one of the few studies that have examined genetic variation in zooplankton. High levels of genetic diversity com-

combined with a lack of genetic structuring enabled us to contrast the two species' present genetic composition. The strength of this approach is that we compared the same gene for two systematically and ecologically related species, which differed mainly in distributional ranges.

Mitochondrial diversity

Sagitta elegans and *S. setosa* both possessed very high levels of genetic diversity with unique haplotypes for every sequenced individual and nucleotide diversities (π) of 0.061 and 0.021, respectively. This is surprising because COII is generally considered a conserved protein-coding gene (e.g. Simon *et al.* 1994). Compared to other studies of mitochondrial diversity in zooplankton, our estimates of π are among the highest ever reported. For example in pelagic crustaceans (copepods and euphausiids) π values ranged from 0.002 to 0.018 (Bucklin *et al.* 1997, 2000; Bucklin and Wiebe 1998; Zane *et al.* 2000; Papadopoulos *et al. in press*, Chapter 3 of this thesis). However, these studies are not strictly comparable because different mitochondrial genes were used, which may differ in mutation rate. High levels of genetic diversity are often explained by large population sizes (Nei 1987; Avise *et al.* 1989) and the enormous population sizes of zooplankton, which are among the largest in the animal kingdom (Bucklin and Wiebe 1998; Norris 2000) may explain the high levels of diversity that we found. A small fraction of the diversity may also be explained by PCR/cloning artefacts. For example, assuming an error rate of 0.004 (see Materials and Methods) a maximum of 5% of the pairwise comparisons for *S. setosa* may have been between identical haplotypes, but none of the pairwise comparisons for *S. elegans*.

Based on π and k , which are corrected for sample size, mitochondrial diversity was about three times higher for *S. elegans* than for *S. setosa*. Comparing our results with COII variability in different populations of *S. setosa* from Peijnenburg *et al.* (2004) indicated lower estimates of nucleotide diversity of the NE Atlantic population of *S. setosa* (0.021) compared to Mediterranean populations (0.026 and 0.027), but higher estimates compared to the Black Sea population (0.015). *Sagitta elegans* displayed higher levels of nucleotide diversity (0.061) than any of the *S. setosa* populations. This finding is congruent with our estimates of a ~8-20 times larger census population size of *S. elegans* compared to *S. setosa*.

Effective versus census population size

We found a large discrepancy between effective population sizes and our estimates of total population sizes (4.4×10^5 – 6.0×10^6 -fold smaller effective than census population sizes). The estimation of effective population size depends on the assumption that the standard 2% clock is correct for the evolution of chaetognath mitochondrial DNA and that our single locus represents the true genealogy of the species. Furthermore, estimates of effective population size depend on estimates of θ , which assume an idealized population and a neutral model of evolution. The fact that our estimates of effective population size based on θ_w and π differ (Table 2B) suggests that one or more assumptions of the neutral model are invalid (see also next section). However, there exists no alternative for the neutral model to estimate effective population sizes. The estimation of census population size for the two species is also approximate because of the scarcity of data on species abundances in different seasons and across their entire geographical ranges. On one hand, abundances may have been overestimated because most estimates were obtained during seasons when chaetognaths are abundant. On the other hand, abundances may have been underestimated because chaetognaths may not have been sampled most efficiently. For example, *S. elegans* is known to be abundant in the hyperbenthic zone, an area that is usually not sampled by plankton-nets (Øresland 1987; Choe and Deibel 2000).

