Genetic Diversity and Cryptic Speciation in Planktonic Foraminiferal Morphotaxa

Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften

der Geowissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

vorgelegt von
Ralf Aurahs
aus Saarbrücken

2010

Tag der Disputation: 20. April 2010

Dekan: Prof. Dr. Peter Grathwohl

1. Berichterstatter: Prof. Dr. Michal Kučera
Institut für Geowissenschaften, Tübingen

2. Berichterstatter: Prof. Dr. Ralf Schiebel
Laboratoire des Bio- Indicateurs Actuels
et Fossiles, Université d'Angers

Content

Authors contributions to the publications 1- 3

1. Using the Multiple Analysis Approach to Reconstruct Phylogenetic Relationships among Planktonic Foraminifera from Highly Divergent and Length-polymorphic SSU rDNA Sequences.
Ralf Aurahs, Markus Göker, Guido W. Grimm, Vera Hemleben, Christoph Hemleben, Ralf Schiebel and Michal Kučera (2009), Bioinformatic Biology Insight, **3**, 155–177.

Own contributions:

Idea/Hypothesis	40%
Data acquisition	70%
Interpretation	25%
Writing	60%

2. Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber.*
Ralf Aurahs, Guido W. Grimm, Vera Hemleben, Christoph Hemleben and Michal Kučera (2009), Molecular Ecology, **18**, 1692–1706.

Idea/Hypothesis	80%
Data acquisition	100%
Interpretation	80%
Writing	90%

3. A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence.
Ralf Aurahs, Yvonne Treis, Michal Kučera,
To be submitted to Marine Micropaleontology.

Idea/Hypothesis	75%
Data acquisition	80%
Interpretation	90%
Writing	80%

Zusammenfassung

Die rezenten Vertreter planktonischer Foraminiferen werden anhand morphologischer Merkmale ihrer kalzitischen Schale in etwa 40-50 Arten unterteilt. Abstammung und Verwandtschaftsbeziehungen innerhalb der planktonischen Foraminiferen sind durch den außergewöhnlich umfangreichen Fossilbefund dieser Protistengruppe sehr gut untersucht. Anhand molekulargenetischer Untersuchungen des Gens, welches für die RNA der 30S Untereinheit des Ribosoms kodiert (die *small subunit ribosomal DNA,* kurz SSU rDNA), konnten die phylogenetischen Verwandtschaftsverhältnisse moderner Foraminiferen in weiten Teilen bestätigt werden. Das untersuchte Fragment der SSU rDNA stellt ein Mosaik aus konservativen, variablen und hochvariablen Sequenzbereichen dar. Die phylogenetische Information aus den konservierten Regionen zeigt eine gute bis sehr gute Übereinstimmung mit den, aus dem Fossilbefund abgeleiteten, verwandtschaftlichen Beziehungen innerhalb der einzelnen Großgruppen (Spinose, Makroperforate und Mikroperforate non-spinose), wie auch in der Mehrzahl der rezenten Arten untereinander. Anhand der variablen und hochvariablen Bereiche wiederum lassen sich innerhalb verschiedener Morphospezies wenige bis mehrere Genotypen voneinander unterscheiden.Wie bei vielen anderen rein morphologisch beschriebenen Arten auch, finden sich mit den Methoden der molekularen Genetik eine hohe Anzahl genetisch distinkter aber morphologisch identischer (oder nahezu identischer) Typen. Aufgrund der sehr langsamen Evolutionsrate der SSU rDNA liegt die Vermutung nahe, dass es sich bei diesen genetischen Typen um eigenständige, ‚kryptische' Arten handelt.

Ziel der vorliegenden Dissertation war es, anhand neuer SSU rDNA Sequenzen die Phylogenie der planktonischen Foraminiferen zu erweitern, die genetische Diversität in ausgewählten Arten zu untersuchen und deren geographische und jahreszeitliche Verteilung aufzunehmen. Mit den Resultaten dieser Untersuchungen soll die Grundlage für eine morphometrisch quantifizierbare Unterscheidung verschiedener Genotypen innerhalb einer Morphospezies erarbeitet werden.

Die Verwendung automatisierter multipler Alignments für die Rekonstruktion der verwandtschaftlichen Verhältnisse planktonischer Foraminiferen ermöglicht eine vollständige Reproduzierbarkeit der Ergebnisse. Die Resultate dieser Methode sind vergleichbar mit der traditioneller manueller Alignments; im Gegensatz zu letzteren sind diese aber vollständig objektiv und unter deutlich geringerem Zeitaufwand erstellbar. Die Methode der multiplen Alignments ist damit eine wichtige Vorrausetzung für eine kontinuierlich und objektiv erweiterbare Phylogenie der planktonischen Foraminiferen, vor allem in Bezug auf die Erkennung und Klassifizierung neuer SSU rDNA Genotypen.

Die Diversität, Häufigkeit und Verbreitung von SSU rDNA Genotypen der spinosen Art *Globigerinoides rube*r wurde im östlichen Atlantischen Ozean und im Mittelmeer untersucht. Die Verteilung der fünf gefundenen Genotypen (Pink, Ia, IIa1, IIa2, IIb) zeigt sowohl saisonale als auch geographische Unterschiede in der relativen Häufigkeit der einzelnen genetischen Typen. Ein hier erstmals beschriebener Genotyp (IIb) scheint nach diesen Ergebnissen im Mittelmeer endemisch zu sein. Zwei sehr nah verwandte Genotypen, IIa1 und IIa2, die im Kanarenstrom zusammen vorkamen, zeigten im Mittelmeer eine strikte Trennung auf je eines der beiden Hauptbecken. Die zeitliche und räumliche Verteilung der Genotypen in *G. ruber* deutet darauf hin, dass genetische Schwester-Typen um eine ökologische Nische konkurrieren und sich in stabilen Habitaten ausschließen.

Anhand des digitalen Bildmaterials der Schalen von *G. ruber* konnten erstmals direkte morphometrische Vergleichsmessungen zwischen Genotypen dieser Art ausgeführt werden. Im Abgleich mit Messdaten an Schalen aus rezenten Sedimenten, Museen und der neueren Literatur wurde der Grad der Kompression in der letzten und vorletzten Kammer untersucht. Die Ergebnisse zeigen, dass sich bereits anhand dieser zwei Faktoren die Gruppen von *G. ruber* im engeren Sinne und *G. ruber* im weiteren Sinne voneinander unterscheiden lassen. Eine Kombination von Molekularer Uhr und den Erkenntnissen aus dem fossilen Befund unterstützt die Ergebnisse der morphometrischen Analysen. Die Art *Globigerinoides ruber* enthält demnach mindestens zwei differenzierte Arten, von denen eine mit der morphologischen Definition der Art *Globigerinoides elongatus* übereinstimmt.

Zusammenfassend sind die Erkenntnisse über die genetische Diversität in *Globigerinoides ruber* ein deutlicher Hinweis dafür, dass das morphologische Artkonzept in planktonischen Foraminiferen auf dem Niveau der biologischen Art in weiten Teilen nicht ausreichend präzise ist.

Abstract

Modern planktonic foraminifera comprise 40-50 species that are defined by morphologic characters of their calcite shells. Due to their exceptionally good fossil record, the phylogenetic relationships in this group are well established. From the analyses of a gene fragment coding for the RNA of the ribosomal small subunit (the SSU rDNA), most of the fossil phylogeny was confirmed, but incomplete taxonomic sampling and large differences in the rate of molecular evolution leave several key evolutionary events unresolved. Further, within several morphospecies, up to seven distinct SSU rDNA genotypes were found. These morphologically inseparable but genetically clearly diverged genetic types are thought to represent cryptic species. The aims of this dissertation were *i)* to augment the SSU rDNA based phylogeny of planktonic foraminifera with new genetic data and new computational methods, *ii)* to analyse the genetic diversity in selected species and their geographical and seasonal distribution and *iii)* to evaluate possible correlations between distinct genotypes and certain phenotypes reported whitin morphospecies.

Traditionally, alignments of planktonic foraminiferal SSU rDNA sequences are manually checked for nucleotide homology and sites where this homology be established discarded. In a new approach tested within this thesis, phylogenetic reconstructions on the basis of unreduced, automated alignments were shown to resolve the phylogeny of this group equally good or even better than manually culled alignments. The multiple alignment approach generates multiple alternative phylogenetic topologies in a time-efficient and objective manner. The topologies from these automated alignments are fully reproducible and represent an important step towards the application of SSU rDNA sequences in a genetic taxonomy of planktonic foraminifera.

In the morphospecies *Globigerinoides ruber*, the diversity, abundance and distribution of five SSU rDNA genotypes in individuals from the Northeast Atlantic Ocean and Mediterranean Sea has been monitored. A newly described genotype (IIb) was found to be endemic in the Mediterranean Sea. Individuals of *G. ruber* pink corresponded to a single genetic type, namely Type Pink. Two genetic subtypes, IIa1 and IIa2 co-occurred in the Canary Current, yet each of the types was found exclusively in one of the Mediterranean basins. The spatial and temporal distribution of the genotypes of *G. ruber* indicate that closely related genotypes compete for very similar niches and that niche-partitioning between these genotypes under stable habitat conditions results in a pattern of mutual exclusion.

In a bid to understand the link between the genetic diversity in *G. ruber* and the morphological variability within this species, morphological characters of shells from genotyped individuals, recent sediments, museum collections and the recent literature have been compared. The statistical analysis of these morphometric data shows that the degree of compression of the ultimate and penultimate chambers can be used to separate the genotypes IIa, Pink and Ib from each other phenotypically. The same degree of morphometric divergence was found between individuals of *G. ruber* and those identified in the past as *G. elongatus* from museum material. In combination with results from molecular clocks based on multiple assumptions of divergence dates from the fossil record, our data strongly suggest that the definition of *G. ruber* as it is used today comprises at least 3 different extant species and an extinct lineage in the early to middle Miocene. One of the extant species, represented by the genetic type IIa, corresponds to the species definition of *G. elongatus*.

Introduction

"... I was much struck how entirely vague and arbitrary is the distinction between species and varieties." Charles Darwin 1859

Planktonic foraminifera are marine protists with ornate, multichambered calcareous shells, found in large number in marine sediments since the Jurassic (~ 170 Ma). Modern planktonic foraminifera comprise 40-50 morphospecies. Some species are abundant in the oceanic plankton and widely distributed in the world oceans; others are restricted to single ocean basins or specific oceanographic provinces (e.g. Hemleben et al. 1989). As each species is adapted to a specific habitat, the distribution of planktonic foraminiferal species in the oceans and their relative abundance in a foraminiferal assemblage is driven by environmental factors such as water temperature and food availability (e.g. Hemleben et al. 1989). The outstanding preservation of foraminiferal shells in marine sediments is therefore not only of great benefit for the study of evolutionary processes in planktonic species and for the stratigraphic dating of sediments, but also for the reconstruction of paleo-ecological properties of past ocean water masses (e.g. Kucera 2007).

Given the variety of different habitats that planktonic foraminiferal species are adapted to, the degree of precision resulting from their applications fundamentally depends, irrespective of the technical approach, on the correct identification of these species. Therefore, the main aim of micropaleontological morphotaxonomy was to establish easily recognizable taxa that provide a reasonable resolution for paleoceanographic and stratigraphic studies. In the process, the morphological definitions of species were chosen relatively robust, i.e. with a wide range of possible phenotypical variations that were believed to be induced by environmental factors (e.g. Kennett 1976).

As a consequence of the increasing number of scientists working on the subject, as well as of the highly different quality of preservation of recent and fossil planktonic foraminiferal material, the taxonomic value of the various shell features and their phenotypical modifications were interpreted inconsistently among workers (e.g. Saito et al. 1981; Kenneth & Srinivasan 1983). Whereas some researchers used slightest variations in specific shell structures to define species from both modern and fossil lineages (e.g. Saito et al. 1981), these “splitters” were opposed by a more conservative fraction of “lumpers”, who defended their more integrative species definitions with the well-established phenomenon of phenotypic plasticity (e.g. Parker 1962). These disagreements in the interpretation of morphological traits led not only to some confusion about the taxonomical status of several species, but also about the number of modern and fossil genera and the species they comprise (e.g. Moullade 1964).

Supported by SEM observations of shell ultrastructure, combined with a revision of the (vast) taxonomic literature available, a general consensus on the number and relation between extant planktonic foraminifera was established in the 80’s by publications of Kenneth and Srinivasan (1983) and Hemleben et al. (1989). Their relatively conservative interpretation of morphological variability in planktonic foraminifera limited the number of extant species to about 50, divided into three major groups: spinose, non-spinose macroperforate and non-spinose microperforate (Hemleben et al. 1989). The planktonic foraminifera in these groups are reported to possess a bilamellar shell ultrastructure (Hemleben et al. 1989) even though this is not established beyond doubt for several microperforate species (C. Hemleben, pers. comm. 2009). The only two extant species reported to build monolamellar

shells, the "spinose" *Hastigerina pelagica* and *Hastigerinella digitata* are classified separately in the family Hastigerinidae (e.g. Hemleben et al. 1989; Fig. 1).

Besides the planktonic foraminiferal taxa with spiral shell growth (planispiral in Hastigerinidae, trochospiral in all other groups; e.g. Bé 1967) there are also two species of foraminifera with serial chamber arrangement that are assigned to the planktonic forms, *Gallitellia vivans* (e.g. Kroon & Nederbragt 1990) and *Streptochilus globigerus* (e.g. Hemleben et al. 1989). These two species resemble a morphology more common among modern benthic foraminifera, and recent genetic and shell chemistry data suggest that these species indeed derive from the benthos, but possess a planktonic life stage (Darling et al. 2009; Kimoto et al. 2009).

As mentioned above, planktonic foraminifera offer a great opportunity for understanding planktonic speciation and evolution on a large spatial and temporal scale. The exceptionally good shell preservation, in combination with their high abundance in the sediments, provides unique insights on important aspects of speciation as, for example, taxon duration, ancestor-descendants relationship, as well as species origination dates and localities on paleontological timescales (e.g. Hills & Thierstein 1989; Lazarus et al. 1995). Yet, even the phylogenetic history of the extant species, mainly recovered by the stratophenetic tracing of first appearance of species (Fig. 2) and their specific shell features (e.g. Wade 1964; Wei 1994) are still subject to scientific debate.

1. Interpretation of the fossil record and the origin of extant planktonic foraminiferal lineages

Cifelli and Scott (1986) pointed out that references on the continuous and excellent fossil record of planktonic foraminifera also call for comments on the quality and sensibility of the methods used to resolve it. Researchers have used morphological traits of varying phylogenetic value, sometimes resulting in highly incompatible phylogenetic hypotheses (e.g. Saito et al. 1981; Kennett & Srinivasan 1983). The iterative appearance of (taxonomically and phylogenetically relevant) shell features in relatively unrelated foraminiferal lineages, in combination with their sometimes high level of plasticity, is a considerable obstacle for the reconstruction of the fossil phylogeny of modern lineages (e.g. Norris 1991).

In this regard, some characteristics of the modern planktonic foraminiferal shells wall structure seem to be relatively stable through time and features as spines, shell ultrastructure and pore size have been found to combine only into the three major groups mentioned above (Fig 1; Olson et al. 1999; Pearson et al. 2006).

The bilamellar spinose and macroperforate non-spinose taxa are considered to share a common ancestor from the genus *Hedbergella* (Liu & Olsson 1994; Olsson et al. 1999). As the earliest spinose species is believed to have developed from a survivor of the Cretaceous-Tertiary extinction, *Hedbergella monmouthensis* (Liu & Olsson 1992), and the macroperforate non-spinose are hypothesized to originate from another survivor species of the Cretaceous-Tertiary extinction, *Hedbergella holmdelensis* (Pearson et al. 2006), the divergence between these two groups is dated to the very late Cretaceous, 70-65 million years ago.

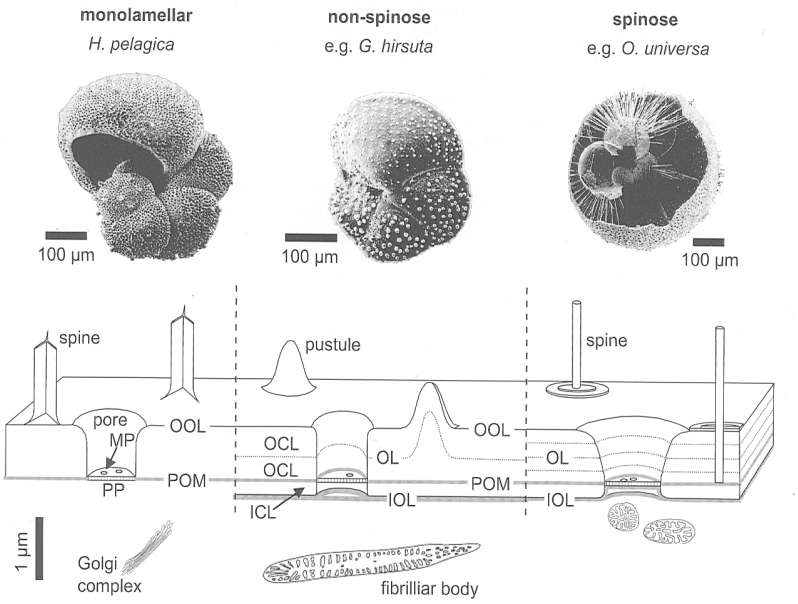

Figure 1 Schematic drawing of the planktonic foraminiferal shell ultrastructure, separating the bilamellar spinose, non-spinose and monolamellar spinose; the scheme for the non-spinose is representative for both macro- and micro-perforates; *OL*- organic lining; *IOL*- inner organic lining; *OOL*- outer organic lining; *OCL*- outer calcite layers; *POM*- primary organic membrane; *PP*- pore plate; *MP*- micropores; from Schiebel and Hemleben (2005).

Yet, direct evidence for the transition from a non-spinose to spinose state is lacking in the fossil record, which so far can only be explained by a transition speed beyond the resolving potential of the fossil record. An analysis of the fossil record following this initial radiation of the spinose taxa indicates that all subsequent lineages of spinose planktonic foraminifera with bilamellar shells can be linked to the genus *Eoglobigerina* (Kennett & Srinivasan 1983; Olsson et al 1992; Olsson et al. 1999; Pearson et al. 2006). After a diversification in the early to middle Eocene, several spinose lineages continued through the Oligocene into the Neogene (Pearson et al. 2006), where most extant spinose species are believed to have developed (Fig. 2).

The extant non-spinose macroperforate lineages are the result of a radiation that started 30 Ma ago (Fig. 2; review in Kucera & Schönfeld 2007). Whereas some phylogenetic relationships in this group are well supported in the fossil record (e.g. the monophyly of the Neogloboquadrinidae; Kennett & Srinivasan 1983), several alternative interpretations of the fossil record exist to explain the relationships within the modern genus *Globorotalia* (Kennett & Srinivasan 1983; Cifelli & Scott 1986).

Another survivor of the Cretaceous-Tertiary extinction, the genus *Guembelitria*, which possessed a microperforate wall texture, is the most likely ancestor of the modern microperforate planktonic foraminifera (Banner & Blow 1960). This fossil-based phylogenetic hypothesis would imply that the modern microperforate foraminifera are monophyletic and distinct from both the spinose and non-spinose macroperforate lineages (Liu & Olsson 1992; Gorgescu 2009).

The origin of the Hastigerinidae is unknown. In comparison to the bilamellar planktonic foraminifera, the monolamellar shells of both *Hastigerina* and *Hastigerinella* are frail and rarely preserved in marine sediments. Reports of Hastigerinid shells from ~ 9 Ma old sediment (Kennett & Srinivasan 1983) are most likely misidentifications of the retrieved specimens, resulting from confusion about the taxonomical status of *Hastigerina pelagica* and *Globigerinella siphonifera* (Banner & Blow 1960; Walker & Vilks 1973). Ultrastructural observations on features such as spine structure, shell lamellarity and the unique 'bubble capsule' of the Hastigerinidae clearly separate the two species from *Globigerinella siphonifera* and *Beella digitata* (e.g. Hemleben et al. 1969).

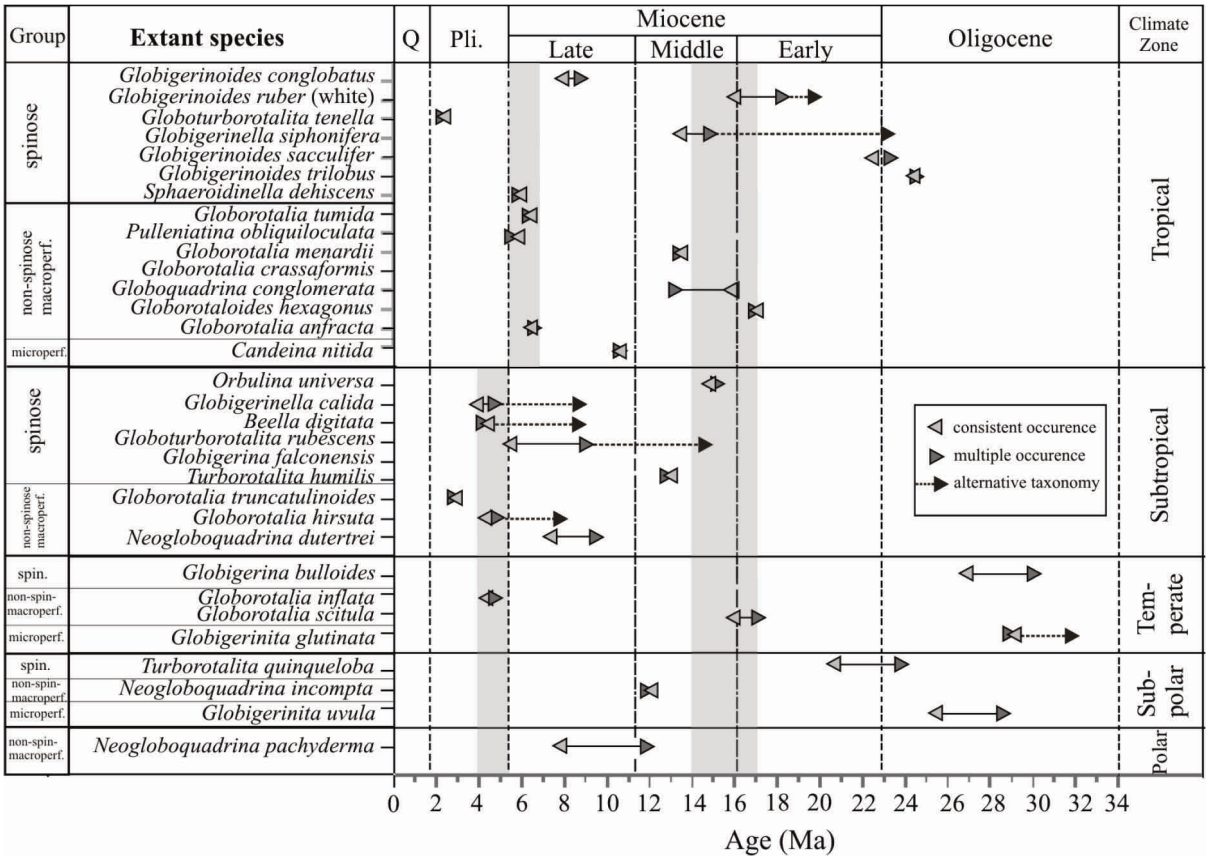

Figure 2 First appearance date (FAD) estimates for 32 extant planktonic foraminifer species; FAD ages of selected lineages using broader taxonomic concepts are shown by dashed lines. Grey bars show periods of major diversification. Species are assigned to faunal provinces after Kucera et al. (2005). Modified after Kucera and Schönefeld (2007).

Even though the general phylogeny of the major extant groups and genera seemed sufficiently well reconstructed, relationships of several modern species remained subject to inconsistencies, e.g. the monophyly of the Globigerinoidae, or the ancestral state of *G. siphonifera* (e.g. Saito et al 1981; Kenneth & Srinivasan 1983). As mentioned above, the resolution of the phylogenetic relationships of the extant planktonic foraminiferal species is not equally well resolved for all taxa. Therefore, with the advent of easy-to-use and reliable DNA amplification and sequencing techniques in the mid 1980's, a set of molecular data became available that had the potential to solve the phylogenetic relationships in the extant foraminiferal lineages.

2. SSU rDNA data and the planktonic foraminiferal fossil phylogeny

When the first ribosomal LSU and SSU rDNA (LSU: large subunit; SSU: small subunit) sequences of planktonic foraminifera became available in the early 90's (Merlé et al. 1994; Darling et al. 1996), micropaleontologists hoped for a fast confirmation (or clarification) on the fossil phylogeny of modern planktonic foraminifera. The first DNA-based phylogenetic reconstructions containing planktonic foraminiferal SSU rDNA were, however, focused on the phylogenetic position of this group relative to other eukaryotic and prokaryotic protists, as only few sequences of even fewer species were available at that time. These phylogenetic reconstructions placed the Foraminiferida relatively basal in a tree of the eukaryotic kingdom, suggesting an early origin of the group even before the separation of the major eukaryotic lineages (Fig. 3; Pawlowski et al. 1994; Pawlowski et al. 1996; Wade et al. 1996).

The phylogenetic signal conserved in the SSU rDNA sequences of planktonic foraminifera resulted in very similar tree reconstructions as those derived from using the LSU rDNA, placing the foraminifera again at the basis of the 'crown' group (Fig. 3). However, the SSU rDNA of planktonic foraminifera possesses some unique features among eukaryotes. The most prominent difference is the relatively large size of the planktonic foraminiferal SSU rDNA gene (Fig. 4) in comparison to the SSU rDNA of most other eukaryotes, e.g. ~4000 nucleotides (nt) in *Orbulina universa* compared to only ~ 2000 nt in a "typical" eukaryotic SSU rDNA gene (de Vargas et al. 1997).

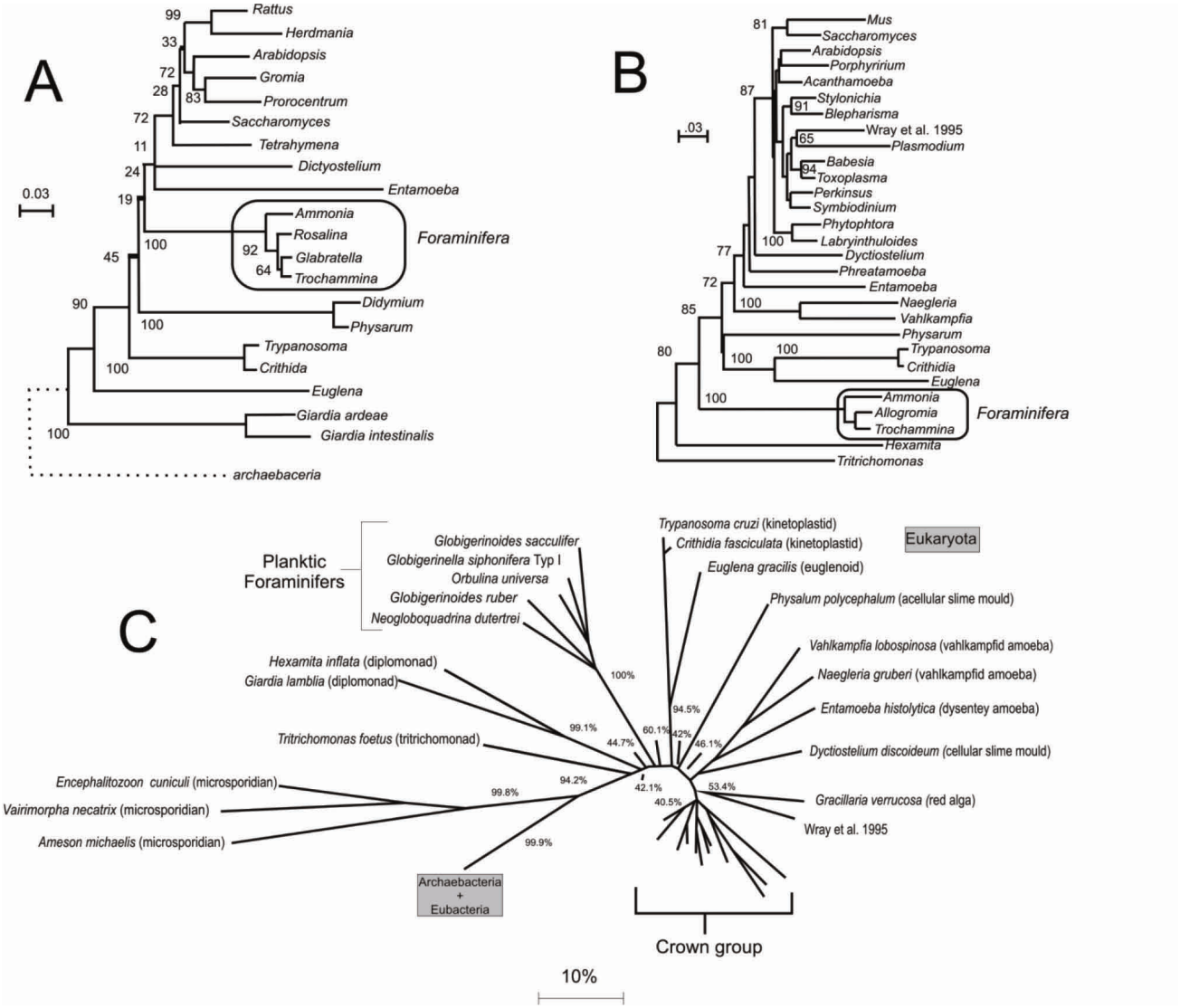

Figure 3 Phylogenetic position of benthic (A and B) and planktonic (C) foraminifera in relation to other eukaryotes reconstructed from LSU (A) and SSU (B and C) rDNA sequences

A. Eukaryotic phylogeny, including foraminifera, inferred from partial LSU rDNA sequences. The tree was constructed by NJ analysis of 610 unambiguously aligned sites; redrawn from Pawlowski et al. (1994).

B. Eukaryotic phylogeny inferred from SSU rDNA sequences; modified after Pawlowski et al. (1996).

C. SSU rDNA phylogeny for the planktonic foraminifera, reconstructed by NJ analyses of 546 unambiguously aligned sites. The "crown group" contains sequences from a wide range of groups, from dinoflagellates to *Homo sapiens*; modified after Wade et al. (1996).

Responsible for this size difference are mostly the highly variable expansion segments of the SSU rDNA gene (Fig. 4; Box 1). These expansion segments alternate with more conservative regions in the gene and vary considerably in size and nucleotide composition between species. The reason for the large variability in nucleotide sequence and length of the expansion segments is not completely understood. The functional analysis of the unique foraminiferal insertions is aggravated by the fact that no information about the final and functional rRNA sequence is available for planktonic foraminifera until today. The SSU rRNA in eukaryotes is known to be extensively processed before (or during) its final structural folding (e.g. Wuyts et al. 2002). The fate of the expansion segments in planktonic foraminiferal SSU rRNA in this processing is unknown. It was however speculated that the surplus of basepairs is excluded during the processing of the rRNA (Pawlowski et al. 1996).

Another feature of the SSU rDNA in planktonic foraminifera is its lack of intra-individual variability. The gene complex of the ribosomal DNA is presented in multiple copies, yet all publications on SSU rDNA sequences in planktonic foraminifera only report a single intra-individual SSU sequence type. However, a considerable intra-individual variability has been found in the ITS sequences of *Globorotalia truncatulinoides* (de Vargas et al. 2001).

As the SSU rDNA sequences of planktonic foraminiferal showed to be more divergent between species as was their LSU rDNA, the SSU rDNA became the standard genetic marker for genotyping and phylogenetic approaches in planktonic foraminifera (e.g. Darling & Wade 2008; Ujiié & Lipps 2009).

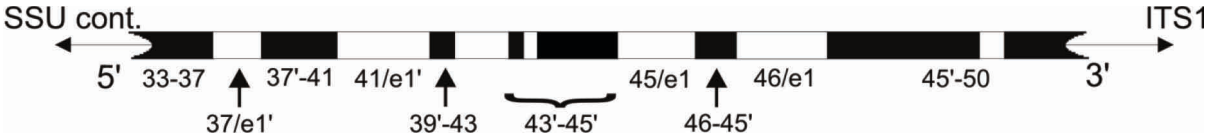

Figure 4 Schematic representation of the SSU rDNA fragment used for the genetic identification of planktonic and benthic foraminifera.
Black areas represent the relatively conservative regions, white regions stand for the more variable parts of the fragment. The numbering refers to a hypothetical secondary structure model for the 3′ SSU rDNA according to Wuyts et al. (2002), labelled after the SSU rRNA helices they are coding for. Modified after Grimm et al. (2007).

The first publications concerning the ingroup phylogeny of the planktonic foraminifera based on SSU rDNA arrived in the late 90's and their results mostly supported, at least for the species available, the generally accepted planktonic foraminiferal phylogeny (Darling et al. 1997; de Vargas et al. 1997). However, some phylogenetic relationships with an ambiguous fossil phylogeny (as that of *Globigerinella siphonifera*) were also weakly supported in the molecular phylogeny (Darling et al. 1997; de Vargas et al. 1997), or were even in contradiction to what was established from the fossil record, e.g. for the non-spinose *Globorotalia inflata* (de Vargas et al. 1998). Apparently, the phylogenetic signal manifested in the SSU rDNA sequences of planktonic foraminifera was not evolving with the same speed or in the same mode as were their morphological traits. Also, the level of divergence within and between the three major groups of planktonic foraminifera (spinose, macroperforate non-spinose and microperforate non-spinose) was found to be largely different. When compared to benthic foraminifera in a phylogenetic reconstruction, the spinose foraminiferal species showed surprisingly long branches, whereas the non – spinose groups were less divergent from their benthic relatives (e.g. Ujiié et al. 2008).

Box 1: SSU rDNA in planktonic foraminifera

The first approaches aiming to sequence the ribosomal DNA in planktonic foraminifera were undertaken in the mid 90's. Langer et al. (1993) report the first successful attempt to sequence planktonic foraminiferal SSU rDNA. Merlé et al. (1994) published the first LSU rDNA based phylogeny comprising planktonic foraminifera, yet had to cope with a number of "contaminated" sequences. As the primers they used were rather universal for the gene region, they had amplified and sequenced yeast and copepods in three out of five cases (Merlé et al. 1994). A similar contamination by Wray et al. (1995) with an alleged sequence of the benthic foraminifera *Ammonia beccari* was later related either to *Plasmodium* (Pawlowski et al. 1996) or red algae (Wade et al. 1996). Aiming to evade such problems in their attempt to sequence SSU rDNA from planktonic foraminifera, Darling et al. (1996) cultured their collected specimens to gametogenesis. By doing so, they minimised any potential contamination from food particles and symbionts. Their results paved the way for more specialised primers for planktonic foraminiferal SSU rDNA (e.g. Darling et al. 1997), facilitating the amplification of the gene from only a single individual of planktonic foraminifera without the effort of further culturing.

