Phylogenetics and Phylogeography in the Planktonic Diatom Genus *Chaetoceros*

Daniele De Luca (M.Sc.)

Personal ID: F3576948
Abstract

The initial aims of this thesis were to assess the systematics of the planktonic diatom genus *Chaetoceros* and the phylogeographic patterns of selected species in this genus across spatial and temporal scales. As expected in every experiment, some initial ideas have been pursued as they were; others have taken a different route and led to different questions. Consequently, the systematics of *Chaetoceros* has become a multigene phylogeny and a revision of the classical taxonomic scheme for the family Chaetocerotaceae (Chapter II). Then, the phylogeographic approach, initially meant as a Sanger sequencing of a few genes from specimens collected around the world, turned into the analysis of the *C. curvisetus* cryptic species complex by using an approach which combines haplotype networks and metabarcoding data (Chapter IV). The mapping of this complex against a temporal metabarcoding dataset (MareChiara, Gulf of Naples, IT) has become a story of concerted evolution and has been extended to different *Chaetoceros* species and supported by a single strain 18S-V4 high throughput sequencing (Chapter V). Amid these experiments, the potential of metabarcoding data for biological recording was explored and tested in the whole genus *Chaetoceros* to assess diversity and distribution (Chapter III). Such data were integrated with classical ones from public repositories and literature and used to produce, among the other results, distribution maps of *Chaetoceros* species.
To my Mom,

For having taken on so many burdens,

allowing me to fully dedicate myself to this thesis

For having endured me during my rainy days

For having surrendered to the constant entry of books in the house:

one day all these books will drive us away.
Acknowledgements

It is almost impossible to thank all the people who made my time at the SZN an unforgettable and great experience. Most of them contributed with a simple “buongiorno”, a smile, or a chat on the terrace, and to them goes my gratitude. Others occupy a more special place in my heart. Therefore, my “special” thanks go…

…To my Director of Studies, Dr. Wiebe Kooistra, who gave me the opportunity to carry out my Ph.D. research and the freedom to develop my ideas, showing me the way, provide guidance and alternative routes, and not dictating the rules. This is the best gift I could have ever received.

… To my internal supervisor, Dr. Diana Sarno, for the taxonomic knowledge she shared with me, and her passion for diatoms, which she transmitted to me and which inspired me to immerse myself fully into the diversity of diatoms and, in particular, Chaetoceros.

… To my external supervisor, Prof. Dr. Christine A. Maggs, for the time spent together during her visits at the SZN, talking about my project, science in general and ordinary life. Her suggestions and enthusiasm were pure positive energy that prompted me to do better and better.

…To Dr. Roberta Piredda, for the invaluable time she invested in me teaching, discussing, listening and for her friendship. Thanks to her, I have domesticated the untamed beasts of “Linux”, “batching”, “server” and “big data”, I was introduced to the –omics world, and looked with new eyes to the ecology of communities. I own her everything I know about metabarcoding and much more about evolutionary methods for the study of biodiversity. Someone said: “together you are war machines”. Nothing more true.

… To Carmen Minucci, for the training in the lab and the constant help in carrying it on.
… To Dr. Grazia Quero, for being my friend and the best office partner I could ever have desired. Thank you for introducing me to the metagenomics world and for the constant help every time the machine took control over the man.

… To Dr. Isabella D’Ambra, for the morning chats and the courage with which she faces life every day. You are an example of success.

… To the old and new people of the Procaccini lab, my first house at the SZN: Gabriele, Miriam, Chiara, Emanuela, Domenico, Gaetano, Lazaro, Marlene, Lucia, Roberto, Jessica... Among you are head of departments, researchers, post-docs, but there is no “Dr” because you will be always my second family first, and then the brilliant scientists and lovely persons I have ever met. I am proud to have been a Procaccinidae.

… To Dr. Graziano Fiorito, my first “official” mentor at SZN for the esteem and affection he has always shown to me. I was proud to be “his compulsive student”.

… To Pamela, Elena, Giovanna, Ruth, Caitlin, Paola and Andrea of the “OctoLab” for being my friends and all the lunches and “mozzarella days” we had together. Pam, we started this adventure almost together several years ago and I am proud of how far you went. The spirit of my “Pam!” along the corridors will accompany your for many years to come.

… To the people of the SBM: the big boss, Dr. Elio Biffali, who made me a better molecular biologist with his original way of teaching; Dr. Pasquale De Luca, for adopting me as “cousin” and witnessing all my feats; Elvira Mauriello and Raimondo Pannone, for all the support in molecular biology techniques and the chats and laughs together. I own you thousands of DNA sequences!

… To Gabriella Grossi and Daniela Consiglio for the support in the life as Ph.D. student and student’s representative the first, and as reference student for the invited speakers the latter. Thanks for the passion you have for your work and the kindness showed.
**Table of Contents**

**Chapter I - Introduction**

1.1. Diversity, the hallmark of living organisms 3
   
   1.1.1. Diatom diversity and evolution 5

1.2. The species problem 8

   1.2.1. Species concepts in diatoms 11

1.3. Do species really exist? 14

1.4. The need for classification 16

1.5. DNA barcoding 18

1.6. From barcodes to metabarcodes 20

   1.6.1. DNA barcoding and metabarcoding in diatoms 21

1.7. Case study: the planktonic diatom family Chaetocerotaceae, with emphasis on the genus *Chaetoceros* 24

   1.7.1. Fossil record of *Chaetoceros* 27

   1.7.2. The ecological and evolutionary importance of *Chaetoceros* 28

   1.7.3. Aim of Ph.D. thesis 30

References 31

**Chapter II - Inferring the evolutionary history of Chaetocerotaceae**

2.1. Introduction 55

   2.1.1. Systematics of Chaetocerotaceae 55

2.2. Materials and methods 58

   2.2.1. Taxon sampling, outgroups selection and DNA extraction 58

   2.2.2. Selection of genes, amplification and sequencing 59

   2.2.3. Sequence editing and alignment 61

   2.2.4. Nucleotide composition and substitution saturation analyses 62
Chapter III - Assessing diversity and distribution in *Chaetoceros*: integration of classical and novel strategies

3.1. Introduction

3.1.1. Primary Biodiversity Data: recording the occurrence of species

3.1.2. Primary Biodiversity Data for planktonic species

3.1.3. Aim of this work

3.2. Materials and methods
3.2.1. Data collected from available public repositories, literature and checklists

3.2.2. Data generated from molecular sources

3.3. Results

3.3.1. Data collected from available public repositories, literature and checklists

3.3.2. Data generated from molecular sources

3.4. Discussion

3.4.1. Global distribution of Chaetoceros

3.4.2. Abundance of Chaetoceros at global scale

3.4.3. Integration of literature and metabarcoding data: three study cases in Chaetoceros

3.4.4. Assessing species distribution in Chaetoceros

3.4.5. Future directions

References

Appendix III

Chapter IV - Resolving the Chaetoceros curvisetus cryptic species complex

4.1. Introduction

4.1.1. Cryptic species complexes: origin, distribution and methodology of study

4.1.2. The Chaetoceros curvisetus species complex

4.1.3. Objectives of the study

4.2. Materials and methods

4.2.1. Download and processing of metabarcoding data
4.2.2. Phylogenetic haplotype network inference

4.2.3. Genetic divergence among species and variability within species

4.2.4. Global distribution of taxa belonging to the C. curvisetus species complex

4.3. Results

4.3.1. Validation of C. curvisetus candidate sequences

4.3.2. Phylogenetic haplotype networks

4.3.3. Genetic differentiation and variability

4.3.4. Global distribution of taxa belonging to the C. curvisetus species complex

4.4. Discussion

4.4.1. Phylogenetic relationships among taxa belonging to the C. curvisetus species complex

4.4.2. Distribution of taxa belonging to the C. curvisetus species complex

4.4.3. Considerations on sequence variation in metabarcoding data

References

Appendix IV

Chapter V - Concerted evolution in Chaetoceros

5.1. Introduction

5.2. Materials and methods

5.2.1. Selection of taxa to study concerted evolution

5.2.2. Analysis of environmental sequences

5.2.3. Single strain HTS

5.2.4. Data pre-processing and analysis of single-strain HTS
5.2.5. Testing the concerted evolution hypothesis  

5.3. Results  
5.3.1. General characteristics of the datasets  
5.3.2. Abundance plots from environmental metabarcoding and single strain HTS  
5.3.3. Blast of environmental haplotypes vs. single strain  
5.3.4. Phylogenetic networks from environmental samples  

5.4. Discussion  
5.4.1. Concerted evolution in Chaetoceros  
5.4.2. Implications for DNA barcoding  
5.4.3. Copy number across the Tree of Life and possible role of rDNA heterogeneity  
5.4.4. Conclusions  

References  
Appendix V  

Chapter VI – Concluding remarks and future perspectives  
6.1. Concluding remarks  
6.2. Future perspectives  

References
List of Figures

Chapter I

Fig. 1.1. Modes of evolution across space and time. 4
Fig. 1.2. Early example of diatom illustrations. 12
Fig. 1.3. Some classification essays from the sixteenth to the nineteenth century. 17
Fig. 1.4. Main target genes utilised for DNA barcoding in diatoms. 22
Fig. 1.5. Main morphological features of Bacteriatrum and Chaetoceros. 26

Chapter II

Fig. 2.1. Different orientation of terminal setae on the terminal valves of a colony of Bacteriatrum sections Isomorpha (A) and Sagittata (B). 56
Fig. 2.2. Chloroplasts disposition in the subgenera Chaetoceros (A) and Hyalochaete (B) of Chaetoceros. 57
Fig. 2.3. Schematic representation of a typical Chaetoceros species, with the main morphological features relevant to this analysis. 64
Fig. 2.4. Individual and concatenated sequence alignments of Chaetocerotaceae dataset. 66
Fig. 2.5. Multigene Maximum Likelihood and Bayesian phylogenetic trees. 69
Fig. 2.6. Chaetoceros danicus (A) and C. rostratus (B), two members of the Section Chaetoceros. 81
Fig. 2.7. Chaetoceros costatus, Section Costata. 82
Fig. 2.8. Chaetoceros minimus (A) and C. throndsenii (B), two members of the Section Minima. 82
Fig. 2.9. Chaetoceros affinis (A) and C. diversus (B), two members of the Section Stenocincta. 83

Appendix II

Fig. A2.1. Light microscopy photographies of Bacteriatrum and Chaetoceros
species utilised in the present study.

Fig. A2.2. Maximum Likelihood (ML) tree of concatenated nuclear genes (18S and 28S).

Fig. A2.3. Maximum Likelihood (ML) tree of concatenated plastid genes (rbcL and psbA).

Fig. A2.4. Maximum Likelihood (ML) tree of mitochondrial COI gene.

Fig. A2.5. Maximum Parsimony (MP) tree.

Chapter III

Fig. 3.1. Graphical representation of the main workflow utilised.

Fig. 3.2. Occurrence of Chaetoceros using (A) GBIF and (B) OBIS data.

Fig. 3.3. Occurrence of Chaetoceros using literature data.

Fig. 3.4. Species richness of Chaetoceros estimated from literature data.

Fig. 3.5. Chaetoceros distribution according to OSD (A) and Tara Oceans (B) data.

Fig. 3.6. Log10 abundance of Chaetoceros reads according to OSD (A) and Tara Oceans (B) datasets.

Fig. 3.7. Distribution of C. tenuissimus (A, B), C. gelidus (C, D) and C. neogracilis (E, F) according to literature (orange dots) and metabarcoding data (blue dots for OSD and red dots for Tara Oceans).

Appendix III

Fig. A3.1. Distribution maps of Chaetoceros species using OSD and Tara Oceans datasets.

Chapter IV

Fig. 4.1. Chaetoceros curvisetus (A) and C. pseudocurvisetus (B).

Fig. 4.2. Light microscopy photographies of the known members of the C. curvisetus species complex.
Fig. 4.3. Occurrence of taxa belonging to the *C. curvisetus* species complex in OSD (A) and Tara Oceans (B) datasets.

Fig. 4.4. TCS haplotype network for the *C. curvisetus* species complex according to OSD data.

Fig. 4.5. TCS haplotype network for the *C. curvisetus* species complex according to Tara Oceans data.

Fig. 4.6. Maximum Likelihood tree of the *C. curvisetus* species complex based on representative sequences of V4 data.

Fig. 4.7. Maximum Likelihood tree of the *C. curvisetus* species complex based on representative sequences of V9 data.

Fig. 4.8. Distribution of the *C. curvisetus* species complex in Longhurst provinces.

Fig. 4.9. Heatmap showing the abundance of *C. curvisetus* spp. in each Longhurst province according to OSD data.

Fig. 4.10. Heatmap showing the abundance of *C. curvisetus* spp. in each Longhurst province according to Tara Oceans data.

**Chapter V**

Fig. 5.1. Abundance plots for each *Chaetoceros* species from validated environmental sequences.

Fig. 5.2. Abundance plots for each strain analysed in different *Chaetoceros* species.

Fig. 5.3. TCS haplotype network for *C. anastomosans* inferred from the MareChiara temporal dataset.

Fig. 5.4. TCS haplotype network for *C. costatus* inferred from the MareChiara temporal dataset.

Fig. 5.5. TCS haplotype network for *C. curvisetus* 2 inferred from the MareChiara temporal dataset.
Fig. 5.6. TCS haplotype network for *Chaetoceros* sp. Na11C3 (left) and Na26B1 (right) inferred from the MareChiara temporal dataset.

Fig. 5.7. TCS haplotype network for *C. tenuissimus* inferred from the MareChiara temporal dataset.
List of Tables

Chapter II

Table 2.1. List of the primers used for phylogenetic inference. 59
Table 2.2. Classification scheme of the family Chaetocerotaceae. 71

Appendix II

Table A2.1. List of taxa (species and strains) utilised in the present study, including sampling localities and dates and accession numbers for each gene amplified. 109
Table A2.2. Tests of substitution saturation for \textit{rbcL} (a), \textit{psbA} (b) and \textit{COI} (c) genes. 114
Table A2.3. Chi-squared test of homogeneity of state frequencies across taxa. 115
Table A2.4. Traditional classification scheme for the family Chaetocerotaceae. 121

Chapter IV

Table 4.1. List of reference sequences utilised for gathering \textit{C. curvisetus}-like taxa. 181
Table 4.2. Pair-wise genetic differentiation between \textit{C. curvisetus} species in OSD (A) and Tara Oceans (B) datasets. 193
Table 4.3. Average evolutionary divergence over sequence pairs within species. 195

Appendix IV

Table A4.1. List of OSD and Tara Oceans sites in which were found metabarcodes validated as \textit{C. curvisetus} spp. 217

Chapter V

Table 5.1. List of outgroup taxa for the validation of \textit{Chaetoceros}-species sequences. 226
Table 5.2. List of strains utilised for single-strain HTS. 228
Table 5.3. Number of environmental sequences and haplotypes utilised in this study. 232
Table 5.4. Number of sequences before and after pre-processing and total number of haplotypes utilised in each strain. 233
Table 5.5. Correspondence between the reference barcode (Sanger sequence) of each species and the dominant haplotypes of the environmental dataset (MareChiara) and single strain HTS.

Table 5.6. Summary of percentage of identity found between environmental haplotypes and single strain in each Chaetoceros species.

Appendix V

Table A5.1. List of the 50 most abundant haplotypes of MareChiara dataset and relative abundance.

Table A5.2. List of the 50 most abundant haplotypes in each strain and relative abundance.

Table A5.3. Percentage of identity between MareChiara haplotypes (query) and single strain ones (subject) after blast analysis.
**List of acronyms and abbreviations**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BS</td>
<td>Bootstrap Support</td>
</tr>
<tr>
<td>COI</td>
<td>Cytochrome Oxydase subunit I</td>
</tr>
<tr>
<td>GBIF</td>
<td>Global Biodiversity Information Facility</td>
</tr>
<tr>
<td>LSU</td>
<td>Large Subunit ribosomal DNA</td>
</tr>
<tr>
<td>OBIS</td>
<td>Ocean Biogeographic Information System</td>
</tr>
<tr>
<td>OSD</td>
<td>Ocean Sampling Day</td>
</tr>
<tr>
<td>PP</td>
<td>Posterior Probability</td>
</tr>
<tr>
<td>psbA</td>
<td>photosystem II protein D1 gene</td>
</tr>
<tr>
<td>rbcL</td>
<td>ribulose bisphosphate carboxylase Large subunit gene</td>
</tr>
<tr>
<td>sp.</td>
<td>species (singular)</td>
</tr>
<tr>
<td>spp.</td>
<td>species (plural)</td>
</tr>
<tr>
<td>SSU</td>
<td>Small Subunit ribosomal DNA</td>
</tr>
</tbody>
</table>
Chapter I

Introduction
1.1. Diversity, the hallmark of living organisms

The most peculiar characteristic of life on Earth is its sheer and unfathomable diversity. It is so huge and widespread that, to date, we are still remarkably uncertain about how many species exist on Earth (May, 1992; Stork, 1993). Guesstimates vary by several orders of magnitude and show remarkably different levels of uncertainty, from 3-100 million species (May, 2010) to around 8.7 million (Mora et al., 2011). In contrast to our persistent uncertainty about the extent of biological diversity, our comprehension of the mechanisms giving rise to it is becoming well understood. The generation of biological diversity (biodiversity; Wilson, 1988), at least at the gene- and species levels, is an intrinsic property of evolution (Wilson, 1992). As outlined by Darwin in the *Origin* (1859), evolution, in its essence, is nothing but “descent with modification”, a dualistic process ruled by chance and anti-chance factors (Mayr, 1997). The former is the modification, i.e. whatever heritable change affecting both the genotype and phenotype, produced “by chance” in the form of mutation, recombination or any other mechanism. One of the latter is natural selection, a mechanism that favours certain individuals over others with particular genetic attributes to fit in a specific environment, i.e., “survival of the fittest.” Such combination of chance and “anti-chance” factors gives evolution flexibility and “goal-directedness” and makes it so powerful (Mayr, 1963).

Darwin (1859) clearly recognised that evolution follows two kinds of trajectories, one across time and another one across space (see also Wilson, 1992). The former, called *phyletic evolution* is a process of gradual change within a single population or metapopulation of a species, resulting in the gradual transition of an ancestral species into a new one (anagenesis). As consequence, phyletic evolution does not imply speciation and can be seen as a line from ancestral to descent taxa (Fig. 1.1A). This is the mode of evolution Darwin (1859) had in mind when explaining the action of natural selection. On the contrary, when changes occur over time and are spread over space, (e.g. if populations of a species become different and occupy different ecological niches or geographic areas),
then an ancestral species splits gradually into two or more daughter species (Fig. 1.1B). This process, called *divergent evolution* (Gulick, 1888) is nothing but the change within a lineage accompanied by speciation (cladogenesis), and can be drawn as a branching tree. Divergent evolution implies that several species may all exist at the same time and is the source of biodiversity that, using the words of E. O. Wilson (1992) is “a collateral effect of evolution”.

![Fig. 1.1. Modes of evolution across space and time.](image)

(A) phyletic evolution (anagenesis); (B) divergent evolution (cladogenesis). Dots represent individuals of a population (species). Colours refer to variation among individuals.

As stated above, modification is an integral part of evolution as common descent. Whatever change not transmitted to the offspring has no consequence for evolution. Furthermore, the rate at which mutations arise (mutation rate) is highly variable across organisms (Drake, 1999; Baer et al., 2007) and the spread of such mutations within a
population through gene flow, drift and selection following the rules of population genetics, eventually determines the evolution of organisms. All these processes affect the way we perceive “species”. The adaptation to different ecological niches due to divergent natural selection and sexual reproduction have been indicated as the main factors responsible of genetic and phenotypic discontinuities between populations (Maynard Smith and Szathmáry, 1995; Coyne and Orr, 1998). Such discontinuities can also be observed at the molecular level. As already pointed out in the nineteenth century by the English geneticist William Bateson in its work *Materials for the study of variation* (1894), the variation of biological characters can be both continuous and discontinuous, and that “variations of a discontinuous nature may play a preponderant part in the constitution of a new species”. But what if characters change slowly over time? What if there are no discontinuities? How do we recognise species in that case?

1.1.1. Diatom diversity and evolution

Diatoms are one of the most successful contemporary groups of photosynthetic eukaryotic microorganisms. The estimated number of species ranges from guesstimates of over 200,000 species (Mann and Droop, 1996) to more conservative, morphology-based estimations of 12,000 species (Guiry, 2012) and metabarcoding-based estimations of 4,748 OTUs (Malviya et al., 2016). Molecular phylogenetic studies (e.g. Medlin and Kaczmarska, 2004; Theriot et al., 2010) group diatoms in three main categories: the ancestral and paraphyletic radial centrics, the likewise paraphyletic multipolar centrics and the most recent and monophyletic pennates. Radial centrics seem to consist of few remnant lineages, with *Leptocylindrus* constituting an important bloom former in coastal regions all over the world (Nanjappa et al., 2014). Multipolar centrics contain two highly diverse clades, the Thalassiosirales and the Chaetocerotales, whilst the pennates are the most diverse group (Not et al., 2012).
Diatoms have a diplontic life cycle (i.e. they spend the most of their life as diploid organisms and form haploid gametes through meiosis) consisting of a long period (up to several years) during which cells divide mitotically and a brief period (a few days) during which sexual reproduction takes place. Mitotic divisions are constrained by the siliceous cell wall (frustule). Indeed, as vegetative growth goes on, two sibling cells with different valve size are produced: one identical to the parent cell and the other one slightly smaller. MacDonald (1869) and Pfitzer (1869) described this size diminution process independently over a century ago. Although some taxa have been shown to possess both physiological and morphological modifications to overcome size diminution (e.g. von Stosch, 1965; Round, 1972; Drebes, 1977; Gallagher, 1983), in most species size restoration occurs only through sexual reproduction (Edlund and Stoermer, 1997). Lewis (1984) argued that size reduction cannot be a mere consequence of the cell division mechanism in presence of a siliceous frustule, but must have an adaptive significance. He suggested that size reduction might act as a chronometer for sex, allowing diatoms to spread the high costs of sexual reproduction over several or many years (Lewis, 1984; Mann, 1989; Mock and Medlin, 2012).

Centric diatoms reproduce sexually through oogamy (i.e. production of non-motile, large cells, the oogonia, and small, motile ones, the sperm cells), whilst pennates do so through isogamy (i.e. gametes of similar morphology differing in allele expression in one or more mating-type regions). In radial centrics the sperm cells do not include chloroplasts whereas in multipolar centrics (such as Bacteriastrum and Chaeroceros) the sperm carries a plastid, but this plastid generally does not contribute to the zygote. Instead in pennates, both gametes usually add a plastid to the zygote (Round et al., 1990). Centrics (including Bacteriastrum and Chaeroceros) are monoecious meaning that single strain can produce male and female gametes, and fertilise itself. This is a setback for crossing experiments and affects strain identity over time. Instead, pennates are dioecious, meaning that strains from the opposite mating type are needed to produce the next generation (Round et al., 1990).
Several diatom lineages can form resting stages in the form of spores or resting cells (McQuoid and Hobson, 1996). Resting cells are cells with condensed cytoplasm, less pigments in shrunken chloroplasts and thicker frustule than vegetative cells, but with the same shape of the vegetative cell (Lund, 1954). Instead, spores have a markedly different morphology from vegetative cells (i.e. a thick frustule, often ornamented with spines and other protuberances; Round et al., 1990).

There is increasing evidence that the evolutionary diversification of diatoms has taken place predominantly within sexual lineages. Indeed, there are no evidences of families or genera, even the most species rich, in which all the species are asexual or parthenogenetic (Mann, 1999). Natural diatom populations often consist of many fewer genotypes than individuals (except perhaps after mass auxosporulation), as a result of mitotic division and colony fragmentation (Richardson, 1995). In this scenario, it is possible that mutations occurring in a single individual are perpetuated quickly and indefinitely, eventually establishing a new species, as hypothesised by Goldschmidt (1940) and Small (1950). However, this mechanism is unlikely to work in sexual species (and most diatoms probably fall into this group) since a new species, arising through a macromutation in a single individual, would initially contain only one sex (Mann, 1999). However, in some species (e.g. *Chaetoceros*) individual strains can form male and female gametes (Round et al., 1990), making this scenario more likely to happen.

Divergence and speciation can apparently take place rapidly in diatoms, over periods of 1,000 - 10,000 years or less (e.g. Theriot, 1992). The availability of several diatom genomes has made it possible to estimate diversification rates at molecular level (Mock and Medlin, 2012). The bipolar centric diatom *Thalassiosira pseudonana* and the pennate *Phaeodactylum tricornutum*, known to have been diverging for about 90 million years, diverged of about 45% in their genomes (differences based on the percentage of amino acid identity of 4267 orthologous gene pairs, Bowler et al., 2008). In multicellular eukaryotes, a similar divergence (about 40%) is found between *Homo sapiens* and the
Pufferfish *Takifugu rubripes*, which have been diverging for the last 550 million years (Bowler et al. 2008). This comparison demonstrates that unicellular eukaryotes diverge faster than multicellular counterparts, which might be related to a higher mutation rate, larger effective population size, and shorter generation times (Mock and Medlin, 2012). In multicellular organisms with small population size, advantageous mutations are rare and disrupted by sexual reproduction (Bromham, 2011); furthermore, the life histories longer than unicellular eukaryotes further reduce the diversification rates (Mock and Medlin, 2012).

1.2. The species problem

Despite the fact that species are the fundamental units of biology, the dispute about how to define them is still ongoing. Mayr (1982) argued that most of the confusion about what constitutes a species is due to the application of the term “species” to two fundamentally different logical categories. The first of these includes the use of the word species as synonymous of “kind of”, to describe natural phenomena or things, like the words 'planet' or 'moon' (Mayr, 1996). In case of living-things, we refer to it as *species as taxon*, i.e. individuals that exists in space and time and have a historical continuity (Hull, 1976; 1978). The other meaning of species is *as taxonomic category*, to which taxa can belong. In this sense, the problem of species refer to the species as category and to the way (attributes) such categories are defined (Mayr, 1982). The pre-Linnaean concept of (biological) species was similar to the one used for non-living things: a species was defined by a set of unchangeable or slightly variable characteristics that allow us to recognise it from other such species. Therefore, for each species there was a model or “type” organism to which all the others must conform to be considered as members of the same species. However, after Darwin’s *Origin*, it became clear that biological species are not immutable entities: they constantly change, in space and time and at any detectable level. Consequently, a good concept of biological species must take into the account this
variability and consider not the difference but the degree of difference as a threshold for delimiting members of the same or different species (Mayr, 1982). But what are these differences (or characteristics), how to choose them and who does the choosing?

Over the centuries, philosophers, physicians, naturalists and many other (categories of) people have spotted and used different sets of properties useful to identify species and distinguish them from others. Wilkins (2011) stated that there currently are seven operational definitions of species (reviewed in de Queiroz et al., 2007) with 27 variations, three more in respect to the 24 counted in Mayden (1997). These seven definitions identify species according to different properties: for the purposes of this thesis, I will focus on two of them: the morphological (also called phenetic) species concept and the phylogenetic species concept. The morphological species concept is the one that all the people, familiar or not with the biological sciences, use in their daily life. It encompasses all the organisms (individuals) that share a similar morphology and assumes that a “type species”, identifiable through a “type specimen”, exists. Nowadays we do not use anymore the term “type species” but still use “type specimen” in taxonomy, although with a different meaning, to designate the specimen to which the name of a genus or subgenus is taxonomically associated. It expresses the way in which classical taxonomists work: they collect different specimens, if possible, from different sites or regions, look at similarities and dissimilarities in their various traits, and make hypotheses on their relatedness. Unless these phenotypic characteristics have a selective advantage (i.e. are determined by the environment), generally they are indicative of a common ancestry due to interbreeding among individuals of the same species (which in turn is the so called “biological species concept”).

On the contrary, the phylogenetic species concept is far less obvious and out of reach of not-professionals. In its simplest version, derived from the original by Eldredge and Cracraft (1980), it indicates “the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent” (Cracraft, 1983). It clearly
shows Darwin’s footprint and the core of its theory: the similarity by descent. However, less clear is how we detect such relationships of ancestry and descent. A few decades ago, this was achieved by means of phylogenetic trees. One school of tree builders applied cladistics using morphological characters and their states, considering only those states that were “derived from a common ancestor and shared across its descendants” (synapomorphies); the relationships among taxa were then determined looking at the clades in the phylogenetic tree. A competing school of tree builders applied distance-based or probabilistic methods to infer phylogenies, taking into the account all the available characters and their states. If variable characters are plentiful, both methods tend to gravitate onto similar tree topologies because phylogenetic signal is additive whereas invariable characters and noise do not add anything to that signal (Lemey et al., 2009).

Nowadays the characters and their states that we use are mostly at molecular level, in the form of nucleotide positions (characters) and their states (the four nucleobases: A, T, C and G) in case of DNA and amino acid positions and their 20 possible states (the 20 amino acids) in case of proteins. Therefore, the characters that we use to identify species are nucleotide / amino acid positions in coding and non-coding DNA regions / proteins. There is no competition between morphological (phenotypic) and molecular characters as there is no universal best marker; each one can be used for a particular aim.

It should be noted that phylogenies inferred from whatever information is at hand are mere incomplete hypotheses of the real, but unknown, evolutionary history. Not all the information at hand is equally useful for phylogeny reconstruction; as stated by Avise (2012), “because phylogeny is “the stream of heredity”, only genetically transmitted characters are informative to phylogenetic estimation”. Even the information on such characters is not, by definition, adding to the phylogenetic resolution because both morphological and molecular characters can be affected by convergence (independent changes in different lineages converging on a similar or identical outcome). Phylogenies
inferred using just a very few characters (no matter their nature) are also of restricted use because they may deviate markedly from the true evolutionary history (Mayr 1982).

1.2.1. Species concepts in diatoms

As for most of the taxa, the morphological species concept has dominated diatom taxonomy and systematics from its beginning (e.g. Van Heurck, 1896). Diatom species began to be described in the first half of the 1800s (Fig. 1.2) based on morphological characters observable in light microscopy (e.g. Agardh, 1830-32; Kützing, 1833; Ehrenberg, 1838; reviewed in Mann, 2010a). At that time, species were considered discrete and immutable entities, and so in those early descriptions there was almost no discussion of species concepts, nor intraspecific variation taken into the account (Mann, 1999). From 1859 onward, when the publication of Darwin’s *Origin* brought to attention the importance of varieties in the formation of new species, diatomologists started describing a huge number of species and varieties (e.g. Grunow, 1879; reviewed in Mann, 1999).
Fig. 1.2. Early example of diatom illustrations. Extracted from the “Synopsis Diatomearum, oder, Versuch einer systematischen Zusammenstellung der Diatomeen” by Kützing (1833).

Nowadays, the importance of intra-specific variation in diatoms is widely recognised and some authors have argued that no assertions should be made at the species level before considering it (Wood and Leatham, 1992). Intra-specific variation has been detected at different levels: i) at the individual (strain) level as clonal diversity (Rynearson and Armbrust, 2005; Ruggiero et al., 2018), heterozygosity (Rynearson and Armbrust, 2000), or phenotypic diversity (Gallagher et al., 1984; Gsell et al., 2012; Canesi and Rynearson, 2016); ii) at the population level as genetic differentiation among populations (Rynearson and Armbrust, 2004; Casteleyn et al., 2010) and phenotypic adaptation to different environments (Kremp et al., 2012). Empirical studies (reviewed in Godhe and Rynearson,
2017) have shown that intraspecific variation in diatoms is important in species’ responses to environmental factors such as light, temperature, salinity and nutrient availability.

Despite the fact that identification of diatom species by means of morphological traits is hampered by such intra-specific diversity as well as by phenotypic plasticity and cryptic speciation, it is the easiest, quickest and cheapest way to identify taxa. The vast majority of diatom species descriptions are based on morphological features such as overall cell shape as well as the shape, size and ultrastructural detail of the siliceous cell wall elements comprising the frustule (Evans et al., 2007; Alverson, 2008).

To overcome the difficulties associated with morphological data, several attempts have been made to apply the biological species concept (BSC, Mayr 1942) to define and delimit diatom species (e.g. Amato et al., 2007; Kaczmarska et al., 2009; Quijano-Scheggia et al., 2009; De Decker et al., 2018). However, carrying out crossing experiments in diatoms is not without risk; for many species, the details of the sexual cycle have not been described, and even for those for which the phase is known, the triggers to commence sex are often not. Only in a few species does the experimenter have control over the process. Therefore, it remains impossible to test the validity of most diatom species under this concept. Furthermore, taxa to be tested are generally chosen based on the assumption that differences in their morphology are indicative of reproductive isolation, which often is not the case (Mann, 2010a).

The true revolution in diatom taxonomy and systematics arrived with the introduction of the phylogenetic species concept (Eldredge and Cracraft, 1980) and the use of molecular tools. The former provided a framework based on homology to analyse informative characters; the latter a quick, objective and cheap way to gather taxa. Homology can be estimated for both morphological and molecular characters: in the first case by a detailed knowledge of the morphological structure in question and its development; in the latter, aligning the bases (states) observed at the same positions (characters) in the nucleotide sequences obtained from different strains or specimens. The analysis of DNA or RNA
marker sequences has become particularly attractive for species discovery and classification in diatoms because homology is ascertained easily. For each of the many nucleotide positions in the sequence, the sequence markers exhibiting the appropriate level of variation can be chosen depending on the questions at hand, and the bases (states) at homologous positions (characters) can be scored easily, unambiguously and cost-effectively (Alverson, 2008; Mann, 2010a).

Nowadays diatom species are described and identified using a combination of morphological and genetic characters and, when available, information about their ecology and distribution. This approach, called integrative taxonomy (Dayrat, 2005) has the advantage of combining different properties of species and so providing a more robust framework for their inference.

1.3. Do species really exist?

Despite all the available species concepts, some authors have even questioned the existence of species. This is not to say that species as taxa are unreal; they are, but, as category, they are as artificial as all the other taxonomic ranks above the species level (Mishler, 1999). Species are the outcome of different evolutionary strategies and environmental factors; this is why an animal species cannot be compared with a plant or fungal species, not even to mention a prokaryotic “species.” According to their motility, mating barriers, mutation rate, population size and many other factors, populations of a certain species are more or less dynamic and prone to changes over time and space. Generally, among botanists the attitude of denying biological species is prevalent (Bachmann, 1998; Mishler, 1999). Indeed, some plants form interspecific hybrids and, in some groups, phenotypic variation does not fit into discrete categories (Rieseberg and Willis, 2007). This was also the opinion of Darwin, who treated species as artificial constructs as genera, families and orders, asserting in the Origin: “I look at the term species as one arbitrarily given, for the sake of convenience, to a set of individuals closely
resembling each other, and that it does not essentially differ from the term variety, which is given to less distinct and more fluctuating forms” (Darwin, 1859). On the other hand, zoologists and especially the ones working on macrofauna, tend to recognise species because the reproductive barriers and morphological discontinuities are stronger or at least better defined. Among them, the most vehement defender of biological species was Ernst Mayr (1963; 1970; 1996; 1999), who happened to be an ornithologist.

Microbiologists have questioned the present species concepts and definitions used in microbiology (i.e. the morphological species concept for eukaryotic microorganisms and the DNA-based species definition for prokaryotic microorganisms) namely whether closely related isolates of bacteria or other microorganisms clustering into discrete groups have to be considered as different species (Spratt et al., 2006).

Within microbial eukaryotes, the sequencing of global samples of individuals of fungi and protists has shown that a vast diversity of genotypes exists and that this diversity is contained within relatively few morphologically recognised species that are globally distributed (Koufopanou et al., 2006; Spratt et al., 2006; Whitaker, 2006). In diatoms, it has been shown that these “phylogenetic species” that cannot be distinguished by morphology, are not simply the product of neutral genetic drift between geographically separate populations, because mating experiments have shown the presence of reproductive barriers (reviewed in Mann, 1999).

Two obvious differences underlying speciation between unicellular and multicellular organisms are that (i) population sizes tend to be much larger in the former and (ii) rates of homologous recombination can vary greatly, and lateral transfer can spread genes across large phylogenetic distances (Gogarten and Townsend, 2005; Spratt et al., 2006).

Since each species has its own evolutionary history, it is up to the specialist to ascertain if in its study system it is better to refer to species, or instead, to consider individuals, strains, populations, or meta-populations as the fundamental units of evolution. There is a common ancestry for all species, but not a common faith or definition.
My personal opinion on the matter embraces all the points discussed so far. I agree that every living organism is the product of unique and different historical, evolutionary and stochastic processes; therefore, in some cases it would be recommendable to refer to some taxa as species (e.g. when they form well distinct, homogeneous and recognisable reproductive units across time and space). In other cases (e.g. when reproductive barriers are labile), “species” are not arranged in discrete units and it would be better to consider lower taxonomic categories as the units of evolution below the species rank (e.g. metapopulations, populations, or even individuals or strains).

1.4. The need for classification

Classifications are arbitrary human constructs meant to group individual objects in categories based on a set of shared characters/properties. They are necessary when dealing with diversity, providing an effective tool for the storage and retrieval of information (Wheelis et al., 1992). However, the role of classifications is not limited to this. In his book “A System of Logic, Ratiocinative and Inductive” (1843), the English philosopher John Stuart Mill argued that a classification should serve to generate hypotheses. Similarly, Mayr (1969) supported this idea, but added that such hypotheses should have a strong likelihood of being true in order to produce reliable inferences.

The nature of a classification strictly depends on its intended function, and so there is no one "correct" classification (Wheelis et al., 1992). Biological classifications are just a kind, and their general meaning has changed profoundly over time. Early classifications of living things were utilitarian, attempting at explaining the plan of Creation (Agassiz, 1859), grouping organisms based on medical properties (Dioscoride, De Materia Medica) or physiological and reproductive traits (Aristotle), their “immutable essence” (Linnaeus) or simply analogies and differences (e.g. Cesalpino, De plantis, 1583) (Fig. 1.3). This has been their prevalent function until the formulation of the theory of common descent by Darwin and Wallace, when classification became phylogenetic (based on genealogy).
Despite some scientists as the French naturalist and mathematician Georges-Louis Leclerc, Comte de Buffon (1707-1788) and the English physician Erasmus Darwin (Charles Darwin’s grandfather, 1731-1802) had considered the hypothesis that similar species could have derived from the same ancestral species, Charles Darwin was the first one to state it unequivocally (Mayr, 1982). An interesting outcome is that, despite being based on different perspectives, phenetic classifications often reflect phylogenetic ones. This is because similarity among organisms is fundamentally the result of common ancestry, as Darwin had understood. However, as outlined by Darwin himself in the *Origin* (1859) some organisms can be markedly different in morphology despite common descent because of radical modifications they underwent during evolution. A typical case is the one of birds and crocodiles (Arcosauromorpha), taxa that share a common ancestor but are extraordinarily different in their aspect due the different evolutionary trajectories they have followed. Nowadays, it is widely accepted that biological classifications should be both practical and phylogenetic, putting together organisms that have the greatest amount of shared characters due to common descent (Mayr, 1942; 1982).
At this point, it is important to highlight the distinction between classification and identification. As outlined by Simpson (1961) and Mayr (1969), classification and identification are two different things. Classification only involves groups, is based on the analysis of several different characters and searches for shared (synapomorphic) character states. On the contrary, identification is an individual-based process, requires the analysis of a few characters, and prefers to work with species-defining (autapomorphic) character states. Even if at the end of the identification process individuals are assigned to a particular group, this process cannot be called “classification” and so identification schemes are not classifications (Mayr, 1982). Both classification and identification are the object of study of taxonomy, whilst the study of the relationships among taxa is the field of systematics (Simpson, 1961; Mayr, 1969). Taxonomy and systematics have both benefited from the introduction of molecular approaches. In particular, in the last decade there has been a “renaissance” of taxonomic research due to introduction of DNA-based identification approaches (DNA barcoding).

1.5. DNA barcoding

The concept of DNA barcoding (i.e. the identification of taxa using short DNA sequences) is linked to one godfather, Paul Hebert, and one marker, the cytochrome oxidase subunit 1 (COI). The idea of identifying species with molecular markers can be traced back to the advent of molecular biology techniques in the early 1980’s (Cristescu, 2014). Following the invention of PCR (Mullis and Faloona, 1987) and the development of universal primers (e.g. Kocher et al., 1989; Taberlet et al., 1991), Arnot et al. (1993) were the first to refer to “DNA barcodes” for species identification, amplifying the *Plasmodium falciparum* circumsporozoite (CS) gene to identify parasite stocks and lineages. However, the real revolution started when Hebert et al. (2003) proposed a system for the identification of animal taxa, called DNA barcoding, based on the use of a single gene marker, a 645 bp portion of the mitochondrial gene cytochrome c oxidase I (COI). A system based on DNA
barcodes provides both a way to identify taxa (e.g. the COI sequence for animals) and a way to delimit them from other such taxa (using a threshold of sequence divergence). In the same paper, Hebert et al. (2003) provided the example of 3% COI threshold dissimilarity value to delimit lepidopteran species and cite the > 2% cytochrome b threshold for vertebrates (Avise and Walker, 1999). The authors stressed multiple times that using a standard COI threshold for species delimitation, though appealing, should merely be considered as aid to the initial steps of the process. Unsurprisingly, the paper had its critics. Will et al. (2005) argued that “the real cutting-edge future for systematics and biodiversity research is integrative taxonomy, which uses a large number of characters, including DNA and many other types of data, to delimit, discover, and identify meaningful, natural species and taxa at all levels”. They have not even spared the use of DNA barcoding for the identification of taxa, stating that “by emphasizing a single gene as a “universal barcode” (Powers, 2004), DNA barcoders are returning to an ancient, typological, single-character-system approach” (Will et al., 2005).

Rubinoff et al. (2006), instead, clarified that the opposition to DNA barcoding must not be intended as an opposition to the use of molecular tools in systematics and taxonomy in general. They argued that if DNA barcoding is intended as identification of species previously defined by other means, definition of new species by interpretation of DNA diversity as indicative of species diversity and operational units for ecological studies, there is no opposition to it. In spite of that, barcoding is actually functioning in a very different way from the original purpose for which it had been intended (i.e. identify known species and reveal those that are undescribed). Indeed, one of the criticisms raised by Rubinoff et al. (2006) is that barcoding papers have focused their attention on case studies where “cryptic species” were already suspected based on other sources of data (e.g. morphological or ecological data), thus violating the initial aim of identifying the unknown biodiversity.
From the practical point of view, an important limitation of DNA barcoding is that it relies on the assumption that speciation is generally accompanied by divergence in the sequence of the target gene. However, since sequence divergence is a stochastic process, some closely related species could not be resolved by barcoding, even if the chosen region of DNA evolves rapidly (Mann et al., 2010). Furthermore, some species might be impossible to barcode using a single gene simply because they are paraphyletic (Meyer and Paulay, 2005).

Whatever the pitfalls or drawbacks of a barcoding approach based on DNA sequences, it is unquestionable the impact that the idea of Paul Hebert and colleagues had on the study of biodiversity. Since its publication in 2003, their article has been cited more than 9000 times and it has opened the way to a new field of research. Even if the original idea of a “universal barcode” for all kingdoms of life has been abandoned, DNA barcodes are nowadays available for a huge number of taxa from all over the tree of life. For plants, two chloroplast genes, the large subunit of the rubisco enzyme (rbcL) and the maturase K (matK) have been chosen (CBOL Plant Working Group, 2009), whilst for fungi the nuclear internal transcribed spacer region (ITS) (Schoch et al., 2012) and the V4 region of the gene coding for the small ribosomal subunit (V4-18S) for protists (Pawlowski et al., 2012) are the markers of choice.

1.6. From barcodes to metabarcodes

The recent technical advancements of massive parallel DNA sequencing technologies (e.g. next-generation sequencing platforms, NGS; Shendure and Ji, 2008; Glenn, 2011) have revolutionised many areas of scientific inquiry, taxonomy included. Providing millions of sequence-reads in a single experiment, NGS platforms have extended the classical, one-specimen-at-a-time Sanger sequencing identification of single specimens to the community level (Taberlet et al., 2012). This approach, called “metabarcoding”, is a multispecies
identification method using massive parallel sequencing of a particular marker in environmental DNA or RNA samples (Cristescu, 2014). The significant decrease in the costs of massive sequencing and the ease of sampling and analysing multiple instead of individual specimens has led to an increase of metabarcoding studies for aquatic, microbial and soil communities (Schmidt et al., 2013; Valentini et al., 2016; Abdelfattah et al., 2018), as well as to its application to biodiversity surveillance and monitoring (Bohmann et al., 2014; Deiner et al., 2017). However, being “blind”, metabarcoding approaches need a comprehensive taxonomic reference database, which is generated with the traditional barcoding approach on morphologically verified and curated specimens (Cristescu, 2014). Furthermore, its blindness is also extended to the unknown amount of species to identify in the community; this requires the primers used for the PCR to be highly versatile (amplify different target molecules with the same efficiency), in order not to miss species whose target sequences do not match well with the primers designed (Taberlet et al., 2012). Despite these and many other issue shared with the classical DNA-barcoding approach (use of a single target gene to identify taxa, PCR errors, etc.), DNA-metabarcoding has a potential that goes beyond biodiversity assessment and monitoring. It has proven to be an effective tool for diet assessment (Leray et al., 2013; De Barba et al., 2014; Kartzinel et al., 2015), species diversity and distribution (Nanjappa et al., 2014; Malviya et al., 2016; dos Santos et al., 2017; Tragin and Vaulot, 2019) and product authentication (Mishra et al., 2016; Raclariu et al., 2017; 2018). All the aforementioned studies show that we are still at the early stages of exploitation of DNA-metabarcoding potential, and it will be a powerful technique for many years to come.

1.6.1. DNA barcoding and metabarcoding in diatoms

The application of DNA barcoding to diatoms is no different, in principle, from that in other organisms i.e. to provide unambiguous identification of a specimen, using a short sequence of coding or noncoding DNA (Mann et al., 2010). Some characteristics found in
diatoms as cryptic speciation, different morphology across life cycle and culture conditions (Mann, 1999) make barcoding particularly advantageous in these organisms over classical morphological examinations (Mann et al., 2010).

To date, no universal barcode region for diatoms exists, but several markers have been considered and proposed within the nuclear, mitochondrial and chloroplast genomes (Moniz and Kaczmarska, 2009, Fig. 1.4).

![Image: Main target genes utilised for DNA barcoding in diatoms. Orange = mitochondrion; green = chloroplast; blue = nucleus.](image)

The classical barcode genes used for animals (COI) and plants (**matK**, **rbcL**) seem not to work well for diatoms and other protists. For COI, the main problem is lack of sufficiently conserved primer target regions across taxa (Evans et al., 2007; Moniz and Kaczmarska, 2009) and occurrence of introns (Ehara et al., 2000; Armbrust et al., 2004; Ravin et al., 2010). Plastid markers have been considered problematic for DNA barcoding due to both uniparental or biparental inheritance (Round et al., 1990; Jensen et al., 2003; Leviaaldi Ghiron et al., 2008). Nonetheless, the **rbcL** has been evaluated both in its entire length (~1400 bp) and as fragment at 3’-end (**rbcL**-3P, ~750 bp, Hamsher et al., 2011; ~540 bp, MacGillivary and Kaczmarska, 2011). Preliminary results suggested that the 3’-region is
more variable than the 5’-one and so discouraged the use of the whole gene (Hamsher et al., 2011). In spite of the fact that ease of amplification, sequencing, and alignment as well as lack of indels and introns make it a promising marker (MacGillivray and Kaczmarska, 2011), the low resolution at discerning closely related species in some groups and the aforementioned uncertain inheritance led to the conclusion of a better use of \textit{rbcL}-3P region as complementary barcoding gene together with 5.8S-ITS2 rDNA region in a dual-locus DNA barcoding system (MacGillivray and Kaczmarska, 2011). This latter region was proposed by Moniz and Kaczmarska (2009, 2010) as candidate barcode based on its use at identifying protist, fungal and plant species (Wayne Litaker et al., 2007; Seifert, 2009; Chen et al., 2010). However, the ITS region is known to be difficult to align even in closely related species (Desdevises et al., 2000; Poisot et al., 2011) and to show infraspecific polymorphism due to non-concerted evolution (Harpke et al., 2006; Zheng et al., 2008), all factors that limit its applications in heterogeneous taxa.

Among nuclear DNA markers, and still within the rDNA cistron, most of the attention has been focused on the genes coding for the nuclear small and large subunit (SSU and LSU) RNAs of the ribosomes, (a.k.a. 18S and 28S rDNA, respectively). Due to its overall length, generally around 3,000 bp, barcoding has focused on the D1-D3 (~ 800 bp) and D2-D3 (~ 613 bp) regions in the LSU (Hamsher et al., 2011). These fragments are considered as variable as the \textit{rbcL}-3P (Hamsher et al., 2011), and therefore, expected to resolve species- and sometimes population-level relationships (Alverson, 2008). However, these markers are unsuitable for current NGS platforms used in metabarcoding approaches because they are too long. Another drawback is that LSU reference sequences are available only for selected groups of organisms; not yet across the entire eukaryotic tree of life, not even across the diatom diversity. On the contrary, the SSU region has been used extensively in diatom phylogenies (Medlin et al., 1993; Kooistra and Medlin, 1996; Medlin et al., 1996; Medlin and Kaczmarska, 2004; Sarno et al., 2005; Sorhannus, 2007) and the huge number of reference sequences stored in public databases (e.g. PR², Guillou et al., 2012) essentially
covers the diversity of the diatoms. The validity of the various variable regions as barcoding target has been evaluated, in particular the V4 and V9 (Nelles et al., 1984). Recent results showed that the V4 region (~ 380-400 bp) can be considered the most promising candidate marker for DNA barcoding in diatoms given its ease of amplification, extensive reference library and variability, and universality of its primer target (Zimmermann et al., 2011; Luddington et al., 2012). It outperforms the V9 region in separating closely related species because of its greater length (~ 380 bp vs. 105 bp) and the fact that the V9 region is located at the very 3’-end of 18S gene, a region that is often sequenced incompletely or poorly (Gaonkar, 2017; Gaonkar et al., 2018). However, currently several V4 (BioMarKs, Massana et al., 2015; the Ocean Sampling Day, Kopf et al., 2015) and V9 (e.g. Tara Oceans, de Vargas et al., 2015) metabarcoding datasets are available to explore diversity and distribution of organisms (diatoms included) in world’s oceans and to test the effectiveness of both regions in discriminating specific taxa. In this thesis, I will use the two global metabarcoding datasets, OSD (V4) and Tara Oceans (V9) to explore the diversity of *Chaetoceros* in the world’s oceans.

