Morphometric and molecular investigations of species limits in *Cyclotella meneghiniana* (Bacillariophyceae) and closely related species

Dissertation
zur Erlangung des Akademischen Grades
eines Doktors der Naturwissenschaften

-Dr. rer. nat.-

im Fachbereich 2 (Biologie/Chemie)
der Universität Bremen

vorgelegt von

Bánk Beszteri

Bremerhaven, März 2005

Erster Gutachter: Prof. Dr. G.O. Kirst
Zweiter Gutachter: Prof. Dr. Ulrich Bathmann

Tag und Ort des öffentlichen Kolloquiums: 9. Mai 2005, Universität Bremen

Hiermit erkläre ich, dass ich die Vorliegende Dissertation selbständig verfasst und keine anederen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die entnommenen Stellen aus benutzten Werken wurden wörtlich oder inhaltlich als solche kenntlich gemacht.

Bánk Beszteri

1. General introduction

1.1. The problematic 'species'

The basic task of the scientific discipline of systematics is to classify organisms. Whereas classification of objects in general is one of the most common mental exercises occurring in all fields of life and science, classification of living organisms in modern biology is in a number of ways special. Classification in its widest sense is simply based on similarities of objects, as was the case with the classification of organisms before the 19th century. However, the recognition of the role of evolutionary history in leading to observed patterns of similarities and dissimilarities among organisms has caused the field of biological systematics to become into close contact with evolutionary studies.

The main practical importance of systematics lies in the fact that the classifications produced by it can be used as predictive tools (Alverson & Theriot 2005): organisms close to each other in a classification are expected to be more similar to each other than ones grouped into different groups at any level. Because recent features of organisms reflect their evolutionary history, producing classifications useful as such predictive tools needs consideration of the evolution of the organisms concerned.

The systematic category 'species' was particularly strongly affected by these conceptual changes. Whereas it continued to be used to refer to taxonomic units, it gained several new aspects as well. Species have become entities dynamically arising through complex evolutionary processes, generally kept together by gene flow and separated from each other by reproductive barriers. The increasing understanding of microevolutionary processes and the range of different perspectives applied in research of 'species' has lead to the 'species problem' becoming one of the most debated topics in biology.

The fact that 'species' has not lost its central role as the 'unit of biodiversity' in spite of this conceptual chaos illustrates its practical importance for a wide range of biological disciplines that continue to use species names as labels for organisms. Recently, the species debate has started to consolidate and it is being noted repeatedly that the different species concepts recommended by different authors often do not have very different practical implications for species delimitation / identification, but rather they have raised a number of (in part complementary) issues that can improve species level systematics (Avise & Wollenberg 1997; Harrison 1998; Mann 1999).

Besides the conceptual difficulties associated with 'species', our way of thinking about them has also been heavily affected by technological developments, which allow us to sample the organismal phenotype more widely or in more detail than before. However, perhaps the most important group of technological developments for this field has been that of molecular genetic methods, now allowing for direct sampling of genetic variation present in nature.

The above conceptual and technological developments have impacted upon species level systematics of microscopic organisms. The view is gaining increasing acceptance that species delimitation in these organisms is often a non-trivial scientific task, which needs to consider as many independent (pheno- and genotypic) lines of evidence as possible and should not ignore microevolutionary processes involved in shaping these patterns of variation (Mann 1999). Data accumulating from different sources – investigations of reproductive compatibilities (Mann 1984; Mann 1989), comparative DNA sequencing studies (De Vargas *et al.* 1999; Lundholm *et al.* 2003; Sáez *et al.* 2003; Sarno *et al.* 2005), population genetic surveys (Rynearson & Armbrust 2004) – are making clear that morphospecies often do not correspond to species in a microevolutionary sense in these organisms. Instead, morphospecies of microorganisms are often either complexes of multiple, reproductively

isolated species, cryptic or sibling species (De Vargas *et al.* 2004), or of clonal lineages (Tibayrenc & Ayala 2002).

1.2. Species concepts

Even though it is impossible to review the topic in full depth, an outline of the some of the most important species concepts is given here, more with the aim of introducing different aspects of the species problem than to arrive at THE species concept to be used. As noted above, the different species concepts represent different perspectives of different research fields (notably, systematics and population genetics); some focus on mechanisms of speciation whereas others on patterns observed; and importantly, they often only focus on organisms within a limited phylogenetic group, often metazoans or plants. Because of these reasons, probably none of the different species concepts has a completely general validity; in part, they raise complementary issues, different subsets of which might be relevant to a particular case (Harrison 1998).

The morphological species concept (**MSC**) is the oldest and still probably most widely used species concept; it seems to be so basic that several reviews of species concepts do not even mention it. The concept is based on the notion that morphological variation observed in nature is not continuous. According to this concept, species are morphologically distinct groups of organisms. The phylogenetic species concept, **PSC** (Cracraft 1989; Nixon & Wheeler 1990), is a more general concept: it emphasises that species are diagnosably distinct from each other, but does not specify what kinds of characters to use for diagnosis. Although the MSC and the PSC in their original form do not make any reference to evolutionary or ecological processes, they are recently also often used in somewhat different ways. For example in diatom taxonomy, morphological distinctness of sympatric (geographically not isolated) populations has often been interpreted as suggestive of reproductive isolation between the morphs (Theriot & Stoermer 1984a; Lange-Bertalot 1990).

The latter aspect, the notion of reproductive isolation, in fact originates from the second most prominent species concept, the biological species concept, **BSC** (Mayr 1963; Dobzhansky 1970), and plays a central role in a number of other species concepts as well (Paterson 1985; Templeton 1989). The BSC is one of the results of the evolutionary genetic synthesis around the middle of the last century. It emphasises the importance of reproductive isolation mechanisms in keeping species distinct. The basic notion that sexual reproduction is the way of sharing genetic information lead to the insight that the acquisition of some reproductive barrier is necessary to prevent genetic homogenisation of incipient species. One of the main practical impacts of the BSC has been the application of mating compatibility experiments in diatom systematics (Mann 1984; Mann 1989), but its impact upon our general understanding of species was more general (Lange-Bertalot 1990).

Another, closely related – complementary – aspect of sexual reproduction is emphasised by the cohesion species concept, **CSC** (Templeton 1989). The CSC focuses attention on the factors leading to a homogenisation of genetic variation within a species, which make a species an evolutionary unit - the latter aspect is in the main focus of the evolutionary species concept (Wiley 1978). Sexual reproduction is one of the most general cohesion mechanisms. However, in prokaryotes for instance, an equally important mechanism leading to cohesion within species is the recurrent occurrence of selective sweeps, which might enforce cohesion even in the lack of sexual reproduction (Majewski & Cohan 1999; Cohan 2002). Another important aspect emphasised by the CSC is demographic exchangeability. Organisms that are demographically exchangeable are ecologically equivalent. This aspect gains importance in some special cases, like with sympatric asexual

lineages: these are considered conspecific by the CSC if they are demographically exchangeable (Harrison 1998).

Alternative mechanisms of cohesion within species and of species divergence besides sexual reproduction and reproductive isolation are recently gaining increased attention. It is becoming clear that sexual reproductive barriers in nature are often not simply 'present' or 'absent', but something in between (Streit *et al.* 1994; Muir *et al.* 2000; Daguin *et al.* 2001; Ortiz-Barrientos *et al.* 2002; Schardl & Craven 2003; Miller & Van Oppen 2003; Wu & Ting 2004): hybridisation might occur between clearly distinct species (Muir *et al.* 2000) and conspecific parapatric populations might diverge in spite of gene flow between them (Wilding *et al.* 2001). Selection is a key player in such scenarios: divergent selection can play an important role in leading to the divergence of reproductively not completely isolated groups of organisms (Wu 2001; Wu & Ting 2004), whereas, as noted above, the role of selective sweeps in the cohesion of even asexual lineages also gains recognition (Cohan 2002).

Concepts like the genealogical species concept (**GSC**) complement the views of the MSC or PSC by making reference to genealogical relationships of organisms (Baum & Shaw 1995). According to the GSC, a species is an exclusive group; its members are more closely related to each other than to any other organism. The origin of this concept is related to the spreading use of molecular methods and of genealogical trees in species studies. Similar is the case with the genotypic species cluster definition: this directs attention to the deficit of intermediates between species in a genetic sense (Mallet 1995). Deficit of intermediates is interpreted at single loci as a deficit of heterozygotes and as "strong correlations or disequilibria between loci that are divergent between clusters" at multiple loci, concepts this thesis makes use of.

The use of different species concepts often does not result in any difference in the practice of species delimitation. As noted, there are in part strong complementarities between different species concepts. E.g., the acquisition of a reproductive isolating mechanism between populations leads to a deficit of intermediates in the population genetic sense (heterozygote deficit and linkage disequilibrium). It also results in the divergence of the populations both in divergent adaptive and neutral traits, thus making them diagnosably distinct both at the genetic and phenotypic level. After some generations, members of one population will be closer relatives of each other than of any member of the other population etc. Such scenarios, synthesising different aspects, have become the basic framework of thinking about species. However, there are several factors making some real life situations more complicated.

One of them is the case of asexual organisms. Clearly, the above scenario would not apply to them. Species concepts based exclusively on the importance of sexual reproduction, like the BSC, also cannot be applied in such cases. Concepts making no reference to evolutionary processes, like the MSC or the PSC, can deal with these cases as well: any group diagnosably distinct from other groups can be a species, whatever the evolutionary background of this distinctness. When also concerning evolutionary processes, the CSC is of great relevance in the case of clonal organisms, because it emphasises that other cohesion mechanisms besides sexuality might be in operation in such cases, and because of its reference to the demographic exchangeability of lineages when defining species.

Other important problematic cases are those of hybridising species, occurring among land plants, animals, and also algae and microorganisms (Streit *et al.* 1994; Barton 2001; Coyer *et al.* 2002; Schardl & Craven 2003; Miller & Van Oppen 2003; Arnold 2004; Riginos & Cunningham 2005), and of intraspecific adaptive divergence (Muir *et al.* 2000; Schlötterer 2002; Schlötterer 2003). It is becoming clear that reproductive isolation is not all or nothing, at least in two senses. On the one hand, simply the frequency of mating between populations might be anything between 0 and 1, sometimes leading to divergence simply because of the low number of individuals exchanged by populations, in spite of their complete reproductive

compatibility (Fraser & Bernatchez 2001). On the other hand, gene flow between populations might be limited to specific regions of the genome. Two extremes are when a small number of advantageous genes 'jump' through species boundaries (Morjan & Rieseberg 2004), and when gene flow between parapatric populations is limited only at a small number of adaptively diverged loci (Beaumont & Balding 2004; Storz 2005). A third kind of problem might arise through polymorphisms retained from ancestral species, the coexistence of which can even be stabilized by balancing selection. An example for this is the famous case of MHC genes: several MHC alleles found in humans are more closely related to chimp alleles than to other human alleles (Harrison 1998).

Summarizing, species delimitation might be straightforward, there are cases where the use of most species concepts would lead to the same result, but more difficult cases also exist. Generally, studies of species delimitation should consider as many factors as possible, not only to arrive at a sensible taxonomic decision about species limits, but also to better understand the nature of the species concerned: their diagnostic features, the permeability of boundaries between them, and the factors leading to cohesion within and divergence between them.

1.3. Microdiversity markers

The potential range of information sources useful for species delimitation is wide. Clearly, organisms are much too complex to allow systematics to deal with all their features. The concept of 'markers' is connected to this difficulty: microdiversity studies only use a limited subset of the features of organisms but try to gain knowledge by their use that is valid about the organisms concerned. Different kinds of markers have different (practical and theoretical) strengths and weaknesses; below the types of markers most relevant for this thesis are reviewed shortly. A general consensus in any case is that independent sources of evidence should ideally be used in species delimitation studies, and congruence between multiple, independent data sets is the strongest argument supporting conclusions about species boundaries in the case of sexual organisms (Mann 1999).

1.3.1. Phenotypic markers

Morphology is the most generally used form of phenotypic markers for systematic purposes, but others exist as well. Another widely used example is the array of biochemical tests routinely used in microbiology for strain identification and species delimitation. Basically, anything observed on organisms is potentially useful for trying to identify distinct groups and diagnostic features, provided the extent is known to which the features concerned are representative of the organism as a whole. If variation in a phenotypic trait is caused by underlying genetic variation, it is representative of the organisms in a different way than traits varying through the life cycle or ones involved in physiological adaptation for instance (Trainor & Egan 1991; Trainor 1991). One of the largest difficulties associated with the use of phenotypic markers lies in the fact that this information (i.e., whether a trait is fixed throughout the life cycle or is it capable of plastic changes in response to environmental changes) is often not readily available.

1.3.1.1. Morphology

Diatom systematics is principally based upon morphological features of the silicate cell walls (frustules). The major reasons for this are the following: the frustules are easily prepared; they show an unparalleled morphological diversity in the microbial world, providing systematics with a large number of characters; and they are often well preserved in sediments, making diatoms well suited for palaeo-limnological studies as well besides investigations of extant assemblages (Stoermer and Smol 1999). One of the major technological advances in diatom systematics was electron microscopy in the second half of the last century, which revealed even more minute morphological detail useful for their identification and systematics (Round, Crawford, and Mann 1990). In spite of these practical advantages, species level diatom systematics is still far from being unproblematic (Mann 1999; Sarno *et al.* 2005). The most widely observed problem is the lack of obvious morphological discontinuities, which gives place for subjective interpretations of minor morphological variation and to debates about species boundaries (Mann 1999). However, more fundamental problems with exclusively morphology-based systematics are also gaining increasing attention. These will be reviewed below, after reviewing methods for quantitative morphological investigations.

1.3.1.2. Quantitative morphology

Diatom systematics based exclusively upon morphology of cleaned frustules tries to identify gaps (discontinuities) in morphological variation (Lange-Bertalot 1990; Mann 1999). Although morphological discontinuity is considered important and its presence / absence is often not obvious, the use of quantitative approaches in diatom systematics is relatively rare. This is true in spite of the fact that advances made in the field of quantitative studies of morphology achieved in the last decades have been enormous (Rohlf & Marcus 1993), and that diatoms possess a large number of morphological features useful for quantitative characterization (du Buf and Bayer 2002).

Several different morphometric methodologies are available now, the major differences among them being in the methods of modelling the morphologies for multivariate analyses (Reyment, Blackith, and Campbell 1984; Bookstein, Chernoff, Elder, Humphries, Smith, and Strauss 1985; Dryden and Mardia 1998). Some of the methods also allow the application of special data analysis methods, but the basic difference among them is in the nature of models applied. The oldest group of methods (now often called conventional or multivariate morphometrics) uses a diverse collection of measurements (lengths, distances, angles) and counts to characterize morphologies and uses multivariate data exploratory (e.g., principal component analysis – PCA, canonical variate analysis – CVA) and statistical methods (discriminant analysis – DA, multivariate analysis of variance – MANOVA) for their analysis. Such methods have in several cases been used with diatoms, both to clarify morphospecies boundaries (Theriot & Stoermer 1984b), and to investigate intraspecific morphological variation (Theriot 1987; Teubner 1995; Droop 1995; Hausmann & Lotter 2001).

Another group of morphometric methodologies, called geometric morphometrics, places more emphasis upon a more complete use of two- or three dimensional geometric information and on models permitting visual representations of morphologies (Rohlf & Marcus 1993). One group of geometric morphometric methods describes outline shape using Legendre polynomials or Fourier analysis, and has been used in quantitative studies of outline shape in pennate diatoms (Stoermer & Ladewski 1982; Stoermer *et al.* 1986; Mou & Stoermer 1992; Pappas *et al.* 2001; Rhode *et al.* 2001; du Buf and Bayer 2002). Other geometric

morphometric approaches model morphologies by recording the coordinates of landmarks (points assumed to be homologous) on specimens. Landmark-based morphometrics has been one of the most dynamically developing field of morphometrics (Bookstein 1991; Dryden and Mardia 1998); still, such methods have not yet been used for diatoms.

The largest endeavours concerned with diatom quantitative morphology have been two recent research projects, ADIAC (Automatic Diatom Identification and Classification, 1998-2001) aimed at exploring the possibilities of automatic morphology-based diatom identification and DIADIST (Diatom and Desmid Identification by Shape and Texture, 2001-2004), which strived for the more general aim of visual indexing of images (including drawings) of biological specimens. These projects have combined several methodologies for an effective quantitative characterisation of diatom morphologies; the use of their 'products' in research concerning species delimitation, are, however, yet to be explored.

1.3.1.3. Advantages / disadvantages of phenotypic markers

The major advantage of some phenotypic markers, especially morphology, is their general and easy availability. Some phenotypic markers are also fossilized easily; in the case of diatoms, frustule morphology is preserved well even on geological time scales and can be used to trace past evolutionary and ecological events (Stoermer and Smol 1999). Another example of fossilized phenotypic markers from diatoms are highly branched isoprenoid alkenes and possibly other chemical compounds (Damste *et al.* 2004). However, there are several difficulties associated with the (exclusive) use of phenotypic markers for delimiting species. The first important issue is their representativeness of the organism as a whole, as mentioned above. In the case of diatoms, morphological information as it is generally available is in the form of morphology of acid-cleaned frustules from natural samples. The assumption that morphologies considered characteristic of distinct species really are such, and not differences induced by changes in the environment for instance, has rarely been tested experimentally. The construction of the diatom frustules from two valves, generally formed at different times and thus often among different environmental conditions, gives a powerful possibility for diatomists to reveal phenotypic plasticity even in acid-cleaned natural samples (Mann 1999). The few experimental studies were also reviewed by Mann (1999); he concludes that although phenotypic plasticity does occur in diatoms, taxonomists in general can continue relying on studies of cleaned frustules. Although in great generality this seems to be true, phenotypic plasticity can lead to problems with species delimitation in some cases, and it is particularly problematic in the group dealt with in this thesis, as will be reviewed below.

Another disadvantage of relying exclusively on morphology for species delimitation is that this practice assumes that speciation is always accompanied by morphological divergence (Knowlton 2000). Although little is known about speciation mechanisms in microalgae, morphology is certainly not the only adaptively relevant feature of their cells. Speciation scenarios involving physiological and ecological, but not necessarily morphological divergence are quite plausible in these organisms. Therefore, microalgal systematics has to consider the possibility that species are often cryptic, i.e., indistinguishable on morphological grounds. The number of reports from a wide range of eukaryotic microbial groups lending support to this notion is steadily increasing (Scholin *et al.* 1995; De Vargas *et al.* 1999; Sáez *et al.* 2003; Sarno *et al.* 2005). It has recently been suggested that this might indeed represent a characteristic of planktonic microorganisms in general and that many morphospecies in such organisms are expected in fact to be 'super-species', complexes of morphologically highly similar biological species (De Vargas *et al.* 2004).

1.3.2. Mating compatibility

Mating compatibility might be considered a special kind of phenotypic marker; although in fact it is not a feature of a single organism but one attached to pairs of organisms. The BSC, which has become one of the most important of the multitude of species concepts proposed in the 20th century, places a strong emphasis on the role of reproductive isolation in keeping species distinct by not allowing them to share genetic information. On the other hand, the CSC (Templeton 1989) has emphasised another aspect of reproductive patterns: that exchange of genetic information through sexual reproduction is one of the most important factors responsible for the cohesion within a species. These two notions have had an important impact upon the way we understand species now, and they also had a strong practical impact upon studying species boundaries in diatoms.

Observations of reproductive behaviour in connection with microdiversity problems have already been made by Geitler (Geitler 1932; Geitler 1975), but became real working tools of diatom systematics through Mann's works (Mann 1984; Mann 1989). He studied occurrences of sexual events between individuals of a morphospecies in cultures and in semi-natural assemblages in a number of morphospecies, and reported observations that cells belonging to slightly different morphotypes of a morphospecies did not mate with each other, only with ones belonging to the same morphotype. The freshwater diatom *Sellaphora pupula* has become the most intensively studied model in such investigations (Mann 1989; Mann *et al.* 1999; Behnke *et al.* 2004). It turned out to be composed of several, morphologically slightly different, reproductively isolated groups, biological species. These species apparently even had different reproductive biologies: whereas some were allogamous, mating has never been observed in others; these possibly represent autogamous or completely asexual lineages (Mann 1999). Among the allogamous species, some are strictly dioecious, with male and female clonal lineages, whereas others are monoecious and capable of intraclonal mating. These studies revealed a previously unexpected amount of species diversity and diversity in reproductive biology in what was traditionally considered a single diatom morphospecies.

1.3.2.1. Advantages / disadvantages

Observations of mating compatibilities of sexual organisms is a conceptually special case of use of a phenotypic marker in the sense that this approach explicitly addresses mechanisms responsible for producing observed variation patterns. Use of this approach has contributed significantly to our recent understanding of species diversity of diatoms; still, it has some disadvantages that do not allow their universal application in diatom systematics. One of them is practical: a prerequisite for such observations is that mating between cells identified at an intra-morphospecific level can be observed. This means that they are not feasible with e.g. oogamous organisms, like centric diatoms, where small flagellated sperm cells are released from the male gametangia, nor in cases where the morphotypes concerned cannot be differentiated alive, only by e.g., striation pattern or ultrastructural features, visible only after frustule preparation. A more theoretical concern is that observations are made in the laboratory, and potential effects of environmental conditions upon reproductive behaviour cannot be ruled out. Further complicating factors are the existence of clearly differentiated species, which nevertheless hybridise (a common situation for example in land plants, but one that also occurs in animals or macroalgae, as noted above), and the problem of allopatric comparisons, where the possibility has to be considered that taxa that never or only rarely mate in nature might still be able to hybridise when put together in the laboratory. However, it should also be noted that the approach outlined above is in fact an approach combining two kinds of evidence: minor morphological differences and observations of reproductive compatibility. If morphologically distinct groups are found, between which no mating can be

observed, whereas it is observed within these groups several times in the same conditions, the evidence supporting the idea that the minor morphological differences between the groups are connected to their reproductive isolation is rather strong, making this approach a powerful one in the cases where it is feasible.

.

1.3.3. Genetic markers

The third group of microdiversity markers sample genetic variation directly. Reproductive patterns, demographic events and population subdivision have genome-wide effects upon patterns of genetic variation. Therefore, patterns of genetic variation can potentially reveal information about these factors in the microevolutionary history of organisms. PCR (polymerase chain reaction) technology allows for the selective amplification and sequencing of DNA fragments from genomic DNA extracts, providing the technical basis for a wide variety of methods for sampling genetic variation. Studies using such approaches for microdiversity explorations are also becoming common with eukaryotic microorganisms. The basic conceptual advantages of sampling genetic variation directly is that no concerns of plasticity or life cycle related changes affect interpretation, and that inferences can be made not only about patterns of variation, but also about microevolutionary processes playing a role in shaping them.

Most large taxonomic groups have their classical, most widely used molecular markers. In metazoans, different portions of the mitochondrial genome have become standard markers (Kessler & Avise 1985; Avise & Walker 1999; Wiens & Penkrot 2002); in land plants, parts of the plastid genome, particularly some hypervariable tRNA spacer regions (Taberlet *et al.* 1991). The common advantages of these genomic regions were their variability, providing information even on the intraspecific level, and their uniparental inheritance, which allowed data analyses not to deal with complications arising from the reticulate evolution of nuclear genes (Avise & Walker 1999). In the meanwhile, data analysis methods capable of dealing with reticulate evolution have been developed (Rosenberg & Nordborg 2002), and limitations of sampling only one small portion of the genomes have been recognized (Hey 1997; Anderson & Kohn 1998; Nordborg & Innan 2002; Rosenberg & Nordborg 2002; Zhang & Hewitt 2003), thus, there is a general trend towards applying multiple nuclear markers besides the classical organellar ones (Zhang & Hewitt 2003).

The most widely used molecular markers perhaps overall, but in any case in microorganisms, are different regions of the ribosomal operon (Coleman 2003; Alvarez & Wendel 2003). The main reasons for this are the following: rRNA molecules are among the few that universally occur, from prokaryotes to multicellular organisms, without a change in their function, making them appropriate for phylogenetic studies on any taxonomic level without problems concerning their homology, and it is not known to be involved in lateral gene transfer (Ludwig & Klenk 2001). The wide range in the amount of variability present in different regions provides the possibility to sample regions with the appropriate resolution from studies of deep phylogeny to microdiversity surveys. In the latter context, their main practical advantage is that PCR primer pairs for their amplification from any organism are available (Medlin *et al.* 1988; White *et al.* 1990; Scholin *et al.* 1994). In microdiversity studies, the most variable regions of the rDNA operon are used. Especially the ITS (internal transcribed spacer) regions are in widespread use (Van Hannen *et al.* 2000; Coyer *et al.* 2001; LaJeunesse 2001; Alvarez & Wendel 2003; Orsini *et al.* 2004), but in microalgae, the hypervariable D1 and D2 regions of the 28S rDNA are also often applied (John *et al.* 2003; Lundholm *et al.* 2003; Sarno *et al.* 2005).

Regions encoding nuclear ribosomal RNAs appear in numerous (up to thousands of) copies in eukaryotic genomes, and this fact has several implications for their use as molecular

markers. Contrary to the simplicity of inheritance mechanisms of organellar genes, not only are they affected by the reticulate evolution characteristic of all nuclear genes in sexual species, but they are also subject to concerted evolution (Schlötterer *et al.* 1994; Schlötterer & Tautz 1994). Concerted evolution is a characteristic of multigene families. Members of such families are subject to intragenomic homogenisation, a phenomenon appearing through unequal crossing over and gene conversion (Li 1997), keeping members of gene families similar within genomes and within populations. Although this has sometimes been advocated as a major advantage of rDNA as a molecular marker, several lines of evidence indicate that it might become problematic in some cases (Alvarez & Wendel 2003). The most important such problems for microdiversity studies are those associated with the presence of non-functional copies (pseudogenes) of rDNA genes in the genome and problems associated with hybridisation events.

In spite of concerted evolution, intragenomic variation in the most variable parts (like ITS regions) of the rDNA operon is common. Several reasons can lead to such intragenomic variation. In the simplest case, homogenisation is not fast enough to keep pace with mutation accumulating in variable regions. However, more problematic cases are also known to occur. Pseudogenes are non-functional copies of genes in the genome; rDNA pseudogenes are known to appear even more often than pseudogenes of other classes of coding sequences, because of the lability of the genomic location of rDNA arrays (Alvarez & Wendel 2003). Such non-functional gene copies are freed from the selective constraints acting upon their functional counterparts, leading to their divergence. As long as the flanking sequences targeted by amplification primers have not diverged sufficiently, however, they are amplified together with the functional copies and are being sampled. Another possibility resulting in the presence of divergent rDNA arrays in single genomes is interspecific hybridisation (Muir *et al.* 2001). Ribosomal DNA has an additive inheritance, making it a powerful tool to detect recent hybridisation events (Sang *et al.* 1995; Quijada *et al.* 1997). However, the evolutionary fate of divergent ribosomal DNA arrays after such events might be different. The different rDNA types might coexist in the genomes for some time, or one or the other might disappear. If different sets disappear from different, closely related lineages, then rDNA phylogenies will not reflect species phylogenies, as has been shown in the case of cotton (Wendel *et al.* 1995a; Wendel *et al.* 1995b). Whereas these problems might lead to false inferences in phylogenetic studies, they are not expected to impact studies of species boundaries so seriously. Such studies must, however, also consider the possibility that intragenomic homogenisation of rDNA arrays is not necessarily complete: if it were, species could be detected based simply on the presence of differences in rDNA sequences.

A further advantage of molecular methods is that they enable one to sample several independent sets of characters for species delimitation purposes. As noted above, combining independent sources of evidence is critical for this; sampling regions from different genomes (nuclear and organellar) is an increasingly used strategy to avoid relying on a single genetic marker in studies of species limits (Zhang & Hewitt 2003).

1.4. The problem of clonality

An important problem when dealing with species delimitation in microorganisms is the possibility of strict clonality or autogamy. Most microorganisms have life cycles basically consisting of clonal reproduction with relatively rare sexual events; diatoms are among the examples for this basic plan (Drebes 1977; Mann 1993; Chepurnov *et al.* 2004). However, variation in reproductive biology has been observed even among closely related species, like in the *Sellaphora pupula* species concept mentioned above. Furthermore, several protozoan

species, even some of those previously known as obligate sexual species, have turned out to show 'effective clonality' (Tibayrenc & Ayala 1987; Tibayrenc & Ayala 1991; Urdaneta *et al.* 2001; Bart *et al.* 2001; Tibayrenc & Ayala 2002; Telleria *et al.* 2004). This means that although these protists are capable of sexual reproduction, genetic exchange among clonal lineages is absent or so rare that it has no strong effects upon patterns of genetic variation. It is also generally assumed that some diatom species have abandoned sexuality (Drebes 1977); however, this conclusion is only based on the failure to observe mating in these species (Mann 1993; Chepurnov *et al.* 2004).

Clonality has, in part, similar effects upon patterns of genetic variation as population structure (reproductive isolation): it leads to a deficit of heterozygotes and to linkage (i.e., congruent variation patterns) among physically unlinked loci (Tibayrenc & Ayala 2002). Tibayrenc´s (1995) distinction of non-structured ('real' sexual) and structured species (complexes of cryptic biological species or of clonal lineages) is based on this observation. It is, however, also important to distinguish clonality from cryptic speciation when studying species boundaries, especially because species concepts making reference to reproductive compatibility (like the BSC) cannot be applied in cases where sex is absent. Furthermore, patterns of intraspecific (both genetic and phenotypic) variation expected from a complex of clonal lineages (agamic complex) are also different from those expected from a complex of cryptic sexual species. Most importantly, whereas in the case of clonal complexes a general correlation among unrelated features is expected at any level, this is the case only above the biological species level in the case of cryptic species: sex is expected to break down within-biological-species correlations in the latter case.

Studies of variation patterns within species complexes, where the possibility of clonality cannot be excluded, can make particularly effective use of genetic markers. If the possibility of panmixia (random mating) is rejected based on patterns of genetic variation (heterozygote deficiency, linkage), further tests can be designed to distinguish which of the possible explanations (clonality vs. cryptic species) applies in the particular case. These further tests can simply test the hypothesis of panmixia within the apparent genetic units revealed, using the same criteria as applied on the level of the whole 'species' or species complex. If significant linkage disequilibrium and heterozygote deficiency is found within these groups as well, clonality is the preferred explanation. However, if evidence for gene flow (sexuality) among lineages within these groups is found, cryptic speciation is the more plausible hypothesis. An important requirement of this approach is that the molecular markers used have sufficient resolution (i.e., amount of variation) even within the subgroups identified.

Several approaches to test the hypothesis of panmixia have been used in the literature (Tibayrenc & Ayala 2002). They can be classified as single-locus vs. multilocus tests. The former are based on the idea that deviations from panmixia result in a deficit of heterozygotes; they include quantitative tests of deviations from Hardy-Weinberg equilibrium or qualitative tests based on identifying lack of segregation genotypes. Multilocus tests also have several forms, including qualitative tests based on the lack of recombinant genotypes or congruent variation at independent loci as well as statistical tests of linkage disequilibrium or of correlation between independent sets of markers. Most classical quantitative tests are applicable with codominant markers only. However, methods are also available that can be used with dominant markers, where heterozygotes cannot be distinguished from dominant homozygotes, like in several PCR-based multilocus fingerprinting methods including RAPDs – random amplified polymorphic DNA – or AFLPs – amplified fragment length polymorphism. Mes (1998) proposed an approach based on phylogenetic incompatibility among loci to test the hypothesis of panmixia using such data.

1.5. *A problematic species:* Cyclotella meneghiniana *Kütz., and its relatives*

C. meneghiniana is one of the most commonly reported freshwater diatom species. It occurs worldwide in a wide range of habitat types from inland lakes and reservoirs (Squires & Rushforth 1986; Makarewicz 1987; Zalocar de Domitrovic 1992; Seeligmann & Tracanna 1994; Romo & Miracle 1994), wetlands (Rojo *et al.* 2000; Tapia *et al.* 2003), streams and rivers (Sabater & Sabater 1988; Moss & Balls 1989; Sabater 1990; Gosselain *et al.* 1994; Ha *et al.* 2002) to estuaries (De Seve 1993; Rijstenbil *et al.* 1993; Muylaert & Sabbe 1999; Shin & Cho 1999; Lapointe 2000) and marine environments with decreased salinities, like the Baltic Sea (Korhola & Blom 1996; Bak *et al.* 2001). It is one of the most commonly occurring diatom species in lakes in general (Alvarez Cobelas & Rojo 1994), and it was featured as the single most often reported diatom species in a recent literature survey (Finlay *et al.* 2002). It is generally considered a pollution tolerant species (Giorgio *et al.* 1991) and one that is characteristic of waters with elevated salinity and / or pH (Van Dam 1979), including alkaline lakes (Hecky & Kilham 1973).

Besides its commonality, its ease of culturing also made *C. meneghiniana* one of the most often used freshwater diatom species in laboratory experiments. It has been used in studies of life cycles (Iyengar & Subrahmanyan 1944; Schultz & Trainor 1968; Rao 1970; Rao 1971), physiology (Tilman & Kilham 1976b; Azariah *et al.* 1978; Kopczynska 1979; Rao *et al.* 1983; Millie 1984; Tuchman *et al.* 1984; Millie 1986; Lewandowska & Kosakowska 2004), experimental ecology (Tilman & Kilham 1976a; Azariah *et al.* 1978; Kopczynska 1979; Rao *et al.* 1983; Millie 1984; Tuchman *et al.* 1984; Millie 1986; Lewandowska & Kosakowska 2004), cell biology (Hoops & Floyd 1979), toxicology (Millie & Mark Hersh 1987; Millie & Mark Hersh 1987; Sicko-Goad *et al.* 1989a; Sicko-Goad *et al.* 1989b; Sicko-Goad *et al.* 1989c; Sicko-Goad *et al.* 1989d; Lazinsky & Sicko-Goad 1990a; Lazinsky & Sicko-Goad 1990b), biochemistry (Millie 1984; Millie 1986; Lohr & Wilhelm 2001), molecular diversity (Bourne 1992; Zechman *et al.* 1994), and molecular biology (Büchel 2003) as often as few other freshwater diatom species. In spite of (or in part, because of) this extensive study, it is still a taxonomically problematic species.

First, *C. meneghiniana* is an example of diatoms with large morphological variation which lacks obvious discontinuities. Several publications have discussed this phenomenon; the most detailed example is Schoeman and Archibald (1980), who illustrated the morphological variability of this species in 156 photomicrographs in their Diatom Flora. In spite of intensive discussion, it is still unclear whether this morphological variability is explained by the presence of multiple species that are morphologically difficult to differentiate within *C. meneghiniana*, by plastic responses to environmental conditions, or do they represent simply neutral morphological variability (Hakansson 1998). Furthermore, the separation from *C. meneghiniana* of a number of validly described species have been questioned based on the notion that the large morphological variability found in natural samples in fact spans supposed species boundaries. The most relevant examples for this thesis are *C. scaldensis* (Muylaert & Sabbe 1996; Kiss & Ector 2000) and the particularly complicated case of *C. cryptica*, described below in detail.

Although, as discussed above, the phenomenon of morphological plasticity is not believed to cause major problems for diatom species delimitation in general, it is known to cause problems with *C. meneghiniana*. One of the well-studied examples of plasticity of diatom frustule morphology comes from *C. cryptica*, a species morphologically resembling *C. meneghiniana* closely. *C. cryptica* is one of the rare examples of diatom species described from a culture (Reimann *et al.* 1963). Its main frustule morphological difference from *C. meneghiniana* was “the presence of an ill-defined border between the central area and the radially striated margin” (Reimann *et al.* 1963). However, in later life cycle studies of

different clones of *C. meneghiniana* and *C. cryptica*, it was found that some clones were capable of producing both frustule morphologies, typical of *C. meneghiniana* or of *C. cryptica*, depending on the salinity of the medium (Schultz & Trainor 1970; Schultz 1971). In particular, strains identified as *C. cryptica* produced mostly the *cryptica*-pattern when grown at higher salinities (full-strength seawater, 28.7 ‰ to about 4.3 ‰). However, they produced a majority of valves showing the pattern characteristic of *C. meneghiniana* when grown in media with low salinities, and even at higher salinities for some time after auxospore production. The proportion of valves showing the two patterns depended on the salinity of the medium and time since auxospore formation, and cultures grown at intermediate salinities contained mixtures of the two valve types (Schultz 1971). This phenomenon, and later reports of observations of *C. meneghiniana* cultures capable of producing valves showing the *C. cryptica* pattern (Desikachary & Rao 1973; Hoops & Floyd 1979) lead some authors to the suggestion that these valve patterns might represent extremes of morphological variation of a single species (Desikachary & Rao 1973; Schoeman & Archibald 1980).

A further study bringing a new aspect into our knowledge of the diversity within *C. meneghiniana* was Boerne's plastid genome RFLP (restriction fragment length polymorphism) study (Bourne *et al.* 1992; Bourne 1992). After preparing a physical map of the *C. meneghiniana* plastid genome (Bourne *et al.* 1992), she surveyed restriction fragment length variation in a collection of strains of *C. meneghiniana* and some closely related species (*C. gamma*, *C. cryptica*). Whereas no variation was detected by this approach in the other species, strains of *C. meneghiniana* showed variation, and multiple, well-supported groups were recovered within this morphospecies. Furthermore, phylogenetic analyses placed the *C. cryptica* strains well within the clade containing all *C. meneghiniana* strains, leaving the latter paraphyletic. Bourne suggested alternative interpretations of these results. One was that *C. meneghiniana* should be considered a complex of morphologically highly similar, but genetically distinguishable species. Another possibility was that chloroplast DNA variation was generally large in diatom species. Concerning the *C. cryptica* problem, she wrote that the results supported the idea that *C. cryptica* was a variant of *C. meneghiniana* rather than a separate species; we note that this would be the conclusion only in combination with the latter explanation (large within-species cpDNA variation).

There are, however, even further possible interpretations of these data. Two major factors influencing the amount of variability in a genomic region are mutation rate and effective population size (Hartl and Clark 1997; Nordborg & Innan 2002). Possibly, the difference in the amount of genetic variability between the species studied by Bourne (1992) was caused by their different effective population sizes – it seems indeed plausible that *C. meneghiniana* has larger effective population size than most other *Cyclotella* species, based on its often dominant worldwide occurrence discussed above. The fact that *C. meneghiniana* was found to be paraphyletic does not necessarily imply either that it is a species complex or that *C. cryptica* is not separate species but a variant of *C. meneghiniana*: newly formed species might be even polyphyletic and genetically indistinguishable at most loci for some time after speciation (Hudson & Coyne 2002). Before genetic variation reaches reciprocal monophyly, one would first expect a stage where one of the species is monophyletic and the other is paraphyletic – again, the species with the smaller effective population size is expected to reach monophyly before the other species (Funk & Omland 2003). So an alternative explanation of the RFLP variation patterns observed by Bourne (1992) might be that insufficient time since the speciation event separating *C. meneghiniana* and *C. cryptica* has elapsed to reach the reciprocal monophyly of their plastid genomes. Indeed, the fact that all *C. cryptica* strains share a common plastid RFLP genotype, different from those found in any *C. meneghiniana* strain, can be interpreted as a strong indication of the lack of gene flow between the two groups, thus would support the idea that *C. cryptica* is a separate species. In

any case, Boerne's (1992) studies further contributed to the idea that *C. meneghiniana* was problematic, but was not able to solve the problems unequivocally.