A large discrepancy between effective and census population size is commonplace for marine species (Grant and Bowen 1998; Avise 2000). Large discrepancies have been found for Antarctic krill ($\sim 10^6$ -fold difference; Zane *et al.* 1998) and two North Atlantic copepods ($\sim 10^6$ -fold difference; Bucklin and Wiebe 1998). Explanations for these discrepancies, none of which are mutually exclusive, have included: (1) a large variance in reproductive success among individuals, (2) overlapping generations, (3) extinction and recolonizations of local subpopulations, (4) historical fluctuations in population size, and (5) selective sweeps. The first explanation is very likely for species with high fecundities such as chaetognaths, resulting in huge variances in progeny numbers among families with consequent lowering of effective population size (Hedgecock 1994; Avise *et al.* 1989). Overlapping generations may also be possible for *S. elegans* and *S. setosa* because it is known that both species can lay several batches of eggs and sometimes over prolonged periods (Dallot 1968; Jakobsen 1971). Even though we assumed a single generation per year for each species, the exact number of breeding cycles in different regions of the NE Atlantic is still a

topic of discussion (Pearre 1991). The third explanation of local subpopulations (Slatkin 1985; Whitlock and Barton 1997; Avise 2000) may be applicable to zooplankton populations as they are often patchily distributed. However, we found no evidence of population structuring over evolutionary timescales for the two chaetognaths in the NE Atlantic, which renders this possibility less likely. The fourth and fifth explanations of historical fluctuations in population size and selective sweeps have been hypothesized for many northern temperate species that are affected by Pleistocene glaciations (e.g. Hewitt 1996) and it has been shown that they can have significant effects on the genetic composition of even highly abundant marine species (Avise *et al.* 1988; Grant and Bowen 1998; Bucklin and Wiebe 1998). However, the genetic signature of a population bottleneck followed by an expansion may be very similar to that of a selective sweep in a constant-sized population (Tajima 1989b; Donnelly and Tavaré 1995), making the two difficult to disentangle. Two of the most powerful neutrality tests (R_s and F^* test) rejected neutrality for both *S. elegans* and *S. setosa*, but the Tajima's D test only significantly rejected neutrality for *S. setosa*. An alternative test, which was specifically designed to test for selection is the McDonald-Kreitman test (McDonald and Kreitman 1991; Skibinski 2000). This test indicated highly significant evidence for positive selection on the COII gene during the evolution of *S. elegans* and *S. setosa*. However, it may be that this test is also affected by fluctuations in population size (Eyre-Walker 2002) and the significance may be inflated because sequences of the two species were saturated with respect to each other (see Whittam and Nei 1991). Thus, assumptions of the neutral model are violated for both species, which may be explained by bottlenecks as well as directional selection.

Demographic history

The genetic signatures of *S. elegans* and *S. setosa* are strikingly different when depicted as phylograms (Fig. 2) and mismatch distributions (Fig. 3). *Sagitta setosa* had mtDNA haplotypes showing a star-like phylogeny and a uni-modal mismatch distribution, whereas *S. elegans* had a much deeper mtDNA phylogeny and a multi-modal mismatch distribution. Although selection has probably played a significant role in determining the two species' present genetic composition, it is not a likely explanation for the difference in genetic signature between the two species because the two chaetognaths are biologically very similar. Assuming that selective effects have been the same during the evolutionary history of the two species, the

deeper genealogical structure of *S. elegans* compared to *S. setosa* may be explained if: (1) *S. elegans* is a much older species than *S. setosa*, (2) genetically differentiated groups were sampled for *S. elegans* and not for *S. setosa*, and (3) *S. elegans* was less affected by bottlenecks compared to *S. setosa*. The first possibility of *S. elegans* being a much older species than *S. setosa* is unlikely because based on a molecular species phylogeny, Telford and Holland (1997) suggested a rapid, recent radiation of different groupings within the genus 'Sagitta', which included the grouping 'Parasagitta' to which both species belonged. The second possibility of genetically differentiated groups within *S. elegans* may be important in the light of three subspecies that have been described; one from the Arctic (*S. elegans arctica* Aurivillius 1896), one from the Subarctic (*S. elegans elegans* Verrill 1873), and another from the Baltic (*S. balthica* Ritter-Zahony 1911). However, most authors consider these as local forms representing only variability in size, which may be a phenotypically plastic response to the water temperature in which they develop (Fraser 1952; Alvarino 1965; Pierrot-Bults and Nair 1991). Moreover, all our samples were taken from areas where only the variant 'elegans' had been reported and we did not observe any apparent differences in the sizes of sampled individuals. It may be possible though, that *S. elegans* is genetically structured over broader geographical scales than we sampled, and that divergent lineages have colonized the NE Atlantic through occasional dispersal. One earlier study using allozymes to study the genetic structure of *S. elegans* from the waters off Japan suggested subtle differentiation between populations, but results were contradictory (Thuesen *et al.* 1993). A more broad-scale genetic survey of *S. elegans* populations is necessary to reveal whether the entire species can be regarded as panmictic or not. Our lack of evidence for genetic structuring of populations in the NE Atlantic, however, taken together with other studies of so-called 'high gene flow' species occurring above the shelf in this region (*e.g.* Shaw *et al.* 1999; Hoarau *et al.* 2002) suggest that population structuring in the NE Atlantic of *S. elegans* is unlikely to be the case.