1.7. Case study: the planktonic diatom family *Chaetocerotaceae*, with emphasis on the genus *Chaetoceros*

Diatoms (from the Greek word *diatomos*, “cut in half”) are unicellular eukaryotes whose hallmark is the ornamented silica cell wall called frustule (Round et al. 1990). In the Tree of Life, diatoms are found in the superphylum Heterokonta (Stramenopiles, Adl et al., 2005), which includes unicellular eukaryotes that produce, at some point in their lifecycles, cells with two unequal flagella (Cavalier-Smith, 1986). Diatoms are one of the largest and ecologically most significant groups of organisms on Earth. They occur almost everywhere they can found adequately amount of light and water for photosynthesis: oceans, lakes, rivers, marshes, rock faces, and even on the feathers of some diving birds (Mann, 2010b).
Because of their abundance in marine plankton, diatoms are estimated to account for as much as 20% of global carbon fixation (Field et al., 1998).

From the ecological point of view, diatoms are generally divided in planktonic (suspended in open waters) and benthic (living on the floor of water basins). Planktonic diatoms dominate the phytoplankton of cold, nutrient-rich waters, such as upwelling areas of the oceans and recently circulated lake waters (Graham et al., 2016). Together with benthic ones, after death they are responsible for carbon sinking and accumulation of silica in sediments, contributing to the flux of nutrients (Smetacek, 1985; Willén, 1991).

The focus of my Ph.D. thesis is the planktonic diatom family Chaetocerotaceae Ralfs in Pritchard, with particular emphasis on the genus Chaetoceros. The genus Chaetoceros Ehrenberg, 1844 is common in the plankton worldwide and, together with the genus Bacteriastrum Shadbolt, 1854 constitutes the family Chaetocerotaceae Ralfs in Pritchard. The hallmark of the family is the presence of siliceous hollow spine-like extensions (setae), which protrude from the valve face or margin of the cell. Chaetocerotaceae belong to the bipolar centric diatoms, i.e., a clade or grade of diatoms with valves exhibiting a bi- or multipolar architecture, a circular pattern centre, a centrally located labiate process and apically located fields of poroids that are ultrastructurally distinct from the poroids in the remainder of the cell wall (frustule) elements. The setae are believed to have evolved from those apical pore fields. Among the differences between the two genera are the following: i) valvar symmetry, which is multipolar in Bacteriastrum (Fig. 1.5A) and bipolar in Chaetoceros (Fig. 1.5C); ii) seta number per valve, generally two in Chaetoceros (Fig. 1.5D) and more than two in Bacteriastrum (Fig. 1.5B); iii) valve outline, oval in the former (Fig. 1.5D) and circular in the latter (Fig. 1.5B); and iv) the number of species, hundreds in Chaetoceros, a few dozens in Bacteriastrum.
Gran (1897) divided the genus *Chaetoceros* in two subgenera, *Phaeoceros* Gran and *Hyalochaete*; the first includes species with multiple chloroplasts in the central body of the cell and in the setae, the second comprises species without plastids in the setae (Kooistra et al., 2010). Hendey (1964) changed the name of the subgenus *Phaeoceros* in *Chaetoceros* since the subgenus that includes the type species of the genus (*Chaetoceros dichaeta* Ehrenberg) has to keep the epithet of the genus, according to the rules of the botanical nomenclature. More recently, Hernández-Becerril (1993) created a third subgenus, *Bacteriastroidea* Hernández-Becerril to include a single species, *C. bacteriastroides*, exhibiting two different types of setae per valve.

Fig. 1.5. Main morphological features of *Bacteriatrum* and *Chaetoceros*. (A) Girdle view of *B. furcatum* sp. 2 strain Na8A3 in LM; (B) Valval view of the same strain in SEM; (C) Girdle view of *C. debilis* sp. 3 strain Ch13A4 in LM; (D) Valval view of the same strain in SEM. Figures are from Gaonkar et al. (2018).
Both genera are homothallic, i.e., micro and macrogametes (analogous to male and female gametes) are formed in one and the same clonal culture, but in different cells. Following gamete fusion, the resulting zygote develops through partial inflation into a specialised cell, the auxospore, which re-establishes the initial vegetative cell size. Furthermore, vegetative cells in many of the species can develop into resting spores anytime during the vegetative part of their life cycle (see Round et al. 1990). Resting spores are highly silicified, and often heavily armoured cells that go senescent and can survive under conditions adverse to growth. The spores sink to the sea floor and germinate whenever favourable conditions are restored. Simultaneous germination of massive numbers of spores can trigger sudden seasonal diatom blooms (McQuoid and Hobson, 1996). Alternatively, the spores can be sequestered in the sediment, where they provide a stratigraphic record (Suto, 2006). In the end, they can get fossilised, thus constituting an important carbon sink (Smetacek, 1985).

1.7.1. Fossil record of Chaetoceros

Vegetative cells of Chaetoceros leave no fossil record because these are weakly silicified and in most cases dissolve after the cell’s death (Ishii et al., 2011). Instead, the heavily silicified resting spores are often preserved in near-shore sediments as fossils, frequently in association with other diatom fossils, providing useful information for reconstructing paleo-productivity and paleo-environmental changes (Akiba, 1986; Itakura, 2000). For these reasons, Chaetoceros fossils have been described as “spore genera” and they may represent extinct taxa (Ishii et al., 2011 and references therein). A large number of spore genera has been described, such as Dicladia Ehrenberg (1854), Xanthiopyxis Ehrenberg (1854), Syndendrium Ehrenberg (1854), Liradiscus Greville (1865) and Monocladia Suto (2003), all of which may be assignable to the genus Chaetoceros (Suto, 2005). These fossils are from the Paleogene (65-23 mya), in particular from the Eocene/Oligocene
boundary (~34 mya), the Oligocene/Miocene boundary (~23 mya) and the early/middle Miocene boundary (~15.9 mya, Suto et al., 2006).

However, age estimates of Chaetoceros from diatom phylogenies calibrated with molecular clocks and diatom fossils are far older than direct fossil evidence. A phylogeny of diatoms inferred using the small subunit of rDNA gene (18S) and calibrated with fossil records dated back the origin of Chaetoceros in the Cretaceous (around 120 mya, Sorhannus 2007). In another study, conducted using four molecular markers (SSU, LSU, rbcL and psbA) and performing molecular clock analysis the split between Chaetoceros and Cymatosira was found in the Jurassic, around 180 mya (Medlin, 2015).

1.7.2. The ecological and evolutionary importance of Chaetoceros

Chaetoceros possesses some characteristics that makes it the prime target for ecological and evolutionary studies in marine phytoplankton. Indeed, it; i) is one of the most species-rich genera among diatoms (Rines and Hargraves, 1988; Hasle and Syvertsen, 1996; Hernández-Becerril, 1996), with about 500 taxa attributable to species or “variants”, and few more than 200 flagged as taxonomically accepted species (Guiry and Guiry, 2017); ii) is globally distributed, especially in upwelling regions (VanLandingham, 1968; Hasle and Syvertsen, 1996); iii) it accounts for 20–25% of the total marine primary production (Werner, 1977), especially in near-shore upwelling regions and coastal areas (Rines and Hargraves, 1988; Rines and Theriot, 2003). Furthermore, some species can be harmful during blooms, getting stuck in fish gills with their setae and causing mass mortality through limited oxygen uptake (Albright et al., 1993).

The success of this genus in terms of number of species, abundance, and global distribution is likely due to the combination of particular aspects of the life cycle (e.g. resting spore formation) and evolutionary novelty (the setae).

Many species of Chaetoceros form resting spores (Blasco, 1970; von Stosch et al., 1973; Hargraves and French, 1975), a strategy that allows them to escape situations in which
nutrient supplies are scarce, sinking to the sea floor and germinating when favourable conditions are restored. This characteristic is considered to be an evolutionary primitive trait (Simonsen, 1979) and typical of current neritic species (Ross and Sims, 1974). The putative adaptive advantages in possessing setae have not been cleared; they might deter grazers, have a role in buoyancy or nutrient and CO$_2$ uptake (Smayda and Boleyn, 1966; Smetacek, 1985; Verity and Smetacek, 1996).

*Chaetoceros* is easy to identify at the generic level because of the setae, but it is difficult to identify at the species level since the morphological criteria used (e.g. colony formation, cell size and shape, intercellular spaces, number of chloroplasts, morphology and orientation of setae, etc.) are quite variable (Hargraves, 1979) and in many smaller species difficult to observe in LM. Factors such as the presence/absence of grazers, salinity changes, nutrient availability or prolonged culture conditions can alter the morphology of the species, thus creating uncertainties in the species identification.

In spite of that, integration of phylogenetic and morphological information on isolated strains has contributed to the characterisation of *Chaetoceros* species and to the discovery of cryptic and pseudo-cryptic species (Kooistra et al., 2010; Degerlund et al., 2012; Huseby et al., 2012; Chammansinp et al., 2013; Li et al., 2017; Balzano et al., 2016; Gaonkar et al., 2017; 2018). More than 80 *Chaetoceros* strains and a dozen of *Bacteriastrum* have been characterised so far by morphological (light, scanning and transmission electron microscopy) and genetic (D1-D4 region of 28S rDNA) means (Gaonkar et al., 2018), thus providing a reference library of strains occurring in the Gulf of Naples and/or in other localities.

The PhD thesis of Gaonkar (2017) focused on: i) the molecular phylogeny of Chaetocerotaceae using 18S and partial 28S rDNA; ii) the diversity of Chaetocerotaceae in the Gulf of Naples (GoN) using a V4-18S metabarcoding approach; and iii) the analysis of the *C. socialis* species complex, with the description of two new species. The goals of i)
were: to understand how thoroughly the species diversity of the genera has been explored and what the relationships are between these genera; how many species are to be discovered yet; how common is cryptic diversity; and if morphological species delimitation has a genetic support. The V4-18S metabarcoding approach (point ii)) aimed at assessing: how many species occur in the GoN; how many species are found in the High-Throughput Sequencing (HTS) data but are still to be morphologically identified; how many of them occurring in the HTS data are known from elsewhere but have never been recorded in the cell counts at the GoN. Among the main results relevant to my thesis are the following: i) the 18S and 28S phylogenies do not resolve the position of *Bacteriastrum* with respect to *Chaetoceros*, and only terminal clades obtain significant support; ii) potential cryptic species exist within several morphologically defined species (e.g. *C. affinis*, *C. curvisetus*, *C. lorenzianus*, *C. socialis*); iii) the 18S-V4 region is generally better than V9 at discriminating terminal clades, with the same resolution of whole 18S gene, revealing to be a candidate target for metabarcoding studies.

### 1.7.3. Aim of Ph.D. thesis

Starting from the points discussed above, my PhD thesis has the following aims:

1. **To produce a multi-gene phylogeny of the family Chaetocerotaceae integrating the pre-existing information of nuclear data with chloroplast and mitochondrial ones in order to assess if adding phylogenetic information helps towards resolving the phylogenetic history of the family (Chapter II);**

2. **To provides an assessment of the diversity and distribution of the genus *Chaetoceros* by integrating classical and novel primary biodiversity data (global metabarcoding dataset) (Chapter III);**

3. **To analyse the *C. curvisetus* species complex using the potential of spatial data contained in global metabarcoding datasets in the form of phylogenetic networks (Chapter IV);**
4. To test the hypothesis of concerted evolution in *Chaetoceros* with an appropriate experimental design (single strain HTS and targeted analyses), starting from the data contained in a temporal metabarcoding dataset (MareChiara) (Chapter V).

**References**


Balzano, S., Percopo, I., Siano, R., Gourvil, P., Chanoine, M., Marie, D., ... Sarno, D. (2016). Morphological and genetic diversity of Beaufort Sea diatoms with high


Hernández-Becerril, D. U. (1993). Note on the morphology of two planktonic diatoms: *Chaetoceros bacteriastroides* and *C. seychellarus*, with comments on their...


Van Heurck, H. F. (1896). *A Treatise on the Diatomaceae, containing introductory remarks on the structure, life history, collection, cultivation and preparation of diatoms, and a description and figure typical of every known genus, as well as a description and figure of every species found in the North Sea and countries bordering it, Including Great Britain, Belgium, & c.* Wheldon & Wesley, London.


Chapter II

*Inferring the evolutionary history of Chaetocerotaceae*
The material presented in this chapter has been published as Research Article:


As the author of this Elsevier article, I retain the right to include it in a thesis or dissertation (figures and tables included) according to the rules on copyright of Elsevier available at https://www.elsevier.com/about/our-business/policies/copyright#Author-rights.
2.1. Introduction

2.1.1. Systematics of Chaetocerotaceae

The planktonic diatom family Chaetocerotaceae Ralfs in Pritchard (1861) is one of the largest and most diverse marine diatom families (Cupp, 1943; Hernández-Becerril, 1996; Jensen and Moestrup, 1998; Rines and Hargraves, 1998; Shevchenko et al., 2006; Bosak and Sarno, 2017). It plays an ecologically important role, representing an important primary producer in coastal and offshore marine environments worldwide (Continuous Plankton Recorder Survey Team, 2004; Leblanc et al., 2012; Malviya et al., 2016). The family includes the extant genera *Bacteriastrum* Shadbolt and *Chaetoceros* Ehrenberg, which differ in the number of setae per valve. The former generally possesses many, regularly arranged along the valve margin whilst the latter exhibits usually just two, one at each end of the apical axis (Hasle and Syvertsen, 1996). Despite the ecological importance, little is known about the systematics of Chaetocerotaceae. *Bacteriastrum* is exclusively marine, with 11 taxonomically accepted species (Guiry and Guiry, 2018). Its cells are cylindrical in valve view and contain numerous plastids; intercalary setae usually fuse over a large part of their length and then bifurcate (i.e., appear to branch) whereas the terminal setae do not branch (Hasle and Syvertsen, 1996). Pavillard (1925) erected two sections, *Isomorpha* and *Sagittata*, based on the orientation of the terminal setae on the opposite terminal valves of a colony: in *Isomorpha* they are each other’s mirror image whereas in *Sagittata* their orientation differs (Fig. 2.1). Within *Bacteriastrum*, *B. hyalinum* Lauder is the only species known to form resting spores (Drebes, 1972).
Fig. 2.1. Different orientation of terminal setae on the terminal valves of a colony of *Bacteriastrum* sections *Isomorpha* (A) and *Sagittata* (B). (A) *B. hyalinum* and (B) *B. elegans*.

*Chaetoceros*, with currently well over 200 taxonomically accepted species, is arguably the most diverse genus of planktonic diatoms in the marine realm (Guiry and Guiry, 2018). Most of current knowledge about its systematics dates back to the 19th century when, after the description of the material from the Antarctic expedition of Captain Rofs (1841-1843) by Ehrenberg (1844), several efforts have been made to fit this huge diversity into different taxonomic categories. The first attempt was made by Gran (1897), who divided *Chaetoceros* in two subgenera, *Phaeoceros* (now *Chaetoceros*) and *Hyalochaete*, basing on the distribution of chloroplasts. *Chaetoceros* has numerous small chloroplasts throughout the body of the cell and the setae, which are thick, very long, and armed with conspicuous spines (Hasle and Syvertsen, 1996). On the contrary, members of *Hyalochaete* have usually one or few chloroplasts only within the cell body, and setae are usually thin and more fragile (Fig. 2.2).
Fig. 2.2. Chloroplasts disposition in the subgenera Chaetoceros (A) and Hyalochaete (B) of Chaetoceros. (A) C. peruvianus 2 and (B) C. decipiens.

In addition, species belonging to the subgenus Chaetoceros exhibit rimoportulae (labiate processes) in both intercalary and terminal valves whereas in Hyalochaete these processes are observed only in terminal valves (Hasle and Syvertsen, 1996). Chaetoceros contains mostly oceanic species in which resting spores are lacking or unknown, except for C. eibenii (Grunow) Meunier (Jensen and Moestrup, 1998). After Gran, the two subgenera were further divided in sections by Ostenfeld (1903), Gran (1905) and, in recent times, by Hernández-Becerril (1991; 1993a; 1996), reaching the current number of 22 (Rines and Theriot, 2003). Furthermore, a third subgenus, Bacteriastroidea, was created for the only species C. bacteriastroides (Hernández-Becerril, 1993b).

Each of these infrageneric taxa are based on one or a few distinctive morphological features rather than on a formal cladistic analysis of all available characters and their states. Rines and Hargraves (1988) and Rines and Theriot (2003) pointed out some of these features are plastic, and so not reliable for a phylogenetic investigation. Cladograms inferred by Rines and Theriot (2003) from morphological information resolved Bacteriastrum inside paraphyletic Hyalochaete, which was resolved in its turn in paraphyletic subgenus Chaetoceros. Kooistra et al. (2010) reported similar topologies between phylogenies inferred from partial 28S rDNA sequences and from morphological information from the same strains. They resolved Bacteriastrum and monophyletic
subgenus *Chaetoceros* inside paraphyletic *Hyalochaete*. Yet, their study included fewer species than that of Rines and Theriot (2003).

Recent studies in Chaetocerotaceae have provided detailed morphological and ultrastructural illustrations as well as sequence data of numerous taxa, many of which are new to science (Kooistra et al., 2010; Li et al., 2013; 2017; Bosak et al., 2015; Gaonkar et al., 2017; 2018; Xu et al., 2019). However, most of these studies generally focused on the diversity within sections, and therefore the phylogenetic status of the investigated taxa remains to be resolved. Many studies used only the partial 28S rDNA as molecular marker, which poorly resolves the basal ramifications and therefore does not clarify relationships among the sections.

In this chapter, I infer a phylogeny of the family Chaetocerotaceae from a concatenated alignment of two nuclear (18S and 28S), two plastid (*rbcL* and *psbA*) and one mitochondrial (COI) gene gathered from 100 strains. Furthermore, I use the obtained tree to assess if the genera and the various infrageneric taxa are monophyletic as well as the validity of traditional classification scheme. This tree will also serve as a template to map characters and their states in future researches in order to reconstruct their evolutionary history. In this way, new insights will be gained on the evolution and diversification of one of the most species-rich and abundant marine planktonic diatom families.

### 2.2. Materials and methods

#### 2.2.1. Taxon sampling, outgroups selection and DNA extraction

For this investigation, I used a total of 100 diatom strains (Table A2.1, Fig. A2.1, Appendix II), from all over the diversity of *Chaetoceros* (Rines and Hargraves, 1988; Guiry and Guiry, 2018) and *Bacteriastrum* (Van Ladingham, 1968; Sarno et al., 1997; Godrijan et al., 2012; Guiry and Guiry, 2018). Most of the strains have been previously isolated from various localities (Table A2.1, Appendix II) and grown as monoclonal cultures in 74 ml polystyrene cell culture flasks (Corning Inc., NY, USA) filled with 30 ml
of f/2 medium at the following conditions: salinity of 36 %, 15 °C, 12:12 h light:dark cycle and a photon flux density of 50 μmol m$^{-2}$ s$^{-1}$ provided by cool white (40 W) fluorescent tubes (Gaonkar, 2017; Gaonkar et al., 2018). For the choice of outgroup sequences, I used the phylogenetic tree of diatoms by Theriot et al. (2015). I have chosen a nested set of taxa within the bipolar centric diatoms close to Chaetocerotaceae and for which there were GenBank sequences available for most, if not all, of the gene regions used in the present study, and from the same strain (Table A2.1, Appendix II). DNA was here extracted only for specimens not available in Gaonkar et al. (2018) and following the same protocol. DNA quantification was done by Nanodrop spectrophotometry.

2.2.2. Selection of genes, amplification and sequencing

To reconstruct the evolutionary history of Chaetocerotaceae, I used the information of five genes: two nuclear, encoding the small rDNA subunit (18S) and the D1 and D3 hypervariable domains of the large rDNA subunit (28S); two plastid, the rubisco large-subunit and the D1 protein-coding gene of the photosystem II (rbcL and psbA respectively); and a portion of the subunit I of cytochrome c oxidase gene (COI, mitochondrial). The sequences of 18S and 28S were mostly obtained from Gaonkar et al. (2018), except for the new strains here extracted and amplified (Table A2.1). All loci except 28S and COI were amplified for virtually their entire length using the primers listed in Table 2.1.

**Table 2.1. List of the primers used for phylogenetic inference.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence (5’-3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>18S</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSU-F</td>
<td>TCYAAGGAAGGCAGCAGGC</td>
<td>Hamsher et al. (2011)</td>
</tr>
<tr>
<td>SSU-R</td>
<td>GTTCAGCCTTGCGACCATACTCC</td>
<td>Ki et al. (2007)</td>
</tr>
<tr>
<td>28S</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D1R | ACCCGCTGAATTTAAGCATA | Scholin et al. (1994)
D3Ca | ACGAACGATTTCGACGTCA | Scholin et al. (1994)

**rbcL**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| F | GTGACCGTTACGAATCTGGTG | Fox and Sorhannus (2003)
| R | CTGTTCAGCGAAATCAGC | Fox and Sorhannus (2003)

**psbA**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| psbA-F | AGTACCACATAATGGTGTCGCC | Yoon et al. (2002)
| psbA-R1 | ACTTCATCAGCGATTTCGAC | Yoon et al. (2002)

**COI**

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| GazF2 | CAACCAYAAAGATATWGGTAC | Evans et al. (2007)
| KEdtmR | CAAATAAAAATRATWGCWCCTAA | Evans et al. (2007)

PCR amplification protocols were adjusted according to the success or yield of amplification in different species. Regardless of the protocol, each reaction was conducted in a final volume of 20 µL consisting of: 5X Phusion HF Buffer, 0.2 mM dNTPs, 0.5 µM forward and reverse primers, 1 U Phusion® DNA Polymerase, approximately 50 ng of DNA and water to volume.

Nuclear genes were amplified at the conditions specified by Gaonkar et al. (2018). For *rbcL* and *psbA* genes, a first protocol including initial denaturation at 98 °C for 3 min and 34 cycles each with denaturation at 98 °C for 30 s, annealing at 62 to 45 °C (lowering the T of 0.5 °C/cycle) for 25 s, and extension at 72 °C for 1 min and 30 s was performed. In case of lack of amplification or poor yield, for *rbcL* the annealing temperature was lowered to 55-51.6 °C in steps of -0.1 °C/cycle, whilst for *psbA* to 52 °C. For the amplification of COI marker, the following protocol was applied: initial denaturation at 98 °C for 5 min, and 30 cycles each with a denaturation step at 98 °C for 1 min, annealing at 50 °C for 1 min, and extension at 72 °C for 45 s. The annealing temperature was lowered to 45 °C in samples providing poor yield. The amplification of 18S and 28S was carried out as specified in Gaonkar et al. (2017).
The success of PCRs was checked by electrophoresis in 1.5% agarose gel and 0.5 X TBE (Tris-Borate-EDTA). PCR products were purified either from agarose gel with the DNA Isolation Spin-Kit Agarose (AppliChem, Darmstadt, Germany) or directly from PCR tubes using the QIAquick® PCR Purification Kit (Qiagen, Hilden, Germany), whether multiple or single bands were observed following electrophoresis, respectively.

Purified DNA was sequenced using the BigDye Terminator v3.1 sequencing kit on a 48 capillaries-3730 DNA Analyzer (Life Technologies, ThermoFisher Scientific) at the Molecular Biology facility available at the SZN. PCR products were sequenced using both forward and reverse primers used for amplification. For 18S, two additional internal primers were used (Ch-528F and Ch-1055R, Gaonkar et al., 2018), whilst only one for \( rbcL \), primer located at about 500 bp downstream the forward primer (\( rbcLinF, 5'\)-GTCGTGTAGTTTTCGAAG-3’, present study).

2.2.3. Sequence editing and alignment

The electropherograms generated by Sanger sequencing were manually checked using Seq Scanner v2.0 (Applied Biosystems, ThermoFisher Scientific) and then, for 18S, 28S, \( rbcL \) and \( psbA \), the resulting reads were assembled in contigs using ChromasPro v2.1.4 (Technelysium, Pty, Ltd) to generate the amplified fragment. For 18S and partial 28S data not generated in the present study, I used the sequences provided in Gaonkar et al. (2018) with introns removed. Sequences were aligned using ClustalX2 (Larkin et al., 2007) setting the parameters of pairwise and multiple alignment as specified in Hall (2004). Data were visualised and graphically edited in the R (R Core Team, 2018) working package apex (Jombart et al., 2017). Each gene matrix was then concatenated using Mesquite v3.51 (Maddison and Maddison, 2018) and visually checked.
2.2.4. Nucleotide composition and substitution saturation analyses

Base composition and substitution saturation are among the main factors known to affect phylogenetic reconstructions (Foster and Hickey, 1999; Moreira and Philippe, 2000; Theriot et al., 2015). Model-based phylogenetic methods usually assume that the aligned nucleotides evolve under homogeneous conditions (e.g. Jayaswal et al., 2005), but the risk of phylogenetic errors increases if these conditions are violated (Ho and Jermiin, 2004; Jermiin et al., 2004). In order to detect putative base compositional heterogeneity in the dataset, I performed a χ² test of homogeneity of state frequencies across taxa on each gene partition (18S, 28S, rbcL, psbA, and COI) using PAUP* v4.0a (build 159) (Swofford, 2002). I also checked if substitution saturation was occurring at 3rd codon position of protein-coding alignments (rbcL, psbA, and COI) using the software DAMBE v6.4.107 (Xia, 2017). I calculated the proportion of invariant sites (P_{inv}) for the 3rd codon position of each gene using the NJ algorithm, and I used the obtained value to implement the saturation test by Xia et al. (2003). This test calculates the index of substitution saturation (I_{ss}) by sampling different subsets of sequences, and compares it to critical I_{ss} value (I_{ss.c}) at which the sequences will begin to fail to recover the true tree (Xia et al., 2003). Sequences are considered to have experienced little saturation when I_{ss} is significantly smaller than I_{ss.c} (Xia et al., 2009).

2.2.5. Model selection and phylogenetic inference

I calculated the best-fitting model of nucleotide sequence evolution for each gene using the corrected Akaike information criterion (AICc) in PartitionFinder v.2.1.1 (Lanfear et al., 2016). The GTR+G+I model was favoured over the other models for all the genes considered. To ascertain if the evolutionary histories inferred from different cellular compartments were congruent, I inferred Maximum Likelihood (ML) trees using RAxML (Stamatakis, 2014) on the concatenated nuclear (18S and 28S) and plastid (rbcL and psbA) datasets and on the mitochondrial, single gene COI matrix. For ML inference, I conducted
100 ML tree searches under the GTR+G+I model of nucleotide substitution and then I calculated bootstrap support values by means of 1000 bootstrap replicates. The resulting nuclear, plastid and mitochondrial trees were checked for possible conflicts in their topology. Subsequently, I concatenated the five genes and inferred multigene phylogenetic trees using Maximum Parsimony (MP), ML and Bayesian Inference (BI). MP inference was conducted in PAUP* v4.0a (build 159) (Swofford, 2002). Heuristic tree searches comprised 10 random-addition replicates, TBR branch swapping, ACCTRAN character-state optimization, and gaps coded as missing data. Branch support was calculated by bootstrap analysis using 1000 bootstrap replicates. ML analysis was performed with IQ-TREE v1.6.8 (Nguyen et al., 2014) using the partition scheme suggested by PartitionFinder (GTR+G+I for each gene, -spp option), empirical base frequencies (+F option) and 1000 bootstrap replicates (-b option). A Bayesian tree was inferred using MrBayes v3.2.6 (Ronquist et al., 2012) using the GTR+G+I model (lset nst=6, rates=invgamma). All nucleotide substitution model parameters were unlinked across partitions and the different partitions were allowed to evolve at different rates (prset ratepr = variable). I ran four concurrent chains (one cold and three heated) for 10,000,000 generations and recorded samples every 1000 generations. Convergence and effective sample sizes (ESS) for all parameters were analysed in Tracer v.1.6 (Rambaut et al., 2014), the latter considered valid above the threshold of 200. Based on the results of Tracer analysis, I discarded the first 25% of the samples as burn-in.

2.2.6. Morphological sections and species assignment

In order to assign the strains here utilised to existing sections, I retrieved the descriptions of the morphological characteristics defining sections from the literature. For Bacteriastrum, I referred to Pavillard (1924; 1925) and Cupp (1943), whilst for Chaetoceros to Ostenfeld (1903), Gran (1897), Cupp (1943) and Hernández-Becerril (1996). I also integrated information from recent emendations or revisions of sections in
Chaetoceros (e.g. Li et al., 2016; Xu et al., 2019). Then, I assigned each taxon considered in the phylogenetic analysis to the relevant section using the morphological information provided in Gaonkar et al. (2018) and references therein. An illustration of a typical Chaetoceros species with the morphological terminology used here is provided in Fig. 2.3.

Fig. 2.3. Schematic representation of a typical Chaetoceros species, with the main morphological features relevant to this analysis.

2.3. Results

2.3.1. Dataset characteristics

Of all the genes here amplified, the highest amplification rate was obtained for psbA (83), followed by rbcL (74), and COI (52). For 18S and 28S, we used in total 92 and 88 sequences respectively from Gaonkar et al. (2018) plus six here amplified (three for 18S and three for 28S). A graphical overview of single gene and concatenated alignments is provided in Fig. 2.2. The low amplification success of COI is likely due to primer mismatches with their intended target regions. Indeed, the primers were developed against a conserved region within an exon of the pennate diatom Sellaphora. The known occurrence of introns in mitochondrial genomes of diatoms (Chaetocerotaceae included) as
well as the high substitution rate of the marker may have hampered primer-fit. The nucleotide sequences of \( rbcL \), \( psbA \) and COI as well as newly generated 18S and 28S are available at the accession numbers listed in Table A2.1. The concatenated dataset (Table A2.2) included 100 strains (6 \textit{Bacteriastrum} and 60 \textit{Chaetoceros} species) and 5138 characters partitioned as follows: 18S (bp 1-1724), 28S (bp 1725-2495), \( rbcL \) (bp 2496-3806), \( psbA \) (bp 3807-4733), and COI (bp 4734-5138). The datasets organised per genomic compartment were as follows: 97 strains and 2495 characters for the nuclear data, 94 strains and 2238 characters for the plastid data, and 52 strains and 405 characters for the mitochondrial one.

I did not find any significant saturation at the 3\textsuperscript{rd} codon positions of \( rbcL \), \( psbA \) and COI genes (\( I_{ss} < I_{ss.c} \), Table A2.3 in Appendix II) and, therefore, I assumed that the 1\textsuperscript{st} and 2\textsuperscript{nd} codon positions, known to evolve slower than the 3\textsuperscript{rd}, are also not saturated. The results of this test indicated that the phylogenetic signal of such genes was not eroded by the substitution rates and that sequence similarity is largely due to homology. The \( \chi^2 \) test of homogeneity of state frequencies across taxa detected no compositional heterogeneity (\( p > 0.05 \), Table A2.4 in Appendix II), so excluding its potential impact on phylogenetic inferences.
Fig. 2.4. Individual and concatenated sequence alignments of Chaetocerotaceae dataset. Each row represents an algal strain. N = undetermined bases, – = missing data.
2.3.2. Assignment of species to sections

The list and the description of the sections including the species included in the present study is provided in Table A2.5. I was able to assign most of the species to an extant section (Table A2.5, Appendix II) with few exceptions. Among the latter, was the group constituted by *C. cf. vixvisibilis* and *C. sp. clades Na11C3, Na26B1, Na28A1 and Va7D2*, encompassing heterogeneous taxa that did not show very distinctive morphological features. Other exceptions were *C. cf. pseudodichaeta, C. costatus* and *C. throndsenii*, which show distinctive morphological features not included in any extant section.

2.3.3. Nuclear, plastid and mitochondrial phylogenies

The concatenated nuclear ML tree (18S and partial 28S; Fig. A2.4) resolved *Bacteriastrum* as sister to the genus *Chaetoceros* with high bootstrap support (99%). The subgenus *Chaetoceros* was found to be monophyletic (73 BP), whilst *Hyalochaete* was paraphyletic. All terminal clades were fully resolved, whilst some internal nodes were poorly resolved (Fig. A2.2, Appendix II). The topology of the concatenated plastid tree (*rbcL* and *psbA*; Fig. A2.3 in Appendix II) was not in conflict with that of the nuclear tree, and where it was not in agreement, bootstrap support for those different relationships was not relevant. The only example for such a different relationship was the position of *Bacteriastrum*, which was recovered inside the genus *Chaetoceros*, though without bootstrap support (Fig. A2.3 in Appendix II). In general, the topology is as in the nuclear tree, but bootstrap support for many of the clades is low compared with the nuclear dataset. The COI tree (Fig. A2.4, Appendix II) was rooted using *Bacteriastrum* because no outgroup sequences were available in GenBank. The general topology of the mitochondrial tree resembles that of the trees inferred from the nuclear and plastid datasets, but the majority of the clades received insufficient bootstrap support. To summarise, there was no conflict in the tree topologies inferred from different genomic compartments.
2.3.4. Concatenated phylogenies

ML and BI phylogenies inferred from all the five gene regions concatenated showed the same topology (Fig. 2.5). *Bacteriastrum* formed a well-supported clade as sister to a clade comprising the genus *Chaetoceros* (72 BS, 0.99 PP). Within *Bacteriastrum*, phylogenetic relationships among taxa were well resolved but inconsistent with the sections (Fig. 2.5).
Fig. 2.5. Multigene Maximum Likelihood and Bayesian phylogenetic trees. Numbers at the basis of each node indicate the bootstrap support and the posterior probability respectively. Colours refer to the morphological section to which each taxon was assigned. In grey are indicated the rejected sections. N.A. = species not assigned to any existing section.

Within Chaetoceros, the first clade to branch off comprised in its turn a clade with taxa of section Protuberantia (C. didymus / C. protuberans) as sister to a clade with taxa of
subgenus *Chaetoceros*. Strong support for *Protuberantia* as sister to the subgenus *Chaetoceros* left *Hyalochaete* paraphyletic (Fig. 2.5). Within the subgenus *Chaetoceros*, section *Borealia* was not monophyletic. The remaining taxa in *Hyalochaete* were recovered in a clade (71 BS, 0.99 PP) in which a monophyletic section *Compressa* was resolved as sister to a clade with all remaining taxa (72 BS, 0.99 PP). This clade branched in its turn into two large and well supported clades. The lower one of these in Fig. 2.5 comprised the monophyletic sections *Laciniosa, Cylindrica, Curviseta, Furcellata, Socialia, Simplicia* and a clade with *C. costatus*, and the upper one included in essence three clades. One of these comprised section *Diversa* (only *C. diversus*) inside a paraphyletic section *Stenocincta*. A second one comprised a clade with *C. anastomosans* (Section *Anastomosantia*), *C. cf. vixvisibilis*, and strains belonging to a series of not yet formally described species for which morphological information is available in Gaonkar et al. (2018) as sister to a clade comprising the monophyletic section *Dicladia* and *C. throndsenii*. The third one contained the monophyletic section *Diadema* as sister to *C. constrictus* (section *Constricta*).

The MP tree was congruent with the ML and BI trees but exhibited a few poorly and unresolved relationships (Fig. A2.5). Nonetheless, the position of *Bacteriastrum* as sister genus to *Chaetoceros* was confirmed as well as monophyly of subgenus *Chaetoceros* within paraphyletic *Hyalochaete*. In summary, the three phylogenetic inference methods provided the same results, reinforcing the hypotheses of evolutionary relationships here inferred.

2.3.5. *Comparison between morphological sections and molecular clades*

Given the morphological assignment of taxa to sections and their phylogenetic positions in the concatenated ML and BI trees, I was able to name 16 clades in *Chaetoceros* and 2 in *Bacteriastrum* using the taxonomic division in sections (Table A2.4, Appendix II). A few taxa were not assigned to any section. These consisted of: i) a clade of species that have
not yet been formally described and for which an in-depth morphological and ultrastructural analysis is still needed (C. cf. *vixvisibilis*, *C*. spp. clades Na11C3, Na26B1, Na28A1 and Va7D2); ii) the minute species *C. throndsenii*; iii) *C. costatus*, and iv) *C. cf. pseudodichaeta*. As result, I emended one section, rejected seven and erected three new ones (Fig. 2.5, Table 2.1; see Discussion). The new classification system for the taxa here investigated is shown in Table 2.1. I also assigned to each section species for which both morphological and molecular information was available in literature (Table 2.2).

**Table 2.2. Classification scheme of the family Chaetocerota**ceae. Only sections including taxa utilised in the present study are shown. “Reference for description” refers to publications in which the section is described or amended. “Reference for assignation” refers to publications in which both morphological and molecular information of the species are available.

<table>
<thead>
<tr>
<th>Genus Bacteriastrum Shadbolt</th>
</tr>
</thead>
<tbody>
<tr>
<td>No sectional division</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genus Chaetoceros Ehrenberg</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section</td>
</tr>
</tbody>
</table>


| Chaetoceros sect. nov. Sarno, D. De Luca and Kooistra | **Description:** species with numerous chloroplasts in the cell body and in the setae. Robust, thick, and often very long setae armed with small, often elongated spines. Rimoportula on every valve with the exception of *C. pseudodichaeta*, which has rimoportula only on terminal valves. **Reference for description:** this study. **Assigned species:** *C. atlanticus*, *C. castracanei*, *C. convolutus*, *C. danicus*, *C. dichaeta*, *C. eibenii*, *C. peruvianus*, *C. pseudodichaeta*, *C. rostratus*. **Species assignation:** Gaonkar et al. (2018). |

<p>| Compressa Ostenfeld; emended by Yang Li and Lundholm (in Xu et al., 2019) | <strong>Description:</strong> valves broadly elliptical to compress. Numerous small chloroplasts in each cell. Apertures usually moderately large. Terminal setae little different from others. Intercalary setae of two types: thin, common setae and heavy special setae. Heavy setae contorted with spiralling rows of spines and poroids, or heavy setae not visually contorted lacking rows of spines and poroids. Resting spores smooth or with a row of spicules. <strong>References for description:</strong> Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996); Xu et al. (2019). <strong>Assigned species:</strong> <em>C. acadianus</em>, <em>C. bifurcatus</em>, <em>C. compressus</em>, <em>C. contortus</em>. |</p>
<table>
<thead>
<tr>
<th>Taxon</th>
<th>Description</th>
<th>Assigned species</th>
<th>Species assignation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constricta Ostenfeld</strong></td>
<td>Description: cells with one or two chloroplasts and a marked constriction at the base of the valve mantle. Girdle at least one-third the length of the cell. Terminal setae mostly thicker than the others. Resting spores, when present, about the middle of the cell with numerous spines on both valves.</td>
<td><em>C. constrictus</em></td>
<td>Gaonkar et al. (2018).</td>
</tr>
<tr>
<td><strong>Costata sect. nov.</strong> Sarno, D. De Luca and Kooistra</td>
<td>Description: chains generally long, without differentiated terminal setae. One chloroplast. Each valve possesses four submarginal flattened protuberances, two on each pole of the valve, joining with those of the sibling valves. Girdle bands with a distinct thickened longitudinal rib at one edge also visible in LM.</td>
<td><em>C. costatus</em></td>
<td>Gaonkar et al. (2018).</td>
</tr>
<tr>
<td><strong>Curviseta Ostenfeld; emended by Gran</strong></td>
<td>Description: chains usually curved, with setae all bent in one direction without special end cells. One chloroplast.</td>
<td><em>C. curviset us, C. debilis, C. pseudocurvisetus, C. tortissimus</em></td>
<td>Gaonkar et al. (2018).</td>
</tr>
<tr>
<td><strong>Cylindrica Ostenfeld</strong></td>
<td>Description: cells with valves nearly circular (cylindrical). Apertures very narrow. Small, numerous chloroplasts. Terminal setae not thicker than others. Resting spores about middle of the cells, smooth or with spines.</td>
<td><em>C. lauderi, C. teres</em></td>
<td>Gaonkar et al. (2018).</td>
</tr>
<tr>
<td><strong>Diadema (Ehrenberg) Ostenfeld; emended by Gran</strong></td>
<td>Description: one chloroplast per cell. Chains long with conspicuous terminal setae. Primary valve of resting spores with branched processes or crown of spines, or sometimes smooth.</td>
<td><em>C. diadema, C. rotosporus, C. seiracanthus, Chaetoceros sp. Clade Na13C1</em></td>
<td>Li et al. (2013); Gaonkar et al. (2018).</td>
</tr>
<tr>
<td><strong>Dicladia (Ehrenberg) Gran; emended by Lebour</strong></td>
<td>Description: multiple chloroplasts per cell and setae with large pores. Terminal and intercalary setae similar. Resting spores, when known, with two horns armed with small branches on primary valves.</td>
<td><em>C. decipiens, C. elegans, C. laevisporus, C. lorenzianus, C. mannaii, C. mitra, C. pauciramosus</em></td>
<td>Li et al. (2017); Chen et al. (2018).</td>
</tr>
<tr>
<td><strong>Furcellata Ostenfeld</strong></td>
<td>Description: chains generally loose, without differentiated terminal setae. One chloroplast. Resting cells eccentrically arranged in mother cell, lying close together two and two, with thick coalesced setae; with smooth valves or with short spines.</td>
<td><em>C. cinctus, C. radicans</em></td>
<td>Gaonkar et al. (2017).</td>
</tr>
<tr>
<td><strong>Laciniosa Ostenfeld</strong></td>
<td>Description: one or two chloroplasts per cell. Girdle rather long. Aperture large. Terminal setae usually thicker than the others, not diverging greatly.</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Minima sect. nov.</strong></td>
<td><strong>Protuberantia</strong></td>
<td><strong>Simplicia</strong></td>
<td><strong>Socialia</strong></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Sarno, D. De Luca and Kooistra</td>
<td>Ostenfeld; emended by Hernández-Becerril</td>
<td>Ostenfeld</td>
<td>Ostenfeld</td>
</tr>
<tr>
<td><strong>Description:</strong> very small, solitary species usually bearing one seta on a valve and one or two on the other. One chloroplast. Rimoportula very reduced in <em>C. throndsenii</em> and absent in <em>C. minimus</em>.</td>
<td><strong>Description:</strong> two chloroplasts per cell, each with a large pyrenoid situated in a protuberance in the middle of the valve surface. Valves with poroids. Resting spores paired with two long setae or free without setae.</td>
<td><strong>Description:</strong> cells small and fragile, generally single or two or three together. In case of chain formation, there is no differentiation of terminal setae.</td>
<td><strong>Description:</strong> chains irregular and curved embedded in mucilage, forming irregularly spherical colonies. One chloroplast. Resting spores smooth or with small spines.</td>
</tr>
<tr>
<td><strong>References for description:</strong> this study.</td>
<td><strong>References for description:</strong> Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996); Gaonkar et al. (2018).</td>
<td><strong>References for description:</strong> Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).</td>
<td><strong>References for description:</strong> Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).</td>
</tr>
<tr>
<td><strong>Species assignation:</strong> Gaonkar et al. (2018).</td>
<td><strong>Species assignation:</strong> Gaonkar et al. (2018).</td>
<td><strong>Species assignation:</strong> Li et al. (2016); Gaonkar et al. (2018).</td>
<td><strong>Species assignation:</strong> Gaonkar et al. (2018).</td>
</tr>
</tbody>
</table>

**References for description:** Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).

**Assigned species:** *C. brevis*.

**Species assignation:** Gaonkar et al. (2018).

---

**2.4. Discussion**

**2.4.1. General comments to the dataset**

The phylogenetic trees inferred from the three genomic compartments (nuclear, plastid and mitochondrial) are congruent, providing no indication of different evolutionary histories.

According to this result, I conclude that during speciation of Chaetocerotaceae, the
corresponding gene copies in each species has been distributed in a pattern reflecting the parent species trees. This phenomenon is not universal, since gene trees and species trees do not always agree because of population-level lineage sorting (Pollard, et al., 2006), hybridization (McBreen and Lockhart, 2006), gene duplication and differential loss, and lateral gene transfer (LGT), where genes are exchanged between lineages (Dagan and Martin, 2006; Beiko et al., 2005; Leigh et al., 2008). However, most of phylogenetic studies dealing with the analysis of multiple genes often do not explicitly deal with the issue of congruence (Rokas et al., 2005; Qiu et al., 2006; James et al., 2006), making difficult to assess its extent across different taxa. In the few studies that analysed such issue in diatoms, no conflict among different gene trees was observed (e.g. Theriot et al., 2010; Souffreau et al., 2011; Kociolek et al., 2013).

In this study, as result of absence of conflicting topologies among trees, the concatenation of all the sequences increased the number of positively informative sites and so the phylogenetic signal (see e.g., Theriot et al., 2010; 2015). Indeed, the multigene tree shows better resolved relationships than the concatenated ones from each of the genomic compartments separately, as well as the trees based on single markers, e.g. in Gaonkar et al. (2018) and in Xu et al. (2019). Moreover, none of the markers included in our analysis shows saturation of the phylogenetic signal. Thus, I assume that the well supported clades in the concatenated tree can be used to make phylogenetically informed taxonomic decisions.

Most of the sections for which strains of multiple species have been included are monophyletic, and the synapomorphies of the clades are here used to validate, describe or emend the sections they belong to. For the purposes of this work, I aimed at a classification that is both supported phylogenetically and retains practical properties (Mayr, 1982; Benton, 2000). This approach is not mutually exclusive, considering that the objects of classifications should share similarities because of common descent (Mayr, 1942). I
retained only monophyletic sections, made emendations whenever possible and erected new sections only were complete and supported information was available.

2.4.2. Phylogenetic position of the genera *Bacteriastrum* and *Chaetoceros*

Results of the present study indicate that *Bacteriastrum* and *Chaetoceros* are each other’s monophyletic sister genera. This finding contrasts with phylogenies inferred exclusively from partial 28S rDNA sequences, which resolve the former inside the latter, though with meagre support, if any (e.g., Bosak et al., 2015; Gaonkar et al., 2018). My results confirm the hypothesis that *Bacteriastrum* constitutes a genus different from *Chaetoceros* (e.g. Pritchard, 1861; Round et al., 1990; Hasle and Syvertsen, 1996). Within *Bacteriastrum*, the sections *Isomorpha* and *Sagittata* proposed by Pavillard (1925) are unsupported because the former is polyphyletic and the latter paraphyletic. This is because *B. jadranum*, placed in the section *Isomorpha* by Godrijan et al. (2012), is sister to members of section *Sagittata* and only distantly related to *B. hyalinum* (*Isomorpha*). The non-monophyly of these sections invalidates them and shows that their defining character states of terminal setae orientation are not synapomorphies. Thus, I reject the two sections since there is neither a phylogenetic reason nor a utilitarian one to maintain them.

2.4.3. Subgeneric division

The subgenus *Hyalochaete* was erected by Gran (1897) to include all the *Chaetoceros* species without chloroplasts in the setae. Therefore, this “catch-all” taxonomic category includes a highly diverse collection of species that basically share general features encountered in all *Chaetoceros* species; their only defining feature, absence of chloroplasts in the setae, is not a phylogenetically sound character state. Indeed, my results show that *Hyalochaete* is paraphyletic, and therefore, I reject it.

The subgenus *Chaetoceros* was formerly described as *Phaeoceros* by Gran (1897) to include species characterised by numerous plastids in both the cell body and the setae.
According to the concatenated phylogeny “plastids in the setae,” (Gran, 1897) is a synapomorphy of subgenus *Chaetoceros*, whereas other features believed to define this subgenus, actually do not. Gran (1897) mentions “spores unknown” for subgenus *Chaetoceros*. None of the studies on species in this subgenus have reported spore formation, with one exception: *C. eibenii* (von Stosch et al., 1973; Jensen and Moestrup, 1998). Since this species is the first to branch off within the subgenus, ability to form resting spores, must have gone lost in the last common ancestor of the sister clade of *C. eibenii* because spore formation has been confirmed in most if not all of the other species in the genus *Chaetoceros* (Ishii et al., 2011). Thus, the absence of spores does not define the subgenus *Chaetoceros*. Another character state, “presence of rimoportulae in terminal as well as in intercalary valves” (see Hasle and Syvertsen, 1996) is a symplesiomorphy because strain El1C1 (Eilat, Israel) here identified as *C. cf. pseudodichaeta* resolves within the clade of the subgenus, but exhibits rimoportulae only in its terminal valves. Therefore, “presence of rimoportulae in terminal as well as in intercalary valves” is not a defining character state of the subgenus, either. Although I cannot strictly reject the subgenus *Chaetoceros*, but having already rejected the subgenus *Hyalochaete* for its paraphyly, I argue that there is no utilitarian reason to keep it. It could be better treated as a new section *Chatoceros*, here proposed, to include all the species with chloroplasts not only in the central cell body, but also in the setae. At this point, the only remaining subgenus is *Bacteriastroidea*. It was erected by Hernández-Becerril (1993b) to include the only species *C. bacteriastroides* for its peculiar morphology, intermediate between *Bacteriastrostrum* and *Chaetoceros* (cylindrical valves, intercalary ones with three pairs of setae, two of which very reduced). We agree that, considering the available data, it deserves a dedicated taxonomic category. However, no DNA is available for this species and, therefore, its molecular phylogenetic position within the Chaetocerotaceae is unknown. Genetic data may either confirm the validity of a dedicated taxonomic category or justify its inclusion into a pre-existing section. Therefore,
considering available information and following our way of action, I reject *Bacteriastroidea* as subgenus and consider it provisionally as a section of the genus *Chaetoceros*.