A further uncertainty concerning *C. meneghiniana* should be noted in the context of its microdiversity research. Although, as noted above, the life cycle of this species has been the subject of numerous investigations, a question of basic importance has not yet been answered by these studies: whether *C. meneghiniana* is capable of allogamous sexual reproduction. Gametogenesis and auxospore production is easily induced in cultures of this species in an inducible size range by elevating the salinity of their medium. Production of egg cells and of flagellated sperm can be observed after two to three days, followed by swelling of the eggs and auxospore production (Iyengar & Subrahmanyan 1944; Schultz & Trainor 1968; Rao 1971; Rao 1996; Hakansson & Chepurnov 1999). However, in spite of the numerous observations made on this experimental system and although self-fertilization has been observed (Iyengar & Subrahmanyan 1944), fertilization of the egg cells by sperm has never been reported (Schultz & Trainor 1968; Drebes 1977). Although (Hakansson & Chepurnov 1999) claim that (Rao 1970) observed oogamy, the latter publication in fact also explicitly states that "Fertilization has not been observed". As a consequence, *C. meneghiniana* was mentioned as an example of diatom species that abandoned allogamous reproduction in reviews of diatom sexuality (Drebes 1977; Mann 1993). As noted above, the consequences of strict clonality or autogamy upon genetic and thus also phenotypic variation are fundamental: in the case of clonality or autogamy, lineages evolve independently from each other, without any exchange of genetic information. Therefore, correlation among independent traits is generally expected in such cases, whereas within a sexual species, correlation between phenotypic traits is only expected as a result of close linkage of genes responsible for them or if selection acts on a combination of traits. Importantly for studies of species limits, patterns of genetic variation in a strictly clonally or autogamously reproducing group of organisms is strictly hierarchical, whereas it is reticulate in a sexual species. Thus, the "boundary between hierarchical and non-hierarchical variation that underlies the concept of species" (Mann 1999) (or the conceptually closely related biological species concept) is not applicable in the case of clonal lineages.

1.6. Aim of the thesis

The topic of this thesis is investigations of pheno- and genotypic variation at, and around, the species limit in one of the classical examples of a 'problematic diatom species', *Cyclotella meneghiniana*. The aim was to contribute to resolving uncertainties concerning its microdiversity, microevolution and systematics using morphometric and molecular methods. Morphometric methods for analysing patterns of valve ultrastructural variation were established. Furthermore, sequence variation in different rDNA regions in a single population was explored to serve as a reference for interpreting variation patterns from a larger, more diverse collection of strains. Genetic variation in a global collection of strains of *C. meneghiniana* and some closely related species, including the problematic *C. cryptica*, was surveyed in one of these regions and at an independent, plastid encoded locus. Finally, a multilocus genotyping method, AFLP (amplified fragment length polymorphism) was used to test the hypothesis of strict clonality in two *C. meneghiniana* populations.

1.7. Outline of the thesis

1.7.1. Morphometrics

The first step in this study of variation within *C. meneghiniana* was the sampling of a single population of this species. This was important to avoid difficulties associated with the interpretation of data obtained from strains with diverse geographic origins, as was the case in Bourne's (1992) work. Possible explanations of (either genetic or morphological) differences found between strains can include limited gene flow or differential local adaptation in such cases. If amounts of within-population variation are unknown, no conclusions can be drawn about the significance of differences between individuals (strains) originating from different geographical sources: would the same amount of variation be found in any of the localities sampled if a larger number of strains had been isolated from that single locality?

For this purpose, twenty strains were isolated from a single locality, the lowest, estuarine stretch of the River Geeste (Bremerhaven, North Germany), from what was believed to be a single *C. meneghiniana* population. The strains were first analysed morphologically using scanning electron microscopy; substantial morphological variation among the twenty cultures was found. Five of the cultures showed a valve morphology that was interpreted as intermediate between *C. meneghiniana* and *C. scaldensis* (referred to here and in **Publications I** and **II** as the 'ambiguous' *C. scaldensis* morph). Patterns of valve ultrastructural variation were analysed using quantitative methods in **Publication I**. As noted above, the validity of the separation of *C. scaldensis* from *C. meneghiniana* has been questioned based on the large morphological variability and on the observations of intermediate forms between these species (Kiss & Ector 2000 and Muylaert pers. comm.). The morphometric comparisons were initiated to clarify first of all whether the five cultures with the 'ambiguous' morph were morphologically distinguishable from the rest of the cultures, which showed more typical *C. meneghiniana* morphologies. Because the twenty strains clearly only represented a small subsample of the populations sampled, material from the original, uncultivated samples was also included in these comparisons. This allowed us to (1) explore how representative of the populations they were isolated from the cultures were morphologically, and (2) make sure that morphological differences found between groups of cultures were neither culture induced artefacts nor artefacts resulting from poor sampling.

Based on the intermediate valve morphology of the 'ambiguous' morph between *C. meneghiniana* and *C. scaldensis*, it was also relevant to compare it morphologically to the latter morphospecies. For this purpose, in the lack of type material for SEM investigations, material provided by K. Muylaert from the type locality of *C. scaldensis* was used. SEM observations of this material revealed the presence of valves closely resembling our 'ambiguous' cultures, as well as ones with valve morphologies more typical of *C. scaldensis* (i.e., they were more different from *C. meneghiniana*; we called them 'extreme' *C. scaldensis* morph in **Publication I**) and typical *C. meneghiniana*. Our morphometric approach was thus also used to test the morphological distinctness of these three morphs in sympatry. Again, the comparisons also allowed exploration of the morphological relationships between representatives of the different morphs with different origins (in the case of *C. meneghiniana* and the 'ambiguous' *C. scaldensis* morph: cultures, Geeste population, and Schelde population).

The major mentionable methodological aspects of **Publication I** were its use of morphometric methods for quantitative comparisons of valve ultrastructure using scanning electron microscopy (SEM) instead of light microscopy (LM) and the first use of landmark-based geometric morphometric methods in diatom morphometrics. The first (SEM) was necessary due to the subject. Morphometric studies of pennate diatoms generally only use

light microscopy. However, the importance of valve ultrastructural characteristics that cannot be resolved by LM in the taxonomy of centric diatoms (Theriot & Stoermer 1984a; Theriot *et al.* 1988) and particularly in the group studied here made the use of electron microscopy necessary for our study.

Concerning landmark-based morphometric methods, we tested their applicability by comparing analyses using conventional morphometric data collected from SEM images with those using landmark coordinates as the primary data. Although, as noted above, methods based on landmark coordinates were among the most dynamically developing areas in biological morphometrics in the last decades, methods using anatomic landmark points have not before been used in diatoms. Our aim was to explore whether and how frustule ultrastructural features of the diatoms studied could be modelled using landmark coordinates, and to compare the effectivity of landmark based and conventional morphometric methods in answering the above questions.

1.7.2. Ribosomal DNA variation at a single locality

The second part of the study involved exploration of sequence variation in different regions of the rDNA operon in the twenty strains isolated from the Geeste estuary (**Publication II**). This part of the study was aimed at answering three questions. First, it was shown in **Publication I** that the 'ambiguous' morph was morphologically distinct from typical *C. meneghiniana* in the same conditions in cultures as well as in sympatry in the field samples of different origins included in the comparisons. This was interpreted as a strong indication that the organisms producing these valve morphologies were not conspecific. Ribosomal DNA comparisons allowed us to test this idea using an independent data set; thus, the first question was whether the morphologically distinct groups of cultures were also distinct in their rDNA sequences.

The second aim of these sequence comparisons was to identify which region of the rDNA operon was the most appropriate for studying intraspecific variation in *C. meneghiniana*. This question has two important aspects. On the one hand, a molecular marker has to show variation in the group of interest in order to provide any useful information. Variability in different regions of the rDNA operon is known to vary in a wide range. Some positions are so conserved that they allow the alignment of sequences from all organisms, whereas other regions often even show substantial intraspecific variation. Because the variability of different regions is determined by the joint effect of several factors (the two most important of them being functional constraints and effective population size), the region showing the necessary amount of variation needs to be chosen experimentally. Conversely may a more variable region also vary not only within species, but even within individuals (clonal strains). To some extent, this can be dealt with using computational methods as long as within-strain variation is constrained to nucleotide substitutions (Clark 1990; Zhang & Hewitt 2003). However, the experimental effort required to obtain DNA sequences substantially increases if variation also includes length variation (insertion-deletion polymorphisms): PCR products cannot be sequenced directly in this case, but they need to be cloned and multiple clones sequenced. Thus, a compromise between information content and practicality would be to choose a region for large-scale comparisons that shows sufficient variation, but no within-strain length variation in the organisms of interest.

The third question to be answered by this study was that if rDNA regions showing variation among the *C. meneghiniana* strains are found, is their variation pattern consistent with expectations from a single panmictic population or does it show indications of genetic structure? As discussed above, at least two possible hypotheses might be considered as alternatives of panmixia: the presence of several (cryptic) species in the 'population' studied,

or the complete (or close to complete) lack of allogamous sexual reproduction (strict clonality or autogamy). Patterns of genetic variation in the rDNA operon deviate from panmictic expectations in similar ways in both cases. With panmixia, random combinations of the rDNA variants present in the population are combined in individuals (clonal strains). Therefore, the chances of finding combinations of rDNA variants within single individuals are independent of the amount of sequence difference between the variants. Concerted evolution is furthermore expected to homogenize rDNA variation within single strains, thus, due to sexual transfer of rDNA variants between lineages, ultimately within the whole population. This leads to the expectation of low within-strain and a comparably low within-population rDNA variation in a panmictic population. In the case of cryptic sexual species, mixing of rDNA variants and their homogenisation is constrained to subsets of individuals. Because of the lack of gene flow among these groups, mutations accumulating within them lead to their divergence from each other. Therefore, in this case, small within-strain rDNA variation and larger within-'population' variation with multiple, well-distinguished groups (i.e., small within-group variation, larger among-group differences) are expected. Third, the diverging lineages are represented by single individuals in the case of strict clonality or autogamy (which cannot be differentiated from each other based on patterns of genetic variation). Mutations occurring in any individual are only inherited by its clonal offspring, leading to a hierarchic pattern of variation not characteristic of sexually reproducing organisms. The expectation is small within-strain and larger within-'population' variation, but without the presence of distinct groups. The amount of rDNA variation is expected to be larger in any randomly chosen group of individuals than in the single individuals, provided there is variation at all.

The data obtained strongly rejected the null hypothesis of panmixia, and indicated that the 'population' studied probably included multiple, reproductively isolated sexual species (although the possibility of strict clonality could not be excluded with great confidence either). A final question that **Publication II** attempted to answer was whether the groups of strains belonging to different rDNA groups were morphologically distinguishable by the conventional morphometric characters recorded for **Publication I** or were they truly 'cryptic'.

1.7.3. Genetic variation within *C. meneghiniana* on a global scale

It was shown in **Publication II** that patterns of rDNA sequence variation in a single 'population' of *C. meneghiniana* suggested the presence of genetic structure in this morphospecies. **Publication III** extended this study in a number of ways. First, a global collection of *C. meneghiniana* strains was included in the comparisons. Second, the taxonomic sampling was further extended by including strains of the closely related species *C. quillensis*, *C. choctawatcheeana*, the problematic *C. cryptica*; and *Discostella* (former stelligeroid *Cyclotella*, Houk & Klee 2004) and *Stephanosdicus* strains to serve as outgroups. Third, an independent molecular marker, a coding region of the chloroplast genome was also sequenced besides a part of the nuclear ribosomal DNA operon. In order to find an appropriate plastid marker, amounts of variation at a number of loci, for the amplification of which PCR primers have been published, was screened, and the region showing the largest amount of variation was chosen. The molecular markers used to screen variation in **Publication III** were the D1/D2 regions of the 28S rDNA (identified in **Publication II** as the best compromise between variability and practicability of the nuclear rDNA regions tested) and a section of the plastid encoded psaA gene (identified in these experiments as the most variable of the plastid genomic regions tested).

The use of multiple, physically unlinked markers allows testing of a further aspect of genetic variation. As noted above, congruence of patterns of variation in independent traits is

generally considered crucial for studies aimed at the clarification of species limits (Mann 1999). The characters provided by single genomic loci are not truly independent characters because of their strong physical linkage. However, use of different genomes (nuclear and organellar) represents a powerful approach to sampling genetic variation at loci the evolution of which was as independent from each other as possible. Therefore, congruence between variation patterns at a nuclear rDNA region and a plastid region is a much stronger indication of underlying phenomena that affect the whole organism than congruence of variation pattern at closely linked loci is (like different parts of the rDNA operon).

The question whether the observed deviations from panmictic expectations were caused by the presence of several cryptic species or by strict clonality could not be answered with confidence in **Publication II** using rDNA sequence data. The reason for this was the relatively small amount of variation within the single rDNA groups found. The combined use of the two loci in **Publication III** could not help to answer this question either, because of the same reason. Thus, in **Publication III** another approach was chosen to test further which of these possibilities applies to *C. meneghiniana*. The approach chosen involved multilocus genotyping of a subsample of strains isolated from two populations using amplified fragment length polymorphism (AFLP) and analysis of the phylogenetic incongruence in the AFLP fingerprints. This approach was chosen for two reasons. The first was that AFLP markers are dominant, i.e., heterozygotes cannot be differentiated from dominant homozygotes using this technique. Classical tests of linkage disequilibrium are not applicable in such cases. The other reason was the comparatively small sample sizes reachable using our sampling. Testing the hypothesis of clonality requires detailed sampling of single populations. However, the majority of strains used in **Publication III** represented different populations, in the sense that they either had different geographic origins or they belonged to different rDNA groups. Only two populations, sampled on a single day at two localities on the River Weser (North Germany), which exclusively included strains belonging to a single genotypic group, provided an appropriate sample for this analysis.

2. Publications

2.1. List of publications

This thesis is based on the following publications:

I. B. BESZTERI, É. ÁCS, L.K. MEDLIN (2005)

Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyceae)

European Journal of Phycology 40: 89-103.

II. B. BESZTERI, É. ÁCS, L.K. MEDLIN

Ribosomal DNA sequence variation within and among clonal strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) from an estuarine locality

Submitted to Protist.

III. B. BESZTERI, U. JOHN, L.K. MEDLIN

Congruent variation at a nuclear and a plastid locus suggests that the diatom *Cyclotella meneghiniana* is a species complex

To be submitted to Molecular Ecology.

Other publications prepared or in preparation with contributions of the candidate from the period of this work:

IV. MAKK, J., BESZTERI, B., ÁCS, É, MÁRIALIGETI, K. & SZABÓ, K. (2003).

Investigation on diatom-associated bacterial communities colonizing an artificial substratum in the River Danube.
Large Rivers 14., Archiv für Hydrobiologie Suppl. 147/3-4: 249-265.

V. JOHN, U., GROBEN, R., BESZTERI, B., MEDLIN, L.(2004).

Utility of Amplified Fragment Length Polymorphisms (AFLP) to analyse genetic structures within the *Alexandrium tamarense* species complex.
Protist 155: 169-179.

VI. M. HOPPENRATH, B. BESZTERI, G. DREBES, H. HALLIGER, S. JANISCH, J. VAN BEUSEKOM & K.H. WILTSHIRE.

Thalassiosira species (Bacillariophyceae, Thalassiosirales) in the waters around the islands of Helgoland and Sylt, German Bight, North Sea – how to monitor the diversity, a first approach.
In preparation.

VII. RIMET, F., BESZTERI, B., BERZANO, M., MASQUELIER, S., NICOT, N., ECTOR, L. & MEDLIN, L.K.

Phylogeny of the family Bacillariaceae (Bacillariophyceae) based on 18S rDNA sequences.
In preparation.

2.2. Statement of the candidate´s contribution to the publications included in this dissertation

Publication I

Sampling and methodological concept was designed, material and data were collected and the manuscript was written by the candidate. É. Ács provided help concerning taxonomic background, É. Ács and L.K. Medlin provided help in writing the manuscript.

Publication II

The work was designed and carried out by the candidate in discussion with the co-authors. The candidate, with help of the co-authors, wrote the publication.

Publication III

The concept was developed and the work carried out by the candidate. U. John provided help in sampling the Weser populations and in the AFLP analyses. The manuscript was written by the candidate in discussion with the co-authors.

2.3. Publication I: Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related Cyclotella species (Bacillariophyceae)

BÁNK BESZTERI[1], ÉVA ÁCS[2], LINDA MEDLIN[1]

[1]*Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany*

[2]*Danube Research Station of the Hungarian Academy of Science, Jávorka S. u. 14, 2131 Göd, Hungary*

2.3.1. Abstract

Conventional and landmark-based geometric morphometric approaches were used to clarify the taxonomic identity of a centric diatom morph. This morph, found in cultures isolated from the Geeste estuary (Northern Germany), showed an intermediate valve morphology between that of typical specimens of *Cyclotella meneghiniana* Kützing and of *C. scaldensis* Muylaert & Sabbe. Its internal valve ultrastructure was compared to that of cells from 1) cultures of *C. meneghiniana* isolated from the same field samples and grown in the same conditions, 2) field material from the samples from which the cultures were isolated, and 3) a sample from the type locality of *C. scaldensis*. The morphometric analyses were used to determine 1) whether morphological variation of these morphs was continuous, or whether there were distinct morphological groups, and 2) how effective the alternative morphometric approaches were in answering this question. Both approaches proved informative and their results complemented each other, supporting the conclusion that three distinct size-reduction series were present in the samples investigated. Since the different morphs occurred sympatrically, we suggest that they probably belong to three reproductively isolated species.

Key words: *Cyclotella*, landmarks, morphometrics, morphospecies

2.3.2. Introduction

Cyclotella meneghiniana Kützing is one of the classically problematic diatom species. To investigate the variation in what we believed to be a single population of this species, we established a collection of clonal cultures from an estuarine locality in Northern Germany. Scanning electron microscopy revealed that some of these strains differed morphologically from typical *C. meneghiniana*. Morphometric analyses were initiated to clarify the taxonomic identity of these cultures.

The valve morphology of these cultures resembled that of *C. scaldensis* Muylaert & Sabbe (1996), but did not differ as markedly from *C. meneghiniana* as the former, based on the above authors' illustrations. We therefore refer to this morph as "ambiguous". SEM

observations of a sample from the River Schelde (the type locality of *C. scaldensis*) showed that it contained valves similar to our "ambiguous" morph, as well as others that we considered characteristic of *C. scaldensis*. However, these observations did not unambiguously reveal whether there was a continuum of morphological variation from typical *C. meneghiniana* through the "ambiguous" morphology to typical ("extreme") *C. scaldensis*, or if these were distinct morphological groups (morphospecies). We therefore attempted to characterize valve morphology quantitatively to clarify this question. Based on our preliminary observations, we assigned specimens to three morphs (*C. meneghiniana*, the "ambiguous" morph, or more characteristic *C. scaldensis*), and also grouped them according to their origin, viz., cultures, original samples from the River Geeste, and from the River Schelde sample. Original field samples were included to determine whether any aspects of our results were culture-induced.

We used both conventional and geometric morphometric approaches to characterize inner valve face morphologies. Conventional morphometric techniques have proven effective for clarifying species limits, especially in the centric genus *Stephanodiscus* (Theriot & Stoermer, 1984 a, b), as well as other diatom genera, and to investigate intraspecific morphological variation (Theriot, 1987; Hausmann & Lotter, 2001; Droop, 1995; Teubner, 1995). Whereas the data used for such analyses can include a diverse collection of measurements, counts and angles, geometric morphometrics place greater emphasis on a more complete use of two- or three-dimensional geometric information (Rohlf & Marcus, 1993). One such group of methods describes outline shape using Legendre polynomials or Fourier analysis, and has been used in quantitative studies of outline shape in pennate diatoms (Pappas *et al.*, 2001; Rhode *et al.*, 2001, du Buf & Bayer, 2002). We used another methodology called thin plate splines (Bookstein, 1991; Dryden & Mardia, 1998). Here, the shapes studied are modelled as geometric configurations of landmarks (reliably identifiable points that are assumed to be homologous in the range of specimens investigated), and their differences as "smooth" deformations. Shape variation within a sample of landmark configurations is described in terms of parameters that describe these deformations and can be analysed with classical multivariate analytical methods (Dryden & Mardia, 1998, for methodological details). To our knowledge, thin-plate splines and landmark-based morphometric analyses of valve ultrastructure have not previously been used for diatoms.

2.3.3. Materials and Methods

2.3.3.1. Samples

Two net plankton and two benthic samples were taken from the lowest, tidal stretch of the River Geeste in Bremerhaven (Northern Germany) on 9 July 2001. Plankton samples were collected from shore using a 20 µm mesh plankton net. Benthic samples were collected by scraping material from the surface of stones and wooden structures close to shore. A sediment sample from the Schelde estuary (Belgium) from December 2002 was provided by K. Muylaert.

2.3.3.2. Cultures

Cultures were initiated from samples taken from the River Geeste using a modified DY-IV medium (Andersen *et al.*, 1997) prepared with filter sterilized water from the sampling

sites. After one to four days, individual cells were isolated from these initial cultures with a micropipette and repeatedly washed in fresh medium. Cultures were maintained at 15 ^{o}C in a growth chamber with a 14/10 light/dark cycle at a photon flux density of 30-40 µmol $m^{-2}s^{-1}$. Twenty cultures were used in the analyses, fifteen of which were identified as *C. meneghiniana* and five as the other, "ambiguous" morph. Cultures or live material of the third morph (*C. scaldensis*) were not available for our study.

2.3.3.3. Electron microscopy

Diatom valves from field samples and cultures were cleaned using 30% H_2O_2 and five drops of HCl, rinsed in distilled water and dried on coverslips, sputter coated in an Emscope SC500 Modular Sputter Coater and examined with a PHILIPS XL30 ESEM operated at 10 kV accelerating voltage. For all morphometric analyses, scanning electron micrographs of inner valve views were used, as these provided the greatest number of possibly useful characters, including alveolar morphology, the frequency of marginal fultoportulae (FPs) and morphology of the rimoportulae (RPs). All SEM photographs and data sheets are available from B. Beszteri upon request.

2.3.3.4. Conventional morphometrics

The Microscope Control software of the SEM was used for measurements. The image of each specimen measured was saved as a tiff file. For all measurements and photographs, each cell was rotated so that the RP was located at the bottom of the image to standardize diameter measurements. Data from 312 specimens were used for the analyses, including cultures as well as field samples. Valves were assigned to one of the three morphs before morphometric analysis.

Variables used are shown in Fig. 1. Diameter measurements (D1, D2, D3) were taken in horizontal and vertical directions on the images and averaged to reduce measurement error and error due to the valves not lying absolutely perpendicular to the electron beam. Ratios of valve diameters (D1) measured in the two directions were calculated and specimens whose ratio was smaller than 0.95 (an arbitrarily chosen limit to exclude tilted specimens) were excluded from further analyses.

Principal component analysis (PCA) was calculated from the correlation matrix of the variables. PCA scores were calculated by postmultiplying the data matrix (with rows corresponding to cases and columns to variables) by the matrix of the eigenvectors of the correlation matrix without any further scaling. Canonical variate analysis was used to ordinate specimens to maximize separation of morphs with respect to their within-group variances. The size ranges of the morphs were similar (see below) and they overlapped, so no correction for this effect was used (dos Reis *et al.*, 1990). The analyses were carried out with MATLAB 6 (MathWorks Inc., Natick, MA, USA), using several MATLAB functions of R. Strauss (available at http://www.biol.ttu.edu/Strauss/Matlab/matlab.htm).

Discriminant analyses were performed using SYSTAT 10 (Version 10, SPSS Inc.) in order to test the distinctness of the groups. A jack-knifed discriminant analysis was also performed with SYSTAT 10, in which one of the specimens in every run was excluded from the calculation of the Mahalanobis distances from group centroids and was subsequently classified using the values obtained from the rest of the sample. This might be considered a more conservative indication of group distinctness.

Diagrams of valve morphologies were constructed in the following way to illustrate the results: three concentric circles with diameters D1, D2 and D3, were drawn to represent the valve outline, the inner rim of the valve mantle and the central area, respectively. The area between the second and third of these circles was divided into NC "triangles", representing the alveoli. The RPs were represented by equilateral triangles with a height of RL and a base length of RW. The value of MFP was used to calculate the reciprocal of the frequency of marginal FPs, (NC-1)/MFP. This value was rounded to the nearest integer, M, and a star, representing a FP, was drawn at every M^{th} costa.

The diagrams were used to reconstruct morphologies corresponding to points in the PCA scatter plot. Coordinates of the points on all except the two first principal component axes were fixed to zero, and points were projected back into the space of the original measurements. For this, row vectors of PCA scores were post-multiplied by the transpose of the matrix of eigenvectors, the entries of the resulting vector multiplied by the standard deviations of the corresponding column of the original matrix and added to the corresponding column means. The scores thus calculated were used to draw valve diagrams as described above.

2.3.3.5. Geometric morphometrics

The images used for the geometric morphometric comparisons were a subset of those used for the above conventional analyses (193 specimens). Landmark coordinates were digitized on the images and scale factors were recorded for each by measuring the length of the scale bar in the picture in pixel units using the software tpsDig (http://life.bio.sunysb.edu/morph/). The approach chosen for this study models "shape" as point (landmark) configurations and obtains a multivariate dataset describing shape variation using smooth deformations (Bookstein, 1991; Dryden & Mardia, 1998). This dataset can then be analysed with conventional multivariate data analysis methods. For digitizing landmarks and performing much of the following calculations, the tps series of programs (by F. J. Rohlf, http://life.bio.sunysb.edu/morph/) was used.

The landmarks recorded are listed in Table 1 and their position on a valve is shown in Fig. 2. Landmark 1, the centre of the valve, was found by digitizing the valve outline and calculating the coordinates of its centroid using the tpsDig program. To be able to use landmark 7, specimens without a FP at the costa adjacent to the RP had to be excluded from the analyses. However, it seemed to be the rule in all three morphs that there was a FP at both costae neighbouring the RP, and only eight specimens were excluded for this reason.

Landmarks were chosen around the RP as a point of correspondence between valves, since only one RP is present in valves of the group studied. In order to characterize alveolar morphology (landmarks 8-10), the alveoli adjacent to the RP might have appeared the most appropriate following this line of thought, but in many valves with large, bent RPs, the inner part of those (where landmark 10 would be) was obscured by the head of the RP. We therefore chose to use the second alveolus.

A multivariate dataset describing shape variability was obtained by the use of partial warp scores (Dryden & Mardia, 1998; Rohlf, 1999). With this technique, the thin-plate spline, an interpolating function minimizing bending energy, was fitted to the landmark configurations of the sample, using the calculated average landmark configuration as reference. Shape variation was described in terms of the parameters of the fitted functions. If shape variation in a sample is sufficiently small, the data thus obtained are a good approximation of the exact non-linear shape space which uses Procrustes distance as a metric (Dryden & Mardia, 1998; Rohlf, 1999). To verify that this was the case here, the slope of the regression line of tangent space distances against Procrustes distances and their uncentered

correlation coefficient were calculated using tpsSmall (http://life.bio.sunysb.edu/morph/). The slope of the regression line was 0.99, the correlation coefficient 0.99996, and inspection of the scatter of points revealed no large deviations of single points. These were accepted as indications of a good fit.

The use of partial warp scores allows for differential weighting of localized or global shape differences in a principal component (also called relative warp) analysis (Bookstein, 1991; Rohlf, 1993). With positive values of the weighting parameter α, partial warps with smaller bending energy, describing more global shape differences are given more weight, whereas with negative α more localized differences can be emphasized. Because we expected that the most interesting variation in our sample would be found in small-scale features, such as the shape of the RP or alveoli, we used $\alpha = -1$ for the relative warp analysis.

To analyse shape differences further, the partial warp scores were used as an ordinary multivariate dataset, in principle as described above for the conventional dataset. Centroid size and valve diameter were used as size measures to describe size-dependent morphological variation. Centroid size is a size measure generally used in geometric morphometric studies - it is calculated as the square root of the sum of squared distances of a specimen´s landmarks from their centroid (Dryden & Mardia, 1998). Valve diameter, on the other hand, is the most readily available measure of cell size in these diatoms. To compare allometry, shape change with size within the three groups, the partial warp scores were regressed on both size measures (centroid size and valve diameter) as well as on their logarithms using tpsRegr (http://life.bio.sunysb.edu/morph/). A multivariate analysis of covariance was performed to test the null-hypothesis of equal slopes of the regression lines in the three morphs. Of the different size measures used, the logarithm of valve diameter explained the most variance of the partial warp scores (data not shown), and was therefore used for the illustrations. Diagrams depicting shape changes from the overall mean landmark configuration to those predicted by the regression for different sizes using the thin-plate spline interpolation (Bookstein, 1991; Dryden & Mardia, 1998) were used to summarize differences in morphological changes with size reduction.

2.3.4. Results

2.3.4.1. Description of the morphs

We refer to the three morphs as “ambiguous” and “extreme” morphs of *C. scaldensis*, and *C. meneghiniana*. The “ambiguous“ morph included the ambiguously identified valves found in cultures isolated from the River Geeste and in the original samples, as well as similar valves from the River Schelde. Valves of the “extreme” morph were predominantly found in the Schelde sample, only four valves were encountered in the Geeste material, and none in culture. *C. meneghiniana* valves included those unambiguously identified as belonging to this species, from cultures and from field samples from both rivers.

All three morphs showed the typical features of the *C. meneghiniana* complex (Figs 3-26; Håkansson, 1990; Håkansson & Chepurnov, 1999; Håkansson 2002). Valves were areolated peripherally but not in the central area (apart from some of the smallest valves, < 7 µm diameter). The central area was smooth on the inside, and smooth to colliculate (rugose) on the outside (Figs 6-8, 14-16, 22-24; Muylaert & Sabbe, 1996). It was tangentially undulate, with the central FPs in the elevated part when observed from the valve exterior. The single RP was found in front of the elevated part of the undulation and the central FPs. The marginal FPs occurred with varying frequency on the costae, spines were often inserted at their outer

openings. All FPs had three satellite pores. When observed from the valve exterior, costae corresponded to the furrows in the undulating striated peripheral area.

The following differences could be observed between the morphs: *C. meneghiniana* had long alveolar chambers (Figs 4-5, 10), with open central portions. In both *C. scaldensis* morphs the central parts of the alveolar chambers were covered by central lamina (Figs 11-13, 17, and Figs 19-21, 26, respectively). The RP lips were radially oriented in *C. meneghiniana* (Fig 10), but twisted in the *C. scaldensis* morphs (Figs 17, 26). In *C. meneghiniana*, marginal FPs were found on almost every costa (Figs 3-5), whereas they occurred on every second to fourth in the *C. scaldensis* morphs (Figs 11-13, 19-21).

The most marked differences between the two *C. scaldensis* morphs were found in the frequency of the marginal FPs and in RP morphology. The "ambiguous" morph had more frequent marginal FPs than the "extreme" (Figs 11-13, 19-21). The number of central FPs also differed in the two morphologies: valves of the "ambiguous" morph had fewer than valves of the same diameter assigned to the "extreme" morph (Figs 12-13 to 20-21). RPs had somewhat undulate lips in the "ambiguous" morph (Fig. 17), whereas in the "extreme" morph, they were straighter and wider (Fig. 26).

Examining the figures in Muylaert & Sabbe (1996), we found that their concept of *C. scaldensis* probably included both morphs as described above. They included a valve interior that we would recognize as the "extreme" morph (Muylaert & Sabbe, 1996: fig. 11), together with ones (Muylaert & Sabbe, 1996: figs 12,13) that we would identify as the "ambiguous" morph. Thus, according to their description, both morphs belong to *C. scaldensis*.

Our morphometric results (below) indicated that the two *C. scaldensis* morphs could be differentiated by their marginal FP frequency and the number of central FPs at any given size. Based on this, we were then also able to compare exterior valve views of the morphs. The valve face was more pronouncedly colliculate in the "extreme" morph (compare Figs 14-16 and 22-24). Spines at the outer openings of processes were straighter in this morph (Fig 25), whereas those of the "ambiguous" morph had a more elaborate shape (Fig 18). Both *C. scaldensis* morphs differed from *C. meneghiniana* by having small emergences at the external process openings, especially marked for the RPs. Such emergences were absent in *C. meneghiniana* (Figs 18, 25; Håkansson & Chepurnov, 1999: figs 34, 35).

2.3.4.2. Conventional morphometrics

Pair-wise scatter plots of variables provide perhaps the simplest approach for looking at multivariate data. Fig. 27 shows the number of marginal FPs (MFP) plotted against number of costae (NC). These variables showed the most marked differences between the two morphs of *C. scaldensis*, showing no overlap in this plot. On the other hand, the "ambiguous" morph overlapped with *C. meneghiniana* at the smaller end of the size range, while being quite distinct at larger sizes.

The range of the ratio of NC and MFP (calculated as (NC-1)/MFP because one costa is always occupied by the RP) in the "extreme" *C. scaldensis* morph (2.44 – 3.92) did not overlap with that of either the "ambiguous" morph (1.24 – 2.25) or *C. meneghiniana* (1.0-2.0) in our sample. Similar ratios are not generally appropriate for compensating for size-related differences and are often less effective in revealing morphological distinctness than an analysis of the interdependence of the two variables (Bookstein *et al.*, 1985; Theriot, 1988). In this particular case, however, (NC-1)/MFP proved to be a useful index that unequivocally discriminated the two morphs of *C. scaldensis*. Assuming that the major axis regression lines, MFP=a*NC+b (a=0.35, b= -1.98 for the "extreme" and a=0.59, b=-2.0 for the "ambiguous" morph), describe the interrelationship of these variables well, we can expect that the value of the (NC-1)/MFP index approaches 1/a at relatively large values of NC, i.e., if NC>10. This is

well below the observed minimum value of NC in *C. scaldensis* (20 for the "ambiguous", 39 for the "extreme" morph). We can therefore expect that for these morphs, the curvilinearity occurring at small sizes (Theriot 1988) will not impede the ability of this ratio to distinguish them. However, whereas the bivariate plot (Fig. 27) indicates that valves of the "ambiguous" morph are more different from *C. meneghiniana* at larger sizes, this does not emerge if the ratio of the two variables is applied.

To summarize the variation in the eight recorded variables, a principal component analysis of their correlation matrix was calculated. The first and second principal components accounted for 75.16 and 13.55 % of the total (standardized) variation, respectively. The scatter plot in Fig. 28 also showed that the two morphs of *C. scaldensis* were distinct from each other with respect to the eight variables used, whereas scatters of the "ambiguous" morph and *C. meneghiniana* again overlapped at the lower end of the size range. Nevertheless, as in Fig. 27, the PCA scatter plot strongly suggests that there were three different size reduction series in our samples. The loadings of the variables on the first principal component (PC1) were all positive (Fig. 28) and PC1 accounted for 75% of total (standardized) variation. Thus, PC1 reflects overall size variation in our sample (Humphries *et al.*, 1981). This suggests that morphometric differences between *C. meneghiniana* and the "ambiguous" morph were clearer at larger sizes (at larger values of PC1), but smaller specimens were more similar to each other. The loadings of the variables on the first two principal component axes and reconstructed morphologies corresponding to some points in the PCA plot (Figs 28-29) illustrate that PC2 contrasted RP width (RW) and NC with MFP. Thus, points with larger PC2 values represent valves with a wider RP, more costae, and fewer marginal FPs. The other variables were not strongly correlated with this axis, which separated the morphs.

Furthermore, Figs 27-28 show that cultured cells of *C. meneghiniana*, and those found in the River Schelde, were practically indistinguishable from cells of this species from the Geeste material. On the other hand, similarity of the groups (cultures, Geeste and Schelde material) of the "ambiguous" morph was less. Nevertheless, these also overlapped to a large extent, and their differences seem to be caused by variation in the size distributions of the three samples. Cultures of this morph tended to be smaller (had smaller values on PC1) than cells of the same morph in the original field samples (River Geeste); those found in the Schelde sample were even larger.

Results of a canonical variate analysis of the eight variables are shown in Fig 30. This technique maximizes the separation of predefined groups (the three morphs) in multivariate space with respect to their within-group variance. The figure shows that even *C. meneghiniana* could be distinguished without overlap from the "ambiguous" morph in the variables used, despite their overlap in the PCA plot. Fig. 30 also shows that valve diameter (D1), diameter measured between the inner edges of the mantle (D2) and RP length (RL) contributed least to the separation of the morphs, and that CV2 was only strongly correlated with the number of central fultoportulae (CFP). The other variables all contributed to the separation of the groups on the first axis (CV1).

Discriminant analyses provided a further test of the distinctness of the morphs. In a regular discriminant analysis, two *C. meneghiniana* specimens were misclassified as the "ambiguous" morph, all other specimens were classified correctly. In a jack-knifed discriminant analysis, two *C. meneghiniana* specimens were misclassified as "ambiguous", and one specimen of the "ambiguous" morph as *C. meneghiniana*. Thus, although the distinction between the "ambiguous" morph and *C. meneghiniana* was somewhat less marked than separation of the "extreme" morph from the others, the former were also distinguishable from each other with minimal error.

2.3.4.3. Geometric morphometrics

Fig. 31 shows the results of a relative warp analysis, a principal component analysis of partial warp scores, using a weighting factor α of -1. This weighting emphasizes small-scale differences, and we used it because we expected that interesting differences among the morphs would be found in configurations of closely spaced landmarks, rather than in large-scale features. This plot (Fig. 31) showed that landmark configurations separated *C. meneghiniana* from the *C. scaldensis* morphs. Separation of the latter morphs was not complete, although they showed negligible overlap. Thus, despite similarity in alveolar and RP morphologies of the *C. scaldensis* morphs, there were considerable differences in these features.

Points in the relative warps plot in Fig. 31 each represent a landmark configuration (see Fig. 2 for the correspondence of landmark points with anatomic features of the valve). The plots (Fig. 32) illustrate changes in these configurations when moving in different directions (up [Fig. 32a], to the left [Fig. 32b], to the right [Fig. 32c], down [Fig. 32d]) from the origin (the overall mean configuration) in the coordinate system of the relative warps (Fig. 31). Fig. 32b is particularly interesting because it shows the difference between an "ambiguous" *C. scaldensis* and *C. meneghiniana*. Changes in the relative positions of landmarks 4, 5 and 10 are the most marked. These changes correspond to a radial orientation of the RP head and long alveolar chambers and show that these are important differences between *C. meneghiniana* and the other morphs. The difference between the *C. scaldensis* morphs (Figs 32c, d) is less easily interpreted using these diagrams.

Canonical variates analysis (not shown) indicated more clearly that even the two morphs of *C. scaldensis* could be distinguished without overlap. The differences are illustrated most clearly by analysing the correlation of shape change with size (Fig. 33). *C. meneghiniana* (Figs 33a-c) has longer alveoli than the other two morphs and its RP head is always orientated radially. The contrasting tendencies between the two morphs of *C. scaldensis* also become more obvious. The "extreme" morph (Figs 33g-i) tends to have smaller alveoli and marginal FPs and a wider RP than the "ambiguous" morph (Figs 33d-f) at the same size (valve diameter).

