Structure at open sea: genetics of zooplankton populations

Peijnenburg, K. T. C. A.

Citation for published version (APA):

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.
Summary

Little is known about the barriers at sea that limit dispersal between zooplankton populations. Yet, if we are to better understand how differentiation processes, and ultimately speciation, take place in the open waters, we need to determine the scales at which zooplankton populations are ecologically and evolutionarily independent. One way to study this would be to individually tag specimens and follow their movements through time and space. However, given the vast size of the ocean and the small size of zooplankton, this would be completely unfeasible. A good alternative is to use genetic markers to study the genetic trail that successful migrants leave behind. In this way, it is possible to indirectly estimate the connectivity of populations, *i.e.* the integrated evolutionary history of gene flow between populations, or, in other words, to determine the degree and scales of genetic structuring.

The central question of this thesis is *what are the scales of genetic structuring of zooplankton populations?*. The spatial scales that I have examined range from comparisons between three major European basins – the North East (NE) Atlantic, Mediterranean, and Black Sea – (Chapters 1, 2, and 3) to comparisons within these basins (Chapters 4 and 5). The temporal scales that I have considered range from vicariance events that took place on geological time-scales (Chapters 1, 2, and 3), to comparisons over several years (Chapters 4 and 5), and comparisons between juveniles and adults present within a single plankton sample (Chapter 5).

In this thesis, I have mainly focussed on the holoplanktonic chaetognath *Sagitta setosa* Müller 1847, because it provides an ideal study species to examine the impact of both present and past barriers on dispersal and therefore on differentiation processes in the pelagic realm. This species has a disjunct neritic distribution pattern (*i.e.* occurring over the continental shelf areas only) in European seas. It has been proposed that Mediterranean and Black Sea populations are glacial relicts, *i.e.* locally adapted remnants of Atlantic populations that colonized the Mediterranean resulting from the lowering of sea level and temperature during Pleistocene glacials. This
would imply a separation between Atlantic and Mediterranean populations of at least 18,000 years, since the last glacial maximum. Furthermore, morphological differences between populations from different basins have been reported.

In Chapter 1, it is shown that considerable morphological variation exists within the distribution range of *S. setosa*; specimens from the Mediterranean were smallest with relatively long caudal segments, and few teeth and hooks, whereas specimens from the Black Sea were largest with relatively short caudal segments and many teeth and hooks. Specimens from the NE Atlantic were intermediate with regards to these characters, but ranges overlapped and there were no differences in allometry. Moreover, I found more variation in quantitative morphological characters within *S. setosa* from different basins than between *S. setosa* and two closely related, but disputed taxa, *S. batava* Biersteker and Van der Spoel 1966 from the Scheldt Estuary (The Netherlands), and *S. eurina* Moltschanoff 1909 from the Black Sea. *Sagitta batava* conformed to *S. setosa* in terms of all morphological characters considered and comparisons with *S. elegans* Verrill 1873 in the original description were found to be based on misidentifications. Therefore, *S. batava* should no longer be considered a separate taxon. For *S. eurina* the data were inconclusive. However, the fact that all reported samples from the Black Sea either consisted entirely of *S. setosa* or *S. euxina* (depending on sampling season and depth) suggests that these samples may represent seasonal variants of one and the same species.