2.4.4. The sectional division

Ostenfeld (1903) was the first to subdivide the genus *Chaetoceros* into sections. Later authors (cit) added sections to accommodate species new to science that did not fit in the sections of Ostenfeld or to split pre-existing sections based on newly defined characters and their states. Newly described species are usually sorted without much ado into those existing sections (e.g. Li et al., 2016) or the sectional description needs to be emended only slightly to accommodate species new to science (Xu et al., 2019). However, some species such as *C. phuketensis* (Rines et al., 2000) are not. The results of my phylogenetic explorations show that the sections *Dicladia, Constricta, Diadema, Laciniosa, Cylindrica, Curviseta, Furcellata, Socialia, Simplicia, Compressa* and *Protuberantia* are monophyletic, and therefore, considered valid. For the species that do not fit in any pre-existing section, a possible course of action would be to create a new section for every one of them showing a unique feature (e.g. section *Anastomosantia* for *C. anastomosans* with its silica bridge linking sibling setae; section *Rostrata* for *C. rostratus* with its fused rimoportulae of sibling valves). However, Rines et al. (2000) pointed out that this would lead *in extremis* to placing every morphologically distinct species in its own section, thereby defying the utilitarian purpose of sections. I recognise that some peculiar characters are important for species identification purposes, but I agree with Rines et al. (2000) to refrain from considering these as reasons to create new sections. Everytime the morphology of a new species does not quite fit the sectional description, I simply decided to emend the latter (see e.g., Xu et al., 2019).

For example, Ostenfeld (1903) placed *C. diversus* in a section called *Diversa* because this species’ intercalary setae are far more robust than its terminal ones. This section was
maintained by Cupp (1943) and Hernández-Becerril (1996). However, the robust intercalary setae and delicate terminal setae of *C. diversus* resemble the robust terminal setae and delicate intercalary setae of the species in its paraphyletic “mother” section *Stenocincta*. Other characteristics such as a single plastid per cell and a narrow aperture are, in fact, shared between the two sections (Gaonkar et al., 2018). The phylogenetic position of *C. diversus* in our multigene phylogeny, inside the clade encompassing all the other taxa belonging to *Stenocincta*, provides further evidence of common ancestry. Therefore, I decided to emend the section *Stenocincta* to include taxa previously within *Diversa* and I hereby propose to reject the section *Diversa*.

Similarly, Ostenfeld (1903) and Hernández-Becerril (1996) considered the presence of spines on the spore valves the defining feature of the section *Diadema*. Instead, Cupp (1943), following Gran’s (1905) emended description, provided a broader definition, which includes also features of the vegetative cells and which accommodates the spiny resting spores of *C. diadema* and *C. seiracanthus* as well as the smooth ones of *C. rotosporus* (Li et al., 2013). My phylogenetic tree supports the findings of Li et al. (2013) and Gaonkar et al. (2018) that the aforementioned taxa in section *Diadema* form a clade. Thus, all these species, including the recently described *C. rotosporus*, fit perfectly fine in the monophyletic section *Diadema*.

Most of the sections in my phylogeny are monophyletic and their defining character states are synapomorphies, but for some of them the taxonomic coverage is still low. Many more species need to be added to confirm their monophyly. For the sections *Compressa, Constricta, Rostrata* and *Simplicia* I was limited in testing their robustness because of the low numbers of species available, but I have no evidence that newly added will falsify the current classification into sections. For instance, in my phylogeny, only two species in section *Compressa* are available, but a phylogeny inferred from 28S rDNA sequences in Xu et al. (2019) shows this section to be monophyletic given far wider taxon coverage. Likewise, in my multigene phylogeny the section *Simplicia* is represented only by *C.
tenuissimus. Li et al. (2016) included in their 18S rDNA phylogeny (fig. 25 of their paper) several other taxa in this section, most of which resolved together with C. tenuissimus in a weakly supported clade (clade IV). However, several strains that were poorly identified at morphological level, made some species polyphyletic and made the authors considering the section non monophyletic. In Gaonkar et al. (2018), C. tenuissimus forms a well-supported clade with taxonomically validated strains of C. neogracilis (section Simplicia; see Balzano et al., 2017). The 18S sequences of C. cf. neogracile in Li et al. (2016) are virtually identical to those in Balzano et al. (2017), and therefore they made me hypothesise that the section Simplicia is monophyletic.

A few species in my tree do not fit in any of the existing sections. For instance, Cupp (1943) placed C. costatus in the section Stenocincta whereas Lebour (1930) put it in Curviseta (under the name C. adhaerens). Instead, in my multigene phylogeny the strains of this species are recovered as nearest neighbour of a clade comprising sections Socialia and Furcellata. Neither morphological nor ultrastructural characters are shared with these neighbour sections. However, this species possesses several peculiar morphological features (e.g. four submarginal flattened protuberances joining with those of the sibling valves) that justify its inclusion into a dedicated section. Therefore, I propose to erect a new section for C. costatus (Section Costata). Such a section does not affect monophyly of the related sections Furcellata, Socialia and Simplicia.

The species description of C. throndsenii does not provide any assignment to a section (Marino et al., 1991). In the 18S and 28S trees by Gaonkar et al. (2018) it forms a clade with the morphologically similar species C. minimus, with which it shares a small cell size, a single cell habit, a reduction in the number of setae per cell (2 to 3) and a similarly shaped resting spore. These morphological similarities were already reported in Marino et al. (1991). I had no access to DNA of C. minimus and therefore this species was not included in my multigene tree. In my phylogeny, C. throndsenii is recovered on a long branch as sister to Dicladia, though it does not share any evident character state with this
section. Therefore, I propose placing *C. minimus* and *C. throndsenii* into a new section here called *Minima*.

Regarding *Anastomosantia*, Ostenfeld (1903) established this section for *C. anastomosans* based on the silica bridges linking sibling setae, which in our multigene tree constitutes an autapomorphy. Yet, this species is recovered in a well-supported clade with *C. cf. vixvisibilis* and a whole series of still undescribed taxa. *Chaetoceros dayaensis* (Li et al., 2015) also belongs to this clade (Gaonkar et al., 2018). Li et al. (2015) did not place *C. dayaensis* in any section nor did they establish a new section for it. In this case, I refrain from erecting a new section for the whole clade because the possible morphological synapomorphies defining this clade are still to be uncovered, and I keep the section *Anastomosantia* exclusively for *C. anastomosans*.

According to my multigene phylogeny, the sections *Borealia* (Ostenfeld 1903) and *Peruviana* (Hernández-Becerril, 1996) are not monophyletic and so I have to reject them. The presence of intercalary processes that link cells in chains in *C. rostratus*, defines the monotypic section *Rostrata* (Hernández-Becerril, 1998). Although these specialised processes are useful for taxonomic identification, they constitute an autapomorphy. Maintaining the section *Rostrata* requires the establishment of a whole series of additional sections in the clade, several of which will be monotypic, and without any clear defining character states, which is not particularly utilitarian. The same accounts for *C. atlanticus* and *C. dichaeta* in section *Atlantica* (Ostenfeld 1903), which are resolved in the phylogenies in Gaonkar et al. (2018) close to *C. peruvianus* (*Peruviana*) and *C. danicus* (*Borealia*). Therefore, I propose to reject not only *Borealia* and *Peruviana*, but also *Rostrata* and *Atlantica* and to erect a new Section *Chaetoceros* for all the taxa sharing the presence of chloroplasts in the central cell body as well as in the setae. The name *Chaetoceros* follows the rules of botanical nomenclature, according to which any
subdivision of a genus that includes the type specimen must adopt the name of the genus to which it is assigned.

**Section Chaetoceros** D. Sarno, D. De Luca and W.H.C.F. Kooistra, sect. nov.

Species with numerous chloroplasts in the cell body and in the setae. Robust, thick, and often very long setae armed with small, often elongated spines. Rimoportula on every valve with the exception of *C. pseudodichaeta*, which has rimoportula only on terminal valves (Fig. 2.6).

![Fig. 2.6](image)

*Fig. 2.6. Chaetoceros danicus* (A) and *C. rostratus* (B), two members of the Section Chaetoceros.*

**Section Costata** D. Sarno, D. De Luca and W.H.C.F. Kooistra, sect. nov.

Chains generally long, without differentiated terminal setae. One chloroplast. Each valve possesses four submarginal flattened protuberances, two on each pole of the valve, joining with those of the sibling valves. Girdle bands with a distinct thickened longitudinal rib at one edge also visible in LM (Fig. 2.7).
Fig. 2.7. *Chaetoceros costatus*, *Section Costata*. The arrow indicates the submarginal flattened protuberance typical of the Section.

**Section Minima** D. Sarno, D. De Luca and W.H.C.F. Kooistra, sect. nov.

Very small, solitary species usually bearing one seta on a valve and one or two on the other. One chloroplast. Rimoportula very reduced in *C. throndsenii* and absent in *C. minimus* (Fig. 2.8).

Fig. 2.8. *Chaetoceros minimus* (A) and *C. throndsenii* (B), two members of the Section *Minima*. Photo credit: Susanne Busch. From Nordic Microalgae (http://www.nordicmicroalgae.org).

**Section Stenocincta** Ostenfeld 1903 emend. D. Sarno, D. De Luca and W.H.C.F. Kooistra

Emended diagnosis: One chloroplast per cell. Usually narrow aperture. Terminal setae generally thicker than the intercalary ones. Instead, *C. diversus* possesses thin terminal
setae and generally two types of intercalary setae, differing in orientation and robustness (Fig. 2.9).

Fig. 2.9. *Chaetoceros affinis* (A) and *C. diversus* (B), two members of the Section *Stenocincta*. Arrows indicate some characteristics of the members of the Section, i.e. thick terminal setae (A) and two types of intercalary setae in *C. diversus* (B).

2.4.5. Future directions

The multigene analysis here inferred using the 18S, 28S, *rbcL*, *psbA* and COI genes has provided a robust phylogenetic hypothesis depicting the evolutionary history of Chaetocerotaceae. The comparison between infrageneric taxa based on morphology and the clades in the tree revealed congruence for most of them, falsified others, and highlighted that future work is needed on unresolved taxa. I rejected the three subgenera within *Chaetoceros* and seven sections (two in *Bacteriastrum* and five in *Chaetoceros*), emended one section and described three new ones. I refrained from elevating the sections into genera of their own. Splitting would be justified by the fact that the genetic distances among the chaetocerotacean sections are comparable with those observed among families or even orders in other diatom lineages. For instance, the Order Thalassiosirales has been split into a large series of narrowly defined genera. However, this has left the genus *Thalassiosira* paraphyletic. Although I have demonstrated that most of the Sections in *Chaetoceros* are monophyletic, I believe that the utilitarian principle has precedence. *Chaetoceros* is easily recognised because of its defining feature, the setae, which are
visible in LM. If the sections are elevated to genera, however, these new genera may not be recognised so easily. The sections can be further supported by ultrastructural features of valves and setae (Chamnansinp et al. 2015; Bosak and Sarno, 2017; Gaonkar et al., 2018; Xu et al., 2019) but this requires in depth comparison of the species in these sections. Future work may include the adding of new species to our phylogenetic tree as well as of new sections, in order to have a better and more complete view of the evolution of such important family of marine planktonic diatom. To date, this study represents a further step towards a better understanding of the evolution of Chaetocerotaceae.

References


Chamnansinp, A., Li, Y., Lundholm, N., Moestrup, Ø. (2013). Global diversity of two widespread, colony-forming diatoms of the marine plankton, Chaetoceros socialis
(syn. *C. radians*) and *Chaetoceros gelidus* sp. nov. *Journal of Phycology, 49*(6), 1128-1141.


85


Hernández-Becerril, D. U. (1993b). Note on the morphology of two planktonic diatoms: Chaetoceros bacteriastroides and C. seychellarus, with comments on their


Li, Y., Lundholm, N., Moestrup, Ø. (2013). Chaetoceros rotosporus sp. nov. (Bacillariophyceae), a species with unusual resting spore formation. Phycologia, 52(6), 600-608.


Appendix II
Fig. A2.1. Light microscopy photographs of *Bacteriastrum* and *Chaetoceros* species utilised in the present study. Pictures are from Gaonkar (2017), Gaonkar et al. (2018) and Dr. Wiebe Kooistra lab collection.

*Bacteriastrum elegans*  
*B. furcatum 2*  
*B. hyalinum*  
*B. jadranum*  
*B. mediterraneum*  
*Chaetoceros affinis*
C. cf. tortissimus  
C. cf. vixvisibilis  

C. cinctus  
C. circinalis  

C. constrictus  
C. contortus  

C. contortus cf. var. contortus  
C. costatus
C. sp. Va7D2
C. sporotrupruccatus
C. teres
C. throtdsenii
C. tortissimus
Fig. A2.2. Maximum Likelihood (ML) tree of concatenated nuclear genes (18S and 28S). Numbers at each node refer to bootstrap support after 1000 replicates.
Fig. A2.3. Maximum Likelihood (ML) tree of concatenated plastid genes (\textit{rbc}L and \textit{psb}A). Numbers at each node refer to bootstrap support after 1000 replicates.
Fig. A2.4. Maximum Likelihood (ML) tree of mitochondrial COI gene. Numbers at each node refer to bootstrap support after 1000 replicates.
Fig. A2.5. Maximum Parsimony (MP) tree. Numbers at each node refer to bootstrap support after 1000 replicates.
Table A2.1. List of taxa (species and strains) utilised in the present study, including sampling localities and dates and accession numbers for each gene amplified. NA = not available.

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Sampling locality</th>
<th>Sampling date</th>
<th>18S</th>
<th>28S</th>
<th>rbcL</th>
<th>psbA</th>
<th>COI</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. elegans</td>
<td>Na25A3</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97220</td>
<td>MG914436</td>
<td>NA</td>
<td>MK64235</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA)</td>
</tr>
<tr>
<td>B. furcatum 2</td>
<td>Na8A3</td>
<td>Gulf of Naples (Italy)</td>
<td>06/02/2014</td>
<td>MG97235</td>
<td>MG914439</td>
<td>MK64249</td>
<td>MK64235</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
</tr>
<tr>
<td>B. hyalinum</td>
<td>Na10B1</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG97220</td>
<td>MG914440</td>
<td>NA</td>
<td>NA</td>
<td>MK64243</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (COI)</td>
</tr>
<tr>
<td>B. jadranum</td>
<td>Na19C1</td>
<td>Gulf of Naples (Italy)</td>
<td>30/07/2014</td>
<td>MG97235</td>
<td>MG914441</td>
<td>NA</td>
<td>MK64236</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA)</td>
</tr>
<tr>
<td>B. jadranum</td>
<td>Na19C3</td>
<td>Gulf of Naples (Italy)</td>
<td>30/07/2014</td>
<td>MG97235</td>
<td>MG914442</td>
<td>MK64249</td>
<td>NA</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL)</td>
</tr>
<tr>
<td>B. mediterraneum</td>
<td>Na1C4</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG97220</td>
<td>MG914444</td>
<td>NA</td>
<td>NA</td>
<td>MK64243</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (COI)</td>
</tr>
<tr>
<td>B. mediterraneum</td>
<td>Na29B3</td>
<td>Gulf of Naples (Italy)</td>
<td>01/12/2014</td>
<td>NA</td>
<td>MG914446</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (28S); this study (rbcL/psbA/COI)</td>
</tr>
<tr>
<td>B. parallelum</td>
<td>newLA2</td>
<td>Gulf of Naples (Italy)</td>
<td>21/05/2013</td>
<td>MG97220</td>
<td>MG914447</td>
<td>NA</td>
<td>NA</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (COI)</td>
</tr>
<tr>
<td>C. affinis</td>
<td>Na49A2</td>
<td>Gulf of Naples (Italy)</td>
<td>26/10/2016</td>
<td>NA</td>
<td>MG914453</td>
<td>NA</td>
<td>MK64236</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA)</td>
</tr>
<tr>
<td>C. anastomosans</td>
<td>Na14C2</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG97235</td>
<td>MG914456</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
</tr>
<tr>
<td>C. anastomosans</td>
<td>Na14C3</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG97235</td>
<td>MG914457</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
</tr>
<tr>
<td>C. brevis 1</td>
<td>Na7B1</td>
<td>Gulf of Naples (Italy)</td>
<td>18/01/2014</td>
<td>MG97221</td>
<td>MG914464</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
</tr>
<tr>
<td>C. brevis 2</td>
<td>Na7C2</td>
<td>Gulf of Naples (Italy)</td>
<td>17/01/2014</td>
<td>MG97221</td>
<td>MG914467</td>
<td>NA</td>
<td>MK64236</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA)</td>
</tr>
<tr>
<td>C. brevis 3</td>
<td>Ch9B3</td>
<td>Concepción (Chile)</td>
<td>29/10/2013</td>
<td>MG97221</td>
<td>MG914468</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S)</td>
</tr>
<tr>
<td>C. cf. convolutus</td>
<td>Ch5C4</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>MG97222</td>
<td>MG914482</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
</tr>
<tr>
<td>C. cf. convolutus</td>
<td>L7-B6</td>
<td>Lohafex experiment</td>
<td>NA</td>
<td>LC466960</td>
<td>NA</td>
<td>MK64249</td>
<td>NA</td>
<td>NA</td>
<td>This study</td>
</tr>
<tr>
<td>C. cf. decipiens</td>
<td>El6B1</td>
<td>Eilat, Red Sea (Israel)</td>
<td>31/01/2016</td>
<td>NA</td>
<td>LC466963</td>
<td>MK64249</td>
<td>MK64236</td>
<td>MK64244</td>
<td>This study</td>
</tr>
<tr>
<td>C. cf. lorenzianus</td>
<td>E11A4</td>
<td>Eilat, Red Sea (Israel)</td>
<td>31/01/2016</td>
<td>NA</td>
<td>NA</td>
<td>MK64250</td>
<td>MK64236</td>
<td>MK64244</td>
<td>This study</td>
</tr>
<tr>
<td>C. cf. pseudodichaeta</td>
<td>E11C1</td>
<td>Eilat, Red Sea (Israel)</td>
<td>31/01/2016</td>
<td>MG97230</td>
<td>MG914586</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL)</td>
</tr>
<tr>
<td>C. cf. tortissimus</td>
<td>Na18C4</td>
<td>Gulf of Naples (Italy)</td>
<td>01/07/2014</td>
<td>MG97227</td>
<td>MG914640</td>
<td>NA</td>
<td>MK64237</td>
<td>NA</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA)</td>
</tr>
<tr>
<td>C. cf. tortissimus</td>
<td>Na28B3</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97227</td>
<td>MG914643</td>
<td>MK64250</td>
<td>MK64236</td>
<td>MK64244</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/COI)</td>
</tr>
<tr>
<td>Species</td>
<td>Accession Numbers</td>
<td>Collection Sites</td>
<td>Dates</td>
<td>Refs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------</td>
<td>---------------------------------------</td>
<td>----------------</td>
<td>-------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. cf. vizzobisbis</td>
<td>Na28A4</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. cf. vizzobisbis</td>
<td>Na28A4</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. cinctus</td>
<td>Ch3C4</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>Gaonkar et al. 2017 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. cincinnalis</td>
<td>Na15C2</td>
<td>Gulf of Naples (Italy)</td>
<td>24/04/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. constrictus</td>
<td>Ch12C1</td>
<td>Las Cruces (Chile)</td>
<td>04/11/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. contortus cf. var. contortus</td>
<td>Na31B2</td>
<td>Gulf of Naples (Italy)</td>
<td>07/04/2015</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. costatus</td>
<td>Na1A3</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. costatus</td>
<td>Ro1B1</td>
<td>Roscoff Estacade (France)</td>
<td>11/08/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 1</td>
<td>Na10C1</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 1</td>
<td>Ro3B2</td>
<td>Roscoff Estacade (France)</td>
<td>11/08/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 2a</td>
<td>Na1C1</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 2b</td>
<td>Ch5B1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 2c</td>
<td>El6A2</td>
<td>Eilat, Red Sea (Israel)</td>
<td>31/01/2016</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>El4A2</td>
<td>Eilat, Red Sea (Israel)</td>
<td>31/01/2016</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>Na3C4</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. danicus</td>
<td>Na9B4</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. debilis 2</td>
<td>MM24-A3</td>
<td>Southern Ocean (Atlantic)</td>
<td>Oct. 2004</td>
<td>Koosstra et al. 2010 (28S); Gaonkar et al. 2018 (18S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. debilis 2</td>
<td>MM24-C3</td>
<td>Southern Ocean (Atlantic)</td>
<td>Oct. 2004</td>
<td>Koosstra et al. 2010 (28S); Gaonkar et al. 2018 (18S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. debilis 3</td>
<td>Ch1A1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. decipiens</td>
<td>Na28A2</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (rbcL/psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diadema 1</td>
<td>Ch4A1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diadema 1</td>
<td>Na13B1</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diadema 2</td>
<td>Ch5C1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>Gaonkar et al. 2018 (18S/28S); this study (psbA/COI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>Location</td>
<td>Date</td>
<td>Accession Numbers</td>
<td>Study Reference</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. dichotoensis</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>KY852272</td>
<td>28S; this study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. didymus 1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>MG972270</td>
<td>Gaonkar et al. 2017 (18S/28S); this study</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. didymus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>29/07/2014</td>
<td>MG972271</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diversus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>10/09/2014</td>
<td>NA</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diversus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG972235</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diversus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>07/11/2016</td>
<td>NA</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diversus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG972236</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. diversus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>29/10/2013</td>
<td>MG972295</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>01/11/2013</td>
<td>NA</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG972237</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 1</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG972238</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>28/07/2015</td>
<td>MG972295</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>26/08/2015</td>
<td>NA</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>01/11/2013</td>
<td>MG972290</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>31/10/2013</td>
<td>MG914567</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>05/11/2013</td>
<td>MG914569</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>16/10/2013</td>
<td>MG972292</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>16/10/2013</td>
<td>MG972294</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>newEA1</td>
<td>28/03/2013</td>
<td>MG972298</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Las Cruces (Chile)</td>
<td>01/11/2013</td>
<td>MG972296</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Concepción (Chile)</td>
<td>29/10/2013</td>
<td>MG972299</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. lorenzianus 2</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG972304</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. radicans</td>
<td>Las Cruces (Chile)</td>
<td>29/10/2013</td>
<td>KY852263</td>
<td>GAONKAR et al. 2018 (18S/28S); this study (psbA)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Species</td>
<td>Accession</td>
<td>Location</td>
<td>Collection Date</td>
<td>GenBank Accession</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. radicans</td>
<td>Ch11A4</td>
<td>Las Cruces (Chile)</td>
<td>01/11/2013</td>
<td>KY852262</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. rostratus</td>
<td>Na1C3</td>
<td>Gulf of Naples (Italy)</td>
<td>26/11/2013</td>
<td>MG97230 7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. rostratus</td>
<td>newDA3</td>
<td>Gulf of Naples (Italy)</td>
<td>28/03/2013</td>
<td>MG97231 0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. rotosporus</td>
<td>Na22B1</td>
<td>Gulf of Naples (Italy)</td>
<td>10/09/2014</td>
<td>MG97235 0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. socialis</td>
<td>Na33B1</td>
<td>Gulf of Naples (Italy)</td>
<td>14/07/2015</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na11C3</td>
<td>Na11C3</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG97232 8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na11C3</td>
<td>Na43A1</td>
<td>Gulf of Naples (Italy)</td>
<td>15/03/2016</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na12A3</td>
<td>Na9A3</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na13C2</td>
<td>Na13C2</td>
<td>Gulf of Naples (Italy)</td>
<td>19/03/2014</td>
<td>MG97234 4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na17B2</td>
<td>Na17B2</td>
<td>Gulf of Naples (Italy)</td>
<td>01/07/2014</td>
<td>MG97233 4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na26B1</td>
<td>Na26B1</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97232 9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Na28A1</td>
<td>Na28A1</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97236 4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Va7D2</td>
<td>Na43A4</td>
<td>Gulf of Naples (Italy)</td>
<td>15/03/2016</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sp. Clade Va7D2</td>
<td>Na44B3</td>
<td>Gulf of Naples (Italy)</td>
<td>15/03/2016</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sporotrichatus</td>
<td>Ch2A4</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>KY852270</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. sporotrichatus</td>
<td>Ch9C4</td>
<td>Concepción (Chile)</td>
<td>29/10/2013</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tenuissimus</td>
<td>Na26A1</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97234 4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tenuissimus</td>
<td>Na36B4</td>
<td>Gulf of Naples (Italy)</td>
<td>26/08/2015</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tenuissimus</td>
<td>Na44A1</td>
<td>Gulf of Naples (Italy)</td>
<td>31/05/2016</td>
<td>MG97231 5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tereis</td>
<td>Ch13B2</td>
<td>Las Cruces (Chile)</td>
<td>05/11/2013</td>
<td>NA</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tereis</td>
<td>Ch5A1</td>
<td>Las Cruces (Chile)</td>
<td>16/10/2013</td>
<td>MG97231 7</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. thronsenii</td>
<td>Na45B3</td>
<td>Gulf of Naples (Italy)</td>
<td>31/05/2016</td>
<td>MG97232 3</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tortissimus</td>
<td>Na25A2</td>
<td>Gulf of Naples (Italy)</td>
<td>07/10/2014</td>
<td>MG97232 5</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detonula confervaceae</td>
<td>CCMP353</td>
<td>Culture collection</td>
<td>NA</td>
<td>HQ912617</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydrosera sp.</td>
<td>CYTX025</td>
<td>NA</td>
<td>NA</td>
<td>HQ912683</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This study
Theriot et al. 2010 (18S, rbcL); 2015 (psbA)
<table>
<thead>
<tr>
<th>Species</th>
<th>Collection ID</th>
<th>Accession 1</th>
<th>Accession 2</th>
<th>Accession 3</th>
<th>Accession 4</th>
<th>Accession 5</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terpsinoe musica</td>
<td>NHOP43</td>
<td>NA</td>
<td>HQ912682</td>
<td>NA</td>
<td>HQ912546</td>
<td>NA</td>
<td>Theriot et al. 2010 (18S, rbcL); 2015 (psbA)</td>
</tr>
<tr>
<td>Thalassiosira pseudonana</td>
<td>CCMP1335</td>
<td>NA</td>
<td>HQ912555</td>
<td>NA</td>
<td>HQ912419</td>
<td>NA</td>
<td>Theriot et al. 2010 (18S, rbcL); 2015 (psbA)</td>
</tr>
</tbody>
</table>
Table A2.2. Tests of substitution saturation for *rbcL* (a), *psbA* (b) and COI (c) genes.

Analyses were performed on all sites.

(a)

<table>
<thead>
<tr>
<th>N OTU</th>
<th>Iss</th>
<th>Iss.cSym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
<th>Iss.cAsym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.266</td>
<td>0.791</td>
<td>23.896</td>
<td>397</td>
<td>0.000</td>
<td>0.758</td>
<td>22.400</td>
<td>397</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.273</td>
<td>0.746</td>
<td>20.824</td>
<td>397</td>
<td>0.000</td>
<td>0.634</td>
<td>15.912</td>
<td>397</td>
<td>0.0000</td>
</tr>
<tr>
<td>16</td>
<td>0.271</td>
<td>0.709</td>
<td>19.567</td>
<td>397</td>
<td>0.000</td>
<td>0.500</td>
<td>10.218</td>
<td>397</td>
<td>0.0000</td>
</tr>
<tr>
<td>32</td>
<td>0.279</td>
<td>0.695</td>
<td>18.871</td>
<td>397</td>
<td>0.000</td>
<td>0.368</td>
<td>4.007</td>
<td>397</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

Note: two-tailed tests are used.

(b)

<table>
<thead>
<tr>
<th>N OTU</th>
<th>Iss</th>
<th>Iss.cSym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
<th>Iss.cAsym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.171</td>
<td>0.781</td>
<td>24.153</td>
<td>246</td>
<td>0.0000</td>
<td>0.756</td>
<td>23.139</td>
<td>246</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.186</td>
<td>0.734</td>
<td>18.902</td>
<td>246</td>
<td>0.0000</td>
<td>0.626</td>
<td>15.164</td>
<td>246</td>
<td>0.0000</td>
</tr>
<tr>
<td>16</td>
<td>0.187</td>
<td>0.680</td>
<td>16.850</td>
<td>246</td>
<td>0.0000</td>
<td>0.474</td>
<td>9.808</td>
<td>246</td>
<td>0.0000</td>
</tr>
<tr>
<td>32</td>
<td>0.203</td>
<td>0.683</td>
<td>15.443</td>
<td>246</td>
<td>0.0000</td>
<td>0.354</td>
<td>4.851</td>
<td>246</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Note: two-tailed tests are used.

(c)

<table>
<thead>
<tr>
<th>N OTU</th>
<th>Iss</th>
<th>Iss.cSym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
<th>Iss.cAsym</th>
<th>T</th>
<th>DF</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.598</td>
<td>0.785</td>
<td>4.129</td>
<td>130</td>
<td>0.0001</td>
<td>0.801</td>
<td>4.479</td>
<td>130</td>
<td>0.0000</td>
</tr>
<tr>
<td>8</td>
<td>0.594</td>
<td>0.757</td>
<td>3.751</td>
<td>130</td>
<td>0.0003</td>
<td>0.681</td>
<td>1.987</td>
<td>130</td>
<td>0.0491</td>
</tr>
<tr>
<td>16</td>
<td>0.602</td>
<td>0.610</td>
<td>0.197</td>
<td>130</td>
<td>0.8438</td>
<td>0.459</td>
<td>3.447</td>
<td>130</td>
<td>0.0008</td>
</tr>
<tr>
<td>32</td>
<td>0.597</td>
<td>0.735</td>
<td>3.504</td>
<td>130</td>
<td>0.0006</td>
<td>0.460</td>
<td>3.462</td>
<td>130</td>
<td>0.0007</td>
</tr>
</tbody>
</table>

Note: two-tailed tests are used.
Table A2.3. Chi-squared test of homogeneity of state frequencies across taxa.

<table>
<thead>
<tr>
<th>Taxon</th>
<th></th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. elegans Na25A3</td>
<td>O</td>
<td>894.00</td>
<td>635.00</td>
<td>781.00</td>
<td>982.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>883.39</td>
<td>625.82</td>
<td>777.39</td>
<td>1005.40</td>
</tr>
<tr>
<td>B. furcatum 2 Na8A3</td>
<td>O</td>
<td>1239.00</td>
<td>863.00</td>
<td>1059.00</td>
<td>1374.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1216.94</td>
<td>862.11</td>
<td>1070.92</td>
<td>1385.03</td>
</tr>
<tr>
<td>B. hyalinum Na10B1</td>
<td>O</td>
<td>769.00</td>
<td>507.00</td>
<td>716.00</td>
<td>845.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>761.29</td>
<td>539.32</td>
<td>669.95</td>
<td>866.44</td>
</tr>
<tr>
<td>B. jadranum Na19C1</td>
<td>O</td>
<td>852.00</td>
<td>604.00</td>
<td>754.00</td>
<td>938.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>844.74</td>
<td>598.44</td>
<td>743.39</td>
<td>961.42</td>
</tr>
<tr>
<td>B. jadranum Na19C3</td>
<td>O</td>
<td>1005.00</td>
<td>633.00</td>
<td>851.00</td>
<td>1043.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>947.79</td>
<td>671.44</td>
<td>834.07</td>
<td>1078.70</td>
</tr>
<tr>
<td>B. mediterraneum Na1C4</td>
<td>O</td>
<td>763.00</td>
<td>512.00</td>
<td>699.00</td>
<td>843.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>755.92</td>
<td>535.52</td>
<td>665.22</td>
<td>860.33</td>
</tr>
<tr>
<td>B. mediterraneum Na29B3</td>
<td>O</td>
<td>1366.00</td>
<td>933.00</td>
<td>1153.00</td>
<td>1560.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1344.94</td>
<td>952.79</td>
<td>1183.57</td>
<td>1530.71</td>
</tr>
<tr>
<td>B. parallelum newLA2</td>
<td>O</td>
<td>772.00</td>
<td>512.00</td>
<td>700.00</td>
<td>838.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>757.26</td>
<td>536.47</td>
<td>666.40</td>
<td>861.86</td>
</tr>
<tr>
<td>C. affinis Na49A2</td>
<td>O</td>
<td>668.00</td>
<td>510.00</td>
<td>636.00</td>
<td>761.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>690.98</td>
<td>489.51</td>
<td>608.08</td>
<td>786.43</td>
</tr>
<tr>
<td>C. anastomosans Na14C2</td>
<td>O</td>
<td>1331.00</td>
<td>942.00</td>
<td>1147.00</td>
<td>1528.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1327.76</td>
<td>940.63</td>
<td>1168.45</td>
<td>1511.16</td>
</tr>
<tr>
<td>C. anastomosans Na14C3</td>
<td>O</td>
<td>1348.00</td>
<td>954.00</td>
<td>1160.00</td>
<td>1546.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1343.86</td>
<td>952.03</td>
<td>1182.62</td>
<td>1529.48</td>
</tr>
<tr>
<td>C. brevis 1 Na7B1</td>
<td>O</td>
<td>1344.00</td>
<td>974.00</td>
<td>1210.00</td>
<td>1570.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1368.01</td>
<td>969.14</td>
<td>1203.87</td>
<td>1556.97</td>
</tr>
<tr>
<td>C. brevis 2 Na7C2</td>
<td>O</td>
<td>848.00</td>
<td>674.00</td>
<td>845.00</td>
<td>957.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>891.97</td>
<td>631.90</td>
<td>784.95</td>
<td>1015.18</td>
</tr>
<tr>
<td>C. brevis 3 Ch9B3</td>
<td>O</td>
<td>644.00</td>
<td>476.00</td>
<td>666.00</td>
<td>667.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>658.25</td>
<td>466.32</td>
<td>579.27</td>
<td>749.17</td>
</tr>
<tr>
<td>C. cf. convolutus Ch5C4</td>
<td>O</td>
<td>1382.00</td>
<td>931.00</td>
<td>1199.00</td>
<td>1573.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1364.53</td>
<td>966.67</td>
<td>1200.80</td>
<td>1553.00</td>
</tr>
<tr>
<td>C. cf. convolutus L7-B6</td>
<td>O</td>
<td>836.00</td>
<td>532.00</td>
<td>719.00</td>
<td>923.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>807.71</td>
<td>572.21</td>
<td>710.80</td>
<td>919.28</td>
</tr>
<tr>
<td>C. cf. decipiens El6B1</td>
<td>O</td>
<td>1013.00</td>
<td>629.00</td>
<td>887.00</td>
<td>1072.00</td>
</tr>
<tr>
<td>Species</td>
<td>E</td>
<td>O</td>
<td>684.56</td>
<td>809.00</td>
<td>995.00</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----</td>
<td>-----</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>C. cf. lorenzianus El1A4</td>
<td>O</td>
<td>910.00</td>
<td>656.00</td>
<td>809.00</td>
<td>1099.77</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>904.32</td>
<td>640.64</td>
<td>795.81</td>
<td>1029.23</td>
</tr>
<tr>
<td>C. cf. pseudodichaeta El1C1</td>
<td>O</td>
<td>1153.00</td>
<td>744.00</td>
<td>989.00</td>
<td>1273.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1116.04</td>
<td>790.64</td>
<td>982.13</td>
<td>1270.19</td>
</tr>
<tr>
<td>C. cf. tortissimus Na18C4</td>
<td>O</td>
<td>1349.00</td>
<td>940.00</td>
<td>1163.00</td>
<td>1555.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1343.59</td>
<td>951.84</td>
<td>1182.38</td>
<td>1529.18</td>
</tr>
<tr>
<td>C. cf. vixvisibilis Na25C3</td>
<td>O</td>
<td>1023.00</td>
<td>718.00</td>
<td>878.00</td>
<td>1167.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1015.95</td>
<td>719.73</td>
<td>894.05</td>
<td>1156.27</td>
</tr>
<tr>
<td>C. cf. vixvisibilis Na28A4</td>
<td>O</td>
<td>1172.00</td>
<td>840.00</td>
<td>1024.00</td>
<td>1286.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1159.78</td>
<td>821.62</td>
<td>1020.62</td>
<td>1319.97</td>
</tr>
<tr>
<td>C. cinetus Ch3C4</td>
<td>O</td>
<td>1252.00</td>
<td>885.00</td>
<td>1096.00</td>
<td>1397.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1242.43</td>
<td>880.17</td>
<td>1093.36</td>
<td>1414.04</td>
</tr>
<tr>
<td>C. circinalis Na15C2</td>
<td>O</td>
<td>1259.00</td>
<td>902.00</td>
<td>1120.00</td>
<td>1411.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1259.07</td>
<td>891.96</td>
<td>1108.00</td>
<td>1432.97</td>
</tr>
<tr>
<td>C. constrictus Ch12C1</td>
<td>O</td>
<td>1357.00</td>
<td>975.00</td>
<td>1202.00</td>
<td>1562.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1367.48</td>
<td>968.76</td>
<td>1203.40</td>
<td>1556.36</td>
</tr>
<tr>
<td>C. contortus cf. var. contortus Na31B2</td>
<td>O</td>
<td>1249.00</td>
<td>928.00</td>
<td>1109.00</td>
<td>1402.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1257.99</td>
<td>891.20</td>
<td>1107.05</td>
<td>1431.75</td>
</tr>
<tr>
<td>C. contortus Ch12A4</td>
<td>O</td>
<td>1353.00</td>
<td>986.00</td>
<td>1189.00</td>
<td>1565.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1366.67</td>
<td>968.19</td>
<td>1202.69</td>
<td>1555.44</td>
</tr>
<tr>
<td>C. costatus Na1A3</td>
<td>O</td>
<td>1347.00</td>
<td>969.00</td>
<td>1184.00</td>
<td>1579.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1362.92</td>
<td>965.53</td>
<td>1199.39</td>
<td>1551.17</td>
</tr>
<tr>
<td>C. costatus Ro1B1</td>
<td>O</td>
<td>1248.00</td>
<td>912.00</td>
<td>1104.00</td>
<td>1410.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1254.24</td>
<td>888.54</td>
<td>1103.75</td>
<td>1427.48</td>
</tr>
<tr>
<td>C. costatus Ro2A2</td>
<td>O</td>
<td>1247.00</td>
<td>900.00</td>
<td>1114.00</td>
<td>1415.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1254.77</td>
<td>888.92</td>
<td>1104.22</td>
<td>1428.09</td>
</tr>
<tr>
<td>C. curvisetus 1 Na10C1</td>
<td>O</td>
<td>968.00</td>
<td>710.00</td>
<td>894.00</td>
<td>1137.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>995.29</td>
<td>705.09</td>
<td>875.87</td>
<td>1132.76</td>
</tr>
<tr>
<td>C. curvisetus 1 Ro3B2</td>
<td>O</td>
<td>1351.00</td>
<td>963.00</td>
<td>1186.00</td>
<td>1565.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1359.16</td>
<td>962.87</td>
<td>1196.08</td>
<td>1546.89</td>
</tr>
<tr>
<td>C. curvisetus 2a Na1C1</td>
<td>O</td>
<td>1317.00</td>
<td>945.00</td>
<td>1166.00</td>
<td>1560.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1338.50</td>
<td>948.23</td>
<td>1177.90</td>
<td>1523.38</td>
</tr>
<tr>
<td>C. curvisetus 2b Ch5B1</td>
<td>O</td>
<td>1351.00</td>
<td>966.00</td>
<td>1181.00</td>
<td>1581.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1362.92</td>
<td>965.53</td>
<td>1199.39</td>
<td>1551.17</td>
</tr>
<tr>
<td>----------------</td>
<td>-------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 2c El6A2</td>
<td>O</td>
<td>775.00</td>
<td>506.00</td>
<td>729.00</td>
<td>845.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>766.12</td>
<td>542.74</td>
<td>674.20</td>
<td>871.94</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 3 El4A2</td>
<td>O</td>
<td>901.00</td>
<td>644.00</td>
<td>791.00</td>
<td>1019.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>900.29</td>
<td>637.79</td>
<td>792.27</td>
<td>1024.64</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 3 Na3C4</td>
<td>O</td>
<td>1153.00</td>
<td>744.00</td>
<td>959.00</td>
<td>1287.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1111.75</td>
<td>787.59</td>
<td>978.35</td>
<td>1265.31</td>
</tr>
<tr>
<td><em>C. danicus</em> Na9B4</td>
<td>O</td>
<td>1376.00</td>
<td>952.00</td>
<td>1147.00</td>
<td>1593.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1359.96</td>
<td>963.44</td>
<td>1196.79</td>
<td>1547.81</td>
</tr>
<tr>
<td><em>C. debilis</em> 2 MM24-A3</td>
<td>O</td>
<td>1377.00</td>
<td>952.00</td>
<td>1146.00</td>
<td>1593.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1359.96</td>
<td>963.44</td>
<td>1196.79</td>
<td>1547.81</td>
</tr>
<tr>
<td><em>C. debilis</em> 2 MM24-C3</td>
<td>O</td>
<td>1334.00</td>
<td>943.00</td>
<td>1162.00</td>
<td>1562.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1341.98</td>
<td>950.70</td>
<td>1180.97</td>
<td>1527.35</td>
</tr>
<tr>
<td><em>C. debilis</em> 3 Ch1A1</td>
<td>O</td>
<td>969.00</td>
<td>732.00</td>
<td>912.00</td>
<td>1112.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>999.58</td>
<td>708.13</td>
<td>879.65</td>
<td>1137.64</td>
</tr>
<tr>
<td><em>C. debilis</em> 3 Ch9A3</td>
<td>O</td>
<td>965.00</td>
<td>728.00</td>
<td>913.00</td>
<td>1119.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>999.58</td>
<td>708.13</td>
<td>879.65</td>
<td>1137.64</td>
</tr>
<tr>
<td><em>C. decipiens</em> Na28A2</td>
<td>O</td>
<td>1349.00</td>
<td>983.00</td>
<td>1204.00</td>
<td>1560.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1367.48</td>
<td>968.76</td>
<td>1203.40</td>
<td>1556.36</td>
</tr>
<tr>
<td><em>C. diadema</em> 1 Ch4A1</td>
<td>O</td>
<td>1372.00</td>
<td>968.00</td>
<td>1170.00</td>
<td>1587.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1367.75</td>
<td>968.95</td>
<td>1203.64</td>
<td>1556.66</td>
</tr>
<tr>
<td><em>C. diadema</em> 1 Na13B1</td>
<td>O</td>
<td>897.00</td>
<td>662.00</td>
<td>819.00</td>
<td>1005.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>907.81</td>
<td>643.12</td>
<td>798.88</td>
<td>1033.20</td>
</tr>
<tr>
<td><em>C. diadema</em> 2 Ch5C1</td>
<td>O</td>
<td>898.00</td>
<td>657.00</td>
<td>821.00</td>
<td>1009.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>908.34</td>
<td>643.50</td>
<td>799.36</td>
<td>1033.81</td>
</tr>
<tr>
<td><em>C. dichatoensis</em> Ch1B3</td>
<td>O</td>
<td>1239.00</td>
<td>897.00</td>
<td>1113.00</td>
<td>1383.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1242.97</td>
<td>880.55</td>
<td>1093.83</td>
<td>1414.65</td>
</tr>
<tr>
<td><em>C. didymus</em> 1 Ch6A3</td>
<td>O</td>
<td>1240.00</td>
<td>895.00</td>
<td>1111.00</td>
<td>1378.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1240.82</td>
<td>879.03</td>
<td>1091.94</td>
<td>1412.21</td>
</tr>
<tr>
<td><em>C. didymus</em> 2 Na20B4</td>
<td>O</td>
<td>1337.00</td>
<td>957.00</td>
<td>1188.00</td>
<td>1547.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1349.50</td>
<td>956.02</td>
<td>1187.58</td>
<td>1535.90</td>
</tr>
<tr>
<td><em>C. diversus</em> 1 Na23B1</td>
<td>O</td>
<td>1356.00</td>
<td>951.00</td>
<td>1207.00</td>
<td>1585.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1368.28</td>
<td>969.33</td>
<td>1204.11</td>
<td>1557.28</td>
</tr>
<tr>
<td><em>C. diversus</em> 1 Na3C1</td>
<td>O</td>
<td>1258.00</td>
<td>881.00</td>
<td>1127.00</td>
<td>1425.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1258.80</td>
<td>891.77</td>
<td>1107.76</td>
<td>1432.67</td>
</tr>
<tr>
<td><em>C. diversus</em> 1 Na50B2</td>
<td>O</td>
<td>1262.00</td>
<td>884.00</td>
<td>1081.00</td>
<td>1411.00</td>
</tr>
<tr>
<td>Species</td>
<td>Strain</td>
<td>Mass Loss 1</td>
<td>Mass Loss 2</td>
<td>Mass Loss 3</td>
<td>Mass Loss 4</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>C. diversus 1 Na5B2</td>
<td>O</td>
<td>943.00</td>
<td>738.00</td>
<td>951.00</td>
<td>1151.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1015.14</td>
<td>719.16</td>
<td>893.34</td>
<td>1155.36</td>
</tr>
<tr>
<td>C. eibenii Ch8C3</td>
<td>O</td>
<td>944.00</td>
<td>738.00</td>
<td>949.00</td>
<td>1152.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1015.14</td>
<td>719.16</td>
<td>893.34</td>
<td>1155.36</td>
</tr>
<tr>
<td>C. eibenii Ro1B2</td>
<td>O</td>
<td>1229.00</td>
<td>915.00</td>
<td>1143.00</td>
<td>1399.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1257.46</td>
<td>890.82</td>
<td>1106.58</td>
<td>1431.14</td>
</tr>
<tr>
<td>C. elegans Ch12A1</td>
<td>O</td>
<td>967.00</td>
<td>723.00</td>
<td>905.00</td>
<td>1140.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1002.26</td>
<td>710.03</td>
<td>882.01</td>
<td>1140.70</td>
</tr>
<tr>
<td>C. lauderi Na13A4</td>
<td>O</td>
<td>1348.00</td>
<td>950.00</td>
<td>1182.00</td>
<td>1566.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1354.06</td>
<td>959.26</td>
<td>1191.59</td>
<td>1541.09</td>
</tr>
<tr>
<td>C. lauderi Na2A1</td>
<td>O</td>
<td>1231.00</td>
<td>869.00</td>
<td>1080.00</td>
<td>1398.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1228.48</td>
<td>870.29</td>
<td>1081.08</td>
<td>1398.16</td>
</tr>
<tr>
<td>C. lauderi Na34C3</td>
<td>O</td>
<td>1229.00</td>
<td>865.00</td>
<td>1078.00</td>
<td>1396.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1225.79</td>
<td>868.39</td>
<td>1078.72</td>
<td>1395.10</td>
</tr>
<tr>
<td>C. lauderi Na36A1</td>
<td>O</td>
<td>1328.00</td>
<td>933.00</td>
<td>1155.00</td>
<td>1567.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1337.15</td>
<td>947.28</td>
<td>1176.72</td>
<td>1521.85</td>
</tr>
<tr>
<td>C. lorenzianus 1 Ch11C1</td>
<td>O</td>
<td>948.00</td>
<td>706.00</td>
<td>884.00</td>
<td>1134.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>985.36</td>
<td>698.06</td>
<td>867.13</td>
<td>1121.46</td>
</tr>
<tr>
<td>C. lorenzianus 1 Ch4C3</td>
<td>O</td>
<td>846.00</td>
<td>545.00</td>
<td>715.00</td>
<td>898.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>806.10</td>
<td>571.07</td>
<td>709.38</td>
<td>917.45</td>
</tr>
<tr>
<td>C. lorenzianus 2 Ch11A1</td>
<td>O</td>
<td>1050.00</td>
<td>680.00</td>
<td>935.00</td>
<td>1104.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1011.39</td>
<td>716.50</td>
<td>890.04</td>
<td>1151.08</td>
</tr>
<tr>
<td>C. lorenzianus 2 Ch13B4</td>
<td>O</td>
<td>1372.00</td>
<td>957.00</td>
<td>1181.00</td>
<td>1584.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1366.94</td>
<td>968.38</td>
<td>1202.93</td>
<td>1555.75</td>
</tr>
<tr>
<td>C. lorenzianus 2 Ch4A3</td>
<td>O</td>
<td>1343.00</td>
<td>974.00</td>
<td>1192.00</td>
<td>1568.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1362.38</td>
<td>965.15</td>
<td>1198.91</td>
<td>1550.56</td>
</tr>
<tr>
<td>C. lorenzianus 2 Ch4C4</td>
<td>O</td>
<td>1368.00</td>
<td>972.00</td>
<td>1157.00</td>
<td>1592.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1365.60</td>
<td>967.43</td>
<td>1201.75</td>
<td>1554.22</td>
</tr>
<tr>
<td>C. peruvianus 1 newEA1</td>
<td>O</td>
<td>1368.00</td>
<td>974.00</td>
<td>1161.00</td>
<td>1592.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1367.21</td>
<td>968.57</td>
<td>1203.17</td>
<td>1556.05</td>
</tr>
<tr>
<td>C. peruvianus 2 Ch11B4</td>
<td>O</td>
<td>1036.00</td>
<td>669.00</td>
<td>933.00</td>
<td>1115.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1007.09</td>
<td>713.45</td>
<td>886.26</td>
<td>1146.20</td>
</tr>
<tr>
<td>C. protuberans Ch8C2</td>
<td>O</td>
<td>1039.00</td>
<td>667.00</td>
<td>933.00</td>
<td>1116.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1007.63</td>
<td>713.83</td>
<td>886.73</td>
<td>1146.81</td>
</tr>
<tr>
<td>C. pseudocurvisetus Na13C4</td>
<td>O</td>
<td>1308.00</td>
<td>939.00</td>
<td>1161.00</td>
<td>1503.00</td>
</tr>
<tr>
<td>Species</td>
<td>Clade</td>
<td>O</td>
<td>E</td>
<td>933.59</td>
<td>1159.71</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td><em>C. radicans</em> Ch10A3</td>
<td></td>
<td></td>
<td></td>
<td>1317.83</td>
<td>1313.00</td>
</tr>
<tr>
<td><em>C. radicans</em> Ch11A4</td>
<td></td>
<td></td>
<td></td>
<td>1321.05</td>
<td>1354.00</td>
</tr>
<tr>
<td><em>C. rostratus</em> Na1C3</td>
<td></td>
<td></td>
<td></td>
<td>862.00</td>
<td>867.29</td>
</tr>
<tr>
<td><em>C. rostratus</em> newDA3</td>
<td></td>
<td></td>
<td></td>
<td>1168.00</td>
<td>1161.39</td>
</tr>
<tr>
<td><em>C. rostratus</em> Na22B1</td>
<td></td>
<td></td>
<td></td>
<td>1247.00</td>
<td>1257.72</td>
</tr>
<tr>
<td><em>C. rostratus</em> Na23A1</td>
<td></td>
<td></td>
<td></td>
<td>1207.00</td>
<td>1218.28</td>
</tr>
<tr>
<td><em>C. socialis</em> Na33B1</td>
<td></td>
<td></td>
<td></td>
<td>1228.00</td>
<td>1241.09</td>
</tr>
<tr>
<td><em>C. sp. Clade Na11C3 - Na11C3</em></td>
<td></td>
<td></td>
<td></td>
<td>977.00</td>
<td>986.43</td>
</tr>
<tr>
<td><em>C. sp. Clade Na11C3 - Na43A1</em></td>
<td></td>
<td></td>
<td></td>
<td>1352.00</td>
<td>1330.45</td>
</tr>
<tr>
<td><em>C. sp. Clade Na12A3 - Na9A3</em></td>
<td></td>
<td></td>
<td></td>
<td>534.00</td>
<td>540.71</td>
</tr>
<tr>
<td><em>C. sp. Clade Na13C2 - Na13C2</em></td>
<td></td>
<td></td>
<td></td>
<td>636.00</td>
<td>648.59</td>
</tr>
<tr>
<td><em>C. sp. Clade Na17B2 - Na17B2</em></td>
<td></td>
<td></td>
<td></td>
<td>1271.00</td>
<td>1259.07</td>
</tr>
<tr>
<td><em>C. sp. Clade Na26B1 - Na26B1</em></td>
<td></td>
<td></td>
<td></td>
<td>1048.00</td>
<td>1010.31</td>
</tr>
<tr>
<td><em>C. sp. Clade Na28A1 - Na28A1</em></td>
<td></td>
<td></td>
<td></td>
<td>1262.00</td>
<td>1253.16</td>
</tr>
<tr>
<td><em>C. sp. Clade Va7D2 - Na43A4</em></td>
<td></td>
<td></td>
<td></td>
<td>1369.00</td>
<td>1361.84</td>
</tr>
<tr>
<td><em>C. sp. Clade Va7D2 - Na44B3</em></td>
<td></td>
<td></td>
<td></td>
<td>1224.00</td>
<td>1243.50</td>
</tr>
<tr>
<td><em>C. sporotrichatus</em> Ch2A4</td>
<td></td>
<td></td>
<td></td>
<td>1238.00</td>
<td>1257.99</td>
</tr>
<tr>
<td><em>C. sporotrichatus</em> Ch9C4</td>
<td></td>
<td></td>
<td></td>
<td>1308.00</td>
<td>1308.00</td>
</tr>
<tr>
<td>Species</td>
<td>E</td>
<td>1297.17</td>
<td>918.95</td>
<td>1141.53</td>
<td>1476.34</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td><em>C. tenuissimus</em> Na26A1</td>
<td>O</td>
<td>1389.00</td>
<td>959.00</td>
<td>1157.00</td>
<td>1577.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1363.72</td>
<td>966.10</td>
<td>1200.10</td>
<td>1552.08</td>
</tr>
<tr>
<td><em>C. tenuissimus</em> Na36B4</td>
<td>O</td>
<td>1043.00</td>
<td>740.00</td>
<td>871.00</td>
<td>1226.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1041.17</td>
<td>737.60</td>
<td>916.25</td>
<td>1184.98</td>
</tr>
<tr>
<td><em>C. tenuissimus</em> Na44A1</td>
<td>O</td>
<td>833.00</td>
<td>527.00</td>
<td>697.00</td>
<td>846.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>779.00</td>
<td>551.87</td>
<td>685.53</td>
<td>886.60</td>
</tr>
<tr>
<td><em>C. teres</em> Ch13B2</td>
<td>O</td>
<td>1048.00</td>
<td>707.00</td>
<td>839.00</td>
<td>1173.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1010.85</td>
<td>716.12</td>
<td>889.56</td>
<td>1150.47</td>
</tr>
<tr>
<td><em>C. teres</em> Ch5A1</td>
<td>O</td>
<td>1031.00</td>
<td>740.00</td>
<td>870.00</td>
<td>1187.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>1027.22</td>
<td>727.71</td>
<td>903.97</td>
<td>1169.10</td>
</tr>
<tr>
<td><em>C. throndsenii</em> Na45B3</td>
<td>O</td>
<td>889.00</td>
<td>630.00</td>
<td>743.00</td>
<td>1093.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>900.29</td>
<td>637.79</td>
<td>792.27</td>
<td>1024.64</td>
</tr>
<tr>
<td><em>C. tortissimus</em> Na25A2</td>
<td>O</td>
<td>798.00</td>
<td>593.00</td>
<td>674.00</td>
<td>930.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>803.69</td>
<td>569.36</td>
<td>707.26</td>
<td>914.70</td>
</tr>
<tr>
<td><em>Detonula confervacea</em> CCMP353</td>
<td>O</td>
<td>800.00</td>
<td>580.00</td>
<td>682.00</td>
<td>914.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>798.59</td>
<td>565.74</td>
<td>702.77</td>
<td>908.89</td>
</tr>
<tr>
<td><em>Hydrosera</em> sp. CYTX025</td>
<td>O</td>
<td>707.00</td>
<td>490.00</td>
<td>542.00</td>
<td>904.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>709.23</td>
<td>502.44</td>
<td>624.13</td>
<td>807.19</td>
</tr>
<tr>
<td><em>Terpsinoe musica</em> NHOP43</td>
<td>O</td>
<td>602.00</td>
<td>443.00</td>
<td>455.00</td>
<td>738.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>600.55</td>
<td>425.45</td>
<td>528.50</td>
<td>683.50</td>
</tr>
<tr>
<td><em>Thalassiosira pseudonana</em> CCMP1335</td>
<td>O</td>
<td>602.00</td>
<td>445.00</td>
<td>459.00</td>
<td>732.00</td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>600.55</td>
<td>425.45</td>
<td>528.50</td>
<td>683.50</td>
</tr>
</tbody>
</table>

Chi-square = 316.520842 (df=297), P = 0.20860911

Warning: This test ignores correlation due to phylogenetic structure.
Table A2.4. Traditional classification scheme for the family Chaetocerotaceae. Only sections including taxa utilised in the present study are shown. “References for description” refers to publications in which the section is described or amended. “Morphologically assigned species” refers to taxa assigned to the sections using information in Gaonkar et al. (2018) and references therein).