As with the conventional morphometric analyses, morphological similarity of the three groups of *C. meneghiniana* (cultures, Geeste and Schelde samples) was closer than that of the three "ambiguous" morph groups (Fig. 31). As discussed above (cf. Fig. 28), this is probably due to the size differences. Valves from cultures showed more variation in the landmark configurations than those from field samples, a pattern not observed in the conventional analyses (Figs 27-28).

In a discriminant analysis using partial warp scores, no case was misclassified, whereas in the jack-knifed discriminant analysis, one *C. meneghiniana* valve was misclassified as "ambiguous" *C. scaldensis*, and five *C. scaldensis* specimens were misclassified as the other morph of this species, but never as *C. meneghiniana*. This again illustrates that *C. meneghiniana* was most distinct, while the two *C. scaldensis* morphs could also be distinguished with minimal error.

2.3.5. Discussion

2.3.5.1. Morphometrics

Electron microscopy has clarified several earlier problems of diatom taxonomy (Round *et al.*, 1990) simply by providing increased resolution of morphological features.

However, increased resolution alone is probably insufficient to resolve all taxonomic questions. One of the important problems when delimiting species is to decide whether morphological groups are distinct (Mann, 1999). Morphometric methods can help decide if morphological variation is continuous and they can identify which characters can be most effectively or easily used to distinguish groups identified (Theriot and Stoermer, 1984a, b). Furthermore, they can help communicate results and information about morphological variation over many specimens, the range of which could not be illustrated photographically. They therefore constitute an important part of the methodological arsenal of diatom taxonomy.

Basic morphometric insights are of fundamental importance to diatom taxonomy. Some of these are already gaining recognition: the increasing use of multivariate data analysis methods instead of single-variable comparisons, and the characterization of allometric trends by analysing the interdependence of size-dependent quantities instead of their reduction to a ratio (Theriot, 1988). Reproducing morphologies based on the analysed parameters is a classical morphometric approach (Bookstein *et al.*, 1985) that might also be useful for diatom studies, as shown in diatom outline shape analyses (Mou & Stoermer, 1992). It can be used to verify the adequacy of recorded characters by comparing specimen morphologies with diagrams based on the data collected. Furthermore, it can also be used to help interpret results in a variety of ways, e.g. presenting diagrams corresponding to different points in morphospace (Figs 29, 32) or to phases of a size reduction series (Fig. 33). This is an often-used advantage of many landmark- and outline-based geometric morphometric methods. Although this is not necessarily readily achievable in conventional morphometric analyses, Tropper (1975) and this study illustrate that it could be feasible.

The use of landmarks, which has become a central concept of morphometrics in the last two decades (Rohlf & Marcus, 1993; Dryden & Mardia, 1998), has not yet established itself in morphometric analyses of diatom valves. Mou & Stoermer (1992) suggest that this is because “most diatoms do not have outline landmarks” and that “internal structures, such as labiate processes, are usually not qualified as landmarks”, because of the lack of knowledge about their function and ontogeny, and because their number usually varies within a genus, and often within a species. We think that, at least in some diatom groups, landmark points can be chosen so that we have as much confidence in their homology between specimens as in that of our conventional morphometric variables. The varying number of most structures found on diatom frustules is, on the other hand, a practical problem. To compare shapes of the repetitive structures themselves (like this study) might enable us to circumvent this problem. The characterization of RP morphology in our study is a good illustration of a case where landmarks characterize morphologies more intuitively and precisely than simple measurements. Using the traditional approach, a difference in the “width” of the RPs between the *C. meneghiniana* and the *C. scaldensis* morphs (Fig 29) could be revealed. The same differences could, on the other hand, be interpreted more precisely as a twist of the RP head in the landmark based analyses (Figs 32-33; compare Figs 9-10, 17, 26).

2.3.5.2. Taxonomic identity of the ambiguous morph

The two different approaches introduced to analyse valve ultrastructural variation proved to some extent complementary: differences between the two morphs of *C. scaldensis* emerged more clearly in the conventional morphometric analyses (Figs 27-30), those between *C. meneghiniana* and the “ambiguous” morph in the geometric ones (Fig. 31). Results from both approaches confirmed that 1) valve morphologies of the cultivated strains corresponded to valve morphologies of the morphs in the field samples, and 2) the three morphs were morphologically distinct. The analyses allowed us to identify our “ambiguous” cultures as a

C. scaldensis morph that has an intermediate valve morphology between the other ("extreme") morph of this species and *C. meneghiniana*.

The biological relevance of the morphological distinctness of samples of diatom valves depends, first of all, on whether the differences can be explained simply by size differences. The characteristic cell size reduction, inherent in many diatom life cycles, can lead to morphological changes that cause smaller and larger specimens of a clonal culture to look quite different (Meyer *et al.*, 2001). The morphological differences found between our three morphs cannot be explained by this phenomenon, because their distinctness was found despite overlapping size ranges (*C. meneghiniana*: 4.4 – 32.5 µm; "ambiguous" morph: 7.1 – 24.3 µm; "extreme" morph: 12.8 – 27.1 µm).

Alternatively, morphological differences might be explained by intraspecific geographical variation or phenotypic responses to different environmental conditions. To date, little is known about geographical variation in diatom valve morphology. On the other hand, morphological plasticity has been observed in a closely related species, *C. cryptica*. Strains of this species exhibited valve morphology characteristic of *C. meneghiniana* when grown in low salinity media, whereas they produced the typical *C. cryptica* morphology at higher salinities (Schultz, 1971). In our study, we did not attempt to study phenotypic plasticity, but any effect of environmental or geographic factors on valve morphology was reduced to a minimum by the sampling design. The compared valves were, 1) grown under the same culture conditions (medium, temperature, light intensity, light-dark cycle), and 2) co-occurred in the same field samples.

Culture studies could reveal cases where morphological variation spanned proposed species boundaries in this genus, as illustrated by Hegewald & Hindáková (1997) for the *C. ocellata*-complex. The occurrence of different morphs in their study was partly size-dependent (connected to life cycle), but also seemed to represent phenotypic variability. We found neither size-dependent morphological changes nor size-independent morphological variability that exceeded morph limits. However, variability was observed in some minor features. In particular, the presence and morphology of spines was highly variable in the "ambiguous" *C. scaldensis* cultures. Valves with regularly and irregularly spaced spines, as well as those without any spines, could be found in a single sample of a clonal culture. Presence of spines seemed to be a more stable character in *C. meneghiniana*: valves always possessed spines, although they were not always present at each FP opening.

The availability of cultures of the "extreme" morph could provide the possibility for comparative experimental studies of phenotypic plasticity and/or for molecular genetic comparison with the other morphs. However, even in the absence of cultures, the co-occurrence of the three morphs in the same field samples at overlapping size ranges strongly suggests that the morphological differences between them result from their genetic distinctness – i.e. the three morphs constitute different morphospecies.

We note that ribosomal DNA sequence comparisons of these strains further reinforce the possibility that they belong to distinct species (see **Publication II**). Furthermore, whereas cultures of *C. meneghiniana* were easy to isolate and grow, even for a long time in a variety of media, cultures of the "ambiguous" morph were more problematic. The latter required the addition of river water and frequent transfers when they reached stationary phase in order to survive. This indicates that the cultures differed not only mophologically, but also physiologically, and further supports their separation into different species. The three morphs also seem to have different ecological preferences, as illustrated by their contrasting abundances in the two estuarine sites.

However, a number of questions remain to be answered about the three morphs, to improve understanding of this particular taxonomic group, not only for its own sake, but also as a model of diatom species complexes in general. With respect to morphology, the most important question is whether these morphs remain morphologically distinct on a global scale,

and whether they show morphological plasticity. On the other hand, the biological and ecological differences between the morphs need to be clarified to permit any "practical" consequences of distinguishing them, for example in water quality monitoring.

2.3.6. Acknowledgements

Thanks to Koenraad Muylaert and Koen Sabbe for providing the sample from the Schelde. This paper was made possible by B. Beszteri´s attendance at the Amsterdam Morphometrics Workshop in 2002, helpful suggestions from Edward Theriot, and the Stony Brook Morphometrics web site (http://life.bio.sunysb.edu/morph/). We would also like to thank Friedel Hinz and Ute Bock for their help with the electron microscopy, Uwe John and Alberto Garcia Sáez for discussions, and Eileen Cox, Marina Montresor, Richard Crawford and two anonymous reviewers for valuable comments on different versions of the manuscript, which helped to improve it greatly. This work was in part supported by the project "Algaterra" of the German Federal Ministry of Education and Research (project ID 01LC0026, http://www.algaterra.org) and by grant no. FKFP-0154/2000 of the Hungarian Ministry of Education.

2.3.7. References

ANDERSEN, R. A., MORTON, S. L., & SEXTON, J. P. (1997). CCMP - Provasoli-Guillard National Centre for Culture of Marine Phytoplankton. *J. Phycol. Suppl.,* **33**: 1-75.

BOOKSTEIN, F.L. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, Cambridge.

BOOKSTEIN, F.L., CHERNOFF, B., ELDER, R.L., HUMPHRIES, J.M., SMITH, G.R. & STRAUSS, R.E. (1985). *Morphometrics in evolutionary biology: the geometry of size and shape change, with examples from fishes*. Academy of Natural Sciences of Philadelphia, Philadelphia.

DOS REIS, S.F., PESSÔA, L.M. & STRAUSS, R. (1990). Application of size-free canonical discriminant analysis to studies of geographic differentiation. *Rev. Brasil. Genet.,* **13**: 509-520.

DROOP, S. J. M. (1995). A morphometric and geographical analysis of two races of *Diploneis smithii* / *D. fusca* (Bacillariophyceae) in Britain. In *Proceedings of the Thirteenth International Diatom Symposium* (Marino, D. & Montresor, M., eds), 347-369.

DRYDEN, I.L. & MARDIA, K.V. 1998. *Statistical shape analysis*. John Wiley & Sons, New York.

DU BUF, H. & BAYER, M. M. (editors) 2002. *Automatic diatom identification.* Series in Machine Perception and Artificial Intelligence, Vol. 51.World Scientific Publishing, London.

HÅKANSSON, H. (1990). *Cyclotella meneghiniana* Kütz. (Bacillariophyceae), its morphology and reappraisal of similar species. *Nova Hedwigia Beih.,* **100**: 19-37.

HÅKANSSON, H. (2002). A compilation and evaluation of species in the genera

Stephanodiscus, *Cyclostephanos* and *Cyclotella* with a new genus in the family Stephanodiscaceae. *Diatom Res.,* **17**: 1-139.

HÅKANSSON, H. & CHEPURNOV, V. (1999). A study of variation in valve morphology of the diatom *Cyclotella meneghiniana* in monoclonal cultures: effect of auxospore formation and different salinity conditions. *Diatom Res.,* **14**: 251-272.

HAUSMANN, S. & LOTTER, A. F. (2001). Morphological variation within the diatom taxon *Cyclotella comensis* and its importance for quantitative temperature reconstructions. *Freshwat. Biol.,* **46**: 1323-1333.

HEGEWALD, E. & HINDÁKOVÁ, A. (1997). Variability of a natural population and clones of the *Cyclotella ocellata*-complex (Bacillariophyceae) from the Gallberg-pond, NW-Germany. *Arch. Hydrobiol. Suppl.,* **120**: 17-37.

HUMPHRIES, J.M., BOOKSTEIN, F.L., CHERNOFF, B., SMITH, G.R., ELDER, R.L. & POSS, S.G. (1981). Multivariate discrimination by shape in relation to size. *Syst. Zool.,* **30**: 291-308.

MANN, D.G. (1999). The species concept in diatoms. *Phycologia*, **38**: 437-495.

MEYER, B., WULF, M. & HÅKANSSON, H. (2001). Phenotypic variation of life-cycle stages in clones of three similar *Cyclotella* species after induced auxospore production. *Diatom Res.,* **16**: 343-361.

MOU, D. & STOERMER, E. F. (1992). Separating *Tabellaria* (Bacillariophyceae) shape groups based on Fourier descriptors. *J. Phycol.,* **28**: 386-395.

MUYLAERT, K. & SABBE, K. (1996). *Cyclotella scaldensis* spec. nov. (Bacillariophyceae), a new estuarine diatom. *Nova Hedwigia,* **63**: 335-345.

PAPPAS, J. L., FOWLER, G. W., & STOERMER, E. F. (2001). Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae). *Phycologia,* **40**: 440-456.

RHODE, K. M., PAPPAS, J. L., & STOERMER, E. F. (2001). Quantitative analysis of shape variation in type and modern populations of *Meridion* (Bacillariophyceae). *J. Phycol.,* **37**: 175-183.

ROHLF, F.J. (1993). Relative warp analysis and an example of its application to mosquito wings. In *Contributions to Morphometrics* (Marcus, L. F., Bello, E. & Garcia-Valdecasas, A., eds), 131-159. Museo Nacional de Ciencias Naturales (CSIC), Madrid.

ROHLF, F. J. (1999). Shape statistics: Procrustes superimpositions and tangent spaces. *J.Class.,* **16**: 197-223.

ROHLF, F. J. & MARCUS, L. F. (1993). A revolution in morphometrics. *Trends Ecol. Evol.,* **8**: 129-132.

ROUND, F. E., CRAWFORD, R. M., & MANN, D. G. (1990). *The diatoms: Biology and morphology of the genera.* Cambridge University Press, Cambridge.

SCHULTZ, M. E. (1971). Salinity-related polymorphism in the brackish-water diatom

Cyclotella cryptica. *Can. J. Bot.*, **49**: 1285-1289.

TEUBNER, K. (1995). A light microscopical investigation and multivariate statistical analyses of heterovalvar cells of *Cyclotella*-species (Bacillariophyceae) from lakes of the Berlin-Brandenburg region. *Diatom Res.*, **10**: 191-205.

THERIOT, E. (1987). Principal component analysis and taxonomic interpretation of environmentally related variation in silicification in *Stephanodiscus* (Bacillariophyceae). *Br. Phycol. J.*, **22**: 359-373.

THERIOT, E. (1988). An empirically based model of variation in rotational elements in centric diatoms with comments on ratios in phycology. *J. Phycol.*, **24**: 400-407.

THERIOT, E. & STOERMER, E. F. (1984a). Principal component analysis of *Stephanodiscus*: observations on two new species from the *Stephanodiscus niagarae* complex. *Bacillaria*, **7**: 37-58.

THERIOT, E. & STOERMER, E. F. (1984b). Principal component analysis of variation in *Stephanodiscus rotula* and *S. niagarae* (Bacillariophyceae). *Syst. Bot.*, **9**: 53-59.

TROPPER, C. B. (1975). Morphological variation of *Achnanthes hauckiana* (Bacillariophyceae) in the field. *J. Phycol.*, **11**: 297-302.

Table 1. Landmark points recorded for the geometric morphometric analyses. Their position is described on a valve the RP of which was oriented towards the bottom of the image. See also Fig. 2.

Number	Description
1	the center of the valve
2	the distal right corner of the alveolus at the left of the rimoportula (RP)
3	the distal left corner of the alveolus at the right of the RP
4	the end of the RP looking towards the upper right corner of the image
5	the other end of the RP head
6	the distal right corner of the alveolus at the right of the RP
7	the tip of the marginal FP at the first costa to the right of the RP
8	the distal left corner of the second alveolus on the right-hand side of the RP
9	the distal right corner of the same alveolus
10	the midpoint of the proximal edge of the same alveolus.

2.3.8. Figures

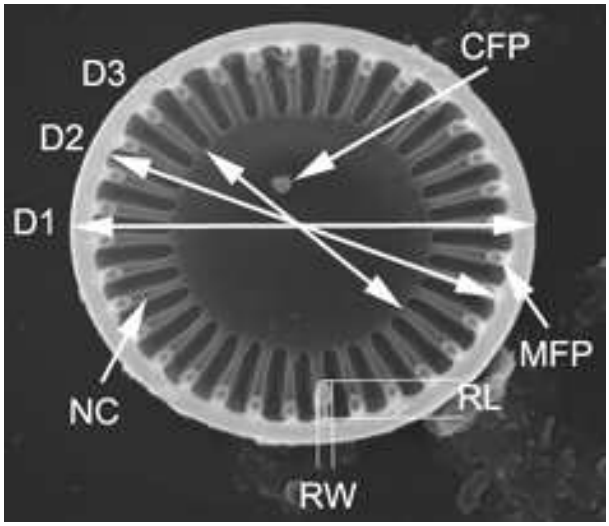


Fig. 1. Variables used for the conventional morphometric analyses, illustrated on a *C. meneghiniana* valve. D1 – valve diameter, D2 – valve diameter measured between the inner edges of the mantle (i.e., D1 minus twice the mantle thickness), D3 – diameter of the central area, RW – width of the rimoportula, RL – length of the rimoportula, NC – number of costae, MFP – number of marginal fultoportulae, CFP – number of central fultoportulae.

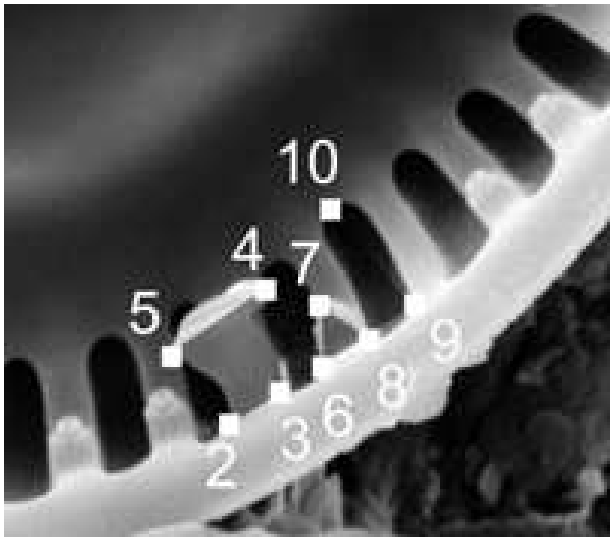
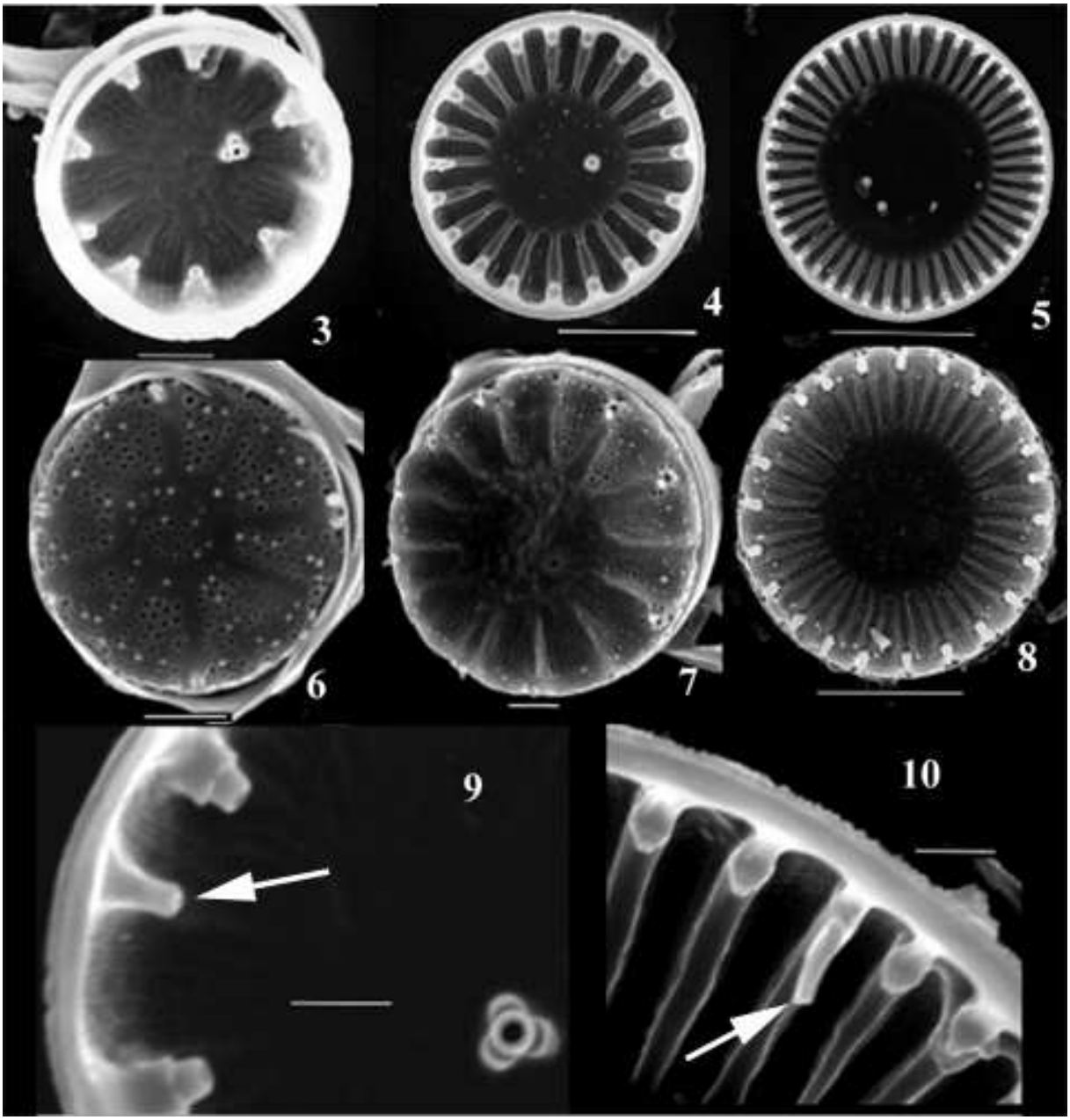
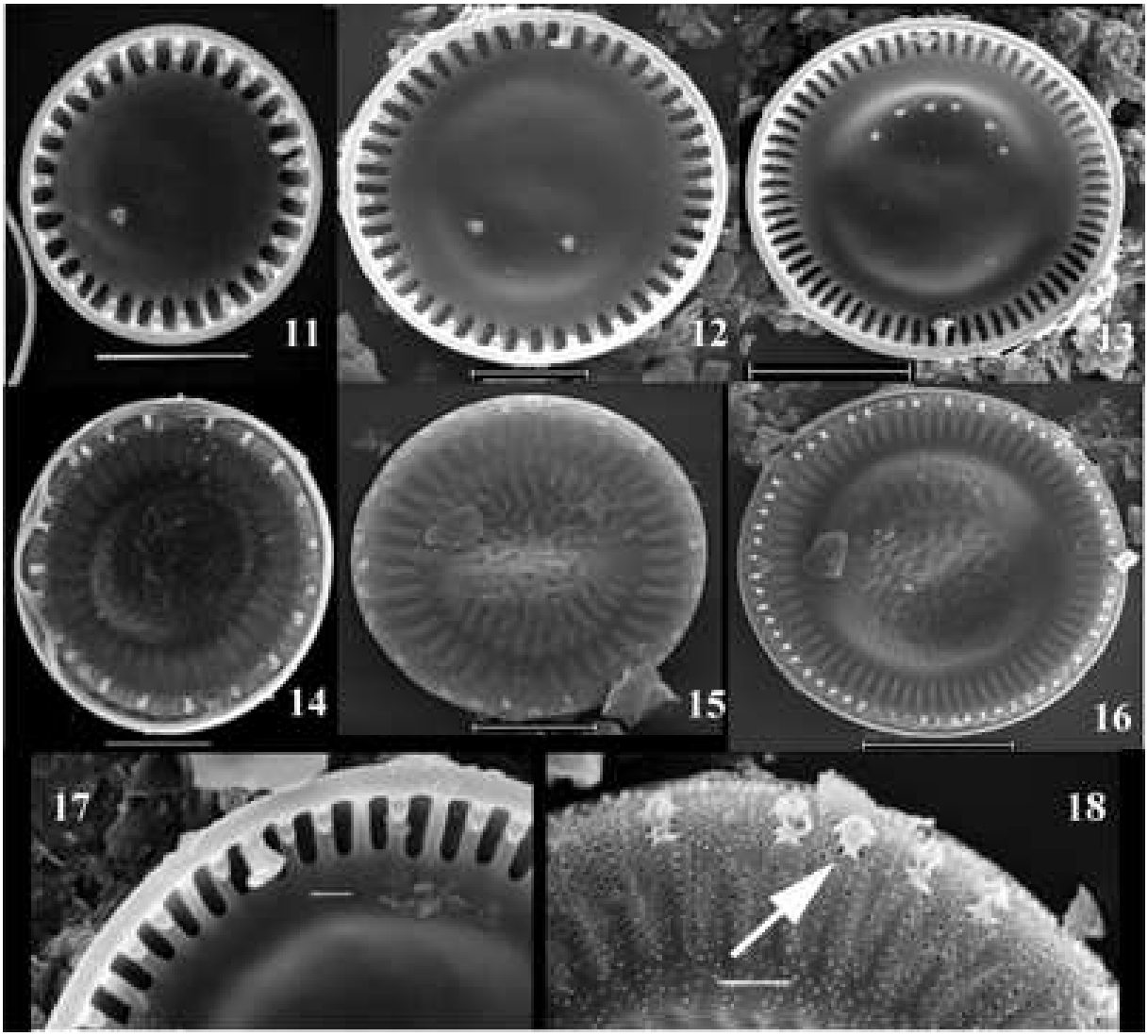


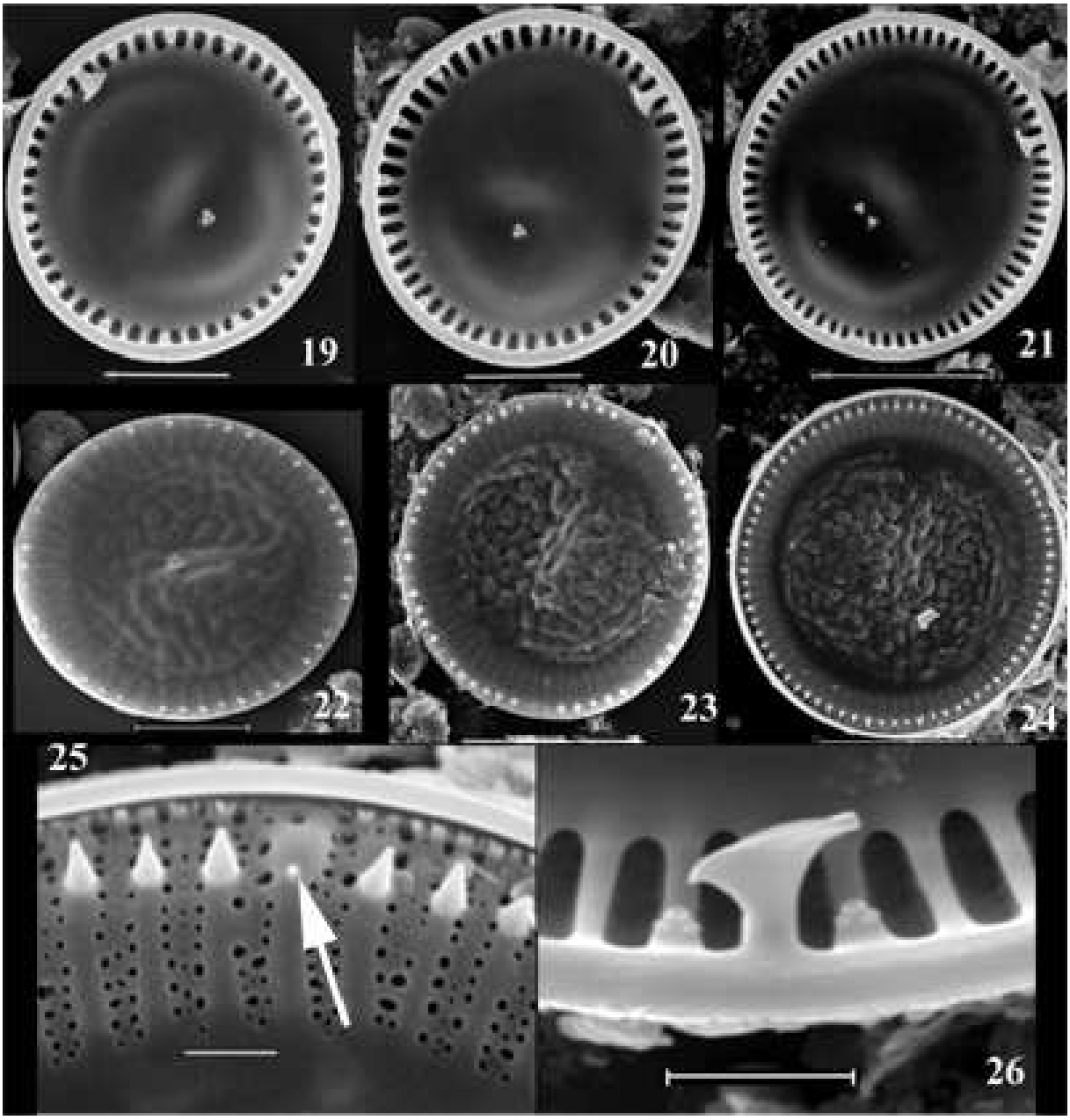
Fig. 2. Positions of landmarks 2 to 9 used for the geometric morphometric analyses shown on a valve from the “ambiguous” morph. Landmark 1, the midpoint of the valve is not shown here. Table 1 describes positioning of the landmarks.



Figs 3-10. Scanning electron micrographs of valves from *Cyclotella meneghiniana* cultures. Figs 3-5. Valve interiors. Figs 6-8. Valve exteriors over the size range. Figs 9-10, Detail of interior margins, showing rimoportulae (arrows) in an extremely small (Fig. 9) and a larger (Fig 10) valve. Scale bars represent 5 µm (Figs 3-8) and 1 µm (Figs 9, 10).



Figs 11-18. SEM images of valves of the "ambiguous" morph of *Cyclotella scaldensis*. Figs 11-13. Valve interiors. Figs 14-16. Valve exteriors over the size range. Fig. 17. Interior margin showing rimoportula. Fig. 18. exterior margin showing process openings; RP opening is marked by an arrow. Scale bars represent 5 µm (Figs 11-16) and 1 µm (Figs 17, 18).



Figs 19-26. SEM images of valves of the "extreme" morph of *Cyclotella scaldensis*. Figs 19-21. Valve interiors. Figs 22-24. Valve exteriors over the size range. Fig. 25. Exterior margin, showing process openings; RP opening marked by an arrow. Fig. 26. Interior margin with RP and two neighbouring FPs. Scale bars represent 5 µm (Figs 19-24) and 2 µm (Figs 25, 26).

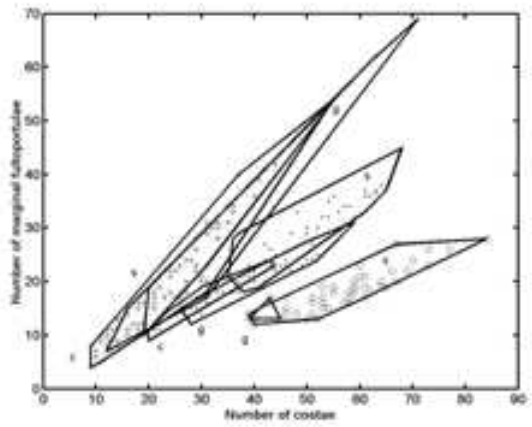


Fig. 27. Scatter plot of number of costae (NC) against number of marginal fultoportulae (MFP). Group outliers are connected by lines. (+ = *C. meneghiniana*, · = "ambiguous" morph of *C. scaldensis*, ° = "extreme" morph of *C. scaldensis*. c = cultures, g = field samples from the River Geeste, s = field samples from the river Schelde).

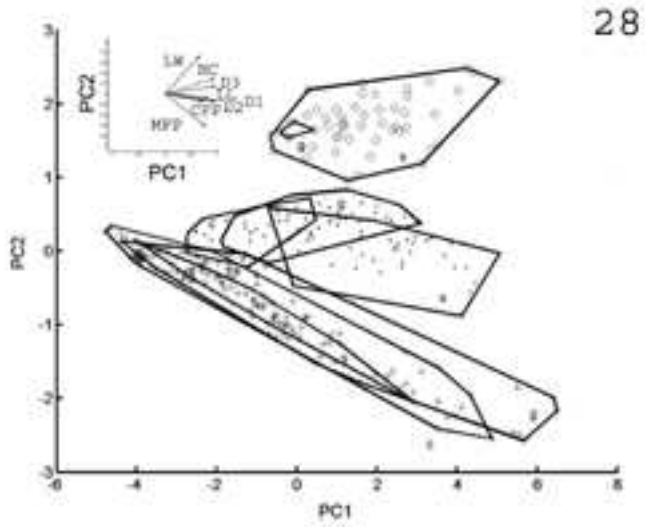


Fig 28. Scatter plot on the first two principal component axes of the conventional morphometric dataset. Group outliers are connected by lines. (+ = *C. meneghiniana*, · = "ambiguous", ° = "extreme" morph of *C. scaldensis*. c = cultures, g = field samples from the River Geeste, s = field samples from the river Schelde). Vector correlations of the original variables with the first two principal components are shown in the upper left hand corner.

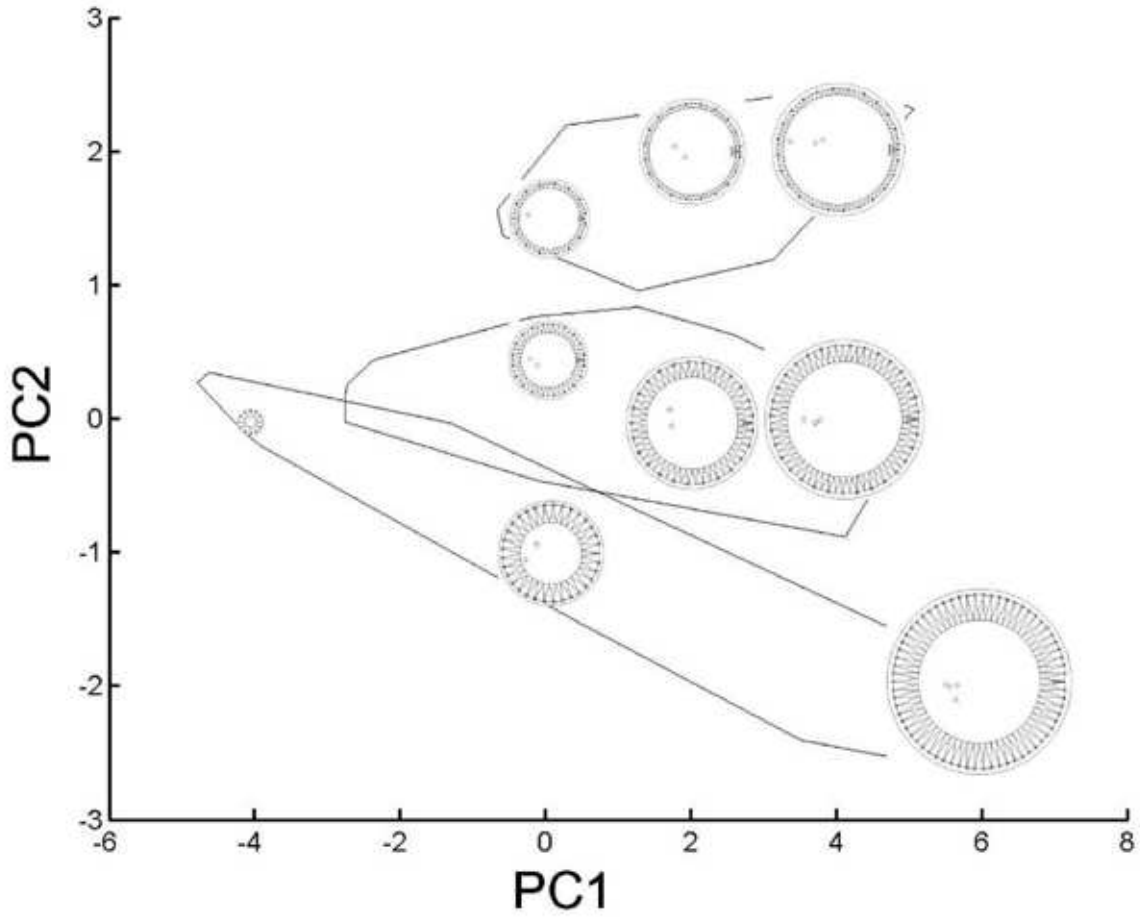
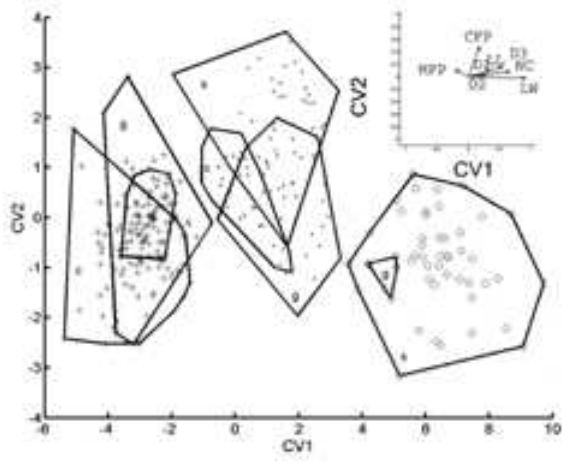


Fig. 29. Reconstructed morphologies of some points in the plane of the first two principal component axes (data shown in Fig 28). Group outliers of the three morphs are shown. Morphologies were "reconstructed" by projecting PCA scores back into the coordinate system of the original variables and drawing diagrams based on the resulting values (see Materials and Methods for details). The diagrams are drawn at the same scale. The morphologies shown correspond to PCA scores (-4, 0), (0, -1), (0, 0.5), (0, 1.5), (2, 0), (2, 2), (4, 0), (4, 2) and (6, -2).



Figs 30. Canonical variate analysis of the conventional morphometric dataset. Canonical variate scores calculated by grouping the specimens according to morph. (+ = *C. meneghiniana*, · = "ambiguous", ° = "extreme" morph of *C. scaldensis*; c = cultures, g = field samples from the River Geeste, s = field samples from the river Schelde). Group outliers are connected by lines. Vector correlations of the canonical variates with the eight original variables are shown in the upper right hand corner.

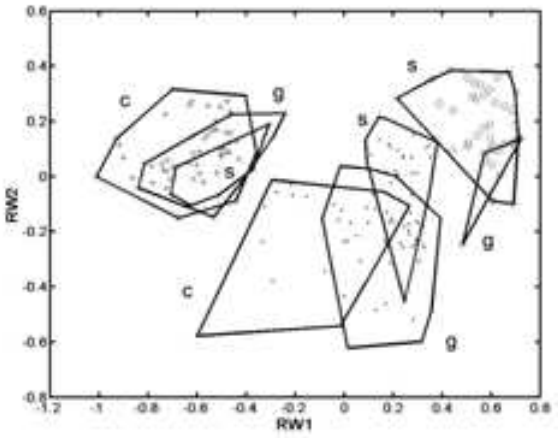


Fig. 31. Relative warps analysis with α=-1, small scale differences weighted more than global ones. Scores of specimens on the first two relative warp axes. Group outliers are connected by lines. (+ = *C. meneghiniana*, · = "ambiguous", ° = "extreme" morph of *C. scaldensis*; c = cultures, g = field samples from the River Geeste, s = field samples from the river Schelde).