As quantitative morphological characters were highly variable within and overlapping between populations of *S. setosa*, I regarded them unsuitable to determine the degree of isolation between populations. Thus, in Chapter 2 a phylogeographic analysis is presented based on mitochondrial DNA (mtDNA) sequences comprising the Cytochrome Oxidase II (COII) region of 86 individuals collected from the NE Atlantic, Mediterranean, and Black Sea. Sequences were highly variable, each individual represented a different haplotype. The analysis also revealed strong phylogeographic structuring with four main groups corresponding to the NE Atlantic, Mediterranean Sea (including Ligurian Sea, Tyrrhenian Sea, and Gulf of Gabès), Adriatic Sea, and Black Sea. Two of these (Atlantic and Black Sea) were resolved as monophyletic clades, indicating that gene flow between disjunct populations of *S. setosa* is probably absent. The pattern of population divergence, deepest split between Atlantic and Mediterranean/Black Sea populations followed by a split between Mediterranean and Black Sea populations, is congruent with an hypothesized biogeographic scenario.
based on the geological and paleoclimatic history of the European basins. This scenario is based on a colonization of the Mediterranean from the Atlantic during one of the Pleistocene glacial periods, followed by a colonization of the Black Sea from the Mediterranean during the Holocene. However, the estimates of population divergence time based on net nucleotide divergences and a standard mitochondrial molecular clock (2% divergence per million years) indicated much deeper population divergences, namely 1.7 and 0.4 million years ago for the Atlantic/Mediterranean and Adriatic/Black Sea split, respectively, than the proposed dates of most recent possible contact based on paleo-data (18,000 and 7,000 years ago, respectively). Basically, this leaves us with two alternative explanations: either the standard rate of mitochondrial evolution is roughly correct and *S. setosa* populations diverged since the beginning and middle Pleistocene, or divergences are more recent and *S. setosa* has a faster mitochondrial clock than ever reported in the literature. I suggest two approaches to try and differentiate between these alternatives: (1) collect data from other unlinked (nuclear) loci to reduce the variance that is typically associated with estimating divergence times based on a single (mitochondrial) marker (*e.g.* microsatellites, see Chapter 5), and (2) study the phylogeography of other (holoplanktonic) species with comparable distribution patterns to *S. setosa* to test whether similar present and past barriers may have been present in unrelated taxa (*e.g.* the copepods *Calanus helgolandicus* Claus 1863 and *C. euxinus* Hulsemann 1991 would be ideal comparative taxa, see Chapter 3).

Chapter 3 presents a joined mitochondrial phylogeography for the calanoid copepods *C. helgolandicus* from the NE Atlantic and the Adriatic Sea and the closely related *C. euxinus* from the Black Sea. Cytochrome Oxidase I (COI) and Cytochrome B (CYTB) haplotypes were shared amongst individuals of the two species, demonstrating a very close genetic relatedness. Coalescent analyses of COI sequences indicated a similar pattern of population divergence compared to *S. setosa*, with the NE Atlantic-Mediterranean divergence being the deepest, and the Adriatic-Black Sea divergence being the most recent. Moreover, these analyses showed that divergences dated back to the middle Pleistocene, and were thus, similar to *S. setosa*, much older than the estimated dates of most recent colonization of the Mediterranean and Black Sea basins based on paleo-scenarios. These results do not rule out that the assumed colonizations took place, but suggest that planktonic populations that colonized these basins were carrying, and have since retained, a large amount of pre-existing mitochondrial divergence. I thus consider the 'deep split' scenario as pro-
posed for *S. setosa* more likely, in which I suggest that several glacial cycles may have contributed to the present genetic divergence between plankton populations, possibly reinforced by local adaptation, divergent selection, bottlenecks, and/or founder-flush.

To test the prediction that neritic species, because of smaller and more fragmented populations, would have been more vulnerable to population size reductions resulting from range contractions and displacements during glacial cycles than oceanic species, I contrasted the mitochondrial signatures of the two dominant chaetognaths of the NE Atlantic, the neritic *S. setosa* and the more oceanic *S. elegans* in Chapter 4. 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 of *S. elegans* compared to *S. setosa*. Mitochondrial signatures of the two species differed strikingly. *Sagitta setosa* haplotypes produced a star-like phylogeny and a unimodal mismatch distribution, indicative of a bottleneck followed by population expansion, whereas *S. elegans* haplotypes produced a deeper phylogeny and a multimodal mismatch distribution, expected for a more stable population. Assuming that selective effects have been similar for *S. setosa* and *S. elegans*, these data can only be explained by contrasting demographic histories of the two species, and are thus consistent with the starting hypothesis. Even though high levels of genetic diversity were present in populations of both species, much more diversity would have been expected based on estimates of their present census population sizes according to predictions based on neutral theory. I estimated a very large ratio between census population size and evolutionary effective population size (*N*/*N_e*) of ~10^−10. One of several factors that could explain this discrepancy is directional selection on mtDNA, and in this chapter, some evidence is presented for this based on results from a McDonald-Kreitman test.