<table>
<thead>
<tr>
<th>Genus</th>
<th>Bacteriastrum Shadbolt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Section</td>
</tr>
<tr>
<td>Isomorpha Pavillard</td>
<td><strong>Description</strong>: terminal setae of like construction and form on both ends of chain (isomorphic). Setae on both ends directed either outward from chain axis or toward the center. The two outer valves are therefore mirror images. <strong>References for description</strong>: Pavillard (1924); (1925); Cupp (1943). <strong>Morphologically assigned species</strong>: <em>B. hyalinum</em>, <em>B. jadranum</em>.</td>
</tr>
<tr>
<td>Sagittata Pavillard</td>
<td><strong>Description</strong>: terminal setae on either end of chain different in form and direction (dimorphic). Setae of posterior valve directed outward from chain and running nearly parallel to chain axis, forming a bell-shaped space. Setae of other or anterior valve curved toward inner part of chain, or on their ends turned back toward the outside or in general deviating little from the valvar plane. <strong>References for description</strong>: Pavillard (1924); (1925); Cupp (1943). <strong>Morphologically assigned species</strong>: <em>B. elegans</em>, <em>B. furcatum</em>, <em>B. mediterraneum</em>, <em>B. parallelum</em>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Genus</th>
<th>Chaetoceros Ehrenberg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Subgenus <em>Chaetoceros (Phaeoceros)</em> Gran</td>
</tr>
<tr>
<td></td>
<td>Section</td>
</tr>
<tr>
<td>Borealia Ostenfeld</td>
<td><strong>Description</strong>: setae diverging in all directions; the directions of the setae of the one valve are often different from those of the other valve; the external process of the rimoportula in the centre of the valve absent. Apertures narrow. <strong>References for description</strong>: Ostenfeld (1903); Cupp (1943). <strong>Morphologically assigned species</strong>: <em>C. cf. convolutus</em>, <em>C. danicus</em>, <em>C. eibenii</em>.</td>
</tr>
<tr>
<td>Peruviana Hernández-Becerril</td>
<td><strong>Description</strong>: cells solitary or in chains, heterovalvar. All setae robust, pointed towards the same end. Rimoportula present in every valve, excentrically placed. <strong>Reference for description</strong>: Hernández-Becerril (1996). <strong>Morphologically assigned species</strong>: <em>C. peruvianus</em> 1-2.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subgenus</th>
<th>Hyalochaete Gran</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section</td>
<td></td>
</tr>
</tbody>
</table>
| **Anastomosantia** Ostenfeld | **Description**: setae united by a bridge. Chains mostly loose.  
**References for description**: Ostenfeld (1903); Hernández-Becerril (1996).  
**Morphologically assigned species**: *C. anastomosantia*. |
| **Compressa** Ostenfeld; emended by Yang Li and Lundholm (in Xu et al., 2019) | **Description**: valves broadly elliptical to compress. Numerous small chloroplasts in each cell. Apertures usually moderately large. Terminal setae little different from others. Intercalary setae of two types; thin, common setae and heavy special setae. Heavy setae contorted with spiralling rows of spines and poroids, or heavy setae not visually contorted lacking rows of spines and poroids. Resting spores smooth or with a row of spicules.  
**References for description**: Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996); Xu et al. (2019).  
**Morphologically assigned species**: *C. cf. var. contortus, C. contortus*. |
| **Constricta** Ostenfeld | **Description**: cells with one or two chloroplasts and a marked constriction at the base of the valve mantle. Girdle at least one-third the length of the cell. Terminal setae mostly thicker than the others. Resting spores, when present, about the middle of the cell with numerous spines on both valves.  
**References for description**: Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996); Gaonkar et al. (2018).  
**Morphologically assigned species**: *C. constrictus*. |
| **Curviseta** Ostenfeld; emended by Gran | **Description**: chains usually curved, with setae all bent in one direction without special end cells. One chloroplast.  
**References for description**: Ostenfeld (1903); Gran (1905); Cupp (1943); Hernández-Becerril (1996).  
| **Cylindrica** Ostenfeld | **Description**: cells with valves nearly circular (cylindrical). Apertures very narrow. Small, numerous chloroplasts. Terminal setae not thicker than others. Resting spores about middle of the cell, smooth or with spines.  
**References for description**: Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species**: *C. lauderi, C. teres*. |
| **Diadema** (Ehrenberg) Ostenfeld; emended by Gran | **Description**: one chloroplast per cell. Chains long with conspicuous terminal setae. Primary valve of resting spores with branched processes or crown of spines, or sometimes smooth.  
**References for description**: Ostenfeld (1903); Gran (1905); Cupp (1943).  
**Morphologically assigned species**: *C. diadema 1-2, C. rotoporus*. |
| **Dicladia** (Ehrenberg) Gran; emended by Lebour | **Description**: multiple chloroplasts per cell and setae with large pores. Terminal and intercalary setae similar. Resting spores, when known, with two horns armed with small branches on primary valves.  
**References for description**: Gran (1905); Lebour (1930); Cupp (1943); Hernández-Becerril (1996); Gaonkar et al. (2018).  
**Morphologically assigned species**: *C. cf. decipiens, C. cf. lorenzianus, C. decipiens, C. elegans, C. lorenzianus 1-2*. |
| **Diversa** Ostenfeld | **Description**: one chloroplast per cell. Short rigid chains. Inner setae of two kinds. Terminal setae less spread out than a special pair of setae in middle of cell.  
**References for description**: Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species**: *C. diversus*. |
| **Furcellata** Ostenfeld | **Description**: chains generally loose, without differentiated terminal setae. One chloroplast. Resting cells eccentrically arranged in mother cell, lying close together two and two, with thick coalesced setae; with smooth valves or with short spines.  
**References for description**: Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species**: *C. cinctus, C. radicans*. |
| **Laciniosa Ostenfeld** | **Description:** one or two chloroplasts per cell. Girdle rather long. Aperture large. Terminal setae usually thicker than the others, not diverging greatly. Resting spores smooth or with minute spines on primary valve, not in the middle of the cell.  
**References for description:** Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species:** *C. brevis* 1-2-3. |
| **Protuberantia Ostenfeld; emended by Hernández-Becerril** | **Description:** two chloroplasts per cell, each with a large pyrenoid situated in a protuberance in the middle of the valve surface. Valves with poroids. Resting spores paired with two long setae or free without setae.  
**References for description:** Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996); Gaonkar et al. (2018).  
**Morphologically assigned species:** *C. didymus* 1-2, *C. protuberans*. |
| **Simplicia Ostenfeld** | **Description:** cells small and fragile, generally single or two or three together. In case of chain formation, there is no differentiation of terminal setae.  
**References for description:** Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species:** *C. tenuissimus*. |
| **Socialia Ostenfeld** | **Description:** chains irregular and curved embedded in mucilage, forming irregularly spherical colonies. One chloroplast. Resting spores smooth or with small spines.  
**References for description:** Ostenfeld (1903); Cupp (1943); Hernández-Becerril (1996).  
**Morphologically assigned species:** *C. dichatoensis*, *C. socialis*, *C. sporotrustcatus*. |
| **Stenocincta Ostenfeld** | **Description:** a single chloroplast per cell. Usually narrow aperture. Terminal setae curved, thicker than other setae.  
**References for description:** Ostenfeld (1903); Hernández-Becerril (1996); Gaonkar et al. (2018).  
Chapter III

Assessing diversity and distribution in Chaetoceros: integration of classical and novel strategies
The material presented in this chapter has been published as Research Article:


The paper has been published under the CC BY 4.0 licence. Therefore, I have all the rights to re-use the material contained in it (text, figures and tables).
3.1. Introduction

3.1.1. Primary Biodiversity Data: recording the occurrence of species

Primary Biodiversity Data can be defined as the basic attributes of observations or records of the occurrence of species (Anderson et al., 2016). For centuries, primary species-occurrence data were mostly obtained from taxonomic descriptions of specimens stored in museums, herbaria and private collections (Chapman, 2005). In the last few years, biological recording has evolved, particularly due to the involvement of citizens and the application of molecular tools (Isaac and Pocock, 2015; Pocock et al., 2015). Indeed, nowadays data are also gathered through satellite tracking and direct or remote observation (Croxall et al., 1993), frozen tissue collections and seed banks (Chapman, 2005), environmental DNA (August et al., 2015), and citizen science initiatives (Devictor et al., 2010; Hochachka et al., 2012).

Regardless their sources, data for biological recording are generally presence-only records (opportunistic incidence records, Peterson et al., 2011) since they do not report any info about the absence of the species into a particular area at the time of the survey. Furthermore, they are subject to bias in space and time, such as uneven sampling due to information gathered in urbanized or easily accessible areas and in suitable weather conditions for citizen science projects (Kéry et al., 2010; Isaac and Pocock, 2015) or a time series data for a small area in the case of checklists.

The uses of primary species-occurrence data in natural sciences are numerous and different, from the monitoring of biodiversity (Soberón and Peterson, 2009) and invasive species (Zanetos et al., 2005) to the identification and management of marine protected areas (Araújo and Williams, 2000) and development of conservation plans (Myers et al., 2000; Rondinini et al., 2006).

Field notes and checklists, such observations from early naturalists, scientific expeditions, and museum records, are among the most traditional data used by biologists to document
past patterns of species’ distribution and abundance (Droege et al., 1998) and due to their source, they are generally highly reliable. These data are generally more biased in space than in time because related to an area chosen for being especially diverse for the taxon of interest and so intensively sampled over time (Prendergast et al., 1993).

The growth of biological records in recent decades led to the establishment of recording schemes and the organization and storing of such data in freely accessible online portals, such as the National Biodiversity Network Gateway (NBN Gateway; http://www.nbn.org.uk/) and GBIF (http://www.gbif.org/) (Isaac and Pocock, 2015; Powney and Isaac, 2015).

In recent years, environmental DNA (eDNA) data, defined as any DNA-containing trace left behind by organisms in the environment, were added to the list of PBDs (August et al. 2015; Lawson Handley, 2015). Despite eDNA metabarcoding can complement and overcome the limitations of conventional methods by targeting different species simultaneously and catching greater diversity, research is still needed to understand the complex spatial and temporal dynamics of the various eDNA types in the environment (Deiner et al., 2017). Among the different approaches for the characterisation of eDNA, DNA metabarcoding revealed useful at assessing species distribution of marine diatoms of the genus *Leptocylindrus* (Nanjappa et al., 2014), whilst metagenomics for the estimation of species abundance, distributions and richness in fungi (Unterreicher et al., 2011).

### 3.1.2. Primary Biodiversity Data for planktonic species

Biodiversity data of planktonic species are traditionally gathered through samples collected over time, either once though opportunity, or many times through long term ecological research (LTER) projects at single sites (e.g. Helgoland Roads, MareChiara; Blanes Bay Microbial Observatory, Hawaii Ocean Time series), or once at each of many sites through expeditions (e.g., Challenger, Plankton-Expedition). A shortcoming of all these sampling
schemes is that they provide incomplete distribution maps of species with many “blank”
regions and seasons for which information is lacking. Furthermore, species distribution
patterns often reflect the distribution of the scientists studying the species, or tracks of
expeditions (Droege et al., 1998). Sampling intensity is often skewed towards areas known
to be diverse for taxa of interest because those areas attract the collectors (Prendergast et
al., 1993). However, despite largely time-biased (i.e. sampling events occurring in single
dates), these data have the advantage of providing information from areas difficult to
access that would otherwise be unexplored (Ji et al., 2013).

Some of the initiatives in the plankton world tried to overcome these issues, such as the Sir
Alister Hardy Foundation for Ocean Science’s (SAHFOS) program of putting plankton
recorders behind ships to sample tracks recurrently (Southward et al., 2005), and the
involvement of the public in citizen science initiatives (Castilla et al., 2015; Busch et al.,
2016). The results are usually available in form of taxonomic monographs, checklists, or
species descriptions.

Among the freely accessible online portals where it is possible to check occurrence data for
planktonic species are the Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) (Isaac and Pocock, 2015; Powney and Isaac, 2015) and the Ocean
Biogeographic Information System (OBIS; http://iobis.org/). GBIF contains occurrence
data for both aquatic and terrestrial species gathered from different sources as natural
history collections, environmental monitoring programmes, recording initiatives and
citizen scientist projects. On the contrary, OBIS only focuses on world’s ocean biodiversity
and biogeographic data but uses the same sources of data as GBIF except for museum
specimens and herbaria collections. Both facilities contain records that are processed
according to the Darwin Core Standard (DwC, Wieczorek et al., 2012). Specific for algae
is AlgaeBase (Guiry and Guiry, 2018), a repository of information with updated taxonomic
info, images, bibliographic items and distributional records of algae curated by
phycologists. It focuses mainly on taxonomy, but provides also taxonomically reliable literature sources on distribution.

Molecular approaches revealed to be particular useful for the study of planktonic species, especially algae. Taxonomic assignment of specimens based on morphology alone can be inaccurate due to cryptic diversity or variation in morphological characters. This is why species identification is often done nowadays using DNA-based methods (Vanormelingen and Souffreau, 2010; Zimmermann et al., 2015). In addition, high throughput sequencing of taxonomically discriminative barcode regions (HTS metabarcoding), has revolutionised our capacity to gather biodiversity data from environmental samples allowing to identify of a plethora of species present in complex sample matrices or from mass collections of specimens.

HTS metabarcoding is particularly common in the study of marine microbial communities, as shown by several recent projects aimed at characterising the diversity and distribution of sea life. Examples are BioMarKs (http://www.biomarks.eu), the Cariaco Microbial Observatory (Edgcomb et al., 2011), Tara Oceans (https://oceans.taraexpeditions.org/en/m/about-tara/), Ocean Sampling Day, OSD (https://www.microb3.eu/osd.html), and time-series at aforementioned LTER stations. These initiatives are in many ways complementary and additive. For instance, Tara Oceans samples have been taken along a global oceanic trajectory on different dates, and the 18S rDNA-V9 region was used as metabarcode (e.g., Malviya et al., 2016), whereas OSD sampled globally as well, but at coastal sites, on a single day (June 21st summer solstice) and used the 18S rDNA-V4 region (e.g., Kopf et al., 2015). Their standardised procedures, including a centralised hub for laboratory work and data processing guaranteed consistency and data interoperability, and the resulting sequences and contextual data are now publicly available. Previous examples of the use of OSD or Tara Oceans datasets to map phytoplankton distribution were performed using only one of two datasets, without
integration of classical sources and at high taxonomic levels (e.g. Malviya et al., 2016; Lopes dos Santos et al., 2017; Penna et al., 2017; Tragin and Vaulot, 2018). As result of all these activities, a wealth of different kinds of plankton biodiversity data is now available from various sources and in different formats, waiting to be applied to fields such as biogeography, biodiversity estimations, conservation and climate change biology. The integration of all these classical data sources and results from HTS metabarcoding may help to improve environmental monitoring, -management and -policy decisions (Kelly et al., 2014; Thomsen and Willerslev, 2015).

3.1.3. Aim of this work

In this work, I highlight the importance of the integration of classical and novel primary biodiversity data and the challenges related to them through the assessment of the global distribution of Chaetoceros. Chaetoceros is a highly diverse genus of marine planktonic diatoms (VanLandingham, 1968; Rines and Hargraves, 1988), and an abundant one globally (Guiry and Guiry, 2018). Molecular studies (e.g., Gaonkar et al., 2018) make it comparable to higher taxonomic categories (e.g. family or orders) in other diatom lineages. Cryptic diversity seems to be extensive in this group (Kooistra et al., 2010; Balzano et al., 2017; Gaonkar et al., 2017; Li et al., 2017) affecting the mapping of species distribution patterns based on morphological data.

I first explore the potential of different sources of occurrence data at assessing distribution and abundance of a highly diverse phytoplankton genus as well as its species richness in various regions all over the world. Then, I assess distribution patterns of Chaetoceros species using metabarcoding data and compare them with literature data in selected species in order to evaluate their potential and limits in biodiversity assessments.
3.2. Materials and Methods

3.2.1. Data collected from available public repositories, literature and checklists

In order to collect comprehensive info about the distribution of *Chaetoceros* species, I have developed a pipeline that is summed up in Fig. 3.1. I started my search consulting AlgaeBase (Guiry and Guiry, 2018). Upon typing “*Chaetoceros*” in the field ‘search genus’, I performed a preliminary filtering, taking into account only the taxonomically accepted species. For these, I retrieved the listed key literature to take note of the occurrence in the given area. In parallel, I searched Google Scholar for main checklists and distributional records in the literature using as keywords “*Chaetoceros*/phytoplankton distribution” and “*Chaetoceros*/phytoplankton checklist as well as “occurrence” and “biogeography”. Papers resulting from cited literature were also considered. This approach allowed retrieving literature data compiled from experts of the field and so limiting misidentification of species.

I used all the papers focusing mainly on taxonomy containing info at the species-level and considered only names of taxonomically accepted species in Algaebase.
To include other sources of occurrence data at the genus level, I checked the Global Biodiversity Information Facility website (GBIF, https://www.gbif.org/) and the Ocean Biogeographic Information System (OBIS, http://iobis.org/). The former is an online tool including occurrence records of both terrestrial and aquatic species gathered from many sources, from museum specimens to geo-tagged smartphone photos. On the contrary, OBIS contains only records of marine species. Although many datasets are published in both, some are only in one (e.g. herbarium or museum collections containing marine species are only available in GBIF). Furthermore, despite the OBIS network is also included in GBIF, differences in updating procedures can cause temporary differences in results.

In both cases, I used the query “Chaetoceros” and downloaded the resulting occurrence data. Occurrence data generated from both databases were plotted using the R (R Core team, 2018) working packages “rgbif” (Chamberlain, 2017) and “robis” (Provoost and
Bosch, 2018) for GBIF and OBIS respectively. Data were plotted using the packages “maps” (Becker et al., 2018) and “ggplot2” (Wickham, 2016).

3.2.2. Data generated from molecular sources
I used the V4-18S metabarcoding data from the Ocean Sampling Day (OSD) initiative and the V9-18S metabarcoding data from the Tara Oceans expedition to obtain new insights on the global distribution of *Chaetoceros*. For the OSD dataset, I downloaded the V4 lgc workable data (e.g. data already pre-processed in order to derive common data sets on which to base follow-up analysis) available at the website https://mb3is.megx.net/osd-files?path=/2014/datasets/workable. Details of sampling protocols and the different kind of molecular data generated are available at https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD-2014-data, whilst details of pre-processing can be found at https://github.com/MicroB3-IS/osd-analysis/wiki/Sequence-Data-Pre-Processing. The workable fasta files, downloaded for each of 144 geographical sampling sites, were pooled and I generated a total fasta file containing the non-redundant (unique) sequences and a table containing their distribution along the sites (Total OSD abundance table) using mothur v1.41.1 (Schloss et al., 2009).

For Tara Oceans dataset, I downloaded the V9-metabarcode dataset (De Vargas et al., 2017; Ibarbalz et al., 2019) available at https://doi.pangaea.de/10.1594/PANGAEA.873277 and at ENA website with acc. numb. PRJEB6610. Then, following the same pipeline described above, from the total 210 sampling sites, I generated a total unique fasta file and a Total Tara Oceans abundance table.

To generate distribution data, I used a high-quality data reference containing a selection of taxonomic validated *Chaetoceros* sequences of the 18S gene (Goankar et al., 2018). In particular, the reference barcode dataset included 202 *Chaetoceros*, 15 *Bacteriastrum* and 29 outgroup taxa. The fragments V4 and V9 were extracted from the full-length 18S genes.
and aligned using MAFFT online (Katoh et al. 2017). In order to avoid mis-assignations at species level, for the two fragments (V4 and V9) I simulated several thresholds of clustering based on genetic distances (commands “dist.seqs” and “cluster” in mothur) (Schloss et al., 2009).

The V4 and the V9 reference sequences were used as queries for a local BLAST against the two global metabarcode datasets OSD and Tara Oceans. For the mapping at genus level, I set the threshold to 90% of identity and from the outputs of BLAST I retained only the metabarcode hits having a query coverage with the reference > 370 bp in the analysis of V4 OSD dataset, and >105 bp for V9 Tara Oceans dataset. The metabarcodes extracted were aligned with the references, including outgroup taxa, using MAFFT online (Katoh et al., 2017) and two phylogenetic trees were then built in FastTree v2.1.8 (Price et al., 2010) using the GTR model and visualised in Archaeopteryx v0.9901 (Han and Zmasek, 2009).

Metabarcode hits clustering within the outgroup clades were excluded from further analyses, whereas the others were considered as validated *Chaetoceros*. Their abundances and distributions were extracted from the Total OSD and Tara Oceans abundance tables to generate the *Chaetoceros*-genus OSD abundance table and *Chaetoceros*-genus Tara Oceans abundance table. For the mapping at species level, I first evaluated the information generated from the analyses described above for the V4 and V9 fragments (calculation of the genetic distances and simulation of several thresholds of clustering). Based on them, I extracted only the BLAST hits assigned in the range 100-99% of similarity. This range was identified as the best compromise between the precision required to an assignation at species level and the intra-species variation that could occur especially at global level. After the BLAST, I applied the same procedure described above for the genus level (alignment and generation of tree) to validate the assignations and we generated the *Chaetoceros*-species abundance table for the OSD and for Tara Oceans datasets.
The Chaetoceros-genus abundance tables were used both in term of occurrence and abundance of V4 and V9 reads in each sampling site. Abundance values were log10-transformed and plotted using ggplot2 (Wickham, 2016).

Finally, to explore in detail the performances of classical and molecular data, I selected three species as case study: i) *C. tenuissimus* as test of cosmopolitan species; ii) *C. gelidus* as species with restricted distribution; iii) *C. neogracilis* as example of putative cryptic species complex.

### 3.3. Results

3.3.1. Data collected from available public repositories, literature and checklists

According to AlgaeBase, the genus *Chaetoceros* contained 370 species names and 172 intraspecific ones, 220 of which have been flagged as taxonomically accepted species based on the available literature (searched on 15/10/2018). This discrepancy is due to the occurrence of many homotypic or heterotypic synonyms in the literature as well as species of uncertain taxonomic status, which need taxonomic revision. I further filtered the 220 taxa flagged as taxonomically accepted (e.g. removing entries occurring twice) obtaining a final table (Table A3.1, Appendix III) with 175 entries at the date of the search. I considered the latter taxa in the count for species richness from literature data (see below).

The distribution map of *Chaetoceros* obtained using GBIF data (Fig. 3.2A) was based on 201,047 occurrence records from 1863 to 2018 (https://www.gbif.org/occurrence/charts?q=chaetoceros). Data were mostly from human observations (75.7 %) and preserved specimens (20.2 %) (GBIF.org, 14 September 2018, GBIF Occurrence Download https://doi.org/10.15468/dl.nofa8w). The definition of records is available at https://gbif.github.io/gbif-api/apidocs/org/gbif/api/vocabulary/BasisOfRecord.html. Filtered occurrence data from GBIF are also available as supplementary info (Table A3.2, Appendix III). No information
from literature was available for Chaetoceros in GBIF data. Most of the observations were from the North Atlantic Ocean between 35° - 60° N and -80° W – 10°E (Continuous Plankton Recorder Dataset, SAHFOS, 83,513 counts; Réseau d'Observation et de Surveillance du Phytoplancton et des Phycotoxines, REPHY, 17,742 counts; QUADRIGE, 12,458 counts), followed by the Pacific coasts of North and Central America and Australia (Fig. 3.2A).

The distribution map obtained searching Chaetoceros in the OBIS database (Fig. 3.2B) contained 389,206 records from 1863 to 2016 (Table A3.2, Appendix III). Most of observations were from the World Ocean Database 2009 (119,592), followed by the Continuous Plankton Recorder (86,309) and the Japan Oceanographic Data Center Dataset (JODC, 31,388).

Chaetoceros occurrence data were found in 435 GBIF datasets and 179 OBIS datasets, of which 20 were shared (Table A3.2, Appendix III).
Fig. 3.2. Occurrence of *Chaetoceros* using (A) GBIF and (B) OBIS data.

The literature search conducted in Google Scholar and the other sources (see Material and Methods) resulted in 84 main bibliographic references reporting data of *Chaetoceros* occurrences (Table A3.3, Appendix III). These data encompassed both single observations and time series across the world, covering a period from 1873 to 2017 (Table A3.3, Appendix III). These data surely represent only a fraction of the whole existing literature (and the literature indexed in Google Scholar) but are representative of the principal checklists/floras compiled by expert taxonomists and of the spatial coverage were *Chaetoceros* is known to occur. None of these bibliographic references (checklists and
papers) was contained in GBIF or OBIS datasets (Table A3.2, Appendix III). According to these data, *Chaetoceros* species mostly occurred in the temperate to equatorial coastal waters of northern hemisphere and in the subtropical to tropical coastal waters of the southern one (Fig. 3.3).

![Fig. 3.3. Occurrence of *Chaetoceros* using literature data.](image)

In terms of species richness, here defined as the number of valid species recorded in each locality’s checklist, I found the highest values in the temperate waters of European coasts (North Sea, Baltic Sea, and middle Adriatic Sea, Fig. 3.4), followed by the tropical and subtropical waters of Brazil, Mozambique Channel and Indonesia (Fig. 3.4). The lowest number of species was found in the subpolar waters alongside the coasts of northern countries (Canada, Greenland, Norway and Russia) as well as in the equatorial ones of southern oceans (Fig. 3.4).
Fig. 3.4. Species richness of *Chaetoceros* estimated from literature data. Colours refer to the different classes of abundance (number of species recorded).

### 3.3.2. Data generated from molecular sources

Based on the generation of distances and simulation of clustering thresholds, the clustering at 100% similarity of the V4 *Chaetoceros* reference dataset (unique or non-redundant sequences) resulted in the collapse of only multiple strains from the same species, whereas the clustering at 99% similarity threshold resulted in the collapse of several species (Table A3.4, Appendix III). On the contrary, in the V9 *Chaetoceros* reference dataset the clustering at 100% of similarity produced the collapse of different taxa generating more limitations in the mapping at species level (Table A3.4, Appendix III).

At genus level, I found occurrences of *Chaetoceros* taxa in 138 out of 144 OSD sampling sites (96%) and 146 out of 210 Tara Oceans stations (70%), highlighting very wide distribution of the genus (Fig. 3.5, Table A3.5, Appendix III).
Fig. 3.5. *Chaetoceros* distribution according to OSD (A) and Tara Oceans (B) data. Dots indicate presence of *Chaetoceros* taxa in the sampling stations, whilst triangles their absence.

The plot of abundances, both in OSD and in Tara Oceans datasets, showed that *Chaetoceros* was equally abundant in the northern as in the southern hemisphere (Fig. 3.6). The highest abundances (in terms of reads) were mostly found in the polar to temperate regions of the two hemispheres, with some exceptions in the equatorial coastal waters of India and Indonesia (Fig. 3.6A). Lowest abundances were found in the subtropical to equatorial zones, especially in open ocean stations in the case of Tara Oceans dataset (Figure 6B), in the Red Sea for both datasets, and other few sites in the OSD dataset (Fig. 3.6A).
Fig. 3.6. Log10 abundance of *Chaetoceros* reads according to OSD (A) and Tara Oceans (B) datasets. Size and colours of the circles refer to the abundance.

At species level I generated, at 99% similarity threshold, a map of occurrence in the OSD and Tara Oceans datasets for each of the 69 *Chaetoceros* species (Figure A3.1, Appendix III). The only exceptions were *C. cf. vixvisibilis* Na16A3 and *C. sp. Clade Na28A1 strain Na26C1*, in which the collapse of barcodes prevented the plot of occurrences in Tara Oceans stations at species level.
The comparison of literature and genetic (metabarcoding) data in selected species of Chaetoceros (Fig. 3.7) showed consistency in the signal for *C. tenuissimus* and *C. gelidus*, and highlighted the occurrence of putative cryptic species in *C. neogracilis*.

In *C. tenuissimus*, literature (Fig. 3.7A) and metabarcoding data (Fig. 3.7B) confirmed a cosmopolitan distribution, with metabarcoding data providing new records for African, Asian and New Zealand coasts (Fig. 3.7B).

For *C. gelidus*, genetic data from OSD and Tara Oceans (Fig. 3.7D) confirmed the distribution area of literature data (field observations, Fig. 3.7C) but also included new records for Canada, North Scotland and Iceland (Fig. 3.7D). The species was also found in one OSD station in the Caribbean side of Panama coasts, but very low abundance (2 reads at 100% similarity).

*C. neogracilis* revealed to be an example of cryptic species complex. According to literature, the species was found both in the northern and southern hemisphere (Fig. 3.7E). On the contrary, occurrence data from metabarcoding revealed instead a distribution limited to the northern hemisphere, so covering just a small part of the distribution range known from literature data (Fig. 3.7F).
Fig. 3.7. Distribution of *C. tenuissimus* (A, B), *C. gelidus* (C, D) and *C. neogracilis* (E, F) according to literature (orange dots) and metabarcoding data (blue dots for OSD and red dots for Tara Oceans). Maps containing the sites considered for literature and metabarcoding data are found in Figure 3 and Figure A3.1 respectively.

### 3.4. Discussion

#### 3.4.1. Global distribution of *Chaetoceros*

The more complete picture of *Chaetoceros* distribution was provided by the GBIF and OBIS platforms, which contain a huge amount of data from different sources (fossils, literature, machine and human observations, museum and herbarium specimens) and cover a wide time scale (in this case more than 150 years). Despite OBIS is a resource dedicated to marine organisms already included in GBIF database, I did not recover the same number of records and datasets from the two sources. Differences in updating data procedures are partially responsible for such temporary differences in results. Furthermore, some kinds of
information as museum collections are only available in GBIF, generating the necessity to interrogate both databases also in the case of marine species to ensure a complete mapping. The overview provided by the Google Scholar search of the main phytoplankton checklists is, despite the obvious limitations, able to provide the main distributional areas of the genus. Google Scholar can be considered as a convenient starting place to start a literature search, not a comprehensive endpoint. It has among its advantages the fact that is easily accessible to retrieve data that are otherwise stored in libraries’ catalogues and databases and goes back in time in the scale of hundreds of years or more. Since this approach is highly sensitive to the kind and order of keywords used for the search, I cannot exclude the possibility of having missed some information, even if multiple searches were performed. However, I have retrieved datasets that are not included in GBIF or in OBIS. This aspect underlines that despite the big effort to generate and update these global databases, a minor part of the information could be lost. Furthermore, it highlights the difficulty for the researches to produce an exhaustive assessment of all the available data of a particular taxon. Probably, more effort is needed by the institutions from around the world to provide and share biodiversity datasets generated.

The two global metabarcoding datasets OSD and Tara Oceans, despite biased in time and space, provided an overall distribution map of the genus that is comparable to the one obtained from the sources discussed above. This clearly highlights that, despite some weaknesses (e.g. Coissac et al., 2012; Ficetola et al., 2015), the metabarcoding approach, in less than a decade from its diffusion, was able to compete with classical morphological records gathered over hundreds of years. At the moment, metabarcoding data cannot replace the classical ones, and should be seen as a powerful complement rather than a substitute of other data sources (Bush et al., 2017). For instance, the Tara Oceans dataset added new occurrence information for equatorial regions and other open ocean sites in the southern hemisphere, contributing to our knowledge in these still poorly investigated areas.
Despite both OSD and Tara Oceans datasets are open access, the extraction of information from these sources is not straightforward and requires some bioinformatic skills that are not common, especially among taxonomists.

My results showed that all data sources (GBIF, OBIS, Google Scholar search, OSD and Tara Oceans) support a cosmopolitan distribution of this genus as suggested by Rines and Hargraves (1988) using only classical sources, and Malviya et al. (2016) using only metabarcoding data. In terms of occurrence, *Chaetoceros* taxa showed a global distribution ranging from coastal areas to open ocean and from polar to tropical regions. However, the different data sources point out a prevalence of taxa in the temperate coastal waters between the temperate waters 60°N and 30°N and in the subtropical and equatorial ones between 30°N and 30°S. This can be due to the presence in such regions of various habitats (upwelling zones, lagoons, oligotrophic as well as eutrophic regions) and the marked seasonality in the water, which offer opportunities of co-existence of species by spatial or seasonal niche partitioning. Boreal regions are poorer probably because there is only the single summer season for phytoplankton growth.

With some exceptions (e.g. Hernández-Becerril and Granados, 1998 for the Gulf of Mexico and Hernández-Becerril, 1996 for the Mexican Pacific), the tropics are generally under-investigated for species diversity, though this is now ameliorated by recent studies in those regions (Li et al., 2013; 2017; Chamnansinp et al., 2015).

3.4.2. Abundance of *Chaetoceros* at global scale

In general, patterns of abundance in both molecular datasets suggest that *Chaetoceros* is equally abundant in the temperate to equatorial waters of northern and southern hemispheres, with the highest abundance in the Arctic region. A paucity of reads was generally observed from many sites located in the open ocean. This observation could reflect the well known hypothesis made on terrestrial ecosystems, according to which cold
to temperate regions contain less species but more abundant. However, multiple variables involved could alter such results. First and obvious is that the picture provided by metabarcoding data is very limited in space and time, and could not represent the real situation. Second, some species may have been collected during a bloom period, which could explain the high values of abundance. Chaetoceros socialis, for example, is known to be an important component of diatom blooms in the Barents Sea (Von Quillfeldt, 2000). Third, data here used (V4 and V9 regions) are from a multi-copy gene and since the copy number in Chaetoceros is unknown, this aspect, combined with a hypothetical bloom, could hamper our conclusions.

A previous mapping of Chaetoceros in Tara Oceans dataset was performed in Malviya et al. (2016) using only 46 stations. In the latter study, Chaetoceros was found to be highly abundant in the Southern Ocean and absent in the polar regions of the northern hemisphere. My analysis, using the complete Tara Oceans dataset (210 stations), showed that Chaetoceros is present also in the polar regions of the northern hemisphere, highlighting the fact that the wider the coverage of sampling and/or the integration from other source the better the resolution of distribution.

3.4.3. Integration of literature and metabarcoding data: three study cases in Chaetoceros

The direct comparison of literature and metabarcoding data in three selected species of Chaetoceros shows the power of novel molecular data coupled with classical occurrence data. In the case of C. tenuissimus, the molecular data allowed to increase the geographic range of distribution of this cosmopolitan species with new records in African, Asian and New Zealand coasts. Yet, in C. gelidus molecular data confirmed the previous knowledge on its restricted distribution in cold water, also adding new records for Canada, North Scotland and Iceland. For this reason, at the moment I interpret the occurrence of two reads found in one OSD station in the Caribbean coasts as a spot occurrence rather than a stable
geographic point. However, global changes could alter limits both in cosmopolitan or restricted species with consequent range expansion or contraction, highlighting the importance to generate baseline studies of the geographic distribution range of taxa to use as bases for future comparisons.

More complex is, instead, the case of *C. neogracilis*. The epithet *C. neogracilis* (*C. gracile* Schütt) has been attributed in the past to many small, unicellular *Chaetoceros* taxa collected worldwide (Rines and Hargraves, 1988). This led to considering the species cosmopolitan. A recent study by Balzano et al. (2017) from the Beaufort Sea (Canadian Arctic) revealed the occurrence of morphologically similar strains sharing identical 18S rDNA sequences, but belonging to four distinct genetic clades based on 28S rDNA, ITS-1 and ITS-2 markers. Since OSD and Tara oceans datasets are based on the 18S gene, I regarded these entities as one single species. In Balzano et al. (2017) they are also reported to co-occur at the stations they visited. The reference barcode from Balzano et al. (2017) blasted against the two datasets found identical sequences only in the cold waters of the northern hemisphere, so covering just a small part of the distribution range known from literature data. My results strongly suggest that under the name *C. neogracilis* there is a species restricted to polar regions of the northern hemisphere (as highlighted also by Balzano et al., 2017). Furthermore, as pointed out by these authors, a closely related species occurs in the cold waters of Antarctica, whilst the status of the *neogracilis* taxon reported in literature from South America and Africa is still to be determined. It could be a complex of taxa with similar morphology and, further samplings in these regions accompanied by genotyping of strains, will help clarifying the taxonomic status. However, the example highlights how the integration of several sources is required to a correct interpretation at species level of the patterns obtained from a metabarcoding sampling.
3.4.4. Assessing species distribution in Chaetoceros

The maps of occurrences generated using the OSD and Tara Oceans datasets for each of the 69 Chaetoceros species, provide new insights on biogeography in marine diatoms. According to available literature, few endemic diatom species are known, and they are mostly freshwater (e.g. Eunophora in Tasmania and New Zealand, Vanormelingen et al., 2008 and Cyclotella minuta in Lake Baikal, Mackay et al., 2006) or from saline inland lakes (e.g. Aulacoseira baicalensis). Claims of putative endemic marine diatoms exist and are discussed in Mann and Vanormelingen (2013). In the specific case of Chaetoceros, Hernández-Becerril (1996) recognised that little efforts have been made to assess its world distribution but, starting from literature data available and personal observations, he grouped taxa according to major regions as inhabitants of cold waters, temperate to subtropical waters, world-wide warm waters and tropical and subtropical waters. Metabarcoding data here analysed suggested that cases of endemism or restricted geographical distributions can be also found in the marine environment. I detected species whose occurrence is limited to single basins as the Mediterranean Sea (C. diversus 1) or part of it (C. throndsenii in the Adriatic Sea) as well as distribution restricted to climatic zones (e.g. the polar to temperate zones for C. constrictus, C. danicus strain RCC2565, C. debilis 1 and C. neogracilis).

3.4.5. Future directions

In this chapter, I have highlighted both the importance of the integration of data and the challenges related to it, generating a comprehensive primary baseline of the geographic distribution range and diversity for Chaetoceros, one of the most diverse and abundant genera of marine planktonic diatoms. I have also stressed out that, at the moment, molecular and classical sources tend to be organised and maintained in separated repositories or infrastructures, forcing the users interested in integration of such sources to
do a not trivial work across the several sources of data and analyses. Certainly, molecular approaches can improve our knowledge both reducing mis-assignments (taxonomic lumping) in cryptic species complexes and helping for rare and small species not easily detected with traditional methods, especially for microbial (protist) species. However, this is not always true and, as I have shown, the short fragments used in metabarcoding can be identical in closely related taxa, not allowing a discrimination at species level (Cowart et al., 2015; Mordret et al., 2018; Piredda et., 2018). Nonetheless, metabarcoding data are a valuable source of primary biodiversity data.

The knowledge of the geographic range of species is a key issue in ecology, conservation and evolutionary biology allowing investigating causes and consequences of the limits. Climate change can alter these limits with consequent range expansion or contraction, and several examples have been reported (Walther et al., 2002; Parmesan and Yohe, 2003; McLachlan et al., 2005). This process is supposed to be underway, stressing the need to collect, integrate and summarise data available to create a Primary Biodiversity Data baseline. These collections provide bases for future comparisons or model predictions to support biodiversity change assessments.

References


Li, Y., Lundholm, N., Moestrup, Ø. (2013). *Chaetoceros rotosporus* sp. nov. (Bacillariophyceae), a species with unusual resting spore formation. *Phycologia, 52*(6), 600-608.


Appendix III
Fig. A3.1. Distribution maps of *Chaetoceros* species using OSD and Tara Oceans datasets.
Note: Tara Ocean's distribution (red dot) might refer to C. sp. Na28A1 due to identical reference sequences in Tara Ocean's dataset.
Note: Tara Oceans distribution (red dot) might refer to *C. cf. vitreous* due to identical reference sequences in Tara Oceans dataset.
For the Supplementary Tables cited in this Chapter, I remind to the online material published alongside the paper (https://peerj.com/articles/7410/#supplemental-information) as follows:

<table>
<thead>
<tr>
<th>Reference in the Chapter</th>
<th>Corresponding reference in the paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table A3.1 - List of taxonomically accepted <em>Chaetoceros</em> taxa according to AlgaeBase</td>
<td>DOI: 10.7717/peerj.7410/supp-5</td>
</tr>
<tr>
<td>Table A3.2 - List of datasets included in GBIF and OBIS platforms</td>
<td>DOI: 10.7717/peerj.7410/supp-6</td>
</tr>
<tr>
<td>Table A3.3 - List of bibliographic references reporting <em>Chaetoceros</em> at given localities</td>
<td>DOI: 10.7717/peerj.7410/supp-7</td>
</tr>
<tr>
<td>Table A3.4 - Clustering test at 100 and 99% of similarity for V4 and V9 fragments</td>
<td>DOI: 10.7717/peerj.7410/supp-8</td>
</tr>
<tr>
<td>Table A3.5 - List of OSD and Tara Oceans stations</td>
<td>DOI: 10.7717/peerj.7410/supp-9</td>
</tr>
</tbody>
</table>
Chapter IV

Resolving the

Chaetoceros curvisetus

*cryptic species complex*
4.1. Introduction

4.1.1. Cryptic species complexes: origin, distribution and methodology of study

Cryptic species is a collective term generally used to indicate taxa that are morphologically indistinguishable to the observer but for which there is evidence (genetic, ecological, behavioural, etc.) of belonging to different evolutionary lineages (Mayr, 1970; Bickford et al., 2007). When many virtually identical species are involved, these groups of organisms are commonly referred to as “cryptic species complexes”. Cryptic species may originate from recent divergence during speciation process (Fišer et al., 2018), which results in the lack of substantial morphological differences, or may be phylogenetically old and reproductively isolated from each other by strong biological barriers (Trontelj et al., 2009). In some cases, cryptic species can be phylogenetically unrelated and resulting from mimicry and convergence (Struck et al., 2018). In any case, they are real biological entities that have been inaccurately identified by taxonomists.

Increasing knowledge showed that cryptic species occur on all the branches of the tree of life and biogeographic regions (Pfenninger and Schwenk, 2007; Trontelj and Fišer, 2009). Furthermore, the frequency with which they are discovered using DNA sequence data calls for the integration of such methods in the process of species discovery and description (Bickford et al., 2007). The study of cryptic species has been approached in different ways, e.g. inferring phylogenies (e.g. Andrews et al., 2016), using species delimitation methods (e.g. Jörger et al., 2012; Crawford et al., 2013; Mills et al., 2017) and integrative taxonomy approaches (e.g. Gomes et al., 2015; Papakostas et al., 2016; Steiner et al., 2018). In general, all these approaches rely on the information gathered from Sanger sequencing of selected genes from a few sampled specimens (Lukhtanov et al., 2015; Saitoh et al., 2015; Iftikhar et al., 2016) and following inferred trees, to which morphological examinations can be added (e.g. integrative taxonomy approach). If genetic distances are large enough to justify an independent evolutionary lineage (e.g. a separate branch in a phylogenetic tree),
the occurrence of cryptic species is hypothesised. However, none of these approaches is free of pitfalls. First, phylogenetic trees may not be the best tool to visualise putative cryptic species, since they are well suited for representing evolutionary histories resulting from bifurcating speciation events and vertical changes within an ancestor-descent lineage (Huson et al., 2010). In the case of cryptic species, especially if they are the product of recent divergence, there could still be ongoing gene flow, which is better modelled by networks rather than phylogenetic trees. Second, it is often difficult to have a picture of the geographic variability of a species across its distributional area and what is indicated as putative cryptic species could also be a geographically isolated population that is undertaking a different evolutionary history.