Progress was also made in the foraminiferal DNA extraction methods. Whereas Merlé et al. (1994) used proteinase K and phenol-chlorophorm (a rather labour intense process) to digest the cells and extract the DNA, Pawlowski et al. (1994) had developed a sodium deoxicholate buffer (DOC) that simplified the extracting procedure and was subsequently used in most foraminiferal DNA surveys. A major drawback of this approach, however, was that the calcite shells of planktonic foraminifera dissolve in the buffer, preventing any further taxonomic or morphometric validation of the specimens after DNA extraction. Therefore, DOC was later replaced in some working groups by a guanidinium thyocyanate buffer (de Vargas et al. 2002) that had no dissolving effect on the foraminiferal shells.

In order to gain a high quality DNA sequence reads, PCR (polymerase chain reaction) amplification products are generally cloned before sequencing. Cloning (i.e. to ligate a PCR product into a plasmid vector and multiply it in competent *E. coli* cells) has the great advantage that different sequence replicates from a PCR product are separately multiplied in "their" clone, resulting in a minimised amount of misreads in the sequence-chromatogram output. Moreover, any existing intra-individual variability is more likely to be detected. Besides the common practice of cloned sequencing, PCR products are also sequenced directly, which usually results in a less well resolved sequence read, yet has the advantage of lesser costs and is also considerably faster (from extraction to sequence). Direct sequencing is generally used, when a sequence type is already established from cloning and large numbers of individuals need to be genotyped. In this regard, the application of RFLP (restriction fragment length polymorphism) analyses is, in comparison, even faster and cheaper. Here, specific restriction enzymes digest the (PCR amplified) SSU rDNA at specific short nucleotide sequences, resulting in a number of sequence fragments of different sizes that show genotype specific patterns in a gel-electrophoresis. Even though this technique allows a large number of samples to be analysed in a relatively short time span (e.g. de Vargas et al. 2001; Morard et al. 2009), it requires a detailed a priori knowledge about the targeted sequences, and cannot be used to detect minor variations between closely related genetic types in the SSU rDNA of planktonic foraminifera.

The SSU rDNA of foraminifera, and of planktonic foraminifera in particular, is unique among eukaryotes in regard to the existence of the variable regions 37/e1', 41/e1' and 46/e1' (Fig. 4). Apparently, the variable region 37/e1' corresponds to a universal variable regions of the prokaryote structure model (Neefs et al. 1990; de Vargas et al. 1997). The other three length-variable regions (Fig. 4) of the SSU rDNA are also known from the SSU rDNA of other eukaryotes. The degree of variability in these variable gene regions greatly varies between the different groups of foraminifera and only few species of planktonic foraminifera (e.g. the non-spinose *Globorotaliidae*) can be aligned in these regions to benthic foraminifera. For the spinose planktonic foraminifera this is only possible in the conserved regions of the gene (e.g de Vargas et al. 1997). Therefore, manual alignments of SSU rDNA of planktonic foraminifera were modified based on the SSU rRNA universal secondary structure model (e.g Van de Peer et al. 1996; Wuyts et al. 2002), in order to include only homologous nucleotide positions in the phylogenetic reconstructions. The effect of this procedure in dealing with the variable regions in planktonic foraminifera on manual alignment-based phylogenetic reconstructions of the group is discussed in great detail in publication 1 of this thesis.

Moreover, intensified sampling and sequencing in the subsequent years unveiled an unexpected high degree of genetic diversity in the SSU rDNA gene for several planktonic foraminiferal morphospecies (e.g. Darling et al. 1999; de Vargas et al. 2001). Numerous distinct genetic types were identified in ~80% of the investigated species (for the most recent summary see Darling & Wade 2008). From the very beginning, these SSU rDNA genotypes were considered to represent a 'cryptic' divergence at the level of species within the established planktonic foraminiferal morphotaxa. Even though the possibility of a cryptic speciation was discussed in micropaleontological circles (e.g. Norris et al. 1996), the prevalence of genetically validated cryptic species in planktonic foraminifera was an unexpected surprise. As cryptic species are, per definition, morphologically undistinguishable from one another (or nearly so), the synergetic effects between genetic research on planktonic foraminifera and the applied branch of micropaleontology are still few and mainly occur when a cryptic divergence in a morphospecies is uncovered to be pseudo-cryptic (as in the case of *Neogloboquadrina pachyderma*; e.g. Hippler et al. 2009).

3. The 'cryptic species' problem

Cryptic species are reported abundantly in scientific publications since the mid 20th century (e.g. Winge 1965; McLaren et al. 1966), yet the term is actually a consequence of a much older dispute, namely the 'species problem'. Essentially, the 'species problem' originates from the confusion and ideological friction in biological science about what a 'species' is and how to define it (e.g. Mallet 1995). Therefore, the term 'species' can be used in a taxonomical sense for a typological classification of different forms of life (creating morphospecies), or it can stand for the definition of an evolutionary unit (representing biological species; e.g. Mayr 1970; Schilthuizen 2000). Moreover, some definitions, as that of a biological species for instance, cannot be applied to all forms of live equally, as they postulate specific requirements not shared by all organisms (e.g. sexual reproduction and diploidy; Mayr 1970). Other definitions are based on virtually unverifiable statements, such as monophyly of a group of individuals or niche separation in an ecosystem (e.g. Coyne & Orr 2004).

Generally speaking, cryptic species are a corollary of the incapability of the classic morphological taxonomic concept to resolve the evolutionary level of species divergence that can only (or at first) be resolved on a physiological or genetic scale (e.g. Amato et al. 2007). Consequently, cryptic species can only be described in morphotaxa, however, their description then bases on characters that are not phenotypical, i.e. not part of the original definition of the species that the cryptic divergence was found in. Frequently, a subsequent re-evaluation of the morphological characters of the discovered cryptic species culminates in the description of, ever so slight, phenotypic differences, transposing the cryptic species into pseudo-cryptic species.

The rate at which cryptic species are discovered in classic morphotaxa is increasing since the late 80's (Fig. 5) mainly induced by the progress in molecular genetics (e.g. Bickford et al. 2007). The term 'cryptic species' is, however, used slightly inconsistently in the literature (Bickford et al. 2007), frequently used synonymous with the older term 'sibling species', referring to two species that are direct sister species and have not been distinguished from one another taxonomically. Both terms describe a similar phenomenon, yet on different evolutionary and temporal scales. Whereas the term 'sibling' addresses two sister species of relatively recent divergence, 'cryptic species' are not necessarily direct sister species and therefore could have separated somewhat more deep in the past (e.g. Saez & Lozano 2005).

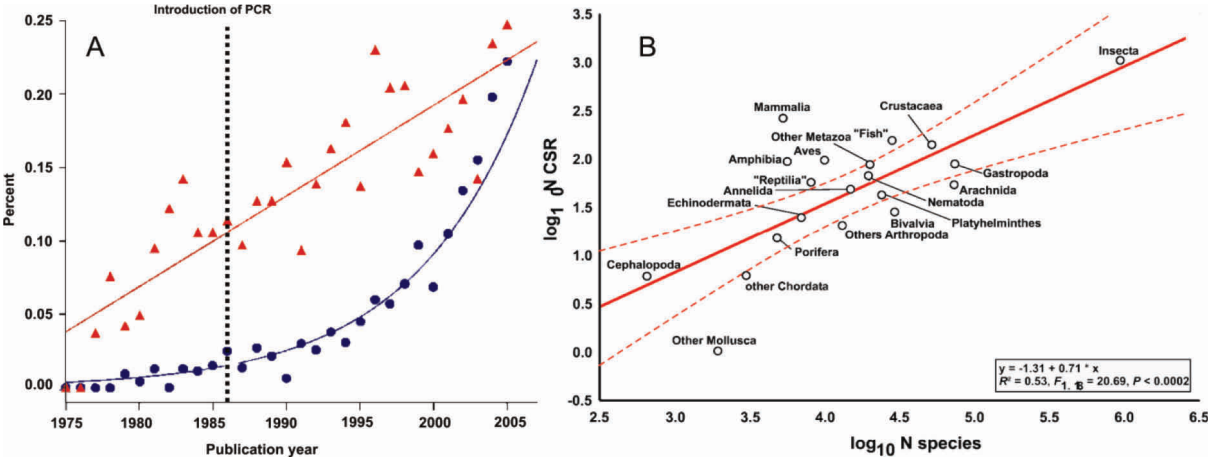

Figure 5 Reports of cryptic species in the scientific literature and their relation to morphospecies.

A. Percentage of peer-reviewed publications in the Zoological Record Plus (CSA) that mention 'cryptic species' (blue circles) or 'sibling species' (red triangles) in the title, abstract, or keywords. Redrawn from Bickford et al. (2007).

B. The $\log_{10}$ of cryptic species reports (CSR) as a function of the $\log_{10}$ number of described species in the respective taxon**.** Deviations from the regression line represent CSR taxon variation. Dashed lines represent 95% confidence intervals. Redrawn from Pfenniger and Schwenk (2007).

Besides the taxonomical implications for morphologically defined species and higher taxa, the existence of a cryptic divergence has other, more practical consequences. As the classification of an ecosystem is based on the best possible resolution of its biodiversity (Hooper at al. 2005; Wittebolle et al. 2009), a hidden species richness will likely result in a more or less severe misinterpretation of ecosystem properties as, for example, its stability, key species or trophic levels (e.g. Bickford et al 2007). This is especially relevant for eukaryotic protists, where morphological characters are often rare, largely absent, or ill-defined (e.g. Fenchel & Finlay 2006; Adl et al. 2007). Yet, for planktonic foraminifera with their character-rich shells and continuous fossil record, the discovery of a hidden species-level divergence was unexpected.

4. Cryptic species in planktonic foraminiferal morphotaxa

As for most other eukaryotic taxa (Fig. 5), the cryptic species phenomenon in planktonic foraminifera can be interpreted as an emulsion of two separate phenomena: for one, the inadequate resolution of morphotaxonomy, resulting in taxa which are of a 'higher order' than the level of biological species (some researches therefore labelled the classic morphospecies as 'species aggregates' or 'super-species'; e.g. de Vargas et al. 2004); second, the existence of a morphologically irresolvable genetic divergence between two separate biological sister species.

In the everyday application of planktonic foraminiferal morphotaxonomy, researches have to deal with the natural variability in species shell formation. The occurence of inter-specific morphological varieties, as well as autecological and physiological differences within planktonic foraminiferal morphospecies is well established in the literature (e.g. Orr 1969; Kahn 1981; Healy-Williams et al. 1985). In the absence of independent genetic information, the phenotypic variation reported in all well-studied planktonic foraminiferal species was mostly interpreted as ecophenotypic (e.g. Hecht 1974; Kennett 1976; Fig. 6). Yet, depending on the level of expertise of the individual researcher, these phenotypic variations were also basis for the naming of numerous varieties and subspecies (e.g. *Globigerina calida* Parker 1962, splitted from *G. siphonifera* d'Orbigny 1839 and followed by the splitting of *G. calida* subsp. *praecalida* Blow 1969).

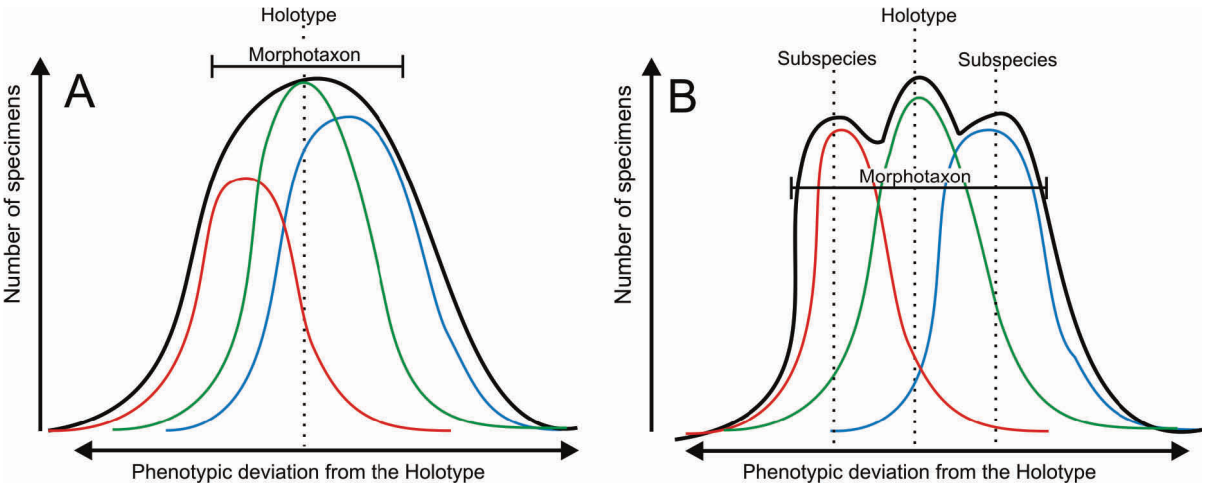

Figure 6 General schemata of cryptic divergence in a morphologically defined species.
A. The phenotypes of a cluster of biological species (coloured lines) are strongly overlapping with no clear abundance maximum of any distinct morphotypic variation. Therefore, the most common phenotype is set as holotype and the species are integrated into a single morphotaxon. Deviation from the holotype is widely considered to be induced by environmental factors. This is the case for most planktonic foraminiferal morphospecies with more than one SSU rDNA genotype.
B. The phenotypes of the species overlap, however with distinct abundance peaks of distinct morphotypes. The taxonomy of such species clusters is generally more controversial in the literature, the morphotypes are either regarded as clear ecophenotypes (*G. ruber* white and pink), labelled as subspecies, or are regarded separate species (e.g. *G. siphonifera* and *G. calida*).

Soon after the first SSU rDNA sequences of planktonic foraminifera were published (Darling et al. 1996), numerous genetic types were found within several species (e.g. de Vargas et. al. 1997; Huber et al. 1997; Darling et al. 1999) and the question arose whether this genetic divergence correlated with the phenotypic intra-species variability recorded in so many species (e.g. Huber et al. 1997; de Vargas et al. 2001; Darling et al. 2006). *Globigerinella siphonifera*, for example, was reported to have two ecotypes with different preferred depth habitats, and an off-phase, full and semi-lunar reproduction cycle between the two types has been speculated (e.g. Huber et al. 1997). However, the correlation with the two genetic types known from individuals of *G. siphonifera* could not be established beyond doubt (Huber et al. 1997; de Vargas et al. 2002), and therefore the extensive surveys performed on *G. siphonifera* led to no taxonomical aftermath.

The genetic diversity uncovered in planktonic foraminifera was found to be unevenly distributed between the various morphospecies. About 50% of the modern planktonic foraminiferal morphospecies have been sequenced to date, the taxa coverage of SSU rDNA sequences being fairly even between spinose (including *Hastigerina pelagica* and *Hastigerinella digitata*), macroperforate and microperforate non-spinose foraminifera. Whereas seven distinct genetic types with pronounced geographical distribution patterns are known from the species *Neogloboquadrina pachyderma* (Darling et al. 2000, 2004, 2007), or six types in *Globigerina bulloides* (Darling et al. 1997; de Vargas et al. 1997; Darling et al. 1999, 2000, 2003), other planktonic foraminifera such as *Globigerinoides sacculifer* and *Globorotalia inflata* are (to date) known only by a single genetic type with global distribution (Darling et al 1996, de Vargas et al. 1997; own unpublished data).

This discontinuous and unpredictable occurrence of genetic diversity in planktonic foraminiferal morphotaxa, in combination with inconsistent sampling efforts and a high degree of provincialism among genetic types (e.g. Darling et al. 2004, 2007; publication 2 of this thesis), so far prevented a

sensible estimate for the total genetic diversity of planktonic foraminiferal SSU rDNA genotypes. Based on the total of available SSU rDNA sequence data today, it appears that cryptic species can be found in all intensively sampled and studied foraminiferal taxa, but the extent of the genetic diversity in a single morphological species could be limited to a maximum of less than ten distinct SSU rDNA genetic types (see Darling & Wade 2008 for the latest review).

Besides the debate about the systematic consequences for the established morphospecies, the discovery of distinct genetic types in planktonic foraminiferal morphotaxa had an impact on another, more fundamental question: What is the principal mode of speciation in planktonic foraminifera? Under the classic concept of morphospecies the mode of speciation is, generally speaking, neo-darwinian, i.e. speciation by the means of natural selection (e.g. Mallet 1995). Here, morphospecies evolve from competition, which infers that individuals of a morphospecies diverge while co-occuring. The planktonic realm appears to be a, more or less, homogenous intermixing water mass. Therefore, it is not hard to imagine that planktonic foraminiferal morphospecies, especially the cosmopolitan representatives, co-occur in global populations, at least on evolutionary time scales.

Biological species, as most likely represented by the SSU rDNA genotypes of planktonic foraminifera, are considered to originate from allopatric populations, i.e. are a result of a geographical isolation or reproductive barrier between distinct groups of a species that over time differentiate into new distinct species (e.g. Hutchinson 1961; Palumbi 1992; Sexton & Norris 2008). However, spatial isolation in largely cosmopolitan planktonic species living in a constantly intermixing environment is hard to prove (e.g. Sexton & Norris 2008; Cermeño & Falkowski 2009). Therefore, the possibility of a sympatric speciation of planktonic foraminifera, the genesis of two species from a common ancestor without spatial isolation, was considered by some authors (e.g. Lazarus et al. 1995; Pearson et al. 1997).

Yet, even though the oceanic realm appears to be without physical barriers, it is far from being homogenous. Especially when considering that reproductive isolation has not only a spatial component, but also a temporal, the opportunities for allopatric speciation are potentially numerous.

For one, the diversity of planktonic foraminiferal species is not equally distributed throughout the world ocean. The highest number of planktonic foraminiferal species can be found in tropic to subtropic waters (Rutherford et al. 1999; Fig. 7), a circumstance shared by many other marine planktonic taxa (e.g. Palumbi 1994). This high species richness is accompanied by a relatively low overall planktonic foraminiferal abundance, resulting in a high diversity and patchy distribution of planktonic foraminiferal assemblages (e.g. Bé 1959; Bé et al. 1971; Tolderlund & Bé 1971). Consequently, the possibility for gene flow between patchy distributed populations of a single species is reduced, leading to an elevated potential for speciation. Further, the adaptation of planktonic foraminiferal species to a variety of different depth habitats in the vertical water column (Schiebel & Hemleben 2005; Fig. 7) indicates that the observation of the two-dimensional geographic planktonic foraminiferal species distribution is not sufficient to determine whether two populations or ecophenotypic variants of a species are allopatric or not. Moreover, geographically and hydrographically completely sympatric populations might be separated by temporally shifted lunar reproduction cycles (e.g. Bijma et al. 1990). If one part of a population has their reproductive peak a few days before or after the other, gene flow is again reduced and a, superficially sympatric, speciation can occur.

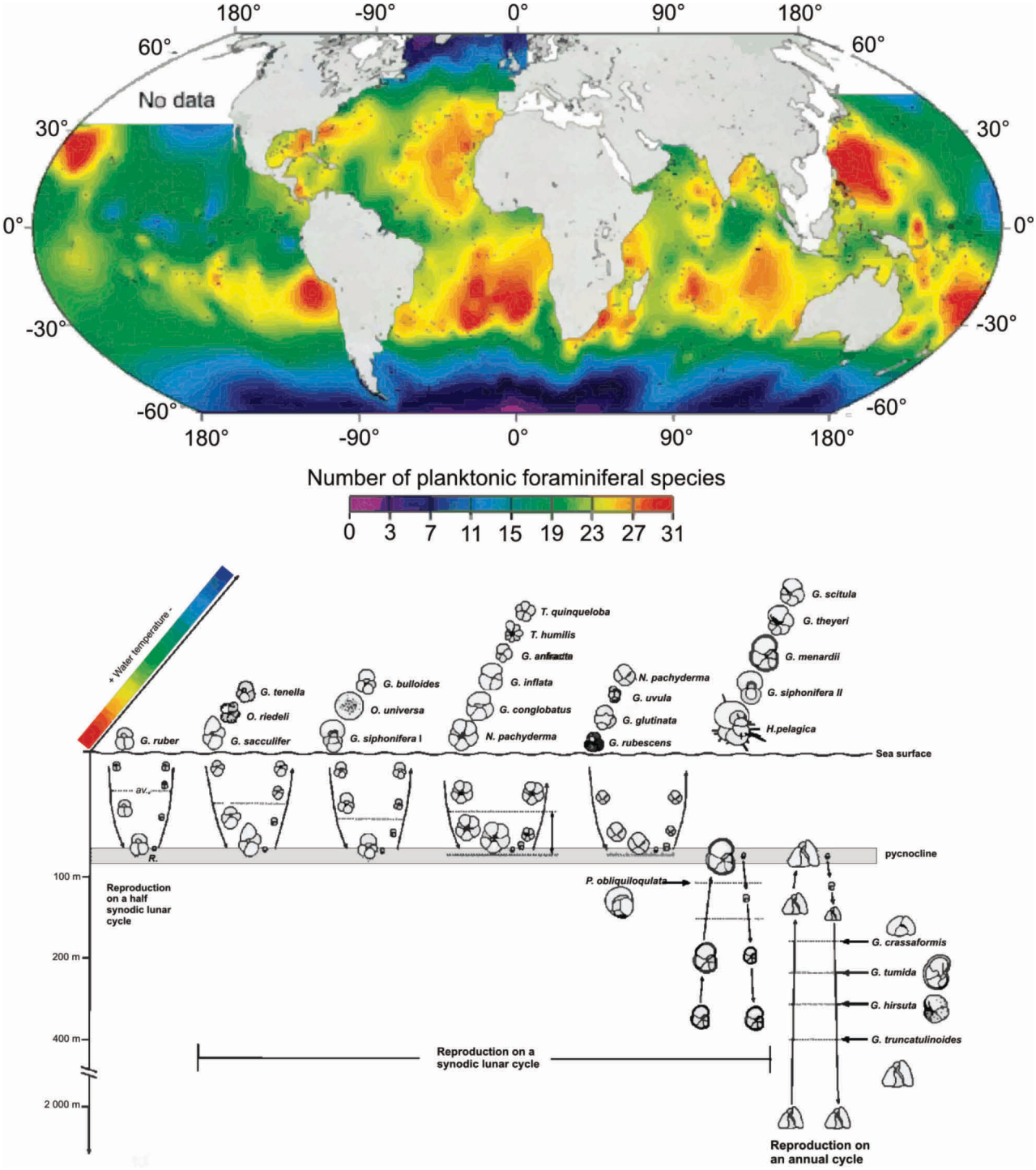

Figure 7 Global distributions of planktonic foraminiferal species richness and depth habitats.
Interpolation of the global distributions of planktonic foraminiferal species richness, redrawn from Rutherford et al. (1999). As found in most other marine planktonic taxa, the species richness of planktonic foraminiferal assemblages is highest in the oligotrophic centres of the world ocean, and lowest in polar waters. The schematic drawing of the vertical depth habitat of the best known planktonic foraminiferal species and their vertical distribution during ontogeny is redrawn from Schiebel and Hemleben (2005).

Apparently, cryptic species are a common phenomenon in planktonic foraminifera (e.g. Darling & Wade 2008), and the existence of a species-level genetic divergence should not have come as a big surprise for the micropaleontological community. For one, the application-oriented species concept produced globally distributed morphospecies with large phenotypic ranges that were treated as ecophenotypic variation. Authors had described and catalogued a range of morphotypes in well established species, or created new species on the basis of ever so slight aberrations in morphological features, sometimes both in the same publication (e.g. Parker 1962). Moreover, a molecular diversity beyond that of the established morphotaxonomic concept had also been reported for other planktonic protist groups as for example chlorophyta (Olsen-Stojkovich et al. 1986) and dinoflagellates (Cembella et al. 1988). Latest ecological (e.g. de Vargas et al. 2001; Darling et al. 2003; see publication 2 of this thesis) and evolutionary (Alizon et al. 2008) studies indicate that a better knowledge of the cryptic diversity in planktonic foraminifera is a necessary precondition for an enhanced understanding of spatial and temporal distribution patterns in planktonic foraminifera. On the basis of genetic and sedimentary data from *G. bulloides* (Kucera & Darling 2002; Fig. 8), it was laid out that the knowledge about the genetic diversity in a planktonic foraminiferal morphospecies could be used to decrease the error in the reconstructions of SST signals considerably.

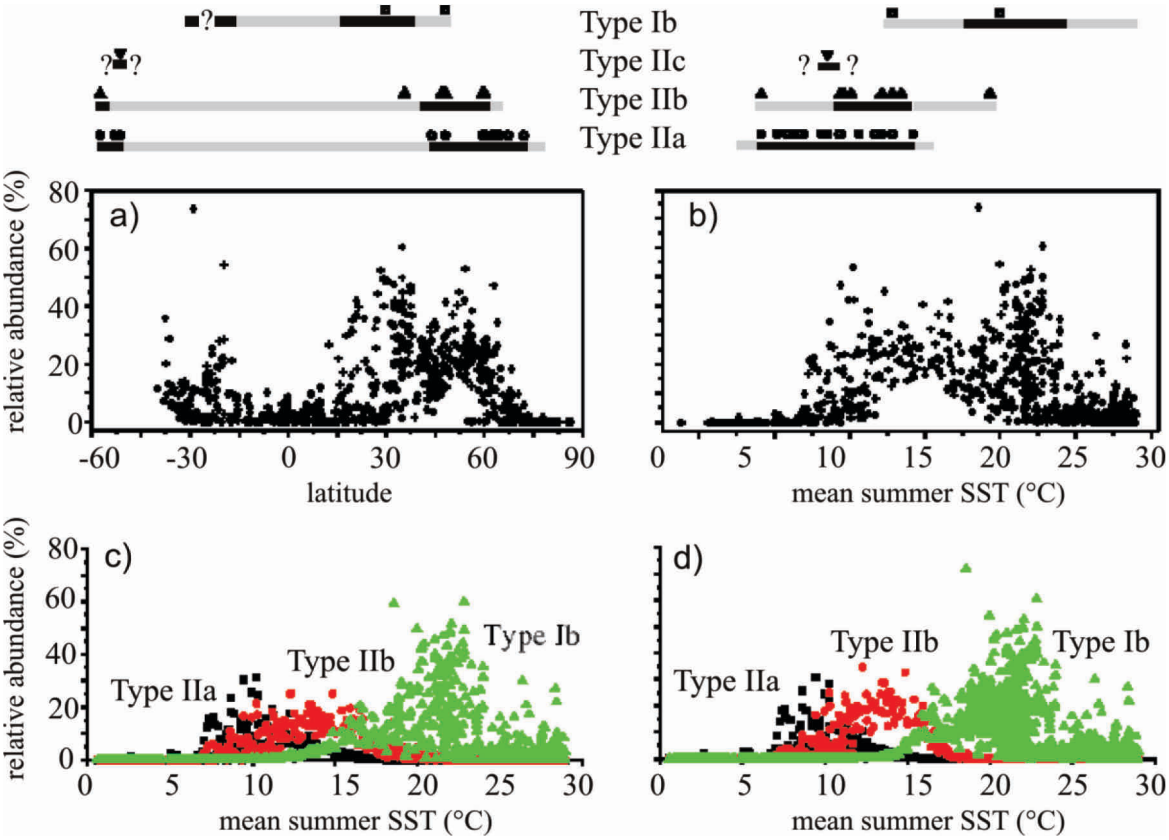

Figure 8 Genetic types in G. bulloides and their potential for improving SST reconstructions
The genetic types of the spinose *G. bulloides* (Ib to IIc) and their reported latitudinal and SST related distribution (symbols) and potential distribution (black bars). The graphs a) and b) show the relative abundance of *G. bulloides* shells in core top sediment samples; c) and d) are a combination of the genotype distribution and the core top abundances, under the assumption of Gaussian bell shaped curve models with more (c) and less (d) overlap and that the genetic types are co-occurring in an assemblage. Redrawn from Kucera and Darling (2002).

Inevitably, for all the insights and progress made in the phylogeny and diversity on planktonic foraminifera, the amount and species coverage of SSU rDNA sequence data is in constant need of improvement. Moreover, the methodological linkage between the results from the genotyping of extant planktonic foraminiferal populations and their sedimentary remains in the recent and fossil marine sediments is still basic at best. Yet very recent publications to that topic provided fresh impulses to the discussion of the cryptic or pseudo-cryptic nature of SSU rDNA genotypes in planktonic foraminifera (e.g. Morard et al. 2009; Kuroyanagi et al. 2009).

This thesis aims to contribute in three different fields to the research of planktonic foraminiferal genetic diversity, publication 1 provides an objective and fast method for phylogenetic reconstructions on all available SSU rDNA sequences of planktonic foraminifera, publication 2 presents a high resolution mapping of the seasonal and spatial distribution of *Globigerinoides ruber* genotypes and publication 3 documents the revalidation of the species status of *G. ruber* by the combination of morphometric analyses, molecular dating and the fossil record.

5. Focus and structure of the included research papers

Publication 1, "*Using the multiple analysis approach to reconstruct phylogenetic relationships among planktonic foraminifera from highly divergent and length-polymorphic SSU rDNA sequences*", deals with a new approach to derive the best possible molecular phylogeny for planktonic foraminifera based on their SSU rDNA sequences. The SSU rDNA gene is a succession of highly conserved and highly variable regions. The variable regions are a great obstacle for sequence alignments of distinctly related morphospecies, as the homology of the single bases in these regions cannot be unambiguously reconstructed. Yet, complete base homology in an alignment is a basic requirement for any phylogenetic analysis. Therefore, the expansion segments had to be excluded from final alignments. The existence of such large expansion segments in planktonic foraminiferal SSU rDNA is unique for eukaryotes (e.g. de Vargas et al. 1997; Grimm et al. 2007). However, as the sequence and size of the functional ribosomal RNA in planktonic foraminifera is still unknown, start and end of the expansion segments in the SSU rDNA are purely speculative and the extent to which these regions are excluded from an alignment depends at equal parts on the range of species that are included and the personal preference of the researcher creating the alignment.

Generally speaking, the more species are involved, the smaller the alignments will be, as the number of homology-ambiguous sites increases, loosing up to ~50% of the actual alignment. This, of course, has a negative effect on the support of the terminal nodes of a phylogeny, which is especially unsatisfying when a number of new genetic types of unknown ancestral state are included. By using automated multiple alignments of all available planktonic foraminiferal sequences, the resulting phylogenies are entirely reproducible, objective and most of all time-saving relative to manual alignments. In comparison to older phylogenies based on manually truncated alignments, the phylogenies and node supports generated by the automated alignments are similarly conclusive; some nodes are even better supported. Another important benefit of multiple alignments is the fast and unbiased phylogenetic assignment of new or unknown sequence types. This was tested in our approach with a number of sequences from unidentified spinose and non-spinose foraminifera, as well as with a new and rather distinct genetic type of *Hastigerina pelagica*. Until the sequence of functional SSU rRNA is known, the automated multiple alignment approach is likely to serve as the most elaborate and scientific method for the reconstruction of an SSU rDNA based foraminiferal phylogeny.

Publication 2, "*Geographical distribution of cryptic genetic types in the planktonic foraminifer* Globigerinoides ruber" is dealing with the diversity and distribution of genetic types found in the morphospecies *Globigerinoides ruber,* sampled from various stations in the eastern Atlantic Ocean and the Mediterranean Sea. The mapping of the spatial (and temporal) distribution of cryptic species in planktonic foraminifera has the potential for an increased resolution in past-ocean SST reconstructions (Kucera & Darling 2002). The distribution of genotypes found in *Orbulina universa*, *Globorotalia truncatulinoides* and *Globigerinella siphonifera* from a transect through the central Atlantic Ocean have been correlated to ambient chlorophyll concentrations (de Vargas et al. 1999, de Vargas et al. 2001, de Vargas et al. 2002). Genetic types in the cold water species *Neogloboquadrina pachyderma* from both Polar Regions were reported to have distinct temperature ranges (e.g. Darling et al. 2004, Darling & Wade 2008). Our data on the genetic types of *G. ruber* suggest that the controlling factors behind their distribution are ecologically complex. The distribution of the G. *ruber* genotypes in the observed area shows a pattern of exclusion and co-occurrence that suggests a gradual influence of species-level competition between closely related sister types. The closer two genetic types are related, the less likely they occur together in the same habitat, given the habitat is "stable" enough. The two most closely related genetic sister types in our survey, named Type IIa1 and IIa2, co-occur in the area of the Canary Islands (Canary Current), but exclude one another from either one of the two main Mediterranean Sea basins. That this pattern is not simply a geographical or seasonally induced signal is highlighted by the existence of a direct sister genotype, named IIb occuring in the western and eastern Mediterranean Basin alike. This type also seems to be endemic in the Mediterranean Sea, a before unknown feature in planktonic foraminifera on such a relatively small scale. In contrast to the distribution patterns between these Types IIa and IIb, the distribution an abundance of the genotype found in *G. ruber* pink is by no means influenced by their occurrence and abundance. As the distribution of the types seemed to correlate to the degree of divergence of their SSU rDNA sequence, we revisited the phylogenetic relationships of the *G. ruber* genotypes and extended it by the new endemic type, highlighting the paraphyly of the morphospecies. When analysed together with the sister species of *G. ruber*, *G. conglobatus*, the genetic types divide into two clusters, one containing types allocated to *G. ruber* in the strictest sense (s.str.), the other cluster containing the types that share a common ancestor with *G. conglobatus*, suggesting a parallel development of the shell features in the genotypes of *G. ruber* s.str. and *G. ruber* in the widest sense (s.l.).

