The Open University, London
Stazione Zoologica Anton Dohrn,
Naples

Diversity, distribution and evolution of the planktonic diatom family Chaetocerotaceae

Chetan Chandrakant Gaonkar

Thesis submitted for the degree of Doctor in Philosophy
(PhD) in Life and Biomolecular Sciences

July 2016

I dedicate this thesis to my mother Savita and my father
Chandrakant Gaonkar

Abstract i
  Acknowledgements ii
  List of tables iii
  List of figures iv
  Abbreviations vii
Chapter 1. General Introduction 1
Chapter 2. Materials and Methods 25
Chapter 3. Phylogeny of the marine planktonic diatom family Chaetocerotaceae
  Abstract 41
  Introduction 42
  Materials and Methods 44
  Results 58
  Discussion 124
  Conclusion and outlooks 133
Chapter 4. Species diversity of Chaetocerotaceae in the Gulf of Naples as revealed by High Throughput Sequencing - meta-barcoding
  Abstract 135
  Introduction 136
  Materials and Methods 138
  Results 147
  Discussion 172
  Conclusion and outlooks 182
Chapter 5. Morphological and molecular diversity of *Chaetoceros socialis* Lauder
  Abstract 185
  Introduction 186
  Materials and Methods 190
  Results 194
  Discussion 218
  Conclusion and outlooks 222
Chapter 6. Characterization of introns in the nuclear encoded ribosomal RNA of the planktonic diatom family Chaetocerotaceae
  Abstract 225
  Introduction 226
  Materials and Methods 231
  Results 233
  Discussion 246
  Conclusion and outlooks 250
Chapter 7. General Conclusion 251
Bibliography 261
Appendix - Glossary

Diversity, distribution and evolution of the planktonic diatom family Chaetocerotaceae

Chetan Chandrakant Gaonkar

Abstract

The number and abundance of diatom species in environmental samples are counted traditionally by means of light microscopy (LM). However, recognizing –let alone, counting-species is often challenging because of the existence of cryptic species and intraspecific phenotypic plasticity. Proper characterization requires isolation of cells, growing them into monoclonal cultures, and characterizing the cultures genetically and morphologically. However, not all species grow in culture, featureless ones are less likely to be isolated, and the procedure is laborious. High-throughput sequencing (HTS) metabarcoding bypasses morphology; DNA is collected from environmental samples, a particular marker sequenced, and the resulting sequences sorted into clusters or terminal clades assumed to represent species. Yet, reference barcodes of taxonomically validated species are needed to identify these clades. This exercise is the main aim of my thesis.

Since it is impossible to do this for all the diversity within a PhD thesis project, we selected Chaetocerotaceae, an abundant and diverse family of marine planktonic diatoms, containing two genera: *Chaetoceros* and *Bacteriastrum*. Its members uniquely share setae; thin siliceous tubes emerging from the valve corners, facilitating detection in samples. Strains were obtained from the Gulf of Naples (GoN), from Central Chile and Roscoff – at sites for which LTER time series data are available.

A total of 270 strains were obtained from these sites, and their 18S- and partial 28S rDNA sequences and morphological information gathered. The strains grouped into 60 genetically distinct species, thus providing a dataset of validated Chaetocerotacean 18S reference barcodes. Inferred molecular phylogenies showed monophyletic Chaetocerotaceae as well as monophyletic *Bacteriastrum* inside paraphyletic *Chaetoceros*, and the presence of cryptic diversity. To start with taxonomic updates, the species *C. sporotruncatus* and *C. dichatoensis* were described within the *C. socialis* species-complex based on spore morphology and sequence differences. Several rDNA sequences contained spliceosomal introns (ca. 100bp) and/or group-I introns (ca. 400bp). Phylogenies inferred from the introns did not corroborate rDNA phylogenies, suggesting horizontal gene transfer. Presence/absence of introns in conspecific strains sampled in different seasons suggests population differentiation between these seasons.

A HTS dataset consisting of V4-sequences (part of 18S) from 48 seawater samples taken over the seasons in the GoN revealed 76 terminal clades of which 46 grouped with a reference barcode. Some of these species occur year-round whereas most others are seasonal. Surprisingly, of the 30 clades belonging to unknown Chaetocerotacean species, two appear to be among the most abundant in the GoN.