Metabarcoding data from environmental samples have proven to be a powerful tool of non-invasive biodiversity assessment from species to community level (Cristescu, 2014; Deiner et al., 2017) and, in more recent times, a new source of biological records (Lawson Handley, 2015). They provide not only a bulk of sequence data for a gene region of interest of all the community sampled, but also information about their relative abundance, genetic variability and distribution. However, their application in ecology and evolution is still largely unexplored, especially as tool for inferring phylogenetic and phylogeographic relationships among taxa.

4.1.2. The Chaetoceros curvisetus species complex

The Chaetoceros curvisetus species complex currently includes several morphologically similar species that all share the straightforward characteristic of having the setae directed toward the outside of the chain spiral (Hasle and Syvertsen, 1996). Under light microscopy, some morphological features can distinguish among them, as the size of aperture between sibling cells (Kooistra et al., 2010), large and elliptical or nearly circular
in curvisetus (Fig. 4.1A) and narrow and oval in pseudocurvisetus (Fig. 4.1B). This is the visible effect of a very different valve morphology and a different type of cell junction.

Fig. 4.1. Chaetoceros curvisetus (A) and C. pseudocurvisetus (B). The size and shape of aperture between sibling cells (see arrows) are useful characters for distinguishing these taxa.

All the species included in Chaetoceros curvisetus species complex form a monophyletic group, the section Curviseta (see Chapter II). To date, the only species that have been formally described are Chaetoceros curvisetus Cleve and C. pseudocurvisetus Mangin. A first molecular analysis using the hypervariable region (D1-D4) of the LSU rDNA gene revealed the occurrence of two distinct genetic clusters within C. curvisetus (Kooistra et al., 2010). A second screening, including more strains and sequences of LSU and SSU rDNA genes (Gaonkar et al., 2018) raised the number of genetic clusters in “curvisetus” to three. Furthermore, both studies highlighted the seemingly paraphyletic status of C. curvisetus due to a closer phylogenetic relationship among some “curvisetus” species to pseudocurvisetus strains than to other conspecifics. According to Gran (1897), C. curvisetus can be found throughout the year in the Atlantic Ocean and the Baltic Sea, but is especially abundant in summer and autumn. Hasle and Syvertsen (1996) indicated C.
curvisetus as a cosmopolitan mainly found in temperate and warm waters and C. pseudocurvisetus as an inhabitant of warm waters. My results of Chapter III (see Fig. A3.1) have shown that most of the C. curvisetus spp. have apparently no specific distribution restricted to particular habitats, with the exception of C. curvisetus 1, which was mostly found in cold-temperate waters. C. pseudocurvisetus was found in the warm waters of Indian coasts and Indonesia as well as in the Mediterranean Sea and nearby Atlantic Ocean (Fig. A3.1, Chapter III).

4.1.3. Objectives of the study

In this chapter, I use an 18S reference library of C. curvisetus species and close out-group taxa and map it against two global metabarcoding datasets: Tara Oceans and The Ocean Sampling Day 2014. The resulting data are used to generate a phylogenetic network in order to: 1) infer the number of the species within the complex; 2) explore the evolutionary relationships and the presence of gene flow among the members of the complex. Furthermore, I assess the distribution of the complex according to OSD and Tara Oceans data as well as the occurrence and abundance of each species delimited from the networks in Longhurst’s biogeographic provinces (Longhurst, 2007).

I also explore the relative impact of sequence variability introduced by PCR and sequencing artefacts on the one hand and inter- and intraspecific variability on the other hand in metabarcoding data as well as the utility of using genetic distances to set boundaries across taxa.

4.2. Materials and Methods

4.2.1. Download and processing of metabarcoding data

To assess the phylogenetic relationships among members of the C. curvisetus species complex on a global scale, I used the V4-18S metabarcoding data from OSD and the V9-
18S ones from Tara Oceans. OSD data were downloaded from https://mb3is.megx.net/osd-files?path=/2014/datasets/workable, whilst Tara Oceans data (De Vargas et al., 2017; Ibarbalz et al., 2019) from https://doi.pangaea.de/10.1594/PANGAEA.873277 and ENA website at acc. numb. PRJEB6610. For the OSD dataset, I pooled together the 144 workable fasta files from each sampling site and generated a total fasta file with the unique sequences and a table containing their abundance across the sites (Total OSD abundance table) using mothur v1.41.1 (Schloss et al., 2009). For the Tara Oceans dataset, I directly extracted a total unique fasta file and a Total Tara Oceans abundance table from the downloaded file containing sequences from 210 sampling sites.

To retrieve sequences of *C. curvisetus*-like taxa from these metabarcoding data, I started from the full length 18S rDNA sequences of *C. curvisetus* and *C. pseudocurvisetus* species and close outgroups (*C. tortissimus* and *C. cf. tortissimus*, Table 4.1) provided in Gaonkar et al. (2018) and used in Chapter II for phylogenetic inference. For V4 region, further barcodes were retrieved from NCBI, in particular for *C. curvisetus* (strain SKLMP_YG033, acc. numb. MG821562) and *C. pseudocurvisetus* (strain IRB, acc. numb. MG385841). In this chapter, numbers after *C. curvisetus* species’ names (1, 2, 2c, 3 and 3e) refer to genetically defined species for which a formal description is not available yet, but that are discussed in Gaonkar et al. (2018) or in this thesis (e.g. Chapter II). Light microscopy photographies of these species are provided in Fig. 4.2. The wording “sp.” followed by number (1, 2, 3 and 4) refers to hypothetical new species here identified.

**Table 4.1. List of reference sequences utilised for gathering *C. curvisetus*-like taxa.**

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Strain</th>
<th>Accession Number</th>
<th>Reference for V4</th>
<th>Reference for V9</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. curvisetus</em></td>
<td>SKLMP YG033</td>
<td>MG821562</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 1</td>
<td>Na10C1</td>
<td>MG972232</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 2</td>
<td>Na1C1</td>
<td>MG972235</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Species</td>
<td>Accession</td>
<td>Sequence</td>
<td>Colony</td>
<td>Clonal Descent</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------------</td>
<td>----------</td>
<td>--------</td>
<td>----------------</td>
</tr>
<tr>
<td>C. curvisetus 2c</td>
<td>El6A2</td>
<td>LC466961</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>newBB2</td>
<td>MG972241</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>C. curvisetus 3e</td>
<td>El4A2</td>
<td>LC466962</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>C. pseudocurvisetus</td>
<td>IRB</td>
<td>MG385841</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>C. pseudocurvisetus</td>
<td>Na13C4</td>
<td>MG972304</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>C. cf. tortissimus</td>
<td>Na18C4</td>
<td>MG972275</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>C. tortissimus</td>
<td>Na25A2</td>
<td>MG972325</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

**Images:**
1. C. curvisetus 1
2. C. curvisetus 2
3. C. curvisetus 2c
4. C. curvisetus 3
I extracted from the 18S region the V4 and V9 regions corresponding with the fragment amplified by the primers used in OSD and Tara Oceans. These fragments were clustered at several thresholds (100-99%) to ensure that different *C. curvisetus* species were not collapsed together (see Chapter III). The reference fragments were used as queries for a local BLAST to recover entries at 95% of similarity against the OSD and Tara Oceans datasets. The combined strategy of using both reference barcodes of close outgroups and a relaxed threshold (95%) allowed gathering in the metabarcoding datasets sequences of *C. curvisetus* like taxa for which reference barcodes could be unavailable.

The metabarcodes extracted were aligned with the references, including outgroup taxa, using MAFFT online (Katoh et al., 2017) and a phylogenetic tree was built in FastTree v2.1.8 (Price et al., 2010), using the GTR model. The resulting tree was visualised and modified in Archaeopteryx v0.9901 (Han and Zmasek, 2009), in order to remove false positive sequences clustering within outgroup clades and gather only *curvisetus*-like metabarcodes. This procedure was followed separately for V4 and V9 fragments. The sequences filtered through the previous procedure, were considered validated as *C. curvisetus*-like. The abundance and distribution of V4 and V9 *curvisetus* metabarcodes were extracted from the Total OSD and Tara Oceans abundance tables. At the end of the validation procedures, I generated four files: 1) the V4_OSD_curvi_validated.fasta file,
containing the sequences validated as *C. curvisetus* from OSD; 2) the V4_OSD_curvi_validated.count_table file, containing the distribution of each haplotype across the OSD sites; 3) the V9_TARA_curvi_validated.fasta file, containing the sequences validated as *C. curvisetus* from Tara Oceans; 4) the V9_TARA_curvi_validated.count_table file, containing the distribution of each haplotype across the Tara Oceans sites.

4.2.2. Phylogenetic haplotype network inference

Phylogenetic haplotype networks were used to circumscribe species within a species-complex in a non-dichotomous approach. For such inference, I used the statistical parsimony algorithm by Templeton et al. (1992) implemented in TCS network (Clement et al., 2002). This agglomerative algorithm collapses sequences in haplotypes and estimates the number of differences among them due to single substitutions and with a 95% statistical confidence (parsimony limit). Then, haplotypes (nodes in the network) are progressively connected among them by edges starting from the ones that differ by one change, then by two, three and so on until all the haplotypes have been connected into a single network or the parsimony limit has been reached. This kind of phylogenetic network was preferred over others (e.g. median joining networks, MJ) because it shows reticulations and includes unobserved haplotypes in the network as the MJ network but is computationally quicker.

Despite displaying the final output as networks, TCS approach differs from Swarm (Mahé et al., 2015). Swarm is a de novo clustering method that uses a clustering threshold (d) of nucleotide difference (a substitution, insertion, or deletion), whilst TCS works on a multi-alignment. Moreover, edges in Swarm networks carry no phylogenetic information, but are only a representation of the parameter d used, so it is not possible to infer relationships among OTUs as in TCS networks. Since I was interested in assessing the internal structure
(phylogenetic relationships) of my species complex and not only in the assessment of OTUs, I have chosen the TCS method over Swarm.

TCS algorithm was inferred and visualised as implemented within PopART v1.7 (Leigh and Bryant, 2015). Abundance of sequences was included in the inference. Each network was exported as table and nexus file.

Using the information contained in the table of haplotype of each TCS network, I delineated species following these criteria: 1) the sequences found within a node including the reference barcode were attributed to that species; 2) the sequences having mutations ≤ 2 in respect to the node with the reference and with abundance ≤ 3 were attributed to the node; 3) nodes without reference and with mutations > 2 and abundance > 3 respect to the ones with reference were considered as hypothetical new taxa. The latter were indicated as C. curvisetus sp. 1, sp. 2, etc.

After species inference, I took the representative sequence of each delimited species and inferred a phylogenetic tree (for V4 and V9 regions) for a rapid and supported visualisation of phylogenetic relationships among taxa. Maximum Likelihood (ML) trees were inferred using IQ-TREE v1.6.8 (Nguyen et al., 2014) under the TN+F+G4 model for V4 and the K2P+G4 model for V9 (suggested by ModelFinder, Kalyaanamoorthy et al., 2017) and 1000 bootstrap replicates for both datasets. The sequences of C. tortissimus and C. cf. tortissimus were used as outgroup.

4.2.3. Genetic divergence among species and variability within species

To quantify the relatedness of each species in terms of distances rather than number of mutations, I calculated the net genetic distances between pairs of species as implemented in MEGA6 (Tamura et al., 2013):

\[ d_A = d_{XY} - (d_X + d_Y)/2, \]
where \( d_{XY} \) is the average distance between groups X and Y, and \( d_X \) and \( d_Y \) are the mean within-group distances.

I used the Jukes-Cantor (JK) model of sequence evolution (Jukes and Cantor, 1969) to calculate the genetic distances across all metabarcodes of each species, which best fitted our data. I also calculated, using the same model, the minimum, maximum and average evolutionary divergence of sequences within nodes (the number of base substitutions per site from averaging over all sequence pairs within each group) using MEGA6 (Tamura et al., 2013). The presence of barcoding gap in the inferred species was explored. The barcoding gap was considered to occur if the maximum distance within species was lower than the minimum distance between species (Meyer and Paulay, 2005).

4.2.4. Global distribution of taxa belonging to the \( C. \) curvisetus species complex

I mapped the distribution of the members of the \( C. \) curvisetus species complex in world’s oceans using the previously inferred species. First, from the abundance tables previously generated (V4.OSD_curvi_validated.count_table and V9.TARA_curvi_validated.count_table files), I summed the abundances of the haplotypes belonging to the same inferred species. Then, I plotted the occurrence of all \( C. \) curvisetus inferred species together on a world map divided in Longhurst’s provinces. I also plotted the abundance of each species (in terms of reads) in Longhurst’s provinces in the form of heatmaps.

To plot the occurrences, I downloaded the shapefiles containing the coordinates of Longhurst provinces (Longhurst, 2007) from the Marine Regions portal (http://www.marineregions.org/downloads.php#longhurst) and plotted them using the R package \texttt{rgdal} (Bivand et al., 2018) and the function \texttt{ssplot} in the daughter process “sp” (Pebesma and Bivand, 2005; Bivand et al., 2013). For abundances, I used the R (R Core}
Team, 2019) working packages *phyloseq* (McMurdie and Holmes, 2013) and *ggplot2* (Wickham, 2016).

### 4.3. Results

After the extraction of the V4 and V9 regions, corresponding to the fragments amplified by the primers utilised in OSD and Tara Oceans datasets, from partial or full-length 18S sequences, I obtained 10 reference sequences for V4 and 8 for V9 (Table 4.1). Such difference was because two 18S sequences (*C. curvisetus* strain SKLMP YG033 and *C. pseudocurvisetus* strain IRB) did not cover the V9 region too.

#### 4.3.1. Validation of *C. curvisetus* candidate sequences

Regarding the OSD dataset, following BLAST analysis I retrieved 4,223 sequences corresponding to 1,428 unique haplotypes including outgroups. After the validation of metabarcodes belonging to the *C. curvisetus* species complex by means of the phylogenetic tree-approach, I gathered 1,232 haplotypes, for a total of 3,804 sequences. Regarding Tara Oceans data, BLAST analysis returned 856,967 sequences corresponding to 2,247 unique haplotypes including outgroups. After validation, I eventually retrieved 68,210 sequences for 772 haplotypes belonging to the complex.

Metabarcodes validated as *C. curvisetus* (1,232 for OSD and 772 for Tara Oceans) were found in 60 out of 144 OSD sampling sites (41.7%) and 117 out of 210 Tara Oceans stations (55.7%) (Fig. 4.3, Table A4.1 in Appendix IV).
Fig. 4.3. Occurrence of taxa belonging to the *C. curvisetus* species complex in OSD (A) and Tara Oceans (B) datasets. Blue dots refer to occurrence in OSD data, whilst red dots in Tara Oceans data. Grey triangles indicate absence in the respective sampling site.

4.3.2. Phylogenetic haplotype networks

The haplotype network based on the OSD dataset (V4 region) contained seven nodes assigned to known species in the *C. curvisetus* complex plus two without a reference (Fig. 4.4). Most of the metabarcodes were assigned to *C. curvisetus* 1, 2 and 3, *C. curvisetus* strain SKLMP_YG033 and *C. pseudocurvisetus* (Fig. 4.4). No sequences were found for the barcode *C. curvisetus* 3e (El4A2) and only one for *C. curvisetus* 2c, both from the Red Sea. Many sequences clustered into two closely related nodes lacking barcodes (Fig. 4.4).
Moreover, the species *C. curvisetus* 3 is more closely related to *C. pseudocurvisetus* than other “curvisetus” species; *C. curvisetus* 3e (from the Red Sea) is closely related to *C. pseudocurvisetus* (two base changes) and distantly to *C. curvisetus* 1 (at least 12 mutations from the main edge). This latter node is separated by eight base changes from the other one referring to *C. curvisetus* strain SKLMP_YG033 from Hong Kong.

The haplotype network based on the Tara Oceans dataset (V9 region) contained six nodes assigned to a known *curvisetus* species plus two without a reference (Fig. 4.5). Most of the metabarcodes were assigned to *C. curvisetus* 2, followed by *C. curvisetus* 3, *C. curvisetus*
3e and all the others with comparable abundances. For both species *C. curvisetus 3* and *C. curvisetus 3e*, a peripheral node with considerable abundance separating from the main one was observed and treated separately for further analyses (named *C. curvisetus* sp. 3 and sp. 4 respectively). Strains isolated from the Red Sea (*C. curvisetus 2c* and *C. curvisetus 3e*) were highly represented in terms of sequences in the Tara dataset when compared with the one of OSD.

![TCS haplotype network for the *C. curvisetus* species complex according to Tara Oceans data.](image)

**Fig. 4.5.** TCS haplotype network for the *C. curvisetus* species complex according to Tara Oceans data. The size of the nodes refers to the abundance of the reads. Numbers or codes after *C. curvisetus* species’ names refer to genetically and morphologically defined species within the *C. curvisetus* complex for which references are available (see Gaonkar et al., 2018). *C. curvisetus* sp. 3 and 4 refer to species in the *C. curvisetus* complex for which no reference sequences are available yet.

The comparison of V4 and V9 networks showed minor differences. In the former, the group encompassing the species *C. curvisetus 3*, *C. curvisetus 3e* and *C. pseudocurvisetus* acted as a bridge between *C. curvisetus 1* and *C. curvisetus 2* and the unassigned nodes, whilst *C. curvisetus 2c* did the same for *C. curvisetus 2* and the two unassigned nodes. In the latter, the node attributed to *C. curvisetus 1* was the pivot around which *C. curvisetus 3*,
the groups *C. curvisetus* 2 - *C. curvisetus* 2c and *C. curvisetus* 3e – *C. pseudocurvisetus* were collocated. In both networks, the relationships among main nodes were generally simple (without complex reticulations), indicating substantial lack of gene flow.

The Maximum likelihood tree inferred using the V4 representative sequences of each newly identified putative species plus the references confirmed that the taxa without reference barcodes, here indicated as *C. curvisetus* sp. 1 and sp. 2, are likely to constitute at least one new species (Fig. 4.6). These are closely related and share a common ancestor with *C. curvisetus* group 2 (Fig. 4.6).

![Fig. 4.6. Maximum Likelihood tree of the *C. curvisetus* species complex based on representative sequences of V4 data. Numbers at the basis of nodes indicate the support to branches after 1000 bootstrap replicates.](image)

For the V9 tree, due to the shortness of the fragment, it was difficult to make hypotheses on the nature of the newly discovered taxa as well as about their phylogenetic relationships with other *curvisetus* species (Fig. 4.7). However, *C. curvisetus* sp. 3 seemed to be more differentiated to its sister taxon (*C. curvisetus* 3, 100 BS) than *C. curvisetus* sp. 4 to its own (*C. curvisetus* 3e, 61 BS, Fig. 4.7).
Fig. 4.7. Maximum Likelihood tree of the *C. curvisetus* species complex based on representative sequences of V9 data. Numbers at the basis of nodes indicate the support to branches after 1000 bootstrap replicates.

4.3.3. *Genetic differentiation and variability*

Genetic distances across inferred species for V4 and V9 regions were different in terms of absolute values, but the proportions were comparable (Table 4.2). For V4, the lowest interspecies genetic distance values were between *C. curvisetus* 3e and *C. pseudocurvisetus* (0.007) and *C. curvisetus* sp. 1 and sp. 2 (0.008, Table 4.2A), whilst the highest between *C. curvisetus* 1 and 2 (0.107) and *C. curvisetus* 2 and *C. curvisetus* strain SKLMP_YG033 (0.105) (Table 4.2A). For V9, values ranged from 0.368 (genetic distance between *C. curvisetus* 2c and 3) to 0.022 (*C. curvisetus* 3e and sp. 4) (Table 4.2B). For both V4 and V9 regions, the highest value of intraspecific divergence (0.105 and 0.049 respectively) was not lower than the minimum value of interspecific divergence (0.007 and 0.022 respectively). The lowest interspecies distances were lower than maxima intraspecies ones. Therefore, no threshold value was found within the complex to distinguish between inter- and intra-specific variability (barcoding gap).
Table 4.2. Pair-wise genetic differentiation between *C. curvisetus* species in OSD (A) and Tara Oceans (B) datasets. Genetic distances were calculated using the Jukes-Cantor model.

(A)

<table>
<thead>
<tr>
<th></th>
<th>C. curvisetus 1</th>
<th>C. curvisetus 2</th>
<th>C. curvisetus 2c</th>
<th>C. curvisetus 3</th>
<th>C. curvisetus 3e</th>
<th>C. curvisetus SKLMP_YG033</th>
<th>C. pseudocurvisetus</th>
<th>C. curvisetus sp. 1</th>
<th>C. curvisetus sp. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. curvisetus 1</td>
<td>-</td>
<td>0.107</td>
<td>-</td>
<td>0.054</td>
<td>0.018</td>
<td>0.062</td>
<td>0.083</td>
<td>0.083</td>
<td>0.085</td>
</tr>
<tr>
<td>C. curvisetus 2</td>
<td>0.098</td>
<td>0.024</td>
<td>-</td>
<td>0.034</td>
<td>0.105</td>
<td>0.063</td>
<td>0.043</td>
<td>0.043</td>
<td>0.046</td>
</tr>
<tr>
<td>C. curvisetus 2c</td>
<td>0.083</td>
<td>0.072</td>
<td>0.069</td>
<td>0.022</td>
<td>0.094</td>
<td>0.049</td>
<td>0.030</td>
<td>0.030</td>
<td>0.032</td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>0.083</td>
<td>0.072</td>
<td>0.069</td>
<td>0.022</td>
<td>0.094</td>
<td>0.049</td>
<td>0.030</td>
<td>0.030</td>
<td>0.032</td>
</tr>
<tr>
<td>C. curvisetus 3e</td>
<td>0.054</td>
<td>0.034</td>
<td>0.022</td>
<td>0.025</td>
<td>0.086</td>
<td>0.023</td>
<td>0.014</td>
<td>0.014</td>
<td>0.018</td>
</tr>
<tr>
<td>C. curvisetus SKLMP_YG033</td>
<td>0.018</td>
<td>0.105</td>
<td>0.094</td>
<td>0.086</td>
<td>0.054</td>
<td>0.063</td>
<td>0.063</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C. pseudocurvisetus</td>
<td>0.062</td>
<td>0.063</td>
<td>0.049</td>
<td>0.023</td>
<td>0.007</td>
<td>0.063</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C. curvisetus sp. 1</td>
<td>0.083</td>
<td>0.043</td>
<td>0.030</td>
<td>0.054</td>
<td>0.014</td>
<td>0.084</td>
<td>0.034</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>C. curvisetus sp. 2</td>
<td>0.085</td>
<td>0.046</td>
<td>0.032</td>
<td>0.057</td>
<td>0.018</td>
<td>0.086</td>
<td>0.037</td>
<td>0.008</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>C. curvisetus 1</td>
<td>C. curvisetus 2</td>
<td>C. curvisetus 2c</td>
<td>C. curvisetus 3</td>
<td>C. curvisetus sp. 3</td>
<td>C. curvisetus 3e</td>
<td>C. curvisetus sp. 4</td>
<td>C. pseudocurvisetus</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>-----------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>---------------------</td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 2</td>
<td>0.155</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 2c</td>
<td>0.189</td>
<td>0.134</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>0.181</td>
<td>0.241</td>
<td>0.368</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus sp. 3</td>
<td>0.181</td>
<td>0.229</td>
<td>0.354</td>
<td>0.038</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus 3e</td>
<td>0.038</td>
<td>0.179</td>
<td>0.154</td>
<td>0.204</td>
<td>0.204</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. curvisetus sp. 4</td>
<td>0.062</td>
<td>0.204</td>
<td>0.165</td>
<td>0.210</td>
<td>0.237</td>
<td>0.022</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. pseudocurvisetus</td>
<td>0.079</td>
<td>0.203</td>
<td>0.191</td>
<td>0.204</td>
<td>0.230</td>
<td>0.079</td>
<td>0.073</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Within each species, the mean evolutionary divergence over sequence pairs ranged from 0.000 (C. curvisetus 2c) to 0.055 (C. curvisetus 3e) for V4 region and from 0.000 (C. curvisetus sp. 3) to 0.017 (C. curvisetus 2) for V9 (Table 4.3).

Table 4.3. Average evolutionary divergence over sequence pairs within species. The number of base substitutions per site from averaging over all sequence pairs within each group are shown. Analyses were conducted using the Jukes-Cantor model. Dashes refer to species absent in that dataset.

<table>
<thead>
<tr>
<th>Species</th>
<th>Divergence V4 region</th>
<th>Divergence V9 region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Min</td>
</tr>
<tr>
<td>C. curvisetus 1</td>
<td>0.009</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus 2</td>
<td>0.008</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus 2c</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus 3</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus 3e</td>
<td>0.010</td>
<td>0.003</td>
</tr>
<tr>
<td>C. curvisetus SKLMP_YG033</td>
<td>0.013</td>
<td>0.003</td>
</tr>
<tr>
<td>C. curvisetus sp. 1</td>
<td>0.008</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus sp. 2</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>C. curvisetus sp. 3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C. curvisetus sp. 4</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C. pseudocurvisetus</td>
<td>0.010</td>
<td>0.003</td>
</tr>
</tbody>
</table>

4.3.4. Global distribution of taxa belonging to the C. curvisetus species complex

The plotting of occurrence data gathered from OSD and Tara Oceans metabarcoding data revealed that the species complex is cosmopolitan, and occurs in both coastal and open ocean waters at all latitudes from northern to southern hemisphere (Fig. 4.8A and B).
Fig. 4.8. Distribution of the C. curvisetus species complex in Longhurst provinces. (A) OSD data; (B) Tara Oceans data. Green dots indicate presence of the taxa in that station, whilst red dots indicate absence.

However, some species showed a specific pattern of occurrence and abundance across the different datasets. For instance, C. curvisetus 1 was found and revealed to be mostly abundant in polar (ARCT, BPLR) and temperate provinces (NADR, NECS, SSTC), whilst C. curvisetus 2 a typical generalist species (Fig. 4.9 and Fig. 4.10). Some species were rare in some datasets (e.g. C. curvisetus 2c in OSD) and completely absent in others (e.g. C. curvisetus sp. 1 and sp. 2 in Tara Oceans).

In the specific case of closely related taxa (e.g. C. curvisetus 1 - C. curvisetus strain SKLMP YG033 and C. curvisetus sp. 1 and sp. 2 in OSD; C. curvisetus 3 – C. curvisetus sp. 3 and C. curvisetus 3e – C. curvisetus sp. 4 in Tara Oceans), a peculiar occurrence
pattern was observed. In each couple, if the two taxa were sharply separated in the network (e.g. *C. curvisetus* 1 - *C. curvisetus* strain SKLMP YG033), in the heatmap I observed one of these occupying different provinces with opposite and generally of different environmental characteristics (e.g. tropical vs. temperate).
Fig. 4.9. Heatmap showing the abundance of *C. curvisetus* spp. in each Longhurst province according to OSD data. Abundance refers to the number of reads. Species are ordered according to phylogenetic position.
Fig. 4.10. Heatmap showing the abundance of *C. curvisetus* spp. in each Longhurst province according to Tara Oceans data. Abundance refers to the number of reads. Species are ordered according to phylogenetic position.
4.4. Discussion

In this chapter, I showed how the study of a cryptic species complex can be enhanced by the combining classical evolutionary approaches and huge amounts of diversity information contained in global metabarcoding datasets. In particular, I used the *C. curvisetus* species complex as case study and combined evolutionary approaches (haplotype networks, phylogenetic relationships, and genetic distances) with the most comprehensive and complementary global metabarcoding datasets available, the predominantly coastal OSD and the mainly oceanic Tara Oceans. The results obtained show the enormous potential of the integration of such methodologies for phylo- and biogeographic studies.

4.4.1. Phylogenetic relationships among taxa belonging to the *C. curvisetus* species complex

The use of haplotype networks allows a clear assessment and visualisation of the relationships among the taxa within the complex, which are not straightforward in the V4 and V9 trees in Gaonkar et al. (2018). The latter were based on references of a few specimens per species whereas the haplotype networks and their relative abundances provide insights in the populations of those species. In addition, phylogenetic trees are constrained in visualising speciation events as bifurcating processes, whereas haplotype networks can model evolution in a reticulated manner, best fitting cases of recent divergence as may occur in species complexes. The V4 and V9 TCS networks were presented slight differences related to different length of the regions (~384 and ~105 bp respectively). These differences are also found in the V4 and V9 phylogenetic trees in Gaonkar et al. (2018). Overall, the signal was consistent between the two datasets, allowing the inference of at least eight different species within the *C. curvisetus* species complex.
The network approach revealed also to be useful at detecting putative new species or isolated populations. These findings are supported by the high bootstrap values recovered in the tree obtained using the reference barcodes of *curvisetus* species and the representatives of unassigned nodes. The same is found in Tara Oceans V9 data for two taxa (*C. curvisetus* sp. 3 and 4), despite the fact that the separation from the other nodes is not as straightforward as in the OSD V4 network. Using the same OSD and Tara Oceans datasets but different taxa and a non-evolutionary approach (swarm OTU clustering), Pargana (2017) found a new clade close to *Leptocylindrus danicus* and several clades within *L. minimus* of uncertain taxonomic identity. Such inference of taxa from signature sequences (metabarcoding data) is just the first step of the process; the next step is to link such anonymous sequences to a reference of a specific taxon in order to be validated. This approach is called “reverse taxonomy” (Markmann and Tautz, 2005). In the case of metabarcoding data, the validation of anonymous sequences in the field is favoured by the use of abundance tables, which contain the information of occurrence and abundance in each sampled locality.

In general, the shape and size of nodes in the network, together with the number and structure of edges connecting them, can be considered as a primary hypothesis for species/population delimitation based on gene flow. In my networks, the signal of active gene flow between inferred species is weak but present. The absence of barcoding gap confirmed that signal, suggesting that the genetic barriers in part of the complex are not complete.

4.4.2. Distribution of taxa belonging to the C. curvisetus species complex

*Chaetoceros curvisetus* was reported by Gran (1897) as a common inhabitant of the Atlantic Ocean and the Baltic Sea, with peaks of abundance in summer and autumn. Hasle and Syvertsen (1996) indicated it as a cosmopolitan species mainly found in temperate and
warm waters. This was confirmed by my results in chapter III. In Chinese waters, the only references about the distribution of such species are the ones related to harmful algal blooms (Wang and Wu, 2009; Zhen et al., 2009), during which the species is particularly abundant. However, no production of toxins is known to date in any “curvisetus” species. Instead, Hasle and Syvertsen (1996) considered *C. pseudocurvisetus* as an inhabitant of warm waters. This finding was partially confirmed by results of my analysis in this chapter and chapter III, in which the species was found not only in the Mediterranean Sea, the nearby Atlantic Ocean and the Indian Ocean, but also in the North Sea.

In general, results of my analysis using OSD and Tara Oceans dataset indicates that the *C. curvisetus* complex is cosmopolitan. Nonetheless, some species showed preference for particular environmental conditions. For example, *C. curvisetus* 1 occurs in cold to temperate waters, with the exception of the Mediterranean Sea. In the Gulf of Naples (Mediterranean Sea), Gaonkar (2017) found this species only during winter, supporting its preferences for cold environments. Similarly, but with an opposite trend, *C. curvisetus* strain SKLMP is only found in tropical seas. This is also interesting from the phylogenetic point of view, since these two taxa are sister species. This marked difference in climate preference between closely related species was also observed for other members of the complex, e.g. *C. curvisetus* 1 - *C. curvisetus* SKLMP, *C. curvisetus* sp. 1 – sp. 2, *C. curvisetus* 3 - *C. curvisetus* sp. 3 and *C. curvisetus* 3e – *C. curvisetus* sp. 4. The aforementioned pattern was more evident for sister taxa that were clearly separated in the network (*C. curvisetus* 1 - *C. curvisetus* SKLMP) than in others where gene flow was still on-going or the separation was recent (*C. curvisetus* sp. 1 – sp. 2, *C. curvisetus* 3 - *C. curvisetus* sp. 3 and *C. curvisetus* 3e – *C. curvisetus* sp. 4).

Other studies involving cryptic species have shown similar results. In the genus *Skeletonema* for example, the widely distributed species *Skeletonema costatum* sensu lato revealed to be a complex of several species (Sarno et al., 2005; 2007; Zingone et al., 2005).
Several of these appeared to be widely distributed as well, but within some broad climatological boundaries (cool-temperate *S. japonicum*; temperate to tropical *S. tropicum*; Kooistra et al., 2008). However, a few others such as *S. grethae* appeared to be more regional and apparently absent in climatologically comparable regions (Kooistra et al., 2008). More in general, Hasle (1976) already noticed that morphologically closely related diatom species were often found in different biogeographic regions. In the genera *Nitzschia* and *Thalassiosira*, she observed species only from the cold-water species of the Northern and Southern Hemispheres as well as from warm-water species, and cosmopolitan ones (Hasle, 1976). In *Leptocylindrus*, most species were found to be widespread across coastal waters (e.g. *L. convexus*, *L. danicus* and *L. hargravesii*) with the only exception of *L. minimus*, which was restricted to cold waters of the Northern Hemisphere (Pargana, 2017).

According to the "everything is everywhere" hypothesis (Baas Becking, 1934), most microbes form populations large enough to migrate efficiently and accumulate mutations that could be beneficial in particular environments (Shapiro et al., 2016). Speciation in the microbial world is therefore expected to involve little drift and geographical separation and more selection (Shapiro et al., 2016). Diatoms, for example, are believed to exhibit high intraspecific variability, which would be key for their adaptation to different environments (Godhe and Rynearson, 2017). It is possible that different strains of a species already possess beneficial mutations allowing them to adapt to different environments due to high intraspecific variability (see Godhe and Rynearson, 2017). Once a different environment is reached, some strains would be favoured by natural selection and, over time, accumulate other mutations that will finally differentiate them from the parental population, leading to speciation. In this context, the adaptation to different environments would be the factor triggering speciation in diatoms. In agreement with Hasle (1976), which surveyed the biogeographic trends of 26 diatom species, in this study I have observed that sister *C. curvisetus* species (e.g. *C. curvisetus* 1 - *C. curvisetus* SKLMP; *C. curvisetus* sp. 1 – sp. 2)
tend to be found in different biogeographic provinces with generally opposite environmental conditions (e.g. cold vs. warm environments). Data are far from conclusive to assert that adaptation to different environmental conditions triggers speciation in diatoms, but I have added other elements to support this hypothesis. Furthermore, all these studies emphasise once more the importance of correct identification of taxa at the species level to make adequate inferences on their distribution and ecology. In this context, metabarcoding data accompanied by a well-represented reference barcode library are a useful tool for primary hypotheses of species distribution.

4.4.3. Considerations on sequence variation in metabarcoding data

In this work, I have used the accepted barcode for protists (V4 region, Pawlowski et al., 2012) and the V9 region to study a cryptic species complex. Instead of a classical, Sanger-based approach of a multitude of geographic strains, I have used metabarcoding datasets (OSD and Tara Oceans), to take advantage of the data available for many sampling localities across the globe, which would have been difficult to sample with a classical sampling approach of establishing strains. As consequence of this choice, I had to work with thousands of sequences. Indeed, differently from a Sanger sequencing, which provides a single sequence as output (a consensus of all the amplified products), high-throughput techniques sequence every single molecule. Furthermore, since the 18S gene occurs in hundreds to thousands of copies within the genome, and sometime on multiple chromosomes (Alvarez and Wendel, 2003), the number of sequences to handle was even bigger. Such rDNA copies are expected to be homogenised by concerted evolution over time, but empirical studies suggest that this process is not perfect and multiple, polymorphic copies can persist within the genome (Alverson and Kolnick, 2005). When using environmental samples, 18S copies from different cistrons, chromosomes and
individuals are mixed together, rendering it difficult to discern between intra- and interspecific variability.

Using the network approach and simple criteria to infer sequences to a species (see M&M section), I have demonstrated that this is not an issue. Indeed, all these sequences resulting from the apparent failure of concerted evolution to achieve complete homogenisation, from geographic variability, from PCR and sequencing errors are arranged around the main node in which the “dominant haplotype” is located. All these dominant and peripheral haplotypes contribute to the definition of the species’ overall genetic variation for this marker region. The dominant haplotype is here defined as the most abundant haplotype for a specific taxon, which also corresponds to the Sanger sequence in the species for which reference barcodes are available.

Furthermore, it is possible that the 18S copies escaping concerted evolution retain ancestral polymorphisms that can help assessing phylogenetic relationships among species.

In this context, I showed that the use of a multi-copy gene is not a disadvantage, but all these copies contribute to the evaluation of inter- and intra-species variation.

References


Steiner, F. M., Csősz, S., Markó, B., Gamisch, A., Rinnhofer, L., Folterbauer, C., ... Schlick-Steiner, B. C. (2018). Turning one into five: Integrative taxonomy
uncover complex evolution of cryptic species in the harvester ant 
Messor "structor". *Molecular Phylogenetics and Evolution, 127*, 387-404.


Table A4.1. List of OSD and Tara Oceans sites in which were found metabarcodes validated as *C. curvisetus* spp.

<table>
<thead>
<tr>
<th>OSD</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Station</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSD2</td>
<td>-3.938</td>
<td>48.778</td>
<td>TARA_004</td>
<td>-6.553</td>
<td>36.563</td>
</tr>
<tr>
<td>OSD4</td>
<td>14.25</td>
<td>40.808</td>
<td>TARA_005</td>
<td>-4.406</td>
<td>36.030</td>
</tr>
<tr>
<td>OSD5</td>
<td>24.99</td>
<td>35.661</td>
<td>TARA_006</td>
<td>-4.251</td>
<td>36.529</td>
</tr>
<tr>
<td>OSD6</td>
<td>2.8</td>
<td>41.667</td>
<td>TARA_007</td>
<td>1.948</td>
<td>37.031</td>
</tr>
<tr>
<td>OSD13</td>
<td>27.909</td>
<td>43.176</td>
<td>TARA_008</td>
<td>3.966</td>
<td>38.011</td>
</tr>
<tr>
<td>OSD14</td>
<td>3.15</td>
<td>42.49</td>
<td>TARA_009</td>
<td>5.820</td>
<td>39.112</td>
</tr>
<tr>
<td>OSD22</td>
<td>5.175</td>
<td>42.226</td>
<td>TARA_010</td>
<td>2.865</td>
<td>40.668</td>
</tr>
<tr>
<td>OSD24</td>
<td>-2.88</td>
<td>35.193</td>
<td>TARA_011</td>
<td>2.798</td>
<td>41.666</td>
</tr>
<tr>
<td>OSD26</td>
<td>-5.75</td>
<td>35.82</td>
<td>TARA_012</td>
<td>7.899</td>
<td>43.348</td>
</tr>
<tr>
<td>OSD29</td>
<td>-80.283</td>
<td>27.469</td>
<td>TARA_014</td>
<td>12.858</td>
<td>39.902</td>
</tr>
<tr>
<td>OSD37</td>
<td>-80.093</td>
<td>26.103</td>
<td>TARA_016</td>
<td>15.454</td>
<td>37.398</td>
</tr>
<tr>
<td>OSD38</td>
<td>-80.784</td>
<td>24.745</td>
<td>TARA_017</td>
<td>14.306</td>
<td>36.258</td>
</tr>
<tr>
<td>OSD43</td>
<td>-117.257</td>
<td>32.867</td>
<td>TARA_018</td>
<td>14.288</td>
<td>35.756</td>
</tr>
<tr>
<td>OSD50</td>
<td>-1.925</td>
<td>43.333</td>
<td>TARA_019</td>
<td>13.865</td>
<td>34.216</td>
</tr>
<tr>
<td>OSD51</td>
<td>-82.266</td>
<td>9.348</td>
<td>TARA_020</td>
<td>14.973</td>
<td>34.451</td>
</tr>
<tr>
<td>OSD54</td>
<td>-69.641</td>
<td>43.844</td>
<td>TARA_022</td>
<td>17.400</td>
<td>39.729</td>
</tr>
<tr>
<td>OSD55</td>
<td>-69.578</td>
<td>43.86</td>
<td>TARA_023</td>
<td>17.729</td>
<td>42.176</td>
</tr>
<tr>
<td>OSD58</td>
<td>-76.671</td>
<td>34.718</td>
<td>TARA_024</td>
<td>17.956</td>
<td>42.457</td>
</tr>
<tr>
<td>OSD101</td>
<td>-16.711</td>
<td>32.742</td>
<td>TARA_025</td>
<td>19.421</td>
<td>39.333</td>
</tr>
<tr>
<td>OSD102</td>
<td>-16.91</td>
<td>32.646</td>
<td>TARA_026</td>
<td>20.188</td>
<td>38.431</td>
</tr>
<tr>
<td>OSD107</td>
<td>-9.38</td>
<td>39.14</td>
<td>TARA_030</td>
<td>32.789</td>
<td>33.929</td>
</tr>
<tr>
<td>OSD108</td>
<td>-8.966</td>
<td>38.757</td>
<td>TARA_031</td>
<td>34.819</td>
<td>27.151</td>
</tr>
<tr>
<td>OSD109</td>
<td>-9.012</td>
<td>38.677</td>
<td>TARA_032</td>
<td>37.254</td>
<td>23.391</td>
</tr>
<tr>
<td>OSD110</td>
<td>-8.869</td>
<td>40.145</td>
<td>TARA_033</td>
<td>38.218</td>
<td>22.057</td>
</tr>
<tr>
<td>OSD116</td>
<td>-9.219</td>
<td>39.415</td>
<td>TARA_036</td>
<td>63.524</td>
<td>20.824</td>
</tr>
<tr>
<td>OSD117</td>
<td>-7.504</td>
<td>37.167</td>
<td>TARA_038</td>
<td>64.576</td>
<td>19.017</td>
</tr>
<tr>
<td>OSD124</td>
<td>135.121</td>
<td>34.324</td>
<td>TARA_039</td>
<td>66.463</td>
<td>18.647</td>
</tr>
<tr>
<td>OSD131</td>
<td>27.401</td>
<td>42.245</td>
<td>TARA_040</td>
<td>67.984</td>
<td>17.500</td>
</tr>
<tr>
<td>OSD145</td>
<td>3.119</td>
<td>51.361</td>
<td>TARA_041</td>
<td>70.011</td>
<td>14.582</td>
</tr>
<tr>
<td>OSD147</td>
<td>81.052</td>
<td>8.522</td>
<td>TARA_042</td>
<td>73.919</td>
<td>5.992</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>OSD148</td>
<td>8.149</td>
<td>53.581</td>
<td>TARA_043</td>
<td>73.489</td>
<td>4.660</td>
</tr>
<tr>
<td>OSD153</td>
<td>-7.973</td>
<td>36.998</td>
<td>TARA_044</td>
<td>71.520</td>
<td>2.806</td>
</tr>
<tr>
<td>OSD154</td>
<td>-1.167</td>
<td>44.667</td>
<td>TARA_045</td>
<td>71.710</td>
<td>0.941</td>
</tr>
<tr>
<td>OSD155</td>
<td>10.599</td>
<td>59.816</td>
<td>TARA_046</td>
<td>73.162</td>
<td>-0.659</td>
</tr>
<tr>
<td>OSD156</td>
<td>10.72</td>
<td>59.9</td>
<td>TARA_047</td>
<td>72.164</td>
<td>-2.042</td>
</tr>
<tr>
<td>OSD157</td>
<td>10.628</td>
<td>59.622</td>
<td>TARA_048</td>
<td>73.162</td>
<td>-0.659</td>
</tr>
<tr>
<td>OSD158</td>
<td>-25.19</td>
<td>37.433</td>
<td>TARA_049</td>
<td>59.504</td>
<td>-16.808</td>
</tr>
<tr>
<td>OSD159</td>
<td>-4.552</td>
<td>48.359</td>
<td>TARA_050</td>
<td>56.795</td>
<td>-21.476</td>
</tr>
<tr>
<td>OSD162</td>
<td>-2.103</td>
<td>56.963</td>
<td>TARA_051</td>
<td>54.283</td>
<td>-21.476</td>
</tr>
<tr>
<td>OSD163</td>
<td>-2.973</td>
<td>58.957</td>
<td>TARA_052</td>
<td>53.508</td>
<td>-17.023</td>
</tr>
<tr>
<td>OSD166</td>
<td>2.9</td>
<td>43.433</td>
<td>TARA_053</td>
<td>46.923</td>
<td>-13.070</td>
</tr>
<tr>
<td>OSD173</td>
<td>3.14</td>
<td>51.441</td>
<td>TARA_054</td>
<td>45.226</td>
<td>-12.813</td>
</tr>
<tr>
<td>OSD177</td>
<td>2.702</td>
<td>51.186</td>
<td>TARA_055</td>
<td>42.742</td>
<td>-17.026</td>
</tr>
<tr>
<td>OSD60</td>
<td>-79.168</td>
<td>33.323</td>
<td>TARA_058</td>
<td>42.320</td>
<td>-17.455</td>
</tr>
<tr>
<td>OSD64</td>
<td>30.776</td>
<td>46.442</td>
<td>TARA_062</td>
<td>40.182</td>
<td>-22.339</td>
</tr>
<tr>
<td>OSD69</td>
<td>12.26</td>
<td>45.457</td>
<td>TARA_064</td>
<td>37.929</td>
<td>-29.508</td>
</tr>
<tr>
<td>OSD70</td>
<td>12.438</td>
<td>45.414</td>
<td>TARA_065</td>
<td>26.334</td>
<td>-35.226</td>
</tr>
<tr>
<td>OSD71</td>
<td>170.771</td>
<td>-45.744</td>
<td>TARA_066</td>
<td>18.016</td>
<td>-34.905</td>
</tr>
<tr>
<td>OSD74</td>
<td>-8.667</td>
<td>41.142</td>
<td>TARA_068</td>
<td>4.620</td>
<td>-31.039</td>
</tr>
<tr>
<td>OSD76</td>
<td>12.935</td>
<td>43.948</td>
<td>TARA_072</td>
<td>-18.006</td>
<td>-8.691</td>
</tr>
<tr>
<td>OSD77</td>
<td>13.073</td>
<td>43.851</td>
<td>TARA_076</td>
<td>-35.231</td>
<td>-21.029</td>
</tr>
<tr>
<td>OSD78</td>
<td>13.595</td>
<td>43.57</td>
<td>TARA_078</td>
<td>-43.323</td>
<td>-30.158</td>
</tr>
<tr>
<td>OSD81</td>
<td>-7.973</td>
<td>37.005</td>
<td>TARA_080</td>
<td>-51.952</td>
<td>-40.698</td>
</tr>
<tr>
<td>OSD91</td>
<td>-9.037</td>
<td>32.747</td>
<td>TARA_081</td>
<td>-52.214</td>
<td>-44.497</td>
</tr>
<tr>
<td>OSD92</td>
<td>-7.701</td>
<td>33.584</td>
<td>TARA_082</td>
<td>-58.012</td>
<td>-47.165</td>
</tr>
<tr>
<td>OSD94</td>
<td>-2.215</td>
<td>35.086</td>
<td>TARA_083</td>
<td>-65.023</td>
<td>-54.418</td>
</tr>
<tr>
<td>OSD95</td>
<td>103.917</td>
<td>1.268</td>
<td>TARA_085</td>
<td>-49.503</td>
<td>-62.176</td>
</tr>
<tr>
<td>OSD97</td>
<td>-28.602</td>
<td>38.53</td>
<td>TARA_088</td>
<td>-56.806</td>
<td>-63.386</td>
</tr>
<tr>
<td>OSD98</td>
<td>-28.13</td>
<td>38.64</td>
<td>TARA_092</td>
<td>-71.977</td>
<td>-33.697</td>
</tr>
<tr>
<td>OSD99</td>
<td>TARA_094</td>
<td>-87.093</td>
<td>-32.765</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD100</td>
<td>TARA_096</td>
<td>-101.268</td>
<td>-29.655</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD101</td>
<td>TARA_098</td>
<td>-110.992</td>
<td>-26.261</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD102</td>
<td>TARA_100</td>
<td>-96.283</td>
<td>-13.162</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD103</td>
<td>TARA_102</td>
<td>-85.270</td>
<td>-5.218</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD104</td>
<td>TARA_106</td>
<td>-84.620</td>
<td>0.037</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSD105</td>
<td>TARA_109</td>
<td>-84.545</td>
<td>1.800</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_110</td>
<td>-84.616</td>
<td>-1.913</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_113</td>
<td>-134.920</td>
<td>-23.114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_114</td>
<td>-134.912</td>
<td>-23.130</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_115</td>
<td>-134.931</td>
<td>-23.216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_116</td>
<td>-134.931</td>
<td>-23.217</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_118</td>
<td>-135.009</td>
<td>-23.129</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_120</td>
<td>-134.912</td>
<td>-23.012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_123</td>
<td>-140.304</td>
<td>-8.878</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_125</td>
<td>-142.610</td>
<td>-8.890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_126</td>
<td>-151.208</td>
<td>-11.975</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_128</td>
<td>-153.305</td>
<td>-0.469</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_130</td>
<td>-152.462</td>
<td>11.265</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_131</td>
<td>-158.052</td>
<td>22.746</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_133</td>
<td>-127.750</td>
<td>35.343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_134</td>
<td>-121.986</td>
<td>32.667</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_135</td>
<td>-121.832</td>
<td>32.983</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_137</td>
<td>-116.699</td>
<td>14.161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_138</td>
<td>-103.017</td>
<td>6.216</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_140</td>
<td>-79.312</td>
<td>7.471</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_141</td>
<td>-80.086</td>
<td>9.834</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_142</td>
<td>-88.417</td>
<td>25.602</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_143</td>
<td>-79.682</td>
<td>29.885</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_144</td>
<td>-72.815</td>
<td>36.369</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_145</td>
<td>-70.076</td>
<td>39.163</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_146</td>
<td>-71.248</td>
<td>34.731</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_147</td>
<td>-66.533</td>
<td>32.954</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_148</td>
<td>-64.145</td>
<td>31.782</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_149</td>
<td>-49.840</td>
<td>34.098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_150</td>
<td>-37.102</td>
<td>35.800</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_151</td>
<td>-28.801</td>
<td>36.194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_152</td>
<td>-16.662</td>
<td>43.668</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_153</td>
<td>-16.564</td>
<td>44.034</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_155</td>
<td>-16.755</td>
<td>54.597</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_158</td>
<td>0.374</td>
<td>67.193</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_163</td>
<td>1.689</td>
<td>76.078</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_168</td>
<td>44.126</td>
<td>72.582</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value1</td>
<td>Value2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_173</td>
<td>75.345</td>
<td>78.939</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_175</td>
<td>66.384</td>
<td>79.343</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_178</td>
<td>73.235</td>
<td>77.234</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_180</td>
<td>75.459</td>
<td>75.172</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_188</td>
<td>91.725</td>
<td>78.304</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_189</td>
<td>116.482</td>
<td>78.022</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_191</td>
<td>160.961</td>
<td>71.549</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_193</td>
<td>174.901</td>
<td>71.115</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_194</td>
<td>-168.518</td>
<td>73.336</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_196</td>
<td>-154.934</td>
<td>71.895</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_201</td>
<td>-85.729</td>
<td>74.329</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_205</td>
<td>-71.952</td>
<td>72.423</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_208</td>
<td>-51.578</td>
<td>69.107</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TARA_210</td>
<td>-55.985</td>
<td>61.544</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter V

Concerted evolution

in Chaetoceros
5.1. Introduction

The first DNA reannealing and hybridisation studies conducted in the mid-1960-70s to unveil the structure and organisation of eukaryotic genomes showed that a large fraction of them was composed of repetitive regions (Britten and Waring, 1965; Britten and Kohne, 1968). The subsequent study of such regions revealed that, when comparing repetitive DNA families, there was greater sequence similarity within species than between species (Brown et al., 1972; Elder and Turner, 1995). Such observation was incompatible with the then common model of divergent evolution, according to which the differences in nucleotide sequence between different repeats of the same species were expected to be as large as those between repeats of different species (Nei and Rooney, 2005). Therefore, there had to be a mechanism responsible for the homogenisation of such sequences within an individual organism. The expression “concerted evolution” (Zimmer et al., 1980) was coined to indicate this phenomenon, by which an individual member of a gene family evolves in the same (concerted) way as all the other members of the family (Graur and Li, 1999).