Publication 3, "A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence", contains the results of an investigation into the paraphyletic nature of the *G. ruber* morphospecies with comparative morphometric measurements. Like most other recent planktonic foraminiferal morphospecies, *G. ruber* has a certain range of morphological plasticity, commonly regarded as ecophenotypic variation (e.g. Parker 1962; Hecht 1974). As indicated by experimental data, parameters as temperature, salinity, oxygen content and food availability are known to have an influence on chamber formation, their size and porosity (e.g. Hemleben et al 1987; Hemleben et al. 1989). With the knowledge about the existence of a species-level genetic diversity in planktonic foraminifera, the question arose to what extend the observed phenotypic variability was actually *eco*phenotypic (Norris 2000). Earlier morphometric analyses have been performed for *G. siphonifera* (Huber et al. 1997), *G. truncatulinoides* (de Vargas et al. 2001) and *O. universa* (Morard et al. 2009), searching for a genotype correlated signal in the phenotype. However, even though the measurements performed in these species were extensive, they were unable to differentiate the correlation between the observed genotypes and distinct features in shell morphology from ecophenotypic variation (Huber et al. 1997; Morard et al. 2009). Consequently, the gap between the level of genetic diversity and phenotypically recognisable species couldn't be bridged. This was only accomplished once, for the coldwater species *N. pachyderma*. Here, the predominant coiling direction correlated with the genetic divergence

recorded in *N. pachyderma*, so that one of the genetic types could be attributed to an (already existing) species name, *N. incompta*. The remaining genetic types remained within the species *N. pachyderma* (Darling et al. 2006).

First superficial observations on the shells of genotyped specimens of *G. ruber* 'pink' and 'white' from the eastern Atlantic Ocean and Mediterranean Sea suggested a recognisable difference between the individuals of the Pink and IIa genotypes. In an attempt to quantify the morphological signal, digital pictures from specimens of Type Pink and IIa were measured and compared with specimens of *G. ruber* 'white' and 'pink' from a recent sediment sample, individuals from a museum collections and a number of images from the recent literature. Led by our first observations and in concordance with the definition of *G. ruber* s.l. (Wang 2000), the focus was laid on measurements of the ultimate and penultimate chamber, using a ratio between chamber height against chamber length to create a measure for the extent of chamber compression. The results indicate that the genetic types of *G. ruber* Type Pink can be statistically separated from Type IIa by the degree of compression of the last and penultimate chamber, thereby corresponding to the separation of the phenotypes of *G. ruber* s.str (Type Pink) and *G. ruber* s.l. (Type IIa). Further, individuals of *G. ruber* s.l. and Type IIa group together with museum specimens of *G. elongatus* (d'Orbigny 1826), a species synonymised with *G. ruber* by Parker (1962).

These findings are supported by literature based reconstructions of the fossil lineages of *G. elongatus*, *G. conglobatus* and *G. ruber*. The fossil phylogenies of *G. ruber* and *G. elongatus* as reconstructed by Cordey (1967) and Perconig (1969) mirror the relationship of the G. ruber s.str. clade (Type Pink and Ia + b) and G. ruber s.l. (Type IIa + IIb) + G. conglobatus clade based on the SSU rDNA phylogenetic reconstruction. Divergence time estimates from a molecular dating approach are in congruence with the fossil dating of the FAD of G. conglobatus and *G. elongatus*, suggesting that Type IIa is in fact not a genotype of *G. ruber* but synonymous with the species definition of *G. elongatus* instead. In general, the combination of original species description, SSU rDNA genotypes and morphometric data strongly suggests a re-evaluation of the species definition of *G. ruber* as it is used today.

General Conclusions

The research on the genetic diversity of planktonic foraminifera is no field of fast progression, compared to the number of publications actually using the group as proxy for environmental reconstructions. Seventeen publications containing genuinely new genetic and morphometric data added to the topic since 1996 (Darling et al. 1996: de Vargas et al 1997; Darling et al. 1997; Huber et al. 1997; de Vargas et al. 1999; Darling et al 1999; de Vargas et al 2001; Stewart et al. 2001; de Vargas et al 2002; Darling et al 2003; Darling et al. 2004; Darling et al. 2006; Kuroyanagi et al. 2008; Ujiié et al 2008; Aurahs et al. 2009a; Morard et al 2009; Ujiié & Lipps 2009), an average of ~ two research publications per year. About 150 research papers using planktonic foraminifera as proxies were published in 2009 alone (searched January 2010 on http://apps.isiknowledge.com, using "planktonic foraminifera" and "planktic foraminifera" as search topic). Moreover, the methods applied in molecular phylogenetic reconstructions in the group are unchanged since 1996 (compare e.g. Wade et al. 1996; Darling & Wade 2008) and modern theories on protist ecology or genotype distribution beyond a correlation with oceanographical parameters as water temperature, salinity and productivity are largely ignored (de Vargas et al 2001; de Vargas et al 2002; Darling et al. 2003; but see Alizon et al. 2008; Aurahs et al. 2009a).

This thesis contains a large set of new SSU rDNA sequences, some representing new genotypes and a comparative morphometric dataset. It further opposes the traditional approaches of SSU rDNA phylogeny by advocating phylogenies based on automated multiple alignments.

Multiple automated alignments as basis for SSU rDNA phylogenies

The phylogenetic approach presented in the publication "*Using the multiple analysis approach to reconstruct phylogenetic relationships among planktonic foraminifera from highly divergent and length-polymorphic SSU rDNA sequences*" is a significant progress in the field of planktonic foraminiferal SSU rDNA based phylogenetic reconstructions. The phylogenetic relationships of the included planktonic foraminifera and their statistical support resulting from the multiple automated alignment approach were found to be comparable to the phylogenetic results from the traditional approach of manually culled alignments. Relationships that are well resolved from the manual alignments are well resolved in the phylogenies from the automated alignments as well (e.g. the *Globigerinoides* cluster), and nodes with low to moderate statistical support are weakly resolved by both the manual and automatic alignment alike. The application of various automated alignment algorithms and the implementation of the "complete" SSU rDNA fragment to phylogenetic reconstructions creates fully reproducible phylogenetic reconstructions. As none of the variable sites in the expansion segments are excluded from the final reconstruction, no a-priory subjective opinion about base homology in these regions is needed. This is certainly the single greatest advantage of this approach as it dispensed any bias from sequence editing and equalises the starting point for the later interpretation of the phylogenetic results.

The automated process of aligning planktonic foraminiferal SSU rDNA sequences considerably eases the addition of completely new sequences, no matter how variable or extensive their expansion segments might be. It further allows the generation of thresholds for the level of genotype divergence, as all genotypes are treated equally in respect to their variable regions. As morphotaxonomy is still in need to catch up with the resolution of genotype recognition, DNA taxonomy could catalogue the cryptic diversity in planktonic foraminifera far more sufficiently. The traditional way of naming "major" genetic types by roman numbers and less diverged type by roman letters and/or Arabic numbers is inadequate in this regard and used inconsistently between researchers (e.g. Darling & Wade 2008; Ujiié & Lipps 2009). A combination between distance analyses and phylogenetic reconstructions from multiple automated alignments is most likely to result in a nomenclatorial system that sufficiently represents the genetic diversity in the group. This represents one of the first requirements for a much pleaded DNA taxonomy (e.g. Tautz et al. 2003).

Distribution patterns of genetic types in Globigerinoides ruber

The spatial distribution and abundance of most planktonic foraminiferal species is traditionally correlated to more or less narrow ranges of physical oceanographic parameters (e.g. Tolderlund & Bé 1971). Publications of planktonic foraminiferal SSU rDNA genotypes have therefore tried to correlate their findings with the ambient oceanographic conditions (e.g. de Vargas et al. 2001; Morard et al. 2009). Here, the genotype distribution patterns in *Globigerinoides ruber* are interpreted on the level of species interactions, rather than on physical habitat parameters as sea surface temperature.

From an organismic perspective, this is slightly ambiguous. Of course the distribution of each species and genetic type is controlled by the environmental factors in its habitat. This angle has been covered

very thoroughly in the literature, at least for morphospecies. Yet as the genetic types in the *G. ruber* cluster are most likely closer to actual "biological" species than the morphospecies they are found in, the role of species-species interaction on their distribution becomes more prominent against the influence of the habitat parameters. The patterns of distribution and abundance of the genetic types in *G. ruber* s.str. (Types Pink and Ia) and *G. ruber* s.l. (Types IIa1, IIa2 and IIb) in the Canary region and the Mediterranean Sea suggest that the closer two genetic types are related, the more unlikely their co-occurrence in a stable habitat like the Mediterranean Basins is. It can be speculated that this phenomenon is based on the almost identical habitat requirements of two sibling species (as a logical consequence of the identical requirements of individuals in a single species), resulting in the incapability of one species to intrude into any established and stable population of its sister species.
In highly perturbed areas as the regions influenced by the Azores Front and the Canary current, co-occurrence is more likely as populations of either sibling species are unstable in size and distribution. This can be seen in the distribution of the two most closely related SSU rDNA genotypes reported in planktonic foraminifera, the *G. ruber* s.l. Types IIa1 and IIa2. As the more distant related genotypes Type Pink and Type IIa show no pattern of spatial exclusion but of season dependent abundance differences, these findings are actually an indirect evidence for a (at least) species level signal in the SSU rDNA fragment generally used in the genotyping of planktonic foraminifera. Moreover, the unique occurrence of the newly found genetic type IIb in the Mediterranean Basins suggests that other semi-enclosed margin seas in the warm water belt of the world ocean might harbour genetic types of *G. ruber* that are yet known.

Morphometric evaluation of the cryptic species present in the Globigerinoides ruber *morphotaxon*

Apparently, the application of *G. ruber* as a paleo-proxy and the need for a universally applicable and simple species description have outweighed the attempts to develop a sensible taxonomy of this species. Starting with the deviation from the original species description of *G. ruber* with regard of its shell colouration, up to the extensive variation in adult chamber form that was assigned to the species, the morphospecies definition of *G. ruber* became increasingly broader. However, the number of described ecophenotypes and subspecies attributed to this species indicate the presence of distinguishable forms and call for a better resolved taxonomy.

Our relatively simple morphometric measurements support the separation of *G. ruber* s.l. and *G. ruber* s.str. with regard to their chamber compression in the final whorl. These data further suggest that the individuals assigned to *G. ruber* that yielded the SSU rDNA genotype IIa correspond to the phenotypic description of *G. ruber* s.l. (Wang 2000). Our measurements present the first quantitative comparison of the species definitions of *G. ruber* and *G. elongatus* from museum material. The data indicate that the morphometric separation between individuals of the two species is of the same nature as between Type IIa and Type Pink individuals. We thus assign the IIa genotype to *G. elongatus*, noting that the methodological bias which resulted from the different sources and image quality of our material prevented us from developing an unambiguous quantitative discriminator for this species.
Combining phylogenetic reconstruction of genetic types of *G. ruber* with several alternative fossil hypotheses of the species origin, the results from the molecular clock approach support the separation of *G. ruber* lineage from *G. obliquus* at ~ 12 Ma, rather than the much older origin from *G. subquadratus* at ~22 Ma. The separation of *G. ruber* 'pink' and *G. ruber* 'white' (s.str.) is estimated at about ~7-5 Ma, and thereby much older than the first appearance of the *G. ruber* 'pink' colourmorphs in the fossil record. The results further support our hypothesis that Type IIa represents the genetic type of the *G. conglobatus* sister species *G. elongatus*, instead of being a *G. conglobatus* genotype. We

conclude that the species definition of *G. ruber* should once again follow the original description of d'Orbigny and be attributed to G. ruber 'pink' individuals alone. Specimens of G. ruber 'white' s.str. morphology are left without a valid taxonomic name.

The final conclusions drawn from either of the approaches and interpretations in this thesis are not necessarily "more correct" than traditional interpretations on the planktonic foraminiferal phylogeny, taxonomy and biogeography. They do, however, provide additional view angles rarely embraced or acknowledged in prior publications. In this respect, a holistical, broadminded and integrative research on the genetic diversity in planktonic foraminifera is necessary if the nature, extend and value of the SSU rDNA genotypes in this group is to be understood.

References

Adl SM, Leander BS, Simpson AGB (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**: 684–689.

Alizon S, Kucera M, Jansen VAA (2008) Competition between cryptic species explains variation in rates of lineage evolution. *Proceedings of the National Academy of Sciences, USA*, **105**: 12382–12386.

Amato A, Kooistra WHCF, Levialdi Ghiron JH, Mann DG, Pröschold T, Montresor M (2007) Reproductive isolation among sympatric cryptic species in marine Diatoms. *Protist*, **158**: 193—207.

Aurahs R, Grimm GW, Hemleben V, Hemleben C, Kucera M (2009) Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*. *Molecular Ecology*, **18**: 1692-1706.

Banner FT, Blow WH (1960) The taxonomy, morphology and affinities of the genera included in the subfamily Hastigerininae, *Micropaleontology*, **6**: 19-31.

Bé AWH (1959) Ecology of recent planktonic foraminifera. I. Areal distribution in the western North Atlantic. *Micropaleontology*, **5**: 77–100.

Bé AWH (1967) *Foraminifera: Families: Globigerinidae and Globorotaliidae*. Fiches d'Identification du Zooplancton. Conseil Permanent International pour l'Exploration de la Mer. Zooplankton Sheet, **108**: 1-8.

Bé AWH, Vilks G, Lott L (1971) Winter distribution of planktonic foraminifera between the Grand Banks and the Caribbean. *Micropaleontology*, **17**: 31-42.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, **22**: 148-155.

Bijma, J, Erez, J Hemleben, C (1990) Lunar and semi-lunar reproductive cycles in some spinose planktonic foraminifers. *Journal of Foraminiferal Research*, **20**: 117-127.

Cembella AD, Taylor FJR, Therriault JC (1988) Cladistic analysis of electrophoretic variants whitin the toxic dinoflagellate genus *Protogonyaulax*. *Botanica Marina*, **31**: 39-51.

Cermeño P, Falkowski PG (2009) Controls on diatom biogeography in the ocean. *Science*, **325**: 1539 – 1541.

Cifelli R, Scott G (1986) Stratigraphic record of the Neogene Globorotalid radiation (planktonic Foraminiferida). *Smithsonian Contributions to Paleobiology,* **58**: 1–101.

Coyne JA, Orr HA (2004) *Speciation.* Sinauer, Sunderland, MA.

Darling KF, Kroon D, Wade CM, Brown AJL (1996) Molecular phylogeny of the planktic foraminifera. *Journal of Foraminiferal Research*, **26**: 324–330.

Darling KF, Wade CM, Kroon D, Leigh Brown AJ (1997) Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Marine Micropaleontology,* **30**: 251–266.

Darling KF, Wade CM, Kroon D, Leigh Brown AJ, Bijma J (1999) The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past ocean circulations. *Paleoceanography,* **14**: 3–12.

Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R, Leigh Brown AJ (2000) Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature,* **405**: 43–47.

Darling KF, Kucera M, Pudsey CJ, Wade CM (2004) Molecular evidence links cryptic diversification in polar planktonic protists to quaternary climate dynamics. *Proceedings of the National Academy of Sciences, USA,* **101**: 7657–7662.

Darling KF, Kucera M, von Langen P, Pak D (2003) Seasonal distribution of genetic types of planktonic foraminifer morphospecies in the Santa Barbara Channel and its paleoceanographic implications. *Paleoceanography*, **18**:1032.

Darling KF, Kucera M, Kroon D, Wade CM (2006) A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*. *Paleoceanography,* **21**: PA2011.

Darling KF, Kucera M, Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of Sciences, USA,* **104**: 5002–5007.

Darling KF, Wade CM (2008) The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotype. *Marine Micropaleontology,* **67**: 216–238.

Darling KF, Thomas E, Kasemann SA, Seears HA, Smart CW, Wade CM (2009) Surviving mass extinction by bridging the benthic/planktic divide. *Proceedings of the National Academy of Sciences, USA,* **106**: 12629-12633.

Darwin C (1859) *On the origin of species by means of natural selection, or the preservation of favoured species in the struggle of life.* John Murray, London.

De Vargas C, Zaninetti L, Hilbrecht H, Pawlowski J (1997) Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *Journal of Molecular Evolution*, **45**: 285-294.

De Vargas C, Pawlowski J (1998) Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Molecular Phylogenetics and Evolution*, **9**: 463–469.

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.

De Vargas C, Renaud S, Hilbrecht H, Pawlowski J (2001) Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: genetic, morphologic, and environmental evidence. *Paleobiology*, **27**: 104–125.

De Vargas C, Bonzon M, Rees NW, Pawlowski J, Zaninetti L (2002) A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Marine Micropaleontology*, **45**: 101–116.

De Vargas C, Sáez AG, Medlin LK, Thierstein HR (2004) Superspecies in the calcareous plankton. In: *Coccolithophores — from Molecular Processes to Global Impact* (Eds Thierstein HR, Young J), Springer-Verlag, Heidelberg, Germany: 271–298.

Fenchel T, Finlay BJ (2006) The diversity of microbes: resurgence of the phenotype. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**: 1965–1973.

Gorgescu MD (2009) On the origins of superfamily Heterohelicacea Cushman, 1927 and the polyphyletic nature of planktic foraminifera. *Revista Espagñola de Micropaleontologica*, **40**: 107-144.

Grimm GW, Stögerer K, Ertan KT, Hemleben V, Hemleben C, Kucera M (2007) Diversity of rDNA in *Chilostomella*: molecular differentiation patterns and putative hermit types. *Marine Micropaleontology*, **62**: 75–90.

Healy-Williams N, Ehrlich R, Williams DF (1985) Morphometric and stable isotopic evidence for subpopulations of *Globorotalia truncatulinoides*. *Journal of Foraminiferal Research*, **15**: 242-253.

Hecht AD (1974) Intraspecific variation in recent populations of *Globigerinoides ruber* and *Globigerinoides trilobus* and their application to paleoenvironmental analysis. *Journal of Paleontology*, **48**: 1217-1234.

Hemleben C, Bronniman P, Renz HH (1969) Ultramicroscopic shell and spine structure of some spinose planktonic Foraminifera. In *Proceedings of the First International Conference on Planktonic Microfossils*, *Geneva 1967*. Brill EJ, Leiden.

Hemleben C, Spindler M, Breitinger I, Ott R (1987) Morphological and physiological responses of *Globigerinoides sacculifer* (Brady) under varying laboratory conditions. *Marine Micropaleontology*, **12**: 305-324.

Hemleben C, Spindler M, Anderson OR (1989). *Modern Planktonic Foraminifera*. Springer, New York.

Hills SJ, Thierstein HR (1989) Plio-Pleistocene calcareous plankton biochronology. *Marine Micropaleontology*, **14**: 67-96.

Hippler D, Kozdon R, Darling KF, Eisenhauer A, Naegler TF (2009) Calcium isotopic composition of high-latitude proxy carrier *Neogloboquadrina pachyderma* (sin.). *Biogeosciences*, **6**: 1-14.

Hooper DU, Chapin III FS, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, **75**: 3-35.

Huber, BT, Bijma J, Darling K (1997) Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Paleobiology*, **23**: 33-62.

Hutchinson GE (1961) The paradox of the Plankton. *American Naturalist*, **95**: 137-145.

Kahn MI (1981) Ecological and paleoecological implications of the phenotypic variation in 3 species of living planktonic foraminifera from the northeast Pacific Ocean 50 degrees North 145 degrees west. *Journal of Foraminiferal Research*, **11**: 203-211.

Kennett JP, Srinivasan MS (1983) *Neogene Planktonic Foraminifera: a Phylogenetic Atlas.* Hutchinson Ross Stroudsburg, New York.

Kennett JP (1976) Phenotypic variation in some Recent and Late Cenozoic planktonic Foraminifera. *Foraminifera*, **2**: 111-170.

Kimoto K, Ishimura T, Tsunogai U, Itaki T, Ujiié Y (2009) The living triserial planktic foraminifer *Gallitellia vivans* (Cushman): Distribution, stable isotopes, and paleoecological implications. *Marine Micropaleontology*, **71**: 71-79.

Kroon D, Nederbragt AJ (1990) Ecology and paleoecology of triserial planktic foraminifera. *Marine Micropaleontology*, **16**: 25–38.

Kucera M, Darling K (2002) Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philosophical Transactions of the Royal Society A: Physical Sciences*, **360**: 695–718.

Kucera M, Weinelt M, Kiefer T, Pflaumann U, Hayes A, Weinelt M, Chen M, Mix AC, Barrows TT, Cortijo E, Duprat J, Juggins S, Waelbroeck C (2005) Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans. *Quaternary Science Review*, **24**: 951–998.

Kucera M (2007) Planktonic foraminifera as tracers of past oceanic environments. In: *Developments in Marine Geology, Volume 1*, *Proxies in Late Cenozoic Paleoceanography* (Eds Hillaire-Marcel C, de Vernal A), Elsevier, Amsterdam: 213–262.

Kucera M, Schönfeld J (2007) The origin of modern oceanic foraminiferal faunas and Neogene climate change. *The Micropalaeontological Society Spec Publ 2*, eds M Williams M, Haywood AM, Gregory FJ, Schmidt DN (The Micropalaeontological Society, London): 409–426.

Kuroyanagi A, Tsuchiya M, Kawahata H, Kitazato H (2008) The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleo-environmental implications. *Marine Micropaleontology*, **68**: 236–243.

Langer MR, Lipps JH, Piller WE (1993) Molecular paleobiology of protists – amplification and direct sequencing of foraminiferal DNA. *Micropaleontology*, **39**: 63-68.

Lazarus D, Hilbrecht H, Spencer-Cervato C, Thierstein H (1995) Sympatric speciation and phyletic change in *Globorotalia truncatulinoides*. *Paleobiology*, **21**: 28-51.

Liu C, Olsson RK (1992) Evolutionary adaptive radiation of microperforate planktonic foraminifera following the K/T mass extinction event. *Journal of Foraminiferal Research,* **22**: 328–46.

Liu C, Olsson RK (1994) On the origin of Danian normal perforate planktonic foraminifera from *Hedbergella*. *Journal of Foraminiferal Research*, **24**: 61–74.

Mayr E (1970) Populations, species, and evolution: an abridgment of Animal species and evolution. Belknap Press Series, Harvard University Press, **7**.

McLaren IA, Woods SM, Shea Jr. JR (1966) Polyteny: A source of cryptic speciation among Copepods. *Science*, **153**: 1641 – 1642.

Merlé C, Moullade M, Lima O, Perasso R (1994) Essai de caractérisation phylogénétique des foraminiféres planctoniques à partir de séquences partielles d'ADNr 28S. Les Comptes Rendus de l'Académie des sciences (Ser. II), **319**:149-153.

Morard R, Quillévéré F, Escarguel G, Ujiié Y, de Garidel-Thoron T, Norris RD, de Vargas C (2009) Morphological recognition of cryptic species in the planktonic foraminifer *Orbulina universa Marine Micropaleontology*, **71**: 148-165.

Moullade M (1964) Pour une simplification de la taxonomie des foraminiferes appartenant à la superfamille des Globigerinacea. *Compte Rendu Somaire des Séances de la Société Géologique de France*, **1964**: 58-60.

Neefs JM, Van de Peer Y, Hendriks L, De Wachter R (1990) Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Research*, **18**: 2237-2317.

Norris RD (1991) Biased extinction and evolutionary trends. *Paleobiology*, **17**: 388-399.

Norris RD, Corfield RM, Cartlidge J (1996) What is gradualism? Cryptic speciation in globorotaliid foraminifera. Paleobiology, **22**: 386-405.

Norris RD (2000) Pelagic species diversity, biogeography, and evolution. *Paleobiology*, **26**: 236-258.

Olsson RK, Hemleben C, Berggren WA, Liu C (1992) Wall texture classification of planktonic foraminifera genera in the Lower Danian. *Journal of Foraminiferal Research,* **22**: 195–213.

Olsson DR, Hemleben C, Berggren WH, Huber B (1999) *Atlas of Paleocene Planktonic Foraminifera.* Smithsonian Contribution to Paleobiology; Smithsonian Institution press, Washington, **85**: 249.

Olsen-Stojkovich J, West JA, Lowenstein JM (1986) Phylogenetics and biogeography in the Cladophorales complex (Chlorophyta) – some insights from immunological distance data. *Botanica Marina*, **29**: 239-249.

Orbigny A D' (1826) Tableau Méthodique de la classe des Céphalopodes: Annales Des Sciences Naturelles, Paris, Ser. 1, tome **7**: 96-314.

Orr WN (1969) Variation and distribution of *Globigerinoides ruber* in the Gulf of Mexico. *Micropaleontology*, **15**: 373-379.

Pawlowski J, Bolivar I, Guiard-Maffia J, Gouy M (1994) Phylogenetic position of foraminifera inferred from LSU rRNA gene sequences. *Molecular Biology and Evolution*, **11**: 929-938.

Pawlowski J, Bolivar I, Fahrni J, Cavalier-Smith T, Gouy M (1996) Early origin of foraminifera suggested by SSU rRNA gene sequences. *Molecular Biology and Evolution*, **13**: 445-450.

Palumbi SR (1992) Marine speciation on a small planet. *Trends in Ecology and Evolution*, **7**: 114-118.

Palumbi SR (1994) Genetic divergence, reproductive isolation and marine speciation. *Annual Review of Ecology and Systematics*, **25**: 547-572.

Parker FL (1962) Planktonic foraminiferal species in Pacific sediments. *Micropaleontology*, **8**: 219–254.

Pearson PN, Shackleton NJ, Hall MA (1997) Stable isotopic evidence for the sympatric divergence of *Globigerinoides trilobus* and *Orbulina universa* (planktonic foraminifera). *Journal of the Geological Society***, 154** : 295-302.

Pearson PN, Olsson RK, Hemleben C, Huber B, Berggren WA (2006) *Atlas of Eocene Planktonic Foraminifera. Cushman Foundation Special Publication*. Cushman Foundation of foraminiferal research Inc.; Fredricksburg, Virginia, USA, **41**.

Pfenninger M, Schwenk K (2007) Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, **7**: 121.

Rutherford S, D'Hondt S, Prell W (1999) Environmental controls on the geographic distribution of zooplankton diversity. *Nature*, **400**: 749-753.

Saez AG, Lozano E, 2005. Body doubles. *Nature*, **433**: 111.

Saito T, Thompson PR, Berger D (1981) *Systematic Index of Recent and Pleistocene Planktonic Foraminifera*. University of Tokyo Press, Tokyo, Japan.

Schiebel R, Hemleben Ch (2005) Modern planktic foraminifera. *Paläontologische Zeitschrift,* **79**:135–148.

Schilthuizen M (2000) Dualism and conflicts in understanding speciation. *Bioessays*, **22**: 1134-1141.

Sexton PF, Norris RD (2008) Dispersal and biogeography of marine plankton: Long-distance dispersal of the foraminifer *Truncorotalia truncatulinoides*. *Geology*, **36**: 899-902.

Stewart IA, Darling KF, Kroon D, Wade CM, Troelstra SR (2001) Genotypic variability in subarctic Atlantic planktic foraminifera. *Marine Micropaleontology,* **43**: 143–153.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution*, **18**: 70-74.

Tolderlund DS, Bé AWH (1971) Seasonal distribution of planktonic foraminifera in the western North Atlantic. *Micropaleontology*, **17**: 297–329.

Ujiié Y. Kimoto K, Pawlowski J (2008) Molecular evidence for an independent origin of modern triserial planktonic foraminifera from benthic ancestors. *Marine Micropaleontology*, **69**: 334-340.

Ujiié Y, Lipps JH (2009) Cryptic diversity in planktic Foraminifera in the northwest Pacific ocean. *Journal of Foraminiferal Research*, **39**: 145-154.

Van De Peer Y, Nicolai S, De Rijk I, De Wachter R (1996) Database on the structure of small ribosomal subunit RNA. *Nucleic Acids Research*, **24**: 86-91.

Wade M (1964) Application of the lineage concept to biostratigraphic zoning based on planktonic Foraminifera. *Micropaleontology*, **10**: 273-290.

Wade CM, Darling KF, Kroon D, Brown AJL (1996) Early evolutionary origin of the planktic foraminifera inferred from small subunit rDNA sequence comparisons. *Journal of Molecular Evolution*, **43**: 672-677.

Walker DA, Vilks G (1973) Spinal ultrastructure of the planktonic foraminifers *Hastigerina* and *Globigerinella*. *Journal of Foraminiferal Research*, **3**: 196-198.

Wang LJ (2000) Isotopic signals in two morphotypes of *Globigerinoides ruber* (white) from the South China Sea: implications for monsoon climate change during the last glacial cycle. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **161**: 381–394.

Wei KY (1994) Stratophenetic tracing of phylogeny using SIMCA pattern recognition technique; a case study of the late Neogene planktic foraminifera *Globoconella* clade. *Paleobiology*, **20**: 52–65.

Winge H (1965) Interspecific hybridisation between the six cryptic species of *Drosophila willistoni* group. Heredity, **20**: 9–19.

Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K, De Vos P, Verstraete W, Boon N (2009) Initial community evenness favours functionality under selective stress. *Nature,* **458**: 623-626.

Wray CG, Langer M R, De Salle R, Lee JJ, Lipps JH (1995) Origin of the foraminifera. *Proceedings of the National Academy of Sciences, USA,* **92**:141-145.

Wuyts, J, De Rijk P, Van de Peer Y, Pison G, Rousseeuw P, De Wachter R (2000) Comparative analysis of more than 3000 sequences reveals the existence of two pseudoknots in area V4 of eukaryotic small subunit ribosomal RNA**.** *Nucleic Acids Research*, **28**: 4698–4708.

Wuyts J, Van de Peer1 Y, Winkelmans T, De Wachter R (2002) The European database on small subunit ribosomal RNA. *Nucleic Acids Research*, **30**: 183-185.

Acknowledgment

This thesis would not have come into being (and most certainly not to an end) without the ideas, help and constructive critisism of many people. I am very greatful for their professional and personal support, and for making the process far more worthwhile than it could have been.

My special thanks go to

My supervisor, Prof . Dr. Michal Kučera, for his scientific guidance through the realm of micropaleontology, his patience when things were not going according to plan. I am particularly thankful that there was always time to talk things through. Sometimes twice.

Prof. Dr. Vera Hemleben and Prof. Dr. Christoph Hemleben, who both were incredibly generous with their time and scientific expertise and for acting as ‘honorary’ supervisors.

Prof. Dr. Ralf Schiebel, for being so kind to evaluate this thesis as second referee on relatively short notice.

Dr. Guido Grimm, who jump started me into my thesis, and managed to teach me one or two things along the way. His ‘tree magic’ was much appreciated.

Karin Stögerer, who set the tone on working moral and kept the lab running while I kept displacing her equipment.

Margret Bayer for her practical help, technical assistance and organisatory talent prior to and during cruises , as well as for her very heathy perspective on the identification of planktonic foraminifera.

Publication 1: Using the multiple analysis approach to reconstruct phylogenetic relationships among planktonic foraminifera from highly divergent and length-polymorphic SSU rDNA sequences

Bioinformatics and Biology Insights

METHODOLOGY

Using the Multiple Analysis Approach to Reconstruct Phylogenetic Relationships among Planktonic Foraminifera from Highly Divergent and Length-polymorphic SSU rDNA Sequences

Ralf Aurahs[1], Markus Göker[2,3], Guido W. Grimm[4], Vera Hemleben[5], Christoph Hemleben[1], Ralf Schiebel[6] and Michal Kučera[1]

[1]Department of Micropaleontology, Institute of Geosciences, Eberhard Karls University of Tübingen, Sigwartstraβe 10, 72076 Tübingen, Germany. [2]Organismic Botany, Eberhard Karls University of Tübingen, Auf der Morgenstelle 1, 72076 Tübingen, Germany. [3]DSMZ—German Collection of Microorganisms and Cell Cultures, Inhoffenstraβe 7 B, 38124 Braunschweig, Germany. [4]Department of Palaeobotany, Natural History Museum, Box 50007, 10405 Stockholm, Sweden. [5]Department of General Genetics, Centre of Plant Molecular Biology (ZMBP), Eberhard Karls University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany. [6]Laboratoire des Bio-Indicateurs Actuels et Fossiles, University of Angers, 2 bd Lavoisier, 49045 Angers, France. Email: ralf.aurahs@uni-tuebingen.de

Abstract: The high sequence divergence within the small subunit ribosomal RNA gene (SSU rDNA) of foraminifera makes it difficult to establish the homology of individual nucleotides across taxa. Alignment-based approaches so far relied on time-consuming manual alignments and discarded up to 50% of the sequenced nucleotides prior to phylogenetic inference. Here, we investigate the potential of the multiple analysis approach to infer a molecular phylogeny of all modern planktonic foraminiferal taxa by using a matrix of 146 new and 153 previously published SSU rDNA sequences. Our multiple analysis approach is based on eleven different automated alignments, analysed separately under the maximum likelihood criterion. The high degree of congruence between the phylogenies derived from our novel approach, traditional manually homologized culled alignments and the fossil record indicates that poorly resolved nucleotide homology does not represent the most significant obstacle when exploring the phylogenetic structure of the SSU rDNA in planktonic foraminifera. We show that approaches designed to extract phylogenetically valuable signals from complete sequences show more promise to resolve the backbone of the planktonic foraminifer tree than attempts to establish strictly homologous base calls in a manual alignment.

Keywords: planktonic foraminifera, phylogeny, fossil record, automated alignment

Bioinformatics and Biology Insights 2009:3 155–177

This article is available from http://www.la-press.com.

Introduction

DNA sequences coding for the 3' segment of the small-subunit ribosomal RNA (SSU rDNA) have been broadly used to evaluate phylogenetic relationships among living planktonic Foraminifera.[1–13] SSU rDNA data stored in international databases are in most cases sufficient to determine the systematic affinity of an unknown SSU rDNA fragment derived from a planktonic foraminifer using the blast algorithm.[14] This is caused by two general characteristics of planktonic foraminiferal SSU rDNA sequences: (i) a higher intraspecific and interspecific variability in SSU rDNA regions which are generally conserved among most other foraminiferal lineages; and (ii) diagnostic sequences in SSU rDNA regions that are highly divergent between and among all major foraminiferal lineages.[8,15,16] Those general characteristics nourished the hope that SSU rDNA data could be useful to address the evolutionary unfolding of all planktonic foraminifers.

