

## Global population structure in selected species of the planktonic diatom genus *Chaetoceros*

Director of studies: Wiebe Kooistra

Department of Integrative Marine Biology

Planktonic diatom species are widely distributed, showing a patchwork of disconnected geographical populations. These populations could share a common gene pool, i.e., belong to a single biological species or they could consist of a complex of genetically and biologically distinct, morphologically cryptic species. However, these possibilities represent the extreme cases along a gradient with fussy cases in between. An inherent issue in speciation is that there should be a gradient between “one species” and “two distinct species.” The very word –speciation- indicates a process moving from one to multiple, biologically separated sister species.

In order to capture different “points” in the speciation process, I propose to use High Throughput Sequencing (HTS) meta-barcode sequences of the V4 region in the 18S rDNA as markers to assess population structuring in a series of widely distributed morpho-species belonging to the planktonic genus *Chaetoceros*. The V4-region is part of nuclear encoded rDNA cistron present in hundreds of copies in each haploid genome. Each population shows a dominant V4-haplotype and several surrounding minor haplotypes; closely related populations are expected to share a similar haplotype make-up whereas more distantly related populations are expected to show a more divergent make-up of minor haplotypes, and even a different dominant haplotype. All the HTS sequences belonging to a series of locally common *Chaetoceros* species have been obtained from the Neapolitan LTER Marechiarra HTS meta-barcode V4-dataset and these need to be sorted into unique haplotypes. Haplotype networks will be inferred and seasonality mapped over these networks to assess if different seasons host different populations of the same species (to assess population differentiation by season). Then, V4-sequences will be collected from what are considered the same species or closely related species from a series of HTS-datasets obtained at different geographic sites (BioMaRKs, TARA, etc.). The sequences are grouped in clades, taxonomically verified using reference Sanger sequences generated from taxonomically authenticated strains, and then sorted into haplotype networks. Geography is mapped over those networks and patterns elucidated. Results are expected to reveal a series of cases along a continuum:

- i) two geographic samples exhibit the same dominant haplotype and a similar distribution of sequences over the minor haplotypes;
- ii) two geographic samples exhibit markedly different distributions of sequences over the shared minor haplotypes, and they show different minor haplotypes;
- iii) two geographic samples exhibit different dominant haplotypes and different minor haplotypes, but all still group within the same cluster or terminal clade;
- iv) two distinct but closely related clusters with minor overlap due to a few ancestral haplotypes
- v) two distinct terminal clades.

Nowhere along this “progress of speciation” a threshold can be drawn beyond which one species can be declared two or more. Speciation implies a transition zone “speciation in progress”; and the core objective of the study is to illustrate cases along such a gradient. The obtained haplotype networks can be used also to test biogeographic hypotheses of events separating the populations into geographically or climatologically disjunct ones, as well as provide ad-Hoc explanations for recent jump dispersal events.

The intellectual merit of the proposed activity is in demonstrating that HTS meta-barcode sequences can be used in many more applications than just biodiversity appraisals, one of the applications being assessment of population genetic structure. The broader impacts resulting from the proposed activity is that it will lead to a series of comparable studies.