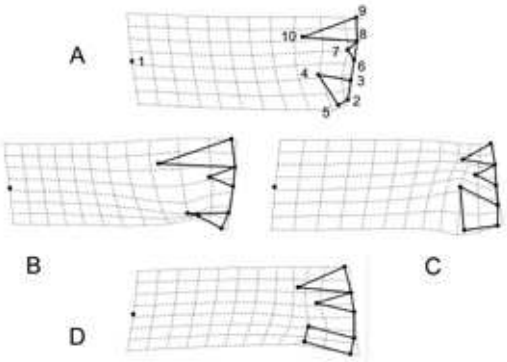


Fig 32. Landmark configurations corresponding to some points of Fig. 31 depicted as deformations from the mean configuration (the origin of the coordinate system in Fig. 31). Fig. 32A corresponds to the point (0, 0.5) of the coordinate system, Fig. 32B to (-0.5,0), Fig. 32C to (0.5,0), and Fig 32D to (0,-0.5). The points connected by thick lines represent the landmark points shown in Fig. 2 but the diagrams are rotated approximately 90^{o} anti-clockwise compared to Fig. 2.

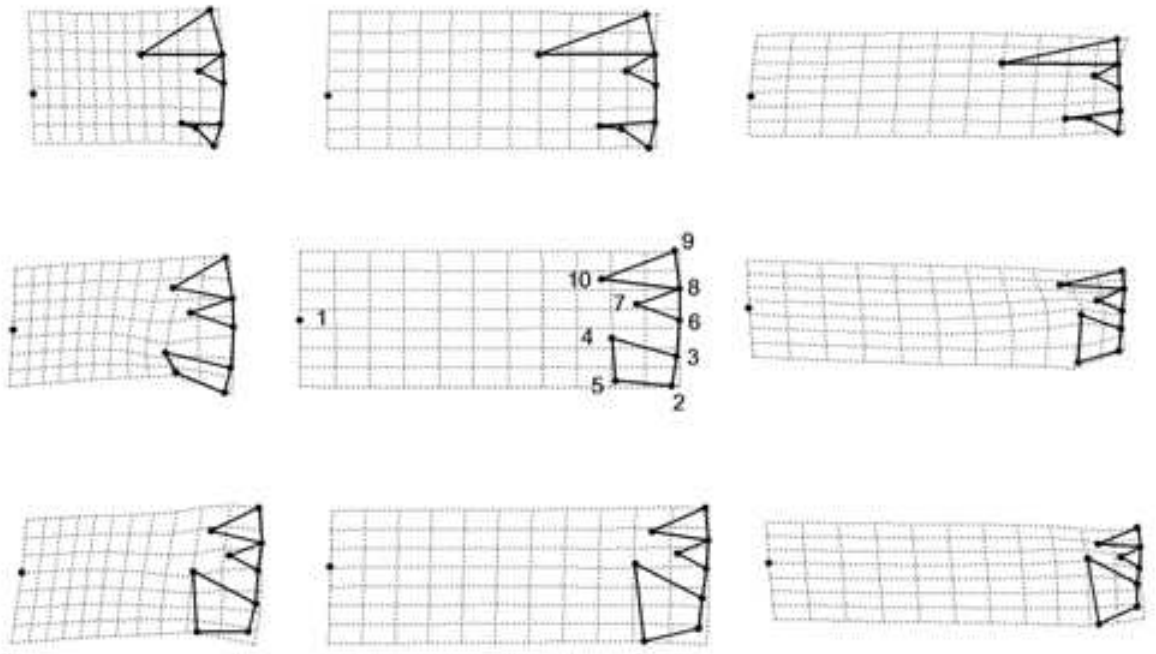


Fig. 33. Patterns of allometric valve shape change in the three morphs with size as calculated from a linear regression of the partial warp scores on the logarithm of valve diameter. Landmarks (see Fig. 2 to link them to the anatomic features of the valve interior) are connected by thick lines but the diagrams are rotated approximately 90^{o} anti-clockwise compared to Fig. 2. The landmark configurations predicted by the regression for three different diameters (D1 = 8, 16 and 24 µm, respectively) in the three morphs are shown in the columns and the different taxa by rows. Row 1 = *Cyclotella meneghiniana*, row 2 = "ambiguous" morph of *C. scaldensis* and row 3 = "extreme" morph of *Cyclotella scaldensis*.

2.4. Publication II: Ribosomal DNA sequence variation within and among clonal strains of the Cyclotella meneghiniana complex (Bacillariophyceae) from an estuarine locality

Bánk Beszteri[a], Éva Ács[b] and Linda K. Medlin[a]

[a] Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany

[b] Danube Research Station of the Hungarian Academy of Science, Jávorka S. u. 14, 2131 Göd, Hungary

2.4.1. Abstract

Cyclotella meneghiniana Kützing is one of the most commonly found and intensively studied freshwater diatom species. However, it is considered taxonomically problematic because of its unusually wide ecological range and large frustule ultrastructural variation. As part of a study to clarify patterns of morphological and genetic variation of this species, we surveyed nucleotide variation in the 18S and 5.8S rDNA, in the hypervariable D1/D2 regions of the 28S rDNA, and in the two ribosomal internal transcribed spacers (ITS) in a collection of twenty sympatric clonal cultures. Synthesis of these data with results of previously published morphometric analyses revealed that (1) a group of five strains were both morphologically and genetically distinct from *C. meneghiniana* (identified as *C.* cf. *scaldensis* Muylaert & Sabbe through the morphometric comparisons), and (2) that our group of *C. meneghiniana* strains contained four, morphologically indistinguishable, but genetically distinct lineages. The ITS sequences indicated that the latter were more likely cryptic sexual species than strictly clonal lineages. The presence of unrecognised diversity within *C. meneghiniana* is probably an important part of the explanation of its wide range of ecological and morphological variation.

Keywords: clonality, *Cyclotella meneghiniana*, microdiversity, rDNA, species boundaries

2.4.2. Introduction

Cyclotella meneghiniana Kützing is one of the most commonly recorded freshwater-brackish diatom species (Finlay et al. 2002) and is among the most extensively studied model species of them as well. Several studies have been published concerning its life cycle (Håkansson and Chepurnov 1999; Iyengar and Subrahmanyan 1944; Meyer et al. 2001; Rao

1970; Rao 1971; Rao 1996; Schultz and Trainor 1968). Furthermore, it has been the subject of a number of experimental physiological (El-Bestawy et al. 1996; Millie and Hersh 1987; Rosen and Lowe 1984), biochemical (Lohr and Wilhelm 2001; Louda et al. 2002) and molecular (Büchel 2003) investigations as a model diatom species.

However, the taxonomy of *C. meneghiniana* is still problematic. It shows extreme variability in frustule morphology when compared to closely related taxa (Håkansson and Chepurnov 1999; Schoeman and Archibald 1980). Also, it is found in an unusually wide variety of habitat types, ranging from oligotrophic lakes to strongly polluted waters and ones with elevated salinities, including brackish parts of estuaries (Håkansson 2002). Third, a study of chloroplast DNA restriction fragment length polymorphism revealed (1) a much greater genetic variation within *C. meneghiniana* than within closely related species, and (2) the paraphyly of *C. meneghiniana* with respect to *C. cryptica* (Bourne 1992).

Based on these facts, and considering Mann´s (1999) notion that species boundaries in diatoms in general are rather poorly understood, the question arises whether *C. meneghiniana* really is a single species. Alternatively, an explanation of its comparatively large morphological and genetic variation and wide ecological range might be that it consists of multiple species, each of them morphologically, genetically and ecologically more narrowly defined. Species identification is critical for e.g. water quality monitoring and palaeoenvironmental reconstructions (Stoermer and Smol 1999; Whitton and Rott 1996), but also for general physiological, biochemical and molecular biological studies. Therefore, the question of species boundaries is of special importance in diatom taxonomy.

Aiming at resolving the taxonomy of *C. meneghiniana*, we established a collection of twenty sympatric clonal cultures from what was originally believed to be a single population of this species for morphometric and molecular genetic characterization. Here we present the results of our survey of nuclear ribosomal DNA variation within and among these twenty cultures. Morphometric analyses (Beszteri et al. 2005) revealed that five of the twenty cultures isolated for this study showed a size reduction series distinct from *C. meneghiniana*. We referred to these cultures as the "ambiguous" *C. scaldensis* morph in Beszteri et al. (2005) and will do so throughout this paper as well. Our aims were (1) to assess whether the rDNA sequences could further support that this "ambiguous" *C. scaldensis* morph is not conspecific with *C. meneghiniana*; (2) to assess whether patterns of rDNA sequence variation within and among the *C. meneghiniana* strains are consistent with expectations for a single, randomly mating population; (3) if they are not, can they reveal whether the "population" is strictly clonal or is it rather a complex of distinct sexually reproducing species.

We sequenced different parts of the ribosomal operon to answer these questions and to identify which rDNA region could be the most useful for more extensive surveys. As this gene is also used as species-level marker in molecular identification of microorganisms (Medlin et al. 1991; Simon et al. 1997; Simon et al. 2000), we also sequenced the complete 18S rRNA gene from several strains.

2.4.3. Methods

Cultures. Sample origins, the isolation of cells into cultures and culture conditions have been described elsewhere (Beszteri et al. 2005). The twenty cultures (listed in Table 1) were isolated from samples taken from the Geeste estuary (Northern Germany) on a single day. Already published scanning electron microscopic (SEM) observations and morphometric comparisons with field material (Beszteri et al. 2005) identified five of these cultures as belonging to a morphotype of the closely related morphospecies *C. scaldensis* that was morphologically distinct from *C. meneghiniana* ("ambiguous" *C. scaldensis*). The other fifteen cultures could unambiguously be identified as *C. meneghiniana* in the SEM. Images of

the cultures and additional information have been made available in the AlgaTerra Information System (www.algaterra.org).

DNA methods. Liquid cultures were harvested by filtration and DNA was extracted using a modified CTAB method (Doyle and Doyle 1990). Amplification and sequencing primers for the 18S rDNA were based on Medlin et al. (1988) and Elwood et al. (1985); for the D1 and D2 regions of the 28S rDNA, on Scholin et al. (1994), and for the ITS regions, on White et al. (1990). PCR products were purified with the QIAQuick PCR Product Purification Kit (QIAGEN, Germany) and directly sequenced on both strands. The 18S rDNA PCR products were sequenced with the Long Read Kit (Biozym, Hessisch Oldendorf, Germany). Sequencing products were electrophoresed on a LICOR 4000L sequencer (MWG, Ebersberg, Germany). The D1/D2 regions of the 28S rDNA and the ITS regions were sequenced using Big Dye Terminator v3.1 sequencing chemistry (Applied Biosystems, CA, USA); sequencing products were electrophoresed on an ABI 3100 Avant sequencer (Applied Biosystems, CA, USA). Whereas the 18S and partial 28S rDNA PCR products could be sequenced directly, the ITS regions had to be cloned. For this, PCR products were ligated into pCR2.1 vectors and cloned following the manufacturer's instruction with the TA Cloning Kit (Invitrogen, Germany). Three to seven clones were sequenced from ITS PCR products containing the ITS1 and 2 regions and the 5.8S rRNA gene from each of 13 cultures.

Sequence assembly and analyses. Sequences exported from corrected electropherograms were assembled using DNASYS (Pharmacia, Germany) in the case of the 18S rDNA sequences. The partial LSU sequences were analysed using SeqScape 2.1 (Applied Biosystems, CA, USA) using the following settings. For base calling, the KB basecaller was used; for the detection of ambiguous positions (within-strain polymorphisms), the lower limit of calling an ambiguity was set at 30 % (i.e., an ambiguous base was called if a smaller peak with at least 30 % the signal intensity of the major peak was detected). After autoanalysis, each polymorphic position was checked manually in all electropherograms, and at last, all singleton mutations and ambiguous positions were re-checked. Sequences of the ITS regions were more variable so that they could not be processed similarly, thus they were assembled one by one using SeqMan (Lasergene package, DnaStar, Madison, WI, USA) and aligned with each other and closely related ITS sequences available from GenBank (acc. numbers U3073-U3077) using ClustalX (Thompson et al. 1997).

After checking basecalling, sequences of rDNA variants were reconstructed from the partial 28S rDNA sequences containing ambiguities following Clark (1990). In sequences available from GenBank, boundaries of different functional regions of the ribosomal operon are assigned in a variety of ways. For our sequences, we accepted the assignments in the sequence that is the closest relative of those determined here, sequenced from a *C. meneghiniana* culture by Zechman et al. (1994) (accession nr. U03073).

Basic statistics for characterizing levels of within-strain and within-morphotype ITS sequence variation were calculated using MEGA version 2.1 (Kumar et al. 2001). Phylogenetic analyses were performed using PAUP* 4.0b10 (Swofford 1998). For maximum likelihood (ML) and distance based tree calculations, models of nucleotide substitution were compared using Akaike Information Criterion (AIC) scores calculated on a neighbour joining tree as implemented in Modeltest 3.0 (Posada 1998). The model thus chosen was used for phylogenetic analyses using ML and neighbour joining (NJ) tree calculation with ML distances. Maximum parsimony (MP) trees were obtained in heuristic searches, with 10 random taxa addition sequences. Bootstrap analyses (NJ with ML distances and MP) were performed in 1000 replicates.

The hypothesis of purely clonal reproduction within *C. meneghiniana* was tested by an analysis of molecular variance (AMOVA, Excoffier et al. 1992) of the ITS data using Arlequin v. 2.000 (http://anthro.unige.ch/arlequin).

The dataset of the sequences of the ITS regions was explored for traces of recombination with the method of phylogenetic correlation as implemented in the program PhylPro (Weiller 1998). Bootscanning analysis (Salminen et al. 1995) was performed with selected candidate sequences using the program SimPlot version 2.5 (Lole et al. 1999) to identify potential parents of candidate recombinants and putative recombination sites. Because SimPlot cannot handle more than 26 sequences, besides candidate recombinant sequences, only one representative from each well-supported and stable clade was included in these analyses. Bootscanning was performed with a variety of window sizes between 100 and 250 bases, with a step size of 20 nucleotides. NJ trees were calculated with Kimura two-parameter distance corrections with a transition-transversion ratio of 2.0 and bootstrapped in 100 replicates. Candidate sequences were subsequently bootscanned against their putative nonrecombinant parent sequences and a fourth ("outgroup") sequence to use the Find Sites feature of SimPlot to identify putative recombination points.

Secondary structure model of the 18S rRNA was based on published diatom SSU secondary structures: The European Ribosomal RNA Database, http://oberon.fvms.ugent.be:8080/rRNA/index.html; The Comparative RNA Website, http://www.rna.icmb.utexas.edu/, and Medlin et al. (1996). For hypervariable loops, thermodynamical folding predicted using the mfold server with default parameters (http://www.bioinfo.rpi.edu/applications/mfold/, Zuker 2003) was used as a guideline.

2.4.4. Results

2.4.4.1. 28S rDNA

Tables 1 and 2 summarize information about the partial 28S rDNA data set. As no within-strain insertion-deletion polymorphisms were encountered, our approach to deal with intragenomic variation in this sequence region was to identify double peaks in electropherograms based on two criteria: relative peak height and sequencing noise in the neighborhood of the position concerned. Positions thus assigned ambiguous with a high confidence were marked by ambiguity codes, and sequences of the rDNA variants which probably occurred in the PCR product concerned, were inferred based on the algorithm of Clark (1990). Ten of the 13 unique sequences obtained contained 0 or 1 ambiguous position, thus variants present in these sequences could unambiguously be resolved. The remaining three sequences contained three (two strains), four, and 22 (both found in single strains) ambiguous positions. These could also be resolved into two variants each, with at least one of them identical to an already unambiguously resolved one (Table 1).

The two morphs (morphospecies). The two morphs (*C. meneghiniana* vs. the "ambiguous" *C. scaldensis* morph) differed markedly in their partial 28S rDNA sequences. Four different sequence variants (designated amb1 – amb4, Table 1) were found in the five strains of the "ambiguous" *C. scaldensis* morph (Fig. 1). The rest of the variants (designated A1 to D2) were found in the fifteen *C. meneghiniana* cultures. Thus, no variants were shared by the two morphs. Moreover, pairwise differences between sequences belonging to different morphs (*C. meneghiniana* vs. the "ambiguous" *C. scaldensis* morph) ranged from 36 to 45 nucleotides (Fig. 1), of which 25 differences were fixed between the two morphs.

Within C. meneghiniana. Four further distinct groups of 28S rDNA variants were found in the *C. meneghiniana* strains. The distribution of the pairwise distances showed a bimodal distribution with two distinct peaks (not shown): the number of substitutional differences between pairs of variants was either smaller than four or larger than 15 (Fig. 1). Based on this observation, variants differing from each other in less than four positions were grouped together. The resulting four distinct rDNA groups were designated A to D in Table 1.

Within-strain sequence variation was much lower than overall (within-"population") variation in *C. meneghiniana*. With the exception of one strain, only variants belonging to the same group were found in every *C. meneghiniana* strain (Table 1, Fig. 1). The exception was strain G184, with two variants (A1 and D1, Table 1) differing in 22 positions. Thus, to simplify following discussions, we will refer to groups of strains by the alphabetic code assigned to their 28S rDNA variants (see Table 1). Groups A to C included at least two strains each, whereas the single strain with exclusively D 28S rDNA variants formed group D, and the above mentioned mixed genotype (A plus D) strain formed group AD.

Phylogenetic analyses could not resolve the relationships of these groups of alleles in detail; the only relationship with bootstrap support > 50 % was the grouping of type C with D (Fig. 2).

2.4.4.2. ITS1, 5.8S rDNA and ITS2

Direct sequencing of the ITS regions (including ITS1, the 5.8S rDNA and ITS2) proved problematic, sequences were not readable because of within-strain length polymorphisms. Therefore, PCR products were cloned for sequencing. 13 strains were selected for this so that at least two strains from each 28S rDNA group (which contained more than one strain) were included. Three to seven clones were sequenced from each of these 13 strains (altogether 63 clones). Information about the ITS data set is summarized in Tables 1 and 2. Among the 63 sequences, 46 different variants were found, which were designated h0 to h45 (Table 1). Their alignment was 982 bps long. Within each strain, two to five different variants were found among the three to seven clones sequenced (Table 1).

The two morphs (morphospecies). The two morphs were also distinct in their ITS sequences. 40 of the 46 variants were obtained from *C. meneghiniana* cultures, whereas the remaining 6 variants came from the "ambiguous" *C. scaldensis* morph (Table 1). Thus, similarly to the 28S rDNA data set, no ITS variant was shared between the two morphs. Fixed differences between sequences from the two morphs included five large insertions in the ITS1 sequences from the "ambiguous" *C. scaldensis* morph (responsible for length differences of ~ 200 bps, Table 2), and 43 fixed mutational differences.

Within C. meneghiniana. Five distinct groups of ITS variants were found in the *C. meneghiniana* strains, as illustrated by the phylogenetic trees (Figs 2, 3). Three of the well separated clades recovered in the phylogenetic analyses corresponded to groups of strains defined above based on their partial 28S rDNA sequences (groups A, B and D in Figs 2, 3). Remarkably, two divergent ITS types (marked C1 and C2 in Figs 2, 3) were obtained from the strains of group C. Furthermore, strain G184 (in which two divergent 28S rDNA types - A and D - were found) also had two divergent ITS types, one (ITS variants h22, h24, h29 and h32) grouping together with group A, and one (h37) grouping with group D ITS sequences (marked in bold in Fig. 3). Sequence of one of the clones from strain G184 was identical to those of five other clones sequenced from cultures from group A (variant h22). ITS variants

were shared by more than one strain also in three further cases (Table 1 and Fig. 3). Strains with identical ITS alleles were in each case members of a single 28S rDNA group (h15 – group B, h22 – group A, and h44 – group E, i.e., the "ambiguous" *C. scaldensis* morph).

Two of the ITS variants proved likely recombinants. Analyses of phylogenetic profiles of the ITS data set revealed that ITS variants h32 and h33 showed particularly low phylogenetic correlations in at least one region (Fig. 4). Bootscanning and comparisons of fragments of these sequences with others in the data set revealed that the first 263 bases of h32 differed only in a single substitution from the same region of several group D sequences (h11, h12, h37 and h39), whereas from position 286 on, it was identical to group A sequences (h1 and h2). Thus, h32 was identified as the likely result of a recombination event between a group D and a group A ITS variant. Whereas neither of the putative parent sequences of h32 were sampled from the same strain (G184) as h32, ITS variants very similar to these putative parents were (h37 – group D, and h22, h24 and h29 – group A).

The other putative recombinant sequence, h33, showed traces of two recombination events (this is also illustrated by the two marked drops in its phylogenetic profile, see Fig. 4). The first 448 bases of this sequence differed only in a single substitution from group C2 variants (h34 and h36), and, from position 833 on, it was identical to variant h36 (group C2). However, between positions 640 and 767, it differed only in a single substitution from group C1 sequences (variants h9 and h10). Thus, one of the putative parent sequences of h33 (h36) was sampled from the same strain (G183K) as h33, whereas the other two putative parent variants were not found in our data set. However, variants very similar to these putative parents were found in the same strain as well (h36 – group C2, and h8 – group C1).

These recombinants were grouped at the base of the clades with which they shared the most parsimony informative sites when included in phylogenetic analyses (h32 appeared at the base of clade A, h33 at the base of clade C2), but they were excluded from the analyses presented (Figs 2, 3) because they violate the assumptions of phylogenetic inference methods used. The best models inferred by Modeltest for data sets including these sequences were indeed much more complex than those obtained when excluding these likely recombinants (not shown).

Regions with weaker, but also marked drops in the phylogenetic correlations could also be seen in the remaining sequences of group C2 (h34, h35 and h36; Fig. 4). Bootscanning analysis was performed with these sequences as well, in order to identify variants that might have given rise to them through recombination. However, the results revealed that informative sites supporting different affiliations (with groups B, D and with the "ambiguous" morph of *C. scaldensis*) were intermixed in these sequences (not shown). Therefore, we could not arrive at an unequivocal conclusion concerning their recombinant origin nor could we identify sequences most likely involved in recombinations resulting in these sequences.

Results of an AMOVA (Table 3) show that total ITS variation in our sample could be partitioned into significant within-strain and among-group covariance components, whereas the among-strains, within-groups component was not significant. This held independently of the inclusion of strain G184 (Table 3 shows the results obtained with G184 included).

Phylogenetic analyses could resolve the relationships of the ITS types when no outgroup sequences were included better than in the case of the partial 28S rDNA data set. However, rooting of the tree with available *Stephanodiscus* ITS sequences proved problematic because of the relatively large divergence between the in- and outgroup sequences (uncorrected p-distances > 0.3). In the analyses without outgroups, the topology shown in Fig. 2 was obtained. Altogether, both the relative branch lengths and the topology were very similar in the analyses of the two data sets (ITS and partial 28S rDNA); the topology was better resolved in the ITS analyses. The most conspicuous difference between the two trees was the presence of the two divergent C types in the ITS tree. When group C2 is removed, the only difference remains the better resolution of among-groups relationships of

the ITS analyses. ITS type C2 was (1) only distantly related to the other group of variants obtained from the same group of strains (type C1), (2) showed a divergence from all other *C. meneghiniana* sequences comparable to the ITS variants from the "ambiguous" *C. scaldensis* morph (see relative branch lengths in Fig. 2).

For rooting the ITS trees, an alignment with published *Stephanodiscus* ITS sequences was prepared using ClustalX, and all regions which could not unambiguously be aligned were removed manually. The resulting alignment had 530 bps. In phylogenetic analyses of this alignment, the rooting of the ingroup tree remained ambiguous (trees not shown). Whereas all analyses rejected a root position within the group (A, (C1, D)), practically no other possible root positions were excluded by these calculations.

The only *C. meneghiniana* ITS region sequence available from Genbank at the time of our study (Genbank acc. U3073) was also included in these rooting trials. Remarkably, it was more distant from our ingroup (including the "ambiguous" *C. scaldensis* sequences, uncorrected p-distances > 0.3; distances with Jukes-Cantor correction > 0.4) than from the *Stephanodiscus* sequences (U3074 – U3077; uncorrected p-distances < 0.3; distances with Jukes-Cantor correction < 0.4); in phylogenetic analyses, it occupied a distinct position outside both the *Stephanodiscus* and remaining *Cyclotella* clades (not shown).

2.4.4.3. 18S rDNA

The complete 18S rDNA was sequenced from eight selected strains (Table 1, Fig 5). The two morphs were also distinct in their 18S rDNA sequences: the two sequences obtained from the "ambiguous" *C. scaldensis* morph were identical, while displaying four fixed differences from those obtained from the six *C. meneghiniana* strains. Among these four fixed differences, two formed a pair of compensatory base changes in loop E10/1 (Figs 5A, 6). The differences observed among the *C. meneghiniana* strains included no such paired substitutions. The two sequences obtained from single 28S rDNA groups (groups A and B of *C. meneghiniana*) differed in one substitution, whereas the differences between 18S rDNA sequences from different groups ranged from 1 to 11 (Fig. 5).

2.4.5. Discussion

2.4.5.1. The morphs (morphospecies)

Morphometric analyses (Beszteri et al. 2005) showed that (1) five of the strains studied here (referred to as the "ambiguous" *C. scaldensis* morph) were distinctly different from the rest of strains morphologically, and that (2) the two morphs also remained morphologically distinct in sympatry in field samples. This was considered a strong indication that they constituted two distinct reproductively isolated groups (biological species). The rDNA sequence data obtained here further reinforced this view: the two morphs were distinct in all of the rDNA regions sampled. No rDNA variants were shared between the morphs. Furthermore, rDNA variants from one of the morphs were more similar to each other than to any of those sequenced from the other morph. Several fixed differences were found between cultures of the two morphs (i.e., the "ambiguous" *C. scaldensis* morph always formed a monophyletic group), including five large insertion-deletions in the ITS1 and a pair of compensatory base changes in the most conserved of the regions studied, the 18S rDNA. Pairs of complementary base differences in the much more variable ITS regions have been shown

to correlate with mating incompatibility in Volvocales green algae (Coleman 2000). In the 18S rDNA, such differences were also used to separate two species of the marine diatom genus *Skeletonema* (Medlin 1997).

2.4.5.2. *C. meneghiniana* – clonality or cryptic species?

The typical diplontic life cycle of diatoms consists of predominantly clonal reproduction, mostly accompanied by a characteristic cell size reduction. Size regeneration – auxosporulation – is tightly coupled to sexual reproduction in many species (Edlund and Stoermer 1997). However, despite the commonality of this basic plan, diatoms have a range of life cycles comparable to bacteria (Drebes 1977), often making the interpretation of "around the species level" phenotypic and genetic variation difficult. Centric diatoms, where the algae studied here are classified, are generally oogamous. However, some species seem to have abandoned allogamy, as indicated by the failure to observe it (Drebes 1977). Data available about *C. meneghiniana* are not fully conclusive in this respect. As gamete production and auxosporulation are easily induced in cultures of this species (Schultz and Trainor 1970), several reports about them have been published (Rao 1971; Schultz and Trainor 1968). It is thus well known that *C. meneghiniana* is monoecious, i.e., a single clonal culture is capable of both sperm and egg cell production, in a size-dependent ratio (Rao 1996). However, in spite of the several culture studies of these events, fertilization has never been observed, with the exception of self-fertilization of the oogonia (Drebes 1977; Håkansson and Chepurnov 1999; Iyengar and Subrahmanyan 1944).

As species concepts that are based on the idea of cohesion within a species though gene flow are not applicable to strictly clonal organisms, this has important implications to studies of species limits. Moreover, "effective clonality" has been observed in bacteria (Maynard Smith et al. 1993) as well as in several of the well studied pathogenic protists, despite their ability to reproduce sexually (Tibayrenc and Ayala 1991; Tibayrenc and Ayala 2002; Tibayrenc et al. 1991). Thus, the possibility of clonality should be considered when approaching "intraspecific" variation even in the case of protists where sexuality has been observed. Here, we try to use the rDNA sequence data obtained to distinguish three possible patterns of reproduction within our *C. meneghiniana* "population": panmixia, strict clonality (or strict autogamy, or "effective clonality", as these cannot be distinguished by using molecular genetic data), and the presence of more than one, sexually reproducing, but reproductively isolated groups (gamodemes, cryptic species).

These different reproductive patterns are expected to result in different patterns of genetic variation:

(1) Panmixia: gene flow is expected to regularly redistribute genetic variation among individuals. Gene variants occuring in a single strain thus represent a random sample from the pool of variants present in the population, i.e., the amount of within-strain genetic variation is expected to be statistically comparable with within-population variation. Furthermore, in the case of rDNA sequences subject to concerted evolution (Li 1997), intragenomic homogenization reduces the variation present in the whole population, leading to an expectation of low within-population rDNA sequence variability.

(2) Cryptic sexual species: the "population" sampled in fact consists of several independent populations (groups of interbreeding organisms). In the lack of gene flow among them, these are expected to diverge with time from each other genetically. However, sexuality within these "cryptic populations" is expected to result in low genetic variation within them and a comparable amount of within-strain variation, as described above (Tibayrenc and Ayala 2002).

(3) Clonality: there is no exchange of genetic information among individuals, each of them is "reproductively isolated" from every other. Accumulation of mutations occuring independently in the clonal lineages is expected to lead to an amount of within-"population" genetic variation exceeding intraclonal variation (similarly to the case where cryptic species are involved). However, in contrast to the presence of cryptic sexual populations, no group of individuals shares their gene pool – in any arbitrary group of individuals, intraindividual genetic variation is not a random sample from the within-group genetic variation but only represents a subset of the latter. If polymorphisms have accumulated in the region studied, this leads to an expectation that genetic variation within a random sample of individuals is larger than variation present within the individuals themselves.

According to these expectations, the data obtained from each of the rDNA regions sequenced strongly indicated that the *C. meneghiniana* "population" from which we sampled was not a panmictic population: the amount of within-"population" variation strongly exceeded within-strain variation in all regions. Variants found in any of the regions were not a random sample from the overall pool of variants found.

However, the hypotheses of strict clonality vs. the presence of more biological species are not so easily distinguished solely using the data we obtained. Unfortunately, the resolution of the markers used limited our ability to test the hypothesis of within-group panmixia (variation within the four distinct groups detected was limited even in the most variable ITS regions). An AMOVA indicated that ITS variation was evenly distributed among the strains of the individual groups, i.e., this test provided no evidence against the null hypothesis of panmixia within the cryptic groups. In any case, the result of this test should be interpreted more as an indication than as strong evidence because of the above mentioned limited resolution; however, to date, these are the only data available to test these possibilities.

Strain G184 was exceptional in our sample in that it had divergent 28S rDNA and ITS region types (types A and D). One possibility explaining this result could be that, in spite of caution taken to isolate single cells several times when establishing the cultures, this culture was not actually clonal. Alternatively, it might be the offspring of a hybrid of group A and group D strains. Unfortunately, we could not test the former possibility by re-isolating single cells from the culture. However, the important conclusions (i.e., evidence both against panmixia and strict clonality) held independently of the inclusion of this strain in the analyses.

The remaining *C. meneghiniana* strains were characterized by four distinct groups of 28S rDNA and ITS region variants. According to the above arguments, these most probably represent four distinct gamodemes (biological species). The presence of morphologically not recognized species within *C. meneghiniana* could also explain Bourne´s (1992) finding, that plastid DNA RFLPs of this morphospecies showed a comparatively large variation and paraphyly with respect to *C. cryptica*. Diatom morphospecies have in several cases turned out to be complexes of more than one reproductively isolated groups (biological species), most commonly in mating experiments – see Mann´s (1999) review for examples – and, in recent years, in molecular genetic surveys similar to ours (Lundholm et al. 2003; Lundholm et al. 2003; Medlin et al. 1991; Orsini et al. 2004; Orsini et al. 2004; Sarno et al. 2005). The situation is similar in other eukaryotic microbes with inorganic outer shells providing a number of taxonomic characters (De Vargas et al. 1999; Montresor et al. 2003; Saez et al. 2003). In some of these cases, minor morphological features distinguishing the previously unrecognized species could be found; in others, morphology on its own proved insufficient for their differentiation.

2.4.5.3. Frustule morphology

With the strains studied here, morphometric comparisons showed that the principal morphological difference among cultures with different rDNA types was in their size distribution. We mapped rDNA group memberships onto a principal component analysis (PCA) plot of the conventional morphometric data used in Beszteri *et al.* (2005) to illustrate the morphometric differences among the *C. meneghiniana* groups (Fig. 7). PC1 was strongly positively correlated with all the original variables (vector correlations ranged from 0.78 to 0.99); i.e., PC1 can be considered as an overall size-vector. Thus, this plot revealed that the main morphological differences among the rDNA groups were in their size distributions, as their largest separation was on PC1. Their scatter overlapped to an even larger extent in the plains of the other significant principal component axes (not shown), indicating that the frustule ultrastructural characters that were recorded by these morphometric variables could not differentiate between strains with different rDNA types.

We cannot assume that the size ranges of our cultures were representative of the natural situation (sample sizes ranged from only one to eight per group), thus, it is quite possible that also their size ranges overlap to a larger extent than in our small sample. In any case, the morphometric data point out that the cultures with different rDNA types probably will not be distinguishable without overlap by the frustule ultrastructural features classically used for species identification in this diatom group, which were reflected by the eight morphometric variables recorded. We also did not find other minor morphological features distinguishing them in detailed scanning electron microscopic observations.

A further complicating factor in the morphological identification of these groups might be the problem of frustule morphological plasticity: *C. meneghiniana* cultures have been shown to react with differential incorporation of silica in their frustules and frustule morphological differences to salinity differences of their medium (Tuchman et al. 1984). Furthermore, the closely related species *C. cryptica* has been shown to display extreme morphological plasticity: its frustules display the morphology considered characteristic of *C. meneghiniana* at low salinities (Schultz 1971). This indicates that “real” species boundaries might not be easily identified solely using morphology in this diatom group.

2.4.5.4. Recombinant ITS sequences

Traces of recombination events among ITS types implying past hybridization events between different rDNA groups have not been found in our sample. Two recombinant sequences could be identified with great confidence, both of which could be explained either as PCR artefacts or as the products of intragenomic recombination. A further group of sequences with possibly recombinant origin was pointed out by marked drops in their phylogenetic correlations (the group of ITS types marked C2, Fig. 4). However, we were unable to trace the origins of these sequences similarly to the above two probable recombinants. The pattern observed might either be explained by several recombination events, or by the occurrence of many homoplasious mutations.

A further interesting feature of this group of sequences (ITS type C2) was that they were only remarkably distantly related to another group of ITS variants found in the same strains (group C1, Fig 2). Type C2 was situated on a spectacularly long branch in the phylogenetic trees, which can be an indication of its accelerated evolution (reduced selective pressure) in connection with pseudogene formation, a phenomenon sometimes encountered in rDNA studies (Álvarez and Wendel 2003). The apparent recombination signals might be

explained by an accumulation of homoplasious mutations in sequences from ITS group C2 during this accelerated evolution.

2.4.5.5. Implications for taxonomy of other diatoms

The above results raise several questions of general interest to the taxonomy of freshwater diatoms. As noted above, *C. meneghiniana* is among the most intensively studied freshwater diatom morphospecies. Molecular diversity studies of freshwater diatom species have hardly been made to date (De Bruin et al. 2004; Lewis et al. 1997; Soudek Jr. and Robinson 1983). However, based on results with marine diatoms and other protists (see above references), we can expect to meet several morphospecies that turn out to be composed of more than one reproductively isolated entities. Molecular genetic methods seem to provide diatom taxonomy with a means to more efficiently reveal such cryptic diversity than the classical mating experiments: they can be applied in a wider range of taxa, a larger number of strains can be screened in less time, and the effects of potential changes in reproductive behaviour caused by cultivation are excluded. In the case of *C. meneghiniana*, we are further investigating the amount of molecular genetic diversity on a larger geographic scale to provide more detailed information on a model species of freshwater diatomologists, with the aim to contribute to a better understanding of whether the presence of cryptic diversity contributes to the extreme morphological and ecological spectrum of this morphospecies.

2.4.6. Acknowledgements

We thank Uwe John, Wiebe Kooistra, René Groben and Klaus Valentin for reading different versions of the manuscript and for their comments, which substantially helped to improve it. This work was in part supported by the project “Algaterra” of the German Federal Ministry of Education and Research (project ID BMBF 01LC0026, http://www.algaterra.org), and by grant no. FKFP-0154/2000 of the Hungarian Ministry of Education.

2.4.7. Tables

Table 1. List of strains. Group: amb – "ambiguous" *C. scaldensis* morph, Cm – *C. meneghiniana*; strains of the latter morphospecies are also grouped according to their 28S rDNA genotypes (shown in the fourth column). Acc. 28S – Genbank accession numbers of the partial 28S rDNA sequences. AP 28S – number of ambiguous positions in the partial 28S rDNA sequences. 28S variants – letter codes of the 28S rDNA variants obtained from the strain. Acc. 18S – Genbank accession numbers of the 18S rDNA sequences (n.s. – not sequenced). ITS variants – names of ITS variants obtained from the individual cultures (variants found in multiple clones from the same strain are listed multiply; n.s. – not sequenced; Genbank accession numbers of sequences of the ITS region are AY906776-AY906829 and AY911294-AY911302). Variants identified as probable recombinants are marked by asterisks.

Strain name	Group	Acc. 28S	Amb.pos. 28S	28S variants	Acc. 18S	ITS variants
G18W53	amb	AY496201	0	E1	AY496208	h42, h42, h44
G8W8	amb	AY496200	1	E1, E2	n.s.	n.s.
G18W42	amb	AY496198	3	E1, E3	n.s.	h40, h41, h43, h43, h44, h45
G18W44	amb	AY496199	3	E1, E3	n.s.	n.s.
G1W11	amb	AY496188	4	E1, E4	AY496209	n.s.
G16W1	Cm / A	AY496202	0	A1	n.s.	h0, h3, h23, h25
G17W8	Cm / A	AY496204	0	A1	AY496213	n.s.
G182	Cm / A	AJ878468	0	A1	n.s.	n.s.
G183N	Cm / A	AY496203	0	A1	n.s.	h1, h2, h2, h22, h22, h22

G8W5	Cm / A	AY496205	2	A1, A2	n.s.	h21, h22, h22
G8W9	Cm / A	AY496190	0	A2	n.s.	n.s.
G8W6	Cm / A	AJ878470	1	A2, A3	n.s.	h5, h26, h27, h28, h31
G8W4	Cm / A	AY496189	1	A2, A4	AY496212	h4, h6, h7, h30
G184	Cm / AD	AJ878469	22	A1, D1	n.s.	h22, h24, h29, h29, h32*, h37, h37
G17W3	Cm / B	AY496196	0	B	AY496210	h13, h15, h17, h19
G8W7	Cm / B	AY496195	0	B	AY496211	h14, h15, h16, h18, h20
G183K	Cm / C	AY496193	0	C	n.s.	h8, h33*, h33*, h35, h36
G188C	Cm / C	AY496192	0	C	n.s.	n.s.
G188D	Cm / C	AY496191	0	C	AY496207	h9, h10, h34, h34
G18W41	Cm / D	AY496194	1	D1, D2	AY496206	h11, h12, h12, h12, h38, h39, h39

Table 2. Overview of the three data sets obtained. Cm – *C. meneghiniana*, Amb – the "ambiguous" *C. scaldensis* morph. In the case of the ITS regions, insertion-deletion polymorphisms were excluded from counting variable positions; the other data sets contained no indels. * - the number of polymorphic positions within a strain in the partial 28S rDNA data set ranged from zero to two, with the exception of strain G184, where there were 22 of them. ** - number of variable positions in the ITS regions were below ten, with the exceptions of strains G184 (group A and D types, 80 variable positions), G183K (type C1 and C2, 135 variable positions) and G188D (type C1 and C2, 129 variable positions).