However, phylogenetic inference has been hindered by the fact that the highly divergent SSU rDNA regions, which are of high taxonomic and phylogenetic value (Fig. 1), cannot be unambiguously aligned for all planktonic foraminifera. As a consequence, only up to 600 of the approximately 1,000 to 1,200 nucleotides of the more informative and thus commonly sequenced 3' segment of the SSU rDNA have been used for phylogenetic studies of higher taxa in planktonic foraminifera (Fig. 1). In general, aligning noncoding sequences such as rDNA is more difficult than using protein-coding DNA fragments which are structured by reading frames and have most variability concentrated at third base positions within codons.[17]

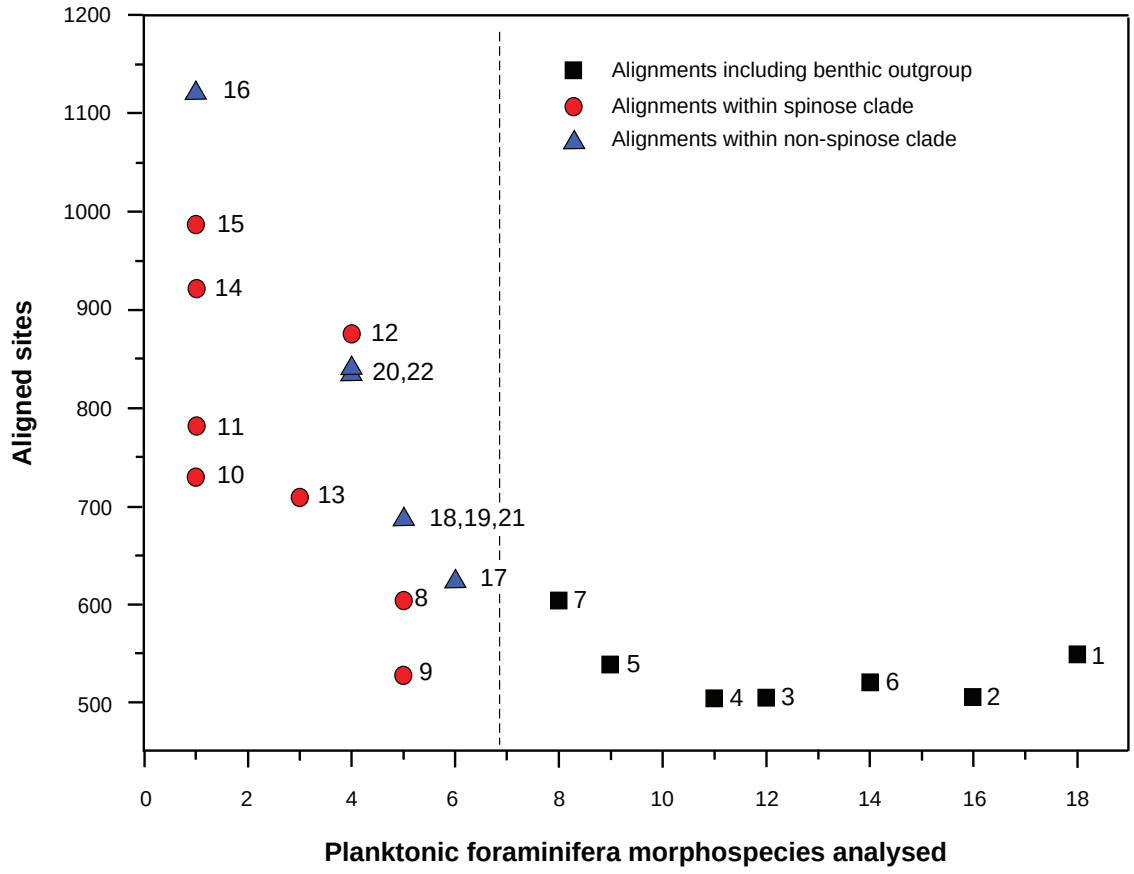

Figure 1. Lengths of manual alignments used to infer the phylogeny of planktonic foraminifera. Summary of planktonic foraminifera molecular phylogenies based on the 3' fragment of the SSU rDNA gene. Almost one half of the ~1000 bp in the analysed fragment are lost when attempting to align "unambiguously" across the entire clade. The remaining variable regions clearly contain phylogenetically useful information, as can be seen by the longer alignments produced for subclades including only selected species. This phylogenetic information is lost when aligning across the three major clades of planktonic foraminifera, or when the alignment includes benthic outgroups. Data sources (in chronological order): 1997, Darling et al[2] [**7**], Huber et al[4] [**8**], de Vargas et al[3] [**3**]; 1999, Darling et al[7] [**5**]; 2000, Darling et al[9] [**4**]; 2001, Stewart et al[11] [**3**], de Vargas et al[10] [**16,17**]; 2002, de Vargas et al[69] [**9**]; 2003, Darling et al[70] [**10,11,18**]; 2004, Darling et al[51] [**19,20**]; 2006, Darling et al[54] [**2,21**]; 2007, Darling et al[71][**22**]; 2008, Kuroyanagi et al[72] [**12**], Ujiié et al[73] [**1**]; 2009, Aurahs et al[74] [**13,14,15**].

Among foraminifera, this situation is aggravated by the fact that their SSU rDNA includes sequence strands ("expansion segments") not found in the SSU of any other eukaryote up to now.[18,19] These expansion segments are of unknown transcriptional fate, as the mature SSU rRNA of foraminifera has not been sequenced to date. Accordingly, any conclusions drawn from the reconstruction of the secondary structure have so far been of limited merits for inferring high-quality sequence alignments in this group of organisms. A further intricacy is that not all planktonic foraminiferal lineages exhibit the same level of sequence divergence from the common foraminiferal SSU rDNA motive. Nonspinose macroperforate and microperforate taxa in general yield SSU rDNA sequences which appear more similar to their benthic relatives than spinose taxa, as illustrated by short branches in phylogenetic trees and a generally low support for all backbone nodes.[5,7,9,11,20]

In this situation, methods are urgently needed that avoid discarding phylogenetically valuable alignment positions but can also cope with the challenge of properly aligning those regions. In fact, the culling of alignment-ambiguous regions does not take into account that different possible alignment solutions do not necessarily imply different topologies or support values.[21] Furthermore, exclusion of characters is often done on subjective grounds and further reduces the reproducibility of the results,[22] which is frequently already low when an alignment is constructed manually, even if the secondary structure is considered. Consequently, Lee[21] advocated an approach based on the generation of several alignments by the same algorithm but under different parameter combinations, which he called "multiple analysis method". In multiple analysis, trees are inferred separately from the respective alignments and only relationships that are well supported in all (or most) of the trees are accepted.[23] Another possibility is to use different alignment algorithms under default values, as did Morrison and Ellis[24] and Kemler et al.[25] The latter approach appears to have several advantages; for instance, one would expect the differences between distinct alignment programs to be higher than those between different parameterisations of the same algorithm. That is, a larger proportion of the alignment space could be explored by running distinct programs. In addition, some parameterisations are simply not biologically reasonable, as, e.g. a scoring matrix that gives higher implicit weight to transversions than to transitions. Furthermore, current alignment algorithms and their default settings are constantly improved using benchmark tests (references for the individual programs are provided in Material and Methods below). Using the most recent version of the software out of the box, i.e. with default settings, is a straightforward approach to the sequence homology problem. In theory, sequence alignment cannot be considered separately from phylogenetic inference (e.g. many alignments programs use a guide tree), but both problems are NP-hard[26,27] and in practice most researchers have regarded tree building as a distinct step (but see[28–30]).

Despite the number of SSU rDNA sequences available, our knowledge of the actual diversity of planktonic foraminiferal SSU rDNA is still very limited (Table 1). Important taxa such as *Globorotalia*, including deep-dwelling species with relatively long reproductive cycles,[32] *Globigerinita*, the to date only sequenced representative of the extant microperforate group, *Hastigerina pelagica*, the largest and morphologically most aberrant modern planktonic foraminifer, and most other spinose taxa save *Globigerinella siphonifera* and *Orbulina universa* are represented by single to few sequences in public databases.[4,6] As a consequence, their genetic variability is not yet known to a sufficient degree. For about 20 planktonic foraminiferal species, i.e. half of the extant diversity in this group, no (reliable) sequence data are available yet (Table 1).

The collection of these species for DNA analyses from plankton samples has been hampered by their small size and relatively low abundance. The taxonomy (and classification; Table 1) of planktonic foraminifera is (still) based on the morphological characters of their calcite shells. Planktonic foraminiferal shells grow by sequential addition of proportionately larger chambers, typically along a trochospiral coil. The shape of individual chambers and the pattern of their addition can change considerably through ontogeny.[33] Current taxonomic concepts are based on shells recovered from surface sediments. Such shells represent mature adult individuals that exhibit specific morphological characters. Living specimens afloat in the plankton, however, represent a range of mostly pre-adult ontogenetic stages that are lacking important taxonomic characters. Thus, it is possible

Table 1. Species of planktonic foraminifers. A list of all planktonic foraminifera species included in this study; and their representation by SSU rDNA data in public databases and newly assembled data.

Species	SSU data available[§]	New data added
Microperforate clade (= Candeinidae Saito and Thompson 1982)		
Candeina nitida d'Orbigny 1839	No	No
Globigerinita glutinata (Egger 1893)	Yes	Yes[†]
G. minuta (Natland 1938)	No	No
G. uvula (Ehrenberg 1861)	Singleton	Yes[†]
Tenuitella fleisheri Li 1987	No	No
T. iota (Parker 1954)	No	No
T. parkerae (Brönnimann and Resig 1971)	No	No
Nonspinose clade (= Globorotaliidae Cushman 1927)		
Berggrenia pumilio (Parker 1962)	No	No
Globoquadrina conglomerata (Schwager 1866)	No	No
Globorotalia anfracta (Parker 1967)	No	No
G. cavernula Bé 1967	No	No
G. crassaformis (Galloway and Wissler 1927)	Singleton*	No
G. hirsuta (d'Orbigny 1839)	Singleton	Yes
G. inflata (d'Orbigny 1839)	Singleton	Yes
G. menardii (d'Orbingy 1826)	Yes	No
G. scitula (Brady 1882)	No	No[‡]
G. truncatulinoides (d'Orbigny 1839)	Yes	Yes[†]
G. theyeri Fleisher 1974	No	No
G. tumida (Brady 1877)	No	No
G. ungulata Bermudez 1960	No	No
Globorotaloides hexagonus (Natland 1938)	No	No
Neogloboquadrina dutertrei (d'Orbingy 1826)	Yes	No
N. incompta (Cifelli 1961)	Yes	Yes
N. pachyderma (Ehrenberg 1861)	Yes	No
Pulleniatina obliquiloculata (Parker and Jones 1862)	Yes	No
Spinose bilamellar clade (= Globigerinidae Carpenter, Parker and Jones 1876)		
Beela digitata (Brady 1879)	No	No
Globigerina bulloides d'Orbigny 1826	Yes	No
G. falconensis Blow 1959	Yes	No
Globigerinella adamsi (Banner and Blow 1959)	No	No
G. calida (Parker 1962)	Singleton*	No
G. siphonifera (d'Orbigny 1839)	Yes	Yes
Globigerinoides conglobatus (Brady 1879)	Yes	No
G. ruber (d'Orbigny 1839)	Yes, biphyletic	No
G. sacculifer (Brady 1877)	Yes	No

(*Continued*)

Table 1. (*Continued*)

Species	SSU data available[§]	New data added
Globoturborotalita rubescens Hofker 1956	No	No
G. tenella (Parker 1958)	No	No
Orbulina universa d'Orbigny 1839	Yes	No
Sphaerodinella dehiscens (Parker and Jones 1865)	No	No
Turborotalita clarckei (Roegl and Bolli 1973)	No	No
T. humilis (Brady 1884)	No	No
T. quinqueloba (Natland 1938)	Yes	No
Spinose monolammelar clade (= Hastigerinidae Saito and Thompson 1976)		
Hastigerina pelagica (d'Orbigny 1893)	Singleton	Yes[†]
Hastigerinella digitata (Rhumbler 1911)	No	No
Orcadia (Hastigerinella) riedeli (Roegl and Bolli 1973)	No	No

*These singletons are possibly not representative for the assigned species.
[†]The new data revealed new sequence (sub)types.
[‡]The new data includes sequences from a globorotaliid specimen, which may be *G. scitula* or not.
[§]Available in public databases at the time of data mining (October 2008). A SSU rDNA sequence of *C. nitida* is available since the end of 2008.[69]

that new, potentially extremely divergent SSU rDNA types will be found among not yet or not sufficiently sampled species, underscoring the need for phylogenetic approaches capable of objective and robust phylogenetic inference from divergent sequences.

In this study, we report new SSU rDNA data of planktonic foraminifera from the Azores Current System and the Mediterranean, including several new sequence types (Table 1). Our data is combined with the SSU rDNA stored in public databases (available until October 2008) and investigated using the multiple analysis approach as described above. This enables us (i) to combine the new and known planktonic foraminiferal SSU rDNA sequence types in reproducible approaches to phylogenetic analysis using all available sequence information in a time-efficient way, and (ii) to re-assess the phylogenetic relationships among planktonic foraminiferal lineages in comparison with earlier manual-alignment based work and evidence from the uniquely complete fossil record of these organisms.

Material and Methods

Sampling and DNA extraction

Live foraminifera in the Northwest Atlantic and the Mediterranean were sampled on RV Poseidon (P283/2, P308) and Meteor (M69/1) cruises using a multiclosing net (100 μm mesh size, sampling down to 700 m) and by filtering surface water from the ship's uncontaminated seawater supply (65 μm mesh size). Specimens were isolated under an incident stereomicroscope (50-fold magnification), and taxonomically identified on board. After mechanical cleaning, single specimens were transferred to Eppendorff cups where the DNA was extracted following the DOC method from Holzmann and Pawlowski.[34] Specimens were crushed in 50 μl of the DOC lysis buffer and incubated on a shaker table at 60 °C for one hour. Samples were than kept at –20 °C until PCR at the home based laboratory. Voucher information including the originally assigned morphotype and collection locality is provided in the Additional file 1.

Data sources

GenBank data

SSU rDNA data of planktonic foraminifers were downloaded from the GenBank/NCBI taxonomy query portal (http://www.ncbi.nlm.nih.gov/; GWG, 28/10/2008).

Newly assembled data

Fragments of the 3' SSU rDNA were amplified by PCR with Vent® (New England Biolabs) polymerase using the primers S14f1,[8] U/T20r1, U/A14f1,[35] for later cloning and the new pelvF (5'TGACTCAACGCGG GAAATCT3') and pelvR (5'CCGGGACATCTAAG

GGCATCAC3') primer pair for direct sequencing of few specimens of *Hastigerina pelagica*. PCR products were purified using the QIAquick gel extraction kits (Qiagen). Ligation and transformation relied on a pUC18/*E. coli* DH5α vector system. Genetic variability within single individuals was determined by sequencing up to five clones per individual and analysing PCR products obtained from several individuals per morphospecies where possible. Nucleotide sequencing was carried out in both directions with ABI 377 automatic sequencer (Perkin Elmer) using the standard vector primers M13uni and M13rev, or by a professional lab (Agowa, Berlin). The newly assembled SSU rDNA sequences have been uploaded to GenBank (accession numbers are provided in the Additional file 1).

Alignments and phylogenetic inference

Multiple sequence alignments were inferred using six different software packages, CLUSTALW version 2.0,[36,37] KALIGN version 2.03,[38] MAFFT version 6.24,[39] MUSCLE,[40] the NRALIGN derivative of MUSCLE which uses an improved scoring function that considers neighbouring residues,[41] and POA.[42] CLUSTALW was run either in default mode or with the gap opening and extension parameters optimized for RNA alignments (using the command-line switches -pwgapopen = 22.5 -gapopen = 22.5 -gapext = 0.83 -pwgapext = 0.83; henceforth referred to as CLWOPT).[43] MAFFT was applied with the command-line switch-maxiterate 1000 and either default settings otherwise (henceforth called MAFFT), -localpair (LINSI), -genafpair (EINSI) or -globalpair (GINSI). POA was run in both default and global scoring mode (applying the command-line switch -do_global; henceforth referred to as POAGLO) using the blosum80_trunc.mat substitution matrix delivered with the software and extended to include the complete nucleotide ambiguity code (the matrix is contained in Additional file 2). Accordingly, a total of eleven alignments were examined (included in Additional file 2).

Phylogenetic trees were inferred from the eleven alignments (without further processing such as a manual re-alignment or manual exclusion of sites) under the maximum likelihood (ML) criterion with RAxML version 7.04.[31,44] RAxML has been specifically designed to efficiently handle large to extremely large datasets and infers phylogenetic trees with ML values at least as large as comparable contemporary programs. To establish node support, we used RAxML's novel fast bootstrap option and 100 replicates in conjunction with the GTRMIX option (command-line switches -m GTRMIX -f a -# 100). GTRMIX applies the fast and memory-efficient GTRCAT model approximation during tree search but estimates the final log Likelihood and branch lengths under GTR + GAMMA.[31,45] The fast bootstrapping has been shown to result in values close to standard bootstrapping, but also in an approximately ten-fold increase in performance.[44] RAxML automatically infers a globally best (best-known) ML tree from the individual bootstrap trees in this running mode.

In the case of alignment-ambiguous data, the effects of different underlying alignment algorithms on phylogenetic reconstruction are usually greater than the effect of the different inference methods.[24] Therefore, one might argue that it is sufficient to apply only the consistent and robust maximum likelihood (ML) criterion to infer phylogenetic trees. Nevertheless, to assess the effect of applying another phylogenetic optimality criterion, we calculated bootstrap support under maximum parsimony (MP) with PAUP* version 4b10.[46] For each of the 100 bootstrap replicates, 10 random sequence addition replicates were conducted, saving only one tree per run. To compare the methods, MP support values were mapped on the corresponding ML trees for each alignment (Additional file 2).

For displaying bootstrap support values, we identified the most representative of the eleven best ML trees inferred from the distinct alignments. This was done by calculating all-against-all Robinson-Foulds distances between the best trees using PAUP* version 4b10 and determining the tree with, on average, the smallest distances to each of the other trees.[46,47] The Robinson-Foulds distance between two trees is defined as the sum of the number of splits (bipartitions) present in one tree but not in the other. Support values from all bootstrap runs were mapped on the most representative tree using RAxML's -f b command-line switch and integrated in one tree file using a UNIX shell script written by MG. For the trees, we also reported the final estimate for the alpha value of the gamma distribution and the log likelihood values of the best trees inferred with RAxML.

In order to quantitatively compare the alignments, we determined their total length. We additionally

classified them using the alignment comparison metric (overlap score) as implemented in MUMSA version 1.0,[48] which also infers UPGMA dendrograms from these similarity values. A corresponding UPGMA classification of the RAxML trees was inferred from their Robinson-Foulds distances with PAUP*.[46,47] To quantify the agreement of the phylogenetic trees with the current taxonomy of planktonic foraminifers, the affiliations of sequences to species were coded as a multi-state pseudocharacter (with one character state per species) for use under the maximum parsimony criterion.[49,50] Newly obtained sequences from undetermined specimens and GenBank accession lacking a valid species name in their organism entry (e.g. "Orbulina sp. 'isolate A102'") were coded as missing data. The parsimony score of each of the best ML trees under this matrix (which we call „T-score") was determined with PAUP*, higher scores indicating lower agreement. The pseudocharacter matrix is contained in Additional File 2.

Results and Discussion

Comparison of multiple sequence alignments

The features of the inferred alignments and ML trees are shown in Table 2. Considerable differences regarding alignment length, estimated alpha values of the gamma distribution and highest obtained likelihood values were observed. This is in accordance with the prediction that the use of different alignment programs, instead of using a single software under a range of parameters, is sufficient to cover a large proportion of the alignment space. Here, CLUSTALW results in the shortest SSU rDNA alignment and MUSCLE in the longest. Classifications of the eleven approaches based on the alignments as well as the inferred trees are shown in Figure 2. The relationships indicated by the Robinson-Foulds distances between the best ML trees do not exactly mirror the relationships between the alignments as measured using the overlap score. For instance, the POA and POAGLO alignments are similar to each other (Fig. 2, right), but the POA-based ML tree is more similar to the CLUSTALW-based trees than to the POAGLO-based tree with respect to Robinson-Foulds distances (Fig. 2, left). On the other hand, the MAFFT-, EINSI-, GINSI- and LINSI-based trees are clustering together, as do their underlying alignments. Our observations on alignment and topological comparison measures are important for future multiple analysis studies as far as they indicate that the

Table 2. Features of the alignments and phylogenetic trees. This table lists features of the eleven sequence alignments constructed and the resulting phylogenetic trees. The entire alignment length is shown. For the resulting best ML trees, the final estimate for the alpha value of the gamma distribution and the log likelihood of the best tree are shown, as well as the sum of the Robinson-Foulds (RF) distances of each tree to the other nine trees and the agreement with the affiliation of sequences to morphospecies (T-score; lower scores indicate better agreement). Note that the likelihood of the best tree cannot directly be used to select the best alignment, because common ML functions as those implemented in RAxML do not consider gaps.

Alignment software	Alignment length	Final alpha value	Highest Log likelihood	Sum of RF distances to other trees	T-score
CLUSTALW	**1384**	**0.93969**	**–3,582,498,665**	**3496**	**23**
CLWOPT	1557	0.97349	–3,598,746,746	3416	25
EINSI	1786	0.48367	–3,012,840,593	3194	23
GINSI	1837	0.48314	–2,849,473,664	3206	23
KALIGN	**1905**	**0.62220**	**–3,251,648,372**	**3482**	**23**
LINSI	1751	0.53379	–3,069,451,219	3226	23
MAFFT	**1965**	**0.54546**	**–3,075,848,970**	**3032**	**23**
MUSCLE	2192	0.82643	–5,422,632,153	4126	25
NRALIGN	**1797**	**0.75213**	**–4,765,997,803**	**3772**	**23**
POA	1856	0.60630	–3,203,410,297	3356	23
POAGLO	**1840**	**0.67321**	**–3,506,284,042**	**3374**	**23**

Alignments considerd for Results and Discussion in bold font.

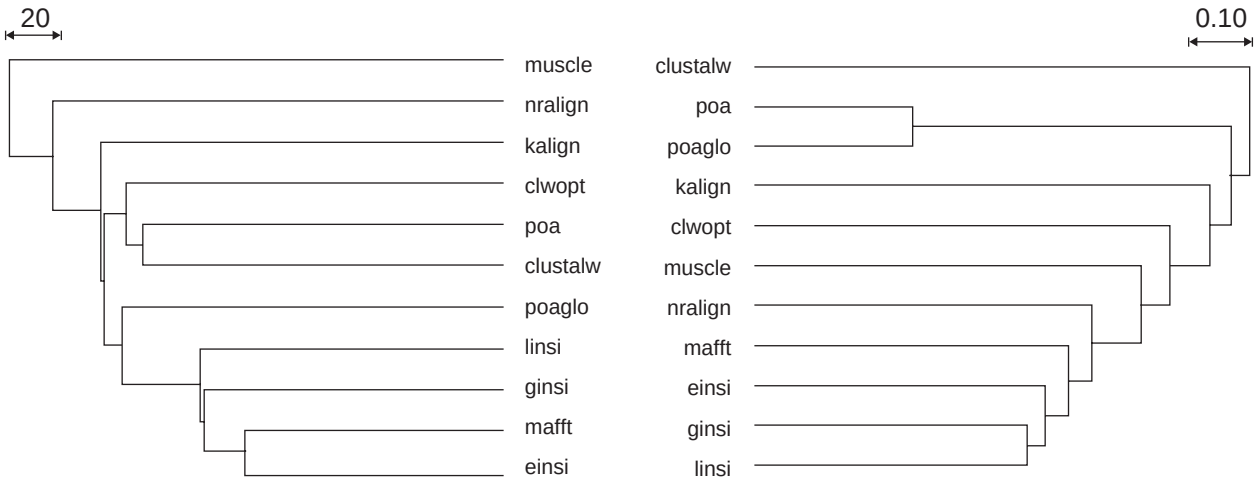

Figure 2. Comparison of alignments and trees. UPGMA dendrograms inferred from overlap scores between sequence alignments (right) and from Robinson-Foulds distances between the corresponding trees (left) are shown. Based on this comparison, EINSI, GINSI and LINSI were not considered further because they are too close to the MAFFT approach. MUSCLE and CLWOPT were omitted because they resulted in some sequences being severely misplaced (see text). Apparently, tree topology can partially (mainly the close relationship of EINSI, GINSI, LINSI and MAFFT) be predicted by the comparison of the underlying sequence alignments.

shape of the tree cannot always be predicted from the descriptive characteristics of the alignment, at least in the case where highly divergent sequences are considered.

Regarding the agreement with morphotaxonomy, the best (minimal) T-score observed is 23, obtained by nine of the eleven alignments (Table 2). This again is in agreement with the prediction that the use of alignment programs under default values, instead of using a single software under a range of parameters, results in biologically reasonable alignments that do not contradict previous taxonomic knowledge. The fact that even the best obtained T-scores are three steps larger than the minimum possible score of 20 (corresponding to 21 pseudocharacter states) is caused by three mislabelled sequences, whereas scores higher than 23 are due to misaligned sequences (shown below). Thus, trees inferred from MUSCLE and CLWOPT achieving T-scores of 25 were not further considered for displaying trees and drawing conclusions on foraminifer evolution. The particularly low likelihood observed for the MUSCLE tree could also be caused by one to several sequences being severely misaligned. However, the likelihood of the best tree cannot directly be used to select the best alignment, because common ML functions, as those implemented in RAxML, do not consider gaps. Also, EINSI, GINSI, and LINSI were not considered further because they were too close to MAFFT regarding both alignment and topological similarity (Fig. 2). ML bootstrap results from the six selected alignments were mapped on the MAFFT tree (Fig. 3), which was the most central one (the least distant from all other trees), irrespective of whether EINSI, GINSI, and LINSI were considered or not.

A comprehensive table of well-supported (ML/MP) and/or systematically relevant phylogenetic splits is provided as supplement (Additional file 3); all alignments and trees are included in Additional file 2. In general, ML and MP support the same phylogenetic splits (bipartitions), although the support under MP is often lower than under ML using the same alignment. At the species level or higher, ML supports 23 bipartitions with high support based on all six alignments ($BS_{ML} \geq 80$), and four more if only five out of the six alignments are considered. Using MP as optimality criterion 22 bipartitions are highly supported based on all six alignments, and an additional one based on five out of six alignments. In all remaining bipartitions, high ML bootstrap support correlates to moderate MP bootstrap support. Only two exceptions were observed: In one case, KALIGN-based ML bootstrap support is low ($BS_{ML} = 12$), and MP high ($BS_{MP} = 100$). In the other, the situation is vice versa ($BS_{ML} = 89$; $BS_{MP} = 12$). In both cases, short sequences are involved. It appears that the portion of missing data, in combination with the KALIGN-generated

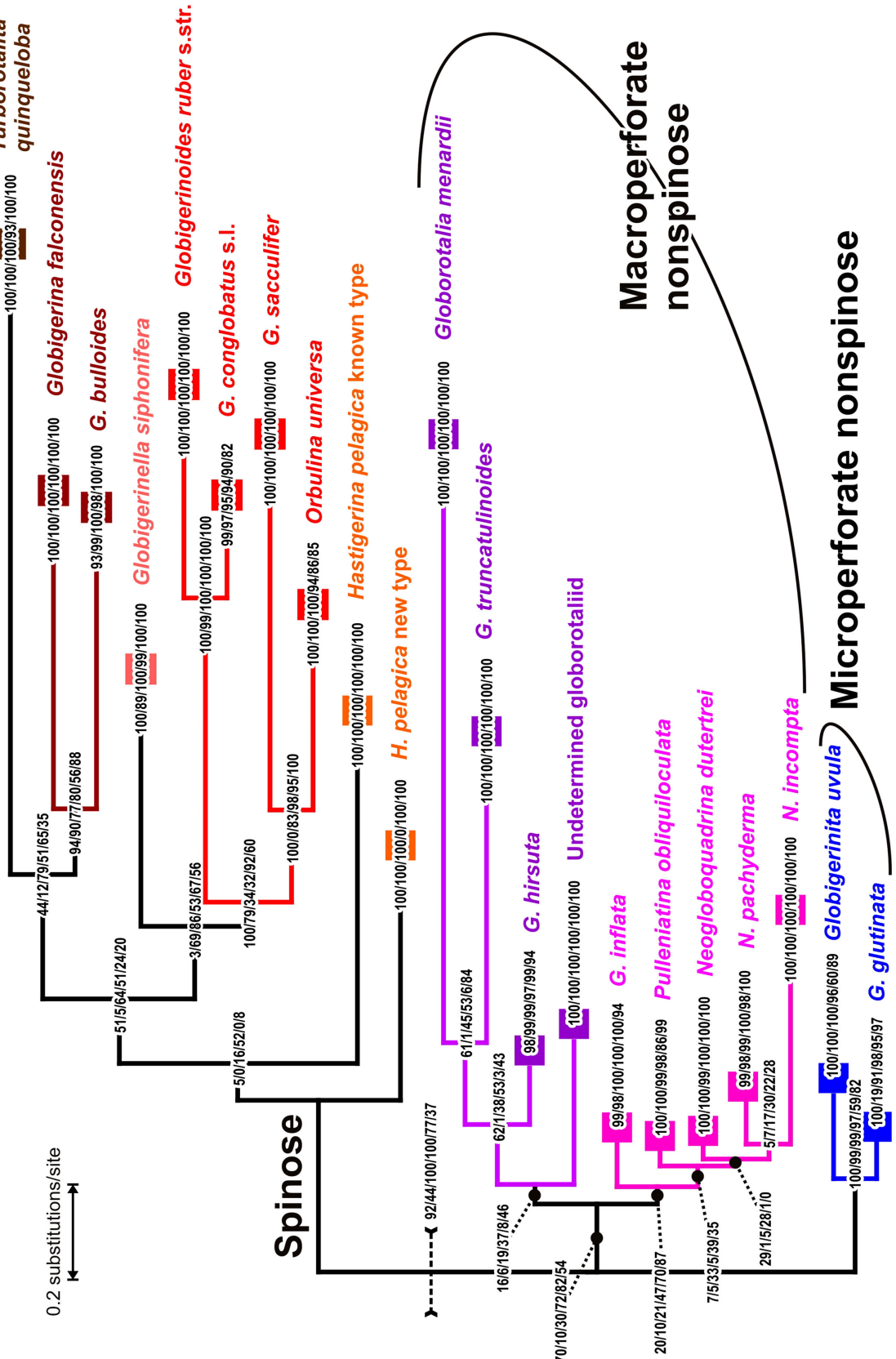

Figure 3. Partly collapsed ML tree inferred from the MAFFT alignment. The best ML tree inferred from the MAFFT alignment is shown. Branches are scaled in terms of the expected numbers of substitutions per site. Subtrees that include only sequences from the same morphospecies are collapsed at their root node and represented by black rectangles. Support, i.e. bootstrap percentages from the CLUSTALW/KALIGN/MAFFT/NRALIGN/POA/POAGLO-based analyses, of the collapsed subtrees and their relationships is indicated on the terminal nodes and on the branches. Not collapsed and accordingly annotated versions of all best known trees are found in the Additional file 2.

alignment, can negatively affect ML and MP inferences, but has little effect elsewhere.

SSU rDNA sequence diversity in planktonic foraminifera, and misidentified or unidentified specimens and sequences

As stated in the introduction the identification of plankton material is challenging and often leads to ambiguities in species determination. This is reflected in several mislabelled sequences found in online databases but also in our collections. The comprehensive evaluation of all database sequences in the course of our study reveals that one GenBank sequence has been mislabelled (Z69600; in GenBank stored as *Globigerinoides sacculifer*, but obtained from a *G. conglobatus* individual[6]) and that the single sequence of *Globorotalia crassaformis* stored in GenBank (AY453134) is 100% identical to sequences of *G. inflata* considering the amplified fragment (newly assembled and public database data). The single *Globigerinella calida* accession (Z83960) is identical to one SSU rDNA type of *G. siphonifera* (Additional files 2, 3). Considering the general level of SSU rDNA divergence within and among morphospecies detected elsewhere (this study,[6,7,9,51]) it is likely that these database sequences have been misidentified on collection, although currently no comparative data exist for *Globorotalia crassaformis* and *Globigerinella calida*.

In our new dataset, two clones of a newly sampled *Globigerina bulloides* specimen (R043) are showing sequence types characteristic for, and well documented in, *Globigerinella siphonifera*. These sequences were placed in all ML trees within the *G. siphonifera* clade. Together with Z69600, the R043 clones were responsible for the best T-scores being three steps larger than the minimum possible score (23 vs. 20). Accordingly, all alignments which resulted in a best ML tree achieving a T-score of 23 were regarded as in agreement with morphotaxonomy (the singletons AY453134 and Z83960 do not have an effect on the T-score of distinct topologies); the two exceptions were CLWOPT and MUSCLE. In trees inferred from the MUSCLE alignment, one (incompletely sequenced: 436 bp) *Globigerinita glutinata* clone (R04903) was placed within *Neogloboquadrina dutertrei*. Trees inferred with CLWOPT even misplaced four *Globigerinita glutinata* sequences (R04903, R04906, R049a1, and AF250105) within *Neogloboquadrina pachyderma*, apparently also an artefact caused by short sequences.

In addition to the identification of mislabelled sequences, ca. 20 sequences in our new dataset obtained from small specimens that could not be properly determined (R021, R034, P155, P125), and gene bank accessions labelled "*Globigerina* sp.", were unambiguously placed in all trees; they nested within existing clades that received high support (Additional file 2). These sequences thus could be identified by their position in the phylogenetic reconstructions and have been treated accordingly for the following discussion.

Monophyly of morphospecies

Figure 3 depicts a reduced ML tree inferred from the MAFFT-generated alignment, together with bootstrap support (BS_{ML}; bootstrap percentages based on 100 replicates) for individual nodes inferred from six selected alignments. For the sake of simplicity, subclades referring to distinct morphotaxa have been collapsed; full, annotated trees can be found in the Additional file 2. Tables 3 and 4 list in addition the bootstrap support of respective bipartitions under MP (BS_{MP}); further details can be found in Additional file 3.

Most terminal nodes received high support from the bootstrap analyses ($BS_{ML/MP} > 80$) independent of the alignment and inference method used; these are the nodes that define molecular clades corresponding to morphologically defined species (Fig. 3; Table 3). Exceptions were *Globigerinita uvula* ($BS_{ML/MP}$ = 60/29, POA; BS_{MP} = 59, POAGLO; $BS_{ML/MP} \geq 89$, others) and *Hastigerina pelagica*. The latter forms a low (under MP) to moderate or high (under ML) supported clade only in the POA-based and POAGLO-based analyses (Table 3). In two cases ML and MP bootstrap support differs strongly as inferred from the KALIGN alignment (*Globigerinita uvula*; *Globigerinella siphonifera*). This is likely due to short sequences which are not optimally aligned by this software (see above).