In Chapter 5, I present data based on a high sampling intensity of *S. setosa* from three European seas, and a combination of mitochondrial and four newly developed nuclear-encoded microsatellite markers. Both marker sets indicated significant differentiation between populations from different basins, confirming earlier results that gene flow is probably absent (Chapter 2). This indicates that contemporary barriers to dispersal must exist between disjunct *S. setosa* populations within seemingly continuous habitat. I suggest that the disjunction is not maintained because of a limited dispersal ability (in fact, the global ocean turns over so rapidly that, by just going with the flow, all planktonic species would become ubiquitous within 500 years!). But rather, it is
SUMMARY

maintained because of the inability to maintain viable populations in suboptimal, geographically intermediate areas (e.g., in the NE Atlantic south of 45° N and the Levantine basin in the eastern Mediterranean Sea). As expected from hydrographic characteristics of the different basins, no evidence was found of population structuring within the NE Atlantic whereas significant differentiation was detected within the Mediterranean basin. Though patterns from mitochondrial and nuclear markers were largely congruent qualitatively, the levels of differentiation were quantitatively very different, with far less pronounced structure detected by microsatellites. This difference may be explained by a larger effective population size of nuclear compared to mitochondrial genomes (reducing drift effects), and several technical problems associated with the use of microsatellites in large populations, including homoplasy, hypervariability, null alleles, and a larger potential for sampling bias. Furthermore, several highly divergent mitochondrial clades were uncovered that did not appear concordant with morphology, geography, or nuclear DNA data. More data are needed to resolve the evolutionary origin of these clades (see 'future challenges' below). To explain the enormous discrepancy between apparent rates of mitochondrial and nuclear evolution as observed for S. setosa, a new hypothesis is put forward which suggests that in very large populations, only slight positive selection pressures (e.g., on mitochondrial genomes) may magnify the observed differences in evolutionary rate compared to neutral loci.

Taken together, the results of my work challenge the commonly held view that for marine holoplanktonic organisms, dispersal with the ocean currents provides high levels of gene flow across large spatial scales, and therefore lead to little or no population genetic structuring and differentiation.

MAIN CONCLUSIONS

- Morphological divergence of Sagitta setosa, as measured by simple quantitative methods, does not reflect population divergence, as inferred from mitochondrial DNA.
- The chaetognath S. setosa is genetically extremely diverse.
- Sagitta setosa populations from the NE Atlantic, Mediterranean Sea, and Black Sea are not connected by gene flow.
- The mitochondrial phylogeography of S. setosa shows that Atlantic and Mediterranean populations diverged first, followed by a split between Mediterranean and Black Sea populations. This is consistent with a colonization scenario of the Black Sea from the Mediterranean.
- Present and past barriers to dispersal of European populations of the copepods *Calanus helgolandicus* and *C. euxinus* are similar to those observed for the chaetognath *S. setosa*. This confirms that the two copepod species have had a similar biogeographic history to that of *S. setosa*.
- Considerably greater mitochondrial divergence is present between Atlantic, Mediterranean, and Black Sea populations of *S. setosa* and *C. helgolandicus/eu.rinus* compared with expectations based on paleoclimatic scenarios that suggested final population splits took place only 18,000 and 7,000 years ago.
- Based on mitochondrial sequence data, the two species *C. euxinus* and *C. helgolandicus* are not more genetically divergent than populations of *C. helgolandicus* from the NE Atlantic and Adriatic are from each other. This calls into question the species status of the Black Sea *C. euxinus*, or alternatively, the conspecific status of the *C. helgolandicus* populations.
- Mitochondrial diversity within NE Atlantic populations of the neritic *S. setosa* is much lower compared to the more widespread *S. elegans*. Assuming that selection pressures have been similar for the two species, this supports the hypothesis that neritic species have been more vulnerable to population size reductions resulting from Pleistocene range shifts than oceanic species.
- Evolutionary effective population sizes of the chaetognaths *S. elegans* and *S. setosa*, as inferred from estimated levels of mitochondrial diversity, are much lower than census population sizes, as estimated from abundance data.
- *Sagitta setosa* constitutes a single panmictic (randomly interbreeding) population in the NE Atlantic.
- Mediterranean populations of *S. setosa* are genetically structured, suggesting that present and past barriers have limited gene flow between populations in different sub-basins.
- Cross-fertilization is the predominant mode of reproduction in natural *S. setosa* populations, as inferred from Hardy-Weinberg proportions of genotypes at two microsatellite loci.
- Hypervariability, null alleles, and homoplasy of microsatellites are nuisance factors, but an inevitable reality when dealing with large marine populations.