The evidence suggests that the difference in genetic signature of *S. elegans* and *S. setosa* is most likely due to a difference in demographic history. We found a clear genetic signature of a population bottleneck followed by a sudden expansion for *S. setosa*, but not for *S. elegans*. Reconstructions of seawater temperatures and current patterns in the North Atlantic during the last glacial maximum (18,000 years ago) based on foraminiferal and coccolithophorid assemblages, show that northern temperate conditions were compressed and shifted southwards compared to present conditions

(McIntyre *et al.* 1976). If we assume that distributional ranges of *S. elegans* and *S. setosa* were similar during warm periods (interglacials) of the Pleistocene compared to the present, then both species were severely affected during glacial periods, with drops in temperature exceeding 10°C (McIntyre *et al.* 1976; see also Fig. 5 in Peijnenburg *et al.* 2004) and drops in sea level of at least 85 m (CLIMAP 1976) resulting in a complete covering with pack ice of the NE Atlantic shelf. Our results support the hypothesis that a neritic species such as *S. setosa*, which presently occurs only above continental shelf areas, has been more affected by bottlenecks resulting from repeated glaciations compared to a more oceanic species with a wider distribution pattern such as *S. elegans*. This is presumably because the distribution of *S. setosa* was not just severely compressed, but ultimately completely displaced. By contrast, even though populations of *S. elegans* were also severely compressed, they were not completely displaced, and may have been buffered against bottlenecks because populations were larger, continuous, and more extensive. We estimated from our mitochondrial data that a major demographic expansion of the NE Atlantic population of *S. setosa* started approximately 685,000 generations ago, during the mid-Pleistocene (assuming one generation per year). However, such estimates should be regarded as very approximate because they are based on a simple demographic model, there is a large variance associated with estimates of τ from a single locus, and estimates are highly reliant on the estimate of mutation rate and generation time of the organism.

In summary, our study of mitochondrial variation in two chaetognaths contributes to the existing literature of other superabundant and high-gene flow species that long term evolutionary (effective) population sizes may be relatively small (reviewed in Avise 2000). The comparison of two species with different distribution patterns has shown that planktonic species may be differentially affected by geological events such as ice ages because of differences in ecological traits. Furthermore, our study adds to a growing body of literature suggesting that mitochondrial markers should not be regarded as strictly neutrally evolving (*e.g.* Ballard and Whitlock 2004).

ACKNOWLEDGEMENTS

We thank all crew of the research vessel Pelagia, R. ter Hofstede, and J.-H. Schotveld for their help in taking plankton samples and C. ten Hallers-Tjabes for inviting KP, and H. van Aken for inviting KP and EH aboard

their research cruises. H. van Aken is also thanked for his calculations of the areas and volumes of the distributions of the two chaetognaths. W. van Ginkel, P. Kuperus, and B. Voetdijk are acknowledged for laboratory assistance and J. van Arkel for his help with figures. We thank G. Lecaillon, P. Meirmans, A. Pierrot-Bults, and A. Southward for discussion. Comments of J. Breeuwer, M. Egas, M. Genner, G. Lecaillon, P. Luttikhuizen, and S. Menken improved the manuscript.