The GenBank sequence of *Globigerinita uvula* (AF387173) is markedly different from other SSU rDNA sequences of planktonic foraminifers in the

expansion segments (not shown, but see Additional file 2). Before this study, three sequences have been documented from its nearest relative, *G. glutinata*. We could amplify SSU rDNA fragments from two small individuals, which were identified upon collection as juveniles of either *Turborotalita quinqueloba* or *Globigerinita uvula*. We obtained and sequenced five clones from these two individuals documenting a new genotype comprising two similar sequence variants (details not shown). This genotype is placed as sister clade to the single *G. uvula* sequence from GenBank ($BS_{ML/MP}$ between 59 and 100; except based on the POA-alignment), and both are placed as a sister clade to *G. glutinata* (Fig. 3; Table 4). We therefore assume that the collected specimens comprise a new sequence type of *G. uvula*. However, it is clear that this group requires much more attention and data (see Table 1).

The most unexpected result of our survey of sequence diversity among the Azores Front planktonic foraminifera was the discovery of a new and highly divergent sequence type isolated from specimens of *Hastigerina pelagica*. Until now, this morphospecies has been represented by a single sequence in the public databases (Z83958;[6]). For this study we had access to SSU rDNA data from eleven specimens of *H. pelagica*, and a total of 38 sequences, mostly clones but also directly sequenced PCR products. Two of these specimens yielded a sequence type consistent with the template Z83958; the remaining nine specimens yielded the new type. The two types differ markedly in their nucleotide sequences (cf. length of the root and placement of both types in Fig. 3). In the ML trees inferred from four of the six alignments, the two sequence types of *H. pelagica* were placed in a grade-like fashion at the root of the spinose group with diminishing support (Fig. 3; refer to Additional file 3 for BS_{MP}). In trees from the POA and POAGLO alignments, *H. pelagica* formed a clade with high to moderate support under ML but not MP (see above; Table 3); and this relationship received little support otherwise (Table 3). None of the alternatives received a considerably higher support than any other based on all six alignments and both optimality criteria (Additional file 3). Thus, our analysis is inconclusive considering the position and relationships of both *H. pelagica* types.

The Hastigerinidae exhibit several morphologically unique features, including triradiate spines, monolamellar shell and a peculiar cytoplasmic "bubble capsule".[32] *Hastigerina pelagica* is one of the easiest identifiable extant species of planktonic foraminifera and a misidentification of the individuals yielding one of the two SSU rDNA genotypes can be largely ruled out. The only other two members of the family Hastigerinidae are *Hastigerinella digitata* and *Orcadia riedeli* (Table 1), which can be distinguished from the latter by chamber shape and spines distribution.[52] With regard to the unique morphology of *H. pelagic*a and considering the morphological variability among other spinose taxa,[32] it also appears unlikely that these characters have evolved in parallel and that they would be indicative of anything else than a common origin. On the other hand, the available SSU rDNA data do not support any scenario that would strongly contradict a common origin of *H. pelagica* (Additional file 3). One explanation why molecular data do not support a monophyly of *H. pelagica* (Table 3) might be a deep divergence followed by a rapid radiation.[53] This situation is analogous to that of *Neogloboquadrina incompta—N. pachyderma.* Both species differ only in their preferred coiling direction and have been traditionally placed in one species, *N. pachyderma.*[54] Like *H. pelagica* this pair is represented by divergent sequence types not supported as sister taxa in phylogenetic trees (Fig. 3; Table 4;[9,51] using limited taxon samplings).

This analysis, like previous work, largely supports the monophyly of SSU rDNA sequences from currently accepted and analysed morphospecies of planktonic foraminifera.[13,55] Save *H. pelagica* as outline above, there is one more exception to this rule, namely the biphyletic nature of sequences collected from specimens identified as *Globigerinoides ruber*. Two main SSU rDNA genotypes have been reported from the white variant of this species, one ("Type II")[7] being placed as a sister taxon to *G. conglobatus* (the clade here referred to as *G. conglobatus* s.l.); the other ("Type Ia", "Ib") forming a distinct clade with the pink-pigmented variant (here referred to as *G. ruber* s.str.; following the common notion that species should mirror monophyla).[7] All analyses have recovered this relationship: Both the *G. conglobatus* s.l. and the *G. ruber* s.str. clades obtained comparably high to very high support ($BS_{ML/MP} \geq 82$ and $BS_{ML/MP} = 100$, respectively;

Table 3. Support of morphotaxa under parsimony. ML bootstrap support (see also Fig. 3) is included for comparison. *Hastigerina pelagica* is, in addition to the known problematic case of *Globigerinoides ruber* (see text) the only morphotaxon that receives no sufficient support.

Alignment used	Nonparametric bootstrap support under ML					
	CLUSTALW	KALIGN	MAFFT	NRALIGN	POA	POAGLO
Microperforate species						
Globigerinita glutinata	100	19	91	98	95	97
G. uvula	100	100	100	96	60	89
Macroperforate nonspinose species						
Globorotalia hirsuta	98	99	99	97	99	94
G. inflata	99	98	100	100	100	94
G. menardii	100	100	100	100	100	100
G. truncatulinoides	100	100	100	100	100	100
Neogloboquadrina dutertrei	100	100	94	100	100	100
N. incompta	100	100	99	98	86	99
N. pachyderma	100	100	100	100	100	100
Pulleniatina obliquiloculata	99	98	99	100	98	100
Spinose species						
Globigerina bulloides	93	99	100	98	100	100
G. falconensis	100	100	100	100	100	100
Globigerinella siphonifera	100	89	100	99	100	100
Globigerinoides ruber s.str.	100	100	100	100	100	100
G. conglobatus s.l.	99	97	95	94	90	82
G. sacculifer	100	100	100	100	100	100
Hastigerina pelagica	4	0	38	9	88	68
Orbulina universa	100	100	100	94	86	85
Turborotalia quinqueloba	100	100	100	93	100	100

Fig. 3, Table 3). The sister group relationship of the two clades was highly supported ($BS_{ML/MP} \geq 99$) in trees from all six selected alignments (Fig. 3; Table 4).

Interclade relationships

Several relationships depicted in the MAFFT-inferred ML tree (Fig. 3) were consistently recovered by all methods. The mutual monophyly of each of the three major lineages of planktonic foraminifera recognized on the basis of their shell ultrastructure,[32] i.e. the microperforate nonspinose, the macroperforate nonspinose, and the spinose groups, was moderately to well supported under ML as the optimality criterion (Fig. 3; Table 4). Support under MP of such 'deep' relationships is, however, markedly decreased (Table 4; see also Additional file 3 for other 'deep' relationships; Additional file 3). An explanation may be that MP becomes statistically problematic, if the rate of change is high.[56]

As noted in the introduction, this is the first comprehensive (full) analysis of SSU rDNA data of planktonic foraminifera since the work of de Vargas et al.[3] That study used 521 "unambiguously aligned" sites among 15 morphospecies and the trees were rooted on several benthic foraminifera species (seven in total, including monothalamids and polythalamous taxa) as outgroups. The analyses identified

Nonparametric bootstrap support under MP										
CLWOPT	CLUSTALW	EINSI	GINSI	KALIGN	LINSI	MAFFT	MUSCLE	NRALIGN	POA	POAGLO
0	100	100	100	100	100	100	1	100	100	100
100	100	77	80	100	95	96	99	94	26	59
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	99	100	100	100	100	100	100
54	100	100	100	100	72	100	0	100	100	100
100	100	100	100	100	100	100	64	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	100	100	12	100	100	98	100	100	100
100	100	100	100	100	100	100	100	100	100	100
100	100	95	99	100	100	100	81	90	99	99
100	100	100	100	100	100	100	100	100	100	100
1	4	5	12	2	8	13	0	6	24	31
100	99	96	87	100	93	100	61	86	91	93
100	100	100	100	100	100	100	90	99	100	100

Moderate and low support values are highlighted.

the same three major planktonic groups, and as in our study, with varying support from nonparametric bootstrapping under different optimality criteria (low to high, a single sequence included representing the microperforate group; Table 4). There have been several later attempts that also included data from all three major lineages (Table 4). They partly found moderate to high support (Table 4) using only the conserved ("unambiguously alignable") sites of the 3' SSU rDNA, however, at the cost that not all SSU rDNA data-covered taxa were included. In the light of the arbitrarily restricted taxon sampling of these studies, they can neither be straightforwardly compared with the results of de Vargas et al[3] nor with this study. From a qualitative point of view, our study agrees with all former analyses in their separation of the three major groups of planktonic foraminifera (but see[7]). Since our focus here was to evaluate the multiple analysis approach to infer a phylogenetic structure *within* planktonic foraminifera and not to place planktonic taxa in an all-foraminiferal phylogeny, we did not include any benthic group. Nevertheless, it could be interesting to see, where the planktonic lineages will be placed in analyses based on matrices, which include *all* available SSU rDNA data of foraminifera.

In addition to relationships recovered by de Vargas et al[3] (morphotaxa generally forming clades,

Table 4. Support for selected phylogenetic scenarios. Comparison of our multiple analysis results (Fig. 3; Additional files 2, 3; BS under ML and MP) with eight previous manual-alignment based phylogenetic reconstructions in terms of the statistical support for relationships that appear to be consistently resolved in the fossil record of planktonic foraminifera. Values of support for each node are given where the respective study have identified the node as the dominant signal; "no" indicates analyses where an alternative topology has been preferred and "N/A" indicates analyses where some of the constituent species of the clade above the node have not been included.

	Microperforate clade	**Macroperforate clade**	***G. truncatulinoides—G. hirsuta* clade**	***Neogloboquadrina—Pulleniatina* clade**	***P. obliquiloculata—N. dutertrei* clade**
Darling et al[2]	N/A	No	N/A	N/A	N/A
De Vargas et al[3]	N/A	46/41/73	N/A	N/A	N/A
De Vargas and Pawlowski[5]	N/A	N/A	47	N/A	N/A
Darling et al[7]	N/A	No	N/A	N/A	N/A
Darling et al[9]	N/A	(76)[‡]	N/A	N/A	N/A
Stewart et al[11]	Unresolved	(69)[§]	N/A	N/A	N/A
Darling et al[54]	Unresolved	<70	N/A	N/A	78 (?)
Ujiié et al[69]	1.00/100	0.88/80	No	No	Unresolved
Multiple analysis					
BS_{ML}	100–59	82–30 (10[†])	78–2	39–5	91–0
BS_{MP}	100–52	20–0	34–0	7–0	99–0

recognition of a macroperforate and spinose clade; microperforate representative distinct from other planktonic foraminifera; a *G. conglobatus—G. ruber* clade; Table 4), some more interspecific relationships can be found, which are addressed in more detail in the following.

The microperforate nonspinose clade

Our analyses include data from two (or possibly three) morphospecies of *Globigerinita*. Their monophyly (distinctiveness) is well supported (Fig. 3; Table 4; POA-based moderate support). Up to now there has been no comprehensive study using the SSU rDNA data of *Globigerinita* (but see[3]).In one earlier analysis, data from both species was included.[11] The distance-based reconstruction used 505 sites from the generally conserved parts of the 3' SSU rDNA. As a result the planktonic lineages were placed along an unresolved polytomy with various benthic taxa. It has to be noted that only two nonspinose taxa were included (*Neogloboquadrina dutertrei* and *N. incompta*) and most of the inferred nodes were unsupported (Table 4).

The macroperforate nonspinose clade

The multiple analysis approach reveals no consistent phylogenetic structure within the macroperforate group, with support for individual nodes being generally low (Fig. 4; see also Additional file 3). *Globorotalia inflata* tends to group with the Neogloboquadrinidae unlike the other *Globorotalia* species (Fig. 4). This result is comparable to culled-alignment analyses of SSU rDNA,[3,10] the only two other studies that used data of all nonspinose taxa that were available at that time. Darling et al,[51] reporting on evolutionary relationships within the Neogloboquadrinidae (*Neogloboquadrina* spp., *Pulleniatina obliquiloculata*), used *Globorotalia inflata* as an outgroup, because it could be better "unambiguously aligned" with the former than the other globorotaliids (685 sites).[51] This is, however, not quantifiable based on the multiple analysis results. Any alternative of interspecific phylogenetic relationships within the nonspinose clade received diminishing support, both under ML and MP (but see Additional file 3 considering the putative sister pair *N. dutertrei—P. obliquiloculata*; Table 4).

N. pachyderma—*N. incompta* clade	Spinose clade	*G. bulloides*—*G. falconensis* clade	*G. ruber*—*G. conglobatus* clade	*O. universa*—*G. sacculifer* clade	*Globigerinoides*—*O. universa* clade
N/A	(No)*	N/A	99	82	87
N/A	No/58/51	N/A	91/100/100	No	No
N/A	(81)*	N/A	100	<50	No
N/A	(57)*	N/A	100	47	No
Unresolved	(86)*	N/A	99	<50	Unresolved
N/A	(88)*	No	98	<50	No
Unresolved	<70	N/A	100	<70	No
N/A	0.87/52	N/A	1.0/100	0.83/80	Unresolved
30–5	100–37	94–56	100–99	100–83 (0†)	100–32
14–0	61–22 (0†)	100–56	100	99–64 (0†)	66–12

*These studies did not include the phylogenetically challenging taxon *Hastigerina pelagica.*
†Based on the KALIGN-generated alignment (see text).
‡No *Globorotalia* species included.
§Only two close relatives included.

The spinose clade

Despite the higher divergence among the spinose lineages, several relationships were consistently recovered by most or all of the analyses (Figs. 3 and 5). A *Globigerinoides conglobatus-G. ruber* clade received the highest support ($BS_{ML/MP} \geq 99$; Fig. 5; Table 4), and has also been found in all former studies based on filtered SSU rDNA data.[3,7,9,11] The sister clade of *G. conglobatus-ruber* comprised *Orbulina universa* and *G. sacculifer* implying a common origin of these four morphospecies; this clade was represented in all six ML trees with BS_{ML} between 32 and 100 (Fig. 5; Table 4). As for the major clades (microperforate, nonspinose macroperforate, and spinose clade; Fig. 3), bootstrap support of this relatively 'deep' relationship is markedly lower under MP than under ML (Table 4). In five of the six analyses *Orbulina universa* appeared as sister group of *G. sacculifer* ($BS_{ML/MP} \geq 82$; Fig. 5; Table 4). Similar relationships have been reported although with low (<50) bootstrap support (Table 4) using filtered SSU rDNA data and distance-based reconstructions (neighbour-joining).[7,9,11] In the more comprehensive study of de Vargas et al,[3] *G. sacculifer* and *O. universa* formed a low to moderately supported clade with *Globigerina bulloides* under ML, distance and parsimony (Table 4).

Globigerina bulloides and *G. falconensis* were supported as sister taxa by bootstrap analysis ($BS_{ML/MP} \geq 53$; Fig. 3; Table 4). They were, however, placed as grade in the POA- and POAGLO-based ML trees (Fig. 5), with *G. bulloides* placed as sister taxon to *Turborotalita quinqueloba*. Such a topology received generally less support than the alternative of *Globigerina* clade (Fig. 3; POA-based ML tree provided in Additional file 2). This underscores the importance of establishing and investigating support (here: nonparametric bootstrapping) in course of multiple analysis (Figs. 3–5; Tables 3, 4), rather than to focus on clades found (or not) in the inferred phylogenetic trees (Figs. 4, 5). A one-alignment-one-tree approach may fail to recover an otherwise supported relationship unless the bipartition tables are investigated, because it is not represented in the inferred tree.

The placement of the extremely long-branched *T. quinqueloba* remains ambiguous. The support for

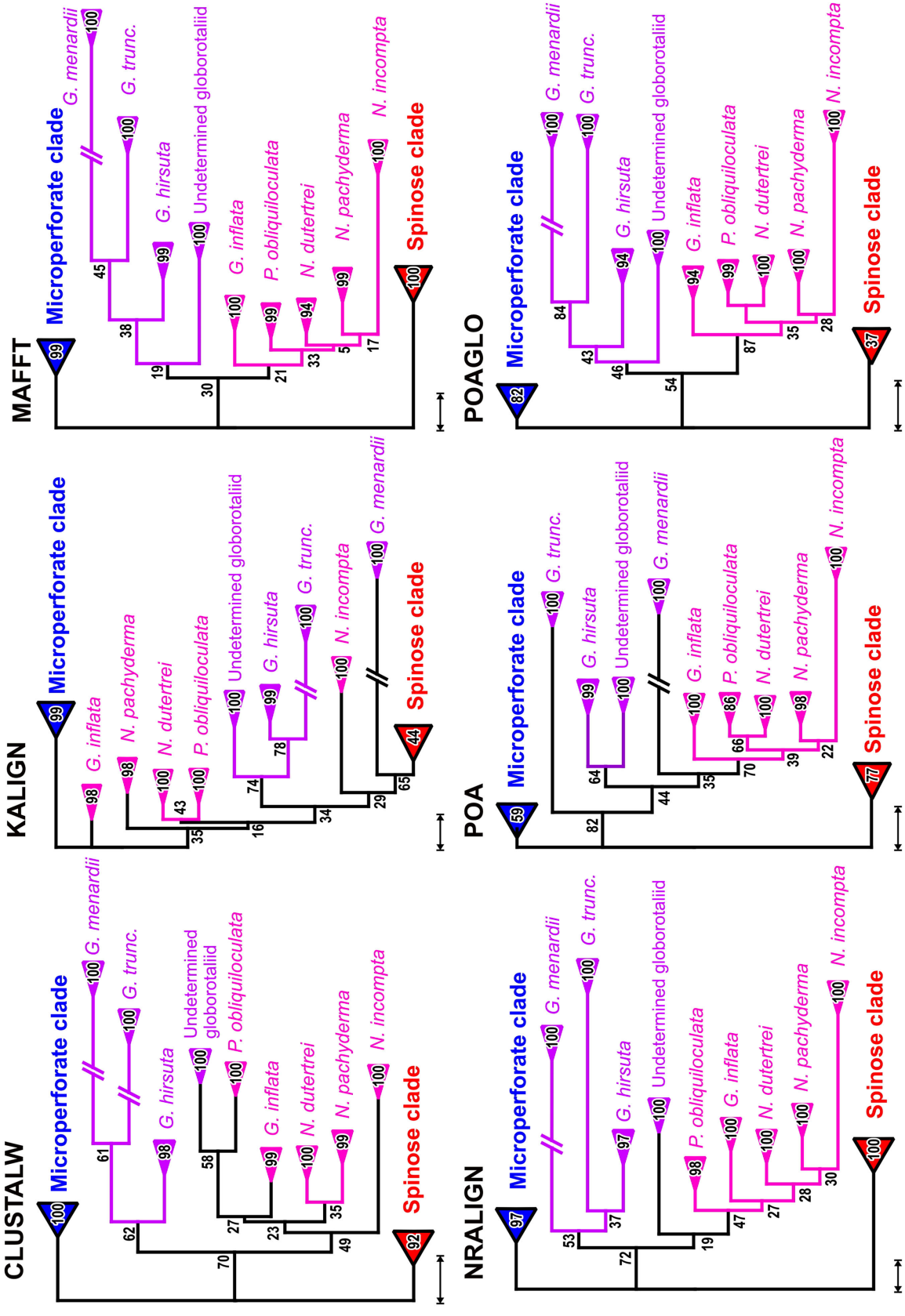

Figure 4. Alternative phylogenetic relationships within the nonspinose macroperforate clade as inferred from the six alignments. Shown are reduced ML phylograms based on the six selected alignments, with bootstrap support under ML annotated on the according branches (MP bootstrap support can be found in Additional files 2, 3). Scale bars are adjusted to 0.1 expected substitutions per site. Where indicated, branches have been broken down to one half of the original length. Subtrees comprising the same morphospecies were collapsed, as in Figure 3, as well as the microperforate (blue triangle) and spinose (red) clades. Not collapsed full ML trees can be found in the Additional file 2.

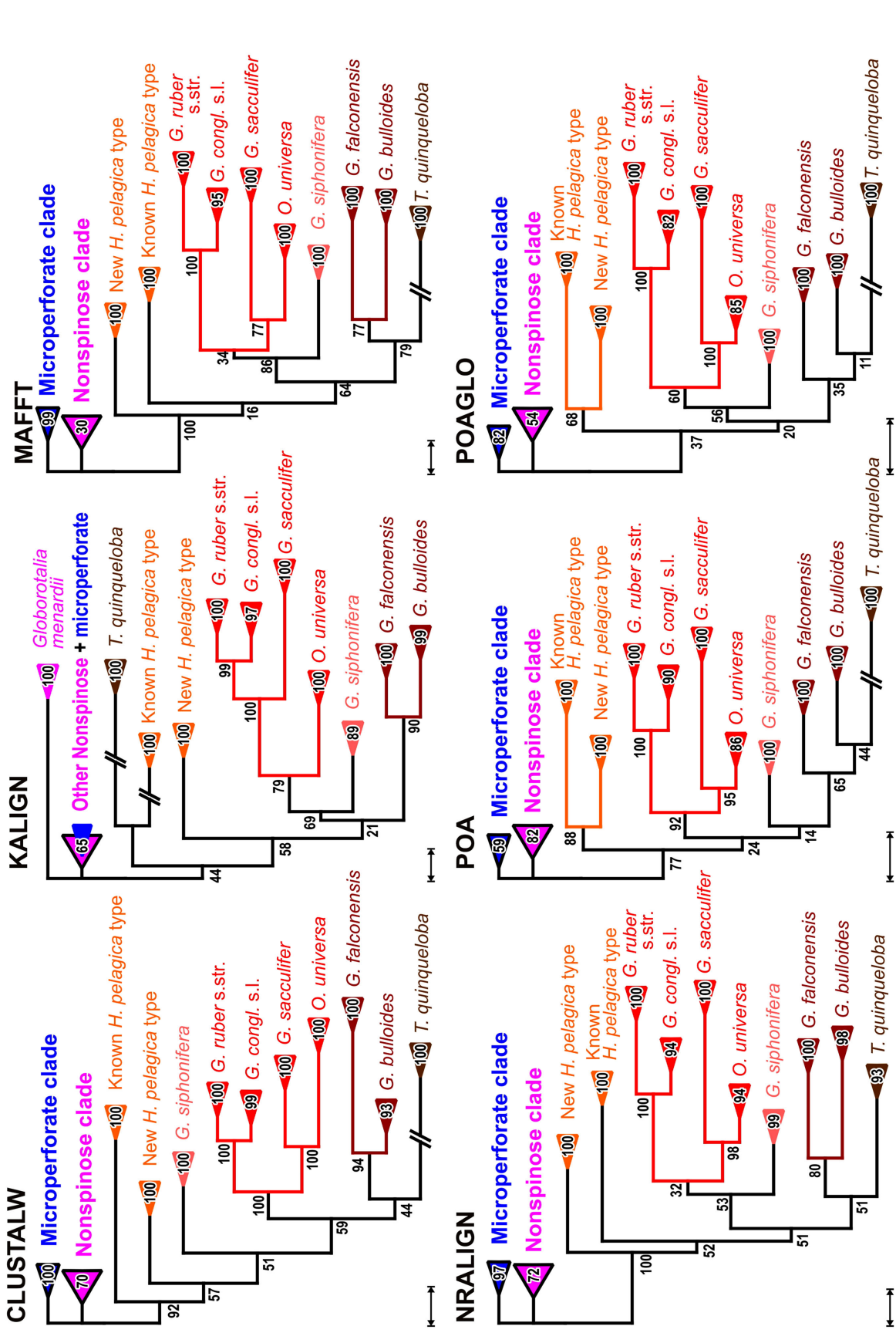

Figure 5. Alternative phylogenetic relationships within the spinose clade inferred from the six alignments. Shown are reduced ML phylograms based on the six selected alignments, with BS_{ML} annotated on the according branches. Scale bars are adjusted to 0.2 expected substitutions per site. Where indicated, branches have been broken down to one half of the original length. Subtrees comprising the same morphospecies as well as the microperforate and nonspinose macroperforate clades were collapsed, analogous to Figures 3 and 4. Not collapsed full ML trees can be found in the Additional file 2.

a common origin of *Globigerina* and *Turborotalita* ranges from very low (KALIGN) to moderate (MAFFT, NRALIGN, POA; Figs. 3, 5; refer to Additional file 3 for BS_{MP}). A sister relationship between *T. quinqueloba* and *G. bulloides* has been found in distance-based analyses,[9,11] which are prone to long-branch attraction more than ML.[56,57] As one alternative, *T. quinqueloba* was placed as sister clade to the known *Hastigerina pelagica* type (KALIGN), which is the longer branching of both *H. pelagica* types. *Hastigerina pelagica* has not been included in most traditional reconstructions that relied on filtered data, except in de Vargas et al.[3] At the time of de Vargas et al,[3] no SSU rDNA data of *T. quinqueloba* was available.

The last spinose taxon to be grouped within the spinose subtree is *Globigerinella siphonifera*. This taxon is placed by four of six alignment methods as a sister to the *Globigerinoides-Orbulina* clade, the according bipartition is moderately supported under ML by five of six alignments (BS_{ML} between 53 and 86; Fig. 5; Table 4). As in the case of the mutual monophyly of the three major groups, a common origin of *Globigerinella* and *Globigerinoides* + *Orbulina* finds support under ML as optimality criterion, but not if MP is used ($BS_{MP} \leq 26$). Alternatively, this clade is placed as sister to the *Globigerina-Turborotalita* clade (POA-based; very low BS under ML and MP); or sister of all spinose taxa except *Hastigerina* (CLUSTALW-based; $BS_{ML/MP} = 51/24$; $BS_{ML/MP} \leq 5$ other; Fig. 5). Based on filtered SSU rDNA data, the position of *G. siphonifera* within the spinose clade remained essentially unresolved ([3,7,9,11], but see[5]).

Comparison with the fossil record

The calcite shells of planktonic foraminifera accumulate in huge quantities on the sea floor, and in deep-sea basins they are a significant constituent of the sediment. The fossil record of planktonic foraminifera is one of the most complete and continuous of all organisms. Most significantly, the palaeontological taxonomy of this group is consistent with that of the living species, as both are based exclusively on the characters of the mineral shell. Because of the rich and continuous fossil record, phylogenetic relationships among fossil lineages of planktonic foraminifera are typically resolved by the method of stratophenetic tracing ([58], among others). Here, the morphology of individual species is traced back through time in short temporal steps until the time of its first appearance, and the ancestor is then determined by tracking of intermediate morphologies at higher temporal resolution. It is important to note that the reconstruction of the phylogeny of the modern species has rarely been the main aim of detailed palaeontological investigations and that many of the phylogenetic relationships remain obscure, but could potentially be linked to the fossil record when appropriate effort and methods were applied.

A synopsis of the multiple analysis results (superspecific clades) and our interpretation of the underlying data together with a schematic compilation of the fossil record of the analysed taxa are shown in Figure 6. Relationships of planktonic foraminifera, which appear well resolved in the fossil record, are included in Table 4, together with a summary of the support given by previous phylogenetic studies and multiple analysis under ML and MP. The characteristics of the wall structure of planktonic foraminiferal shells proved to be highly conserved through time (e.g. there have never been any microperforate foraminifera with spines and none of the spinose lineages is known to have lost spines) and support the existence of three main groups,[59,60] which also find support in SSU rDNA sequence analyses (de Vargas et al[3] and this study). The macroperforate spinose and nonspinose groups are considered to have shared a common ancestor in the Cretaceous—Paleocene genus *Hedbergella*.[7,59,61] The earliest spinose species is considered to have evolved from *Hedbergella monmouthensis*, one of the few survivors of the Cretaceous-Tertiary extinction.[62] However, the transition from the nonspinose to spinose state has never been observed, indicating that it must have been a rapid event associated with the filling of planktonic niches vacated after the mass extinction. Such an ancient and rapid divergence may not leave a conclusive signal in the genes of modern descendants,[53] as mentioned in the case of the two divergent types of *Hastigerina pelagica*. The (common) ancestry of the macroperforate nonspinose group is less well constrained, but the hypothesis presented in Pearson et al[60] links this group with another survivor species of the Cretaceous-Tertiary extinction, *Hedbergella holmdelensis*. The divergence between the two groups would thus be dated to the latest Cretaceous, 70–65 million years ago.

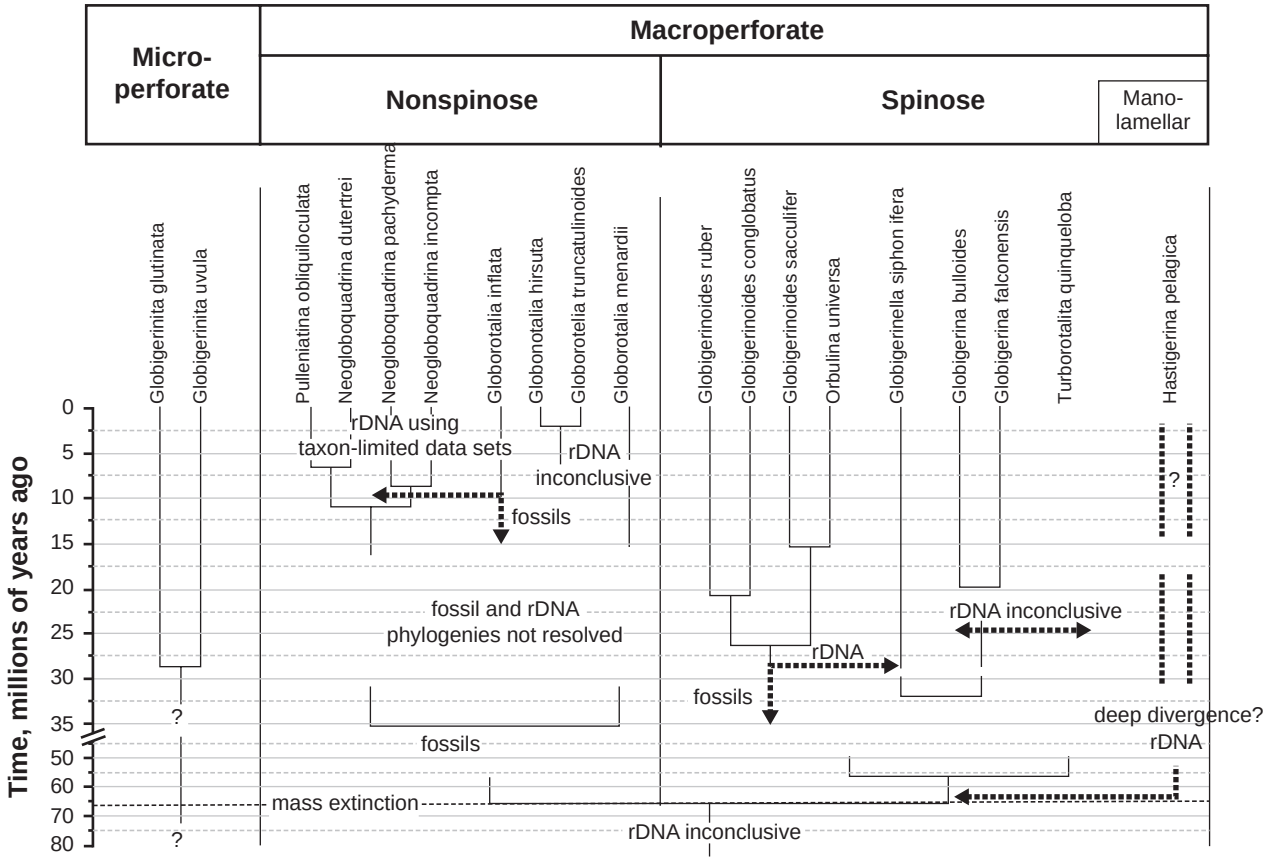

Figure 6. Comparison to the fossil record. A compilation of the fossil record of modern lineages.[59,60,64] Solid lines represent known fossil ranges of species or lineages leading to these species. Incongruence between the molecular-based hypothesis and the fossil record is highlighted; fossil evidence that is contradictory to molecular phylogenies but poorly resolved is also indicated.

The most likely ancestor of the modern microperforate planktonic foraminifera is the genus *Guembelitria*, a survivor of the Cretaceous-Tertiary extinction which possessed a microperforate wall texture,[52] although it must be noted that the link between the modern *Tenuitella* and *Globigerinita* forms and the Paleocene progeny of the *Guembelitria* lineage remains unresolved.[59,60] This fossil-based phylogenetic hypothesis implies that the modern microperforate foraminifera represent a monophyletic clade, which is distinct from both the spinose and nonspinose macroperforate lineages. The origin of the Guemblitriidae in the late Cretaceous remains unclear and it is entirely possible that the clade represents an independent colonisation of the planktonic niche by a different group of benthic foraminifera.

The extant nonspinose macroperforate lineages are the result of a radiation in the last 30 million years (review in[63]). The monophyly of the Neogloboquadrinidae is strongly supported in the fossil record,[64] the well documented lineage leading to *Globorotalia inflata* is clearly distinct from the Neogloboquadrinidae.[65] The common origin of these lineages in SSU rDNA trees (Figs. 3, 4) receives little support (Table 4), and the preferred ML topology could be erroneous. There is equally ample fossil evidence for sister relationships between *N. incompta—pachyderma* and *N. dutertrei—Pulleniatina obliquiloculata*.[64] These relationships are only weakly supported in our analyses as well as in all previous manual-alignment based analyses (Table 4; Additional file 3); they appear to be better resolved in taxonomically reduced datasets, in particular when the long-branching *N. incompta* is not included.[61] Such eclectic sampling obviously cannot solve the issue of the phylogeny of the foraminifera; it can only be used to discuss specific relationships within clades. Several alternative interpretations of the fossil record exist to explain the relationships within the modern genus *Globorotalia*,[64,66] but the genus is generally considered monophyletic with a common ancestor in the Oligocene

around 35–30 million years ago. As in the case of *Neogloboquadrina*, this cannot be supported based on SSU rDNA data to date (Fig. 4; Additional file 3).

The spinose condition in planktonic foraminifera evolved within the genus *Eoglobigerina* in less than 100,000 years after the Cretaceous-Tertiary extinction event some 65 million years ago.[59,67] An analysis of the fossil record following the initial radiation of the spinose taxa indicates that all subsequent lineages of spinose planktonic foraminifera with bilamellar shells (Table 1) can be linked to this one common ancestor.[59,60,64] The origin of the extant family Hastigerinidae possessing monolamellar shells (Table 1), and represented by *H. pelagica* herein (Figs. 3, 5), remains unknown. Earlier attempts to ally *Hastigerina* with *Globigerinella siphonifera* on the basis of similarities in spine architecture have been shown to be misleading.[67,68] In comparison to all other planktonic foraminifera, the monolamellar shells of both *Hastigerina* and *Hastigerinella* are extremely fragile and often partially resorbed during reproduction. As a result, they are only rarely preserved in marine sediments (a questionable report of *H. pelagica* is from the Miocene $<$10 million years ago)[64] and the fossil record therefore bears little further evidence on their phylogenetic position. However, several extinct, fragile monolamellar taxa are known from the early Cainozoic, but no *H. pelagica* or any other monolamellar spinose species have been observed in the sediment. Given the position of *H. pelagica* in SSU rDNA trees (Figs. 3, 5), one could even speculate that this species might represent the latest colonisation of the planktonic niche from a completely different group of benthic foraminifera.