	28S	ITS region	18S
No. of strains sequenced	20	13	8
Length of sequenced region (bps)	613	733 – 748 (Cm) / 957 – 958 (Amb)	1799
No. of variants in Cm	8	40	6
No. of variants in Amb	4	6	2
Variable positions overall	70	256	16
Variable pos. within strains, Cm	0-2, 22*	0-10, 80-135**	0
Variable pos. in Cm	43	218	12
Variable pos. within strains, Amb	0-4	0-17	0
Variable pos. in Amb	5	17	0
Fixed substitutions (Cm vs. Amb)	25	43	4

Table 3. Results of an analysis of molecular variance AMOVA performed with ITS variants, grouped according to their 28S rDNA-based grouping (see Table 1, Fig. 1). Statistically significant (at $p < 0.001$) variance components are marked by asterisks. d.f., degrees of freedom

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation
Among groups	4	3476.147	81.53258	90.54*
Among strains within groups	7	68.023	0.36067	0.40
Within strains	44	358.990	8.15887	9.06*
Total	55	3903.161	90.05212	

2.4.8. Figures

Fig. 1. Number of pairwise nucleotide differences among the 28S rDNA variants. Table 1 lists in which cultures the individual variants were found.

	A1	A2	A3	A4	B	C	D1	D2	E1	E2	E3	E4
A1	-											
A2	2	-										
A3	3	1	-									
A4	3	1	2	-								
B	24	24	25	24	-							
C	18	18	19	18	25	-						
D1	22	22	23	22	26	16	-					
D2	23	23	24	23	27	17	1	-				
E1	36	36	37	36	42	39	43	44	-			
E2	37	37	38	37	43	40	44	45	1	-		
E3	36	36	37	36	42	39	43	44	3	4	-	
E4	37	37	38	37	42	40	44	45	4	5	1	-
	C. meneghiniana								"ambiguous" *C. scaldensis*			

Fig. 2. Phylogenetic trees of the partial 28S rDNA sequence variants (left) and of the groups of ITS types (right). The trees were calculated by bootstrapping the alignments in 1000 replicates and constructing phylogenetic trees by the neighbour joining method using maximum likelihood distances. Model parameters were estimated using Modeltest; for the 28S data set: base frequencies: A – 0.254, C – 0.199, G – 0.300; substitution rates: 0.43 (AC), 2.40 (AG), 2.50 (AT), 0.46 (CG), 10.30 (CT), 1.00 (GT); proportion of invariant sites: 0.702; and equal substitution rates among variant sites. For the ITS data set: equal base frequencies, substitution rates: 1.00 (AC), 3.34 (AG), 0.58 (AT), 0.58 (CG), 3.34 (CT), 1.00 (GT); proportion of invariant sites: 0.48, and no further among-site rate heterogeneity. All branches that were not recovered in at least 60 % of the bootstrap replicates were collapsed. Branch lengths on the topology thus obtained were estimated by maximum likelihood. Numbers above the branches are bootstrap values from the neighbour joining and from an unweighted maximum parsimony analyses. In the ITS tree, the group of variants from the "ambiguous" *C. scaldensis* morph (marked "amb") is connected by dashed lines to the two alternative positions where it was grouped in different analyses. The tree shown was calculated from the full-length alignment of the ingroup sequences; arrows in this tree mark positions of the root suggested by analyses of the data set with outgroups included, but unambiguously aligned positions excluded. The arrow marked "NJ" shows the root position suggested by neighbor joining analyses with ML distances; the one marked "MP" shows the root position suggested by parsimony analyses.

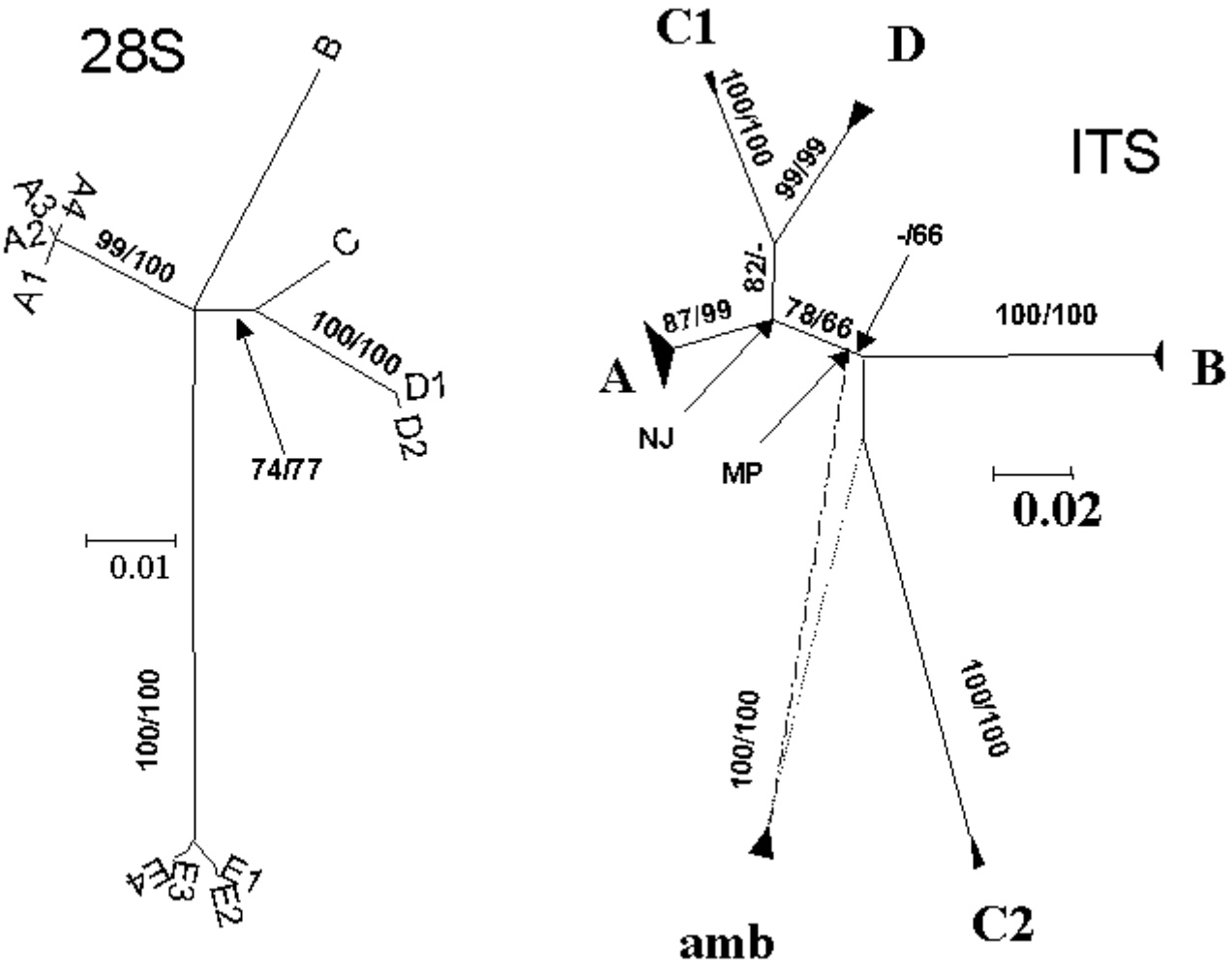


Fig. 3. Unrooted neighbor joining phylogenetic tree of the variants of the ITS regions using maximum likelihood distances (for model parameters, see the legend of Fig. 2). Clades with bootstrap values below 60 % were collapsed and branch lengths were estimated on the topology thus obtained by maximum likelihood. Terminal nodes are variants h0 to h45; after each variant name, the ITS clones belonging to them are listed. Clone names consist of the strain name, followed by a number identifying the clone among those sequenced from the same culture; see also Table 1. G184 (clone names in bold) contained variants from both ITS types A and D.

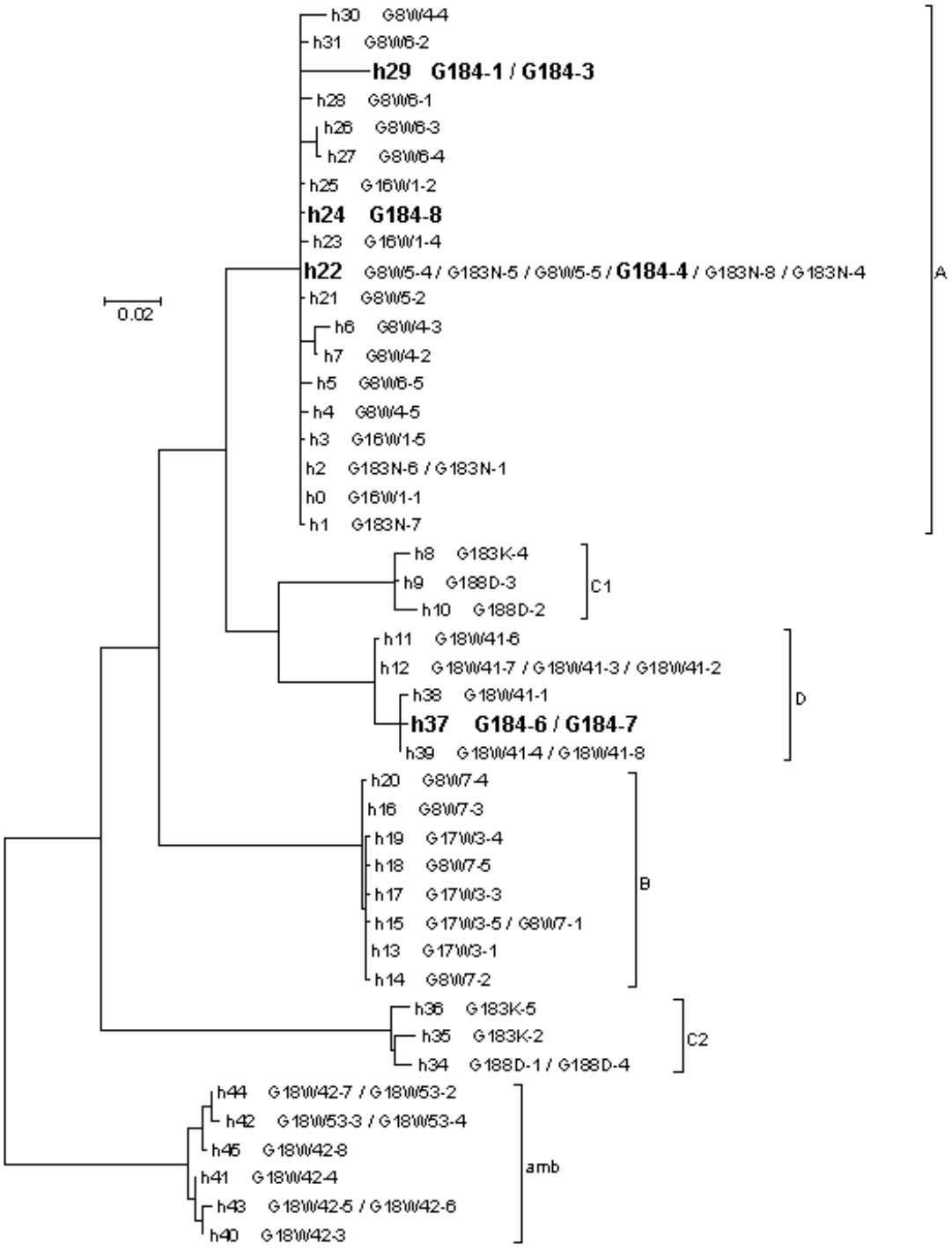


Fig. 4. Phylogenetic profiles of the ITS variants. 40 comparisons were made in the length of the alignment. The x-axis represents variable alignment positions, the y-axis the correlation coefficient between the vectors of pairwise distances calculated from the alignment segment preceding and following the respective position (phylogenetic correlations). The three most marked drops in the phylogenetic correlations are caused by two sequences, h32 and h33; further signals can be seen in the sequences from group C2.

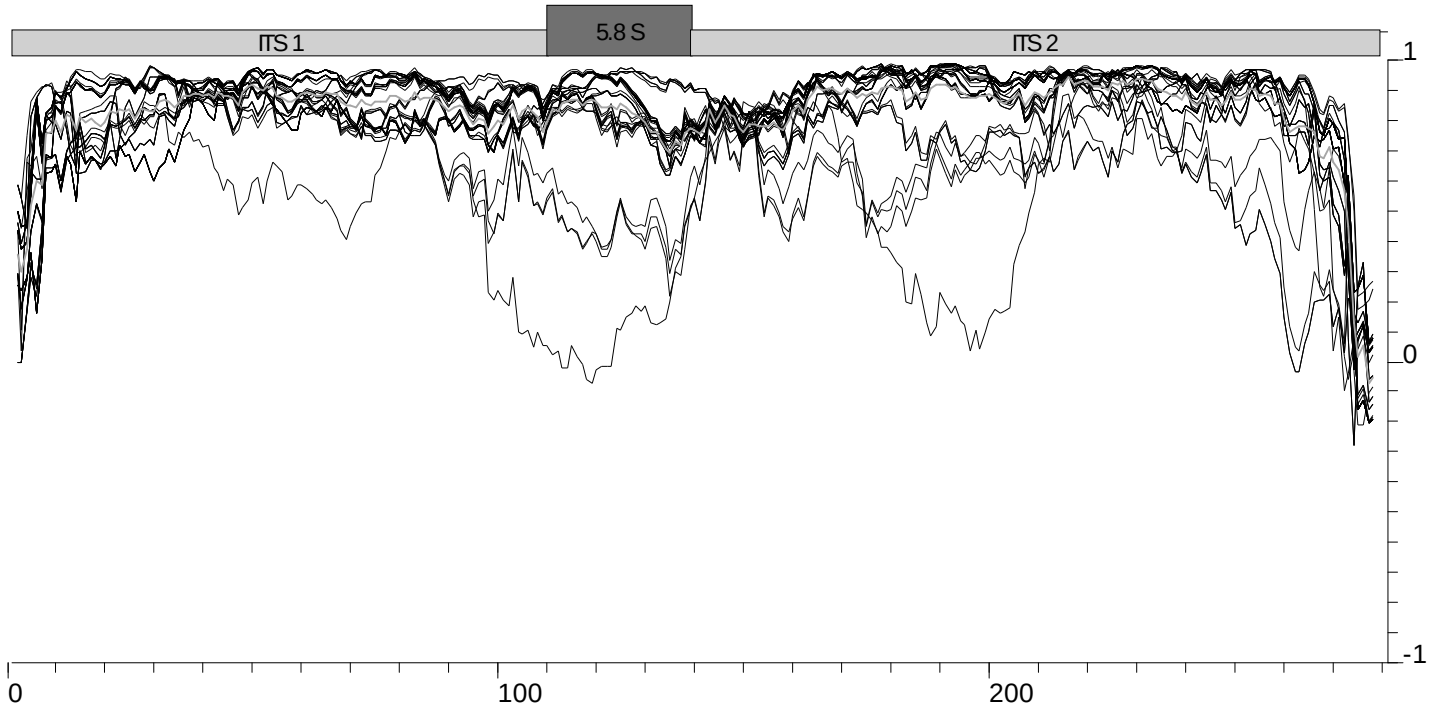


Fig. 5. A: Variable positions in the 18S rDNA sequences determined in this study. Fixed differences of the "ambiguous" *C. scaldensis* morph from the *C. meneghiniana* strains are highlighted in gray. The positions involved in the complementary base differences between the two morphs (Fig 6) are marked by asterisks. B: Pairwise number of nucleotide differences among the 18S rDNA sequences. Group: amb – "ambiguous" *C. scaldensis*, Cm – *C. meneghiniana*. For the *C. meneghiniana* strains, under "Group" the 28S rDNA group membership of the strains is also indicated (see Table 1).

A

		Position															
Strain	Morphotype	211*	215	222*	246	266	640	646	647	653	1025	1600	1678	1702	1705	1707	1768
G8W4	m	A	C	U	A	U	C	C	G	U	A	A	A	G	C	C	G
G17W8	m	A	C	U	A	U	C	U	G	U	A	A	A	G	C	C	G
G188D	m	A	C	U	A	C	U	C	A	U	A	A	G	G	C	C	G
G17W3	m	A	U	U	U	U	U	C	A	U	G	A	G	A	U	U	G
G8W7	m	A	U	U	U	U	U	C	A	U	G	A	G	A	U	U	U
G18W41	m	A	C	U	A	C	U	C	A	U	A	A	G	G	C	U	G
18W53	a	U	U	A	A	C	C	C	A	A	G	G	G	G	U	C	G
G1W11	a	U	U	A	A	C	C	C	A	A	G	G	G	G	U	C	G

B

Strain	Group	G8W4	G17W8	G17W3	G8W7	G188D	G18W41	G18W53	G1W11
G8W4	Cm / A	-							
G17W8	Cm / A	1	-						
G17W3	Cm / B	9	10	-					
G8W7	Cm / B	10	11	1	-				
G188D	Cm / C	4	5	7	8	-			
G18W41	Cm / D	5	6	6	7	1	-		
G18W53	amb	10	11	9	10	8	9	-	
G1W11	amb	10	11	9	10	8	9	0	-

Fig. 6. Secondary structure of loop E10-1 of the 18S rRNA of *C. meneghiniana*, with the arrows showing a pair of fixed complementary base changes between *C. meneghiniana* and the “ambiguous” morph.

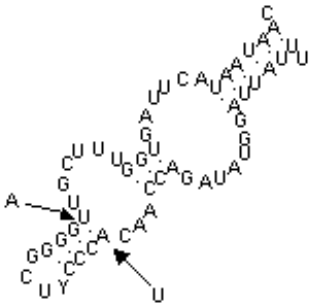
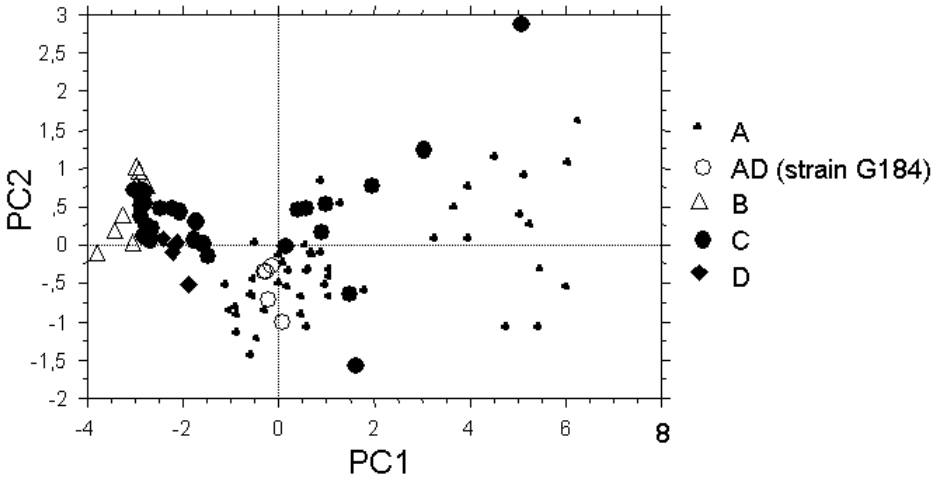


Fig. 7. Principal component analysis of eight morphometric variables measured on at least five frustules from each *C. meneghiniana* culture. All variables have large positive loadings on PC1, thus scatter of points on this axis mainly reflects size variation. The plot illustrates that the main morphological differences among the groups of strains with different rDNA genotypes were in their size distributions i.e., their largest separation is on PC1; the groups were not separated on any of the higher order ("size-independent") principal component axes.



2.4.9. References

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Mol Phylogenet Evol **29**: 417-434

Beszteri B, Ács É, and Medlin LK (2005) Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyceae) . Eur J Phycol **in press**

Bourne CEM (1992) Chloroplast DNA structure, variation and phylogeny in closely related species of *Cyclotella*. PhD Dissertation, University of Michigan

Büchel C (2003) Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states. Biochemistry **42**: 13027-34

Clark A (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. Mol Biol Evol **7**: 111-122

Coleman AW (2000) The significance of a coincidence between evolutionary landmarks found in mating affinity and a DNA sequence. Protist **151**: 1-9

De Bruin A, Ibelings BW, Rijkeboer M , Brehm M, van Donk E (2004/10/01) Genetic variation in *Asterionella formosa* (Bacillaryophyceae): is it linked to frequent epidemics of host-specific parasitic fungi? J Phycol **40**: 823-830

De Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. Proc Natl Acad Sci USA **96**: 2864-2868

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus **12:** 13-15

Drebes, G. (1977) Sexuality. In D Werner (ed) The biology of diatoms. Blackwell Scientific Publications, Oxford, pp 250-283

Edlund MB, Stoermer EF (1997) Ecological, evolutionary, and systematic significance of diatom life histories. J Phycol **33**: 897-918

El-Bestawy E, Bellinger EG, Sigee DC (1996) Elemental composition of

phytoplankton in a subtropical lake: X-ray microanalytical studies on the dominant algae *Spirulina platensis* (Cyanophyta) and *Cyclotella meneghiniana* (Bacillariophyceae). Eur J Phycol **31**: 157-166

Elwood H, Olsen G, Sogin M (1985) The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates Oxytricha nova and Stylonychia pustulata. Mol Biol Evol **2**: 399-410

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131**: 479-491

Finlay BJ, Monaghan EB, and Maberly SC (2002) Hypothesis: the rate and scale of dispersal of freshwater diatom species is a function of their global abundance. Protist **153**: 261-73

Håkansson H (2002) A compilation and evaluation of species in the genera *Stephanodiscus, Cyclostephanos* and *Cyclotella* with a new genus in the family Stephanodiscaceae. Diatom Res **17**: 1-139

Håkansson H, Chepurnov V (1999) A study of variation in valve morphology of the diatom *Cyclotella meneghiniana* in monoclonal cultures: effect of auxospore formation and different salinity conditions. Diatom Res **14**: 251-272

Iyengar MOP, Subrahmanyan R (1944) On reduction division and auxospore formation in *Cyclotella meneghiniana* Kütz. J Ind Bot Sci **23**: 125-152

Kumar S, Tamura K, Jakobsen I, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, Arizona, USA

Lewis RJ, Jensen SI, DeNicola DM, Miller VI, Hoagland KD, Ernst SG (1997) Genetic variation in the diatom *Fragilaria capucina* (Fragilariaceae) along a latitudinal gradient across North America. Pl Syst Evol **204**: 99-108

Li, W.-H. (1997) Concerted evolution of multigene families. Molecular Evolution. Sinauer Ass., Sunderland, MA., pp 309-334

Lohr M and Wilhelm C (2001) Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. Planta **212**: 382-91

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, and Ray SC (1999) Full-length human immunodeficiency virus

type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol **73**: 152-60

Louda JW, Liu L, Baker EW (2002) Senescence- and death-related alteration of chlorophylls and carotenoids in marine phytoplankton. Org Geochem **33**: 1635-1653

Lundholm N, Moestrup O, Hasle GR, Hoef-Emden K (2003) A study of the *Pseudo-nitzschia pseudodelicatissima*/*cuspidata* complex (Bacillariophyceae): What is *P. pseudodelicatissima*? J Phycol **39**: 797-813

Mann DG (1999) The species concept in diatoms. Phycologia **38** : 437-495

Maynard Smith J, Smith NH, O´Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci USA **90**: 4384-4388

Medlin LK (1997) Can molecular techniques help define species limits? Diatom **13**: 19-23

Medlin LK, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene **71:** 491-499

Medlin LK, Elwood HJ, Stickel S, Sogin ML (1991) Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyceae): Evidence for a new species, *Skeletonema pseudocostatum*. J Phycol **27**: 514-524

Medlin LK, Kooistra WHCF, Gersonde R , Wellbrock U (1996) Evolution of the diatoms (Bacillariophyta): III. Molecular evidence for the origin of the Thalassiosirales. Nova Hedw Beih **112**: 221-234

Meyer B, Wulf M, Håkansson H (2001) Phenotypic variation of life-cycle stages in clones of three similar *Cyclotella* species after induced auxospore production. Diatom Res **16**: 343-361

Millie DF, Hersh CM (1987) Statistical characterizations of the atrazine-induced photosynthetic inhibition of *Cyclotella meneghiniana* (Bacillariophyta). Aquat Toxicol **10**: 239-249

Montresor M, Sgrosso S, Procaccini G , Kooistra W (2003) Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): Evidence for cryptic species. Phycologia **42**: 56-70

Orsini L, Procaccini G, Sarno D, Montresor M (2004) Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative

abundances across a spring bloom in the Gulf of Naples. Mar Ecol Prog Ser **271**: 87-98

Posada D (1998) Modeltest 3.06. Department of Zoology, Brigham Young University, Provo, USA

Rao VNR (1970) Studies on *Cyclotella meneghiniana* Kütz. I. Sexual reproduction and auxospore formation. Proc Ind Acad Sci **72**: 285-287

Rao VNR (1971) Studies on *Cyclotella meneghiniana* Kütz. II. Induction of auxospore formation. Phykos **10**: 84-98

Rao VNR (1996) Size dependent reproductive behaviour in *Cyclotella meneghiniana* (Bacillariophyta). Nova Hedw Beih **112**: 235-238

Rosen BH, Lowe RL (1984) Physiological and ultrastructural responses of *Cyclotella meneghiniana* (Bacillariophyta) to light intensity and nutrient limitation. J Phycol **20**: 173-182

Sáez AG, Probert I, Geisen M, Quinn P, Young JR, and Medlin LK (2003) Pseudo-cryptic speciation in coccolithophores. Proc Natl Acad Sci USA **100**: 7163-7168

Salminen MO, Carr JK, Burke DS, and McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type 1 by bootscanning. AIDS Res and Hum Retroviruses **11**: 1423-1425

Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A (2005) Pseudocryptic diversity in the genus *Skeletonema* (Bacillariophyceae): *Skeletonema costatum* consits of several genetically and morphologically distinct species. J Phycol **in press**

Schoeman FR, Archibald REM (1980) The diatom flora of Southern Africa. Natl Inst Water Res, CSIR, Special Report 34 pp.

Scholin C, Herzog M, Sogin M, Anderson D (1994) Identification of Group- and Strain-Specific Genetic Markers for Globally Distributed *Alexandrium* (Dinophyceae). 2. Sequence Analysis of a Fragment of the LSU rRNA Gene. J Phycol **30**: 999-1011

Schultz ME (1971) Salinity-related polymorphism in the brackish-water diatom *Cyclotella cryptica*. Can J Bot **49**: 1285-1289

Schultz ME, Trainor FR (1968) Production of male gametes and auxospores in the centric diatoms *Cyclotella meneghiniana* and *C. cryptica*. J Phycol **4**: 85-88

Schultz ME, Trainor FR (1970) Production of male gametes and auxospores in a

polymorphic clone of the centric diatom *Cyclotella*. Can J Bot **48**: 947-951

Simon N, Brenner J, Edvardsen B, Medlin LK (1997) The identification of *Chrysochromulina* and *Prymnesium* species (Haptophyta, Prymnesiophyceae) using fluorescent or chemiluminescent oligonucleotide probes: a means for improving studies on toxic algae. Eur J Phycol **32**: 393-401

Simon N, Campbell L, Ornolfsdottir E , Groben R, Guillou L, Lange M, Medlin LK (2000) Oligonucleotide probes for the identification of three algal groups by dot blot and fluorescent whole-cell hybridization. J. Eukaryot Microbiol **47**: 76-84

Soudek Jr. D, Robinson GGC (1983) Electrophoretic analysis of species and population structure of the diatom *Asterionella formosa*. Can J Bot **61**: 418-433

Stoermer, E.F. and Smol, J.F. 1999. The diatoms. Applications for the environmental and earth sciences. Koeltz Scientific Books, Königstein, Germany.

Swofford, D. L.(1998) PAUP*. Phylogenetic analysis using Parsimony (* and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts

Thompson J, Gibson T, Plewniak F, Jeanmougin F, Higgins D (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl Acids Res **25**: 4876-4882

Tibayrenc M and Ayala FJ (1991) Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa. Parasitol Today **7**: 228-32

Tibayrenc M and Ayala FJ (2002) The clonal theory of parasitic protozoa: 12 years on. Trends Parasitol **18**: 405-10

Tibayrenc M, Kjellberg F, Arnaud J, Oury B, Breniere SF, Darde ML, and Ayala FJ (1991) Are eukaryotic microorganisms clonal or sexual? A population genetics vantage. Proc Natl Acad Sci USA **88**: 5129-33

Tuchman ML, Theriot E, Stoermer EF (1984) Effects of low level salinity concentrations on the growth of *Cyclotella meneghiniana* Kütz. (Bacillariophyta). Arch Protistenkd **128**: 319-326

Weiller GF (1998) Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. Mol Biol Evol **15**: 326-35

White, T.J., Bruns, T., Lee, S., and Taylor, J. (1990) Amplification and direct

sequencing of fungal ribosomal RNA genes for phylogenetics. In MA Innis, DH Gelfland, JJ Sninsky, TJ White (editors) PCR protocols. Guide to methods and application. Academic Press, San Diego, pp 315-322

Whitton, B.A. and Rott, E. 1996. Use of algae for monitoring rivers. Studia Student GmbH., Innsbruck, Austria.

Zechman FW, Zimmer EA, Theriot EC (1994) Use of ribosomal DNA internal transcribed spacers for phylogenetic studies in diatoms. J Phycol **30**: 507-512

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucl Acids Res **31**: 3406-3415

2.5. Publication III: Congruent variation at a nuclear and a plastid locus suggests that the diatom Cyclotella meneghiniana is a species complex

Bánk Beszteri, Uwe John, Linda K. Medlin

Alfred Wegener Institute for Polar and Marine Research, Am Handelshafen 12, 27570, Bremerhaven, Germany

2.5.1. Abstract

Our previous nuclear rDNA sequence analyses in a "population" of one of the most intensively studied freshwater diatom morphospecies, *Cyclotella meneghiniana*, revealed genetic variation incongruent with the expectation for a single panmictic population. The data suggested that this morphospecies probably consists of several, reproductively isolated sexual species, although strict clonality could also not be ruled out. Here, we extend this work in the following ways: (1) we screened a larger collection of strains, among them strains of *C. meneghiniana* with diverse geographic origins and strains of some closely related species; (2) we surveyed genetic variation in two different genomes of the strains: in the D1 / D2 regions of the nuclear encoded 28S rDNA and in a segment of the plastid encoded psaA gene; (3) to further test the hypotheses of cryptic sexual species vs. strict clonality, we performed AFLP analyses of a subsample of the strains belonging to one genotypic group and isolated from two closely situated localities. A strict association of nuclear rDNA and psaA variants was observed; the results suggested the presence of eight genetically distinct cryptic lineages within *C. meneghiniana*. The large phylogenetic incompatibility among loci in the AFLP data set contradicted the idea that sexual reproduction in this morphospecies would be so rare that it could not break down linkage among parts of the genome. Therefore, we suggest that *C. meneghiniana* is a complex of several, reproductively isolated, but morphologically (to date at least) indistinguishable taxa, cryptic species.

2.5.2. Introduction

In recent years, intraspecific comparative DNA sequencing studies have in many cases lead to discoveries of cryptic diversity (morphologically not distinguishable or at least not yet distinguished species) in different groups of eukaryotic microbes possessing inorganic shells (Scholin *et al.* 1994; De Vargas *et al.* 1999; Sáez *et al.* 2003). Diatoms (Bacillariophyceae) constitute the most morphologically feature rich group among microbial eukaryotes (Round *et al.* 1990). Still, surveys of nuclear ribosomal DNA sequence variation among multiple strains of a species have in several cases lead to such discoveries even in this group (Medlin *et al.* 1991; Manhart *et al.* 1995; Lundholm *et al.* 2003).

Cyclotella meneghiniana Kützing is a recent example of such cases (Beszteri *et al.* 2005a). It is one of the most extensively studied model species of freshwater diatoms (Giri & Devi 1992; Bourne *et al.* 1992; Lohr & Wilhelm 2001; Tedrow *et al.* 2002; Büchel 2003). However, it has been considered taxonomically problematic based on several facts: (1) its large frustule morphological variability (Schoeman & Archibald 1980; Hakansson & Chepurnov 1999); (2) the unusually wide range of habitat types among which it occurs

(Hakansson 2002); (3) the extreme morphological plasticity observed in cultures of the closely related species *C. cryptica* (frustules of this species are capable of displaying the morphology considered characteristic of *C. meneghiniana*, Schultz 1971); (4) the large (much greater than in closely related species) variation and the paraphyly of *C. meneghiniana* (with respect to *C. cryptica*) observed in a study of plastid genome RFLP variation (Bourne 1992). Furthermore, in our recent survey of rDNA sequence variation in a collection of strains from what was originally believed to be a single population of this species, large among-strain rDNA sequence variation was found, accompanied by much smaller within-strain variation (Beszteri *et al.* 2005a). Here, we are extending the results of this latter study in a number of ways to better understand the genetic diversity present in *C. meneghiniana*.

First, we sampled a larger number of cultures here, representing most cultures of this species available at public culture collections and further strains isolated at a number of geographically diverse localities. We also included strains of closely related species, including the above mentioned *C. cryptica*, strains of which species are capable of producing valve morphologies characteristic of *C. meneghiniana* (Schultz 1971).

Second, we sampled another, independent marker besides nuclear ribosomal DNA from these strains. The agreement of different sources of information is of crucial importance in studies of species limits (Mann 1999; Knowlton 2000). In diatom studies, rDNA sequence data are often accompanied by detailed morphological analyses, identifying fixed morphological characters between the groups revealed based on rDNA sequence data (Lundholm *et al.* 2003; Sarno *et al.* 2005). The number of strains included in rDNA surveys is, however, often relatively low and / or they cannot be considered morphologically representative of the whole natural variability and / or morphological or morphometric differences showing no overlap among the cryptic groups cannot be found (Beszteri *et al.* 2005a). In such cases, surveying genetic variation at unlinked loci can yield independent data sets useful for testing the distinctness of "candidate species". Thus, besides using a part of the rDNA operon, the variable D1 and D2 regions of the 28S rDNA, we extended our study to patterns of variation at the gene psaA, encoded in the plastid genome.

Third, the pattern observed in the above "population" study of *C. meneghiniana*, i.e. large among-strain differences compared to within-strain variation (Beszteri *et al.* 2005a), can also be explained by a complete or close to complete lack of allogamous sexual reproduction – i.e., strict selfing or clonality – besides the presence of cryptic species. The two phenomena – cryptic species and clonality – lead to similar deviations from patterns of genetic variation expected from a panmictic population, e.g., a relative lack of heterozygosity and linkage between physically unlinked loci (Tibayrenc & Ayala 2002). This possibility is rarely tested in molecular studies aimed at the clarification of species boundaries in diatoms, although such effective clonality has been observed even in protists earlier known as obligate sexual species (Ajzenberg *et al.* 2002). Diatoms show a wide range of life cycles (Chepurnov *et al.* 2004), and in some species, including *C. meneghiniana*, allogamous sexual reproduction has never been observed in spite of the numerous studies about its life cycle (Iyengar & Subrahmanyan 1944; Schultz & Trainor 1968; Rao 1970; Rao 1971; Drebes 1977; Rao 1996; Hakansson & Chepurnov 1999). Thus, the possibility that a lack of allogamous sexual reproduction is responsible for deviations of genetic variation patterns from panmictic expectations deserves experimental testing within this morphospecies. For this purpose, we generated amplified fragment length polymorphism (AFLP) fingerprints from a subset of strains, isolated from two, closely located populations and belonging to the same genotypic group based on their nuclear rDNA and psaA sequences. The hypothesis of strict clonality was tested using this data set, by exploring phylogenetic incompatibility among loci (Mes 1998).

2.5.3. Materials and methods

2.5.3.1. Cultures

Table 1. lists the cultures used and their origins. Cultures established and used in the course of this work were isolated and grown as described previously (Beszteri *et al.* 2005b). Cultures of *Stephanodiscus hantzschii* and *Discostella* strains were isolated and grown similarly, sequences were obtained from them for use as outgroups. *Discostella* has recently been separated from the genus *Cyclotella* (Houk & Klee 2004). We will refer to the group of cultures previously called "ambiguous" *C. scaldensis* morph (Beszteri *et al.* 2005a; Beszteri *et al.* 2005b) as *C.* cf. *scaldensis* in this publication. Based on valve morphological features, the three morphospecies *C. meneghiniana*, *C. quillensis* and *C. cryptica* will be referred to as *C. meneghiniana* complex. The other two *Cyclotella* morphospecies included, *C.* cf. *scaldensis* and *C. choctawatcheeana*, can be distinguished from them by their possession of closed alveolar chambers. Morphological data and images of the cultures used are available under www.algaterra.org.

2.5.3.2. DNA extraction, PCR, sequencing

Liquid cultures were harvested by filtration and DNA was extracted using a modified CTAB protocol (Doyle & Doyle 1990) or the PAN Plant Kit (PAN Biotech, Aidenach, Germany). The D1 / D2 regions of the nuclear 28S rDNA were amplified and sequenced using primers from Scholin *et al.* (1994). As a chloroplast genetic marker, a ~700 basepairs (bps) long fragment of the psaA gene was chosen based on previously screening the level of variation at a number of plastid loci, for the amplification of which PCR primers were available (tufA – Sáez et al 2003; psaA, psaB, psbA - Yoon 2002). PCR products were purified with the QIAQuick PCR Product Purification Kit (QIAGEN, Germany) and directly sequenced on both strands using Big Dye Terminator v3.1 sequencing chemistry (Applied Biosystems, CA, USA). Sequencing products were electrophoresed on an ABI 3100 Avant sequencer (Applied Biosystems, CA, USA).

2.5.3.3. Sequence assembly and analyses

Sequence assembly was done using SeqScape 2.1 (Applied Biosystems, CA, USA) as described previously (Beszteri *et al.* 2005a). "High confidence ambiguities" were identified based on relative heights of two peaks (0.3 or more) occurring in the same position in the nuclear rDNA electropherograms. Sequences containing such ambiguities were considered representing within-strain polymorphisms. Individual 28S rDNA variants producing the observed ambiguities were reconstructed from these sequences following Clark (1990).

Phylogenetic analyses were performed using PAUP* 4.0b10 (Swofford 1998). For maximum likelihood (ML) and distance based tree calculations, likelihood scores of different nucleotide substitution models were compared on a neighbour joining tree using Modeltest 3.0 (Posada 1998). The best fit model according to the Akaike Information Criterion (AIC) was used for phylogenetic analyses using ML and neighbour joining (NJ) tree inference with ML distances. Maximum parsimony (MP) and ML trees were obtained in heuristic searches,

with 10 random taxa addition sequences. Bootstrapping (BS) of MP and NJ analyses was made in 1000 replicates to assess confidence in clades recovered.