The best-known example of concerted evolution is the rDNA cistron (Ganley and Kobayashi, 2007), but also other genes and non-coding regions (e.g. globins, immunoglobulins, heat-shock genes, histones) are known to evolve in this way (Long and Dawid, 1980; Liebhaber et al., 1981; Coen et al., 1982; Gojobori and Nei, 1984).

The exact mechanisms determining concerted evolution are still unclear. However, two processes, gene conversion and unequal crossing-over, are considered responsible for sequence homogeneity, the latter also causing fluctuations in number over evolutionary time (Lindegren, 1953; Holliday, 1964; Charlesworth et al., 1986). Despite this, the mechanism is not perfect and cases of deviations from such homogenisation have been detected in animals (Nikolaidis and Nei, 2004; Andrea et al., 2006), fungi (Li et al., 2013),
and especially in plants (Harpke and Peterson, 2006; Zheng et al., 2008; Xiao et al., 2010; Vilnet et al., 2012; Xu et al., 2017).

The extent of such non-homogenisation is particularly important in the case of the rDNA cistron, since it is the classical target for DNA barcoding studies in some taxa as fungi and protists (Pawlowski et al., 2012; Schoch et al., 2012; Stoeck et al., 2014). Therefore, understanding the inheritance of ribosomal genes and spacers is vital for taxonomic and systematic studies involving them.

So far, exceptions to concerted evolution have been spotted detecting noise in electropherograms and then cloning and sequencing subsamples of amplified products (e.g. Pillet et al., 2012; Naidoo et al., 2013). The resulting sequences were then put on a phylogenetic tree together with the ones from closely related species to ascertain the degree of similarity within and among species.

This approach has two main limitations: first, the number of detectable variants is constrained by the number of clones that are sequenced; second, there is no information about the abundance of each variant. Nowadays, metabarcoding techniques allow sequencing thousands of copies of a target region from environmental samples, bulk communities and even single specimens. The latter approach can be particularly useful to study concerted evolution.

A temporal metabarcoding analysis conducted in the LTER MareChiara (Gulf of Naples, Italy) across three years (48 dates) to unveil species diversity within the diatom family Chaetocerotaceae (Gaonkar, 2017) showed the following pattern. When a phylogenetic tree based on V4-18S metabarcodes was inferred, many terminal clades contained from few to tens of haplotypes, one of which was far more abundant than the others. Such a sequence, called “dominant haplotype”, was identical or nearly identical to the reference sequence (Sanger), when available, for that clade (Gaonkar, 2017). Furthermore, the relationship among “dominant” and “minor” haplotypes across species was consistent: when plotted on
a logarithmic scale, the dominant haplotype was of two orders more abundant than the others were. The number of detectable minor haplotypes in the environmental sample was a function of the abundance of the dominant one: the more abundant the latter, the bigger the number of minor haplotypes.

However, the author did not discuss if such “minor” haplotypes were PCR or sequencing errors as well as intra- or inter-individual (strain) variation, but argued that such pattern can be considered as “the result of an equilibrium between the appearance of novel haplotypes, random drift, and the homogenizing effect of concerted evolution” (Gaonkar, 2017).

Based on the theory of concerted evolution, I formulated the hypothesis that the patterns observed at temporal scale in the 48 samples of MareChiara dataset were related to this phenomenon. To confirm or reject that hypothesis, I designed an experiment based on HTS of V4 region of 18S gene from single strains of different Chaetoceros species to test:

i) If the proportion between dominant and minor haplotypes in the environmental samples is also observed within individual strains;

ii) The identity between the sequence of the dominant haplotype in the HTS single strain both with the Sanger reference and with the sequence of the dominant haplotype in environmental metabarcoding for each species;

iii) The identity between the sequences with low abundance (minor haplotypes) found in the HTS single strain with the sequences found in the environmental samples;

iv) The pattern of phylogenetic networks for each Chaetoceros species using the metabarcoding dataset generated from the temporal distribution (48 dates).
5.2. Materials and Methods

5.2.1. Selection of taxa to study concerted evolution

In order to answer the aforementioned questions, I used part of the data from the thesis of Chetan Gaonkar (Gaonkar, 2017) and the metabarcoding data of Chaetocerotaceae from the LTER MareChiara (Gulf of Naples) deposited in GenBank at the accession numbers MK938374-MK940235 (414,041 reads). I started analysing the species *C. curvisetus* 2, from which the pattern of concerted evolution was first hypothesised (see Preface). Then, I used the HTS phylogenetic tree in Gaonkar (2017) inferred from the 48 dates of MareChiara to select other *Chaetoceros* species. In particular, I have chosen: i) a species occurring at high abundance all over the year and so displaying many minor haplotypes (*C. tenuissimus*); ii) a species with a marked seasonality displaying also a few minor haplotypes at high abundances (*C. costatus*); iii) a species displaying a single, lowly abundant, dominant haplotype (*C. anastomosans*); iv) two species without a clear delimitation that occurred in the same clade despite having different reference barcodes, and so with mixed minor haplotypes (*Chaetoceros* sp. Na11C3 and Na26B1). For each species, I selected outgroup taxa (Table 5.1) for subsequent validation of sequences gathered from BLAST analysis. The undescribed species *C*. sp. Na11C3 and *C*. sp. Na26B1 were analysed together because they were in the same clade in the NGS tree of Gaonkar (2017) despite having different barcodes.

Table 5.1. List of outgroup taxa for the validation of *Chaetoceros*-species sequences.

<table>
<thead>
<tr>
<th>Species</th>
<th>Outgroups</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td><em>C</em>. cf. <em>vixvisibilis</em> Na16A3</td>
<td>MG972367</td>
</tr>
<tr>
<td></td>
<td><em>Chaetoceros</em> sp. Na11C3</td>
<td>MG972328</td>
</tr>
<tr>
<td><em>C. costatus</em></td>
<td><em>C. cinctus</em> Ch6A2</td>
<td>KY852264</td>
</tr>
<tr>
<td></td>
<td><em>C. radicans</em> Ch2A2</td>
<td>KY852259</td>
</tr>
</tbody>
</table>
C. curvisetus 2

C. cf. tortissimus Na18C4
C. tortissimus

MG972275
MG972325

Chaetoceros sp. Na11C3 / Na26B1

C. anastosomans Na14C2
C. cf. vivvisibilis Na16A3
Chaetoceros clone HM347543

MG972358
MG972367
HM347543

C. tenuissimus

C. neogracilis 1 RCC2507
C. neogracilis 2 RCC2318
C. neogracilis 4 RCC2016
Chaetoceros sp.
Chaetoceros sp.

KT860998
JN934684
JF794049
AF145226
X85390

5.2.2. Analysis of environmental sequences

I used the metabarcoding data corresponding to 48 environmental samples collected in the LTER-MareChiara (Gulf of Naples, Italy) produced and processed by Gaonkar (2017). These data were sequenced in paired end (2 × 250 bp) on an Illumina MiSeq platform (see Gaonkar 2017 for further details) and are available in GenBank at the accession numbers MK938374-MK940235.

The procedure followed to retrieve sequences of selected species of Chaetoceros in the MareChiara dataset is similar to the one adopted in the previous chapter. In brief, I used the V4 region of my target species and close outgroups as queries for a local BLAST at 95%.

The metabarcodes extracted were then aligned with the references and the outgroup taxa using MAFFT online (Katoh et al., 2017) and a phylogenetic tree was built in FastTree v2.1.8 (Price et al., 2010), using the GTR model. The resulting tree was visualised and modified in Archaeopteryx v0.9901 (Han and Zmasek, 2009) in order to remove sequences clustering within outgroup clades and gather only metabarcodes of the species of interest. The sequences retrieved were considered validated and used to retrieve the info of abundance using mothur v1.41.1 (Schloss et al., 2009).
5.2.3. Single strain HTS

Single strain metabarcoding was performed on: two strains of *C. anastomosans*, four strains of *C. costatus*, four strains of *C. curvisetus* sp. 2, one of *Chaetoceros* sp. Na26B1, two of *Chaetoceros* sp. Na11C3 and three strains of *C. tenuissimus* (Table 5.2).

### Table 5.2. List of strains utilised for single-strain HTS.

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td>Na14C2</td>
</tr>
<tr>
<td></td>
<td>Na14C3</td>
</tr>
<tr>
<td><em>C. costatus</em></td>
<td>Na1A3</td>
</tr>
<tr>
<td></td>
<td>Na32B1</td>
</tr>
<tr>
<td></td>
<td>Ro1B1</td>
</tr>
<tr>
<td></td>
<td>Ro2A2</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 2</td>
<td>Ch5B2</td>
</tr>
<tr>
<td></td>
<td>Na1C1</td>
</tr>
<tr>
<td></td>
<td>Na19A2</td>
</tr>
<tr>
<td></td>
<td>Na20A4</td>
</tr>
<tr>
<td><em>Chaetoceros</em> sp. Na11C3</td>
<td>Na11C3</td>
</tr>
<tr>
<td></td>
<td>Na43A1</td>
</tr>
<tr>
<td><em>Chaetoceros</em> sp. Na26B1</td>
<td>Na26B1</td>
</tr>
<tr>
<td><em>C. tenuissimus</em></td>
<td>GB2a</td>
</tr>
<tr>
<td></td>
<td>Na26A1</td>
</tr>
<tr>
<td></td>
<td>Na44A1</td>
</tr>
</tbody>
</table>

Abbreviations are as follows: Ch = Chile; Na = Naples; Ro = Roscoff. GB2a is a strain from the Gulf of Naples.

For each sample, I performed individual PCR in two steps: a first reaction for the amplification of the target sequence, and a second reaction (using the PCR product of the former one as template) to ligate proprietary adaptor sequence (P1) and unique 10–12 bp long identifier nucleotide key tags (barcodes) compatible with the GeneStudio S5 Ion
Torrent (Life Technologies). The obtained fragment contained all the information required for sequencing and differentiation of samples. The first amplification was conducted using the primers targeting the 18S-V4 region by Stoeck et al. (2010) modified by Piredda et al. (2016). PCRs were conducted in a final volume of 25 μL each containing: 3 ng of DNA, 1x Buffer HF, 0.2 mM dNTPs, 0.5 μM of each primer, 1U of Phusion High-Fidelity DNA polymerase (New England Biolabs Inc, Ipswich, MA) and water to volume. The thermal cycling profiles started with 98 °C for 30 s, followed by 10 cycles of denaturation at 98 °C for 10 s, annealing at 44 °C for 30 s, extension at 72 °C for 15 s, and then additional 15 cycles of denaturation at 98 °C for 10 s, annealing at 62 °C for 30 s and extension at 72 °C for 15 s, with a final extension at 72 °C for 7 min. PCR products (∼470 bp) were visualised on 1.2% agarose gel and purified using the AMPure XP Beads kit (Agencourt Bioscience Corp., Beverly, MA, USA), at a concentration of 1.2× vol/vol, according to manufacturer's instructions. The second PCR was conducted in the same volume and using the same concentrations of reagents (DNA, dNTPs, Buffer and Taq). Adapter P1 was added at a concentration of 50 μM, whilst each barcode of 20 μM. The amplification profile was as follows: initial denaturation at 98 °C for 30 s; 5 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s, extension at 72 °C for 15 s, and then a final extension at 72 °C for 7 min. The success of insertion of adapter and barcode in PCR products was checked by electrophoresis on 1.2% agarose gel (increase of size). Amplified products were purified as above and quantity and quality were determined with the Agilent DNA High Sensitivity Kit on the Bioanalyzer (Agilent) following the manufacturer’s recommendations. Since not all PCRs amplified only the fragment of interest, prior to emulsion PCR an equal amount of all COI products was pooled and processed for fragment size selection (around 500 bp). This was done by running the pooled samples on 1.2% agarose gel together with a size standard and cutting the band of interest, which was then purified using the GenElute™ Gel Extraction Kit (Sigma-Aldrich). Emulsion PCR was
conducted in the Ion Chef System (Life Technologies) using 0.1 fmol/µL of the pool into a reaction volume of 50 µL. Massive-parallel sequencing was carried out using the Ion GeneStudio™ S5 System (Life Technologies).

5.2.4. Data pre-processing and analysis of single-strain HTS

From raw fastq data, adapters and primers were removed with cutadapt (Martin, 2011), allowing a maximum of three mismatches. All reads with a length < 350 bp and quality score < 20 were discarded.

Data obtained with Ion Torrent technology are known to have indel errors of an order of magnitude more frequent than substitution errors (Laehnemann et al., 2016), with most of indel errors caused by homopolymers. Furthermore, the Ion Torrent platform is known to have a higher indel error rate associated with the homopolymer region than the Illumina platform (Loman et al., 2012). In order to overcome this issue, I corrected indel errors using ICC v2.0.1 (Deng et al., 2013). This software starts filtering sequences based on length and quality and then blasts them against a reference. Successively, it retrieves the sequences in windows and proceeds with the correction, which is performed in clusters differing by homopolymer indels. As reference for BLAST, I used, for each species, the V4 region generated by Sanger sequencing of one of the strains listed in Table 5.2 since they are all identical.

5.2.5. Testing the concerted evolution hypothesis

Patterns of concerted evolution were detected by means of abundance plots, BLAST analysis and haplotype networks.

I plotted the abundance of the first most abundant 50 haplotypes for both environmental and single-strain samples in order to render the plots clearly readable. If the hypothesis of concerted evolution in action was correct, I expected to see a steep decrease in abundance.
of minor haplotypes with respect to the dominant one. Plots were made in R (R Core Team, 2019) using the packages `ggplot2` (Wickham, 2016), `gridExtra` (Auguie, 2017) and `scales` (Wickham, 2018).

As second strategy, I blasted the validated environmental metabarcodes of each *Chaetoceros* species and the reference barcodes against those of the merged single-strains of each species. This was done in order to ascertain if: i) the most abundant haplotype in each single strain matched the reference barcode of that strain obtained with Sanger sequencing and with the dominant environmental haplotype; ii) the minor haplotypes in the strain were also found in the environmental samples.

Finally, as further check, I inferred haplotype networks for each species from environmental data (MareChiara) using the TCS method (Clement et al., 2000) implemented in PopART v1.7 (Leigh and Bryant, 2015). I only used metabarcodes with abundance $\geq 2$ in order to reduce the number of sequences to be processed for network inference. Furthermore, for *C. costatus* and *C. tenuissimus*, I further reduced the number of haplotypes analysed considering only the ones with abundance $\geq 10$ and $\geq 50$ respectively, in order to obtain a clearer graphical visualisation of networks. Metabarcodes, spanning from 2011 to 2013, were pooled together in months, and a different colour was assigned to each of them. This was done to test the following hypothesis: if concerted evolution was in action, I would have observed not only a major node surrounded by smaller ones, but also a congruence in the temporal pattern (colour pattern in the nodes). If not, I could have observed multiple dominant haplotypes (multiple major nodes), without a correspondence between the temporal pattern in peripheral and major nodes.
5.3. Results

5.3.1. General characteristics of the datasets

The number of haplotypes retrieved for each species from the environmental dataset of MareChiara after the validation procedure described in section 5.3.2 is provided in Table 5.3. In this thesis, the term “haplotype” indicates the non-redundant (unique) sequences. The number of haplotypes utilised ranged from 15 (C. anastomosans) to 527 (C. sp. Na11C3). In C. tenuissimus, considering only the first 50 most abundant haplotypes, I recovered 121,321 sequences (Table 5.3).

Table 5.3. Number of environmental sequences and haplotypes utilised in this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>N sequences utilised</th>
<th>N haplotypes utilised</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. anastomosans</td>
<td>287 (abundance ≥ 2)</td>
<td>14 (abundance ≥ 2)</td>
</tr>
<tr>
<td>C. costatus</td>
<td>8,220 (abundance ≥ 10)</td>
<td>38 (abundance ≥ 10)</td>
</tr>
<tr>
<td>C. curvisetus 2</td>
<td>9,763 (abundance ≥ 2)</td>
<td>369 (abundance ≥ 2)</td>
</tr>
<tr>
<td>Chaetoceros sp. Na11C3</td>
<td>12,924 (abundance ≥ 2)</td>
<td>527 (abundance ≥ 2)</td>
</tr>
<tr>
<td>Chaetoceros sp. Na26B1</td>
<td>1,154 (abundance ≥ 2)</td>
<td>59 (abundance ≥ 2)</td>
</tr>
<tr>
<td>C. tenuissimus</td>
<td>121,321 (abundance ≥ 50)</td>
<td>102 (abundance ≥ 50)</td>
</tr>
</tbody>
</table>

For single strain HTS, the number of raw sequences ranged from 32,112 (C. curvisetus 2 Na1C1) to 516,766 (Chaetoceros sp. Na11C3 and, after pre-processing, from 19,185 (C. curvisetus 2 Na1C1) to 94,449 (Chaetoceros sp. Na11C3). The number of haplotypes used for following analyses ranged from a minimum of 2,002 (C. curvisetus 2 strain Na1C1) to a maximum of 4,696 (C. costatus strain Na32B1) (Table 5.4).
Table 5.4. Number of sequences before and after pre-processing and total number of haplotypes utilised in each strain. Pre-processing refers to removal of adapters, primers and correction with ICC.

<table>
<thead>
<tr>
<th>Species/strains</th>
<th>N raw sequences</th>
<th>N sequences after pre-processing</th>
<th>N haplotypes after pre-processing</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na14C2</td>
<td>427,364</td>
<td>62,284</td>
<td>4,310</td>
</tr>
<tr>
<td>Na14C3</td>
<td>431,665</td>
<td>62,183</td>
<td>3,970</td>
</tr>
<tr>
<td><em>C. costatus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na1A3</td>
<td>238,922</td>
<td>34,226</td>
<td>3,634</td>
</tr>
<tr>
<td>Na32B1</td>
<td>421,807</td>
<td>50,407</td>
<td>4,696</td>
</tr>
<tr>
<td>Ro1B1</td>
<td>274,436</td>
<td>37,489</td>
<td>4,170</td>
</tr>
<tr>
<td>Ro2A2</td>
<td>230,989</td>
<td>32,394</td>
<td>3,622</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ch5B2</td>
<td>161,145</td>
<td>39,735</td>
<td>2,985</td>
</tr>
<tr>
<td>Na1C1</td>
<td>32,112</td>
<td>19,185</td>
<td>2,002</td>
</tr>
<tr>
<td>Na19A2</td>
<td>120,545</td>
<td>34,287</td>
<td>2,794</td>
</tr>
<tr>
<td>Na20A4</td>
<td>117,234</td>
<td>34,149</td>
<td>2,738</td>
</tr>
<tr>
<td><em>Chaetoceros sp. Na11C3</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na11C3</td>
<td>516,766</td>
<td>94,449</td>
<td>5,055</td>
</tr>
<tr>
<td>Na43A1</td>
<td>259,525</td>
<td>54,973</td>
<td>4,444</td>
</tr>
<tr>
<td><em>Chaetoceros sp. Na26B1</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Na26B1</td>
<td>273,039</td>
<td>56,985</td>
<td>3,360</td>
</tr>
<tr>
<td><em>C. tenuissimus</em></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GB2a</td>
<td>211,777</td>
<td>39,516</td>
<td>3,986</td>
</tr>
<tr>
<td>Na26A1</td>
<td>147,806</td>
<td>34,726</td>
<td>3,024</td>
</tr>
<tr>
<td>Na44A1</td>
<td>202,198</td>
<td>32,467</td>
<td>3,024</td>
</tr>
</tbody>
</table>
5.3.2. Abundance plots from environmental metabarcoding and single strain HTS

The plotting of the 50 most abundant haplotypes (Table A5.1 in Appendix V) from environmental metabarcoding data versus their abundance (log10 transformed) in each species (Fig. 5.1) showed a characteristic pattern. Indeed, in each species analysed, of all the haplotypes attributed to a particular species (environmental samples) there was one (the “dominant haplotype”) that was far more abundant of all the others, of at least one order of magnitude (Fig. 5.1). All the other copies occurred in the environment at lower abundance.

Patterns of abundance distribution in the HTS of single strains showed the same trend observed in the metabarcoding data of environmental samples (Fig. 5.2). Indeed, in each strain there was the same steep decrease in abundance of minor haplotypes in respect to the
dominant one. Furthermore, within the same species, the distribution of abundance of haplotypes was congruent (Fig. 5.2). The list of the 50 most abundant haplotypes from single strain HTS in each species, used for the plots, is available in the Appendix V as Table A5.2.

Fig. 5.2. Abundance plots for each strain analysed in different *Chaetoceros* species. (A) *C. anastomosans*; (B) *C. costatus*; (C) *C. curvisetus* 2; (D) *Chaetoceros* sp. Na11C3; (E) *Chaetoceros* sp. Na26B1; (F) *C. tenuissimus*. Data are from single strain high throughput sequencing. Only the first 50 most abundant haplotypes were plotted.

5.3.3. Blast of environmental haplotypes vs. single strain

Within each species, the dominant haplotypes of each strain were identical to each other. Therefore, for showing the results of BLAST analyses of single strains, I referred to just one haplotype (Table 5.5).
The result of BLAST analysis showed that the most abundant haplotype from environmental data as well as single strain HTS matched at 100% identity with the reference barcode (obtained with Sanger sequencing) of the species/strain it belonged (Table 5.5).
Table 5.5. Correspondence between the reference barcode (Sanger sequence) of each species and the dominant haplotypes of the environmental dataset (MareChiara) and single strain HTS. Since the reference sequences of the strains are identical to each other within the same species, only one has been chosen.

<table>
<thead>
<tr>
<th>Species</th>
<th>Reference sequence Accession number</th>
<th>Matching haplotype in MareChiara</th>
<th>% identity</th>
<th>Matching haplotype in single strain</th>
<th>% identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td>MG972358</td>
<td>M00390_81_0000000000-AA7DR_1_2109_10899_14476</td>
<td>100</td>
<td>97KSI_03703_04635</td>
<td>100</td>
</tr>
<tr>
<td><em>C. costatus</em></td>
<td>KY852258</td>
<td>M00390_81_0000000000-AA7DR_1_1112_20701_25092</td>
<td>100</td>
<td>97KSI_03062_04287</td>
<td>100</td>
</tr>
<tr>
<td><em>C. curvisetus</em></td>
<td>MG972239</td>
<td>M00390_81_0000000000-AA7DR_1_1101_24335_7294</td>
<td>100</td>
<td>97KSI_04187_04119</td>
<td>100</td>
</tr>
<tr>
<td><em>Chaetoceros</em> sp. Na11C3</td>
<td>MG972328</td>
<td>M00390_81_0000000000-AA7DR_1_1101_6410_5509</td>
<td>100</td>
<td>97KSI_03663_01512</td>
<td>100</td>
</tr>
<tr>
<td><em>Chaetoceros</em> sp. Na26B1</td>
<td>MG972329</td>
<td>M00390_81_0000000000-AA7DR_1_1101_16198_12414</td>
<td>100</td>
<td>97KSI_01986_05212</td>
<td>100</td>
</tr>
<tr>
<td><em>C. tenuissimus</em></td>
<td>MG972311</td>
<td>M00390_81_0000000000-AA7DR_1_1101_19390_3055</td>
<td>100</td>
<td>97KSI_00416_02071</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 5.6. Summary of percentage of identity found between environmental haplotypes and single strain in each *Chaetoceros* species. Single strains have been merged together before BLAST analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>N hap</th>
<th>100</th>
<th>99.74-99.73</th>
<th>99.48-99.47</th>
<th>99.21-99.20</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td>14</td>
<td>42.9 %</td>
<td>50.0 %</td>
<td>7.1 %</td>
<td>-</td>
</tr>
<tr>
<td><em>C. costatus</em></td>
<td>38</td>
<td>73.7 %</td>
<td>26.3 %</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><em>C. curvisetus</em> 2</td>
<td>369</td>
<td>53.6 %</td>
<td>41.5 %</td>
<td>4.9 %</td>
<td>-</td>
</tr>
<tr>
<td><em>C. sp. Na11C3</em></td>
<td>527</td>
<td>56.9 %</td>
<td>38.3 %</td>
<td>4.2 %</td>
<td>0.6 %</td>
</tr>
<tr>
<td><em>C. sp. Na26B1</em></td>
<td>59</td>
<td>45.8 %</td>
<td>49.2 %</td>
<td>5.0 %</td>
<td>-</td>
</tr>
<tr>
<td><em>C. tenuissimus</em></td>
<td>102</td>
<td>60.8 %</td>
<td>39.2 %</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

In most of the species, more than half of environmental haplotypes attributed were also found in single strains HTS at 100 % of identity. Overall, a match between environmental and single strain haplotypes was found for each species within the threshold of 99.20 % of identity (Table 5.6). This result support the hypothesis that the sequence variability observed in the environmental metabarcoding samples is part of infraspecific variation.

5.3.4. Phylogenetic networks from environmental samples

The inference of haplotype networks from the MareChiara dataset provided a graphical evidence to the occurrence of concerted evolution in the *Chaetoceros* species here analysed. Of all the species here analysed (Fig. 5.3 to Fig. 5.7), the temporal pattern observed in the node containing the dominant haplotype corresponded the temporal pattern of the other nodes containing haplotypes with lower abundance. This was particularly straightforward for *C. curvisetus* 2 (Fig. 5.5), *C. sp. Na11C3* (Fig. 5.6, left network) and *C. tenuissimus* (Fig. 5.7). These were also the species with the highest number of haplotypes.
utilised (369, 527 and 102 respectively). In *C. anastomosans* (Fig. 5.3) the pattern is almost absent due to the low number of sequences validated from the MareChiara dataset. However, in the HTS analysis of single strains (Fig. 5.2A), I have observed the expected pattern for concerted evolution in action.

Fig. 5.3. TCS haplotype network for *C. anastomosans* inferred from the MareChiara temporal dataset. A total of 14 haplotypes with abundance ≥ 2 across 2011 and 2013 was used. Sample in the legend refers to the number of reads.

The removal of all the haplotypes with abundance ≤ 9 in *C. costatus* allowed a better visualisation of minor haplotypes (Fig. 5.4).
Fig. 5.4. TCS haplotype network for *C. costatus* inferred from the MareChiara temporal dataset. A total of 38 haplotypes with abundance ≥ 10 across 2011 and 2013 was used. Sample in the legend refers to the number of reads.

In *C. curvisetus* 2 (Fig. 5.5), I observed at least ten nodes (minor haplotypes) with many reads whose temporal distribution patterns mimic that of the node comprising the dominant haplotype (the big one).
Fig. 5.5. TCS haplotype network for *C. curvisetus* 2 inferred from the MareChiara temporal dataset. A total of 369 haplotypes with abundance ≥ 2 across 2011 and 2013 was used. Sample in the legend refers to the number of reads.

For the closely related species *C*. sp. Na11C3 and *C*. sp. Na26B1, I have inferred a common TCS network since in the HTS phylogenetic tree in Gaonkar (2017) they were in the same clade with mixing minor haplotypes. In the network here inferred, they are on separated nodes, each with their minor haplotypes. The pattern of nodes expected in the case of concerted evolution is more evident for *C*. sp. Na11C3, where more sequences (527) were used; however, it is also observable, in reduced manner, in *C*. sp. Na26B1 (59 sequences utilised). An interest characteristic of such network is the fact that the two closely related species are differentiating each other in the occurrence across the year. The species *C*. sp. Na26B1 was exclusively found, during the years 2011-2013, in the months from August to October, whilst *C*. sp. Na11C3 is particularly abundant in June and July and less in the other months (Fig. 5.6).
Fig. 5.6. TCS haplotype network for *Chaetoceros* sp. Na11C3 (left) and Na26B1 (right) inferred from the MareChiara temporal dataset. A total of 586 haplotypes (527 for *C.* sp. Na11C3 and 59 for *C.* sp. Na26B1) with abundance ≥ 2 across 2011 and 2013 was used. Sample in the legend refers to the number of reads.

*Chaetoceros tenuissimus* is perhaps the species in which the pattern of concerted evolution is more evident (Fig. 5.7). Indeed, almost all the nodes around the central one containing the dominant haplotype have a temporal pattern mimicking it. The visualisation of the first 50 most abundant haplotypes has reduced the noise due to haplotypes at low abundances (e.g. less than 10) that is observable in the networks of other species (e.g. Fig. 5.3, Fig. 5.5 and Fig. 5.6).
Fig. 5.7. TCS haplotype network for *C. tenuissimus* inferred from the MareChiara temporal dataset. A total of 102 haplotypes with abundance \(\geq 50\) across 2011 and 2013 was used. Sample in the legend refers to the number of reads.

5.4. Discussion

5.4.1. Concerted evolution in Chaetoceros

Since the first explanation of the process of concerted evolution in the rDNA cistron of *Xenopus* by Brown et al. (1972) using DNA-RNA hybridisation, this phenomenon has been observed and studied over years in different organisms using different techniques. Among the latter, Sanger sequencing of rDNA copies followed by phylogenetic analysis has been the most common approach (e.g. Vogler and DeSalle, 1994; Buckler et al., 1997; Li and Zhang, 2002; Xiao et al., 2010). In recent times, concerted evolution has also been
revealed by whole-genome shotgun sequence data (e.g. Ganley and Kobayashi, 2007) and chromosomal and array approaches (e.g. Kuhn et al., 2011; Bueno et al., 2016). However, to date there are no examples of studies that have dealt with concerted evolution using metabarcoding data or single-strain high throughput sequencing. Thanks to the experimental design presented in this chapter, I have confirmed my hypothesis that the 18S gene is under concerted evolution in the *Chaetoceros* species here analysed. Furthermore, I have shown that it is possible to use a temporal metabarcoding dataset (with an adequate number of samples) to seek a first signal of this evolutionary phenomenon. Phylogenetic haplotype networks and the plots showing the distribution of the abundance of each haplotype were in accordance with the expectations of homogenisation. In particular, the occurrence in each strain and, more general, in each species of a haplotype (the “dominant” haplotype) far more abundant than all the others, confirmed my hypothesis of concerted evolution in action. In addition, the generation of single strain high throughput sequencing allowed me to prove at molecular level the patterns previously observed at level of ecological community (Gaonkar, 2017). This validation allowed distinguishing the presence of a real biological phenomenon due to infraspecific variation, instead of an artefact due to PCR errors or by-product of massive parallel sequencing. Based on the results obtained, I excluded that the variation found in the environment is an artefact of the methodology used. All the analyses here performed confirmed that the variation occurring in the temporal metabarcoding dataset is due to real variation present in the population and in representative individuals from that population. I did not perform any single-cell analysis, but instead, used a monoclonal culture of each *Chaetoceros* strain to perform high throughput sequencing, I am confident in asserting that the observed variation is intraindividual. This is because I have analysed the pattern of a multicopy gene that occurs in thousands of copies in the genome, and the probability that
any mutation possibly occurring during culturing condition could have hampered the experiment is insignificant.

Minor variation among haplotypes is no sequencing artefact but results from concerted evolution not entirely succeeding in eliminating the emerging microvariation resulting from mutations and recombination. Indeed, BLAST analysis has shown that the haplotypes found in the environment also occur in the strains (are infraspecific variation). The abundance plots demonstrated also that both haplotypes from environmental metabarcoding and single strain HTS exhibit the same distribution pattern, with a dominant haplotype surrounded by several minor haplotypes. Furthermore, the dominant haplotypes of all the strains analysed were identical within the same species, as well as to these strains’ Sanger sequences, and to the dominant metabarcode of that species in the environmental metabarcodes. This observed identity is in accordance with the way HTS and Sanger technologies work. A Sanger sequence can be considered as a consensus of all the targeted copies of a gene amplified. In this “consensus sequence”, most of the weight will be carried by the most abundant sequence and therefore the Sanger sequence will read as the dominant haplotype. On the contrary, in massive parallel sequencing, every single copy present in the reaction tube will be sequenced, the only limit being constituted by reagents and sequencer characteristics. The dominant haplotype in the massive parallel sequencing is therefore the sequence that is “dominating” the aspect of the electropherogram in Sanger sequencing.

Based on my results, I hypothesise that in species in which besides concerted evolution other events have occurred, such as recent merging of two distinct populations, there might be multiple co-dominant haplotypes and their recombinants, a situation likely to result in messy, unreadable electropherograms. However, double peaks in electropherograms can also be due to different alleles occurring at similar frequencies in nuclear markers or to
heteroplasmy in the case of uniparental markers (e.g. mitochondrial and plastid genes). In this context, massive parallel sequencing can be of help at discriminating such situations.

5.4.2. Implications for DNA barcoding

Different regions of the rDNA cistron are targeted for DNA barcoding in several taxa. For example, the V4 region in the 18S gene is the currently recommended barcoding region for protists (Pawlowski et al., 2012), whilst the ITS region serves as such for fungi (Schoch et al., 2012). Some authors (e.g. Chase et al., 2007; Sonnenberg et al., 2007; Spooner, 2009) have argued that the concerted evolution process, known to affect ribosomal genes, may not be sufficiently effective to ensure complete sequence homogeneity. Therefore, knowing the extent of infraspecific variation and modality of evolution of such regions is vital to barcoding studies (Kane et al., 2012). The classical approach to the study of variants in rDNA genes is based on the cloning and Sanger sequencing of amplified products that produces noisy electropherograms. Studies targeting this region in different organisms revealed the occurrence of several different copies within each organism analysed and highlighted the potential risk for barcoding studies (e.g. Naidoo et al., 2013; Dakal et al., 2016). Indeed, one of the characteristics of a good DNA barcode is to have high interspecific divergence and low intraspecific variability (Kress and Erickson, 2008). Dakal et al. (2016) argued that the presence of several ribotypes within an individual shortens the barcoding gap and should be taken into consideration in barcoding studies of yeasts. However, what is lacking in these studies is information about the abundance of these “alternative” rDNA copies. Pillet et al. (2012) tried to predict the number of ribotypes in each specimen of *Elphidium macellum* (Foraminifera) correlating the number of clones screened with the number of ribotypes found. The authors argued that although some of less abundant ribotypes could be due to PCR artefacts, the high Spearman
correlation coefficient suggested that the real number of ribotypes in each individual could be underestimated (Pillet et al., 2012).

In this study, I have demonstrated that within each strain of several *Chaetoceros* species occur thousands of 18S ribotypes, one of which is far more abundant than all the others (the “dominant” haplotype). Because of such huge differences in abundance, the probability that a “minor” haplotype is sequenced with Sanger chemistry is almost null. In turn, this means that there is no risk associated to the use of the rDNA cistron as target gene in classical DNA barcoding studies. However, in metabarcoding studies these minor haplotypes can create a false rare diversity and therefore produce artefacts in diversity assessments.

My study also demonstrated that, when conducting metabarcoding experiments (from both environmental samples and bulk communities) or single strain HTS, the most abundant haplotype that is recovered for each species corresponds to the sequence that would be obtained by Sanger sequencing. Therefore, in case of a taxon for which a reference sequence is not available yet, the dominant haplotype retrieved from a metabarcoding dataset can be considered as such, and subsequently validated using Sanger sequencing when the specimen has been sampled.

5.4.3. Copy number across the Tree of Life and possible role of rDNA heterogeneity

The copy number of rDNA cistron has been estimated in different taxa along the Tree of Life. These studies have demonstrated that this number is highly variable: from 60 to 220 copies in fungi (Simon et al., 2005), 39-19,300 in animals (Prokopowich et al., 2003) and 150-26,048 in plants (Prokopowich et al., 2003). Among protists, ciliates harbour the highest number of rDNA copies, between 3,000 and 400,000 (Gong et al., 2013), followed by diatoms (1,057 to 12,812, Godhe et al., 2008) and dinoflagellates (200 to 1,200, Galluzzi et al., 2004). High variation among copies has been detected using a cloning and
sequencing approach in fungi (Simon and Weiß, 2008), dinoflagellates (Gribble and Anderson, 2007; Miranda et al., 2012), and Foraminifera (Pillet et al., 2012), as well as with genome sequencing in the plant genus *Asclepias* (Weitemier et al., 2015). However, the biological relevance of having many rDNA haplotypes is largely unknown. Part of such variation could be due imperfection of the mechanism that should homogenise all the copies among them. Another explanation, complementary to the former, is that there could be a selective advantage in possessing all these different copies. Indeed, in bacteria it has been shown that the number of copies of small rDNA gene correlates with the rate at which phylogenetically diverse bacteria respond to resource availability, with a high copy number leading to rapid colony formation (Klappenbach et al., 2000). In eukaryotes, the copy number of rDNA genes is unstable (Ganley and Kobayashi, 2014) and its stabilisation extends lifespan in yeast (Howitz et al., 2003). Always in yeast, it has been recently demonstrated that DNA replication stress induces a reduction in rDNA copy number in yeast (Salim et al., 2017). The possible role of rDNA heterogeneity in protists is to be unveiled yet.

5.4.4. Conclusion

In this chapter, I have shown how the analysis of ecological data by evolutionary approach can open unexpected scenarios. In this case, the analysis of temporal metabarcoding data analysed by phylogenetic networks, showed a pattern compatible with the theory of concerted evolution. The next experiment designed (the HTS of the strains) and the sets of analyses performed (plot of haplotype distribution, analysis of sequence similarity, evolutionary networks) confirmed the hypothesis in all the *Chaetoceros* species tested here, providing the first robust proof of concerted evolution in diatoms. Moreover, the simple approach to produce HTS of the strains can also be applied to other genera of diatoms or protists in order to understand the evolution of such gene region in different
marine taxa. In this sense, the use of metabarcoding or HTS data in general here shown is novel and powerful. However, the repercussions of this finding on metabarcoding studies are conflicting. On the one hand, I have demonstrated that the dominant haplotype perfectly matches with the Sanger reference sequence, validating the use of the metabarcoding technique for ecological studies. On the other hand, the high number of sequences occurring at low abundances (minor haplotypes) inflate the diversity assessments. In this study, I showed that at 99% of identity, all infraspecific variability is collapsed together. This is true for Chaetoceros, but the validity across other genera is to be tested yet. For studies using metabarcoding data at genus level, the clustering of sequences could be easily guided by evolutionary networks or trees, but for studies at community level the solution is more complicated. However, a possible course of action for future research could be to compare the results obtained in this study in Chaetoceros with other diatom and protist species, in order to understand the evolution of such gene region as well as the applicability of metabarcoding and high throughput sequencing in ecological and evolutionary studies in other marine organisms.

References


Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D., Breiner, H. W., Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a


Table A5.1. List of the 50 most abundant haplotypes of MareChiara dataset and relative abundance. (A) *C. anastomosans*; (B) *C. costatus*; (C) *C. curvisetus* 2; (D) *Chaetoceros* sp. Na11C3; (E) *Chaetoceros* sp. Na26B1; (F) *C. tenuissimus*.

(A) *C. anastomosans*

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_0000000000-AA7DR_1_2109_10899_14476</td>
<td>251</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_15106_24806</td>
<td>7</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1102_14009_18673</td>
<td>5</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1104_18421_26995</td>
<td>4</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1102_14691_18024</td>
<td>2</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1112_20701_25092</td>
<td>7371</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_17729_14855</td>
<td>58</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1103_19872_21486</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1107_8089_22888</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1101_25301_21125</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2102_5911_13557</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1109_4878_20074</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1107_11931_10743</td>
<td>2</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2103_22630_12031</td>
<td>2</td>
</tr>
</tbody>
</table>

(B) *C. costatus*

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1112_20701_25092</td>
<td>7371</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1101_15660_16312</td>
<td>63</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_17729_14855</td>
<td>58</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1103_9214_24775</td>
<td>53</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1106_5105_16625</td>
<td>49</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1105_3711_16331</td>
<td>47</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2103_10847_8395</td>
<td>46</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_2101_22469_20255</td>
<td>45</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1104_6258_7154</td>
<td>42</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1106_23082_19031</td>
<td>42</td>
</tr>
</tbody>
</table>
(C) *C. curvisetus* 2

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_24335_7294</td>
<td>6944</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_20896_15345</td>
<td>453</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_16288_26989</td>
<td>316</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2103_28131_14780</td>
<td>200</td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_11450_2553</td>
<td>178</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1101_4123_20747</td>
<td>118</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1103_9271_16586</td>
<td>60</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_2104_16946_23798</td>
<td>48</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1113_21922_15727</td>
<td>44</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1112_21269_25157</td>
<td>34</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1103_3248_11772</td>
<td>32</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1108_10322_9155</td>
<td>28</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1103_11972_18123</td>
<td>27</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1101_6719_20035</td>
<td>20</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1114_21412_18288</td>
<td>18</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1106_11976_23254</td>
<td>17</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2107_23256_10864</td>
<td>16</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1104_25381_15393</td>
<td>14</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1105_18022_3841</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1107_28263_20903</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1108_21222_12699</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_2105_8525_7938</td>
<td>14</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2104_20058_6516</td>
<td>13</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1114_19386_5392</td>
<td>12</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2107_25970_19729</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_10957_18509</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1102_2708_17479</td>
<td>11</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1101_22451_8109</td>
<td>10</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1108_17827_14144</td>
<td>10</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1108_21804_18503</td>
<td>9</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1109_4803_21661</td>
<td>9</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2108_15746_20584</td>
<td>9</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1104_26914_13137</td>
<td>8</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1113_27091_11483</td>
<td>8</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_2112_12719_20529</td>
<td>8</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1102_4688_18121</td>
<td>8</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_21225_11210</td>
<td>8</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1111_23982_25410</td>
<td>8</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_2107_9396_11171</td>
<td>8</td>
</tr>
<tr>
<td>M00390_80_0000000000-AA759_1_1107_6136_20046</td>
<td>7</td>
</tr>
</tbody>
</table>
### (D) Chaetoceros sp. Na11C3

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_6410_5509</td>
<td>9723</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_20571_7220</td>
<td>304</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24763_6366</td>
<td>143</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_11196_27974</td>
<td>139</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_2424_15795</td>
<td>112</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_25739_19440</td>
<td>105</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_22194_6232</td>
<td>95</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1110_26826_22971</td>
<td>53</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_27566_10283</td>
<td>48</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_17728_10049</td>
<td>46</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24856_10898</td>
<td>44</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_4716_11773</td>
<td>34</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_9998_3612</td>
<td>24</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_25605_20745</td>
<td>24</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_25478_16104</td>
<td>18</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2103_19587_7983</td>
<td>17</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_9541_20449</td>
<td>17</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_16225_9778</td>
<td>17</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_25799_8358</td>
<td>16</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_12261_15569</td>
<td>16</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_24721_11822</td>
<td>16</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2103_24105_11792</td>
<td>15</td>
</tr>
</tbody>
</table>
(E) *Chaetoceros* sp. Na26B1

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_17994_1598</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_9058_1598</td>
<td>15</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_23857_36378</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2103_16137_11494</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_12929_19411</td>
<td>14</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_27443_14488</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_7128_23668</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_13368_5663</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_19739_2477</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_15583_2449</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_29235_17394</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_11002_6236</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_15319_18036</td>
<td>13</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_24270_18806</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_17825_8140</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19218_3762</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_13352_5601</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_20284_5220</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_8323_22034</td>
<td>12</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2102_20462_22578</td>
<td>11</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_11222_26655</td>
<td>11</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_22190_7531</td>
<td>11</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_17105_13417</td>
<td>10</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_9523_10659</td>
<td>10</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_10927_8701</td>
<td>10</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_27964_16114</td>
<td>10</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_9732_26522</td>
<td>10</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_16616_9226</td>
<td>10</td>
</tr>
</tbody>
</table>

265
(F) *C. tenuissimus*

<table>
<thead>
<tr>
<th>MareChiara haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_19390_3055</td>
<td>102608</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_17418_16093</td>
<td>3567</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_22470_4248</td>
<td>1696</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_19124_2835</td>
<td>704</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_9398_12278</td>
<td>686</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_25018_11322</td>
<td>673</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_10404_7536</td>
<td>614</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_19019_24975</td>
<td>578</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_13180_26564</td>
<td>422</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_24620_23287</td>
<td>337</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_4023_9561</td>
<td>279</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_21482_26412</td>
<td>272</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1110_14154_2636</td>
<td>261</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_7389_18078</td>
<td>259</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_22681_24682</td>
<td>237</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_3967_7974</td>
<td>204</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_12520_13351</td>
<td>203</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_7026_20095</td>
<td>183</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_20566_15022</td>
<td>183</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_5213_17961</td>
<td>172</td>
</tr>
</tbody>
</table>
Table A5.2. List of the 50 most abundant haplotypes in each strain and relative abundance. (A) *C. anastomosans*; (B) *C. costatus*; (C) *C. curvisetus* 2; (D) *Chaetoceros* sp. Na11C3; (E) *Chaetoceros* sp. Na26B1; (F) *C. tenuissimus*.