Within the spinose species, the sister relationships *Globigerina bulloides—G. falconensis*, *Globigerinoides ruber—G. conglobatus* and *Globigerinoides sacculifer—Orbulina universa* (Figs. 3, 5; Table 4; Additional file 3) are in agreement with the fossil record and largely congruent with former SSU rDNA phylogenies (Table 4).[3,7,9,11,64] Furthermore, the *Globigerinoides-Orbulina* clade (Figs. 3, 5) is characterized by several potential morphological synapomorphies (supplementary apertures along the spiral suture, modifications of the last chamber) and the fossil record can be interpreted in favour of its monophyly.[64] The *Turborotalita* lineage can be traced to the Eocene, at least 45 million years ago,[60] and therefore it should have diverged closer to the root of the spinose subtree. Here, we found no unambiguous support for the placement of *T. quinqueloba* as sister group of *Globigerina falconensis* and/or *G. bulloides* and thus no evidence for an actual conflict between molecular and palaeomorphological data.[9,11]

The origin of the *Globigerinella siphonifera* lineage is not resolved in the fossil record. Based on its wall texture and the morphology of the first representatives of the lineage, it appears more closely related to *Globigerina* than *Globigerinoides*.[64] In analogy to *Hastigerina*, neither the fossil evidence nor the molecular (SSU rDNA) support is sufficient to unambiguously identify the sister clade to this species. In contrast to other 'deep' divergences, the according bipartition received only moderate support under ML (CLUSTALW-based none; Figs. 3, 5) and diminishing support under MP (details not shown, Additional file 3).

Conclusion

As depicted in Figure 3, SSU rDNA sequences extracted from morphologically defined species of planktonic foraminifera can be supported as clades (monophyla) by phylogenetic analysis of *complete* fragments of SSU rDNA despite the large divergence and length polymorphism in the expansion segments. Using a reproducible approach based on automated alignments without a priori filtering of nucleotides, we were able to infer several phylogenetic relationships, which obtain significant support from bootstrap analyses of all underlying data matrices (Figs. 3–5, Tables 3, 4; Additional files 2, 3). Thus, these relationships are supported *independently* of alignment ambiguity. The newly reported relationships are at least as congruent with the evidence from the fossil record as those inferred from time-consuming manual alignments after manual exclusion of not unambiguously alignable regions. This indicates that the need to establish nucleotide homology is not the most important obstacle when exploring the phylogenetic structure of the SSU rDNA in planktonic foraminifera. In our multiple analysis approach, important clades were recovered with much less effort than before, and in many cases, with higher support. Importantly, the lower alignment effort enabled us to include *all* available SSU rDNA sequences of planktonic foraminifers in the analyses; to the best of our knowledge, this was done for the first time in the present study.

Regarding the phylogenetic backbone of the planktonic foraminifera tree, many relationships remained ambiguous. The clarification of the relationships within the groups of nonspinose macroperforate planktonic foraminifera and between spinose subclades requires a reinvestigation of the fossil (sediment) record, a re-evaluation of the morphological traits uniting these clades, and additional molecular data covering all known planktonic species. Such combination of molecular, morphological and fossil data has the potential to provide an unprecedented level of understanding of the evolutionary unfolding within planktonic foraminifera.

It is apparent that future efforts in reconstructing the phylogeny of planktonic or other foraminifera with large divergences in SSU rDNA sequences should focus on exploring the effect of distinct alignments on the phylogenetic signal from the SSU rDNA without prior subjective filtering of the data. The same recommendation is likely to apply to other organisms and other alignment-ambiguous loci.[24,25] Use of up-to-date versions of several alignment programs under default values appears reasonable, while at least some potential artefacts as caused by, e.g. incompletely known sequences can be recognized by automated filtering using the comparison with previous information on probable taxonomic relationships.

Abbreviations

BS_{ML}, bootstrap support under ML; BS_{MP}, bootstrap support under MP; ML, maximum likelihood; MP, maximum parsimony.

Author's Contributions

RS, CH and MK collected the plankton samples, RA processed most of the samples in the molecular lab, and MG conducted the multiple analysis and made the necessary implementation of scripts and automated pipelines. This research has been initiated by VH, CH and GWG. RA, MG and GWG prepared the results for publication and drafted the manuscript. RA, CH, MK interpreted the results; and all authors participated in writing the final version of manuscript.

Acknowledgements

We thank Topaç Ertan and the master and crew of the research vessels POSEIDON (cruises P283/2, P308 and P321) and METEOR (cruise M69/1) for their help in taking the plankton samples. Technical assistance of Karin Stögerer and Margret Bayer is gratefully acknowledged. We are grateful to two anonymous referees for their valuable suggestions. This study was financially supported by the Deutsche Forschungsgemeinschaft (grants HE 697/36 and KU 2259/2).

Disclosures

The authors report no conflicts of interest.

References

1. Darling KF, Kroon D, Wade CM, Leigh Brown AJ. Molecular phylogeny of the planktic foraminifera. *J Foram Res.* 1996;26:324–30.
2. Darling KF, Wade CM, Kroon D, Leigh Brown AJ. Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Mar Micropaleontol.* 1997;30:251–66.
3. de Vargas C, Zaninetti L, Hilbrecht H, Pawlowski J. Phylogeny and rates of molecular evolution of planktonic Foraminifera: SSU rDNA sequences compared to the fossil record. *J Mol Evol.* 1997;45:285–94.
4. Huber BT, Bijma J, Darling K. Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Paleobiology.* 1997;23:33–62.
5. de Vargas C, Pawlowski J. Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Mol Phylogenet Evol.* 1998;9:463–9.
6. de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. *Proc Natl Acad Sci U S A.* 1999;96:2864–8.
7. Darling KF, Wade CM, Kroon D, Leigh Brown AJ, Bijma J. The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past ocean circulations. *Paleoceanography.* 1999;14:3–12.
8. Pawlowski J. Introduction to the molecular systematics of foraminifera. *Micropaleontol.* 2000;46:1–12.
9. Darling KF, Wade CM, Stewart IA, Kroon D, Dingle R, Leigh Brown AJ. Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature.* 2000;405:43–7.
10. de Vargas C, Renaud S, Hilbrecht H, Pawlowski J. Pleistocene adaptive radiation in *Globorotalia truncatulinoides:* genetic, morphologic and environmental evidence. *Paleobiology.* 2001;27:104–25.
11. Stewart IA, Darling KF, Kroon D, Wade CM, Troelstra SR. Genotypic variability in subarctic Atlantic planktic foraminifera. *Mar Micropaleontol* 2001;43:143–53.
12. Knowlton N. Sibling species in the sea. *Annu Rev Ecol Syst.* 1993;24: 189–216.
13. Darling KF, Wade CA. The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Mar Micropaleontol.* 2008;67:216–38.
14. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10.
15. Pawlowski J, Holzmann M. Molecular phylogeny of foraminifera—A review. *Eur J Protistol.* 2002;38:1–10.
16. Holzmann M, Pawlowski J. Taxonomic relationships in the genus *Ammonia* (Foraminifera) based on ribosomal DNA sequences. *Micropaleontol.* 2000;19:85–95.
17. Mindell DP, Dick CW, Baker RJ. Phylogenetic relationships among megabats, microbats, and primates. *Proc Natl Acad Sci U S A.* 1991;88: 10322–6.
18. Pawlowski J, Bolivar I, Fahrni JF, de Vargas C, Bowser SS. Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. *J Eukaryot Microbiol.* 1999;46:612–7.

19. Grimm GW, Stögerer K, Ertan KT, et al. Diversity of rDNA in *Chilostomella*: molecular differentiation patterns and putative hermit types. *Mar Micropaleontol.* 2007;62:75–90.
20. Pawlowski J, Bolivar I, Fahrni J, de Vargas C, Gouy M, Zaninetti L. Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol Biol Evol*. 1997;14:498–505.
21. Lee MSY. Unalignable sequences and molecular evolution. *Trends Ecol Evol.* 2001;16:681–5.
22. Gatesy J, DeSalle R, Wheeler W. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol Phylogenet Evol.* 1993;2: 152–7.
23. Farris JS, Källersjö M, Crowe TM, Lipscomb DL, Johansson U. Frigatebirds, Tropicbirds, and Ciconiida: Excesses of confidence probability. *Cladistics.* 1999;15:1–7.
24. Morrison DA, Ellis JT. Effects of nucleotide sequence alignment on phylogeny estimation: a case study of 18S rDNAs of apicomplexa. *Mol Biol Evol.* 1997;14:428–41.
25. Kemler M, Göker M, Oberwinkler F, Begerow D. Implications of molecular characters for the phylogeny of the Microbotryaceae (Basidiomycota: Urediniomycetes). *BMC Evol Biol.* 2006;6:35.
26. Wang L, Jiang T. On the complexity of multiple sequence alignment. *J Comput Biol*. 1994;1:337–48.
27. Elias I. Settling the intractability of multiple alignment. *J Comput Biol.* 2006;13:1323–39.
28. Janies DA, Wheeler WC. Theory and practice of parallel direct optimization. In *Molecular Systematics and Evolution: Theory and practice*. Birkhäuser Verlag, 2002:115–23.
29. Fleissner R, Metzler D, von Haeseler A. Simultaneous statistical multiple alignment and phylogeny reconstruction. *Syst Biol*. 2005;54: 548–61.
30. Ogden H, Rosenberg MS. Alignment and topological accuracy of the direct optimization approach via POY and traditional phylogenetics via ClustalW + PAUP*. *Syst Biol.* 2007;56:182–93.
31. Stamatakis A. RAxML-VI-HPC: Maximum-Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22:2688–90.
32. Hemleben C, Spindler M, Anderson O. *Modern Planktonic Foraminifera.* Springer-Verlag, Heidelberg, Tokyo, New York 1989.
33. Brummer GJA, Hemleben C, Spindler M. Planktonic foraminiferal ontogeny and new perspectives for micropalaeontology. *Nature*. 1986;3 19:50.
34. Holzmann M, Pawlowski J. Preservation of foraminifera for the DNA extraction and PCR amplification. *J Foram Res.* 1996;26:264–7.
35. Ertan KT, Hemleben V, Hemleben C. Molecular evolution of some selected benthic foraminifera as inferred from sequences of the small subunit ribosomal DNA. *Mar Micropaleontol.* 2004;53:367–88.
36. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22:4673–80.
37. Larkin MA, Blackshields G, Brown NP, et al. ClustalW and ClustalX version 2. *Bioinformatics.* 2007;23:2947–8.
38. Lassmann T, Sonnhammer EL. Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics.* 2005;6:298.
39. Katoh K, Kuma K, Toh H, Miyata T. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 2005; 33:511–8.
40. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32:1792–97.
41. Lu Y, Sze SH. Improving accuracy of multiple sequence alignment algorithms based on alignment of neighboring residues. *Nucleic Acids Res.* 2008: DOI:10.1093/nar/gkn945.
42. Lee C, Grasso C, Sharlow MF. Multiple sequence alignment using partial order graphs. *Bioinformatics.* 2002;18:452–64.
43. Wilm A, Mainz I, Steger G. An enhanced RNA alignment benchmark for sequence alignment programs. *Algorithms Mol Biol*. 2006;1:19.
44. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML web-servers. *Syst Biol.* 2008;57:758–71.
45. Stamatakis A. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In *Proceedings 20th IEEE International Parallel and Distributed Processing Symposium.* 2006:278.
46. Swofford DL. *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods)*, Sunderland, Sinauer Associates Inc. 2002.
47. Robinson DF, Foulds RL. Comparison of phylogenetic trees. *Math Biosci.* 1981;53:131–47.
48. Lassmann T, Sonnhammer ELL. Automatic assessment of alignment quality. *Nucleic Acids Res.* 2005;33:7120–8.
49. Farris JS. Formal definition of paraphyly and polyphyly. *Syst Zool.* 1974;23:548–54.
50. Fitch WM. Towards defining the course of evolution: minimal change for a specified tree topology. *Syst Zool.* 1971;20:406–16.
51. Darling K, Kučera M, Pudsey CJ, Wade CM. Molecular evidence links cryptic diversification in polar planktonic protists to Quaternary climate dynamics. *Proc Natl Acad Sci U S A.* 2004;101:7657–62.
52. Banner FT, Blow WH. The taxonomy, morphology and affinities of the genera included in the subfamily Hastigerininae. *J Micropaleontol.* 1960;6:19–31.
53. Whitfield JB, Lockhart PJ. Deciphering ancient rapid radiations. *Trends Ecol Evol.* 2007;22:258–65.
54. Darling KF, Kučera M, Kroon D, Wade CM. A resolution for the coiling direction paradox in *Neogloboquadrina pachyderma*. *Paleoceanography.* 2006;21:PA2011 10.1029/2005PA001189.
55. Kučera M, Darling KF. Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philos Trans Roy Soc Lond A.* 2002;360:695–718.
56. Felsenstein J. *Inferring Phylogenies.* Sunderland, Sinauer Associates Inc. 2004.
57. Sanderson MJ, Wojciechowski MF, Hu JM, Sher Khan T, Brady SG. Error, bias, and long-branch attraction in data of two chloroplast photosystem genes in seed plants. *Mol Biol Evol.* 2000;17:782–97.
58. Wei KY. Stratophenetic tracing of phylogeny using SIMCA pattern recognition technique; a case study of the late Neogene planktic foraminifera *Globoconella* clade. *Paleobiol*. 1994;20:52–65.
59. Olsson DR, Hemleben C, Berggren WH, Huber B. *Atlas of Paleocene Planktonic Foraminifera. Smiths Contrib. Paleobiol*. 1999;85:249.
60. Pearson PN, Olsson RK, Hemleben C, Huber B, Berggren WA. *Atlas of Eocene Planktonic Foraminifera.* Cushman Foundation Special Publication. 2006;41:514.
61. Liu C, Olsson RK. On the origin of Danian normal perforate planktonic foraminifera from *Hedbergella*. *J Foram Res*. 1994;24:61–74.
62. Liu C, Olsson RK. Evolutionary adaptive radiation of microperforate planktonic foraminifera following the K/T mass extinction event. *J Foram Res.* 1992;22:328–46.
63. Kucera M, Schönfeld J. The origin of modern oceanic foraminiferal faunas and Neogene climate change. In *Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies.* Edited by Williams M, Haywood AM, Gregory FJ, Schmidt DN. London: The Geological Society. 2007;2:409–26.
64. Kennett JP, Srinivasan MS. *Neogene Planktonic Foraminifera: a Phylogenetic Atlas.* Hutchinson Ross Stroudsburg, New York 1983.
65. Wei KY, Kennett JP. Phyletic gradualism and punctuated equilibrium in the late Neogene planktonic foraminiferal clade *Globoconella*. *Paleobiol.* 1988;14:345–63.
66. Cifelli R, Scott G. Stratigraphic record of the Neogene Globorotalid radiation (planktonic Foraminiferida). *Smithson Contrib Paleobiol.* 1986; 58:1–101.
67. Olsson RK, Hemleben C, Berggren WA, Liu C. Wall texture classification of planktonic foraminifera genera in the Lower Danian. *J Foram Res.* 1992;22:195–213.
68. Hemleben C, Bronniman P, Renz HH. Ultramicroscopic shell and spine structure of some spinose planktonic Foraminifera. In *Proceedings of the First International Conference on Planktonic Microfossils, Geneva 1967*. Brill EJ, Leiden 1969.

69. de Vargas C, Bonzon M, Rees NW, Pawlowski J, Zaninetti L. A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Mar Micropaleontol.* 2002;45: 101–16.
70. Darling KF, Kucera M, Wade CM, von Langen P, Pak D. Seasonal distribution of genetic types of planktonic foraminifer morphospecies in the Santa Barbara Channel and its paleoceanographic implications. *Paleoceanography*. 2003;1032:10.1029/2001PA000723.
71. Darling KF, Kucera M, Wade CM. Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proc Natl Acad Sci U S A.* 2007;104:5002–7.
72. Kuroyanagi A, Tsuchiya M, Kawahata H, Kitazato H. The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleo-environmental implications. *Mar Micropaleontol.* 2008;68:236–43.
73. Ujiié Y, Kimoto K, Pawlowski J. Molecular evidence for an independent origin of modern triserial planktonic foraminifera from benthic ancestors. *Mar Micropaleontol*. 2008;69:334–40.
74. Aurahs R, Grimm GW, Hemleben V, Hemleben C, Kucera M. Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*. *Mol Ecol.* 2009;18:1692–706.

Additional Files

Additional file 1—Sequence list

List of sequences newly obtained in the current study, including accession numbers and affiliations to morphospecies.

Additional file 2—Alignments and phylogenetic trees

Contains all alignments in FASTA and all phylogenetic trees in Newick format inferred in the course of the study. The trees are not collapsed and appropriately annotated. The pseudocharacter matrix for the comparison with the current taxonomy is also contained, as well as the substitution matrix used for running POA and POAGLO.

Additional file 3—Bootstrap support under ML and MP of selected bipartitions

Comparative list of bootstrap support under ML and MP based on the six selected alignments; along with a qualitative rating of the listed bipartitions, and according phylogenetic relationships.

Publication 2: Geographical distribution of cryptic genetic types in the planktonic foraminifer Globigerinoides ruber

Molecular Ecology (2009) **18**, 1692–1706 doi: 10.1111/j.1365-294X.2009.04136.x

Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*

RALF AURAHS,* GUIDO W. GRIMM,‡ VERA HEMLEBEN,† CHRISTOPH HEMLEBEN* and MICHAL KUCERA*

**Department of Micropalaeontology, Institute of Geosciences, Sigwartstrasse 10, †Department of General Genetics, Auf der Morgenstelle 28, Eberhard Karls Universität Tübingen, 72076 Tübingen, Germany, ‡Swedish Museum of Natural History, Department of Palaeobotany, Box 50007, 10405 Stockholm, Sweden*

Abstract

We present SSU rDNA data resolving the seasonal and geographical distribution of 'cryptic' genetic types of the planktonic foraminifer morphospecies *Globigerinoides ruber* in the eastern Atlantic Ocean and the Mediterranean Sea. Analysis of 262 sequences revealed the presence of five genetic types belonging to two distinct lineages. Although the morphospecies *G. ruber* occurs throughout the investigated region, its constituent 'cryptic' genetic types show a pattern of widespread exclusion, which is difficult to reconcile with the concept of ubiquitous dispersal. One of the newly discovered genetic types was exclusively found at stations in the Mediterranean Sea, possibly representing the smallest-scale example of endemism known in planktonic foraminifera. In general, our results suggest that the geographical scale of mutual exclusion between the genotypes is negatively correlated with their phylogenetic relatedness: the most similar and most recently diverged pair of siblings showed the strongest evidence for small-scale competitive exclusion. This pattern is consistent with the concept of niche partitioning, implying decreasing level of competition between genetic types with increasing degree of genetic divergence.

Keywords: competition, cryptic diversity, niche partitioning, phylogeography, planktonic foraminifera, SSU rDNA

Received 30 October 2008; revision received 16 January 2009; accepted 29 January 2009

Introduction

Assessments of species ecology and biodiversity in the plankton are often confronted with the phenomenon of cryptic sister species (Knowlton 1993; Irigoien *et al*. 2004; Chen & Hare 2008). Yet, studies on ecosystem functioning heavily rely on accurate estimates of their diversity (Weisse 2008). In many groups, the assessment of biodiversity is complicated by the lack of morphological characters. As a result, the number of described species and modelled diversity estimates may vary considerably among taxonomic group, a phenomenon well documented among protists (Adl *et al*. 2007).

Modern planktonic foraminifera are traditionally classified into about 50 species based on the morphology of their calcite shells (Hemleben *et al*. 1989). Considering the abundance and global distribution of this group, this comparably low number of species has been attributed to their high dispersal and limited potential for isolation (Norris 2000). Minor morphological modifications and biological observations indicated the presence of distinct types within the commonly accepted (morphological) species of planktonic foraminifera, but until genetic information became available, this 'intraspecific variation' was traditionally interpreted as ecophenotypic (e.g. Parker 1962; Hecht 1976).

The discovery of distinct genetic types within several planktonic foraminifera morphospecies (Darling *et al*. 1996, 1997, 1999, 2000, 2003; de Vargas *et al*. 1997, 1999, 2001; Huber *et al*. 1997; Darling & Wade 2008) implied that much of what was thought about speciation patterns and biogeography of foraminifera had to be revised. The diversity found in the SSU rDNA sequences obtained from specimens of traditional morphospecies throughout the world ocean was unexpected and, in most cases, individuals of different genetic types could not be distinguished by commonly

Correspondence: Ralf Aurahs, Fax: +49 7071 295727; E-mail: ralf.aurahs@uni-tuebingen.de

used morphological means. Therefore, the SSU rDNA genotypes were used to define 'cryptic species', referring to the fact that the traditional morphological species concept was not sufficient in capturing the true diversity in this group (e.g. Huber *et al*. 1997).

Whether these cryptic species are reproductively isolated is still difficult to assess. Direct cross-breeding experiments are (yet) not possible in planktonic foraminifera due to their complex reproductive cycle (Hemleben *et al*. 1989). However, there are three main lines of evidence suggesting that many, if not all, of the cryptic SSU rDNA genotypes represent distinct biological species. First, multiple examples of biogeographical distinctness and ecological adaptations of the genetic types have been observed (see reviews in Kucera & Darling 2002; de Vargas *et al*. 2004; Darling & Wade 2008). Second, despite the exclusively sexual reproduction in planktonic foraminifera (Hemleben *et al*. 1989) and extensive sequencing of the biparentally inherited multicopy SSU rRNA gene regions of hundreds of specimens (e.g. Darling *et al*. 2004, 2007), there is no compelling molecular genetic evidence for hybridization and/or incomplete lineage sorting in this gene so far. Third, the level of divergence associated with mutations in the SSU rRNA gene is in many groups of organisms indicative of long-lasting isolation and does not represent population-level variability (Rynearson & Armbrust 2004; Logares *et al*. 2007). Even among planktonic foraminifera, which show extremely high substitution rates in this gene, all molecular clock estimates suggest divergence times between the cryptic genetic types which are in the order of hundreds of thousands to millions of years (e.g. Darling *et al*. 1999, 2003, 2004; de Vargas *et al*. 2001).

The majority of planktonic foraminiferal morphospecies are cosmopolitan within their preferred temperature range. Considering the connection between this temperature-related latitudinal pattern and the vertical extension of the foraminiferal habitat in the water column, together with ecological aspects such as the presence of symbiotic algae, it would be reasonable to assume that morphospecies of planktonic foraminifera are characterized by a high degree of niche partitioning. The biogeography of the cryptic species (genetic types) within these morphospecies is, however, far more complex. Whereas there is evidence for global gene flow in some species (Darling *et al*. 1999, 2000), in many cases genetic types within the same morphospecies are geographically restricted (de Vargas *et al*. 1999, 2001, 2002; Darling *et al*. 2004, 2007). Since the genetic types largely represent siblings (sublineages) of a monophyletic, morphologically defined species, the distribution patterns of these types have been interpreted as if they followed the same environmental parameters as the morphospecies they belong to, although on a finer scale (Darling *et al*. 2004). The possible ecological and evolutionary consequences of the presence of such distinct lineages within morphologically defined species of planktonic foraminifera have not yet been fully considered, but first attempts show great promise in explaining the dynamics of species evolution in these organisms (Alizon *et al*. 2008).

Latest theories on community ecology increasingly operate with the concept of a distinctly higher degree of niche competition between closely related species (Leibold 2008). Supported by data from diverse ecosystems like tropical rainforests (Kelly *et al*. 2008), these models maintain that the more closely related two species are, the more their distribution is influenced by their competition and population dynamics. Given the cosmopolitan distribution of their morphospecies, this concept would not seem to apply to planktonic foraminifera. However, the degree of niche separation and genetic divergence has never been investigated among their cryptic genetic types.

Foraminifera have an excellent fossil record. Divergences in morphologically distinct species can be traced back in time, providing a temporal and spatial framework for the more recent, morphologically cryptic, divergences. In addition, the modern distribution of planktonic foraminiferal morphospecies has been extensively studied by palaeoceanographers, who use environmental calibrations of the examined species' abundances to reconstruct past ocean properties (e.g. Kucera *et al*. 2005; Kucera 2007). Here, we take advantage of this potential by investigating one of the most abundant and ecologically important planktonic foraminifer species, *Globigerinoides ruber*, in the Mediterranean Sea and the eastern Atlantic Ocean off North Africa. *G. ruber* (d'Orbigny 1839) is one of the major foraminiferal proxies for Neogene sea surface temperature reconstructions. It is a dominant species in planktonic foraminifera in the warm temperate Atlantic Ocean (e.g. Tolderlund & Bé 1971). The seasonality in its abundance is recorded by sediment traps (e.g. Zaric *et al*. 2005), and its present and past geographical distribution is mirrored in sediment samples (e.g. Schiebel *et al*. 2002). A prominent feature of *G. ruber* is the existence of two colour variations. The 'pink' form at present only lives in the Atlantic Ocean, with abundance maximum during the warmer season. The 'white' form shows high abundances during the colder months and in more oligotrophic oceanic waters (Bé 1959; Tolderlund & Bé 1971).

The region of the eastern Atlantic Ocean and the Mediterranean Sea offers an ideal setting to investigate distribution patterns in planktonic species due to the limited water exchange through the Strait of Gibraltar linked with a large ecological gradient across a short geographical distance, both factors being conducive for isolation. Earlier studies have indicated the presence of several genetic types within *G. ruber*, but their ecology and biogeography remained unclear (de Vargas *et al*. 1997; Darling *et al*. 1999; Kuroyanagi *et al*. 2008). By comprehensive sampling throughout the seasons across the gradient in the target region, we aim to determine whether or not there is a correlation between the

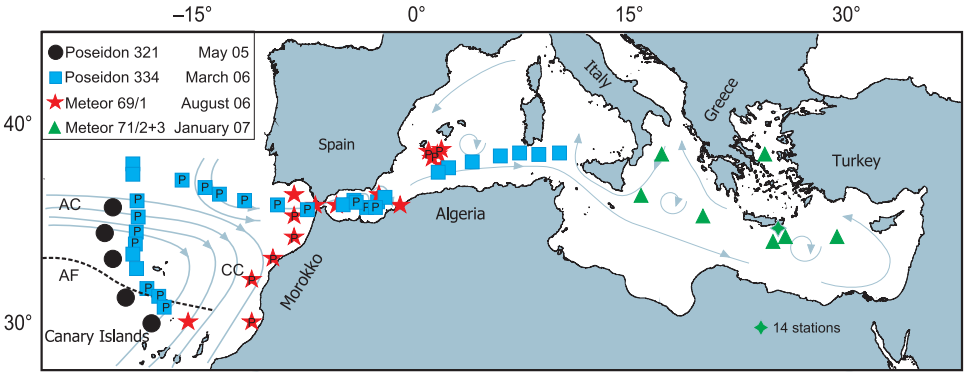

Fig. 1 Oceanographic features of the investigated region together with the location of the sampling stations. Grey lines indicate dominant surface currents: AC, Azores Current; CC, Canary Current; redrawn from Hernandez-Leon *et al*. (2007), in the Mediterranean Sea from Millot (1999) and Hamad *et al*. (2005). The position of the Azores Front (AF) is redrawn from Rogerson *et al*. (2004). Stations marked with 'P' represent surface water pumping stations; all other stations were sampled by multiple opening-closing nets.

degree of ecological or spatial rivalry and the extent of genetic divergence among the cryptic genetic types.

Materials and methods

Oceanography of the sampling area

The Azores Front (AF), located in the Eastern Atlantic Ocean between 30° and 40°N (Fig. 1), marks the boundary between the subtropical mode waters and the colder and fresher water masses of the transitional to subpolar zone (Schiebel *et al*. 2002). South of the Azores Islands, the AF coincides with the Azores Current (AC). The AC flows eastwards with a width of about 50 km and extends to water depth of at least 1000 m, accompanied by eddies and cyclonic recirculation to the north and anticyclonic recirculation to the south. Regional upwelling and downwelling is common. Towards the east Atlantic margin, the geographical position of the AC and AF are separated. The AC flows into the Gulf of Cadiz to replace water lost during water mass transformation in the Mediterranean Sea. The AF resides farther to the south between the Canary Islands and Madeira (Rogerson *et al*. 2004). The AC continues eastwards to the Strait of Gibraltar, then turns south and becomes the Canary Current (CC). The CC is a surface current (0–800 m), transporting NACW (North Atlantic Central Water) to the south. In spring, the CC is located close to the African coast while in summer it broadens, flowing through the Canary Islands (Hernandez-Leon *et al*. 2007). Plankton assemblages change substantially across the AF due to higher productivity and a deeper chlorophyll maximum north of the front (Schiebel *et al*. 2002; Rogerson *et al*. 2004).

The Mediterranean Sea is characterized by its own thermohaline circulation and an anti-estuarine exchange with the Atlantic Ocean through the Strait of Gibraltar. Here, the negative hydrological balance of the Mediterranean is compensated by the inflow of relatively cold, low saline and nutrient rich Atlantic surface waters. Relatively salty and warm deep water masses exit the Mediterranean Sea at depth. The westerly Trade Winds strengthen the Atlantic inflow during the summer months, whereas the inflow into the western Mediterranean during the winter is weaker. The Mediterranean Sea is separated into western and eastern basins; the water exchange between the two is limited by the Strait of Sicily. Towards the east, as the influence of the Atlantic inflow weakens, the surface waters in the Mediterranean Sea become increasingly salty, warm and oligotrophic. During the last glacial cycle, the warm water fauna in the eastern Mediterranean Basin was cut off from the Atlantic Ocean, as the western Mediterranean Basin cooled considerably (Hayes *et al*. 1999). The warm-water fauna isolated in the western Basin was then reconnected with the Atlantic province during the delectations. The last such event occurred at about 10 000 years ago, following a period of isolation lasting at least 20 000 years (Hayes *et al*. 1999).

Sampling

Specimens of the planktonic foraminifer *Globigerinoides ruber* were sampled in the northeastern Atlantic Ocean near the Canary Islands and in the western and eastern Mediterranean Basin (Fig. 1) onboard the RV Poseidon (cruises Pos 321 in May 2005 and Pos 334 in March 2006) and RV Meteor (cruises M 69/1 in August 2006 and M 71/2–3 in January 2007). Specimens were collected by vertical multinet tows (intervals between 0 and 700 m depth; 100 µm mesh) and pumping surface water (~6 m depth) through a 63-µm filter. Using a binocular stereomicroscope, specimens of both colour varieties of *G. ruber* were collected from all samples randomly, irrespective of their size. Only healthy, live

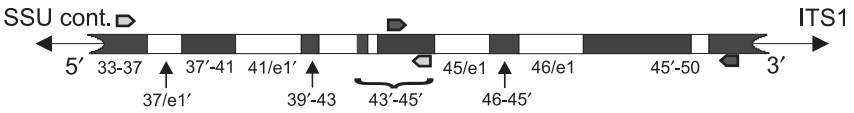

Fig. 2 Schematic illustration showing the 3′-end of the SSU rDNA. Marked are the approximate positions of the two primer pairs for the rear and front fragment used in this approach. Conservative regions of the SSU rDNA are coloured in black, the highly variable extension segments are drawn in white. Redrawn from Grimm *et al*. (2007).

Table 1 Primers used in this work

Primer	Sequence (5′–3′)	Product size (nt)	Ta (°C)	Amplified genotype
rubvF	TGACTCAACGCGGGAAATCTT	580	57	Type pink
rubvR	GAATCCCCGACGGCATACTGAC			
rubv3F	ACAAGCGCGTGGAGCAT	525	56	Type Ia, types IIa and IIb
rubv3R	AGCCCGGGACATCTGAGG			
rubh3F	GGTGGATAAACTCGGGGGACTGC	465	60	All types
rubh3R	GTAAGAGCGACGGGCGGTGTG			

Primer sets used for amplification of the 3′ SSU rDNA fragment: for the front ~550 nt of the fragment, two different sets had to be created (rubv and rubv3), while for the rear part a single set of primers (rubh3) was sufficient. Ta (°C), annealing temperature.

specimens containing cytoplasm were selected for genetic analyses. These specimens were freed carefully from debris with a brush, digitally photographed and transferred individually into 1.5-mL tubes for DNA isolation.

DNA extraction and sequencing

Isolation of DNA followed the DOC extraction method of Holzmann & Pawlowski (1996). We amplified the 3′-end of the gene coding for the small subunit ribosomal RNA (SSU rDNA) via polymerase chain reaction (PCR) using a proofreading Vent® polymerase (New England Biolabs). For genotyping, we used direct sequencing of the front and rear segments of the approx. 1000 nucleotides (nt) long amplicon of the SSU (Fig. 2). These two approximately 500 nt-long segments overlapped for *c*. 50 nt. Due to a unique substitution in the pink genotype, we used two sets of primers for the front segment (Table 1). For all 262 specimens included in this study, at least one of the two segments has been successfully sequenced. For each genetic type, multiple specimens were sequenced for both front and rear segment to confirm the sequence homogeneity within the type. PCR products were purified (DNA/Gel purification kit, QIAGEN Roche) and sequenced by a professional laboratory (AGOWA). The chromatogram of each sequence was screened by eye for the occurrence of ambiguous base calls that may indicate the occurrence of intra-individual variability in the SSU rDNA. All obtained sequences have been uploaded to the EMBL nucleotide database (http://www.ebi.ac.uk/embl/; Accession numbers FM865978–FM866240).