**Future challenges**

As my PhD work has been one of few studies into the levels and patterns of mitochondrial and nuclear genetic diversity in zooplankton, at least as
many questions have been raised as have been answered. I would like to end this thesis by mentioning a few of these questions and indicate some ideas for future research.

- An interesting observation for both *S. setosa* and *C. helgolandicus/euxinus* populations from European basins is that morphologically, Mediterranean and Black Sea populations are at opposite ends of the spectrum whilst genetically, they are more closely related to each other than to Atlantic populations. The Mediterranean and Black Sea basins are also at opposite ends of the environmental spectrum (*e.g.* in terms of temperature and salinity). This raises the question of whether the observed morphological differences are phenotypically plastic or have a genetic basis. Our finding that populations from different basins are not connected by (high levels of) gene flow leaves the possibility open that morphological differences are the result of adaptation to local environmental conditions. Testing this will not be an easy task, however. One approach would be to rear different populations under similar laboratory conditions in a ‘common garden experiment’. This should be feasible for the copepods, but will be very difficult, if not impossible, for planktonic chaetognaths which are extremely vulnerable to sampling and transport and generally die within a few hours of captivity.

- One of the surprising outcomes of my work is the discovery of several extremely divergent mitochondrial clades within *S. setosa* samples that do not appear concordant with morphology, geography, and nuclear DNA data. Individuals from these highly divergent clades may either be present in currently interbreeding populations of *S. setosa*, or they may represent cryptic species. The first option seems more likely because congruent divergence at nuclear loci would have been expected given the enormous levels of observed mitochondrial divergences. However, to test this, more data are needed, particularly DNA sequence data from single-copy nuclear coding genes (*e.g.* elongation factor) as well as from nuclear non-coding regions (*e.g.* introns). To more thoroughly examine the possibility of cryptic species, scanning electron micrography of chaetognath teeth and hooks may be a promising approach as a great number of different, species-specific, surface structures have been revealed this way. In any case, to place the extremely divergent mitochondrial lineages within any kind of systematic context, a (molecular) species phylogeny of the Chaetognatha, and more particularly of the genus *Sagitta*, is desperately needed. Ideally, such a phylogeny should be based on mitochondrial and nuclear data, and should consist of the
widest possible taxon sampling, both in terms of the number of different species and geographical sampling within species.

- At the end of Chapter 5, I expound a theory that may explain the enormous discrepancies between rates of mitochondrial and nuclear evolution as observed for chaetognaths. Basically, I propose that in very large populations, small positive selection pressures on mitochondrial genomes may magnify the observed differences in substitution rate compared to neutral loci (e.g., microsatellites). It would be interesting to examine nuclear variation in copepods to verify this theory. However, it is an interesting result in itself that levels of mitochondrial diversity and divergence (from similar protein-coding genes) in the copepods *C. helgolandicus* and *C. euxinus* were much lower than those observed in the chaetognaths *S. setosa* and *S. elegans*. As copepods have even larger estimated census population sizes than chaetognaths, we would have expected the opposite. One potential explanation may be that chaetognaths have a faster mitochondrial evolutionary rate. It may be that more mutations arise in chaetognath mitochondrial genomes because of a 'sloppy' DNA repair mechanism. This may also be an explanation for the fact that chaetognath mitochondrial genomes contain such few genes (see general introduction of this thesis). It would therefore be extremely interesting to try and calibrate a molecular clock for chaetognaths. As the chaetognath fossil record is virtually non-existent, one possibility may be to study the genetic divergence between populations or species pairs on opposite sides of the Panama Isthmus in order to place some bounds around evolutionary rates in chaetognaths (e.g., *S. friderici* and *S. tenuis* would be extremely interesting candidates, as they have been proposed to be closely related to *S. setosa*, and have disjunct populations in the South East Atlantic Ocean as well).

To further examine the possible link between evolutionary rates and large population sizes, it would be necessary to study other of the many unstudied groups of zooplankton. A particularly interesting group to look at from a genetics perspective would be pteropods. As they are the only group of macroplankton with a good fossil record, thanks to their shells, they may be the perfect candidates to solve this issue.