2.5.3.4. AFLP: strains, methods

AFLP was used to test the hypothesis of strict clonality within one of the apparently genetically distinct lineages within *C. meneghiniana* that were revealed by the DNA sequence data. 12 strains (marked AT-D… and AT-N…, see Table 1), all belonging to group A based on their partial 28S rDNA sequences (see Results and Beszteri *et al.* 2005a), were used for the AFLP genotyping. These strains were isolated from samples taken at two localities from the River Weser (Northern Germany) on the same day (21.07.2003). The two sampling sites are situated ~ 40 km from each other (near Nienburg, 52° 38,54 N; 9° 12,41 E, and Daverden, 52° 57,96 N; 9° 9,38 E). DNA extraction was performed using the PAN Plant Kit (PAN Biotech, Aidenach, Germany). DNA quality was verified by agarose gel electrophoresis; concentration was measured spectrophotometrically.

Digestion of genomic DNA, ligation, and preamplification was performed as described in John *et al.* (2004) starting from 250 ng genomic DNA. Selective amplifications were also performed as described (John *et al.* 2004), with the following modifications: four primer combinations were used (EcoRI + AC – MseI + CG; EcoRI + AC – MseI + CCT; EcoRI + AAG – MseI + CCT), with EcoRI-specific primers marked with the fluorescent dye 6-FAM. 1 µl of the product and 0.5 µl of GeneScan-500 ROX Size Standard (Applied Biosystems) internal DNA size marker was diluted to 10 µl with Hi-Di Formamide and run on an ABI 3100 Avant sequencer (all Applied Biosystems, CA, USA). AFLP bands were sized and manually scored using GeneMapper v3.5 (Applied Biosystems, CA, USA).

2.5.3.5. AFLP data analyses

The hypothesis of strict clonality of the *C. meneghiniana* complex was tested by compatibility analyses. Compatibility analyses were performed using the following programs from the package PICA 95 (Wilkinson 1995). The matrix conflict permutation tail probability test was performed using MATRIX.EXE. This test compares the incompatibility counts observed in the data set with those observed in 100 random permutations of the characters among strains. The proportion of cases having as low or lower incompatibility counts than the original data set is interpreted as the probability that the observed amount of incompatibility is comparable to that resulting from a random assignment of characters to terminal taxa, i.e., here, panmixia. Second, JACTAX.EXE was used to explore the possibility that one or a few strains are responsible for a large proportion of the incompatibilities present in the data set. Here, differences in incompatibility counts in the original data set are compared with those observed when excluding each terminal taxon from the counting in turn. The taxon contributing the most to overall incompatibility in the data is then excluded, and the analyses iterated, as long as the remaining data set contains incompatibilities. Thus, the possibility can be explored that the conflict in the data set is caused by one or a few individuals. In our case, this would mean that although the population is basically clonal, rare sexual events lead also to the presence of “hybrid” genotypes.

2.5.4. Results

2.5.4.1. Nuclear rDNA

Partial 28S rDNA sequences analysed included those obtained by Beszteri et al. (2005a) from the cultures isolated from a Geeste "population" (Genbank accession numbers AY496188-AY496195, AY496198-AY496205, AJ878468-AJ878470) and those sequenced in this study (GenBank accession numbers AJ878462-AJ878467, AJ878471-AJ878502). General information about this region is summarized in Table 2.

The D1/D2 regions of the 28S rDNA were invariably 613 basepairs long in strains from the *Cyclotella* species excluding amplification primer binding sites. The PCR products from the *Discostella* strains (*D. pseudostelligera* and *D. woltereckii*) were 605, that from *Stephanodiscus hantzschii* 603 bases long. From sequences containing mixed base calls, variants could be resolved unambiguously using a parsimony-based method (Clark 1990) with the exception of the following two *C. meneghiniana* strains (Table 1). The sequence obtained from strain CCMP 335 contained eight mixed base calls; it could in three different ways be decomposed into two alleles so that one of them was identical to an already unambiguously identified variant. The sequence from strain Cm7 contained 12 mixed bases; it could similarly be decomposed in two different ways. In both cases, the possible haplotypes differed in no more than two nucleotide positions from each other. For subsequent analyses, the most commonly found one of the already identified variants (this was in both cases variant D1) and the corresponding complementary variants, designated D4 and D5, were used.

Within-morphospecies variation in the partial 28S rDNA sequences was small in all morphospecies studied with the exception of *C. meneghiniana* (Table 2). In the three *C. quillensis* strains, as well as in the two *C. choctawatcheeana* strains, a single variant was observed. Polymorphisms were observed in all other morphospecies represented by more than one strain (Table 2). In the three strains designated *C. cryptica*, two different variants were found, one of which, together with a very similar variant, also was found in strain CCMP 336. (The latter strain was designated *C. meneghiniana* by CCMP, but, as discussed below, probably also belongs to *C. cryptica*.) Up to four nucleotide differences separated the three variants found in these four strains. In the five strains designated *C. scaldensis*, four different variants were found, with up to five nucleotide differences among them.

On the contrary, sequence variation in this region within *C. meneghiniana* was much larger. 21 different variants were observed, with 72 positions showing variation of the 613 (11.7 %) among the *C. meneghiniana* strains. The number of nucleotide differences between pairs of these variants ranged from 1 to 32.

Within-strain sequence variation was, however, much smaller also in *C. meneghiniana*. Nucleotide differences between variants found together in single strains of *C. meneghiniana* were less than three, with the following three exceptions (Fig. 1). In strain G184, as shown previously (Beszteri *et al.* 2005a), two divergent variants (designated A1 and D1, differing at 22 positions) were found. Second, the two above mentioned strains from group E, CCMP 335 and Cm7, contained combinations of alleles differing at 8 (variants D1 and D5) and 12 positions (variants D1 and D4), respectively.

The partial 28S rDNA variants obtained from *C. meneghiniana* could be grouped into eight distinct groups based on their similarities and co-occurrences in single strains (and leaving strain G184 out of consideration, Fig. 1). With the exception of group D, each of these groups displayed no more than four polymorphic nucleotides (Fig. 1, Table 2). The nucleotide differences among variants belonging to different groups ranged from 8 to 33. These groups also appeared as monophyletic in phylogenetic analyses (Fig 2) with strong bootstrap support (> 97 %); the only exception was again group D: the monophyly of this group had lower bootstrap support (63 % NJ, 70 % MP).

Sequences obtained from the other morphospecies of the *C. meneghiniana*-complex (*C. quillensis* and *C. cryptica*) did not differ more from sequences from *C. meneghiniana* than the latter among each other. This could be observed based on either simple counts of nucleotide differences (Fig. 1) as well as in phylogenetic analyses (Fig. 2).

The phylogenetic relationships among the above groups of rDNA variants remained largely unresolved (Fig. 2), only a few nodes had significant bootstrap support. Variants from the *Cyclotella* strains constituted a strongly supported clade, as well as those from the *C. meneghiniana* complex (BS: 100 in both MP and NJ). However, the monophyly of *C. meneghiniana* with the exclusion of *C. quillensis* and *C. cryptica* was not supported. In the single best trees recovered in NJ and ML analyses, *C. quillensis* and *C. cryptica* were always grouped among *C. meneghiniana* variants in optimal trees found, leaving the latter para- or polyphyletic. Also, in the strict consensus of the 92 most parsimonious trees obtained (length: 253 steps, consistency index: 0.727), the LSU sequence of the *C. quillensis* strains was sister to group A of *C. meneghiniana*. Thus, the single optimal trees support the paraphyly of the *C. meneghiniana* LSU alleles. However, bootstrap analyses did not further increase confidence in these results; most nodes within the *C. meneghiniana* complex were not supported by bootstrap values above 60 % (Fig. 2).

2.5.4.2. psaA

The length of the PCR products sequenced was 715 nucleotides in every cases. No mixed base calls were observed in any of the sequences. The number of variable positions was 133 positions in the whole dataset including outgroups, 96 within *Cyclotella*, 52 within the *C. meneghiniana* complex, and 45 within *C. meneghiniana*. Sixteen of the 238 amino acid positions were variable in the whole dataset, 14 within *Cyclotella*, 7 within the *C. meneghiniana* complex, and 6 within *C. meneghiniana*.

Patterns of variation within the morphospecies investigated were similar to the case of the nuclear LSU sequences, with most morphospecies showing small or no variation with the exception of *C. meneghiniana*. Within the latter, however, variation within groups identified based on LSU sequence variation was also minor (Table 2).

Results of phylogenetic analyses were similar to those obtained from the LSU rDNA data (Fig. 2): the monophyly of *Cyclotella* was strongly supported (BS 100 % in both MP and NJ), as was the monophyly of the *C. meneghiniana* complex (BS: 74 % NJ, 84 % MP), but the branching order of clades within the latter remained unresolved. In the strict consensus of the 80 most parsimonious trees obtained, the only clades recovered within the *C. meneghiniana* complex were the same as those shown in Fig. 2. Similarly to the nuclear rDNA data set, neither the mono-, nor the paraphyly of *C. meneghiniana* was supported by bootstrap values above 60 %.

Strict association between the nuclear rDNA and the psaA variants was observed. Within most groups identified based on the partial 28S rDNA sequence variation above, identical psaA sequences were obtained from each strain (Table 2). Exceptions were the *C. cryptica* strains (three different psaA types with four variable positions), and groups A (two psaA types, cm-I and cm-II, differing at three positions) and D (cm-V and cm-VI, with a single nucleotide difference) of *C. meneghiniana* (Tables 1 and 2).

2.5.4.3. AFLP

232 markers, 174 of them polymorphic, were scored for the 12 strains included in the AFLP analysis using three primer pairs. All 12 fingerprints were unique.

Although the two populations from which the 12 strains were isolated were so close geographically that no significant differentiation between them was expected, they turned out to have been highly differentiated. Only 61 (Daverden population, seven strains) and 79 (Nienburg population, five strains) of the 174 polymorphic markers showed polymorphism within the populations. Fifty-nine of the variable loci represented fixed differences between the two populations. The latter positions were by definition also compatible with each other. This explains that a matrix incompatibility permutation tail test showed highly significant (at $p < 0.01$) deviation from panmictic expectations for the combined data set.

However, the number of incompatible character pairs in the data set was substantial (2943), and none of the two populations deviated from panmictic expectations when analysed separately. The number of incompatible pairs was 97 (in the Daverden population, permutation tail probability: 0.89) and 217 (in the Nienburg population, permutation tail probability: 0.25), respectively. Thus, the two populations themselves showed patterns of genetic variation not distinguishable from that expected assuming random mating within the populations.

To test whether all strains contributed in comparable amount to the phylogenetic incompatibility present in the two populations, the contribution of individual strains was assessed by omitting them from the data set and comparing incompatibility counts of data sets with and without each of them. The strain that appeared to contribute the most to incompatibility present in the data was removed, and the test repeated until no further incompatibilities remained in the data set. The results showed that in both populations, the removal of the first strain substantially decreased the amount of phylogenetic incompatibility in the data (Fig. 3). However, also in both populations, a complete lack of incompatibilities could not be reached: some remained even when including only four strains. Four terminal taxa constitute the limit of testing phylogenetic incompatibility in a data set in the case of non-polarized characters (i.e., by definition, no incompatibilities can be present in a data set of three taxa).

2.5.5. Discussion

2.5.5.1. Cryptic diversity in *C. meneghiniana*

The data obtained in this study further reinforced the idea that the diatom morphospecies *C. meneghiniana* is not a genetically homogeneous / irreducible taxon. The following facts strongly indicate the presence of genetic structure within *C. meneghiniana*:

(1) within-species sequence variation substantially exceeded average within-strain variation in the nuclear rDNA segment sequenced. Groups of rDNA variants were clearly distinct (Table 2, Fig. 2). Even group D, showing the most variation, was monophyletic in phylogenetic analyses (Fig. 2). Whereas within-strain variation was encountered in the nuclear rDNA sequences, variants belonging to different groups never occurred together besides the single exception of strain G184, discussed previously (Beszteri *et al.* 2005a).

(2) groups of strains defined based on their nuclear rDNA genotypes had identical or close to identical psaA sequences, whereas psaA sequence differences between rDNA groups were always larger (Table 2).

Sequence variation at both loci within the *C. meneghiniana* groups was comparable to variation in all other morphospecies studied here (Table 2). In the case of the D1/D2 regions

of the 28S rDNA, they were also comparable to amounts of intraspecific variation observed in other algal groups (John *et al.* 2003; Sarno *et al.* 2005).

The observed genetic structure within *C. meneghiniana* cannot be explained by geographic isolation. Strains belonging to different genotypic groups occurred in sympatry in several cases (Table 1). Strains from two localities, from which more strains were sampled (Geeste Estuary, Bremerhaven, and River Weser by Hasenbüren, see Table 2 and Beszteri *et al.* 2005a) showed a large genetic diversity; strains from other localities (River Weser by Nienburg and Daverden) all belonged to the same genotypic group (group A in both cases).

The sequence variants found in *C. meneghiniana* did not form a monophyletic group in the case of either data set. Whereas the *C. meneghiniana* complex (including *C. cryptica* and *C. quillensis* besides *C. meneghiniana*) was monophyletic with strong bootstrap support in all analyses, the monophyly of *C. meneghiniana* was not supported by bootstrap values > 60 %. Both *C. cryptica* and *C. quillensis* were grouped well among clades consisting of variants from *C. meneghiniana* in best trees recovered using different optimality criteria, leaving the latter morphospecies para- or polyphyletic.

2.5.5.2. Clonality or cryptic species?

The above results cannot answer the question whether the variation pattern observed can be explained by a complete lack of sexual reproduction in *C. meneghiniana*, or by the presence of multiple sexual species within it. Although patterns of ribosomal DNA variation might also be of some limited use for answering this question (Beszteri *et al.* 2005a), they are not ideally suited for this purpose because of their low amount of variation within the groups revealed. AFLP fingerprints have two important advantages in this context: they show more variation, and they provide information about genetic variation at multiple loci. Using multilocus genotypes, linkage among loci can be explored. In the case of strict clonality or selfing, genome-wide linkage is expected among loci, whereas if sexual reproduction occurs, even if occasionally, it is expected to break down this linkage. In the case of the dominant AFLPs, allele frequency based tests of linkage disequilibrium are not appropriate, and our low sample size does not permit the use of such tests either. However, tests based on phylogenetic compatibility among loci are not so sensitive to these factors, thus they allowed us testing the hypothesis of clonality in a small subsample.

The AFLP data showed high amounts of incompatibility among loci within both populations sampled for this test. The data strongly suggest that at least occasionally, allogamous sexual reproduction occurred in the past of these populations, leading to a reshuffling of genetic variation among lineages. Our small sample size does not allow us to draw a similarly strong conclusion about the frequency of such sexual events. As illustrated by Fig. 3, one of the strains contributed to a disproportionately large extent to the phylogenetic incompatibility present in the data in both populations. However, to what extent this deviates from patterns expected supposing frequent allogamous events is unclear. A more thorough population sampling might be able to answer this question in the future.

2.5.5.3. Taxonomic conclusions: *C. cryptica* and *C. meneghiniana*

As shown previously, quantitative morphological features classically used for species identification in this diatom group were unable to differentiate a subset of the genetically distinct groups revealed within *C. meneghiniana* (Beszteri *et al.* 2005a). Here we have not even attempted a morphometric comparison because of the large number of factors

complicating the interpretation of morphological variation in this group. The most important of these factors is the possibility of valve morphological plasticity. As noted in the Introduction, one of the morphospecies included in our study, *C. cryptica* is known to be able to produce valve morphology typical of *C. meneghiniana* when grown at low salinities and after auxospore production (Schultz 1971). Later, also *C. meneghiniana* cultures capable of producing the *C. cryptica* morphology were reported (Desikachary & Rao 1973; Hoops & Floyd 1979), and smaller morphological changes in *C. meneghiniana* cultures caused by salinity changes have also been shown (Hoops & Floyd 1979; Tuchman *et al.* 1984). The frames of this work did not allow for an extensive study of the effects salinity on valve morphologies. Clearly, any trial aimed at identifying fixed morphological differences between the to date cryptic taxa apparently existing within the *C. meneghiniana* complex will need to consider this issue.

Our results indicated that strains identified as *C. cryptica* were genetically homogenous and different from *C. meneghiniana*. (We also grouped strain CCMP 336 into *C. cryptica*, based on its nuclear rDNA and psaA sequence, although according to CCMP it is a *C. meneghiniana* strain – as discussed above, the morphological boundary between these species is quite unclear.) Although their rDNA and psaA variants were not more different from other *C. meneghiniana* strains than the latter among each other, they formed one of the distinct groups recovered within the *C. meneghiniana* complex in both data sets (Fig. 2, Table 2). A comparable pattern was observed by Bourne (1992) in her plastid DNA RFLP study: a clade of three *C. cryptica* strains grouped well within the much deeper clade including strains identified as *C. meneghiniana* besides *C. cryptica*.

The situation with *C. quillensis*, the morphological differences of which species from *C. meneghiniana* are also minor (Battarbee & Keister 1982; Hakansson & Kling 1994), was similar in our results. It showed no variation in either region sequenced, was genetically distinct from all other groups, but was part of the well supported clade including *C. meneghiniana* and *C. cryptica* as well. These results suggest that the *C. meneghiniana* complex might include a large number of species, some of which can be distinguished from others, perhaps even other described species. This might also imply that future morphology based taxonomic studies in this group will need to test the above discussed problematic effects of salinity upon morphology also when dealing with other, morphologically similar species like *C. quillensis*.

2.5.5.4. Population differentiation on small geographic scale or even more cryptic species?

One of the most unexpected results of this study was the substantial differentiation observed between the two subsamples of strains used in the AFLP analyses. The samples, from which these cultures were isolated, were taken from the same river, at sites located about 40 kms from each other, and on the same day, with a difference of about two hours. Although sample sizes were very small, the genetic differences between strains from the two localities were enormous. Of the 174 polymorphic loci, 59 showed a fixed difference between the two groups of strains, and further 64 loci were compatible with this grouping. Only 29 parsimony informative loci were incompatible with the split separating the two populations.

The small distance and the hydrological connection between the sampling sites excludes the possibility of simple isolation-by distance, the large genetic difference between the two populations is clearly a result of different selective pressures acting on them at the two localities. In the case of strict clonality, the effect of divergent selection could be expected to result in similar genome-wide population differentiation: selection acting upon any locus would decrease within-population variation also at all other loci in a population,

because the whole genome is in linkage with the locus under selection. However, as discussed above, the amount of phylogenetic incompatibilities within these two populations strongly suggested that allogamous sexual reproduction has occurred so frequently in their past that it could break down genome-wide linkage among loci, making this explanation implausible.

Therefore, the question that arises whether the two populations really were two, highly differentiated populations of a single species, or did they in fact represent two, reproductively isolated taxa. We think two facts seem to favour the second hypothesis. The first is simply the amount of fixed genetic differences between these populations seen in the AFLP data, accompanied by the strong indications of the presence of sexuality, as discussed in the last paragraph. The second is that patterns of variation at both loci sequenced are congruent with this idea: psaA variants occurring in strains from the two populations also showed three fixed mutational differences (Table 1). Their 28S rDNA variants were also different, although variant A2 occurred in both groups – but this might also represent an ancient polymorphism retained from the common ancestor of the two species.

The above dilemma seems to be a recurrent question posed at the end of research involving population genetic data collection from diatoms. The first detailed population genetic study of a diatom species encountered a yearly alternation of highly differentiated populations in *Skeletonema costatum* using multilocus enzyme electrophoresis (Gallagher 1980). We know since that this morphospecies is in fact a complex of several genetically distinct entities (Sarno *et al.* 2005). However, it is still unclear whether the "summer" and "winter" populations detected by Gallagher were representatives of one or more of these species. Another study, performed with the "taxonomically unproblematic" species *Dytilum brightwellii* using microsatellite markers, detected genetically distinct populations in different parts of a single estuarine system (Rynearson & Armbrust 2004). It remained unclear also in this case whether the substantial differentiation was caused by the differential growth of distinct species (reproductively isolated groups), or whether rather different subsets of the genetic variation present in a single species were favoured in the different environments. Based on the reoccurrence of this question in several of the few population genetic investigations that have been made with diatoms to date, one can predict that diatomists will be confronted with this problem in the future in several cases, with the widening application of high resolution molecular markers in this group.

We think this question deserves further attention because of at least two reasons besides its potential frequency. The first is that the alternatives of cryptic diversity vs. strong ecotypic differentiation represent our current limit of understanding of "around-the-species-level" genetic variation in diatoms. The second is that these two hypotheses have different implications for approaches trying to make use of diatom microdiversity for refining water quality monitoring: if the incidence of species, that can only be differentiated using high resolution molecular markers, is much higher in diatoms than expected, then diatomists simply face the technical challenge of using such markers more widely than before to recognize species. However, if strong intraspecific ecotypic differentiation is a characteristic feature of diatoms, then microdiversity research also needs to take into account the effects of gene flow among such ecotypes, making their identification also conceptually more challenging.

2.5.6. Acknowledgements

Thanks to E. Hegewald, É. Ács, E. Theriot & A. Alverson (U.S. National Science Foundation grant DEB 0111883) for providing cultures, to Katrin Hamann for the water chemistry measurements, and Alberto Garcia-Sáez for discussions. This work was in part

supported by the project “Algaterra” of the German Federal Ministry of Education and Research (project ID BMBF 01LC0026, http://www.algaterra.org).

2.5.7. References

Ajzenberg D, Banuls AL, Tibayrenc M, Darde ML (2002) Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. *International Journal of Parasitology*, **32**, 27-38.

Battarbee RW, Keister CM (1982) The frustular morphology and taxonomic relationships of *Cyclotella quillensis* Bailey. *7th Diatom Symposium*, 173-184.

Beszteri B, Ács É, Medlin LK (2005a) Ribosomal DNA sequence variation within and among clonal strains of the *Cyclotella meneghiniana* complex (Bacillariophyceae) from an estuarine locality. *Protist*, **submitted**

Beszteri B, Ács É, Medlin LK (2005b) Conventional and geometric morphometric studies of valve ultrastructural variation in two closely related *Cyclotella* species (Bacillariophyceae). *European Journal of Phycology*, **40**, 73-88.

Bourne CEM (1992) Chloroplast DNA structure, variation and phylogeny in closely related species of *Cyclotella*. PhD Dissertation, University of Michigan

Bourne CEM, Palmer JD, Stoermer EF (1992) Organization of the chloroplast genome of the freshwater centric diatom *Cyclotella meneghiniana*. *Journal of Phycology*, **28**, 347-355.

Büchel C (2003) Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states. *Biochemistry*, **42**, 13027-34.

Chepurnov VA, Mann DG, Sabbe K, Vyverman W (2004) Experimental studies on sexual seproduction in diatoms. *International Review of Cytology*, **237**, 91-154.

Clark A (1990) Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution*, **7**, 111-122.

De Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J (1999) Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences USA*, **96**, 2864-2868.

Desikachary TV, Rao VNR (1973) Studies on *Cyclotella meneghiniana* Kütz. III. The frustule. *Proceedings of the Indian Academy of Sciences*, **77B**, 78-91.

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13-15.

Drebes, G. (1977) Sexuality. In D Werner (ed) The biology of diatoms. Blackwell Scientific Publications, Oxford, pp 250-283

Gallagher JC (1980) Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. *Journal of Phycology*, **16**, 464-474.

Giri BS, Devi NA (1992) Inhibition of cytokinesis by lead (Pb2+) in a centric diatom. *Indian Journal of Experimental Biology*, **30**, 201-4.

Hakansson H (2002) A compilation and evaluation of species in the genera *Stephanodiscus*, *Cyclostephanos* and *Cyclotella* with a new genus in the family Stephanodiscaceae. *Diatom Research*, **17**, 1-139.

Hakansson H, Chepurnov V (1999) A study of variation in valve morphology of the diatom *Cyclotella meneghiniana* in monoclonal cultures: effect of auxospore formation and different salinity conditions. *Diatom Research*, **14**, 251-272.

Hakansson H, Kling H (1994) *Cyclotella agassizensis* nov. sp. and its relationship to *C. quillensis* Bailey and other prairie *Cyclotella* species. *Diatom Research*, **9**, 289-301.

Hoops HJ, Floyd GL (1979) Ultrastructure of the centric diatom *Cyclotella meneghiniana*: vegetative cell and auxospore development. *Phycologia*, **18**, 424-435.

Houk V, Klee R (2004) The stelligeroid taxa of the genus *Cyclotella* (Kützing) Brébisson (Bacillariophyceae) and their transfer into the new genus *Discostella* gen. nov. *Diatom Research*, **19**, 203-228.

Iyengar MOP, Subrahmanyan R (1944) On reduction division and auxospore formation in *Cyclotella meneghiniana* Kütz. *The Journal of the Indian Botanical Society*, **23**, 125-152.

John U, Groben R, Beszteri B, Medlin L (2004) Utility of Amplified Fragment Length Polymorphisms (AFLP) to analyse genetic structures within the *Alexandrium tamarense* species complex. *Protist*, **155**, 169-79.

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia*, **420**, 73-90.

Lohr M, Wilhelm C (2001) Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. *Planta*, **212**, 382-91.

Lundholm N, Moestrup O, Hasle GR, Hoef-Emden K (2003) A study of the *Pseudo-nitzschia pseudodelicatissima*/*cuspidata* complex (Bacillariophyceae): What is *P. pseudodelicatissima*? *Journal of Phycology*, **39**, 797-813.

Manhart JR, Fryxell GA, Villac C, Segura LY (1995) *Pseudo-nitzschia pungens* and *P. multiseries* (Bacillariophyceae): Nuclear ribosomal DNAs and species differences. *Journal of Phycology*, **31**, 421-427.

Mann DG (1999) The species concept in diatoms. *Phycologia*, **38**, 437-495.

Medlin LK, Elwood HJ, Stickel S, Sogin ML (1991) Morphological and genetic variation within the diatom *Skeletonema costatum* (Bacillariophyceae): Evidence for a new species, *Skeletonema pseudocostatum*. *Journal of Phycology*, **27**, 514-524.

Mes THM (1998) Character compatibility of molecular markers to distinguish asexual and sexual reproduction. *Molecular Ecology*, **7**, 1719-1727.

Posada, D (1998) Modeltest 3.06. Department of Zoology, Brigham Young University, Provo, USA

Rao VNR (1970) Studies on *Cyclotella meneghiniana* Kütz. I. Sexual reproduction and auxospore formation. *Proceedings of the Indian Academy of Sciences*, **72**, 285-287.

Rao VNR (1971) Studies on *Cyclotella meneghiniana* Kütz. II. Induction of auxospore formation. *Phykos*, **10**, 84-98.

Rao VNR (1996) Size dependent reproductive behaviour in *Cyclotella meneghiniana* (Bacillariophyta). *Beiheft Zur Nova Hedwigia*, **112**, 235-238.

Round FE, Crawford RM, Mann DG (1990) The diatoms: Biology and morphology of the genera. 747 pp.

Rynearson TA, Armbrust EV (2004) Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology*, **40**, 34-43.

Sarno D, Kooistra WHCF, Medlin LK, Percopo I, Zingone A (2005) Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *Journal of Phycology*, **in press**, DOI: 10.1111/j.1529-8817.2005.04067.x.

Schoeman FR, Archibald REM (1980) The diatom flora of Southern Africa. *National Institute for Water Research, Council for Scientific and Industrial Research, Special Report*, 34 pp.

Scholin C, Herzog M, Sogin M, Anderson D (1994) Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). 2. Sequence analysis of a aragment of the LSU rRNA gene. *Journal of Phycology*, **30**, 999-1011.

Schultz ME (1971) Salinity-related polymorphism in the brackish-water diatom *Cyclotella cryptica*. *Canadian Journal of Botany*, **49**, 1285-1289.

Schultz ME, Trainor FR (1968) Production of male gametes and auxospores in the centric diatoms *Cyclotella meneghiniana* and *C. cryptica*. *Journal of Phycology*, **4**, 85-88.

Swofford, DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (* and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Massachusetts

Sáez AG, Probert I, Geisen M, Quinn P, Young JR, Medlin LK (2003) Pseudo-cryptic speciation in coccolithophores. *Proceedings of the National Academy of Sciences USA*, **100**, 7163-7168.

Tedrow O, Julius ML, Schoenfuss HL (2002) The Effects of Biogenically Active Compounds on *Cyclotella meneghiniana* (Bacillariophyta). *Journal of Phycology*, **38, S1**, 34-35.

Tibayrenc M, Ayala FJ (2002) The clonal theory of parasitic protozoa: 12 years on. *Trends in Parasitology*, **18**, 405-10.

Tuchman ML, Theriot E, Stoermer EF (1984) Effects of low level salinity concentrations on the growth of *Cyclotella meneghiniana* Kütz. (Bacillariophyta). *Archiv Für Protistenkunde*, **128**, 319-326.

Wilkinson, M (1995) PICA 95: Software and Documentation. School of Biological Sciences, University of Bristol, Bristol, UK

Yoon HS, Hackett JD, Bhattacharya D (2002) A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proceedings of the National Academy of Sciences USA*, **99**, 11724-11729.

2.5.8. Tables

Table 1. List of strains used, their origins, and LSU rDNA and psaA haplotypes. In the case of the 28S rDNA haplotypes, the presence of one allele is indicated when no ambiguous positions were observed; in the presence of ambiguous positions, haplotypes were reconstructed assuming the presence of two different alleles, these are listed separated by slashes. Origin: FDCC – Loras College Freshwater Diatom Collection, CCMP – Provasoli-Guillard National Center for the Culture of Marine Phytoplankton, AT – isolated during this work, within the frame of the project Algaterra. * - strain CCMP336 is listed as *C. meneghiniana* by the culture collection; based on its molecular genetic affinities to *C. cryptica* strains, we consider it as belonging to *C. cryptica*. ** - sequences obtained previsouly (Beszteri *et al.* 2005a). *** - 28S variants were resolved with ambiguity.

strain name	species	origin	origin	28S haplotypes	psaA
L1840	*C. choctawatcheeana*	Salton Sea, CA, USA	FDCC	Cchoct	choct
L1844	*C. choctawatcheeana*	Salton Sea, CA, USA	FDCC	Cchoct	choct
CCMP 331	*C. cryptica*	unknown	CCMP	Cry/Cry2	ccr-I
CCMP 332	*C. cryptica*	Martha´s Vineyard, MA, USA	CCMP	Cry	ccr-II
CCMP 333	*C. cryptica*	Martha´s Vineyard, MA, USA	CCMP	Cry	ccr-II
CCMP 336	*C. cryptica* *	Myakka River, FL, USA	CCMP	Cry/Cry3	ccr-III
G16W1	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1**	cm-I
G17W8	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1**	cm-I
G182	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1**	cm-I
G183N	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1**	cm-I

G8W4	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A2, A4**	cm-II
G8W5	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1, A2**	cm-I
G8W6	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A2, A3**	n.s
G8W9	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A2**	cm-II
AT-D2	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D3	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D4	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D5	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D8	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D12	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-D13	*C. meneghiniana*	River Weser by Daverden, Germany	AT	A2, A5	cm-I
AT-N22	*C. meneghiniana*	River Weser by Nienburg, Germany	AT	A3, A4	cm-II
AT-N25	*C. meneghiniana*	River Weser by Nienburg, Germany	AT	A2	cm-II
AT-N30	*C. meneghiniana*	River Weser by Nienburg, Germany	AT	A3, A4	cm-II
AT-N32	*C. meneghiniana*	River Weser by Nienburg, Germany	AT	A3, A4	cm-II
AT-N34	*C. meneghiniana*	River Weser by Nienburg, Germany	AT	A3, A4	cm-II

G184	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	A1, D1**	cm-I
G17W3	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	B**	cm-III
G8W7	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	B**	cm-III
G183K	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	C**	cm-IV
G188C	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	C**	cm-IV
G188D	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	C**	cm-IV
AT-cscal	*C. meneghiniana*	River Moselle, Luxemburg	É. Ács	C	cm-IV
AT-51.12	*C. meneghiniana*	River Ems by Papenburg, Germany	AT	D1	cm-V
AT-67.5	*C. meneghiniana*	River Geeste by Bramel, Germany	AT	D1	cm-V
G18W41	*C. meneghiniana*	Geeste estuary, Bremerhaven, Germany	AT	D1, D2**	cm-V
AT-cm4	*C. meneghiniana*	Köttinger See, Germany	E. Hegewald	D1, D3	cm-V
AT-cm7	*C. meneghiniana*	Köttinger See, Germany	E. Hegewald	D1, D4***	cm-V
CCMP335	*C. meneghiniana*	Loch Leven, Scotland, UK	CCMP	D1, D5***	cm-VI
AT-10.4.0	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	F	cm-VII
AT-10.5.5	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	F	cm-VII
CCMP 334	*C. meneghiniana*	Lake Ohrid, Macedonia	CCMP	F	cm-VII

AT-10.4.1	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	G1	cm-VIII
AT-10.5.2	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	G1, G2	cm-VIII
AT-10.5.2a	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	G1, G2	cm-VIII
AT-10.5.0	*C. meneghiniana*	River Weser by Hasenbüren, Germany	AT	G3, G4	cm-VIII
AT-cm3	*C. meneghiniana*	Hanoi, Vietnam	E. Hegewald	H1, H2	cm-IX
Bpp02-9	*C. meneghiniana*	Bayou Petite Prairie, LA, USA	E. Theriot	I1, I2	cm-X
L776	*C. quillensis*	Devils Lake, ND, USA	FDCC	Quil	quil
L829	*C. quillensis*	Devils Lake, ND, USA	FDCC	Quil	quil
L835	*C. quillensis*	Devils Lake, ND, USA	FDCC	Quil	quil
G18W53	*C.* cf. *scaldensis*	Geeste estuary, Bremerhaven, Germany	AT	E1**	scal
G8W8	*C.* cf. *scaldensis*	Geeste estuary, Bremerhaven, Germany	AT	E1, E2**	n.s
G18W42	*C.* cf. *scaldensis*	Geeste estuary, Bremerhaven, Germany	AT	E1, E3**	scal
G18W44	*C.* cf. *scaldensis*	Geeste estuary, Bremerhaven, Germany	AT	E1, E3**	n.s
G1W11	*C.* cf. *scaldensis*	Geeste estuary, Bremerhaven, Germany	AT	E1, E4**	scal
AT-10.1.1	*D. pseudostelligera*	River Weser by Hasenbüren, Germany	AT	Cpseu1	pseu-I
L435	*D. woltereckii*	Montezuma Well Natl. Mon., AZ, USA	FDCC	Cpseu1	pseu-I

L434	*D. pseudostelligera f. parva*	Montezuma Well Natl. Mon., AZ, USA	FDCC	Cpseu1	pseu-II
A14	*D. pseudostelligera*	Borrow Pit, west Ames, IA, USA	FDCC	Cpseu2	pseu-III
AT-N2	*Stephanodiscus hantzschii*	River Weser by Nienburg, Germany	AT	Shan	shan

Table 2. Summary of information on genetic variation observed within the rDNA groups of *C. meneghiniana* and in the other *Cyclotella* morphospecies included in our study. Nr. of strains sequenced: number of strains for which partial LSU rDNA sequences were obtained, followed by the number of strains for which the psaA fragment was sequenced. Number of variants: the number of unique variants found in the respective groups. Min nr. of differences: the smallest number of nucleotide differences observed between a member of the corresponding group and any other strain. * - *C. cryptica*, including strain CCMP 336.

			LSU			psaA		
	nr. of strains sequenced	number of variants	variable positions	min. nr. of differences	number of alleles	variable positions	min. nr. of differences	
C. meneghiniana, group A	20 / 19	5	5	13	2	3	11	
C. meneghiniana, group B	2 / 2	1	-	8	1	-	4	
C. meneghiniana, group C	4 / 4	1	-	10	1	-	2	
C. meneghiniana, group D	6 / 6	5	14	9	2	1	2	
C. meneghiniana, group F	3 / 3	1	-	13	1	-	10	
C. meneghiniana, group G	4 / 4	4	2	8	1	-	4	
C. meneghiniana, group H	1 / 1	2	1	9	1	-	8	
C. meneghiniana, group I	1 / 1	2	1	22	1	-	11	
C. cryptica *	4 / 4	3	5	13	3	4	13	
C. quillensis	3 / 3	1	-	16	1	-	9	
C. choctawatcheeana	2 / 2	1	-	27	1	-	45	
C. cf. *scaldensis*	5 / 3	4	5	31	1	-	39	

Table 3. Summary of the AFLP data. The informative markers were grouped according to their compatibility with the split separating the two populations.

	No. of loci
monomorphic	58
uninformative	22
fixed differences	59
informative, compatible	64
informative, incompatible	29
altogether	232 loci

2.5.9. Figures

Fig. 1. Matrix of pairwise nucleotide differences (lower left half) and Jukes-Cantor distances (upper right half) of the nuclear LSU rDNA variants. Gray background marks that the corresponding variants were found co-occuring in a single strain.