Phylogenetic analyses

The genetic identity (genotype) of each sequence was established by comparing a new sequence to an alignment database using BioEdit7.0.5.3 (Hall 1999). Sequences that were found to be 100% identical in the sequenced fragment are represented by a single operational taxonomic unit (OTU) in the subsequent phylogenetic analyses. We constructed 999–1017 nt-long summary (concatenated strict consensus) sequences for each SSU genotype. The genotypes were compared to available data from gene banks; if they differed from our summary sequences they were included as additional OTUs. The SSU genotypes were aligned using the ClustalW algorithm implemented in MegAlign® (DNASTAR) and manually re-aligned at a few positions. Calculation of genetic divergence and phylogenetic reconstructions relied on three matrices (Tables 2 and 3): (i) a matrix using all genotypes of *G. ruber* and its sister species *G. conglobatus* as OTUs (17 OTUs, 709 sites; 66% of nucleotides aligned); (ii) a matrix including only genotypes of *G. ruber* s.str. (*G. ruber* species aggregate; defined below; 8×921, 90% aligned; (iii) a matrix including only genotypes of *G. conglobatus* s.l. (*G. conglobatus* species aggregate, defined below; 9×987, 94% aligned). We calculated model-based pairwise distances based on each matrix. DT-ModSel (Minin *et al*. 2003) was used to find the best-fit substitution model under a Bayesian information criterion (BIC); free substitution parameters were estimated by PAUP* 4b10 (Swofford 2002) to compute ML-based distances. Phylogenetic networks were inferred with the neighbour-net algorithm (Bryant & Moulton 2002) implemented in

Table 2 Alignment properties and presettings for the phylogenetic reconstruction

	OTUs	Sites	Proportion of aligned sites (%)*	Proportion of included and defined characters (%)†	Template lengths (nt)	Model selected by BIC
All SSU genotypes	17	709	65.8	65.9–70.1	999‡–1046	K80+Γ
Only genotypes of *Globigerinoides ruber* s.str.	8	921	90.0	90.4–91.6	999‡–1012	TrNef
Only genotypes of *G. conglobatus* s.l.	9	987	93.5	93.6–96	1019–1046	K81uf

Amount of characters included in the distance calculations and phylogenetic reconstructions: *based on alignment length; †only characters (sites) have been counted that exhibit a defined nucleotide state; ‡estimated, shortest genotypes lack data for the central part of V9 (3 to 5 nt long in *G. conglobatus–ruber* complex).

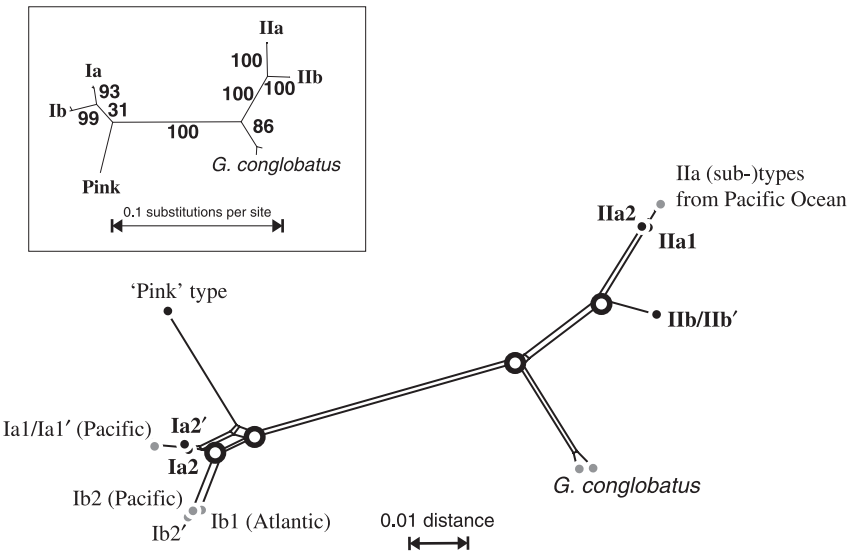

Fig. 3 Phylogenetic unfolding of the genotypes of *G. ruber* and *G. conglobatus* in our data set complemented with 10 sequences representing NCBI GenBank data. Shown is a neighbour-net (NN) splits graph, which is generally treelike, indicating a high degree of compatibility in the phylogenetic information throughout the analysed gene. The edge lengths illustrate the different levels of relatedness between the genetic types. The inset shows an ML phylogram based on the same data, and bootstrap support annotated along the branches. Both reconstructions are based on an alignment of 709 nt (matrix 1, Materials and Methods). The apostrophes at Ia and IIb mark a single polymorphic site found in individuals of the types by manually screening the chromatogram files.

SplitsTree 4.8 (Huson & Bryant 2006) based on model-dependent and uncorrected pairwise distances. Phylogenetic trees were inferred under maximum likelihood (ML) as optimality criterion with a DOS-based version of RAxML-VI-HPC 2.2.3 (Stamatakis 2006), recently renamed to RAxML 7.0. RAxML chose the best tree among 1000 inferences under a general-time-reversible (GTR) model allowing for site variation modelled by a gamma distribution (+Γ; alpha parameter estimated using 25 categories). Each inference used a parsimony tree as a start, which was optimized under the GTRMIX model using fabric settings. Node support was established with nonparametric bootstrapping (Felsenstein 1985), using 10 000 replicates computed with RAxML. Median-joining networks (Bandelt *et al*. 1999) of closely related types relied on the software Network 4.5 (Fluxus Technology Ltd) using fabric defaults. The simple mutational patterns differentiating between closest related types did not require a weighting of characters prior to analysis, re-running of the analysis with higher epsilon values than 0, or post-analysis purging of superfluous links. As input, we used the matrices 2 and 3, excluding the divergent taxa (type pink) and *G. conglobatus* genotypes, regions with missing data (5′- and 3′-end), and positions conserved (invariant) among all included taxa. Gaps were treated as fifth base.

Results

SSU rDNA genotypes and their phylogenetic relationships

The SSU rDNA sequences from a total of 262 specimens, morphologically identified as *Globigerinoides ruber* Pink or white, can be divided into five sequence types: type Pink, type Ia, types IIa1 and IIa2, and the here newly described type IIb (Fig. 3). The types pink and Ia are known from previous studies (de Vargas & Pawlowski 1998; Darling *et al*. 1999). The two types IIa1 and IIa2 differ by exactly two substitutions from the 'California type II' (here: IIa; Fig. 4) described by Darling *et al*. (1999) and are distinguished

Table 3 Mean genetic distances between all genetic types

	G. conglobatus U80790	California Type (IIa)	Type IIa1	Type IIa2	Type IIb/IIb′	Pink	Type Ia (Pacific) U80789	Type Ia/Ia′	Type Ib1 Z69599	Type Ib2 EU012461	Type Ib2 EU012460
G. conglobatus U80790	—	0.123	0.126	0.125	0.124						
California Type (IIa)	0.082	—	0.001	0.001	0.002						
Type IIa1	0.077	0	—	0.002	0.003						
Type IIa2	0.077	0	0	—	0.001						
Type IIb/IIb′	0.069	0.035	0.035	0.035	—						
Type Pink	0.150	0.191	0.167	0.167	0.153	—	0.097	0.077	0.096	0.095	0.096
Type Ia (Pacific) U80789	0.139	0.169	0.144	0.144	0.140	0.062	—	0	0.035	0.041	0.043
Type Ia/Ia′	0.108	0.151	0.151	0.151	0.146	0.032	0	—	0.029	0.036	0.038
Type Ib1 Z69599	0.137	0.163	0.141	0.141	0.142	0.057	0.030	0.018	—	0.015	0.017
Type Ib2 EU012461	0.142	0.166	0.145	0.145	0.145	0.051	0.029	0.018	0.001	—	0.002
Type Ib2 EU012460	0.145	0.169	0.148	0.148	0.148	0.052	0.031	0.020	0.003	0.001	—

Mean genetic distances between major SSU genotypes of *Globigerinoides ruber* s.str. and *G. conglobatus* s.l based on three different alignments. Distances above the diagonal are referred to within group alignments, whereas below the diagonal distances are drawn from a between group alignment of 709 from 1046 sites. For full table, see Supporting Information.

from one another by a single pair of compensatory substitutions (Fig. 5). None of the individuals in our survey carried a sequence 100% identical with the original 'California type'.

The new genotype IIb was found only at stations in the Mediterranean Sea (Figs 6 and 7). It has been labelled IIb due to its relative sequence similarity with the original type IIa from California and the subtypes IIa1 and IIa2 (Fig. 3). SSU sequences of neither *Globigerinoides conglobatus* nor *G. ruber* type Ib (and subtypes; see Supporting Information) were found among the sampled specimens. Specimens of types Ia and IIb frequently show a single site with intra-individual nucleotide polymorphism, reflected by doublet peaks in the chromatogram (Ia/Ia′; IIb/IIb′; Figs 3 and 4). For some specimens, only one of the two fragments could be sequenced. Given the degree of divergence confirmed by multiple sequences of the complete amplified fragment, single fragments were sufficient to distinguish most genetic types except of the type IIa1 from IIa2. In total, about a third of type IIa sequences could hence not be assigned to either IIa1 or IIa2.

Phylogenetic analyses including all sequence types ascribed to *G. ruber* and *G. conglobatus* reveal that the sequences of *G. ruber* type Ia, Pink and the subtypes of Ib share a direct common origin, whereas the *G. ruber* type II genotypes group together with the *G. conglobatus* sequences stored in gene banks (Fig. 3). The major phylogenetic splits received high bootstrap support (Fig. 3, inset). The degree of sequence similarity between and within *G. ruber* s.str. (excluding type II sequences) and *G. conglobatus* s.l. (including type II sequences) is strikingly different (Table 3). Hence, within each lineage, additional positions can be unambiguously aligned (up to 90–96% of all positions; Table 2) and used for analyses and distance calculation. Based on matrix 1, maximum intralineage genetic distances were 0.062 (*G. ruber* s.str.) and 0.085 (*G. conglobatus* s.l); the maximum interlineage distance between the two clusters was 0.191 (*G. ruber* pink to 'California type' AF102230). Based on matrices 2 and 3 including only sequence types of *G. ruber* s.str. and *G. conglobatus* s.l., respectively, both clusters show a higher degree of genetic divergence (0.128, *G. conglobatus* s.l.; 0.096, *G. ruber* s.str.). Median-joining networks for each sublineage are shown in Fig. 4.

Genotype distribution and abundance

Type pink. All specimens identified in the field as the pink phenotype of *G. ruber* exhibit the same sequence type. This type, pink (n = 148), is nearly identical to the two sequences of the pink phenotype from the Caribbean stored in NCBI GenBank. They differ from each other in a few randomly distributed positions, which we believe represent either sequencing errors or editing artefacts. (The NCBI GenBank sequences are comparably old.) We thus assume that all

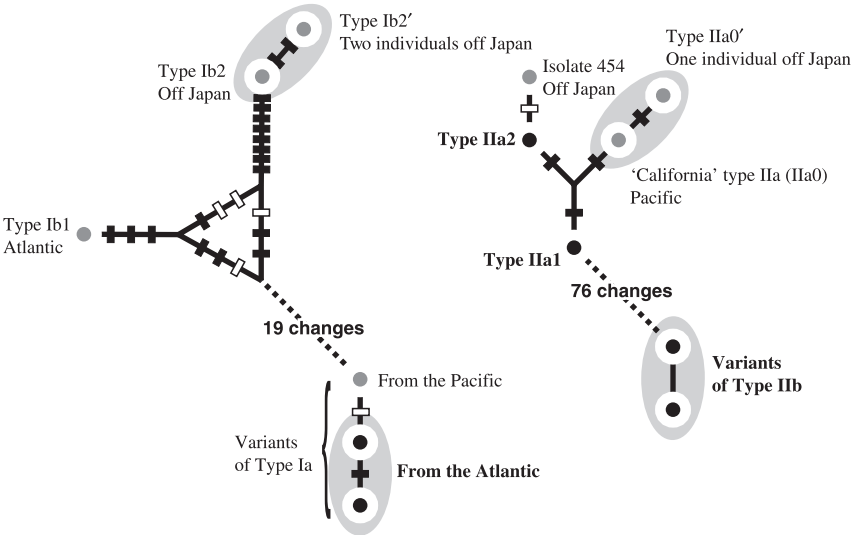

Fig. 4 Median-joining networks of closest related genotypes including new and downloaded data. Left, type I sequences (*Globigerinoides ruber* s.str), highly divergent type Pink (Fig. 3) not included. Right, type II sequences (*G. conglobatus* s.l., *G. conglobatus* accession not included). Black bars indicate inferred point mutations, white bars the deletion of a single nucleotide. Reported and newly found intra-individual variability (dimorphism) is indicated by light grey fields encircling two subtypes.

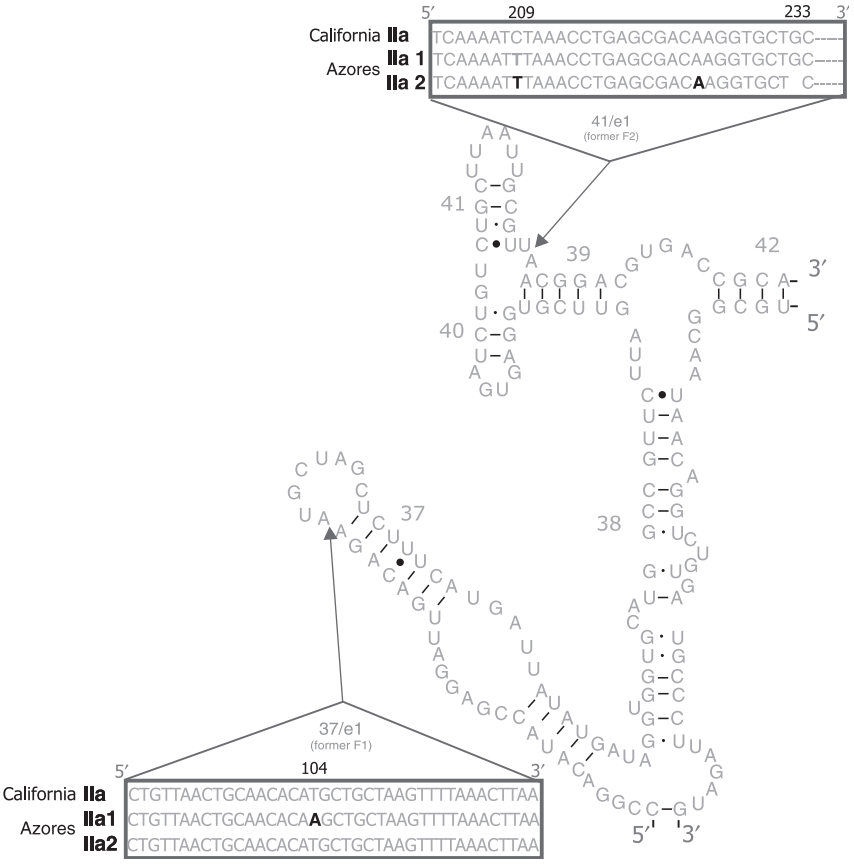

Fig. 5 Two-dimensional SSU rDNA structural reconstruction based on the 3′-end of the original type IIa sequence, the extension segments 37/e1 (complete) and 41/e1 (in part) are represented by their sequences. Type-characteristic substitutions are highlighted in bold.

specimens of *G. ruber* pink in the North Atlantic sequenced to date carry the same SSU genotype or highly similar genotypes. In addition, a considerable number of specimens which we assigned by light microscopic observations to the *G. ruber* white phenotype (n = 39 or 26%), carried the Pink SSU type, i.e. are genotypically *G. ruber* pink.

We found a strong correlation between seasonal changes and the pink genotype's abundance in the region of the Azores and Canary Islands. While the Pink genotype dominated our samples of *G. ruber* in August 2006, it went undetected in the same area during March 2006 (Figs 6 and 7). Instead, types IIa1 and IIa2 were the only genotypes

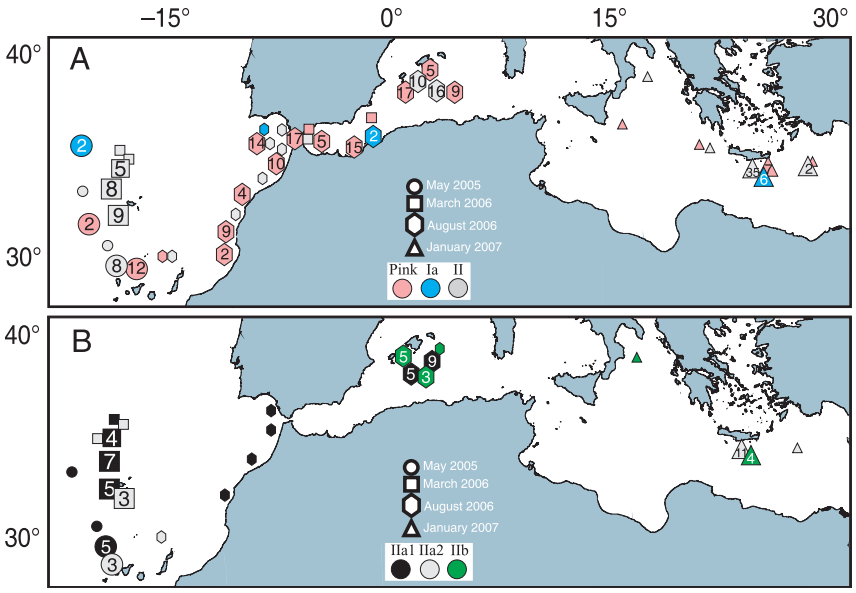

Fig. 6 Geographical sampling location of (A) the genetic types Pink, Ia and the cluster of type II from all four cruises. In (B), the distribution of the genotypes belonging to the type II cluster are shown in detail. Symbols without number mark single individuals.

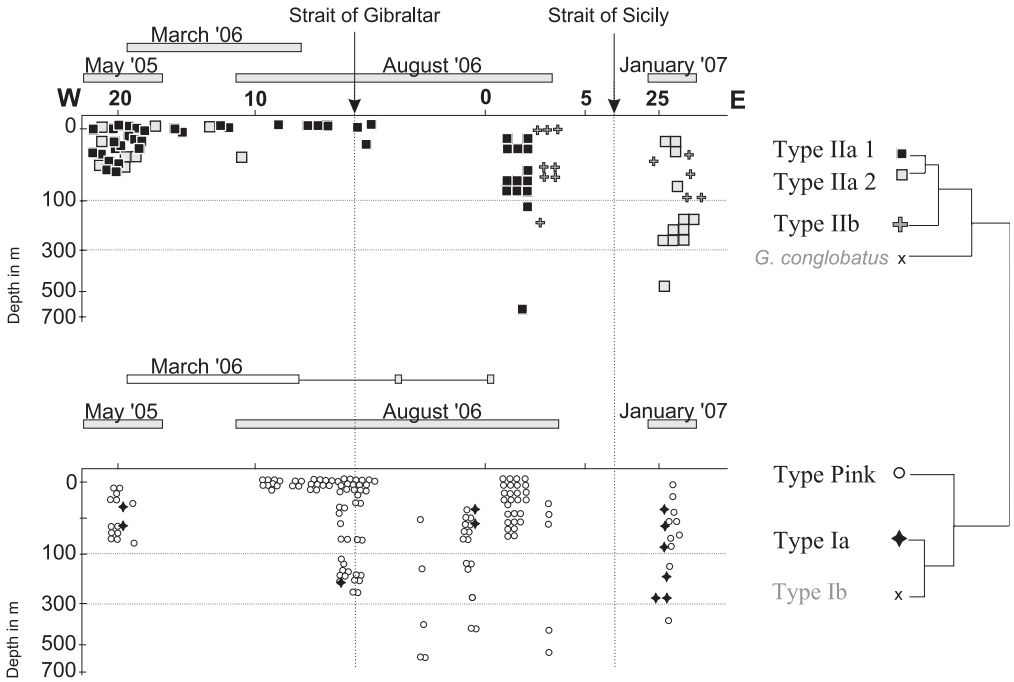

Fig. 7 Vertical distribution of SSU rDNA genotypes in *G. ruber* along a hypothetical West–East transect through our sampling area, ranging from the Canary Islands at about 20°W towards the Levantine Basin at about 25°E longitude. The bars above the diagram indicate the longitudinal range of the sampling from the respective cruises indicated in Fig. 1. The March 2006 cruise found pink genotypes only at two stations in the Mediterranean Sea. The relationships depicted in the phylogram on the right are schematic representations of the topology of the phylograms in Fig. 3 and do not accurately display the genetic distances, which can be found in Table 3.

obtained during that cruise. Few specimens of type pink were found in the Mediterranean Sea in March 2006. In May 2005, the pink genotype was found in the Canary Current in abundances comparable to types IIa1 and IIa2.

Type Ia. The *G. ruber* type Ia ($n = 11$) was the least abundant genotype in our survey (Figs 6 and 7). In May 2005, two individuals were found in the AC and in August 2006 three specimens were caught in the Strait of Gibraltar and close to the Algerian coast. Most individuals ($n = 6$) of type Ia

were found in January 2007 in the eastern Mediterranean basin, equally abundant as specimens of *G. ruber* pink. No individuals of type Ia were found in the Eastern Atlantic south of 35°N or in the central western Mediterranean Sea.

Types IIa1 and IIa2. All but one specimen that carried type IIa were identified by original collectors as '*G. ruber* white'. One individual with the type IIa sequence from the AC was labelled as '*G. conglobatus*'. The distribution pattern of type IIa (IIa1, IIa2, not subtyped IIa, in sum $n = 90$) shows a comparably weak seasonal and regional signal (Fig. 6). This type prevails in the region at comparably stable abundances during the winter and shows no pronounced abundance peak during the warm season. However, a strong geographical signal emerges in the subdivision of type IIa (Figs 6 and 7). The two subtypes, IIa1 ($n = 42$) and IIa2 ($n = 21$) co-occurred in the Canary Current in March 2006 and May 2005, whereas subtype IIa1 was the only representative in August 2006 in the western Mediterranean Sea, and only the subtype IIa2 was found during January 2007 in the eastern Mediterranean Basin. The number of un-subtyped IIa individuals was highest in the Mediterranean Sea, yet showed neither a seasonal nor regional correlation (see Supporting Information). Based on the data, we assume that these specimens either represent IIa1 or IIa2, and follow the same patterns as the subtyped specimens.

Type IIb. The newly discovered genotype IIb was found at stations in the Mediterranean Sea (Figs 6 and 7). Several type IIb specimens occurred at two stations close to the Balearic Islands ($n = 8$, in August 2006) and at three stations in the eastern Mediterranean Basin ($n = 5$, in January 2007). All specimens of type IIb were collected in the upper 100 m of the water column and the shell size of the sampled specimens of this type was comparably small (about 100 µm) relative to adult *G. ruber* specimens, independent of season and station. The original collectors considered these specimens either to belong to the species *Globoturborotalita tenella* or to be pre-adult individuals of *G. ruber* white. Extensive sampling in the Eastern Atlantic and the Strait of Gibraltar across different seasons failed to yield any sequences of this type, indicating that type IIb may be endemic to the Mediterranean Sea or at least absent from the adjacent Northeast Atlantic.

Discussion

SSU rDNA sequences are commonly applied for identifying genetic variability in various protist taxa (e.g. Lara *et al*. 2008; Medlin *et al*. 2008; for a review see Moreira & López-Garcia 2002). Due to the relatively conservative nature of the SSU rDNA, in comparison to other marker genes as ITS or microsatellites, sequence differences in the SSU rDNA between individual members of a taxon are generally considered as being an interspecific signal, rather than being indicative of population dynamics (Rynearson & Armbrust 2004; Logares *et al*. 2007). Relatively broadly sampled and well accessible, the SSU rDNA has become the primary choice to analyse distribution patterns of genetic diversity and assemblage composition in planktonic foraminifera (e.g. Darling & Wade 2008; but see de Vargas *et al*. 2001). Previous analyses of this marker revealed the existence of an extensive cryptic genetic diversity within morphological species; in all cases, this diversity appears to be linked to reproductive isolation, and the cryptic genetic types thus appear to represent biological species (Kucera & Darling 2002; de Vargas *et al*. 2004; Darling & Wade 2008). Like all earlier studies, we find no evidence for hybridization even among the most closely related of the *Globigerinoides ruber* cryptic genetic types and we can clearly distinguish genotype-indicative mutations from intra-individual variability. Therefore, we conclude that the cryptic genetic types of *G. ruber* are most likely associated with reproductive isolation.

Our phylogenetic reconstructions aimed at inferring the degree of relatedness between the genetic types of the *G. ruber-conglobatus* aggregate (Fig. 3). By focusing on the ingroup, we were able to maximize the alignable sites between the genetic types, approaching their real genetic distances as far as is possible with an alignment-based method. In accordance with the analyses by Darling *et al*. (1999) and Kuroyanagi *et al*. (2008), we find strong evidence that the current broad morphological species concept of *G. ruber* as introduced by Parker (1962) is incorrect and that the species is in fact diphyletic: type IIa (first subtype described from the Santa Barbara Channel, California, by Darling *et al*. 1999) and type IIb (newly found) SSU rDNA sequences represent a sister lineage of *G. conglobatus*, whereas sequences of the other types (Pink, Ia and Ib) form a distinct clade. The mutual monophyly of *G. conglobatus* and type IIa (*G. conglobatus* s.l.) vs. types Ia and Pink (*G. ruber* s.str.) is clearly indicated both in the ML phylogram and NN splits graph (Fig. 3), and it is mirrored by the degree of sequence alignability among the types (Table 2). The genetic divergence is higher in the *G. conglobatus* s.l. cluster (Table 3), but the branching of the individual genotypes in the ML phylogram (Fig. 3) appears very alike in both species clusters. If this is an indication for a similar mode of speciation in both clusters or an artefact of the methods we applied and the used molecular marker is open to debate.

Considering our data from the eastern Atlantic and the two sequences generated by de Vargas *et al*. (1997) and Darling *et al*. (1996) in the Caribbean, it appears reasonable to assume that all phenotypically pink *G. ruber* in the warm water belt of the Atlantic Ocean contain the same genetic SSU type, or at least, form one biological species. The pink phenotype of *G. ruber* first appears in the sediment record in the Pleistocene (Thompson *et al*. 1979). Initially found

throughout the warm water belt of the world ocean, the pink phenotype disappeared from the Indo-Pacific ~125 000 years ago (Thompson *et al*. 1979). The nature and function of the red pigment are not fully understood. It seems to be a carotene, derived from the foraminifer's algal symbionts (Thompson *et al*. 1979), but the process behind the colouration of the foraminifers' shells is unknown. Today, specimens of *G. ruber* pink can only be found in the central Atlantic Ocean and its marginal seas. Its temperature preference of 22–26 °C (Hecht 1976) limits *G. ruber* pink to the tropical and subtropical Atlantic between 35°N and 25°S, being less abundant towards the subtropical gyre centres (Tolderlund & Bé 1971).

In comparison to the genetic divergence between its two closest relatives, the types Ia and Ib, the Pink genotype is remarkably distinct (Fig. 3; Table 3). This could be an indication for an accelerated substitution rate in the SSU rDNA, or maybe an effect of the reduced population size of the Pink genotype compared to that of types Ia and Ib. Also, the homogeneity of the *G. ruber* pink population in the Atlantic might be attributed to a recent genetic bottleneck, when the species became extinct from the rest of the world ocean, and a relict population repopulated the Atlantic Ocean. Apparently, the dependence of the pink form of *G. ruber* on temperatures above 19 °C prevents its reinvasion of the Indian and Pacific Ocean around the southern tips of Africa and South America, respectively, although the low-latitude water masses of these two oceans are warm enough to potentially host populations of *G. ruber* pink.

In contrast to the Atlantic provincialism of *G. ruber* pink, specimens of '*G. ruber* white' are found throughout the world ocean's warm water belt, within a temperature optimum of between 18 °C and 25 °C (Hecht 1976). Several morphological variants have been described in '*G. ruber* white' (Saito *et al*. 1981; Robbins & Healy Williams 1991; Löwemark *et al*. 2005), which appear to show different habitat preferences (Kuroyanagi & Kawahata 2004; Lin & Hsieh 2007) as well as differences in the stable isotopic composition (Wang 2000; Lin *et al*. 2004; Kawahata 2005; Löwemark *et al*. 2005) and Mg/Ca geochemistry (Steinke *et al*. 2005) of their shells. The first indication for a possible correlation between shell morphology and genetic differences within the white colour variety of *G. ruber* was recently presented by Kuroyanagi *et al*. (2008) from the Pacific Ocean, highlighting the need for a clarification of the taxonomic status of the morphological variants in this morphospecies.

Two very divergent genetic types were found in specimens identified as *G. ruber* white in our survey. Type IIa was the more abundant and most widely distributed genetic type, whereas type Ia, one genotype of the actual *G. ruber* 'white' s.str., had a rare and more disjunctive appearance in our samples. Type Ia was found so rarely at stations in the Atlantic Ocean that we speculate the ecological optimum of this genotype to be farther towards the central Atlantic. The stations where we found specimens of type Ia in the Atlantic and Strait of Gibraltar suggest a passive transport by eastward currents as far as into the eastern Mediterranean Basin. Planktonic protists are commonly transported by currents into regions of unfavourable conditions and this process, known as expatriation, is well documented for planktonic foraminifera (e.g. Weyl 1978). Sampling of drifted populations will certainly lead towards a misinterpretation of the ecological range of a given species, overestimating the actual home range of such species or genetic type (Weisse 2008). On the other hand, the equal abundances of type Ia and type Pink in the eastern Mediterranean might be an indication for a steady population of type Ia in this remote basin, with conditions more similar to the putative home range of this type in the more oligotrophic North Atlantic subtropical gyre. Considering the global distribution of type Ia (Darling *et al*. 1997; Darling & Wade 2008; Kuroyanagi *et al*. 2008), as well as the provincialism of the Pink genotype in the Atlantic Ocean, the co-occurrence of these two types in our sampling area might be explained in a similar way as the distribution of the type II cluster. The rare findings of type Ia in the Canary region and western Mediterranean Sea in the summer and spring months could be the result of an ecological exclusion by the large population of the Pink genotype, whereas in the eastern Mediterranean, the Pink genotype may not be able to outcompete its type Ia sister species throughout the year, allowing a population of type Ia to be maintained in this basin. The observation that *G. ruber* pink is generally less abundant in planktonic foraminifer assemblages towards the oligotrophic centres of the subtropical gyres (Tolderlund & Bé 1971) provides a possible explanation for its inability to consistently outcompete its type Ia sister in the extremely oligotrophic eastern Mediterranean basin.

About 84% of all the specimens identified as *G. ruber* white (type pink individuals not taken into account) sampled in our survey carried type IIa sequences. Although the two genetic subtypes IIa1 and IIa2 have a sequence similarity of more than 99.9%, and are co-occurring in the region of the Canary Current throughout the seasons, we found no individual showing both subtypes, and obtained no type IIa sequence with ambiguous base calls. This, and the exclusive occurrence of either one of the two genetic types in one of the Mediterranean basins indicates that the individuals bearing either one of the variants are members of recently diverged sister species. The vicariant pattern of both cryptic sister species could be (i) a seasonal signal, limiting genotype IIa1 to the warmer seasons and IIa2 to the winter months (Figs 6 and 7) or (ii) due to niche competition between the two.

When considering the implications of the observed disjunct distribution of these genetic types, the possible bias of insufficient sampling, the lead argument of the idea that protist taxa are generally cosmopolitan (e.g. Fenchel

Table 4 Probability estimates for the detection of genetic types

Region	Undetected type	*N*	*q* in % (max. at 95%)	*p* in % (max. at 95%)
North Atlantic	IIb	43	6.8	0.2
Western Mediterranean	IIa2	14	20.6	1.6
Eastern Mediterranean	IIa1	11	26.1	2.7
Probability of failing to detected the rare type in all areas simultaneously: 0.4%				

Probability estimates for the detection of the genetic types from the *G. conglobatus* cluster in our sampling. *N* is the number of sequenced individuals, *q* is the probability of not finding a rare genetic type among these individuals; *p* is the relative abundance of such rare type.

& Finlay 2006), must be considered. However, planktonic foraminifera reproduce exclusively sexually and show a population response within seasons, so that each genetic type should maintain a reasonably sized standing stock throughout its home range. Since we have consistently collected from all depths in the water column and specimens of all sizes and shapes, and since all genotypes of the type II clade were recovered by our sequencing at comparable number, we believe it is safe to assume that our sampling is not biased towards any of the genotypes. Under these assumptions, it is possible to estimate the probability that the observed distribution is biased by overlooking rare specimens of the presumably absent genetic type. If p is the abundance of a rare type out of all morphologically identified *G. ruber* white specimens, then the probability q of not collecting a single specimen of this type among N individuals is equal to the probability of collecting only specimens of the remaining types, which can be expressed as $q = 1 - (1 - p)^N$. q can be approximated from the observational data as $q < 1/N \pm 1.96\sqrt{[r(1-r)/N]}$ (95% confidence interval), where $r = 1/N$. Since we are interested in estimating the maximum possible abundance of the rare type, we will only apply the upper bound of the 95% confidence interval for the estimate of q; then, with 95% confidence, it is correct to assume that the hypothetical highest abundance of a rare 'missed' type at the time of the collection in individual regions must have been:

$$p < 1 - \sqrt[N]{\frac{1}{N}\left(N - 1 - 1.96\sqrt{1 - \frac{1}{N}}\right)}$$

Thus, with 95% level of confidence, we can conclude that the abundance of type IIb would have to be below 0.2% of the entire *G. ruber* white population west of the of Gibraltar Strait to avoid detection and the abundance of the mutually exclusive types IIa1 and IIa2 in the Mediterranean would have to be below 2% (Table 4). At the same level of confidence, the probability of having failed to detect the rare type in all three regions simultaneously is less than 0.4%.

These results support a possible endemism of type IIb in the Mediterranean Sea, the first endemism on such small geographical scale found in planktonic foraminifera to date. At the same time, the occurrence of type IIb in both parts of the Mediterranean (just as the occurrence of both *G. ruber* s.str. types) without the slightest differentiation in the analysed SSU sequence speaks against a hydrographical barrier that generally prevents gene flow and could have caused the mutually exclusive distribution of types IIa1 and IIa2 in the two basins of the Mediterranean. In fact, the single intra-individual nucleotide polymorphism found in most of our type IIb sequences (Fig. 4) obtained from *both* basins indicates the potential for unhindered gene flow (intermixing) between the two basins.

The discovery of the genetic type IIb in individuals placed within the *G. ruber* morphospecies, is a good example of the nature of 'cryptic' pelagic protist species. As we found specimens of this genotype only at the size of about 100 µm, independent of season and geographical location, we have to consider the possibility that type IIb is a morphological variation inside the *Globigerinoides conglobatus* cluster. All specimens we found of type IIb were smaller than 125 µm. Hence, they belong to a size fraction of the sediment record that is rarely analysed by micropalaeontologists as it is dominated by pre-adult and juvenile specimens and as such are lacking the diagnostic characters present in shells of adult specimens. Therefore, our specimens might represent pre-adult stages of a yet undetermined species belonging to the *G. conglobatus* s.l. cluster, which could be indistinguishable from other *G. ruber* morphotypes found in the Mediterranean Sea.