(A) *C. anastomosans*

<table>
<thead>
<tr>
<th>Strain Na14C2 haplotype</th>
<th>abundance</th>
<th>Strain Na14C3 haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>97KSI_03703_04635</td>
<td>415688</td>
<td>97KSI_00154_01105</td>
<td>425386</td>
</tr>
<tr>
<td>97KSI_01752_02923</td>
<td>8450</td>
<td>97KSI_00577_02517</td>
<td>2896</td>
</tr>
<tr>
<td>97KSI_04955_04530</td>
<td>2642</td>
<td>97KSI_03131_02564</td>
<td>2552</td>
</tr>
<tr>
<td>97KSI_04310_02070</td>
<td>2614</td>
<td>97KSI_03319_07327</td>
<td>2229</td>
</tr>
<tr>
<td>97KSI_01980_02933</td>
<td>2211</td>
<td>97KSI_01243_00573</td>
<td>1720</td>
</tr>
<tr>
<td>97KSI_04654_05661</td>
<td>1758</td>
<td>97KSI_02186_03427</td>
<td>1394</td>
</tr>
<tr>
<td>97KSI_01552_02038</td>
<td>1326</td>
<td>97KSI_03667_05941</td>
<td>1313</td>
</tr>
<tr>
<td>97KSI_03326_03053</td>
<td>1197</td>
<td>97KSI_02019_04661</td>
<td>1310</td>
</tr>
<tr>
<td>97KSI_05279_05020</td>
<td>1137</td>
<td>97KSI_01145_00779</td>
<td>1279</td>
</tr>
<tr>
<td>97KSI_03965_02929</td>
<td>1108</td>
<td>97KSI_04919_03953</td>
<td>1213</td>
</tr>
<tr>
<td>97KSI_00248_04839</td>
<td>1103</td>
<td>97KSI_05265_04107</td>
<td>1097</td>
</tr>
<tr>
<td>97KSI_01964_01784</td>
<td>1013</td>
<td>97KSI_03046_04251</td>
<td>1026</td>
</tr>
<tr>
<td>97KSI_03816_05769</td>
<td>976</td>
<td>97KSI_03775_05001</td>
<td>995</td>
</tr>
<tr>
<td>97KSI_03105_03043</td>
<td>965</td>
<td>97KSI_05249_04750</td>
<td>843</td>
</tr>
<tr>
<td>97KSI_04175_01929</td>
<td>796</td>
<td>97KSI_03376_00849</td>
<td>793</td>
</tr>
<tr>
<td>97KSI_00165_05792</td>
<td>787</td>
<td>97KSI_03553_02332</td>
<td>766</td>
</tr>
<tr>
<td>97KSI_00568_01197</td>
<td>706</td>
<td>97KSI_04764_06003</td>
<td>641</td>
</tr>
<tr>
<td>97KSI_00336_03283</td>
<td>613</td>
<td>97KSI_03031_02892</td>
<td>608</td>
</tr>
<tr>
<td>97KSI_03717_06068</td>
<td>596</td>
<td>97KSI_00478_01384</td>
<td>558</td>
</tr>
<tr>
<td>97KSI_02447_07481</td>
<td>531</td>
<td>97KSI_01772_02651</td>
<td>521</td>
</tr>
<tr>
<td>97KSI_03196_02045</td>
<td>529</td>
<td>97KSI_03655_04959</td>
<td>484</td>
</tr>
<tr>
<td>97KSI_04154_05725</td>
<td>477</td>
<td>97KSI_00857_03833</td>
<td>418</td>
</tr>
<tr>
<td>97KSI_02035_05239</td>
<td>455</td>
<td>97KSI_00955_01738</td>
<td>396</td>
</tr>
<tr>
<td>97KSI_01560_03047</td>
<td>404</td>
<td>97KSI_03270_04978</td>
<td>379</td>
</tr>
<tr>
<td>97KSI_03238_05718</td>
<td>385</td>
<td>97KSI_01084_01758</td>
<td>371</td>
</tr>
<tr>
<td>97KSI_02548_06343</td>
<td>369</td>
<td>97KSI_03644_05435</td>
<td>342</td>
</tr>
<tr>
<td>97KSI_04665_00700</td>
<td>355</td>
<td>97KSI_01172_01434</td>
<td>336</td>
</tr>
<tr>
<td>97KSI_02269_07091</td>
<td>328</td>
<td>97KSI_04142_04404</td>
<td>314</td>
</tr>
</tbody>
</table>
(B) C. costatus

<table>
<thead>
<tr>
<th>Strain Na1A3</th>
<th>Strain Na32B1</th>
<th>Strain Ro1B1</th>
<th>Strain Ro2A2</th>
</tr>
</thead>
<tbody>
<tr>
<td>haplotype</td>
<td>abundance</td>
<td>haplotype</td>
<td>abundance</td>
</tr>
<tr>
<td>97KSI_03062</td>
<td>218582</td>
<td>97KSI_00628</td>
<td>379948</td>
</tr>
<tr>
<td>_04287</td>
<td>_05777</td>
<td>_06963</td>
<td></td>
</tr>
<tr>
<td>97KSI_04461</td>
<td>2428</td>
<td>97KSI_01839</td>
<td>7684</td>
</tr>
<tr>
<td>_01525</td>
<td>_05734</td>
<td>_07118</td>
<td></td>
</tr>
<tr>
<td>97KSI_00359</td>
<td>2236</td>
<td>97KSI_03980</td>
<td>3834</td>
</tr>
<tr>
<td>_04771</td>
<td>_04497</td>
<td>_03185</td>
<td></td>
</tr>
<tr>
<td>97KSI_00934</td>
<td>2161</td>
<td>97KSI_00245</td>
<td>3200</td>
</tr>
<tr>
<td>_04772</td>
<td>_02587</td>
<td>_01176</td>
<td></td>
</tr>
<tr>
<td>97KSI_03738</td>
<td>1526</td>
<td>97KSI_03252</td>
<td>2012</td>
</tr>
<tr>
<td>_05059</td>
<td>_07546</td>
<td>_05587</td>
<td></td>
</tr>
<tr>
<td>97KSI_03733</td>
<td>1488</td>
<td>97KSI_02626</td>
<td>1948</td>
</tr>
<tr>
<td>Code</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>97KSI_04115_01060</td>
<td>185</td>
<td>97KSI_04485_00803</td>
<td>307</td>
</tr>
<tr>
<td>97KSI_04778_06180</td>
<td>173</td>
<td>97KSI_00919_05841</td>
<td>306</td>
</tr>
<tr>
<td>97KSI_02430_06720</td>
<td>166</td>
<td>97KSI_01110_05077</td>
<td>305</td>
</tr>
<tr>
<td>97KSI_02255_01180</td>
<td>161</td>
<td>97KSI_00217_03921</td>
<td>298</td>
</tr>
<tr>
<td>97KSI_02255_01267</td>
<td>153</td>
<td>97KSI_01101_04050</td>
<td>280</td>
</tr>
<tr>
<td>97KSI_00425_02383</td>
<td>147</td>
<td>97KSI_03229_03447</td>
<td>272</td>
</tr>
<tr>
<td>97KSI_02036_04306</td>
<td>142</td>
<td>97KSI_04137_01183</td>
<td>255</td>
</tr>
<tr>
<td>97KSI_03261_03215</td>
<td>140</td>
<td>97KSI_03622_06607</td>
<td>249</td>
</tr>
<tr>
<td>97KSI_04172_02622</td>
<td>135</td>
<td>97KSI_04444_05213</td>
<td>229</td>
</tr>
<tr>
<td>97KSI_01125_01348</td>
<td>127</td>
<td>97KSI_03587_00348</td>
<td>224</td>
</tr>
<tr>
<td>97KSI_04637_00929</td>
<td>126</td>
<td>97KSI_00183_01315</td>
<td>221</td>
</tr>
<tr>
<td>97KSI_03306_02496</td>
<td>124</td>
<td>97KSI_03610_04478</td>
<td>202</td>
</tr>
<tr>
<td>97KSI_04767_05381</td>
<td>118</td>
<td>97KSI_02433_02259</td>
<td>196</td>
</tr>
<tr>
<td>97KSI_02697_01205</td>
<td>118</td>
<td>97KSI_01688_05159</td>
<td>194</td>
</tr>
<tr>
<td>97KSI_01280_03473</td>
<td>112</td>
<td>97KSI_00874_07345</td>
<td>194</td>
</tr>
<tr>
<td>97KSI_01124_06743</td>
<td>108</td>
<td>97KSI_04196_05067</td>
<td>190</td>
</tr>
<tr>
<td>97KSI_04242_05794</td>
<td>107</td>
<td>97KSI_01336_05679</td>
<td>190</td>
</tr>
<tr>
<td>97KSI_02500_06487</td>
<td>107</td>
<td>97KSI_04771_01799</td>
<td>168</td>
</tr>
<tr>
<td>97KSI_04999_03479</td>
<td>105</td>
<td>97KSI_04712_03820</td>
<td>166</td>
</tr>
<tr>
<td>97KSI_01644_01248</td>
<td>103</td>
<td>97KSI_02553_03747</td>
<td>161</td>
</tr>
<tr>
<td>97KSI_00512_02828</td>
<td>100</td>
<td>97KSI_00163_02311</td>
<td>160</td>
</tr>
<tr>
<td>97KSI_02519_03431</td>
<td>100</td>
<td>97KSI_05224_02816</td>
<td>152</td>
</tr>
<tr>
<td>Strain Ch5B2</td>
<td>Strain Na1C1</td>
<td>Strain Na19A2</td>
<td>Strain Na20A4</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>haplotype</td>
<td>abundance</td>
<td>haplotype</td>
<td>abundance</td>
</tr>
<tr>
<td>97KSI_04187</td>
<td>169486</td>
<td>97KSI_02988</td>
<td>38657</td>
</tr>
<tr>
<td>_04119</td>
<td></td>
<td>_02860</td>
<td></td>
</tr>
<tr>
<td>97KSI_01106</td>
<td>2193</td>
<td>97KSI_02471</td>
<td>1828</td>
</tr>
<tr>
<td>_03045</td>
<td></td>
<td>_00580</td>
<td></td>
</tr>
<tr>
<td>97KSI_00309</td>
<td>1250</td>
<td>97KSI_04759</td>
<td>1220</td>
</tr>
<tr>
<td>_01364</td>
<td></td>
<td>_01311</td>
<td></td>
</tr>
<tr>
<td>97KSI_01166</td>
<td>1244</td>
<td>97KSI_00792</td>
<td>712</td>
</tr>
<tr>
<td>_02524</td>
<td></td>
<td>_03430</td>
<td></td>
</tr>
<tr>
<td>97KSI_04760</td>
<td>1145</td>
<td>97KSI_03343</td>
<td>300</td>
</tr>
<tr>
<td>_02997</td>
<td></td>
<td>_07160</td>
<td></td>
</tr>
<tr>
<td>97KSI_02598</td>
<td>951</td>
<td>97KSI_01161</td>
<td>282</td>
</tr>
<tr>
<td>_04820</td>
<td></td>
<td>_05983</td>
<td></td>
</tr>
<tr>
<td>97KSI_01200</td>
<td>898</td>
<td>97KSI_00302</td>
<td>258</td>
</tr>
<tr>
<td>_03715</td>
<td></td>
<td>_02313</td>
<td></td>
</tr>
<tr>
<td>97KSI_04592</td>
<td>749</td>
<td>97KSI_04935</td>
<td>248</td>
</tr>
<tr>
<td>_04855</td>
<td></td>
<td>_03621</td>
<td></td>
</tr>
<tr>
<td>97KSI_01198</td>
<td>734</td>
<td>97KSI_00304</td>
<td>222</td>
</tr>
<tr>
<td>_07171</td>
<td></td>
<td>_02419</td>
<td></td>
</tr>
<tr>
<td>97KSI_02817</td>
<td>547</td>
<td>97KSI_03970</td>
<td>205</td>
</tr>
<tr>
<td>_06374</td>
<td></td>
<td>_02605</td>
<td></td>
</tr>
<tr>
<td>97KSI_01153</td>
<td>539</td>
<td>97KSI_03163</td>
<td>159</td>
</tr>
<tr>
<td>_01578</td>
<td></td>
<td>_06734</td>
<td></td>
</tr>
<tr>
<td>97KSI_04523</td>
<td>518</td>
<td>97KSI_01093</td>
<td>146</td>
</tr>
<tr>
<td>_06290</td>
<td></td>
<td>_01329</td>
<td></td>
</tr>
<tr>
<td>97KSI_00800</td>
<td>511</td>
<td>97KSI_04278</td>
<td>141</td>
</tr>
<tr>
<td>_05590</td>
<td></td>
<td>_06681</td>
<td></td>
</tr>
<tr>
<td>97KSI_04002</td>
<td>431</td>
<td>97KSI_04914</td>
<td>119</td>
</tr>
<tr>
<td>_02065</td>
<td></td>
<td>_06693</td>
<td></td>
</tr>
<tr>
<td>97KSI_03983</td>
<td>401</td>
<td>97KSI_02864</td>
<td>108</td>
</tr>
<tr>
<td>_07206</td>
<td></td>
<td>_04008</td>
<td></td>
</tr>
<tr>
<td>97KSI_01092</td>
<td>355</td>
<td>97KSI_02828</td>
<td>98</td>
</tr>
<tr>
<td>_03272</td>
<td></td>
<td>_07250</td>
<td></td>
</tr>
<tr>
<td>97KSI_00312</td>
<td>340</td>
<td>97KSI_03777</td>
<td>97</td>
</tr>
<tr>
<td>_01166</td>
<td></td>
<td>_00812</td>
<td></td>
</tr>
<tr>
<td>97KSI_04002</td>
<td>337</td>
<td>97KSI_04824</td>
<td>93</td>
</tr>
<tr>
<td>_06564</td>
<td></td>
<td>_04867</td>
<td></td>
</tr>
<tr>
<td>97KSI_02219</td>
<td>303</td>
<td>97KSI_00384</td>
<td>93</td>
</tr>
<tr>
<td>_02561</td>
<td></td>
<td>_04781</td>
<td></td>
</tr>
<tr>
<td>97KSI_02526</td>
<td>294</td>
<td>97KSI_01749</td>
<td>89</td>
</tr>
<tr>
<td>_05461</td>
<td></td>
<td>_07525</td>
<td></td>
</tr>
<tr>
<td>97KSI_01095</td>
<td>273</td>
<td>97KSI_05239</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>97KSI_03777</td>
<td></td>
</tr>
</tbody>
</table>

273
<p>| _   _   _     _   _   _   _   _   _       |
|---|---|---|---|---|---|---|---|
| _03786| _01905| _04242| _04057|
| 274| 97KSI_04945_03799| 71| 97KSI_02284_05741| 203| 97KSI_02902_01662| 185|
| 232| 97KSI_04797_02731| 70| 97KSI_02522_01005| 201| 97KSI_01313_04662| 184|
| 193| 97KSI_04369_00832| 68| 97KSI_00627_03203| 179| 97KSI_04159_05252| 163|
| 189| 97KSI_03520_01532| 64| 97KSI_02241_07073| 168| 97KSI_03857_04134| 155|
| 189| 97KSI_01235_00479| 63| 97KSI_04204_05991| 168| 97KSI_02892_05702| 154|
| 186| 97KSI_01745_01180| 60| 97KSI_00788_05695| 161| 97KSI_04199_06533| 140|
| 182| 97KSI_01129_02122| 60| 97KSI_03233_02899| 140| 97KSI_05261_05641| 128|
| 163| 97KSI_00233_01226| 57| 97KSI_01150_06389| 124| 97KSI_04066_02721| 118|
| 160| 97KSI_05082_02758| 54| 97KSI_01472_03020| 123| 97KSI_02366_02585| 109|
| 146| 97KSI_04969_01211| 50| 97KSI_00670_03093| 118| 97KSI_03596_00164|  94|
| 142| 97KSI_01567_06490| 46| 97KSI_03840_05921| 111| 97KSI_02399_01540|  93|
| 131| 97KSI_04918_01993| 45| 97KSI_00436_03352|  99| 97KSI_01192_01527|  92|
| 127| 97KSI_00291_01293| 44| 97KSI_04117_05593|  92| 97KSI_05062_06515|  92|
| 125| 97KSI_04516_06891| 44| 97KSI_04751_05373|  92| 97KSI_05027_04129|  92|
| 123| 97KSI_00354_04116| 43| 97KSI_00821_01684|  90| 97KSI_03248_02447|  91|
| 120| 97KSI_05059_01647| 41| 97KSI_03880_03291|  87| 97KSI_00965_06607|  88|
| 113| 97KSI_01749_03291| 38| 97KSI_04942_05066|  85| 97KSI_04268_06469|  87|
| 107| 97KSI_03466_02986| 38| 97KSI_05268_02917|  84| 97KSI_03288_02240|  82|
| 105| 97KSI_02187_05081| 38| 97KSI_04321_05569|  84| 97KSI_02123_02627|  82|
|  99| 97KSI_03037_03176| 37| 97KSI_03837_00354|  79| 97KSI_01259_01169|  77|
|  99| 97KSI_00500_04676| 36| 97KSI_04046_05963|  75| 97KSI_01327_01761|  72|
|  95| 97KSI_01635_03769| 34| 97KSI_04046_05735|  74| 97KSI_00512_06372|  71|</p>
<table>
<thead>
<tr>
<th>Strain Na11C3</th>
<th>Strain Na43A1</th>
</tr>
</thead>
<tbody>
<tr>
<td>haplotype</td>
<td>abundance</td>
</tr>
<tr>
<td>97KSI_03663_01512</td>
<td>520646</td>
</tr>
<tr>
<td>97KSI_01342_06051</td>
<td>2986</td>
</tr>
<tr>
<td>97KSI_02348_07517</td>
<td>2877</td>
</tr>
<tr>
<td>97KSI_05274_03658</td>
<td>2642</td>
</tr>
<tr>
<td>97KSI_03157_04337</td>
<td>2887</td>
</tr>
<tr>
<td>97KSI_03666_02005</td>
<td>2508</td>
</tr>
<tr>
<td>97KSI_02256_07312</td>
<td>2479</td>
</tr>
<tr>
<td>97KSI_01133_03016</td>
<td>1015</td>
</tr>
<tr>
<td>97KSI_00933_03016</td>
<td>1500</td>
</tr>
<tr>
<td>97KSI_01782_01635</td>
<td>769</td>
</tr>
<tr>
<td>97KSI_00695_03999</td>
<td>697</td>
</tr>
<tr>
<td>97KSI_05130_06212</td>
<td>610</td>
</tr>
</tbody>
</table>
(E) *Chaetoceros* sp. Na26B1

**Strain Na26B1**

<table>
<thead>
<tr>
<th>haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>97KSI_01986_05212</td>
<td>277568</td>
</tr>
<tr>
<td>97KSI_03665_06329</td>
<td>3153</td>
</tr>
<tr>
<td>Code</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------</td>
</tr>
<tr>
<td>97KSI_04190_02727</td>
<td>1589</td>
</tr>
<tr>
<td>97KSI_00244_01964</td>
<td>1475</td>
</tr>
<tr>
<td>97KSI_03873_06785</td>
<td>1355</td>
</tr>
<tr>
<td>97KSI_03445_07568</td>
<td>1151</td>
</tr>
<tr>
<td>97KSI_02227_03525</td>
<td>1011</td>
</tr>
<tr>
<td>97KSI_03148_07131</td>
<td>932</td>
</tr>
<tr>
<td>97KSI_03876_04249</td>
<td>793</td>
</tr>
<tr>
<td>97KSI_04041_07453</td>
<td>763</td>
</tr>
<tr>
<td>97KSI_04283_00952</td>
<td>670</td>
</tr>
<tr>
<td>97KSI_00030_04326</td>
<td>624</td>
</tr>
<tr>
<td>97KSI_02890_03453</td>
<td>586</td>
</tr>
<tr>
<td>97KSI_04274_01473</td>
<td>525</td>
</tr>
<tr>
<td>97KSI_04440_02001</td>
<td>512</td>
</tr>
<tr>
<td>97KSI_03286_02649</td>
<td>430</td>
</tr>
<tr>
<td>97KSI_02922_01051</td>
<td>366</td>
</tr>
<tr>
<td>97KSI_04530_03070</td>
<td>362</td>
</tr>
<tr>
<td>97KSI_00963_05797</td>
<td>361</td>
</tr>
<tr>
<td>97KSI_04837_05019</td>
<td>314</td>
</tr>
<tr>
<td>97KSI_04106_03770</td>
<td>311</td>
</tr>
<tr>
<td>97KSI_00956_04969</td>
<td>305</td>
</tr>
<tr>
<td>97KSI_03789_01597</td>
<td>282</td>
</tr>
<tr>
<td>97KSI_03156_00523</td>
<td>255</td>
</tr>
<tr>
<td>97KSI_01224_05079</td>
<td>253</td>
</tr>
<tr>
<td>97KSI_02360_01917</td>
<td>253</td>
</tr>
<tr>
<td>97KSI_01739_00466</td>
<td>240</td>
</tr>
<tr>
<td>97KSI_04093_05675</td>
<td>234</td>
</tr>
<tr>
<td>97KSI_01022_00717</td>
<td>208</td>
</tr>
<tr>
<td>97KSI_02078_01146</td>
<td>202</td>
</tr>
<tr>
<td>97KSI_02771_00919</td>
<td>200</td>
</tr>
<tr>
<td>97KSI_02581_00543</td>
<td>188</td>
</tr>
<tr>
<td>97KSI_01203_00783</td>
<td>186</td>
</tr>
<tr>
<td>97KSI_00396_06508</td>
<td>175</td>
</tr>
<tr>
<td>97KSI_03062_03130</td>
<td>172</td>
</tr>
<tr>
<td>97KSI_04257_04121</td>
<td>166</td>
</tr>
<tr>
<td>97KSI_03025_01007</td>
<td>158</td>
</tr>
<tr>
<td>97KSI_03486_05518</td>
<td>148</td>
</tr>
</tbody>
</table>
(F) *C. tenuissimus*

<table>
<thead>
<tr>
<th>Strain GB2a haplotype</th>
<th>abundance</th>
<th>Strain Na26A1 haplotype</th>
<th>abundance</th>
<th>Strain Na44A1 haplotype</th>
<th>abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>97KSI_00416_02071</td>
<td>173565</td>
<td>97KSI_03171_05542</td>
<td>152018</td>
<td>97KSI_03436_06826</td>
<td>194285</td>
</tr>
<tr>
<td>97KSI_04642_06542</td>
<td>35170</td>
<td>97KSI_04894_03826</td>
<td>12696</td>
<td>97KSI_04590_06245</td>
<td>3252</td>
</tr>
<tr>
<td>97KSI_02170_01835</td>
<td>2362</td>
<td>97KSI_01619_03486</td>
<td>2362</td>
<td>97KSI_00987_03631</td>
<td>2640</td>
</tr>
<tr>
<td>97KSI_00830_02955</td>
<td>1849</td>
<td>97KSI_03111_06431</td>
<td>1931</td>
<td>97KSI_01520_02095</td>
<td>1240</td>
</tr>
<tr>
<td>97KSI_01788_02370</td>
<td>1163</td>
<td>97KSI_01659_06282</td>
<td>1077</td>
<td>97KSI_03196_06342</td>
<td>1089</td>
</tr>
<tr>
<td>97KSI_02477_05763</td>
<td>1097</td>
<td>97KSI_04937_05270</td>
<td>939</td>
<td>97KSI_01644_07473</td>
<td>726</td>
</tr>
<tr>
<td>97KSI_01971_07513</td>
<td>628</td>
<td>97KSI_03701_07245</td>
<td>583</td>
<td>97KSI_02090_05953</td>
<td>672</td>
</tr>
<tr>
<td>97KSI_00424_05431</td>
<td>605</td>
<td>97KSI_00722_03534</td>
<td>461</td>
<td>97KSI_03952_04583</td>
<td>615</td>
</tr>
<tr>
<td>97KSI_00326_00927</td>
<td>573</td>
<td>97KSI_04639_01300</td>
<td>452</td>
<td>97KSI_04492_01277</td>
<td>589</td>
</tr>
<tr>
<td>97KSI_04044_01178</td>
<td>571</td>
<td>97KSI_00576_03603</td>
<td>404</td>
<td>97KSI_04250_03866</td>
<td>587</td>
</tr>
<tr>
<td>97KSI_01112_02654</td>
<td>562</td>
<td>97KSI_01878_01695</td>
<td>343</td>
<td>97KSI_04749_02314</td>
<td>556</td>
</tr>
<tr>
<td>97KSI_01372_06128</td>
<td>494</td>
<td>97KSI_03866_06205</td>
<td>334</td>
<td>97KSI_04316_02987</td>
<td>454</td>
</tr>
<tr>
<td>97KSI_02195_0521</td>
<td>480</td>
<td>97KSI_05158_0358</td>
<td>324</td>
<td>97KSI_04034_0535</td>
<td>436</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>-------------------</td>
<td>-----</td>
<td>-------------------</td>
<td>-----</td>
</tr>
<tr>
<td>97KSI_05189_0222</td>
<td>403</td>
<td>97KSI_00651_0498</td>
<td>314</td>
<td>97KSI_03501_0648</td>
<td>354</td>
</tr>
<tr>
<td>97KSI_04045_0668</td>
<td>379</td>
<td>97KSI_01196_0664</td>
<td>295</td>
<td>97KSI_01303_0349</td>
<td>347</td>
</tr>
<tr>
<td>97KSI_04829_0172</td>
<td>351</td>
<td>97KSI_02081_0328</td>
<td>278</td>
<td>97KSI_04917_0415</td>
<td>298</td>
</tr>
<tr>
<td>97KSI_00301_0434</td>
<td>339</td>
<td>97KSI_03506_0651</td>
<td>254</td>
<td>97KSI_01488_0316</td>
<td>297</td>
</tr>
<tr>
<td>97KSI_01056_0031</td>
<td>333</td>
<td>97KSI_01741_0717</td>
<td>246</td>
<td>97KSI_02770_0332</td>
<td>254</td>
</tr>
<tr>
<td>97KSI_02347_0199</td>
<td>264</td>
<td>97KSI_01789_0208</td>
<td>222</td>
<td>97KSI_03487_0361</td>
<td>250</td>
</tr>
<tr>
<td>97KSI_02988_0663</td>
<td>242</td>
<td>97KSI_01767_0445</td>
<td>221</td>
<td>97KSI_04433_0503</td>
<td>247</td>
</tr>
<tr>
<td>97KSI_03746_0575</td>
<td>229</td>
<td>97KSI_03939_0695</td>
<td>211</td>
<td>97KSI_03404_0230</td>
<td>241</td>
</tr>
<tr>
<td>97KSI_00295_0418</td>
<td>229</td>
<td>97KSI_01102_0034</td>
<td>211</td>
<td>97KSI_04607_0250</td>
<td>227</td>
</tr>
<tr>
<td>97KSI_01187_0248</td>
<td>228</td>
<td>97KSI_04115_0442</td>
<td>177</td>
<td>97KSI_02123_0482</td>
<td>227</td>
</tr>
<tr>
<td>97KSI_02624_0574</td>
<td>226</td>
<td>97KSI_03788_0568</td>
<td>174</td>
<td>97KSI_02978_0641</td>
<td>219</td>
</tr>
<tr>
<td>97KSI_03283_0575</td>
<td>218</td>
<td>97KSI_01082_0472</td>
<td>161</td>
<td>97KSI_01510_0146</td>
<td>188</td>
</tr>
<tr>
<td>97KSI_02572_0160</td>
<td>216</td>
<td>97KSI_03767_0567</td>
<td>160</td>
<td>97KSI_02847_0541</td>
<td>169</td>
</tr>
<tr>
<td>97KSI_04520_0337</td>
<td>210</td>
<td>97KSI_05211_0145</td>
<td>145</td>
<td>97KSI_04978_0589</td>
<td>166</td>
</tr>
<tr>
<td>97KSI_03192_0546</td>
<td>199</td>
<td>97KSI_02376_0476</td>
<td>140</td>
<td>97KSI_02805_0517</td>
<td>160</td>
</tr>
<tr>
<td>97KSI_04613_0670</td>
<td>185</td>
<td>97KSI_05165_0414</td>
<td>130</td>
<td>97KSI_04346_0609</td>
<td>156</td>
</tr>
<tr>
<td>97KSI_01283_0060</td>
<td>181</td>
<td>97KSI_01005_0712</td>
<td>124</td>
<td>97KSI_00260_0626</td>
<td>155</td>
</tr>
<tr>
<td>97KSI_00716_0632</td>
<td>169</td>
<td>97KSI_00852_0705</td>
<td>122</td>
<td>97KSI_02069_0330</td>
<td>150</td>
</tr>
<tr>
<td>97KSI_03836_0387</td>
<td>164</td>
<td>97KSI_00826_0584</td>
<td>118</td>
<td>97KSI_03547_0334</td>
<td>146</td>
</tr>
<tr>
<td>97KSI_04152_0223</td>
<td>162</td>
<td>97KSI_01687_0554</td>
<td>113</td>
<td>97KSI_01793_0186</td>
<td>146</td>
</tr>
<tr>
<td>97KSI_02857_0343</td>
<td>161</td>
<td>97KSI_04583_0273</td>
<td>110</td>
<td>97KSI_05247_0553</td>
<td>143</td>
</tr>
<tr>
<td>97KSI_02055_0231</td>
<td>156</td>
<td>97KSI_01300_0638</td>
<td>106</td>
<td>97KSI_00065_0260</td>
<td>142</td>
</tr>
<tr>
<td>Index</td>
<td>KSI</td>
<td>Index</td>
<td>KSI</td>
<td>Index</td>
<td>KSI</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>97KSI_04590_0454</td>
<td>150</td>
<td>97KSI_00362_0289</td>
<td>100</td>
<td>97KSI_01470_0588</td>
<td>135</td>
</tr>
<tr>
<td>97KSI_03686_0146</td>
<td>146</td>
<td>97KSI_04243_0114</td>
<td>100</td>
<td>97KSI_02523_0661</td>
<td>113</td>
</tr>
<tr>
<td>97KSI_04968_0322</td>
<td>143</td>
<td>97KSI_01489_0615</td>
<td>100</td>
<td>97KSI_00662_0496</td>
<td>111</td>
</tr>
<tr>
<td>97KSI_00793_0462</td>
<td>136</td>
<td>97KSI_02287_0427</td>
<td>100</td>
<td>97KSI_03533_0671</td>
<td>95</td>
</tr>
<tr>
<td>97KSI_00181_0172</td>
<td>131</td>
<td>97KSI_01058_0384</td>
<td>98</td>
<td>97KSI_03799_0579</td>
<td>95</td>
</tr>
<tr>
<td>97KSI_01418_0115</td>
<td>130</td>
<td>97KSI_02747_0350</td>
<td>97</td>
<td>97KSI_04125_0142</td>
<td>93</td>
</tr>
<tr>
<td>97KSI_00647_0127</td>
<td>129</td>
<td>97KSI_01051_0410</td>
<td>92</td>
<td>97KSI_00098_0375</td>
<td>92</td>
</tr>
<tr>
<td>97KSI_03612_0618</td>
<td>117</td>
<td>97KSI_03769_0580</td>
<td>91</td>
<td>97KSI_02907_0293</td>
<td>91</td>
</tr>
<tr>
<td>97KSI_00585_0433</td>
<td>116</td>
<td>97KSI_01363_0700</td>
<td>88</td>
<td>97KSI_01146_0368</td>
<td>87</td>
</tr>
<tr>
<td>97KSI_01798_0717</td>
<td>114</td>
<td>97KSI_00876_0659</td>
<td>86</td>
<td>97KSI_03794_0217</td>
<td>83</td>
</tr>
<tr>
<td>97KSI_00711_0112</td>
<td>114</td>
<td>97KSI_05231_0196</td>
<td>79</td>
<td>97KSI_04582_0650</td>
<td>82</td>
</tr>
<tr>
<td>97KSI_01741_0726</td>
<td>109</td>
<td>97KSI_01808_0660</td>
<td>79</td>
<td>97KSI_03682_0270</td>
<td>77</td>
</tr>
<tr>
<td>97KSI_00779_0718</td>
<td>108</td>
<td>97KSI_01199_0586</td>
<td>66</td>
<td>97KSI_01707_0639</td>
<td>73</td>
</tr>
<tr>
<td>97KSI_00236_0310</td>
<td>108</td>
<td>97KSI_01177_0595</td>
<td>64</td>
<td>97KSI_01801_0444</td>
<td>73</td>
</tr>
<tr>
<td>97KSI_03215_0643</td>
<td>107</td>
<td>97KSI_01867_0363</td>
<td>63</td>
<td>97KSI_01465_0542</td>
<td>71</td>
</tr>
</tbody>
</table>
Table A5.3. Percentage of identity between MareChiara haplotypes (query) and single strain ones (subject) after BLAST analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>MareChiara haplotype</th>
<th>Single haplotype</th>
<th>strain identity</th>
<th>% identity</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. anastomosans</em></td>
<td>M00390_80_0000000000-AA759_1_1101_25301_21125</td>
<td>97KSI_02594_00890</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_80_0000000000-AA759_1_1102_14009_18673</td>
<td>97KSI_04102_07418</td>
<td></td>
<td>99.73</td>
</tr>
<tr>
<td></td>
<td>M00390_80_0000000000-AA759_1_1103_20661_12123</td>
<td>97KSI_00508_05944</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_80_0000000000-AA759_1_1104_25301_21125</td>
<td>97KSI_02594_00890</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_80_0000000000-AA759_1_1105_20661_12123</td>
<td>97KSI_00508_05944</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA759_1_2102_5911_13557</td>
<td>97KSI_03016_04504</td>
<td></td>
<td>99.47</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA759_1_2103_22630_12031</td>
<td>97KSI_03068_02457</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA759_1_2104_14691_18024</td>
<td>97KSI_01530_01155</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA759_1_2106_24067_19555</td>
<td>97KSI_02722_02195</td>
<td></td>
<td>99.73</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA759_1_2107_15106_24806</td>
<td>97KSI_03390_05249</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1101_15660_16312</td>
<td>97KSI_03703_04635</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1102_19338_9858</td>
<td>97KSI_01923_04779</td>
<td></td>
<td>99.74</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1103_13921_27562</td>
<td>97KSI_00109_06212</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1104_6258_7154</td>
<td>97KSI_00373_06383</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1105_12696_21050</td>
<td>97KSI_00508_03753</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>M00390_81_0000000000-AA7DR_1_1105_3711_16331</td>
<td>97KSI_00371_06560</td>
<td></td>
<td>100.00</td>
</tr>
</tbody>
</table>

*C. costatus*
| M00390_81_000000000-000000000 | 97KSI_00757_05950 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_05086_06023 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_02188_05262 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_05254_01755 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03905_04827 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03947_04173 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03946_04448 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_03054_04565 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_00710_04425 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03089_07525 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03062_04287 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_04115_01060 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_00611_03023 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_00447_01854 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_00459_02674 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_00232_04621 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_00530_01306 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_00134_02557 | 100.00 |
| C. curvisetus 2 | M00390_81_000000000-000000000 | 97KSI_00743_05902 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_04781_05500 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_03132_05290 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_03868_07408 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_05189_05534 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_04229_03933 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_04700_05670 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_03212_04472 | 100.00 |
| M00390_81_000000000-000000000 | 97KSI_03028_04597 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_02148_01247 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_01340_02411 | 99.74 |
| M00390_81_000000000-000000000 | 97KSI_00134_02557 | 100.00 |

282
<p>| M00390_80_000000000- AA759_1_1103_10733_19662 | 97KSI_00698_06928 | 99.74 |
| M00390_80_000000000- AA759_1_1103_20985_3495 | 97KSI_05102_02709 | 99.74 |
| M00390_80_000000000- AA759_1_1103_5353_19137 | 97KSI_03868_07408 | 99.74 |
| M00390_80_000000000- AA759_1_1104_16447_20731 | 97KSI_02376_07020 | 100.00 |
| M00390_80_000000000- AA759_1_1104_17935_15234 | 97KSI_01122_03542 | 99.74 |
| M00390_80_000000000- AA759_1_1104_25381_15393 | 97KSI_03132_05290 | 100.00 |
| M00390_80_000000000- AA759_1_1104_26195_19039 | 97KSI_04516_04950 | 99.74 |
| M00390_80_000000000- AA759_1_1104_26914_13137 | 97KSI_02785_06860 | 100.00 |
| M00390_80_000000000- AA759_1_1105_10346_14685 | 97KSI_00461_03779 | 100.00 |
| M00390_80_000000000- AA759_1_1105_11673_22262 | 97KSI_03408_02291 | 100.00 |
| M00390_80_000000000- AA759_1_1105_18022_3841 | 97KSI_04187_04119 | 99.74 |
| M00390_80_000000000- AA759_1_1106_18744_19999 | 97KSI_03155_05120 | 99.74 |
| M00390_80_000000000- AA759_1_1106_20541_6865 | 97KSI_04187_04119 | 99.74 |
| M00390_80_000000000- AA759_1_1106_4543_9571 | 97KSI_03612_05711 | 99.74 |
| M00390_80_000000000- AA759_1_1107_14518_27111 | 97KSI_01833_06973 | 100.00 |
| M00390_80_000000000- AA759_1_1107_19330_2927 | 97KSI_04567_02701 | 99.48 |
| M00390_80_000000000- AA759_1_1107_23028_11214 | 97KSI_03304_05267 | 100.00 |
| M00390_80_000000000- AA759_1_1107_26025_8835 | 97KSI_01648_05844 | 99.74 |
| M00390_80_000000000- AA759_1_1107_5485_13612 | 97KSI_03052_03627 | 100.00 |
| M00390_80_000000000- AA759_1_1108_1727_14144 | 97KSI_01225_01807 | 99.74 |
| M00390_80_000000000- AA759_1_1108_18885_3870 | 97KSI_04167_01691 | 100.00 |
| M00390_80_000000000- AA759_1_1108_21041_24244 | 97KSI_00653_00804 | 100.00 |
| M00390_80_000000000- AA759_1_1108_21804_18503 | 97KSI_01225_01807 | 99.48 |
| M00390_80_000000000- AA759_1_1109_10112_22292 | 97KSI_03449_02999 | 100.00 |
| M00390_80_000000000- AA759_1_1109_12487_7144 | 97KSI_03474_03329 | 99.74 |
| M00390_80_000000000- AA759_1_1109_22338_15912 | 97KSI_04167_01691 | 99.74 |
| M00390_80_000000000- AA759_1_1109_22400_3193 | 97KSI_04516_04950 | 99.74 |
| M00390_80_000000000-AA759_1_1109_23473_17636 | 97KSI_01613_07083 | 100.00 |
| M00390_80_000000000-AA759_1_1109_4803_21661 | 97KSI_01376_05247 | 99.74 |
| M00390_80_000000000-AA759_1_1111_11786_2903 | 97KSI_01745_01180 | 100.00 |
| M00390_80_000000000-AA759_1_1111_19795_21223 | 97KSI_00831_05888 | 100.00 |
| M00390_80_000000000-AA759_1_1111_20150_4074 | 97KSI_04567_02701 | 99.74 |
| M00390_80_000000000-AA759_1_1111_24228_17282 | 97KSI_02911_07583 | 100.00 |
| M00390_80_000000000-AA759_1_1111_2812_18777 | 97KSI_04488_06894 | 99.74 |
| M00390_80_000000000-AA759_1_1111_2829_18786 | 97KSI_04488_06894 | 100.00 |
| M00390_80_000000000-AA759_1_1111_8624_21360 | 97KSI_04740_01204 | 99.74 |
| M00390_80_000000000-AA759_1_1112_11357_13565 | 97KSI_00867_07123 | 100.00 |
| M00390_80_000000000-AA759_1_1112_11663_13037 | 97KSI_00930_00962 | 100.00 |
| M00390_80_000000000-AA759_1_1112_22787_23939 | 97KSI_02977_04061 | 100.00 |
| M00390_80_000000000-AA759_1_1113_13540_17170 | 97KSI_04599_02055 | 99.48 |
| M00390_80_000000000-AA759_1_1113_19343_13398 | 97KSI_03764_04133 | 100.00 |
| M00390_80_000000000-AA759_1_1113_27091_11483 | 97KSI_03505_03492 | 100.00 |
| M00390_80_000000000-AA759_1_1113_27098_11499 | 97KSI_03505_03492 | 99.74 |
| M00390_80_000000000-AA759_1_1114_17684_18347 | 97KSI_04911_00819 | 100.00 |
| M00390_80_000000000-AA759_1_1114_19386_5392 | 97KSI_02927_04010 | 99.74 |
| M00390_80_000000000-AA759_1_1114_21412_18288 | 97KSI_04740_01204 | 100.00 |
| M00390_80_000000000-AA759_1_2102_14339_11186 | 97KSI_01137_04815 | 100.00 |
| M00390_80_000000000-AA759_1_2103_15394_7041 | 97KSI_04898_06275 | 99.74 |
| M00390_80_000000000-AA759_1_2104_10763_17096 | 97KSI_02194_02911 | 100.00 |
| M00390_80_000000000-AA759_1_2104_11244_3581 | 97KSI_00809_01739 | 100.00 |
| M00390_80_000000000-AA759_1_2104_18131_13043 | 97KSI_03612_05711 | 100.00 |
| M00390_80_000000000-AA759_1_2104_18695_10584 | 97KSI_04369_00832 | 100.00 |
| M00390_80_000000000-AA759_1_2104_19422_28482 | 97KSI_01336_03485 | 99.74 |
| M00390_80_000000000-AA759_1_2104_20058_6516 | 97KSI_00086_01422 | 99.74 |
| M00390_80_000000000-AA759_1_2105_12470_6531 | 97KSI_03569_02826 | 99.74 |
| M00390_80_000000000-AA759_1_2105_15923_6293 | 97KSI_04397_03614 | 100.00 |
| M00390_80_000000000-AA759_1_2106_11644_17478 | 97KSI_02828_02782 | 100.00 |
| M00390_80_000000000-AA759_1_2106_20114_18359 | 97KSI_01699_04103 | 100.00 |
| M00390_80_000000000-AA759_1_2106_23048_15729 | 97KSI_01958_06937 | 99.74 |
| M00390_80_000000000-AA759_1_2106_7808_13667 | 97KSI_04519_01603 | 100.00 |
| M00390_80_000000000-AA759_1_2107_13577_17164 | 97KSI_01635_03769 | 100.00 |
| M00390_80_000000000-AA759_1_2107_18121_26034 | 97KSI_04585_00974 | 100.00 |
| M00390_80_000000000-AA759_1_2107_23035_13452 | 97KSI_04514_04780 | 100.00 |</p>
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Material Property</th>
<th>Property Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_15906_14284</td>
<td></td>
<td>97KSI_01553_01460</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_17413_12794</td>
<td></td>
<td>97KSI_01648_05844</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19050_21812</td>
<td></td>
<td>97KSI_04738_06689</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19881_12816</td>
<td></td>
<td>97KSI_04537_05455</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_24260_11901</td>
<td></td>
<td>97KSI_00362_02486</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_11776_23421</td>
<td></td>
<td>97KSI_01033_04984</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_12813_20116</td>
<td></td>
<td>97KSI_01261_03143</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_17846_11250</td>
<td></td>
<td>97KSI_04452_06333</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_23437_11795</td>
<td></td>
<td>97KSI_04256_05357</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_12698</td>
<td></td>
<td>97KSI_01376_05247</td>
<td>99.48</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_8241_13607</td>
<td></td>
<td>97KSI_04704_05753</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_10037_24355</td>
<td></td>
<td>97KSI_01376_05247</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_5005_12698</td>
<td></td>
<td>97KSI_02492_01578</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_11382_10477</td>
<td></td>
<td>97KSI_01131_01904</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_11976_23254</td>
<td></td>
<td>97KSI_04213_06207</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_12010_23897</td>
<td></td>
<td>97KSI_00500_04676</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_13620_7801</td>
<td></td>
<td>97KSI_00500_04676</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_14632_23952</td>
<td></td>
<td>97KSI_04740_01204</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_19291_13940</td>
<td></td>
<td>97KSI_04740_01204</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_19693_18158</td>
<td></td>
<td>97KSI_00500_04676</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_20593_21372</td>
<td></td>
<td>97KSI_00500_04676</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_20985_7779</td>
<td></td>
<td>97KSI_04567_02701</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_2628_14595</td>
<td></td>
<td>97KSI_02881_04736</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_4722_10235</td>
<td></td>
<td>97KSI_04694_05800</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_6581_15713</td>
<td></td>
<td>97KSI_02077_05387</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_7821_12016</td>
<td></td>
<td>97KSI_01051_05261</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_12787_9145</td>
<td></td>
<td>97KSI_00233_01226</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_16021_15870</td>
<td></td>
<td>97KSI_05130_02585</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_20388_17913</td>
<td></td>
<td>97KSI_04599_02055</td>
<td>99.48</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_24621_22614</td>
<td></td>
<td>97KSI_00382_04346</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_25117_20516</td>
<td></td>
<td>97KSI_01537_06394</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1110_7895_25543</td>
<td>97KSI_04187_04119</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_10820_10057</td>
<td>97KSI_03419_06418</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_11450_2553</td>
<td>97KSI_03419_06418</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_11655_22760</td>
<td>97KSI_05153_02135</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_12745_18011</td>
<td>97KSI_01936_07377</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_14758_9950</td>
<td>97KSI_01795_00693</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_17102_24469</td>
<td>97KSI_03419_06418</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_21225_11210</td>
<td>97KSI_03419_06418</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1111_23982_25410</td>
<td>97KSI_03419_06418</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_10694_20471</td>
<td>97KSI_01973_03126</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_21269_25157</td>
<td>97KSI_04905_06288</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_23142_16163</td>
<td>97KSI_00681_01119</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1112_25078_12140</td>
<td>97KSI_00996_02030</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_10609_21251</td>
<td>97KSI_01843_06084</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_10765_18473</td>
<td>97KSI_00681_01119</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_12685_22420</td>
<td>97KSI_04771_06275</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_14513_14532</td>
<td>97KSI_00996_02030</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_18348_7289</td>
<td>97KSI_01219_04037</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_22337_17115</td>
<td>97KSI_01225_01807</td>
<td>99.48</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1113_24438_25877</td>
<td>97KSI_04599_02055</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_13133_7846</td>
<td>97KSI_03666_03024</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_18358_7289</td>
<td>97KSI_01219_04037</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_22337_17115</td>
<td>97KSI_01225_01807</td>
<td>99.48</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_24438_25877</td>
<td>97KSI_04599_02055</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_25565_18230</td>
<td>97KSI_03684_06839</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_25609_5817</td>
<td>97KSI_04363_05641</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1114_7843_16895</td>
<td>97KSI_00300_06550</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_10014_19409</td>
<td>97KSI_00942_03578</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>Description</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_12564_15378</td>
<td>97KSI_03958_00212</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_16946_23798</td>
<td>97KSI_00567_05066</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_18077_27447</td>
<td>97KSI_01225_01807</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_21699_15198</td>
<td>97KSI_02356_06598</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_4704_10529</td>
<td>97KSI_03083_01515</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_5389_13932</td>
<td>97KSI_01743_04579</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_11652_17368</td>
<td>97KSI_00189_01245</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_12150_4581</td>
<td>97KSI_02406_00831</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_14502_17429</td>
<td>97KSI_02111_02597</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_14886_21243</td>
<td>97KSI_00538_04716</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_3066_19474</td>
<td>97KSI_01422_06975</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_6558_12191</td>
<td>97KSI_02301_04052</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_8525_7938</td>
<td>97KSI_04781_05500</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_14426_18041</td>
<td>97KSI_01797_07274</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_16661_24663</td>
<td>97KSI_00936_03491</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_17213_17299</td>
<td>97KSI_00475_06811</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_19924_20621</td>
<td>97KSI_03294_06151</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_8835_23021</td>
<td>97KSI_04042_02077</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_10446_7832</td>
<td>97KSI_03958_00212</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_11725_16578</td>
<td>97KSI_03958_00212</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_13036_14378</td>
<td>97KSI_01244_07404</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_14537_22178</td>
<td>97KSI_02742_07076</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_14756_21371</td>
<td>97KSI_01851_04108</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_17842_13515</td>
<td>97KSI_04567_02701</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_20255_15389</td>
<td>97KSI_05184_04990</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_9396_11171</td>
<td>97KSI_03303_06741</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_12821_27054</td>
<td>97KSI_04599_02055</td>
<td>99.48</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_15229_13188</td>
<td>97KSI_04567_02701</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_16729_15068</td>
<td>97KSI_04516_04950</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_17265_11023</td>
<td>97KSI_04975_03734</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_3833_11029</td>
<td>97KSI_03885_03542</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_4505_22110</td>
<td>97KSI_01081_00420</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_6822_9219</td>
<td>97KSI_02082_05813</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_9341_20249</td>
<td>97KSI_03919_06615</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_13907_11340</td>
<td>97KSI_02136_06907</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_14423_9938</td>
<td>97KSI_04296_03028</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_16921_13436</td>
<td>97KSI_00446_02613</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_19339_19157</td>
<td>97KSI_00407_00871</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_4746_14889</td>
<td>97KSI_02023_02420</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_10351_12783</td>
<td>97KSI_03803_06221</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_14206_25147</td>
<td>97KSI_00632_03290</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_17018_9908</td>
<td>97KSI_00291_01293</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_6835_21029</td>
<td>97KSI_04599_02055</td>
<td>99.48</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_10281_13264</td>
<td>97KSI_05182_04204</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_10486_7906</td>
<td>97KSI_04516_04950</td>
<td>99.48</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_16150_12284</td>
<td>97KSI_03840_03784</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_20084_7863</td>
<td>97KSI_04388_06583</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_22522_16753</td>
<td>97KSI_03620_01563</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_24941_11256</td>
<td>97KSI_00473_05399</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_3921_17933</td>
<td>97KSI_01033_04984</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_6272_21662</td>
<td>97KSI_02082_05813</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_9782_11142</td>
<td>97KSI_00440_06051</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_9310_9000</td>
<td>97KSI_01735_04967</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_14241_20548</td>
<td>97KSI_04567_02701</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_14745_11643</td>
<td>97KSI_02492_01578</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_16329_13134</td>
<td>97KSI_00717_00658</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_16913_20967</td>
<td>97KSI_04410_06275</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_16196_13436</td>
<td>97KSI_01648_05844</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_4950_10754</td>
<td>97KSI_04190_05506</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_7363_14186</td>
<td>97KSI_00293_03871</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_15051_13851</td>
<td>97KSI_00293_03871</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_15116_23862</td>
<td>97KSI_01888_03164</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_17389_18710</td>
<td>97KSI_04707_04348</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_20908_10400</td>
<td>97KSI_01340_02432</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