	A1	A2	A3	A4	A5	B	C	D1	D2	D3	D4	D5	F	G1	G2	G3	G4	H1	H2	I1	I2	CRY	CRY2	CRY3	QUIL
A1	-	0.003	0.005	0.005	0.005	0.040	0.030	0.037	0.038	0.035	0.023	0.033	0.033	0.045	0.044	0.047	0.045	0.023	0.022	0.047	0.045	0.030	0.032	0.028	0.032
A2	2	-	0.002	0.002	0.002	0.040	0.030	0.037	0.038	0.035	0.023	0.033	0.033	0.045	0.044	0.047	0.045	0.023	0.022	0.047	0.045	0.030	0.032	0.028	0.032
A3	3	1	-	0.003	0.003	0.042	0.032	0.038	0.040	0.037	0.025	0.035	0.035	0.047	0.045	0.049	0.047	0.025	0.023	0.049	0.047	0.032	0.033	0.030	0.033
A4	3	1	2	-	0.003	0.040	0.030	0.037	0.038	0.035	0.023	0.033	0.033	0.045	0.044	0.047	0.045	0.023	0.022	0.047	0.045	0.030	0.032	0.028	0.032
A5	3	1	2	2	-	0.038	0.028	0.035	0.037	0.033	0.022	0.032	0.035	0.044	0.042	0.045	0.044	0.025	0.023	0.045	0.044	0.028	0.030	0.027	0.033
B	24	24	25	24	23	-	0.042	0.044	0.045	0.042	0.037	0.040	0.035	0.013	0.015	0.015	0.016	0.035	0.033	0.052	0.051	0.040	0.042	0.038	0.047
C	18	18	19	18	17	25	-	0.027	0.028	0.025	0.016	0.027	0.032	0.049	0.047	0.051	0.049	0.022	0.020	0.044	0.042	0.032	0.033	0.030	0.035
D1	22	22	23	22	21	26	16	-	0.002	0.002	0.020	0.013	0.040	0.049	0.047	0.051	0.049	0.028	0.027	0.049	0.047	0.040	0.042	0.038	0.045
D2	23	23	24	23	22	27	17	1	-	0.003	0.022	0.012	0.042	0.051	0.049	0.052	0.051	0.030	0.028	0.051	0.049	0.042	0.044	0.040	0.047
D3	21	21	22	21	20	25	15	1	2	-	0.018	0.012	0.038	0.049	0.047	0.051	0.049	0.027	0.025	0.049	0.047	0.038	0.040	0.037	0.044
D4	14	14	15	14	13	22	10	12	13	11	-	0.016	0.027	0.044	0.042	0.045	0.044	0.015	0.016	0.040	0.038	0.023	0.025	0.022	0.032
D5	20	20	21	20	19	24	16	8	7	7	10	-	0.037	0.047	0.045	0.049	0.047	0.025	0.023	0.047	0.045	0.037	0.038	0.035	0.042
F	20	20	21	20	21	21	19	24	25	23	16	22	-	0.038	0.040	0.040	0.042	0.023	0.022	0.047	0.045	0.028	0.030	0.027	0.032
G1	27	27	28	27	26	8	29	29	30	29	26	28	23	-	0.002	0.002	0.003	0.042	0.040	0.054	0.052	0.040	0.042	0.038	0.054
G2	26	26	27	26	25	9	28	28	29	28	25	27	24	1	-	0.003	0.002	0.040	0.038	0.052	0.051	0.038	0.040	0.037	0.052
G3	28	28	29	28	27	9	30	30	31	30	27	29	24	1	2	-	0.002	0.044	0.042	0.056	0.054	0.042	0.044	0.040	0.056
G4	27	27	28	27	26	10	29	29	30	29	26	28	25	2	1	1	-	0.042	0.040	0.054	0.052	0.040	0.042	0.038	0.054
H1	14	14	15	14	15	21	13	17	18	16	9	15	14	25	24	26	25	-	0.002	0.040	0.038	0.025	0.027	0.023	0.028
H2	13	13	14	13	14	20	12	16	17	15	10	14	13	24	23	25	24	1	-	0.038	0.037	0.023	0.025	0.022	0.027
I1	28	28	29	28	27	31	26	29	30	29	24	28	28	32	31	33	32	24	23	-	0.002	0.049	0.049	0.047	0.049
I2	27	27	28	27	26	30	25	28	29	28	23	27	27	31	30	32	31	23	22	1	-	0.047	0.047	0.045	0.047
CRY	18	18	19	18	17	24	19	24	25	23	14	22	17	24	23	25	24	15	14	29	28	-	0.005	0.005	0.040
CRY2	19	19	20	19	18	25	20	25	26	24	15	23	18	25	24	26	25	16	15	29	28	3	-	0.007	0.042
CRY3	17	17	18	17	16	23	18	23	24	22	13	21	16	23	22	24	23	14	13	28	27	3	4	-	0.038
QUIL	19	19	20	19	20	28	21	27	28	26	19	25	19	32	31	33	32	17	16	29	28	24	25	23	-

Fig. 2. Phylogenies of the nuclear LSU (left) and the psaA (right) variants. Both trees were calculated by bootstrapping the alignments in 1000 replicates and calculating neighbour joining (NJ) trees with maximum likelihood (ML) distances using the best fit model chosen using the Akaike Information Criterion. Branches recovered in less than 60 % of the bootstrap replicates were collapsed into polytomies. Lines connecting groups of LSU and psaA variants indicate that the corresponding variants at the two loci were found in the same group of strains. A single co-occurrence is not shown, that of rDNA variant D1 and of psaA variant cm-I, in the possibly non-clonal strain G184 (see discussion in Beszteri *et al.* 2005a).

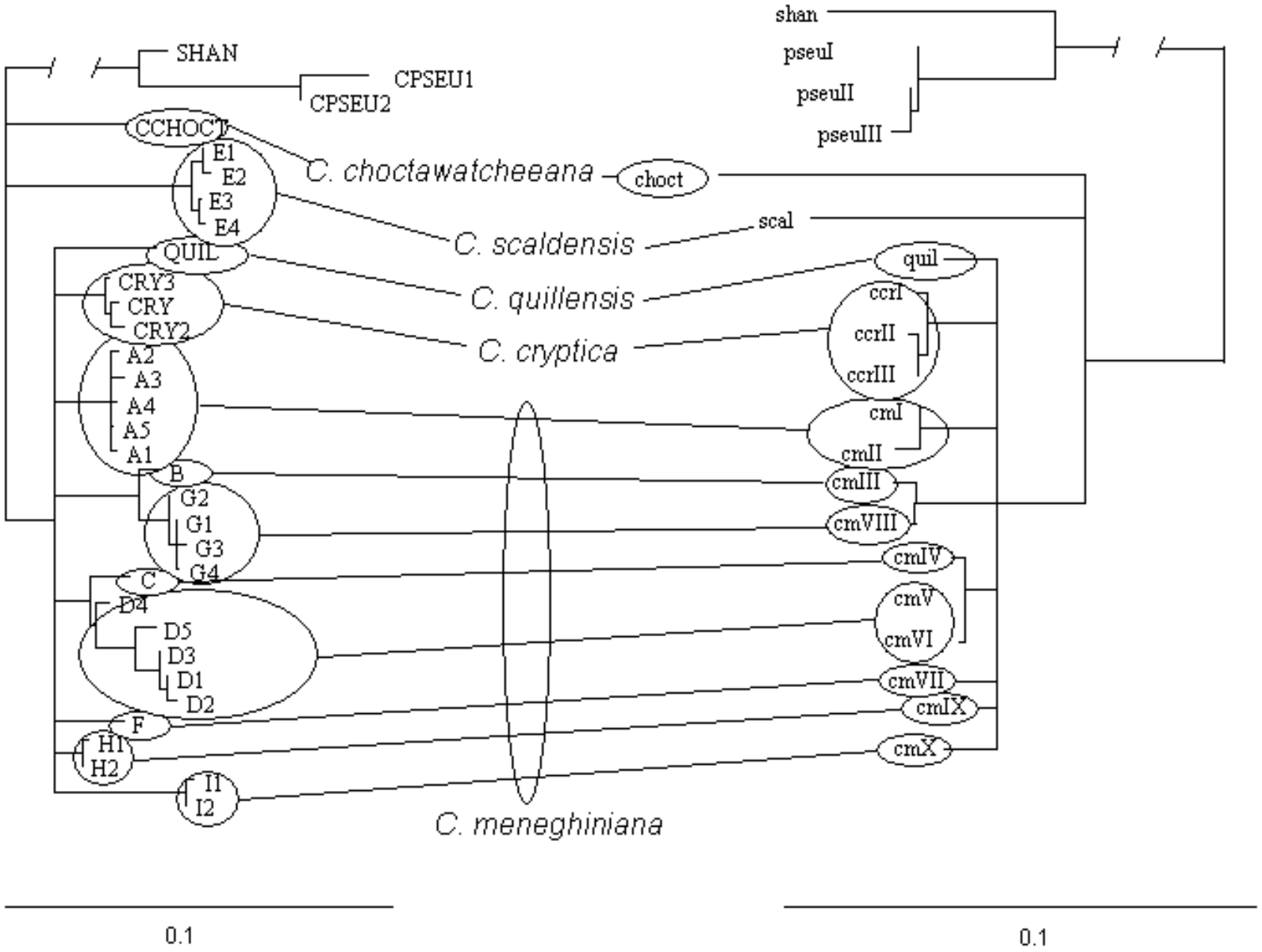
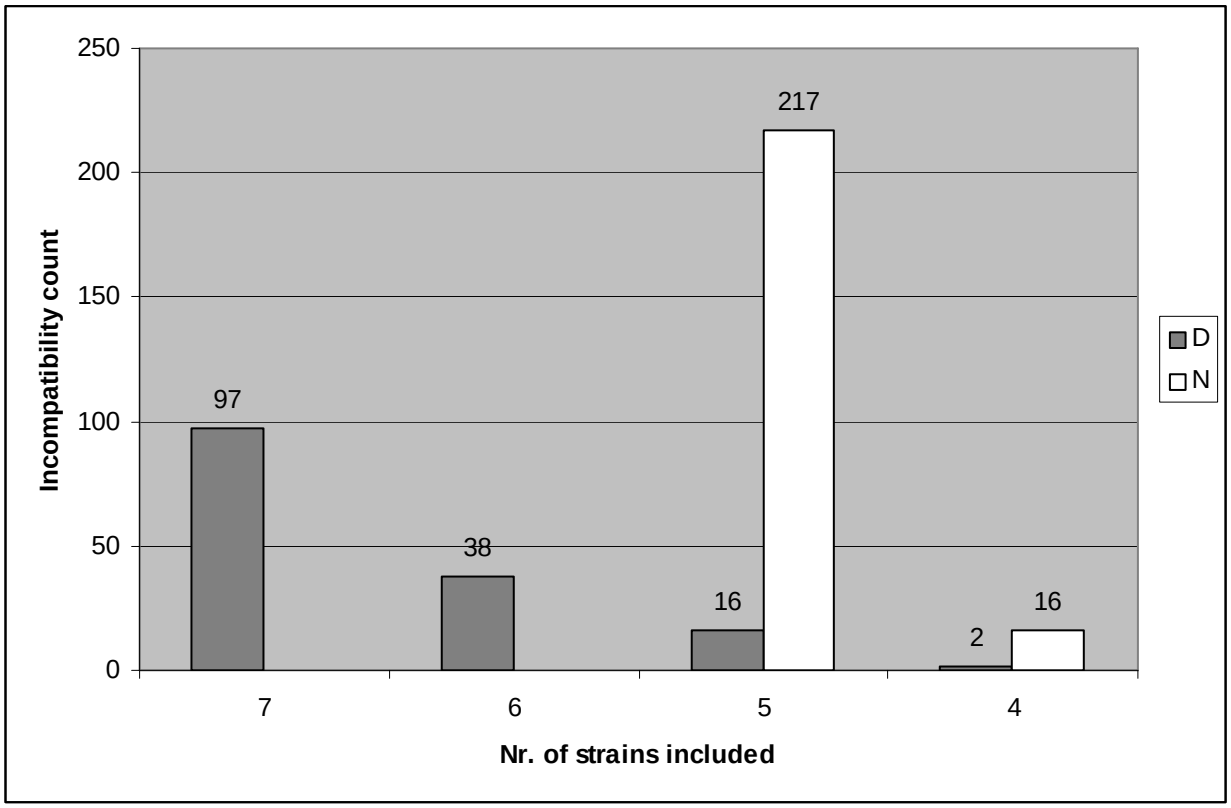

SHAN
CPSEU1
CPSEU2
CCHOCT
E1
E2
E3
E4
QUIL
CRY3
CRY
CRY2
A2
A3
A4
A5
A1
B
G2
G1
G3
G4
C
D4
D5
D3
D1
D2
F
H1
H2
I1
I2
C. choctawatcheeana
C. scaldensis
C. quillensis
C. cryptica
C. meneghiniana
shan
pseuI
pseuII
pseuIII
choct
scal
quil
ccrI
ccrII
ccrIII
cmI
cmII
cmIII
cmVIII
cmIV
cmV
cmVI
cmVII
cmIX
cmX
0.1
0.1

Fig. 3. Decrease of incompatibility counts following removal of strains contributing the most to phylogenetic incompatibility in the two populations. D – Daverden (7 strains), N – Nienburg (5 strains) population.



3. Synthesis

3.1. Morphometric approaches to study diatom valve ultrastructural variation

In **Publication I**, two different morphometric approaches were applied to study valve ultrastructural variation in cultures and field samples of *C. meneghiniana* and *C. scaldensis*. The distinction between these species has previously been questioned based on the notion that intermediate morphologies between those typical of these species have been encountered in field samples (Kiss & Ector 2000, and Muylaert, pers. comm.). Our observations also supported the existence of such intermediates. However, both morphometric approaches convincingly demonstrated that the intermediate morphologies observed in our cultures and in samples from two North European estuaries were morphologically distinct from typical *C. meneghiniana* as well as from typical ('extreme') *C. scaldensis*, and these morphs displayed three distinct size reduction series.

The biological relevance of this morphological distinctness is underlined by the following facts. First, size ranges of the morphs overlapped to a large extent, excluding the possibility that distinct phases of the size reduction series of a single species were sampled. Second, the two cultivated morphs were morphologically distinct in cultures in the same culture conditions. Third, they were also distinct in sympatry in the field samples; in the samples from the Schelde estuary, where all three morphs were found in abundance in sympatry, they were all distinct. Fourth, morphological correspondence of different populations of the morphs (cultures and the two estuarine localities) was good; the largest differences observed were between different populations of the 'ambiguous' *C. scaldensis* morph. These were, however, also explained by differences in the size distributions of the different samples. The conclusion drawn from these facts was that the three morphs represented three different species instead of life cycle phases or environmentally induced morphological variation of a single species.

Both methodologies applied proved effective and useful for drawing the above conclusions. Results from them complemented each other to some extent. The distinctness of *C. meneghiniana* from the other morphs was shown most markedly in the geometric morphometric analyses, whereas the distinctness of the 'extreme' *C. scaldensis* from the other morphs was so in the conventional morphometric analyses. This difference was caused by the fact that in spite of our effort that the two morphometric approaches characterize the same morphological features, there was some complementarity between them. This complementarity proved useful in this case. It made clear that the morphological feature allowing the most straightforward differentiation of the two *C. scaldensis* morphs was the frequency of marginal strutted processes, a feature we could not account for in the geometric morphometric approach. On the other hand, the feature allowing the easiest differentiation of *C. meneghiniana* from the *C. scaldensis* morphs was the radial extension of the alveoli. Although intuitively we consider that this feature was well characterized using both morphometric approaches, the differences shown by the geometric morphometric analyses were larger. This was probably because the other features characterized by this approach differed also markedly between the morphs.

Besides the morphological distinctness revealed by the morphometric comparisons, cultures of the 'ambiguous' *C. scaldensis* morph also had a distinct set of rDNA variants from the *C. meneghiniana* cultures isolated from the same locality at the same time, as shown in

Publication II. The congruence of these two, independent sets of information gives further strong support to the idea that these morphs are not conspecific.

Conventional and landmark-based morphometric methods were used to explore valve ultrastructural variation quantitatively in cultures and field samples of *C. meneghiniana*, *C. scaldensis*, and a morph that displayed an intermediate morphology between them. The results indicated that the intermediate morph displayed a distinct size reduction series from typical forms of the two nominal species instead of linking their morphological variation, suggesting that the three morphs were not conspecific. Ribosomal DNA variation in the two cultured morphs further reinforced this hypothesis.

3.2. Deviations from panmictic expectations within C. meneghiniana

Publication II provided the first strong evidence for the hypothesis that *C. meneghiniana* was a 'structured species', i.e., one consisting of either multiple, cryptic species or a complex of strictly clonal lineages. The evidence was provided by the strong deficiency of clonal strains with combinations of divergent rDNA variants, which were found in the population. This phenomenon was observed in both variable regions used in **Publication II**, the D1/D2 regions of the 28S rDNA as well as in the ribosomal ITS (internal transcribed spacer) regions. Furthermore, the groups identified on the basis of on one of these regions were congruent with those identified on the basis of on the other, i.e. the two data sets were completely congruent, although amounts of variation present in the two regions were different. The 18S rDNA cannot be considered an effective molecular marker for such intraspecific studies because of its conservation (it is about 1800 bases long with less than 10 positions varying within *C. meneghiniana*, as opposed to the 613 bases long D1/D2 regions of the 28S rDNA, with 43 polymorphic positions). However, variation in this conserved region was also congruent with that observed in the two other regions.

Having revealed that the *C. meneghiniana* population sampled was a complex of lineages, among which gene flow was absent or at least so strongly reduced that it could not be detected, the next question arising was whether this could be explained by the presence of multiple cryptic sexual species, or rather by the complete lack of allogamous sexual reproduction in this population. We also attempted to test the hypothesis of panmixia within the single rDNA groups using the ITS sequences to answer this question. An analysis of molecular variance (AMOVA) indicated that ITS sequence variation among the *C. meneghiniana* strains could be partitioned into significant within-strain and among-group variance components, without a significant within-group, among-strain component remaining. This was interpreted as an indication that genetic variation was homogenously distributed among strains belonging to the separate rDNA groups. However, as also noted in the publication, we did not consider this evidence conclusive for two reasons. Firstly, number of strains belonging to individual rDNA type was low, ranging from one to eight. Secondly, even the most variable of the regions compared, the ITS regions, showed a limited variation within the rDNA groups.

An important aspect of **Publication II** was that it aimed at identifying which of the rDNA regions could be used the most effectively to explore patterns of within-morphospecies genetic variation in *C. meneghiniana*. As noted above, 18S rDNA showed too little variation to be useful for this purpose. The amount of variation provided by the second most conserved region considered, the hypervariable D1/D2 regions of the 28S rDNA, was sufficient to resolve the independent groups. This region also lacked insertion/deletion polymorphisms,

making its direct sequencing feasible. On the other hand, whereas the ITS regions provided the most information (i.e., they were the most variable) of the three regions sequenced, they often showed within-strain length polymorphisms, which necessitated cloning of the PCR products for sequencing. Therefore, the D1/D2 regions of the 28S rDNA were chosen to be applied for the following larger scale comparisons.

The results of these comparisons, described in **Publication III**, further reinforced the notion that *C. meneghiniana* was a structured species. The diversity of 28S rDNA variants obtained in this study was larger than that found in the single population studied in **Publication II**, but the basic pattern was similar in that no divergent variants were found in single strains. Although one of the rDNA groups identified in this part of the work (referred to as group D) showed somewhat larger amount of variation than the others, even this group of variants was monophyletic.

The variation at the plastid encoded locus psaA was completely congruent with that found in the nuclear rDNA. Strains belonging to individual rDNA groups usually shared a single psaA variant. In two cases, within-group variation in the psaA sequences was observed, but it was confined to a single position in one case, and to three positions in the other. At the same time, the number of polymorphic positions in the sequenced psaA region within the whole morphospecies was 45. The congruence of variation at the two, physically unlinked genomic locations (one situated in the nuclear, the other in the chloroplast genome) gave strong further support to the idea that the organisms belonging to different rDNA groups within *C. meneghiniana* do not exchange genes with each others.

Inclusion of strains of closely related species revealed a further interesting phenomenon. Whereas species distinguishable from *C. meneghiniana* and its close relatives by the presence of closed alveoli (the feature also distinguishing from *C. meneghiniana* both morphs of *C. scaldensis* studied in **Publications I** and **II**) were also clearly distinguished from *C. meneghiniana* at both genomic locations sequenced, this was not the case with two other species. *C. quillensis* and *C. cryptica* both share the character of open alveoli with *C. meneghiniana*; therefore, they were referred to as members of the *C. meneghiniana* complex in **Publication III**. They were also not more different from the latter species in their rDNA and psaA variants than strains of *C. meneghiniana* were from each other. Although phylogenetic information of neither of these regions was sufficient to resolve the evolutionary relationships of variants within the *C. meneghiniana* complex with confidence, C. meneghiniana appeared paraphyletic in phylogenetic analyses with most methods. Although the monophyly of the *C. meneghiniana* complex was strongly supported in these analyses, the separation of *C. quillensis* and *C. cryptica* from *C. meneghiniana* was not. This result practically repeats those obtained by Bourne (1992) in her chloroplast genome RFLP study: she also found that *C. cryptica* grouped well within the heterogeneous group of *C. meneghiniana* strains. Whereas the interpretation of her results remained uncertain, we can conclude with a larger confidence that this is caused by *C. meneghiniana* being a structured species, either consisting of multiple, reproductively isolated sexual species, or of strictly clonal lineages. Based on the indications (**Publications II** and **III**) that the former seems to apply to *C. meneghiniana*, the most plausible conclusion seems to be that the *Cyclotella meneghiniana* complex probably represents a complex of several distinct species, some of which diverged morphologically from a 'general', characteristic morphology sufficiently allowing for their recognition, whereas others did not. The resolution of the systematic problems around the polymorphic *C. cryptica* seems to be according to these results that it is indeed a separate species from *C. meneghiniana*, but also a member of the complex of species including all extant descendants of the most recent common ancestor of all biological species showing the typical *C. meneghiniana* morphology.

Genetic variation in a single 'population' of *C. meneghiniana* as well as in a global collection of strains of this species strongly supported the hypothesis that *C. meneghiniana* is a structured species, consisting of multiple, reproductive isolated lineages (cryptic species or strictly clonal lineages). The monophyletic group containing *C. meneghiniana* also contains other morphospecies. Nor the nuclear rDNA sequences, neither the plastid encoded psaA sequences were fully conclusive in deciding whether the hypothesis of cryptic species or that of clonality explains this genetic structuring.

3.3. Clonality vs. cryptic species

A high-resolution multilocus fingerprinting approach, AFLP, was applied in **Publication III** to answer the question whether clonality or the presence of cryptic species was responsible for the genetic structure observed in *C. meneghiniana*. The results of these analyses strongly indicated that sexual reproduction was not absent in the evolutionary history of the two populations sampled. This was interpreted as supporting the view that *C. meneghiniana* is a complex of multiple, reproductively isolated sexual species rather than that of independently evolving, effectively clonal lineages. However, the diversity of apparent life cycle strategies in other species complexes, like the *Sellaphora pupula* complex, indicates that this result must cautiously be extrapolated to the other, with these fingerprinting methods as yet unsampled genetic subgroups within *C. meneghiniana*. The possibility cannot be excluded that some of these groups do indeed reproduce sexually, whereas in others, allogamy might be completely or effectively absent.

To our knowledge, **Publication III** represents the first experimental testing of the hypothesis of strict clonality in a diatom species using molecular techniques. The major part of knowledge available about diatom life cycle strategies has come from observations of sexual reproduction and auxospore production. As noted in the Introduction, such observations are not necessarily able to answer the question as to what extent the presence or the lack of sexuality affects patterns of genetic (and thus also phenotypic) variation in a species. As an example, effective clonality has been revealed e.g. in *Toxoplasma gondii*, a parasitic protist earlier known to undergo an obligatory sexual stage in cat (Ajzenberg *et al.* 2002). On the other hand, allogamous fertilization has never been observed in *C. meneghiniana*, in spite of the numerous life cycle investigations performed with this species (cited in the Introduction). Because of this, reviews of diatom sexuality have referred to *C. meneghiniana* as an example of diatom species that abandoned allogamous sexual reproduction (Drebes 1977; Mann 1993). However, the effect of sexuality upon genetic variation patterns within a species might be substantial even if allogamous sexual reproduction occurs only relatively rarely. It is still unclear how often sexuality appears even within the single rDNA-group where the hypothesis of clonality was tested, but it seems clear that it occurs often enough to shuffle genetic variation among lineages, an unexpected finding when considering the experimental studies mentioned above.

Data collected using AFLP, a high-resolution multilocus genetic fingerprinting method, indicated that allogamous sexual reproduction was not absent in the two sampled populations of *C. meneghiniana*. This suggested that the presence of several, reproductively isolated sexual species might be responsible for the genetic structure observed in this morphospecies, rather than strict clonality or autogamy.

3.4. Population differentiation on small geographic scale

Perhaps the most unexpected result of this work was the strong genetic differentiation between the two populations included in the AFLP analyses. The two (planktonic) populations were sampled on the same day, at localities situated around 40 kms from each other in the River Weser (North Germany). The strains isolated from these populations all belonged to a single genotypic group (group A) according to their nuclear rDNA and psaA sequences. The geographic distance between the two localities was so small and lacked any hydrographic barriers that we expected to sample effectively the same population at these localities. This expectation proved wrong. The differentiation between the two populations was so strong that the genotypes of strains grouped into exclusive groups according to their origin; a major part of the overall variation in their genotypes represented differences that were fixed between the two populations (59 loci) and patterns that were compatible with the grouping of strains according to their locality of origin (64 informative loci; overall: 152 informative loci). Only 29 loci showed variation patterns incompatible with a grouping according to origin.

It remains unclear whether this marked differentiation can be explained supposing that the two populations were actually further cryptic species, unrevealed by the DNA sequences used, or that this was rather a case of strong population differentiation within a single species. Our results indicated that within-population variation patterns indicated the presence of allogamous sexuality within the populations. Therefore, the alternative of further cryptic species vs. strong population differentiation can be translated as whether a lack of gene flow between these populations is responsible for their differentiation, or strongly divergent selective pressures acting upon them. As discussed in **Publication III**, some indications of the lack of gene flow between the populations can be found in our data. Three fixed differences were observed between the psaA variants of the two populations. Also, different rDNA variants were typical of the two populations (variants A2 and A5 in the Daverden population, variants A3 and A4 in the Nienburg population). Although one of the rDNA variants typical of the Daverden population (A2) also appeared in one of the Nienburg strains, other scenarios than gene flow (retained ancient polymorphism from the common ancestor species) might as well be invoked to explain this. In any case, our data are insufficient to settle this issue with confidence.

Remarkably, the fingerprinting method applied to test the hypothesis of clonality revealed a marked genetic differentiation between two geographically close populations. The differentiation was so large that the conspecifity of the populations became questionable. Similar observations have also been made in other population genetic surveys of diatoms, further underlining our ignorance concerning their microdiversity and microevolution.

3.5. Conclusion and prospects

This thesis has shown that one of the most common freshwater diatom morphospecies, *C. meneghiniana*, is not a single species. It is rather a complex of several, genetically distinct lineages, at least some of which also reproduce allogamously, whereas others might be strictly clonal or autogamous. These lineages / species seem to be cryptic, i.e., at least to our present knowledge, they are not distinct morphologically. The next questions arising in this context are (1) what are the biological – ecological, distributional, physiological, etc. – differences

between these cryptic lineages, and (2) can other, as yet untested, minor morphological features distinguish them, in order to make their routine identification possible. The second question is not as straightforward to answer with these organisms as it might seem: because of the problems brought about by phenotypic plasticity, this would probably also require experimental approaches. However, both questions need to be answered to be able to fully appreciate the consequences of this (up to now) hidden diversity.

Besides consequences concerning this taxonomic groups, some more general conclusions can also be drawn from this work. Species names are important everyday working tools of biologists in a wide diversity of research areas. However, significant effort might be needed to understand the nature of different species in order for these names to be biologically truly meaningful. Studies of species-level (phenotypic and genetic) variation using multiple approaches are necessary to arrive at a better understanding of this problem. Especially in the case of diatoms, the role of morphological information for species identification and systematics can be expected to retain its central role due to the unique feature-richness of these organisms and to the wealth of information their fossilized frustules can provide about past environmental conditions. Still, the use of alternative approaches (notably studies of reproductive compatibilities and molecular approaches) indicates that several of the diatom morphospecies are not species in the evolutionary sense. The most widespread problem appears to be that taxonomic species defined based on morphological criteria are often composed of multiple species.

Discoveries of cryptic species are becoming a trend not only in diatom studies (Sarno *et al.* 2005), but in other eukaryotic microorganismal groups providing 'above-average' amounts of morphological features as well, including foraminifers (De Vargas *et al.* 1999), coccolithophores (Sáez *et al.* 2003) and dinoflagellates (Montresor *et al.* 2003). In a recent review, De Vargas and co-authors propose that it might be the rule rather than an exception that morphological species of planktonic taxa are in fact complexes of cryptic or pseudo-cryptic species (those that are also identifiable morphologically, but were only recognized to represent separate species through molecular surveys) (De Vargas *et al.* 2004). They introduce the concept of 'planktic super-species' to refer to these complexes, and suggest that strong stabilizing selection acting on morphologies contributes to their maintenance through speciation events. Significantly, and in accord with predictions from ecological competition theory (Sáez & Lozano 2005), the sibling species belonging to such 'super-species' are generally found to occupy subdivisions of the geographic, temporal or ecological range attributed to the morphospecies. A major practical implication of this fact is that once ways to identify such cryptic species routinely are found, the performance of ecological monitoring might be increased. In several cases, minor morphological differences between the 'cryptic species' involved can be found after their discovery (Sáez *et al.* 2003); in such cases, the enhanced performance can even be transferred to palaeo-environmental studies. However, knowledge about this cryptic diversity is not irrelevant for other disciplines either. In laboratory physiological studies, cultures investigated are labelled with species names. In such cases, knowledge of species boundaries can help identifying to which group of organisms particular results can be extrapolated to.

Molecular genetic methods seem to be the most generally applicable tools to address such problems. Furthermore, they have the important capability of informing not only about variation patterns, but to some extent also about microevolutionary processes acting in their background more directly than phenotypic markers or life cycle observations can. For instance, they can be used to measure gene flow actually occurring between populations or demes, or they can be used to draw inferences about life cycle characteristics (e.g. the presence or lack of sexual reproduction) of the populations concerned. Molecular methods are becoming a standard tool in microdiversity investigations of diatoms, although these latter

potentials are perhaps not yet fully exploited (Orsini *et al.* 2002; Lundholm & Moestrup 2002; Lundholm *et al.* 2003; Orsini *et al.* 2004; Sarno *et al.* 2005).

Population genetic surveys of diatoms applying high-resolution markers have come to similar surprising results in a number of cases. As discussed in **Publication III**, strong population differentiation over the year was observed in Gallagher's multilocus enzyme electrophoresis study of *Skeletonema costatum* (Gallagher 1980). *S. costatum* is since known to be a species complex (Sarno *et al.* 2005), although the question whether the winter and summer populations observed by Gallagher (1980) corresponded to different cryptic species is unanswered. Similarly, large genetic differentiation on a small geographic scale was detected in a microsatellite study of *Ditylum brightwellii*, a species that was considered taxonomically unproblematic before this finding (Rynearson & Armbrust 2004).

4. Summary

This thesis was aimed at the clarification of long standing problems concerning species level diversity in one of the most commonly reported and intensively investigated freshwater diatom species *Cyclotella meneghiniana* Kütz. On the one hand, this species is an example of diatom species causing taxonomic problems simply with their large morphological variation that lacks obvious discontinuities. On the other hand, a number of not standard investigations carried out with this species have also contributed to the recognition that it was taxonomically problematic. Plasticity of valve morphology spanning morphospecies boundaries has been observed in this and a closely related morphospecies, *C. cryptica* Reimann, Lewin & Guillard. Furthermore, the *C. meneghiniana* strains included in a study of chloroplast genome variation formed a paraphyletic group with respect to *C. cryptica*. The delimitation of a number of closely related species from *C. meneghiniana* was problematic, and the question was unanswered whether *C. meneghiniana* itself was in fact a single, unstructured species, a complex of multiple, reproductively isolated sexual species, or a complex of strictly clonally or autogamously reproducing lineages.

The first part of the thesis included the sampling of a single *C. meneghiniana* 'population' by isolating clonal cultures. Frustule ultrastructure of the cultures and of field samples was analysed using two different morphometric methodologies, one of them (landmark based geometric morphometric methods) used for the first time with diatoms. These morphometric analyses contributed significantly to clarifying a problem concerning the delimitation of *C. meneghiniana* from a closely related species, *C. scaldensis* Muylaert & Sabbe. The results revealed the existence of a morph with an intermediate valve morphology between those considered typical of these species, and also that this morph represented a distinct size reduction series from both of them instead of linking their morphological variation. The facts that the three morphs were distinct in sympatry, and cultures of two morphs (*C. meneghiniana* and the intermediate morph) were also morphologically distinct when grown in the same conditions, were interpreted as strong indications of that they were not conspecific. Numerous fixed differences in DNA sequences obtained from different regions of the ribosomal operons of cultures of the cultivated morphs (including a pair of complementary base changes in the 18S rDNA) further reinforced this notion. Cultures of the third morph (typical *C. scaldensis*) could not be obtained for similar comparisons.

Next, patterns of ribosomal DNA sequence variation among the fifteen cultures displaying the characteristic *C. meneghiniana* morphology were explored. The patterns observed indicated that the population sampled was actually not a single panmictic population, suggesting that *C. meneghiniana* was indeed a genetically structured morphospecies. The data obtained were also applied in an attempt to test whether this structuring can be explained by the presence of multiple, reproductively isolated sexual species, or by the complete or nearly complete lack of allogamous sexual reproduction. The results suggested the former possibility was the more plausible explanation, but the sequenced regions showed limited variation at the level of the distinct rDNA groups found, thus this conclusion remained tentative.

The following step was surveying sequence variation in the D1/D2 regions of the nuclear ribosomal DNA and in a region of the plastid encoded psaA gene in a collection of strains including *C. meneghiniana* strains with diverse geographic origins and cultures of some closely related species (*C. cryptica*, *C. quillensis* Bailey, *C. choctawatcheeana* Prasad, and the 'ambiguous' *C. scaldensis* morph). Again, a large diversity of rDNA variants was found in the collection of strains, but single strains never contained combinations of divergent variants. Patterns of variation in the psaA region sequenced were congruent with those

observed in the nuclear rDNA region. These facts further reinforced the conclusion that *C. meneghiniana* is a structured species.

However, these data were also unable to decide whether the genetic structuring was caused by strict clonality or the presence of cryptic species. To answer this question, amplified fragment length polymorphism (AFLP) fingerprints from two populations, which contained members of a single rDNA/psaA group, were obtained, and phylogenetic incompatibility in the AFLP data was explored. The results indicated a strong deviation from clonal expectations in both populations. Unexpectedly, genetic differentiation between strains isolated from the two populations (sampled on the same day, situated about 40 kms apart in the River Weser) was so substantial that the possibility that even these populations represented reproductively isolated (biological) species could not be excluded.

In conclusion, the work carried out in this thesis lead to the insight that *C. meneghiniana*, one of the most often encountered and most commonly used model diatom species in experimental culture studies of freshwater diatoms, was not a single species. Instead, it appears to be a complex of several, genetically distinct taxa, some of which reproduce allogamously. The possibility cannot be excluded that some lineages are strictly clonal. Monophyly of *C. meneghiniana* was not supported by either the nuclear rDNA or the psaA region. The data obtained suggested that the *C. meneghiniana* complex (including at least *C. cryptica* and *C. quillensis* besides *C. meneghiniana* itself) was a complex of genetically distinct species, some of which have also morphologically diverged sufficiently to allow their recognition as separate morphospecies. Other species within the complex show considerable morphological overlap in their features classically used for species identification in these taxa. The occurrence of the phenomenon of morphological plasticity within this group might make it a particularly complex challenge to define morphological species boundaries among the, as yet, cryptic species within *C. meneghiniana*.

5. Zusammenfassung

Die Abgrenzung von Arten in der Diatomeensystematik basiert auf der Morphologie der Silikatschalen dieser Mikroalgen. Obwohl die Diatomeenschalen eine Vielzahl von morphologischen Merkmalen haben, die bei anderen mikroskopischen Organismen kaum zu beobachten sind, ist die Abgrenzung vieler Diatomeenarten problematisch. *Cyclotella meneghiniana* Kütz. ist eins der klassischen Beispiele dafür. Einerseits ist ihre Abgrenzung anhand der großen morphologischen Variabilität, die keine eindeutigen Diskontinuitäten aufzeigt, problematisch. Andererseits wurden mit dieser Art zusätzliche Studien durchgeführt, die den Eindruck, dass diese Art taxonomisch problematisch sei, weiter verstärkten. Schalenmorphologische Plastizität, welche die vermuteten Artengrenzen überschritt, wurde in *C. meneghiniana* und in der nah verwandten *C. cryptica* Reimann, Lewin & Guillard beobachtet. Außerdem hat eine Studie von Plastidengenomvariation gezeigt, dass *C. meneghiniana* bezüglich *C. cryptica* paraphyletisch war. Zusammengefasst ist die Abgrenzung mehrerer nah verwandter Arten von *C. meneghiniana* nach wie vor problematisch. Auch die Frage wurde bisher nicht beantwortet, ob *C. meneghiniana* selbst eine einzelne, unstrukturierte Art, ein Komplex von mehreren, reproduktiv isolierten, sich sexuell fortpflanzenden Arten, oder ein Komplex von rein klonalen Abstammungslinien ist.

In dieser Dissertation wurde die schalenmorphologische und molekulare Variabilität in dieser taxonomisch problematischen Süßwasserdiatomeenart und ihrer nahen Verwandten untersucht. Ziel war dabei, zur Klärung der Artengrenzen in dieser taxonomischen Gruppe beizutragen.

Im ersten Teil dieser Arbeit wurden klonale Kulturen aus einer einzigen *C. meneghiniana* ‚Population' isoliert. Von diesen Kulturen und von nicht-kultiviertem Material wurde mittels zwei unterschiedlicher morphometrischer Ansätze die Schalenultrastruktur quantitativ verglichen. Diese morphometrischen Analysen trugen zur Klärung der morphologischen Abgrenzung einer *C. meneghiniana* nah verwandten Art, *C. scaldensis* Muylaert & Sabbe, bei. Ein Morphotyp mit einer intermediären Schalenmorphologie zwischen den für diese beiden Arten typischen erschien in unseren Proben. Die morphometrischen Analysen zeigten, dass dieser Morphotyp keine morphologische Überlappung mit den für die beiden Arten charakteristischen Morphologien gezeigt hat. Die drei Morphotypen (typische *C. scaldensis* und *C. meneghiniana* und der Übergangsmorphotyp) unterschieden sich sowohl sympatrisch als auch in Kulturen. Das verdeutlicht, dass sie nicht konspezifisch waren. Diese Schlussfolgerung wird durch eine Vielzahl von Sequenzunterschieden in unterschiedlichen Bereichen des zellkern-kodierten ribosomalen Operons weiter unterstützt.

Die nächste Fragestellung der Arbeit war, ob genetische Strukturierung innerhalb der Population, repräsentiert durch die Kulturen mit der charakteristischen *C. meneghiniana* Morphologie, nachgewiesen werden kann. Dafür wurde die Sequenzvariabilität in unterschiedlichen Bereichen des zellkern-kodierten ribosomalen Operons untersucht. Die Ergebnisse zeigten, dass die ‚Population', aus der die Kulturen stammten, genetisch / reproduktiv strukturiert war. Die Daten dienten weiter dazu, die Natur der genetischen Strukturierung (reine Klonalität oder kryptische, reproduktiv isolierte Arten) aufzuklären. Die Hypothese der kryptischen Arten wurde als die plausiblere Alternative betrachtet, weil unterhalb der einzelnen rDNA-Gruppen keine weitere hierarchische Strukturierung der Sequenzvariation gefunden wurde. Allerdings bleiben die Schlussfolgerungen aufgrund der geringen Sequenzvariabilität innerhalb der gefundenen rDNA-Gruppen diesbezüglich provisorisch.

Zur Erfassung der globalen Diversität in dieser und nah verwandten Morphospezies, wurden die Untersuchungen mit einer größeren Sammlung klonaler Kulturen, sowohl von

Stämmen von *C. meneghiniana*, isoliert aus verschiedenen geographischen Lokalitäten, als auch von anderen nah verwandten Arten (*C. cryptica*, *C. quillensis* Bailey, *C. choctawatcheeana* Prasad, und der Übergangsmorphotyp von *C. scaldensis*) erweitert. Außerdem wurde neben einer Region des kernkodierten ribosomalen Operons auch ein Teil eines plastidenkodierten Gens (psaA) aus diesen Stämmen sequenziert. Auch in diesem Fall wurde eine große Diversität von ribosomalen Sequenzen gefunden, und die einzelnen Stämme enthielten keine Kombinationen von divergenten Varianten. Die Variationsmuster in der psaA-Region stimmten mit der, die in der ribosomalen Region beobachtet wurde, komplett überein. Diese Ergebnisse verdeutlichen, dass die genetische Strukturierung außer im Zellkerngenom auch im Plastidengenom nachgewiesen werden kann. Weiterhin zeigte es sich, dass die Morphospezies *C. meneghiniana* global eine genetische Strukturierung aufweist.