**C. sp. Na11C3**

<p>| M00390_40_000000000-A6D16_1_1108_17304_25880 | 97KSI_03021_03598 | 100.00 |
| M00390_80_000000000-AA759_1_1101_24925_5895 | 97KSI_03096_05080 | 99.73 |
| M00390_80_000000000-AA759_1_1102_11036_5510 | 97KSI_01656_03252 | 99.73 |</p>
<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_80_000000000-AA759_1_1102_11862_19658</td>
<td>97KSI_00774_05835</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1102_22265_13042</td>
<td>97KSI_002846_03329</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1103_18744_24933</td>
<td>97KSI_00993_01026</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1103_24919_24925</td>
<td>97KSI_04387_04031</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1103_4831_10564</td>
<td>97KSI_00321_05775</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1103_6746_6120</td>
<td>97KSI_00460_06255</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1104_18700_4291</td>
<td>97KSI_01067_00604</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1104_2328_18588</td>
<td>97KSI_03817_04140</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1105_12350_25470</td>
<td>97KSI_01943_03397</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1105_12930_17652</td>
<td>97KSI_01165_00618</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1105_19903_27035</td>
<td>97KSI_03369_02217</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1105_24919_24925</td>
<td>97KSI_04017_02014</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1106_6160_18009</td>
<td>97KSI_04258_02565</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1107_13696_15050</td>
<td>97KSI_03473_05700</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1107_25024_7083</td>
<td>97KSI_04469_05210</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1107_9830_24499</td>
<td>97KSI_02750_07324</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1108_8881_16788</td>
<td>97KSI_00487_03761</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1108_8552_5622</td>
<td>97KSI_02911_03818</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1110_1111_11445_6058</td>
<td>97KSI_00893_05900</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1110_12930_26967</td>
<td>97KSI_01165_00618</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1111_13044_25175</td>
<td>97KSI_00595_04086</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1111_24180_24201</td>
<td>97KSI_04796_03244</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1112_17929_5348</td>
<td>97KSI_01837_05566</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1112_29011_17675</td>
<td>97KSI_00453_01744</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1112_6328_6889</td>
<td>97KSI_01122_03589</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1112_6677_23858</td>
<td>97KSI_00976_05951</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1113_27929_14998</td>
<td>97KSI_03191_01259</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1114_10356_13250</td>
<td>97KSI_04550_07000</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1114_20338_13729</td>
<td>97KSI_00289_02270</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1114_2317_12526</td>
<td>97KSI_02212_04011</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1114_25768_22209</td>
<td>97KSI_01002_05814</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1_1114_7144_11894</td>
<td>97KSI_04095_00209</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_17107_13439</td>
<td>97KSI_00533_03380</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_18409_11139</td>
<td>97KSI_03488_06956</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_18743_18805</td>
<td>97KSI_03833_00784</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_20003_21142</td>
<td>97KSI_00890_05779</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_21290_7677</td>
<td>97KSI_00774_00457</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_24432_18027</td>
<td>97KSI_05250_05331</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_25478_16104</td>
<td>97KSI_00748_06170</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_27613_11799</td>
<td>97KSI_03488_06956</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_27935_22163</td>
<td>97KSI_00748_06170</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_6410_5509</td>
<td>97KSI_03663_01512</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1101_7376_5104</td>
<td>97KSI_02216_04247</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_12058_20294</td>
<td>97KSI_04981_02568</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_12871_5484</td>
<td>97KSI_01214_00826</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_13228_10139</td>
<td>97KSI_01758_00919</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_13427_6631</td>
<td>97KSI_01427_02313</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_15210_26766</td>
<td>97KSI_00890_05779</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_15583_2449</td>
<td>97KSI_02189_04009</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_16054_27848</td>
<td>97KSI_05057_04395</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_17728_10049</td>
<td>97KSI_01062_02647</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_17746_10054</td>
<td>97KSI_03757_00347</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_18329_3761</td>
<td>97KSI_02189_04009</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_19523_26650</td>
<td>97KSI_03496_05087</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_23857_26378</td>
<td>97KSI_04568_05809</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24082_17124</td>
<td>97KSI_03663_01512</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24563_10548</td>
<td>97KSI_04778_03296</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24763_6366</td>
<td>97KSI_00844_05835</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_24856_10898</td>
<td>97KSI_00321_05775</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_25954_22369</td>
<td>97KSI_03894_07053</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_27534_22303</td>
<td>97KSI_03139_06556</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_27566_10283</td>
<td>97KSI_02911_03818</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_3969_15352</td>
<td>97KSI_00533_03380</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_6352_12255</td>
<td>97KSI_00890_05779</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_9046_16993</td>
<td>97KSI_03139_06556</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>------------------</td>
<td>--------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1102_9058_25666</td>
<td>97KSI_01733_00992</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_10959_18065</td>
<td>97KSI_03820_01798</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_11912_2490</td>
<td>97KSI_00378_06507</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_13104_26209</td>
<td>97KSI_01230_04601</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_13537_14886</td>
<td>97KSI_00533_02475</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_14016_18079</td>
<td>97KSI_03360_01646</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_15051_28199</td>
<td>97KSI_03820_01798</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_16291_4409</td>
<td>97KSI_00719_05982</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_16292_4430</td>
<td>97KSI_02552_02912</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_20280_15740</td>
<td>97KSI_02296_00693</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_22194_6232</td>
<td>97KSI_05069_01868</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_23853_17280</td>
<td>97KSI_00313_02395</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_24081_22178</td>
<td>97KSI_01969_05288</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_24082_20791</td>
<td>97KSI_03820_01798</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_24556_22354</td>
<td>97KSI_03988_01657</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_25739_19440</td>
<td>97KSI_03817_04140</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_26170_13373</td>
<td>97KSI_00363_01769</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_4701_17747</td>
<td>97KSI_03114_00768</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_10127_18000</td>
<td>97KSI_03988_01657</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_11340_20596</td>
<td>97KSI_00221_01329</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_14676_18009</td>
<td>97KSI_02374_02761</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_15985_17789</td>
<td>97KSI_04576_01800</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_16767_17706</td>
<td>97KSI_03516_04747</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_17994_3083</td>
<td>97KSI_03988_01657</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_18910_10184</td>
<td>97KSI_00338_04739</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19012_23767</td>
<td>97KSI_04475_04635</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19218_3762</td>
<td>97KSI_04884_05044</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_19364_12375</td>
<td>97KSI_01039_05703</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_22190_7531</td>
<td>97KSI_03913_02426</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_25563_18750</td>
<td>97KSI_02203_06434</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_25731_13884</td>
<td>97KSI_05069_01868</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_27007_19260</td>
<td>97KSI_01166_04899</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_27341_15601</td>
<td>97KSI_00844_05835</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>Date</td>
<td>Time</td>
<td>Temperature</td>
<td>Location</td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>--------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_6825_21363</td>
<td>97KSI_01814_03706</td>
<td>99.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_8517_8323</td>
<td>97KSI_03497_07247</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_11045_15726</td>
<td>97KSI_03606_06975</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_11940_23531</td>
<td>97KSI_02035_06902</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_14740_11197</td>
<td>97KSI_00784_05608</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_18468_5300</td>
<td>97KSI_00924_07082</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_20428_22822</td>
<td>97KSI_00525_06461</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_20439_16228</td>
<td>97KSI_01945_03846</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_23214_8852</td>
<td>97KSI_03817_04140</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_24774_19922</td>
<td>97KSI_03820_01798</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_25605_20745</td>
<td>97KSI_02154_04878</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_26175_14727</td>
<td>97KSI_00556_05581</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_27585_16880</td>
<td>97KSI_00143_02567</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_27967_17783</td>
<td>97KSI_00844_05835</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_3268_17783</td>
<td>97KSI_03114_00768</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_3321_19115</td>
<td>97KSI_02216_06531</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_7167_22593</td>
<td>97KSI_02154_04878</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_8517_22664</td>
<td>97KSI_00525_06461</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_10610_5650</td>
<td>97KSI_00556_05581</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_11577_10270</td>
<td>97KSI_02152_04729</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_16338_21693</td>
<td>97KSI_00844_05835</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_21055_5887</td>
<td>97KSI_02154_04878</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_22057_8800</td>
<td>97KSI_00525_06461</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_23551_25375</td>
<td>97KSI_00369_06436</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_2424_15795</td>
<td>97KSI_03820_01798</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_25771_7692</td>
<td>97KSI_05095_02527</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_27392_21231</td>
<td>97KSI_02067_02314</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_6645_12895</td>
<td>97KSI_00532_02482</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_9051_6587</td>
<td>97KSI_02082_03344</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_9364_16076</td>
<td>97KSI_04830_04581</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_9998_3612</td>
<td>97KSI_01222_07450</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_14821_4905</td>
<td>97KSI_00266_01420</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

297
<table>
<thead>
<tr>
<th>Document Code</th>
<th>Description</th>
<th>Date</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_17406_12355</td>
<td>97KSI_02853_00993</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_1776_15506</td>
<td>97KSI_03188_00866</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_18348_21528</td>
<td>97KSI_02881_01443</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_19739_2477</td>
<td>97KSI_04276_03724</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_20279_27480</td>
<td>97KSI_03397_00992</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_21130_10078</td>
<td>97KSI_03765_01443</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_22475_25010</td>
<td>97KSI_02332_06237</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_23588_16444</td>
<td>97KSI_01195_03946</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_25678_10393</td>
<td>97KSI_00863_01684</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_27207_11882</td>
<td>97KSI_01217_01798</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_2986_21960</td>
<td>97KSI_01217_02170</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_3171_18890</td>
<td>97KSI_04158_00868</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_3484_19190</td>
<td>97KSI_01262_05868</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_4779_15550</td>
<td>97KSI_00477_04066</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_6217_8406</td>
<td>97KSI_04958_03810</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_6332_24750</td>
<td>97KSI_02154_04878</td>
<td>99.47</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_6975_8420</td>
<td>97KSI_01107_06748</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_8730_9843</td>
<td>97KSI_01262_05868</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_11108_11603_11868</td>
<td>97KSI_00863_01684</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_11619_11906</td>
<td>97KSI_05154_03279</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_13180_7077</td>
<td>97KSI_00434_06523</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_16881_27613</td>
<td>97KSI_05069_01868</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_20835_27333</td>
<td>97KSI_02027_03223</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_21114_11099</td>
<td>97KSI_01359_00277</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_24721_11822</td>
<td>97KSI_00355_00902</td>
<td>99.74</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_3479_14144</td>
<td>97KSI_01817_02528</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_3955_11068</td>
<td>97KSI_03515_01655</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_4832_11089</td>
<td>97KSI_03663_01512</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_9798_3001</td>
<td>97KSI_00687_06967</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_10311_3106</td>
<td>97KSI_05069_01868</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Value</td>
<td>Price</td>
<td>Original Price</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>----------------------------</td>
<td>---------</td>
<td>----------------</td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_11107_25766</td>
<td>97KSI_04233_07078</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_11125_25767</td>
<td>97KSI_00804_06204</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_12261_15569</td>
<td>97KSI_00290_02888</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_12394_10561</td>
<td>97KSI_02835_02126</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_15314_9521</td>
<td>97KSI_02722_02279</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_15429_18641</td>
<td>97KSI_03101_05043</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_18421_5919</td>
<td>97KSI_02835_02126</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_22201_12066</td>
<td>97KSI_00196_02555</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_22429_18036</td>
<td>97KSI_04451_01423</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_23147_6138</td>
<td>97KSI_02911_03818</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_26584_8352</td>
<td>97KSI_00460_06255</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_3355_19727</td>
<td>97KSI_03820_01798</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_3918_8950</td>
<td>97KSI_01714_04568</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_4716_11773</td>
<td>97KSI_03663_01512</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_5908_24290</td>
<td>97KSI_02911_03818</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_7800_4262</td>
<td>97KSI_04158_05183</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1109_8052_12092</td>
<td>97KSI_04040_04857</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_12634_14213</td>
<td>97KSI_00906_05965</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_13121_24818</td>
<td>97KSI_00078_05612</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_13803_24457</td>
<td>97KSI_04376_07129</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_14924_26720</td>
<td>97KSI_02034_05599</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_16850_27753</td>
<td>97KSI_01819_01022</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_19679_13356</td>
<td>97KSI_03345_04692</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_19763_2849</td>
<td>97KSI_01511_05514</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_20110_10982</td>
<td>97KSI_02615_01104</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_24225_5889</td>
<td>97KSI_04219_07235</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_24768_13379</td>
<td>97KSI_03684_05943</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_25292_12984</td>
<td>97KSI_03285_02895</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_25756_19987</td>
<td>97KSI_05069_01868</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_26826_22971</td>
<td>97KSI_03663_01512</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_28181_21380</td>
<td>97KSI_02246_03171</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_2836_12986</td>
<td>97KSI_04625_06294</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_3332_21204</td>
<td>97KSI_03663_01512</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_5330_8573</td>
<td>97KSI_02154_04878</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_5925_18140</td>
<td>97KSI_02848_01222</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_0000000000-AA7DR_1_1110_5995_9355</td>
<td>97KSI_00138_05029</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>Code</td>
<td>Description</td>
<td>Value</td>
<td>Percentage</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1110_6357_22219</td>
<td>97KSI_00701_06124</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1110_7052_18315</td>
<td>97KSI_00160_02389</td>
<td>99.21</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1110_7268_15251</td>
<td>97KSI_00645_05634</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1110_7314_13849</td>
<td>97KSI_03013_02536</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1110_9986_27720</td>
<td>97KSI_02077_06508</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1111_10416_18691</td>
<td>97KSI_03195_01710</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1111_1110_7052_18315</td>
<td>97KSI_00160_02389</td>
<td>99.21</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1111_1110_7268_15251</td>
<td>97KSI_00645_05634</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1111_1110_7314_13849</td>
<td>97KSI_03013_02536</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000</td>
<td>AA7DR_1_1111_1110_9986_27720</td>
<td>97KSI_02077_06508</td>
<td>100.00</td>
</tr>
</tbody>
</table>

300
<p>| M00390_81_000000000-AA7DR_1_1112_17859_8164 | 97KSI_02057_02771 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_18973_7862 | 97KSI_01574_04337 | 100.00 |
| M00390_81_000000000-AA7DR_1_1112_20622_21848 | 97KSI_01172_05928 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_21063_5655 | 97KSI_05069_01868 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_21557_5655 | 97KSI_03663_01512 | 100.00 |
| M00390_81_000000000-AA7DR_1_1112_21936_7863 | 97KSI_03665_02501 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_22279_16278 | 97KSI_04170_06566 | 100.00 |
| M00390_81_000000000-AA7DR_1_1112_2632_13931 | 97KSI_00108_02998 | 99.74 |
| M00390_81_000000000-AA7DR_1_1112_26948_23469 | 97KSI_02399_03900 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_29292_16098 | 97KSI_00460_06255 | 99.73 |
| M00390_81_000000000-AA7DR_1_1112_3447_20796 | 97KSI_00363_01769 | 99.47 |
| M00390_81_000000000-AA7DR_1_1112_4447_20796 | 97KSI_00363_01769 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_11527_9270 | 97KSI_03998_01614 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_14267_3433 | 97KSI_04670_06771 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_16579_4869 | 97KSI_03013_02536 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_16616_9226 | 97KSI_04839_05450 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_17568_13740 | 97KSI_01744_01568 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_20112_16241 | 97KSI_00509_03810 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_21216_24600 | 97KSI_02154_04878 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_2240_17861 | 97KSI_00122_02414 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_22428_24990 | 97KSI_02898_05450 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_22706_5313 | 97KSI_01107_06748 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_24059_6884 | 97KSI_04465_06527 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_27443_14488 | 97KSI_03663_01512 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_29235_17394 | 97KSI_02154_04878 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_5250_15472 | 97KSI_05069_01868 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_6961_16472 | 97KSI_04599_03647 | 99.73 |
| M00390_81_000000000-AA7DR_1_1113_7128_23668 | 97KSI_03360_01646 | 100.00 |
| M00390_81_000000000-AA7DR_1_1113_7303_19521 | 97KSI_03820_01798 | 99.47 |
| M00390_81_000000000-AA7DR_1_1114_111218_23596 | 97KSI_04078_01285 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_11222_26655 | 97KSI_00483_01396 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_11759_4336 | 97KSI_03663_01512 | 99.73 |
| M00390_81_000000000-AA7DR_1_1114_11963_10649 | 97KSI_01969_05288 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_12633_1077 | 97KSI_03525_00856 | 99.73 |
| M00390_81_000000000-AA7DR_1_1114_13097_17018 | 97KSI_01642_06778 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_13281_21120 | 97KSI_03525_00856 | 99.73 |
| M00390_81_000000000-AA7DR_1_1114_13862_7325 | 97KSI_01924_01514 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_17435_20606 | 97KSI_00163_03224 | 99.73 |
| M00390_81_000000000-AA7DR_1_1114_17987_9687 | 97KSI_01924_01514 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_24743_8698 | 97KSI_00284_01155 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_27048_16944 | 97KSI_04945_00944 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_27901_17468 | 97KSI_00163_03224 | 99.73 |
| M00390_81_000000000-AA7DR_1_1114_3976_9516 | 97KSI_00163_03224 | 100.00 |
| M00390_81_000000000-AA7DR_1_1114_8755_21908 | 97KSI_02146_01514 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_10559_24670 | 97KSI_02146_01514 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_10960_18184 | 97KSI_03114_00768 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_11179_10541 | 97KSI_00074_04973 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_11461_26930 | 97KSI_00284_01155 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_11474_28568 | 97KSI_05069_01868 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_13016_5366 | 97KSI_00290_02888 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_15617_6895 | 97KSI_01117_04819 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_17640_3858 | 97KSI_04387_04031 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_19155_24278 | 97KSI_02876_02931 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_19403_26289 | 97KSI_01222_07450 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_21889_17023 | 97KSI_01222_07450 | 99.47 |
| M00390_81_000000000-AA7DR_1_2101_24497_5984 | 97KSI_04737_01816 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_5271_10562 | 97KSI_04797_06930 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_5652_21179 | 97KSI_00647_03319 | 100.00 |
| M00390_81_000000000-AA7DR_1_2101_9413_13422 | 97KSI_00252_04181 | 99.73 |
| M00390_81_000000000-AA7DR_1_2101_9925_11464 | 97KSI_01207_04749 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_10796_16083 | 97KSI_01503_02896 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_14629_27271 | 97KSI_00844_05835 | 99.73  |
| M00390_81_000000000-AA7DR_1_2102_15928_24133 | 97KSI_03976_06338 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_16304_20120 | 97KSI_02922_04379 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_17962_19458 | 97KSI_03363_07025 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_19871_14399 | 97KSI_03292_07193 | 99.73  |
| M00390_81_000000000-AA7DR_1_2102_20221_17635 | 97KSI_01355_02233 | 99.74  |
| M00390_81_000000000-AA7DR_1_2102_20462_22578 | 97KSI_04141_05765 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_21439_19560 | 97KSI_01541_01528 | 100.00 |
| M00390_81_000000000-AA7DR_1_2102_21454_6492  | 97KSI_04722_01603 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_16137_11494 | 97KSI_03663_01512 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_17071_11070 | 97KSI_01122_03589 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_19587_7983  | 97KSI_03663_01512 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_20177_16769 | 97KSI_00907_04235 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_21453_17225 | 97KSI_03512_07397 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_23032_14402 | 97KSI_00907_04235 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_24105_11792 | 97KSI_02622_01430 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_24798_23009 | 97KSI_04839_06993 | 99.74  |
| M00390_81_000000000-AA7DR_1_2103_25106_12067 | 97KSI_04825_01640 | 100.00 |
| M00390_81_000000000-AA7DR_1_2103_6765_6807   | 97KSI_04852_06284 | 99.73  |
| M00390_81_000000000-AA7DR_1_2103_7329_8598   | 97KSI_04576_06137 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_10106_24750 | 97KSI_04712_01509 | 99.47  |
| M00390_81_000000000-AA7DR_1_2104_10279_18534 | 97KSI_03663_01512 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_13368_5663  | 97KSI_03496_05087 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_13530_11913 | 97KSI_04822_06010 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_16410_14060 | 97KSI_02200_00828 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_17095_27177 | 97KSI_00534_04066 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_17292_20281 | 97KSI_04462_02230 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_18396_11569 | 97KSI_03249_05481 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_19307_2758  | 97KSI_04387_04031 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_21076_21928 | 97KSI_00781_05708 | 100.00 |
| M00390_81_000000000-AA7DR_1_2104_22220_11603 | 97KSI_04586_06236 | 99.73  |</p>
<table>
<thead>
<tr>
<th>Reference</th>
<th>Start Time</th>
<th>Duration</th>
<th>End Time</th>
<th>Pressure</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_5893_23497</td>
<td>97KSI_02161_00967</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_7192_20096</td>
<td>97KSI_03019_07346</td>
<td>99.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_8323_22034</td>
<td>97KSI_03663_01512</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_8410_19743</td>
<td>97KSI_03663_01512</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_8697_14774</td>
<td>97KSI_02881_05835</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_9244_24116</td>
<td>97KSI_02789_03576</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_9651_25453</td>
<td>97KSI_04072_04602</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_9651_25453</td>
<td>97KSI_04072_04602</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_9651_25453</td>
<td>97KSI_04072_04602</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_12545_11221</td>
<td>97KSI_04730_00564</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_13761_25321</td>
<td>97KSI_02144_02403</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_15404_9580</td>
<td>97KSI_01534_00636</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_22035_24001</td>
<td>97KSI_02450_02524</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_24423_17837</td>
<td>97KSI_02041_00872</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_2829_12406</td>
<td>97KSI_00349_04546</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_4621_18479</td>
<td>97KSI_03682_04245</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_4686_7578</td>
<td>97KSI_03663_01512</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_5362_18382</td>
<td>97KSI_00212_02403</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_7220_16086</td>
<td>97KSI_04737_01816</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_8038_17551</td>
<td>97KSI_05267_04117</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_9332_3887</td>
<td>97KSI_01206_04859</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_9332_3887</td>
<td>97KSI_01206_04859</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_9332_3887</td>
<td>97KSI_01206_04859</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_12662_19654</td>
<td>97KSI_00877_07286</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_13978_16069</td>
<td>97KSI_01865_02366</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_14327_12220</td>
<td>97KSI_04946_01213</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_17536_14372</td>
<td>97KSI_04244_03016</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_22862_17949</td>
<td>97KSI_04095_06760</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_23326_14538</td>
<td>97KSI_02135_03690</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_23367_19635</td>
<td>97KSI_04734_01105</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_24966_14488</td>
<td>97KSI_03585_00511</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_26240_17823</td>
<td>97KSI_05069_01868</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_27384_10333</td>
<td>97KSI_02154_04878</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_27964_16114</td>
<td>97KSI_04520_04321</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_6324_15238</td>
<td>97KSI_01807_03311</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2109_8450_15790</td>
<td>97KSI_03820_01798</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_10348_21199</td>
<td>97KSI_04818_02139</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_11897_25381</td>
<td>97KSI_03308_05912</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document Reference</td>
<td>Conversion Factor</td>
<td>Value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_12101_15963</td>
<td>97KSI_00270_02577</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_12819_22347</td>
<td>97KSI_02872_03000</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_14301_8821</td>
<td>97KSI_03622_02935</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_15881_27179</td>
<td>97KSI_04805_04634</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_22066_18385</td>
<td>97KSI_00844_05835</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_7360_6899</td>
<td>97KSI_03991_07212</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2110_9273_13717</td>
<td>97KSI_01479_03480</td>
<td>99.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_15139_12520</td>
<td>97KSI_00468_02426</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_16225_9778</td>
<td>97KSI_04271_02812</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_21122_42228</td>
<td>97KSI_01206_04859</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_27232_15351</td>
<td>97KSI_00844_05835</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_5228_18238</td>
<td>97KSI_02152_02890</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_6691_25308</td>
<td>97KSI_00893_06529</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2111_9175_9245</td>
<td>97KSI_01206_04859</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_11002_6236</td>
<td>97KSI_03975_01169</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_11893_4010</td>
<td>97KSI_02227_02888</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_13597_23479</td>
<td>97KSI_03256_02872</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_14035_13127</td>
<td>97KSI_04009_00303</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_16269_12869</td>
<td>97KSI_03663_01512</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_19587_25970</td>
<td>97KSI_04147_05597</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_2070_14066</td>
<td>97KSI_02366_03370</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_20747_7120</td>
<td>97KSI_02781_04932</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_21936_11046</td>
<td>97KSI_04852_06284</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_23769_13792</td>
<td>97KSI_03041_04640</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_2708_10279</td>
<td>97KSI_01005_02504</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_28846_15117</td>
<td>97KSI_02337_07107</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_6003_19825</td>
<td>97KSI_03876_06624</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_6092_23830</td>
<td>97KSI_00775_03521</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2112_6103_23850</td>
<td>97KSI_05069_01868</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_10012_16875</td>
<td>97KSI_00742_03662</td>
<td>99.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2113_10927_8701</td>
<td>97KSI_02898_05450</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_12507_23692</td>
<td>97KSI_03561_05871</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------</td>
<td>------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_12929_19411</td>
<td>97KSI_00993_01026</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_15500_21922</td>
<td>97KSI_05130_06212</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_16191_16892</td>
<td>97KSI_02913_03689</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_19441_18399</td>
<td>97KSI_00775_01027</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_21239_18948</td>
<td>97KSI_00529_01262</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_23441_12993</td>
<td>97KSI_01185_04616</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_2379_18430</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_5142_16083</td>
<td>97KSI_00529_01262</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_6080_24551</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2113_6687_19954</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_666_9774</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_5142_16083</td>
<td>97KSI_00529_01262</td>
<td>99.73</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_17524_25803</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_18666_9774</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_22473_15974</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_22648_22014</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_23836_19289</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_25330_24954</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>2114_9691_5295</td>
<td>97KSI_00529_01262</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

**C. sp. Na26B1**

<table>
<thead>
<tr>
<th>M00390_80_000000000-AA759_1</th>
<th>1104_19320_19832</th>
<th>97KSI_01885_05851</th>
<th>99.47</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>1107_13645_27564</td>
<td>97KSI_01885_05851</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>1107_15354_12167</td>
<td>97KSI_01911_00721</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>1111_7314_18049</td>
<td>97KSI_00361_06437</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>1113_17535_10310</td>
<td>97KSI_00493_05675</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>1114_14905_11144</td>
<td>97KSI_00794_03497</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>2108_18456_25450</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>2111_9841_2910</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_80_000000000-AA759_1</td>
<td>2112_10861_13680</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1101_16198_12414</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1101_21520_5635</td>
<td>97KSI_04233_02052</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1101_28091_16249</td>
<td>97KSI_04826_01611</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1101_9470_25874</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1102_17020_5965</td>
<td>97KSI_03486_05518</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1</td>
<td>1102_6690_11966</td>
<td>97KSI_00666_01179</td>
<td>100.00</td>
</tr>
</tbody>
</table>

307
<table>
<thead>
<tr>
<th>Document Code</th>
<th>Model</th>
<th>Specification</th>
<th>Part Number</th>
<th>Material Grade</th>
<th>Tensile Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_17202_26051</td>
<td>97KSI_01364_03881</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1103_22901_8165</td>
<td>97KSI_01348_06387</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_22570_9473</td>
<td>97KSI_01986_05212</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_27372_19474</td>
<td>97KSI_01885_05851</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1104_5102_14191</td>
<td>97KSI_00610_01398</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1105_28183_17335</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_19721_26938</td>
<td>97KSI_01612_01703</td>
<td>99.47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1106_22711_22894</td>
<td>97KSI_03374_05196</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_20021_20682</td>
<td>97KSI_04167_01063</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_25259_9417</td>
<td>97KSI_03931_01058</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1107_25259_9417</td>
<td>97KSI_03931_01058</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1108_19232_2903</td>
<td>97KSI_00130_03851</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_21257_16345</td>
<td>97KSI_00130_03851</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_1109_21257_16345</td>
<td>97KSI_00130_03851</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_11669_5727</td>
<td>97KSI_04875_05703</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_11669_5727</td>
<td>97KSI_04875_05703</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_2101_22817_8892</td>
<td>97KSI_01251_06803</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2101_2101_22817_8892</td>
<td>97KSI_01251_06803</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_11527_27123</td>
<td>97KSI_00136_06038</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_11527_27123</td>
<td>97KSI_00136_06038</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_20122_6543</td>
<td>97KSI_03758_05617</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2104_20122_6543</td>
<td>97KSI_03758_05617</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_20409_3391</td>
<td>97KSI_03921_02093</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_20409_3391</td>
<td>97KSI_03921_02093</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_24095_5901</td>
<td>97KSI_01251_06803</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_24095_5901</td>
<td>97KSI_01251_06803</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_25848_16533</td>
<td>97KSI_02800_01959</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2105_25848_16533</td>
<td>97KSI_02800_01959</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_19266_2682</td>
<td>97KSI_00136_06038</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2106_19266_2682</td>
<td>97KSI_00136_06038</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2107_6490_11438</td>
<td>97KSI_03758_05617</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2108_10756_6054</td>
<td>97KSI_03923_05450</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accession</td>
<td>Genbank</td>
<td>Gene</td>
<td>Chromosome</td>
<td>Start</td>
<td>Stop</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>------</td>
<td>------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2108_15634_22907</td>
<td>97KSI_01393_04532</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2109_22304_14960</td>
<td>97KSI_02828_01344</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2110_24767_14132</td>
<td>97KSI_01986_05212</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2111_19997_13380</td>
<td>97KSI_04190_02727</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2112_8563_10293</td>
<td>97KSI_02913_07422</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_2113_12146_19815</td>
<td>97KSI_00915_03044</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. tenuissimus</td>
<td>M00390_81_000000000-CA7DR_1_1101_11040_7536</td>
<td>97KSI_04069_00269</td>
<td>99.74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_12020_2561</td>
<td>97KSI_02748_05003</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_12750_12735</td>
<td>97KSI_02983_06904</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_12972_26204</td>
<td>97KSI_00416_02071</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_13180_26564</td>
<td>97KSI_02658_06876</td>
<td>99.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_17418_16093</td>
<td>97KSI_04894_03826</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19019_24975</td>
<td>97KSI_00416_02071</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19124_2835</td>
<td>97KSI_03794_07273</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19390_3055</td>
<td>97KSI_00416_02071</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19499_23881</td>
<td>97KSI_01509_04247</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19579_23505</td>
<td>97KSI_00974_05580</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_19596_23537</td>
<td>97KSI_01970_01571</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_20308_10428</td>
<td>97KSI_00368_02076</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_21482_26412</td>
<td>97KSI_01763_06817</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_22682_4982</td>
<td>97KSI_00447_06375</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_22809_8498</td>
<td>97KSI_00596_07083</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_24859_6431</td>
<td>97KSI_002268_05990</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_25018_11322</td>
<td>97KSI_00093_05704</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_25410_22880</td>
<td>97KSI_00814_03179</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_27830_16975</td>
<td>97KSI_03912_02912</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_28042_10043</td>
<td>97KSI_00404_01043</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_28592_8549</td>
<td>97KSI_04782_03918</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1101_7389_17078</td>
<td>97KSI_00652_03222</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1102_13482_6478</td>
<td>97KSI_01911_03168</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1102_18446_3596</td>
<td>97KSI_01987_05213</td>
<td>100.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1102_20566_15022</td>
<td>97KSI_03923_05350</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-CA7DR_1_1102_21169_3932</td>
<td>97KSI_01798_07171</td>
<td>99.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ID</td>
<td>FileID</td>
<td>Location</td>
<td>Distance</td>
<td>KSI Value</td>
<td>Tolerable</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_21445_15640</td>
<td>97KSI</td>
<td>01290</td>
<td>05422</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_23013_11309</td>
<td>97KSI</td>
<td>03894</td>
<td>07381</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_25920_11969</td>
<td>97KSI</td>
<td>02564</td>
<td>05554</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_28179_19330</td>
<td>97KSI</td>
<td>02564</td>
<td>05554</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_4023_9561</td>
<td>97KSI</td>
<td>00416</td>
<td>02071</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_7026_20095</td>
<td>97KSI</td>
<td>02003</td>
<td>04853</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1102_8293_25185</td>
<td>97KSI</td>
<td>02113</td>
<td>03202</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_11989_21537</td>
<td>97KSI</td>
<td>01150</td>
<td>04900</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_16869_20837</td>
<td>97KSI</td>
<td>01982</td>
<td>02174</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_20846_6606</td>
<td>97KSI</td>
<td>05161</td>
<td>07286</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_20219_18162</td>
<td>97KSI</td>
<td>04538</td>
<td>02642</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_21707_22085</td>
<td>97KSI</td>
<td>03738</td>
<td>02676</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_26982_11585</td>
<td>97KSI</td>
<td>03391</td>
<td>02324</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_30469_22966</td>
<td>97KSI</td>
<td>02642</td>
<td>05819</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_6137_19542</td>
<td>97KSI</td>
<td>02347</td>
<td>01990</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1103_9573_27386</td>
<td>97KSI</td>
<td>03871</td>
<td>06777</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1104_20219_18162</td>
<td>97KSI</td>
<td>05087</td>
<td>02289</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1104_25687_10430</td>
<td>97KSI</td>
<td>00416</td>
<td>02071</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1104_3967_7974</td>
<td>97KSI</td>
<td>05087</td>
<td>02289</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1105_12334_10159</td>
<td>97KSI</td>
<td>0476</td>
<td>02252</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1105_14401_23759</td>
<td>97KSI</td>
<td>03045</td>
<td>03726</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1105_19354_7845</td>
<td>97KSI</td>
<td>01283</td>
<td>00600</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1105_25599_18512</td>
<td>97KSI</td>
<td>00543</td>
<td>01739</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1106_19455_13394</td>
<td>97KSI</td>
<td>00373</td>
<td>04914</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1106_24238_20243</td>
<td>97KSI</td>
<td>03847</td>
<td>07014</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1107_22105_11345</td>
<td>97KSI</td>
<td>02624</td>
<td>05741</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1107_22838_6403</td>
<td>97KSI</td>
<td>04102</td>
<td>01861</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1108_11446_5613</td>
<td>97KSI</td>
<td>02124</td>
<td>07354</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1108_16719_27399</td>
<td>97KSI</td>
<td>01118</td>
<td>03038</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1108_24620_23287</td>
<td>97KSI</td>
<td>01788</td>
<td>02370</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1110_13450_22152</td>
<td>97KSI</td>
<td>00416</td>
<td>02071</td>
<td>99.74</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1110_14154_2636</td>
<td>97KSI</td>
<td>02805</td>
<td>02859</td>
<td>100.00</td>
</tr>
<tr>
<td>M00390_81_0000000000</td>
<td>AA7DR_1_1110_20392_15399</td>
<td>97KSI</td>
<td>04483</td>
<td>06080</td>
<td>99.74</td>
</tr>
<tr>
<td>Material ID</td>
<td>Lot ID</td>
<td>97KSI</td>
<td>Price</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------------------</td>
<td>-----------------------------</td>
<td>-------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_12713_17929</td>
<td>97KSI_01150_06546</td>
<td></td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_24850_7921</td>
<td>97KSI_02564_05554</td>
<td></td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M00390_81_000000000-AA7DR_1_2114_25314_9203</td>
<td>97KSI_02622_03603</td>
<td></td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter VI

Concluding remarks and future perspectives
6.1. Concluding remarks

This Ph.D. thesis embodies my contribution to the understanding of the evolution of the marine diatom family Chaetocerotaceae and, in particular, of the genus *Chaetoceros* by means of molecular data. In some cases, molecular data have been used in their canonical way and proved to be conclusive for the purposes they were intended to. This was, for example, the case of the multigene phylogeny inferred in Chapter II to assess the evolutionary history of Chaetocerotaceae. In other cases, molecular data (especially in the form of metabarcoding data), have been used in a new, different way and played the role of main actors in stories that went beyond *Chaetoceros* or diatoms in general. This is what happened in Chapters III, IV and V, in which I have designed a series of experiments that have shown the potential of metabarcoding data in so far unexplored contexts.

For **Chapter II**, I started my experiments with the initial idea of inferring a multigene phylogeny of the family Chaetocerotaceae to resolve terminal or internal relationships that were poorly supported in previous nuclear phylogenies (e.g. Kooistra et al., 2010; Gaonkar et al., 2018). Then, considering that in our lab we had reached a considerable number of strains of Chaetocerotaceae belonging to different species around the world and that there was a renewed interest in revision of sections triggered by the discovery of new species (e.g. Li et al., 2013; 2016; Xu et al., 2019), I decided to change my plans. I kept the initial idea of inferring a multigene phylogeny for the family Chaetocerotaceae, but I also decided to test the traditional classification scheme based in generic and infrageneric (subgenera and sections) divisions using the inferred phylogeny as backbone. Taxonomies are not neutral, but they reflect (or even create) the hypothesis on the structure of living world (Gould and Vrba, 1982). When one looks at how people classify things, one also understands how they think (Foucault, 1970). Therefore, I aimed at a classification scheme that was supported phylogenetically but also retained practical properties, following the thinking of Mayr (1982) and Benton (2000). My classification of Chaetocerotaceae had to
group together species similar because of common descent (phylogenetically informative) and in the meantime allow these groups to aid people in the identification of new or already known species (utilitarian principle, practical purpose). I dusted off the traditional classification scheme and, with some adjustments (emendation of one section, rejection of seven and erection of three new ones) and I made it fit to the clades of the inferred multigene phylogeny. This work made it possible to keep most of the traditional systematic terminology but in the light of a modern and updated interpretation. I tried to avoid leaving clades nameless wherever and whenever I could, because I believe that things without a name tend to be disregarded. Furthermore, giving priority to the utilitarian criterion, I refrained from classifying the major, well-supported clades within *Chaetoceros* into their own genera, since *Chaetoceros* species are easily recognised by their defining feature, the setae, whereas each of such more narrowly defined genera would not be recognised so easily. Splitting would have created a series of genera that are not always easy to distinguish.

In **Chapter III** I have shown how the integration of classical occurrence data and new ones (metabarcoding data) can be used to obtain a comprehensive assessment of the distribution of species, especially of microscopic ones such as protists. Classical occurrence data as reports of scientific expeditions, floras and faunas and checklists have formed the main sources of primary biodiversity data for inferring species distribution (Droege et al., 1998; Chapman, 2005). In recent years, occurrence data have also been gathered from a large variety of sources as satellite tracking and direct or remote observation (He et al., 2015), frozen tissue collections and seed banks (Chapman, 2005), environmental DNA (August et al., 2015), and citizen science initiatives (Devictor et al., 2010; Hochachka et al., 2012). However, a big step forward has been done with the adding of DNA information to classical approaches. This kind of data, have revolutionised the study of protistan diversity (Leray and Knowlton, 2016; Caron and Hu, 2018) that was
before exclusively based on morphological studies. For marine protists indeed, there are several challenges related to the assessment of diversity and distribution at different taxonomic levels, the species one being particularly difficult. Cryptic diversity is widespread (Smayda, 2011; Amato et al., 2019) and traditional analyses based on microscopy are time-consuming and require taxonomic expertise (Culverhouse, 2007). The availability of global metabarcoding datasets as Ocean Sampling Day (OSD, Kopf et al., 2015) and Tara Oceans (de Vargas et al., 2015) has offered a valuable source of sequence and occurrence data that fostered the assessment of diversity and distribution of several marine taxa (de Vargas et al., 2015; Malviya et al., 2016; Tragin and Vaulot, 2018; 2019). In contrast to Tara Oceans, which sampled different marine regions at different times of the year, OSD is a simultaneous sampling of coastal regions (mostly Northern Hemisphere), which allows analysis of spatial distribution patterns of species without the impact of seasonality (Tragin and Vaulot, 2019). For my thesis work, I decided to use the information available in these metabarcoding datasets together with other stored in public repositories (GBIF and OBIS) as well as phytoplankton checklists or floras to show how the integration of these data can contribute to insight in the biogeography and diversity at the genus- and species-level in *Chaetoceros*. I extracted *Chaetoceros* records from GBIF and OBIS, collected literature data by means of a Google Scholar search and mapped *Chaetoceros* references barcodes against OSD (144 sites) and Tara Oceans (210 sites). I compared the resolution of these different data sources in determining the global distribution of the genus and provided examples, at the species level, of detection of cryptic species, endemism and cosmopolitan or restricted distributions. Of all the non-molecular data, the most complete picture of *Chaetoceros* distribution was provided by the GBIF and OBIS platforms, which contain a huge amount of data from different sources and cover a wide time scale. The search on Google Scholar could be considered as a convenient starting place to commence a literature search but not an endpoint. The two
global metabarcoding datasets OSD and Tara Oceans provided an overall distribution of
the genus that was comparable to the one obtained from GBIF and OBIS. This proved that,
despite their bias in space and time, metabarcoding data can compete with classical
occurrence data gathered over hundreds of years. I also produced maps for the genus
containing info about occurrence, species richness and abundance, as well as Chaetoceros
species distribution maps from OSD and Tara Oceans data. Finally yet importantly, in this
chapter I have provided a pipeline to study occurrence and diversity of taxa for which
reference barcodes and metabarcoding data are available.

As stated in the Abstract of this Ph.D. thesis, the initial aim of Chapter IV was to infer the
phylogeographic pattern of selected Chaetoceros species by means of Sanger sequencing
of a few genes from specimens collected around the world. Then, it turned into the analysis
of the C. curvisetus species complex inferring haplotype networks from metabarcoding
data. The choice of changing strategy was made to take advantage of the global
metabarcoding datasets of OSD and Tara Oceans, which together covered about 350
sampling sites across coastal and open ocean waters of both hemispheres. Reaching even a
small fraction of such sampling localities would have been hard considering the duration of
a Ph.D. program, and the costs related to the selection and sequencing of target gene
regions quite high. Then, the change of the subject, from the comparison of the
phylogeographic patterns of different Chaetoceros species to the analysis of a species
complex, was a consequence of the results I obtained from the multigene phylogeny
inferred in Chapter II. Indeed, some phylogenetic relationships among C. curvisetus
species were not fully resolved even including more loci, which made me suppose that the
relationships among them were more complex than simple dichotomies. Therefore, I
decided to explore the patterns of genetic variation of 18S gene (V4 and V9 regions) across
space to reconstruct the evolutionary relationships of the aforementioned species.
Metabarcoding data have been used so far in the form of phylogenetic trees or OTU clustering (Nanjappa et al., 2014; Gaonkar, 2017; Pargana, 2017; Tragin and Vaulot, 2019) to delimit species in protists, but none has built haplotype networks with them.

For my experiment, I started from a set of reference barcodes of *C. curviset us* spp. produced by Gaonkar et al. (2018) and myself (e.g. strains of the Red Sea, see Chapter II) and two global metabarcoding datasets (OSD and Tara Oceans). The latter datasets allowed me to explore the genetic diversity within my cryptic species complex in a way that would have been hard to reach with classical Sanger sequencing data. Since the object of my study was a species complex, I supposed that the best way to analyse it was by means of phylogenetic haplotype networks rather than phylogenetic trees. Therefore, I set up several criteria to delimit species from my networks. Then, I validated at molecular level the species inferred above by means of inference of Maximum Likelihood phylogenetic trees and calculation of genetic distances. After this, I moved to an ecological level, and I have mapped the inferred *C. curviset us* species in the biogeographic provinces of Longhurst (2007) using the information contained in the two global metabarcoding datasets. This latter exercise allowed me to test from the ecological perspective the species I have inferred from genetic data.

In conclusion, I confirmed as species the initial taxa for which I had reference barcodes and that there are four more molecularly defined taxonomic units (MOTUs) that need further investigation, some of which are likely to constitute species new to science. Furthermore, within the *C. curviset us* species complex it seems to still be gene flow.

The final experiment of this Ph.D. thesis, described in Chapter V, initially was not planned at all and resulted from some preliminary results of Chapter IV. It is a story of concerted evolution of 18S gene in several *Chaetoceros* species and inferred from metabarcoding data. Concerted evolution is the mode of evolution of some genes and non-coding regions across all the major branches of the Tree of Life and was first detected by
hybridisation studies and successively by phylogenetic approaches (Graur and Li, 1999). Here, for the first time, I showed how metabarcoding data and single strain high throughput sequencing (HTS) could be used to study this biological phenomenon. Using such data in the form of abundance plots, BLAST analysis and haplotype networks, I have demonstrated that concerted evolution is occurring in all of the investigated species, and all the methodologies here used for its detection are conclusive and easy to perform.

The work presented in this chapter also demonstrated that there are no consequences for DNA barcoding due to the occurrence, within each Chaetoceros strain, of thousands of 18S ribotypes. Indeed, one of the copies, identified as the dominant haplotype, is far more abundant than all the others that the probability that a “minor” haplotype is sequenced with Sanger chemistry is almost null. I have also demonstrated that, when conducting metabarcoding experiments (from both environmental samples and bulk communities) or single strain HTS, the most abundant haplotype that is recovered for each species corresponds to the sequence that would be obtained by Sanger sequencing. However, this study also highlighted that the high number of sequences occurring at low abundances (minor haplotypes) can inflate diversity assessments inferred from metabarcoding data.

In conclusion, my Ph.D. thesis:

- is a contribution to the systematics of the family Chaetocerotaceae (Chapter II);
- provides an assessment of the diversity and distribution of the genus Chaetoceros by integrating classical and novel primary biodiversity data (Chapter III);
- shows a new way to analyse a cryptic species complex using the potential of spatial data contained in global metabarcoding datasets in the form of phylogenetic networks (Chapter IV);
- illustrates how starting from the data contained in a temporal metabarcoding dataset (MareChiara), I have formulated the hypothesis of concerted evolution in
Chaetoceros that was successively tested with an appropriate experimental design (single strain HTS and targeted analyses).

6.2. Future perspectives

During the 3 years of my Ph.D. program, I performed several experiments that have contributed to my understanding of the diversity and evolution of the family Chaetocerotaceae and, in particular, of the genus Chaetoceros. However, no experiment is to be considered definitive and, in this sense, my Ph.D. opens several research perspectives. Strictly related to the work performed in this thesis, I see the following possibilities. The multigene phylogeny (Chapter II), although taxonomically comprehensive regarding the known diversity, leaves a few additional sections still to be investigated. Future work needs to include species not treated here to test the validity of my proposed classification system, and to provide a more comprehensive view of the evolutionary history of this family. In particular, the addition of molecular data for *C. bacteriastroides* will clarify its phylogenetic position in the family Chaetocerotaceae. This species exhibits features of *Bacteriastrum* and *Chaetoceros*. Hernández-Becerril (1993) placed this species into its own subgenus (*Bacteriastroidea*), and here I considered it as section until new data become available.

Another course of action is to infer a cladogram from morphological characters and their states, especially ultrastructural ones, to ascertain if and in how far its topology agrees with that of the molecular phylogeny. Ultrastructural details of valves and setae are increasingly becoming available for Chaetocerotaceae (e.g. Chamnansinp et al., 2015; Bosak and Sarno, 2017; Gaonkar et al., 2018; Xu et al., 2019). The sections here investigated could be further supported by these data, providing a more robust basis to the hypothesis of evolutionary relationships here inferred.
About the inference of genus and species distributions through integration of classical and novel strategies (Chapter III), future research could include the application of this pipeline to different taxa. Integration of metabarcoding occurrence data with classical ones can be used for conservation planning (Rondinini et al., 2006; Newbold, 2010), species distribution models (Elith et al., 2006; Lütolf et al., 2006), and many other ecological and evolutionary applications. Similarly, metabarcoding data could be used to infer phylogeography or analyse other marine species complexes, as I did in Chapter IV. In addition, the molecular data analysed in this thesis, especially in Chapter III and IV (analysis of the genus Chaetoceros and the C. curvisetus species complex respectively) could be integrated with the large amount of imaging data produced by the TARA Oceans initiative. These imaging data, collectively included in the T.A.O.M.I (TArA Oceans Marine biology Imaging) platform, refer to observations of plankton organisms (from a few micrometres to one centimetre) gathered from flow analysis, microscopy and macro-photography. Images from flow analysis were obtained using the FlowCam, an equipment consisting of a cytometer and a microscope enabling to swiftly follow organisms of very different sizes, whilst microscopy pictures were taken by means of stereomicroscopy, fluorescent microscopy and fluorescent microscopy with phase. Finally, T.A.O.M.I. platform also includes a video and macro-photographies of large planktonic organisms (e.g. larvae, jellyfish, etc.), as well as corals and macroscopic algae. Further information is available at https://oceans.taraexpeditions.org/en/m_science/news/imaging-during-tara-oceans/.

The work on concerted evolution in Chaetoceros illustrated in Chapter V could be integrated with an assessment of rDNA copy number in the species here investigated using the combination of single-cell approaches and Digital PCR. Indeed, starting with the extraction of RNA from a single cell, using Digital PCR will be possible to count all the copies of rDNA genes (or a specific gene of the cistron). These data will be then compared.
with the ones here obtained from high throughput sequencing to determine the number of rDNA copies occurring within each Chaetoceros strain. Besides this, it would be interesting to analyse the distribution of 18S-V4 haplotypes here obtained over time in many species to assess how concerted evolution interplays with other processes (drift, geographic patterning, migration, gene flow, spore capital).

Apart from establishing HTS metabarcode time series in many coastal and oceanic regions and connecting the obtained data for meta-analysis, I believe there are several topics related to the ones treated in my thesis worthy of being investigated. For example, I believe that time has come to sequence the genomes of several Chaetoceros species, taking advantage of reduction of time and costs of sequencing technologies. The occurrence within the genus of species with different life-strategies and habits (e.g. spore formers vs. non spore formers; coastal vs. oceanic species), morphological features (chloroplasts migrating in the setae vs. only in the cell body) and different usage of silica for setae formation (thin vs. thick), just to cite a few examples, allows for comparative genomics studies. Indeed, comparing the genomes of closely related species with different ecological and/or morphological traits (e.g. the members of the sections Chaetoceros versus Protuberantia) may reveal genetic factors responsible for these characteristics. Furthermore, the comparison of the genomes of Chaetoceros species versus other diatoms (publicly available), will shed light on the structure and function of core genes responsible for silica production and translocation, putative new genes involved in the formation of setae as well as carbon metabolism (e.g. C3 vs. C4).

References


324
Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., ...
distributions from occurrence data. *Ecography, 29*(2), 129-151.


Gaonkar, C. C., Piredda, R., Minucci, C., Mann, D. G., Montresor, M., Sarno, D., Kooistra,
the planktonic diatom family Chaetocerotaceae. *PLoS ONE, 13*(12), e0208929.

*Paleobiology, 8*(1), 4-15.


He, K. S., Bradley, B. A., Cord, A. F., Rocchini, D., Tuanmu, M. N., Schmidtlein, S., ...
Pettorelli, N. (2015). Will remote sensing shape the next generation of species

Hernández-Becerril, D. U. (1993). Note on the morphology of two planktonic diatoms:
*Chaetoceros bacteriastroides* and *C. seychellarus*, with comments on their
taxonomy and distribution *Botanical Journal of the Linnean Society, 111*(2), 117-
128.

(2012). Data-intensive science applied to broad-scale citizen science. *Trends in
Ecology & Evolution, 27*(2), 130-137.

Montresor, M. (2010). Comparative molecular and morphological phylogenetic


Li, Y., Lundholm, N., Moestrup, Ø. (2013). *Chaetoceros rotosporus* sp. nov. (Bacillariophyceae), a species with unusual resting spore formation. *Phycologia*, 52(6), 600-608.