Anhand dieser Daten konnte jedoch nicht entschieden werden, ob diese genetische Strukturierung durch fehlende sexuelle Fortpflanzung oder durch die Präsenz von mehreren kryptischen Arten innerhalb der Morphospezies zu erklären ist. Um diese Frage zu beantworten, wurden aus Isolaten zweier nah gelegener Populationen mittels AFLP (*A*mplified *F*ragment *L*ength *P*olymorphism) molekulare Fingerabdrücke erzeugt. Die große phylogenetische Inkompatibilität in den AFLP-Fingerabdrücken wies auf den starken Einfluss von sexueller Fortpflanzung hin. Die genetische Differenzierung zwischen den Populationen, die aus zwei geographisch nah gelegenen Lokalitäten im Laufe eines Tages isoliert wurden, und die zudem alle zur gleichen rDNA/psaA-Gruppe gehörten, war unerwartet hoch, so dass selbst die Erklärung, dass auch diese Populationen nicht konspezifisch waren, nicht ausgeschlossen werden kann.

Zusammenfassend hat es sich gezeigt, dass *C. meneghiniana*, eine der häufigsten und in experimentellen Kulturstudien meist verwendeten Morphospezies bei den Süßwasserdiatomeen, in der Tat keine einzelne Art ist. Stattdessen scheint sie ein Komplex mehrerer, genetisch distinkten Taxa zu sein. Obwohl es nicht ausgeschlossen werden kann, dass manche dieser Taxa rein klonale Abstammungslinien sind, hat diese Arbeit sexuelle Fortpflanzung in einer der Gruppen nachgewiesen. Die Morphospezies *C. meneghiniana* zeigte sich in beiden untersuchten genomischen Regionen als nicht monophyletisch. Die Ergebnisse weisen darauf hin, dass der monophyletische *C. meneghiniana* Komplex, der mindestens *C. quillensis* und *C. cryptica* außer *C. meneghiniana* einschließt, ein Komplex mehrerer, genetisch unterschiedlicher Arten ist, von denen manche auch morphologisch von den anderen zu unterscheiden sind. Andere Arten des Komplexes überschneiden sich in den Merkmalen, die klassisch für Artenabgrenzung in dieser Gruppe angewendet werden.

6. References

AJZENBERG, D., BANULS, A.L., TIBAYRENC, M., DARDE, M.L. (2002). Microsatellite analysis of *Toxoplasma gondii* shows considerable polymorphism structured into two main clonal groups. *International Journal of Parasitology* **32**: 27-38.

ALVAREZ COBELAS, M., ROJO, C. (1994). Factors influencing the share of planktic diatoms in lakes. *Archiv Für Hydrobiologie, Supplement* **104**: 73-104.

ALVAREZ, I., WENDEL, J.F. (2003). Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**: 417-434.

ALVERSON, A.J., THERIOT, E.C. (2005). Comments of recent progress toward reconstructing the diatom phylogeny. *Journal of Nanoscience and Nanotechnology* **5**: 57-62.

ANDERSON, J.B., KOHN, L.M. (1998). Genotyping, gene genealogies and genomics bring fungal population genetics above ground. *Trends in Ecology & Evolution* **13**: 444-449.

ARNOLD, M.L. (2004). Natural hybridization and the evolution of domesticated, pest and disease organisms. *Molecular Ecology* **13**: 997-1007.

AVISE, J.C., WALKER, D. (1999). Species realities and numbers in sexual vertebrates: perspectives from an asexually transmitted genome. *Proceedings of the National Academy of Sciences of the USA* **96**: 992-5.

AVISE, J.C., WOLLENBERG, K. (1997). Phylogenetics and the origin of species. *Proceedings of the National Academy of Sciences USA* **94**: 7748-55.

AZARIAH, J., RAO, V.N.R., SRIDHARAN, V.T. (1978). Effect of the epidermal secretions of Hemichordate, *Ptychodera flava* on growth of *Amphora coffeaeformis* and *Cyclotella meneghiniana* (diatoms). *Proceedings of the Indian Academy of Sciences, Section B* **87** : 145-149.

BAK, M., WAWRZYNIAK-WYDROWSKA, B., WITKOWSKI, A. (2001). Odra River discharge as a factor affecting species composition of the Szczecin Lagoon diatom flora, Poland. *Studies on Diatoms* : 491-506.

BART, A., BARNABE, C., ACHTMAN, M., DANKERT, J., VAN DER ENDE, A., TIBAYRENC, M. (2001). The population structure of *Neisseria meningitidis* serogroup A fits the predictions for clonality. *Infection, Genetics and Evolution* **1**: 117-22.

BARTON, N.H. (2001). The role of hybridization in evolution. *Molecular Ecology* **10**: 551-68.

BAUM, D.A. and SHAW, K.L. (1995) Genealogical perspectives on the species problem. In PC Hoch, AG Stevenson (eds) Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden, St. Louis, Missouri, pp 289-303.

BEAUMONT, M.A., BALDING, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology* **13**: 969-80.

BEHNKE, A., FRIEDL, T., CHEPURNOV, V.A., MANN, D.G. (2004). Reproductive compatibility and rDNA sequence analyses in the *Sellaphora pupula* species complex (Bacillariophyta). *Journal of Phycology* **40**: 193-208.

BOOKSTEIN, F.L. (1991). Morphometric Tools for Landmark Data: Geometry and Biology. Cambridge University Press, Cambridge (UK).

BOOKSTEIN, F.L., CHERNOFF, B., ELDER, R.L., HUMPHRIES, J.M., SMITH, G.R., and STRAUSS, R.E. (1985). Morphometrics in evolutionary biology : the geometry of size and shape change, with examples from fishes. Academy of Natural Sciences of Philadelphia, Philadelphia, Pa.

BOURNE, C.E.M. (1992) Chloroplast DNA structure, variation and phylogeny in closely related species of *Cyclotella*. PhD Dissertation, University of Michigan

BOURNE, C.E.M., PALMER, J.D., STOERMER, E.F. (1992). Organization of the chloroplast genome of the freshwater centric diatom *Cyclotella meneghiniana*. *Journal of Phycology* **28**: 347-355.

BÜCHEL, C. (2003). Fucoxanthin-chlorophyll proteins in diatoms: 18 and 19 kDa subunits assemble into different oligomeric states. *Biochemistry* **42**: 13027-34.

CHEPURNOV, V.A., MANN, D.G., SABBE, K., VYVERMAN, W. (2004). Experimental studies on sexual seproduction in diatoms. *International Review of Cytology* **237**: 91-154.

CLARK, A. (1990). Inference of haplotypes from PCR-amplified samples of diploid populations. *Molecular Biology and Evolution* **7**: 111-122.

COHAN, F.M. (2002). What are bacterial species? *Annual Review of Microbiology* **56**: 457-87.

COLEMAN, A.W. (2003). ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* **19**: 370-375.

COYER, J.A., PETERS, A.F., HOARAU, G., STAM, W.T., OLSEN, J.L. (2002). Inheritance patterns of ITS1, chloroplasts and mitochondria in artificial hybrids of the seaweeds *Fucus serratus* and *F. evanescens* (Phaeophyceae). *European Journal of Phycology* **37**: 173-178.

COYER, J.A., SMITH, G.J., ANDERSEN, R.A. (2001). Evolution of *Macrocystis* spp. (Phaeophyceae) as determined by ITS1 and ITS2 sequences. *Journal of Phycology* **37**: 574-585.

CRACRAFT, J. (1989) Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In D Otte, JA Endler (eds) Speciation and its consequences. Sinauer, Sunderland, Massachusettes, pp 28-59.

DAGUIN, C., BONHOMME, F., BORSA, P. (2001). The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus *mac-1*. *Heredity* **86**: 342-354.

DAMSTE, J.S., MUYZER, G., ABBAS, B., RAMPEN, S.W., MASSE, G., ALLARD, W.G., BELT, S.T., ROBERT, J.M., ROWLAND, S.J., MOLDOWAN, J.M., BARBANTI, S.M., FAGO, F.J., DENISEVICH, P., DAHL, J., TRINDADE, L.A., SCHOUTEN, S. (2004). The rise of the rhizosolenid diatoms. *Science* **304**: 584-7.

DE SEVE, M.A. (1993). Diatom bloom in the tidal freshwater zone of a turbid and shallow estuary, Rupert Bay (James Bay, Canada). *Hydrobiologia* **269-270**: 225-233.

DE VARGAS, C., NORRIS, R., ZANINETTI, L., GIBB, S.W., PAWLOWSKI, J. (1999). Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proceedings of the National Academy of Sciences USA* **96**: 2864-2868.

DE VARGAS, C., SÁEZ, A.G., MEDLIN, L.K., and THIERSTEIN, H.R. (2004) Super-species in the calcareous plankton. In HR Thierstein, JR Young (eds) Coccolithophores. From molecular processes to global impact. Springer, Berlin; Heidelberg; New York, pp 271-298.

DESIKACHARY, T.V., RAO, V.N.R. (1973). Studies on *Cyclotella meneghiniana* Kütz. III. The frustule. *Proceedings of the Indian Academy of Sciences* **77B**: 78-91.

DOBZHANSKY, T. (1970). Genetics of the evolutionary process. Columbia University Press, New York.

DREBES, G. (1977) Sexuality. In D Werner (ed) The biology of diatoms. Blackwell Scientific Publications, Oxford, pp 250-283.

DROOP, S.J.M. (1995). A morphometric and geographical analysis of two races of Diploneis smithii / D. fusca (Bacillariophyceae) in Britain. *Proceedings of the Thirteenth International Diatom Symposium* : 347-369.

DRYDEN, I.L. and MARDIA, K.V. (1998). Statistical shape analysis. John Wiley & Sons, New York.

DU BUF, H. and BAYER, M. (2002). Automatic diatom identification. World Scientific Publishing, Singapore.

FINLAY, B.J., MONAGHAN, E.B., MABERLY, S.C. (2002). Hypothesis: the rate and scale of dispersal of freshwater diatom species is a function of their global abundance. *Protist* **153**: 261-73.

FRASER, D.J., BERNATCHEZ, L. (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology* **10**: 2741-2752.

FUNK, D.J., OMLAND, K.E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Reviews in Ecology, Evolution and Systematic* **34**: 397-423.

GALLAGHER, J.C. (1980). Population genetics of *Skeletonema costatum* (Bacillariophyceae) in Narragansett Bay. *Journal of Phycology* **16**: 464-474.

GEITLER, L. (1932). Der Formwechsel der pennaten Diatomeen (Kieselalgen). *Archiv Für Protistenkunde* **78**: 1-226.

GEITLER, L. (1975). Formwechsel, sippenspezifischer Paarungsmodus und Systematik bei einigen pennaten Diatomeen. *Plant Systematics and Evolution* **124**: 7-30.

GIORGIO, P., VINOCUR, A.L., LOMBARDO, R.J., TELL, H.G. (1991). Progressive changes in the structure and dynamics of the phytoplankton community along a pollution gradient in a lowland river - A multivariate approach. *Hydrobiologia* **224**: 129-154.

GOSSELAIN, V., DESCY, J.P., EVERBECQ, E. (1994). The phytoplankton community of the River Meuse, Belgium: Seasonal dynamics (year 1992) and the possible incidence of zooplankton grazing. *Hydrobiologia* **289**: 179-191.

HA, K., JANG, M., JOO, G. (2002). Spatial and temporal dynamics of phytoplankton communities along a regulated river system, the Nakdong River, Korea. *Hydrobiologia* **470**: 235-245.

HAKANSSON, H. (1998). Phenotypic plasticity in the diatom *Cyclotella meneghiniana* or a new species? *Nova Hedwigia* **66**: 187-196.

HAKANSSON, H., CHEPURNOV, V. (1999). A study of variation in valve morphology of the diatom *Cyclotella meneghiniana* in monoclonal cultures: effect of auxospore formation and different salinity conditions. *Diatom Research* **14**: 251-272.

HARRISON, R.G. (1998) Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In D Howard, S Berlocher (eds) Endless forms. Oxford University Press, pp 19-31.

HARTL, D.L. and CLARK, A.G. (1997). Principles of population genetics. Sinauer, Sunderland, Massachusetts.

HAUSMANN, S., LOTTER, A.F. (2001). Morphological variation within the diatom taxon *Cyclotella comensis* and its importance for quantitative temperature reconstructions. *Freshwater Biology* **46**: 1323-1333.

HECKY, R.E., KILHAM, P. (1973). Diatoms in alkaline, saline lakes: ecology and geochemical implications. *Limnology and Oceanography* **18**: 53-71.

HEY, J. (1997). Mitochondrial and nuclear genes present conflicting portraits of human origins. *Molecular Biology and Evolution* **14**: 166-72.

HOOPS, H.J., FLOYD, G.L. (1979). Ultrastructure of the centric diatom *Cyclotella meneghiniana*: vegetative cell and auxospore development. *Phycologia* **18**: 424-435.

HOUK, V., KLEE, R. (2004). The stelligeroid taxa of the genus *Cyclotella* (Kützing) Brébisson (Bacillariophyceae) and their transfer into the new genus *Discostella* gen. nov. *Diatom Research* **19**: 203-228.

HUDSON, R.R., COYNE, J.A. (2002). Mathematical consequences of the genealogical species concept. *Evolution International Journal of Organic Evolution* **56**: 1557-65.

IYENGAR, M.O.P., SUBRAHMANYAN, R. (1944). On reduction division and auxospore formation in *Cyclotella meneghiniana* Kütz. *The Journal of the Indian Botanical Society* **23**: 125-152.

JOHN, U., FENSOME, R.A., MEDLIN, L.K. (2003). The application of a molecular clock based on molecular sequences and the fossil record to explain biogeographic distributions within the *Alexandrium tamarense* "species complex" (Dinophyceae). *Molecular Biology and Evolution* **20**: 1015-27.

KESSLER, L.G., AVISE, J.C. (1985). A comparative description of mitochondrial DNA differentiation in selected avian and other vertebrate genera. *Molecular Biology and Evolution* **2**: 109-25.

KISS, K.T., ECTOR, L. (2000). Morphology and ultrastructure of *Cyclotella scaldensis* Muylaert & Sabbe: comparison between Moselle R. (Luxembourg) and Escaut Estuary (Belgium). *Cryptogamie: Algologie* **21**: 225-226.

KNOWLTON, N. (2000). Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* **420**: 73-90.

KOPCZYNSKA, E.E. (1979). Chloride effect on the growth of *Cyclotella meneghiniana* Kütz. and *Melosira granulata* (Ehr.) Ralfs. *Polar Archives of Hydrobiology* **26**: 587-594.

KORHOLA, A., BLOM, T. (1996). Marked early 20th century pollution and the subsequent recovery of Toeoeloe Bay, central Helsinki, as indicated by subfossil diatom assemblage changes. *Hydrobiologia* **341**: 169-179.

LAJEUNESSE, T.C. (2001). Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *Symbiodinium* using the ITS region: In search of a "species" level marker. *Journal of Phycology* **37**: 866-880.

LANGE-BERTALOT, H. (1990). Current biosytematic research on diatoms and its implications for the species concept. *Limnetica* **6**: 13-22.

LAPOINTE, M. (2000). Modern diatom assemblages in surface sediments from the Maritime Estuary and the Gulf of St. Lawrence, Quebec (Canada). *Marine Micropaleontology* **40**: 43-65.

LAZINSKY, D., SICKO-GOAD, L. (1990a). Morphometric analysis of phosphate and chromium interactions in *Cyclotella meneghiniana*. *Aquatic Toxicology* **16**: 127-140.

LAZINSKY, D., SICKO-GOAD, L. (1990b). Morphometric analysis of phosphate and chromium interactions in *Cyclotella meneghiniana*. *Aquatic Toxicology* **16**: 127-139.

LEWANDOWSKA, J., KOSAKOWSKA, A. (2004). Effect of iron limitation on cells of the diatom *Cyclotella meneghiniana* Kützing. *Oceanologia* **46**: 269-287.

LI, W.-H. (1997) Concerted evolution of multigene families. Molecular Evolution. Sinauer Ass., Sunderland, MA., pp 309-334.

LOHR, M., WILHELM, C. (2001). Xanthophyll synthesis in diatoms: quantification of putative intermediates and comparison of pigment conversion kinetics with rate constants derived from a model. *Planta* **212**: 382-91.

LUDWIG, W. and KLENK, H.P. (2001) Overview: a phylogenetic backbone and taxonomic framework for prokaryotic systematics. Bergey´s Manual of Systematic

Bacteriology. Springer, New York, pp 49-65.

LUNDHOLM, N., MOESTRUP, O. (2002). The marine diatom *Pseudo-nitzschia galaxiae* sp. nov. (Bacillariophyceae): Morphology and phylogenetic relationships. *Phycologia* **41**: 594-605.

LUNDHOLM, N., MOESTRUP, O., HASLE, G.R., HOEF-EMDEN, K. (2003). A study of the *Pseudo-nitzschia pseudodelicatissima*/*cuspidata* complex (Bacillariophyceae): What is *P. pseudodelicatissima*? *Journal of Phycology* **39**: 797-813.

MAJEWSKI, J., COHAN, F.M. (1999). Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. *Genetics* **152**: 1459-74.

MAKAREWICZ, J.C. (1987). Phytoplankton composition, abundance, and distribution: Nearshore Lake Ontario and Oswego River and harbor. *Journal of Great Lakes Research* **13**: 56-64.

MALLET, J. (1995). A species definition for the modern synthesis. *Trends in Ecology and Evolution* **10**: 294-299.

MANN, D.G. (1984). Observations on copulation in *Navicula pupula* and *Amphora ovalis* in relation to the nature of diatom species. *Annals of Botany* **54**: 429-438.

MANN, D.G. (1989). The species concept in diatoms: evidence for morphologically distinct, sympatric gamodemes in four epipelic species. *Plant Systematics and Evolution* **164**: 215-237.

MANN, D.G. (1993). Patterns of sexual reproduction in diatoms. *Hydrobiologia* **269-270**: 11-20.

MANN, D.G. (1999). The species concept in diatoms. *Phycologia* **38**: 437-495.

MANN, D.G., CHEPURNOV, V.A., DROOP, S.J.M. (1999). Sexuality, incompatibility, size variation, and preferential polyandry in natural populations and clones of Sellaphora pupula (Bacillariophyceae). *Journal of Phycology* **35**: 152-170.

MAYR, E. (1963). Animal species and evolution. Belknap Press, Cambridge, Massachusettes.

MEDLIN, L.K., ELWOOD, H.J., STICKEL, S., SOGIN, M.L. (1988). The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**: 491-499.

MES, T.H.M. (1998). Character compatibility of molecular markers to distinguish asexual and sexual reproduction. *Molecular Ecology* **7**: 1719-1727.

MILLER, D.J., VAN OPPEN, M.J.H. (2003). A 'fair go' for coral hybridization. *Molecular Ecology* **12**: 805-807.

MILLIE, D.F. (1984). The effects of nutrient limitation on the biochemical composition of *Cyclotella meneghiniana* Kuetz. *Journal of Phycology Supplement* **20**: 20.

MILLIE, D.F. (1986). Nutrient-limitation effects on the biochemical composition of

Cyclotella meneghiniana (Bacillariophyta): An experimental and statistical analysis. *Canadian Journal of Botany* **64**: 19-26.

MILLIE, D.F., MARK HERSH, C. (1987). Statistical characterizations of the atrazine-induced photosynthetic inhibition of *Cyclotella meneghiniana* (Bacillariophyta). *Aquatic Toxicology* **10**: 239-249.

MONTRESOR, M., SGROSSO, S., PROCACCINI, G., KOOISTRA, W. (2003). Intraspecific diversity in *Scrippsiella trochoidea* (Dinophyceae): Evidence for cryptic species. *Phycologia* **42**: 56-70.

MORJAN, C.L., RIESEBERG, L.H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**: 1341-56.

MOSS, B., BALLS, H. (1989). Phytoplankton distribution in floodplain lake and river system. 2. Seasonal changes in the phytoplankton communities and their control by hydrology and nutrient availability. *Journal of Plankton Research* **11**: 839-867.

MOU, D., STOERMER, E.F. (1992). Separating Tabellaria (Bacillariophyceae) shape groups based on Fourier descriptors. *Journal of Phycology* **28**: 386-395.

MUIR, G., FLEMING, C.C., SCHLÖTTERER, C. (2001). Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. *Molecular Biology and Evolution* **18**: 112-9.

MUIR, G., FLEMING, C.C., SCHLÖTTERER, C. (2000). Species status of hybridizing oaks. *Nature* **405**: 1016.

MUYLAERT, K., SABBE, K. (1996). Cyclotella scaldensis spec. nov. (Bacillariophyceae), a new estuarine diatom. *Nova Hedwigia* **63**: 335-345.

MUYLAERT, K., SABBE, K. (1999). Spring phytoplankton assemblages in and around the maximum turbidity zone of the estuaries of the Elbe (Germany), the Schelde (Belgium/The Netherlands) and the Gironde (France) . *Journal of Marine Systems* **22**: 133-149.

NIXON, K.C., WHEELER, Q.D. (1990). An amplification of the phylogenetic species concept. *Cladistics* **6**: 211-223.

NORDBORG, M., INNAN, H. (2002). Molecular population genetics. *Current Opinion in Plant Biology* **5**: 69-73.

ORSINI, L., PROCACCINI, G., SARNO, D., MONTRESOR, M. (2004). Multiple rDNA ITS-types within the diatom *Pseudo-nitzschia delicatissima* (Bacillariophyceae) and their relative abundances across a spring bloom in the Gulf of Naples. *Marine Ecology Progress Series* **271**: 87-98.

ORSINI, L., SARNO, D., PROCACCINI, G., POLETTI, R., DAHLMANN, J., MONTRESOR, M. (2002). Toxic *Pseudo-nitzschia multistriata* (Bacillariophyceae) from the Gulf of Naples: morphology, toxin analysis and phylogenetic relationships with other *Pseudo-nitzschia* species. *European Journal of Phycology* **37**: 247-257.

ORTIZ-BARRIENTOS, D., REILAND, J., HEY, J., NOOR, M.A. (2002). Recombination and the divergence of hybridizing species. *Genetica* **116**: 167-78.

PAPPAS, J.L., FOWLER, G.W., STOERMER, E.F. (2001). Calculating shape descriptors from Fourier analysis: shape analysis of *Asterionella* (Heterokontophyta, Bacillariophyceae). *Phycologia* **40**: 440-456.

PATERSON, H.E.H. (1985) The recognition concept of species. In ES Vrba (ed) Species and speciation. Transvaal Museum, Pretoria, pp 21-29.

QUIJADA, A., LISTON, A., ROBINSON, W., ALVAREZ-BUYLLA, E. (1997). The ribosomal ITS region as a marker to detect hybridization in pines. *Molecular Ecology* **6**: 995-996.

RAO, V.N.R. (1970). Studies on *Cyclotella meneghiniana* Kütz. I. Sexual reproduction and auxospore formation. *Proceedings of the Indian Academy of Sciences* **72**: 285-287.

RAO, V.N.R. (1971). Studies on *Cyclotella meneghiniana* Kütz. II. Induction of auxospore formation. *Phykos* **10**: 84-98.

RAO, V.N.R. (1996). Size dependent reproductive behaviour in *Cyclotella meneghiniana* (Bacillariophyta). *Beiheft Zur Nova Hedwigia* **112**: 235-238.

RAO, V.N.R., DURAISAMY, A., KANNAN, V. (1983). Growth behaviour of some diatoms to changes in the concentration of K, Ca, and Mg of the medium. *Phykos* **22**: 136-140.

REIMANN, B.E.F., LEWIN, J.M.C., GUILLARD, R.R.L. (1963). *Cyclotella cryptica*, a new brackish-water diatom species. *Phycologia* **3**: 75-84.

REYMENT, R.A., BLACKITH, R.E., and CAMPBELL, N.A. (1984). Multivariate Morphometrics. Academic Press, New York.

RHODE, K.M., PAPPAS, J.L., STOERMER, E.F. (2001). Quantitative analysis of shape variation in type and modern populations of Meridion (Bacillariophyceae). *Journal of Phycology* **37**: 175-183.

RIGINOS, C., CUNNINGHAM, C.W. (2005). Local adaptation and species segregation in two mussel (*Mytilus edulis* x *Mytilus trossulus*) hybrid zones. *Molecular Ecology* **14**: 381-400.

RIJSTENBIL, J.W., BAKKER, C., JACKSON, R.H., MERKS, A.G.A., DE VISSCHER, P.R.M. (1993). Spatial and temporal variation in community composition and photosynthetic characteristics of phytoplankton in the upper Westerschelde Estuary (Belgium, SW Netherlands). *Hydrobiologia* **269-270**: 263-273.

ROHLF, F.J., MARCUS, L.F. (1993). A revolution in morphometrics. *Trends in Ecology and Evolution* **8**: 129-132.

ROJO, C., ORTEGA-MAYAGOITIA, E., RODRIGO, M.A., ALVAREZ-COBELAS, M. (2000). Phytoplankton structure and dynamics in a semiarid wetland, the National Park "Las Tablas de Daimiel" (Spain). *Archiv Für Hydrobiologie* **148**: 397-419.

ROMO, S., MIRACLE, M.R. (1994). Population dynamics and ecology of subdominant phytoplankton species in a shallow hypertrophic lake (Albufera of Valencia, Spain). *Hydrobiologia* **273**: 37-56.

ROSENBERG, N.A., NORDBORG, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3**: 380-90.

ROUND, F.E., CRAWFORD, R.M., and MANN, D.G. (1990). The diatoms: Biology and morphology of the genera. Cambridge University Press, Cambridge (UK) .

RYNEARSON, T.A., ARMBRUST, E.V. (2004). Genetic differentiation among populations of the planktonic marine diatom *Ditylum brightwellii* (Bacillariophyceae). *Journal of Phycology* **40**: 34-43.

SABATER, S. (1990). Composition and dynamics of a highly diverse diatom assemblage in a limestone stream. *Hydrobiologia* **190**: 43-53.

SABATER, S., SABATER, F. (1988). Diatom assemblages in the River Ter. *Archiv Für Hydrobiologie* **111** : 397-408.

SÁEZ, A.G., LOZANO, E. (2005). Body doubles. *Nature* **433**: 111.

SANG, T., CRAWFORD, D.J., STUESSY, T.F. (1995). Documentation of reticulate evolution in peonies (*Paeonia*) using internal transcribed spacer sequences of nuclear ribosomal DNA: implications for biogeography and concerted evolution. *Proceedings of the National Academy of Sciences of the USA* **92**: 6813-7.

SARNO, D., KOOISTRA, W.H.C.F., MEDLIN, L.K., PERCOPO, I., ZINGONE, A. (2005). Diversity in the genus *Skeletonema* (Bacillariophyceae). II. An assessment of the taxonomy of *S. costatum*-like species with the description of four new species. *Journal of Phycology* **in press**: DOI: 10.1111/j.1529-8817.2005.04067.x.

SCHARDL, C.L., CRAVEN, K.D. (2003). Interspecific hybridization in plant-associated fungi and oomycetes: a review. *Molecular Ecology* **12**: 2861-73.

SCHLÖTTERER, C. (2002). Towards a molecular characterization of adaptation in local populations. *Current Opinion in Genetics and Development* **12**: 683-7.

SCHLÖTTERER, C. (2003). Hitchhiking mapping - functional genomics from the population genetics perspective. *Trends in Genetics* **19**: 32-8.

SCHLÖTTERER, C., HAUSER, M.T., VON HAESELER, A., TAUTZ, D. (1994). Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. *Molecular Biology and Evolution* **11**: 513-22.

SCHLÖTTERER, C., TAUTZ, D. (1994). Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Current Biology* **4**: 777-83.

SCHOEMAN, F.R., ARCHIBALD, R.E.M. (1980). The diatom flora of Southern Africa. *National Institute for Water Research, Council for Scientific and Industrial Research, Special Report* : 34 pp.

SCHOLIN, C.A., HALLEGGRAEFF, G.M., ANDERSON, D.M. (1995). Molecular evolution of the *Alexandrium tamarense* 'species complex' (Dinophyceae): dispersal in the North and West Pacific Regions. *Phycologia* **34**: 472-485.

SCHOLIN, C., HERZOG, M., SOGIN, M., ANDERSON, D. (1994). Identification of group- and strain-specific genetic markers for globally distributed *Alexandrium* (Dinophyceae). 2. Sequence analysis of a aragment of the LSU rRNA gene. *Journal of Phycology* **30**: 999-1011.

SCHULTZ, M.E. (1971). Salinity-related polymorphism in the brackish-water diatom *Cyclotella cryptica*. *Canadian Journal of Botany* **49**: 1285-1289.

SCHULTZ, M.E., TRAINOR, F.R. (1968). Production of male gametes and auxospores in the centric diatoms *Cyclotella meneghiniana* and *C. cryptica*. *Journal of Phycology* **4**: 85-88.

SCHULTZ, M.E., TRAINOR, F.R. (1970). Production of male gametes and auxospores in a polymorphic clone of the centric diatom *Cyclotella*. *Canadian Journal of Botany* **48**: 947-951.

SEELIGMANN, C.T., TRACANNA, B.C. (1994). Limnology of El Cadillal Reservoir (Tucuman, Argentina). 2: Qualitative study of phytoplankton. *Cryptogamie: Algologie* **15**: 19-35.

SHIN, J.K., CHO, K.J. (1999). Diurnal variation of environmental factors and freshwater algae in the Naktong River Estuary. *Korean Journal of Limnology* **32**: 341-348.

SICKO-GOAD, L., EVANS, M.S., LAZINSKY, D., HALL, J., SIMMONS, M.S. (1989a). Effects of chlorinated benzenes on diatom fatty acid composition and quantitative morphology. IV. Pentachlorobenzene and comparison with trichlorobenzene isomers. *Archives of Environmental Contamination and Toxicology* **18**: 656-68.

SICKO-GOAD, L., HALL, J., LAZINSKY, D., SIMMONS, M.S. (1989b). Effects of chlorinated benzenes on diatom fatty acid composition and quantitative morphology. III. 1,2,3-Trichlorobenzene. *Archives of Environmental Contamination and Toxicology* **18**: 647-55.

SICKO-GOAD, L., HALL, J., LAZINSKY, D., SIMMONS, M.S. (1989c). Effects of chlorinated benzenes on diatom fatty acid composition and quantitative morphology. II. 1,3,5-Trichlorobenzene. *Archives of Environmental Contamination and Toxicology* **18**: 638-46.

SICKO-GOAD, L., LAZINSKY, D., HALL, J., SIMMONS, M.S. (1989d). Effects of chlorinated benzenes on diatom fatty acid composition and quantitative morphology. I. 1,2,4-Trichlorobenzene. *Archives of Environmental Contamination and Toxicology* **18**: 629-37.

SQUIRES, L.E., RUSHFORTH, S.R. (1986). Winter phytoplankton communities of Utah Lake, Utah, USA. *Hydrobiologia* **131**: 235-248.

STOERMER, E.F., LADEWSKI, T.B. (1982). Quantitative analysis of shape variation in type and modern populations of *Gomphoneis herculeana*. *Nova Hedwigia Beihefte* **73**: 347-373.

STOERMER, E.F. and SMOL, J.F. (1999). The diatoms. Applications for the

environmental and earth sciences. Koeltz Scientific Books, Königstein, Germany.

STOERMER, E.F., YU-ZAO, Q., LADEWSKI, T.B. (1986). A quantitative investigation of shape variation in *Didymosphenia* (Lyngbye) M. Schmidt (Bacillariophyta). *Phycologia* **25**: 494-502.

STORZ, J.F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* **14**: 671-88.

STREIT, B., STADLER, T., SCHWENK, K., ENDER, A., KUHN, K., SCHIERWATER, B. (1994). Natural hybridization in freshwater animals. Ecological implications and molecular approaches. *Naturwissenschaften* **81**: 65-73.

SÁEZ, A.G., PROBERT, I., GEISEN, M., QUINN, P., YOUNG, J.R., MEDLIN, L.K. (2003). Pseudo-cryptic speciation in coccolithophores. *Proceedings of the National Academy of Sciences USA* **100**: 7163-7168.

TABERLET, P., GIELLY, L., PAUTOU, G., BOUVET, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* **17**: 1105-1109.

TAPIA, P.M., FRITZ, S.C., BAKER, P.A., SELTZER, G.O., DUNBAR, R.B. (2003). A Late Quaternary diatom record of tropical climatic history from Lake Titicaca (Peru and Bolivia). *Palaeogeography, Palaeoclimatology, Palaeoecology* **194** : 139-164.

TELLERIA, J., BARNABE, C., HIDE, M., BANULS, A.L., TIBAYRENC, M. (2004). Predominant clonal evolution leads to a close parity between gene expression profiles and subspecific phylogeny in *Trypanosoma cruzi*. *Molecular and Biochemical Parasitology* **137**: 133-41.

TEMPLETON, A.R. (1989) The meaning of species and speciation: a genetic perspective. In D Otte, JA Endler (eds) Speciation and its consequences. Sinauer, Sunderland, Massachusettes, pp 3-27.

TEUBNER, K. (1995). A light microscopical investigation and multivariate statistical analyses of heterovalvar cells of *Cyclotella*-species (Bacillariophyceae) from lakes of the Berlin-Brandenburg region. *Diatom Research* **10**: 191-205.

THERIOT, E. (1987). Principal component analysis and taxonomic interpretation of environmentally related variation in silicification in *Stephanodiscus* (Bacillariophyceae). *British Phycological Journal* **22**: 359-373.

THERIOT, E., HAKANSSON, H., STOERMER, E.F. (1988). Morphometric analysis of *Stephanodiscus alpinus* (Bacillariophyceae) and its morphology as an indicator of lake trophic status. *Phycologia* **27**: 485-493.

THERIOT, E., STOERMER, E.F. (1984a). Principal component analysis of *Stephanodiscus*: observations on two new species from the *Stephanodiscus niagarae* complex. *Bacillaria* **7**: 37-58.

THERIOT, E., STOERMER, E.F. (1984b). Principal component analysis of variation in *Stephanodiscus rotula* and *S. niagarae* (Bacillariophyceae). *Systematic Botany* **9**: 53-59.

TIBAYRENC, M. (1995). Population genetics of parasitic protozoa and other microorganisms. *Advances in Parasitology* **36**: 47-115.

TIBAYRENC, M., AYALA, F.J. (1987). *Trypanosoma cruzi* populations: More clonal than sexual. *Parasitology Today* **3**: 189-90.

TIBAYRENC, M., AYALA, F.J. (1991). Towards a population genetics of microorganisms: The clonal theory of parasitic protozoa. *Parasitology Today* **7**: 228-32.

TIBAYRENC, M., AYALA, F.J. (2002). The clonal theory of parasitic protozoa: 12 years on. *Trends in Parasitology* **18**: 405-10.

TILMAN, D., KILHAM, P. (1976a). Sinking in freshwater phytoplankton: some ecological implications of cell nutrient status and physical mixing processes. *Limnology and Oceanography* **21**: 409-417.

TILMAN, D., KILHAM, S.S. (1976b). Phosphate and silicate growth and uptake kinetics of the diatoms *Asterionella formosa* and *Cyclotella meneghiniana* in batch and semicontinuous culture. *Journal of Phycology* **12**: 375-383.

TRAINOR, F.R. (1991). *Scenedesmus* plasticity: Facts and hypotheses. *Journal of Phycology* **27**: 555-556.

TRAINOR, F.R., EGAN, P.F. (1991). Discovering the various ecomorphs of *Scenedesmus* : The end of a taxonomic era. *Archiv Für Protistenkunde* **139**: 125-132.

TUCHMAN, M.L., THERIOT, E., STOERMER, E.F. (1984). Effects of low level salinity concentrations on the growth of *Cyclotella meneghiniana* Kütz. (Bacillariophyta). *Archiv Für Protistenkunde* **128**: 319-326.

URDANETA, L., LAL, A., BARNABE, C., OURY, B., GOLDMAN, I., AYALA, F.J., TIBAYRENC, M. (2001). Evidence for clonal propagation in natural isolates of *Plasmodium falciparum* from Venezuela. *Proceedings of the National Academy of Sciences of the USA* **98**: 6725-9.

VAN DAM, H. (1979). Diatoms and water quality in lowland streams in the province of Northern Brabant (The Netherlands). *Hydrobiological Bulletin* **13**: 13-21.

VAN HANNEN, E.J., LUERLING, M., VAN DONK, E. (2000). Sequence analysis of the ITS-2 region: A tool to identify strains of *Scenedesmus* (Chlorophyceae). *Journal of Phycology* **36**: 605-607.

WENDEL, J.F., SCHNABEL, A., SEELANAN, T. (1995a). Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences of the USA* **92**: 280-4.

WENDEL, J.F., SCHNABEL, A., SEELANAN, T. (1995b). An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic, intergenomic introgression. *Molecular Phylogenetics and Evolution* **4**: 298-313.

WHITE, T.J., BRUNS, T., LEE, S., and TAYLOR, J. (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In MA Innis, DH Gelfland, JJ Sninsky, TJ White (editors) PCR protocols. Guide to methods and application. Academic

Press, San Diego, pp 315-322.

WIENS, J.J., PENKROT, T.A. (2002). Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Systematic Biology* **51**: 69-91.

WILDING, C.S., BUTLIN, R.K., GRAHAME, J. (2001). Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* **14**: 611-619.

WILEY, E.O. (1978). The evolutionary species concept reconsidered. *Systematic Zoology* **27**: 17-26.

WU, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**: 851-865.

WU, C.I., TING, C.T. (2004). Genes and speciation. *Nature Reviews Genetics* **5**: 114-22.

ZALOCAR DE DOMITROVIC, Y. (1992). Phytoplankton of flooded environments of the Parana River (Argentina). *Revue D'Hydrobiologie Tropicale* **25** : 177-188.

ZECHMAN, F.W., ZIMMER, E.A., THERIOT, E.C. (1994). Use of ribosomal DNA internal transcribed spacers for phylogenetic studies in diatoms. *Journal of Phycology* **30**: 507-512.

ZHANG, D.-X., HEWITT, G.M. (2003). Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**: 563-584.

7. Acknowledgements

First I would like to say thanks to Linda Medlin for the possibility to carry out this work in her group. Extra special thanks to her for her support to participate in training courses, which gave me the possibility to learn a lot. Furthermore, I´d like to thank Prof. Dr. rer. nat. Gunter O. Kirst and Prof. Dr. rer. nat. Ulrich Bathmann for reviewing this dissertation. Thanks to Kerstin Töbe, Richard Crawford, Uwe John and Dirk Mengedoht for reading and helpful criticism of this work. Thanks to Éva Ács for motivating me to start molecular research of diatoms, and for pointing out *C. meneghiniana* as a problematic species worth studying.

Thanks also to several extant and former members of the AG Medlin and other groups – especially the AG Mensa – for the friendly atmosphere, help and advice throughout this work: Alberto, Andrea, Andreas, Christine, Dick, Friedel, Gundula, Helga, Henrik, Ines, Jessica, Katrin, Katja, Kerstin, Klaus, Madda, Monica, Niko, Olaf, René, Sabine, Shinya, Sonja, Steffi, Stephan, Susanne, Thomas, Uwe.

Special thanks: to Alberto for discussions and illuminating reviews of different writings; to Katrin, Sabine, Olaf, and Ines for our work together in Algaterra; to Klaus for help with several things of life from Ausländerbehörde to ZwoNullFünf; René for guidance in Bremerhaven and for reading and constructive criticisms of manuscripts; Thomas (and Claudia and Paul) for sharing an enthusiasm for jazz; Uwe for his constant attention, the discussions, and for motivation in several, hm, deeps.

Diese Arbeit entsand in Rahmen des BMBF Porjektes Algaterra 01LC0026 und durch die Förderung der Stiftung Alfred-Wegener-Institut für Polar- und Meeresforschung.