Overall, the distribution patterns of the four genetic types found in our survey appears to be remarkably unaffected by the physical barriers such as the Straits of Gibraltar and Sicily. The pink genotype thrives between the Azores Current and the Eastern Mediterranean Basin, showing only seasonal changes in its abundance. The distribution of type IIa sequences appears also unlimited by geographical barriers and even less than type Pink by seasonal changes. The subdivision into IIa1 and IIa2, however, reveals that a strong geographical signal in this group is associated with the lowest level of genetic divergence. The observation that types IIa1 and IIa2 co-occur in the Atlantic but exclude each other in the Mediterranean represents in

our opinion a biological signal rather than a physically or geographically controlled vicariance. This biological signal, however, cannot be explained by differences in ecological adaptations between the types. The ecological properties of the eastern and western Mediterranean would be sufficiently different to invoke niche divergence as an explanation, but the area of the Canary Current where the two types co-occur is ecologically completely different from the Mediterranean. Thus, any ecologically relevant factor (or a combination of factors) that might separate the two genetic types in the Mediterranean Sea would have to have the opposite effect in the Atlantic Ocean. We find a different explanation more suitable.

Considering that the two Mediterranean basins represent comparably stable environments with respect to their hydrological conditions relative to the area around the Canary Islands (see Materials and methods for description) and that the exclusion of the two types from either basin is not caused by a hydrogeographical barrier to gene flow, we speculate that a stable population of one sister type in one basin prevents the other type from establishing a population there. This would imply that the niche both types depend on is virtually identical. In this hypothesis, the niche of the two types can be ecologically much broader, including conditions of the Canary Current, where episodic upwelling, changing position of fronts and water masses in the proximity to the African land mass cause a comparatively heterogenic habitat, allowing neither type IIa1 nor type IIa2 to establish a dominant population and to crowd out its sister type.

The very opposite competition pattern can be seen in the distribution of the pink type and type IIa (1 and 2). These two types occur in complete sympatry throughout the region, showing neither geographical, hydrographical (depth habitat in the water column) nor seasonal exclusion (Figs 6 and 7). These two types have the highest genetic distance (Table 3) of any pair of types in our study, and we conclude that they have diverged to such degree that they are now adapted to entirely different niche requirements.

Although the substitution rate in the planktonic foraminiferal SSU rDNA is considered to be relatively high (de Vargas & Pawlowski 1998), especially in the expansion segments (Fig. 2), it is unlikely that the patterns we observe developed during the last 1000 or even 10 000 years. However, considering the glacial history of the eastern Mediterranean Basin, especially the complete separation of the eastern basin from the warm water body of the Atlantic Ocean during the last glacial maximum (Hayes *et al*. 1999) and the high similarity of both types in comparison to others, it would be reasonable that the genetic separation between types IIa1 and IIa2 took place during the glacial period. Type IIa2 could have developed as a smaller population isolated in the eastern Mediterranean during a glacial maximum. When the northern Atlantic and the western Mediterranean warmed again, the Atlantic subtype of IIa (IIa1) reinvaded through the Strait of Gibraltar into the western Mediterranean but was unable to interbreed with the now distinct type IIa2 population, which also prevented it from establishing a population in the eastern basin. In this scenario, it remains to be explained, how the eastern Mediterranean type IIa2 managed to escape into the Atlantic following the warm-water reconnection. An origin in the Mediterranean of type IIb is also likely, but this must have been linked to a much earlier isolation event than that of types IIa1 and IIa2. Independent of the exact timing and mechanism of the separation among the types, the pattern persists until today and is most likely connected with reproductive isolation.

Due to the high dispersal capacity of planktonic foraminifera, any exclusion is unlikely on a global scale. Local displacement of one type by the other can be reversed rather quickly if the conditions alter, and types can re-spawn from a more favourable hideout as might have been happened for the Pink genotype in the Atlantic Ocean. A more thorough sampling, especially in the region of the Strait of Sicily, will eventually show where the distribution border of types IIa1 and IIa2 lies in the Mediterranean Sea and if type IIb is in fact restricted to the Mediterranean Sea.

Conclusion

Since most large sampling surveys for genetic types within one morphospecies of planktonic foraminifera targeted large geographical scales (de Vargas *et al*. 2002; Darling *et al*. 2004), population dynamic effects such as seasonal succession have remained rarely addressed (Darling *et al*. 2003). By sampling a relatively narrow geographical area over several years, we were able to gain a data set for *G. ruber* that depicts changes in the population structure over both time and space.

Our findings suggest that the described seasonal succession between the pink and white phenotypes is an interspecies signal, as most of the phenotypically white '*G. ruber*' refer to a lineage (type II) sharing a direct common origin with *G. conglobatus* rather than *G. ruber* pink. In addition, in contrast to the ubiquitous occurrence of the morphospecies *G. ruber* throughout the investigated region, a complex distributional pattern is evident among the individual 'cryptic' genetic types.

The patterns of co-occurrence and exclusion of the different *Globigerinoides* genotypes appear to mirror the degree of relatedness between them. Closely related genotypes exclude each other when the habitat is stable enough to establish a superior population size. The discovery of multiple new genetic types in *G. ruber* considerably improves the resolution of distribution patterns as for the seasonality reported in *G. ruber* in the Atlantic Ocean (e.g. Zaric *et al*. 2005).

Clearly, a broader geographical sampling is required to determine in which regions the morphologically identified *G. ruber* white corresponds to *G. ruber sensu stricto* (Ia, Ib and pink), and which regions are dominated by representatives of the genetically very distinct type II lineage.

We conclude that the distribution and abundance of planktonic foraminifera and possibly other marine pelagic protists is not determined by hydrographical factors alone, but that it is modulated by the competition between sibling species with similar ecological demands. Until now, the factor of competition between planktonic foraminiferal species has been largely overlooked, assuming that on an interspecies level, these protists do not have the incentive or the means to compete for resources in the oceanic realm.

Acknowledgements

We thank the master and crew of the research vessels Poseidon (cruises 321 and 334) and Meteor (cruise M69/1 and M71/2–3) for their help in taking the plankton samples. Technical assistance of Karin Stögerer and Margret Bayer is gratefully acknowledged. Critique of two anonymous reviewers and the editor helped to enhance the manuscript. This study was financially supported by the Deutsche Forschungsgemeinschaft (grants HE 697/36 and KU 2259/2).

References

Adl SM, Leander BS, Simpson AGB (2007) Diversity, nomenclature, and taxonomy of protists. *Systematic Biology*, **56**, 684–689.

Alizon S, Kucera M, Jansen VAA (2008) Competition between cryptic species explains variation in rates of lineage evolution. *Proceedings of the National Academy of Sciences, USA*, **105**, 12382–12386.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Bé AWH (1959) Ecology of recent planktonic foraminifera. I. Areal distribution in the western North Atlantic. *Micropaleontology*, **5**, 77–100.

Bryant D, Moulton V (2002) NeighborNet: an agglomerative method for the construction of planar phylogenetic networks. In: *Algorithms in Bioinformatics, Second International Workshop, WABI. Rome, Italy* (eds Guigó R, Gusfield D), Springer-Verlag, Heidelberg, Germany, **2452**, 375–391.

Chen G, Hare MP (2008) Cryptic ecological diversification of a planktonic estuarine copepod, *Acartia tonsa*. *Molecular Ecology*, **17**, 1451–1468.

Darling KF, Kroon D, Wade CM, Brown AJL (1996) Molecular phylogeny of the planktic foraminifera. *Journal of Foraminiferal Research*, **26**, 324–330.

Darling KF, Kucera M, Pudsey CJ, Wade CM (2004) Molecular evidence links cryptic diversification in polar planktonic protists to quaternary climate dynamics. *Proceedings of the National Academy of Sciences, USA*, **101**, 7657–7662.

Darling KF, Kucera M, von Langen P, Pak D (2003) Seasonal distribution of genetic types of planktonic foraminifer morphospecies in the Santa Barbara Channel and its paleoceanographic implications. *Paleoceanography*, **18**, 1032–1042.

Darling KF, Kucera M, Wade CM (2007) Global molecular phylogeography reveals persistent. Arctic circumpolar isolation in a marine planktonic protist. *Proceedings of the National Academy of Sciences, USA*, **104**, 5002–5007.

Darling KF, Wade CM (2008) The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. *Marine Micropaleontology*, **67**, 216–238.

Darling KF, Wade CM, Kroon D, Brown AJL (1997) Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. *Marine Micropaleontology*, **30**, 251–266.

Darling KF, Wade CM, Kroon D, Brown AJL, Bijma J (1999) The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracer of present and past ocean circulations. *Paleoceanography*, **14**, 3–12.

Darling KF, Wade CM, Stewart IA *et al.* (2000) Molecular evidence for genetic mixing of Arctic and Antarctic subpolar populations of planktonic foraminifers. *Nature*, **405**, 43–47.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Fenchel T, Finlay BJ (2006) The diversity of microbes: resurgence of the phenotype. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **361**, 1965–1973.

Grimm GW, Stögerer K, Ertan KT *et al.* (2007) Diversity of rDNA in Chilostomella: Molecular differentiation patterns and putativ hermit types. *Marine Micropaleontology*, **62**, 75–90.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hamad N, Millot C, Taupier-Letage I (2005) A new hypothesis about the surface circulation in the eastern basin of the Mediterranean sea. *Progress in Oceanography*, **66**, 287–298.

Hayes A, Rohling EJ, De Rijk S, Kroon D, Zachariasse WJ (1999) Mediterranean planktonic foraminiferal faunas during the last glacial cycle. *Marine Geology*, **153**, 239–252.

Hecht AD (1976) An ecologic model for test size variation in recent planktonic foraminifera: applications to the fossil record. *Journal of Foraminiferal Research*, **6**, 295–311.

Hemleben C, Spindler M, Anderson OR (1989) *Modern Planktonic Foraminifera*. Springer-Verlag, Heidelberg, Germany.

Hernandez-Leon S, Gomez M, Aristegui J (2007) Mesozooplankton in the Canary Current System: the coastal-ocean transition zone. *Progress in Oceanography*, **74**, 397–421.

Holzmann M, Pawlowski J (1996) Preservation of foraminifera for DNA extraction and PCR amplification. *Journal of Foraminiferal Research*, **26**, 264–267.

Huber BT, Bijma J, Darling KF (1997) Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Paleobiology*, **23**, 33–62.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Irigoien X, Huisman J, Harris RP (2004) Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature*, **429**, 863–867.

Kawahata H (2005) Stable isotopic composition of two morphotypes of *Globigerinoides ruber* (white) in the subtropical gyre in the north pacific. *Paleontological Research*, **9**, 27–35.

Kelly CK, Bowler MG, Pybus O, Harvey PH (2008) Phylogeny, niches, and relative abundance in natural communities. *Ecology*, **89**, 962–970.

Knowlton N (1993) Sibling species in the sea. *Annual Review of Ecological Systems*, **24**, 189–216.

Kucera M (2007) Planktonic foraminifera as tracers of past oceanic environments. In: *Developments in Marine Geology, Volume 1, Proxies in Late Cenozoic Paleoceanography* (Eds Hillaire-Marcel C, de Vernal A), pp. 213–262. Elsevier, Amsterdam.

Kucera M, Darling K (2002) Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. *Philosophical Transactions of the Royal Society A: Physical Sciences*, **360**, 695–718.

Kucera M, Weinelt M, Kiefer T *et al*. (2005) Reconstruction of sea-surface temperatures from assemblages of planktonic foraminifera: multi-technique approach based on geographically constrained calibration data sets and its application to glacial Atlantic and Pacific Oceans. *Quaternary Science Reviewss*, **24**, 951–998.

Kuroyanagi A, Kawahata H (2004) Vertical distribution of living planktonic foraminifera in the seas around Japan. *Marine Micropaleontology*, **53**, 173–196.

Kuroyanagi A, Tsuchiya M, Kawahata H, Kitazato H (2008) The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleo-environmental implications. *Marine Micropaleontology*, **68**, 236–243.

Lara E, Heger TJ, Ekelund F, Lamentowicz M, Mitchell EAD (2008) Ribosomal RNA genes challenge the monophyly of the hyalospheniidae (Amoebozoa: Arcellinida). *Protist*, **159**, 165–176.

Leibold MA (2008) Return of the niche. *Nature*, **454**, 39–41.

Lin HL, Hsieh HY (2007) Seasonal variations of modern planktonic foraminifera in the South China Sea. *Deep Sea Research Part II: Tropical Studies in Oceanography*, **54**, 1634–1644.

Lin HL, Wang WC, Hung GW (2004) Seasonal variation of planktonic foraminiferal isotopic composition from sediment traps in the South China Sea. *Marine Micropaleontology*, **53**, 447–460.

Logares R, Rengefors K, Kremp A *et al*. (2007) Phenotypically different microalgal morphospecies with identical ribosomal DNA: a case of rapid adaptive evolution? *Microbial Ecology*, **53**, 549–561.

Löwemark L, Hong WL, Yui TF, Hung GW (2005) A test of different factors influencing the isotopic signal of planktonic foraminifera in surface sediments from the northern South China Sea. *Marine Micropaleontology*, **55**, 49–62.

Medlin LK, Sato S, Mann DG, Kooistra WHCF (2008) Molecular evidence confirms sister relationship of Ardissonea, *Climacosphenia*, and *Toxarium* within the bipolar centric diatoms (Bacillariophyta, Mediophycea), and cladistic analyses confirm that extremely elongated shape has arisen twice in the diatoms. *Journal of Phycology*, **44**, 1340–1348.

Millot C (1999) Circulation in the Western Mediterranean Sea. *Journal of Marine Systems*, **20**, 423–442.

Minin V, Abdo Z, Joyce P, Sullivan J (2003) Performance-based selection of likelihood models for phylogeny estimation. *Systematic Biology*, **52**, 674–683.

Moreira D, López-Garcia P (2002) The molecular ecology of microbial eukaryotes unveil a hidden world. *Trends in Microbiology*, **10**, 31–38.

Norris RD (2000) Pelagic species diversity, biogeography, and evolution. *Paleobiology*, **26**, 236–258.

d'Orbigny A (1839) Foraminifèrs. In: *Histoire Physique et Naturelle de L'ile de Cuba* (ed. de la Sagra Ramon), p. 82. Bertrand A, Paris.

Parker FL (1962) Planktonic foraminiferal species in Pacific sediments. *Micropaleontology*, **8**, 219–254.

Robbins LL, Healy-Williams N (1991) Toward a classification of planktonic-foraminifera based on biochemical, geochemical, and morphological criteria. *Journal of Foraminiferal Research*, **21**, 159–167.

Rogerson M, Rohling EJ, Weaver PPE, Murray JW (2004) The Azores Front since the Last Glacial Maximum. *Earth and Planetary Science Letters*, **222**, 779–789.

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

Saito T, Thompson PR, Berger D (1981) *Systematic Index of Recent and Pleistocene Planktonic Foraminifera*. University of Tokyo Press, Tokyo, Japan.

Schiebel R, Schmuker B, Alves M, Hemleben C (2002) Tracking the recent and Late Pleistocene Azores front by the distribution of planktic foraminifers. *Journal of Marine Systems*, **37**, 213–227.

Stamatakis A (2006) RAxML-VI-HPC: maximum-likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.

Steinke S, Chiu H, Yu P *et al*. (2005) Mg/Ca ratios of two *Globigerinoides ruber* (white) morphotypes: implications for reconstructing past tropical/subtropical surface water conditions. *Geochemistry Geophysics Geosystems*, **6**, Q11005.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Sinauer Associates, Sunderland, Massachusetts.

Thompson PR, Bé AWH, Duplessy JC, Shackleton NJ (1979) Disappearance of pink-pigmented *Globigerinoides ruber* at 120 000 yr BP in the Indian and Pacific Ocean. *Nature*, **280**, 554–558.

Tolderlund DS, Bé AWH (1971) Seasonal distribution of planktonic foraminifera in the western North Atlantic. *Micropaleontology*, **17**, 297–329.

de Vargas C, Bonzon M, Rees NW, Pawlowski J, Zaninetti L (2002) A molecular approach to biodiversity and biogeography in the planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). *Marine Micropaleontology*, **45**, 101–116.

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.

de Vargas C, Pawlowski J (1998) Molecular versus taxonomic rates of evolution in planktonic foraminifera. *Molecular Phylogenetics and Evolution*, **9**, 463–469.

de Vargas C, Renaud S, Hilbrecht H, Pawlowski J (2001) Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: genetic, morphologic, and environmental evidence. *Paleobiology*, **27**, 104–125.

de Vargas C, Sáez AG, Medlin LK, Thierstein HR (2004) Superspecies in the calcareous plankton. In: *Coccolithophores — from Molecular Processes to Global Impact* (Eds Thierstein HR, Young J), pp. 271–298. Springer-Verlag, Heidelberg, Germany.

de Vargas C, Zaninetti L, Hilbrecht H, Pawlowski J (1997) Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. *Journal of Molecular Evolution*, **45**, 285–294.

Wang LJ (2000) Isotopic signals in two morphotypes of *Globigerinoides ruber* (white) from the South China Sea: implications for monsoon climate change during the last glacial cycle. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **161**, 381–394.

Weisse T (2008) Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. *Biodiversity and Conservation*, **17**, 243–259.

Weyl PK (1978) Micropaleontology and ocean surface climate. *Science*, **202**, 475–481.

Zaric S, Donner B, Fischer G, Mulitza S, Wefer G (2005) Sensitivity of planktic foraminifera to sea surface temperature and export production as derived from sediment trap data. *Marine Micropaleontology*, **55**, 75–105.

RA research focusses on the extend, patterns, and cause of genetic diversity in planktonic foraminifera species. This paper is a part of his PhD thesis. GWG and VH are interested in molecular taxonomy and evolution, in particular considering the nuclear encoded ribosomal RNA. GWG's research concentrates on species concepts in extant and extinct Northern Hemispheric tree genera, and their evolutionary unfolding in space and time; VH is an expert on the evolution and function of nuclear genes. CH and MK are both micropalaeontologists. CH's research interests focus on the taxonomy, phylogeny and ecology of planktonic and benthic foraminifera. MK is interested in the ecology and evolution of marine microplankton and the appplication of their fossil record for reconstructions of past climate.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

File S1 This Excel file includes three sheets with supplementary (tabulated) information. Table S1 lists all individuals that have been collected and analysed (genotyped) in course of our survey. It includes information about the original collector, identity and locality. Table S2 list the gene bank data (accessions and references) that has been used for comparison. The columns 'Sequence type' and 'Sequence name' refer to the main text and input matrices (NEXUS Files given Supporting File S2). Table S3 lists the ML-based distances inferred from the three input matrices. Sequence names refer to according columns in Table S1 (sequence type) and S2 (sequence name).

File S2 This zipped file contains NEXUS-formatted data matrices, which were used for analyses (see Tables 2 and 3). <Allssu.nex> refers to matrix (1) in Materials and Methods and includes all sequence types. <Ruber_s_str.nex> and <Cong_s_l.nex> (matrices 2 and 3 in Materials and methods) include sequences representing either genotypes of *Globigerinoides ruber* s.str. or *G. conglobatus* s.l.

File S3 Full ML tree in Newick format including information about ML bootstrap support, inferred from matrix 1.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Publication 3: A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence

A revised taxonomic and phylogenetic concept for the planktonic foraminifer species *Globigerinoides ruber* based on molecular and morphometric evidence

Ralf Aurahs, Yvonne Treis, Michal Kucera
Institute of Geosciences, Eberhard-Karls Universität Tübingen; Sigwartstr. 10, 72074 Tübingen, Germany; *corresponding author: ralf.aurahs@uni-tuebingen.de*

Abstract

Recent SSU rDNA sequence data of *G. ruber* showed that out of five genetic Types so far recognized in individuals assigned to *G. ruber* (white), only half (namely the genetic types Ia, Ib and Ib2) is closely related to *G. ruber* (pink), represented by a single genetic type. The types IIa1, IIa2 and IIb clustered together with *G. conglobatus*, a morphologically (and ecologically) distinct clade of *G. ruber*. Here we present a combined molecular clock and morphological data that shed light on the taxonomy and phylogeny of this ecologically and paleoceanographically important species. Our molecular clock approach based on SSU rDNA sequence data of *G. ruber* and *G. conglobatus* suggests a rather recent origin of *G. ruber* in the late Miocene and a split between *G. ruber* (pink) and the "white" genotypes Ia, Ib and Ib2 around 6 Ma. These results justify the separate treatment of the two colour variants of *G. ruber* in paleoceanographic studies and indicate that a) all records of *G. ruber* prior to the *G. ruber* "pseudo-extinction" event at 8 Ma refer to an unrelated species (*G. subquadratus*) and b) paleoenvironmental analyses based on *G. ruber* after 6 Ma but prior to the first occurrence of the pink pigmentation in the sediment have been potentially aggregating specimens of two distinct lineages. Guided by first observations on differences in gross morphology between individuals of Type IIa and Type Pink from the Eastern Atlantic Ocean and the Mediterranean Sea, we conducted morphometric measurements on (i) pictures of specimens with known genetic identity, (ii) shells from sediment samples, (iii) identified specimens of *G. ruber* and *G. elongatus* from a museum collection and (iv) pictures of *G. ruber* sensu lato from recent literature. Our results suggest that specimens of Type IIa are morphologically identical to the concept of *G. ruber* sensu lato in recent literature, and that these morphotypes are consistent with the description of *G. elongatus*, a species synonymised with *G. ruber* and largely abandoned in the 1960s. Given the degree of divergence among the individual types, the current morphological definition applied for *G. ruber* thus includes at least three distinct species. The name *G. ruber* (sensu d'Orbigny) should be reserved for specimens of the Pink genotype and morphotype, the name *G. elongatus* (sensu d'Orbigny) should be reinstated and used for the genetic type "IIa" and the morphotype *G. ruber* s.l.. Specimens of Types Ia, Ib and Ib2 require a new species name, but our data are not sufficient to provide a morphological character separating these specimens from their sister *G. ruber* pink, other than by their shell colouration.

Introduction

Globigerinoides ruber (d'Orbigny, 1839) is an abundant planktonic foraminiferal species often used in the reconstructions of sea surface conditions in the global oceans (e.g. Zaric et al., 2005; Sadekov et al., 2009). This cosmopolitan species is common in tropical-subtropical waters with temperatures above 19-20° C and is known to remain in the upper 100 m of the water column during its entire life cycle (e.g. Tolderlund and Bé, 1971; Hemleben et al., 1989). Within the morphospecies *G. ruber*, two variations in shell colour are recognized, *G. ruber* "white" with a pale, uncoloured shell as in most other planktonic foraminifera, and *G. ruber "*pink" with reddish (or 'pink') coloured chambers. While *G. ruber* (white) is distributed globally, the pink chromotype is today limited to the Central Atlantic Ocean and its adjunct seas (e.g. Thompson et al., 1979). As these two chromotypes also show differences in ecological requirements and seasonal abundance in the Atlantic Ocean (e.g. Tolderlund and Bé, 1971), most researchers handle white and pink individuals separately for the purpose of paleoceanographic reconstructions (e.g. Schmidt and Muliza 2002; Anand et al. 2003; Chiessi et al. 2007).

The French naturalist Alcide Desallines d'Orbigny first described the species in 1839 as *Globigerina rubra* from recent sediment samples from Cuba (d'Orbigny, 1839). As the name indicates, d'Orbigny, in the description of his species, highlighted the reddish colouration of its test. The original species definition was thus limited to the pink chromotype. In 1927, Cushman used the species as the type of the genus *Globigerinoides*, separating it from the genus *Globigerina*, mainly by the existence of at least one supplementary aperture in the final adult chamber (Cushman, 1927). Already before becoming the genotype of the new genus, the name *Globigerina rubra* has been used for specimens that were homeomorphic to the species description, but were lacking the red colouration (e.g. Cushman, 1914). As a reddish colouration can be found in some other planktonic and several benthic foraminiferal species, it was argued that shell colour is no valid taxonomical character to define a foraminiferal species and the species concept of *G. ruber* since then includes both coloured and colourless specimens with the same basic morphology (e.g. Banner and Blow, 1960).

In this broadened taxonomic concepts, adult individuals assigned to *G. ruber* showed a considerable range of phenotypical plasticity, in particular within the white chromotype. Parker (1962) reported a correlation between the abundance of certain phenotypes and geographical latitude in recent sediment from the Pacific Ocean. The gradual nature of the phenotypic variation she found in *G. ruber* made her to conclude that extant specimens described as belonging to a morphologically very similar species, *Globigerinoides elongatus* (*Globigerina elongata*, d'Orbigny, 1826) were in fact a part of the large ecophenotypic range of *G. ruber*. Parker (1962) did not question the existence of this d'Orbigny's species, but she conjectured that because its holotype is a reworked specimen from an older formation near Rimini, the species name *G. elongatus* should not be used for extant specimens. Consequently,

extant specimens with *G. elongatus* morphology were declared synonymous with *G. ruber* (Parker, 1962; but see Cordey, 1967). After the taxonomic revision by Parker (1962), the name *G. elongatus* indeed ceased to be used in modern and Quaternary planktonic foraminifera. Thus, the species concept of *G. ruber* was becoming progressively broader through time.

However, in addition to the two easily recognisable chromotypes, most researchers acknowledged the existence of distinct morphological types within *G. rube* "white", considering them either subspecies, such as *G. ruber pyramidalis*, (Sadekov et al., 2008), or furnishing the phenotypic variants with non-taxonomic labels (Parker, 1962; Hecht, 1974). In his morphometric study on the distribution of the phenotypic variation in *G. ruber* from the Atlantic Ocean, Hecht (1974) reported a correlation between the morphological variations and ambient water temperatures. More recent studies, focusing on the chemical properties of these different morphotypes, most often labelled as *G. ruber* sensu lato (s.l.) and *G. ruber* sensu stricto (s.str.) (Wang, 2000), found a significant offset in Mg/Ca – ratios (Steinke et al., 2005) and isotopic values (Wang, 2000; Lin et al., 2004; Kawahata, 2005; Löwenmark et al., 2005; Numberger et al., 2009) between morphotypes. The question whether these differences were represented a species specific signal or were instead of pure environmental origin could not be answered.

Results from molecular phylogenetic analyses based on a fragment of the gene coding for the ribosomal small subunit RNA (SSU rDNA) supported the separate treatment of the two chromotypes in *G. ruber*. The single genotype found in individuals of *G. ruber* "pink" was divergent from the genotypes found in specimens of *G. ruber* "white" (Darling et al., 1997, 1999; de Vargas et al., 1997; Aurahs et al., 2009a). In the most extensive survey of the species, all specimens of the *G. ruber* "pink" chromotype in the Northeast Atlantic Ocean were found to belong to the same genetic type (Aurahs et al., 2009a). However, the four genotypes isolated from specimens of *G. ruber* "white" chromotype were divided into two distinct groups. Two of these types, namely Ia and Ib (including subtype Ib2) share a common ancestor with the pink genotype, forming a cluster that is here named *G. ruber* sensu stricto cluster (Fig.1). The genetic types IIa (including the subtypes IIa1 and IIa2) and IIb, however, were found to share a common ancestor with SSU rDNA sequences derived from individuals of *Globigerinoides conglobatus* (Fig 1.). As *G. conglobatus* is morphologically very distinct from *G. ruber* and is considered only distantly related to it, the *G. ruber* morphospecies was consequently considered to be paraphyletic (Darling et al., 1999; Aurahs et al., 2009a). However, no comprehensive morphological data were available to relate these genetic results to the taxonomical concept used in the group.

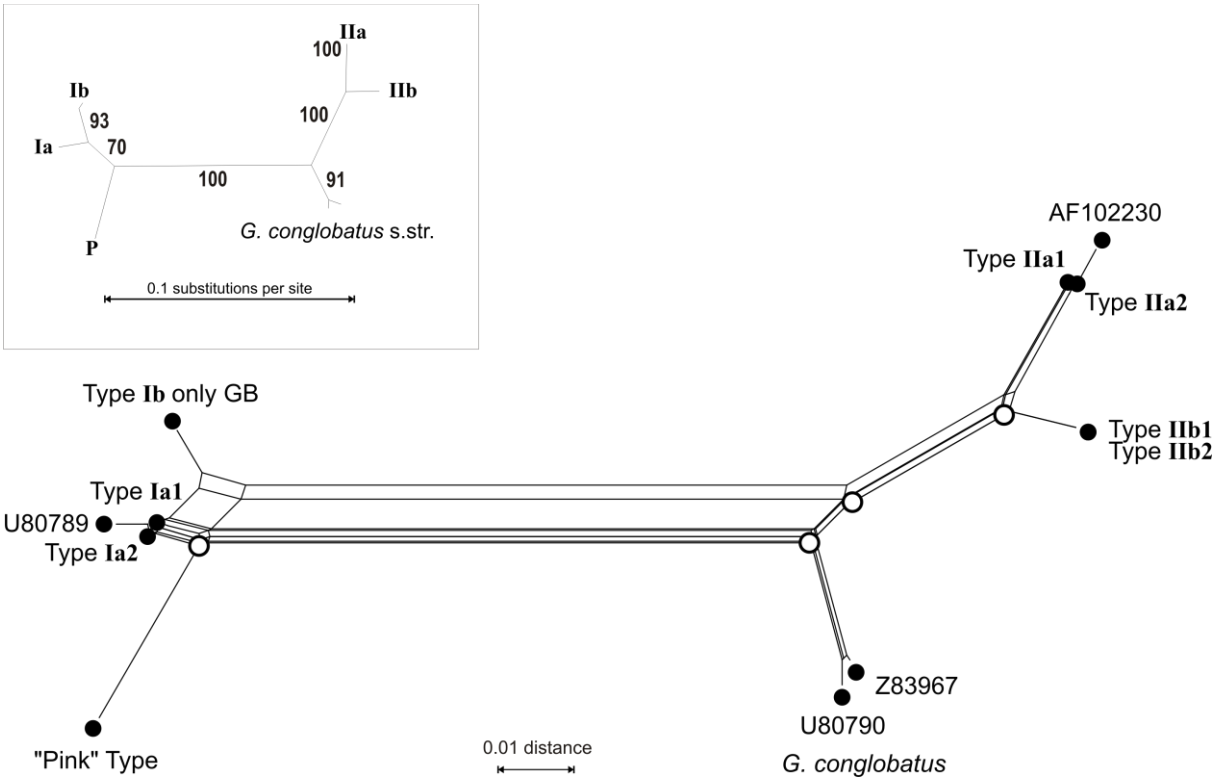

Figure 1 Unrooted Neighbour-Net splits graph tree with genetic types of *G. ruber* and *G. conglobatus.* The insert is showing a Maximum likelihood phylogram based on the same data. The reconstruction is based on an alignment of 709 nt and is redrawn from Aurahs et al. (2009a.).

The challenge of combining molecular genetic results with a morphological species concept lies mainly within the subjective character of a morphological species definition. A combination between genetic divergence and the descriptive morphological species concept in planktonic foraminifera is therefore only possible if quantitative morphometric analyses are applied on defined shell features in specimens with known genetic identity. Morphometric analyses conducted for *Globigerinella siphonifera* (Huber et al., 1997), *Globorotalia truncatulinoides* (de Vargas et al., 2001) and *Orbulina universa* (Morard et al., 2009) have dealt with the challenge of attributing variations in phenotypical characters to the underlying genetic diversity. Even though the researchers found correlations between certain aspects of the shell features they had measured and the genetically defined types, the evidence was not conclusive enough to result in a taxonomical revision of any of the investigated species. Only in the case of *Neogloboquadrina pachyderma*, a morphological validation of the genetic signal resulted in the splitting of a planktonic foraminiferal species based on SSU rDNA sequences. Darling et al. (2006) were able to correlate the coiling direction in specimens of *N. pachyderma* to two divergent genetic types present in the species and assigned dextrally coiled specimens to the existing, but rarely used species name *Neogloboquadrina incompta*.

Even though the approaches on the other planktonic foraminiferal species did not result in newly defined species, the potential of combining morphometric measurements and genetic data for a refined resolution in classical planktonic foraminiferal morphospecies was demonstrated. In an survey of *G. ruber* from the Western Pacific Ocean, Kuroyanagi et al. (2008) first reported a congruence between two different genetic types in *G. ruber* white (type I and type II) with the definitions of the morphotypes *G. ruber* s.str. and *G. ruber* s.l., following the concept of Wang (2000). If it could be shown that the morphological variants in the white chromotype of *G. ruber* correspond to ecologically distinct species, this would have potentially large implications for the interpretation of abundance patterns and geochemical signals in data where *G. ruber* (white) has been treated as a single species (see Numberger et al., 2009).

Here, we present results from morphometric measurements on (i) specimens assigned to *G. ruber* "white" and "pink" in sediment samples from the Alboran Sea (ii) genotyped individuals from plankton tows from the Northeast Atlantic, Mediterranean Sea and Arabian Sea, identified as the genetic types Pink, Ib and IIa based on SSU rDNA sequences and (iii) individuals from the collection of the Smithsonian National Museum of Natural History (NMNH) in Washington, DC, identified as *G. ruber* and *G. elongatus* by various researchers. We then extracted data on species occurrence from the CHRONOS database and use these in combination with literature data and the available SSU rDNA sequence types assigned to *G. ruber* to revise the phylogeny of this group and to estimate the date of divergence between the *G. ruber* sensu stricto cluster and the *G. conglobatus/G. ruber* sensu lato cluster in a molecular clock approach.

Material and Methods

Sampling

In this study, we analysed and compared four different sets of images of shells of *Globigerinoides ruber*, taken from various sources, including a) the live plankton, b) recent sediment, c) a museum collection and d) the recent literature. Live specimens of *Globigerinoides ruber* where isolated from stratified plankton tows in the North-Eastern Atlantic Ocean around the region of the Canary Island and in the western Mediterranean Sea in 2006 (RV METEOR 69/1) and in the eastern Mediterranean Sea in 2007 (RV METEOR 72/1 and 72/3). We used a multiple closing net (mesh size 100 µm) for vertical sampling of the water column, as well as surface water (mesh size 68 µm) from the ship's uncontaminated seawater supply. The specimens were isolated under a binocular microscope on board, identified and digitally photographed (see next section) and then processed for DNA analysis (Aurahs et al. 2009a). The majority of the genetically analysed specimens belonged to the SSU rDNA genotypes Pink and IIa; the other two genetic types Ia and IIb either did not yield a sufficient number of individuals to allow a sound statistical treatment, or were extracted from small (~100 µm), potentially preadult individuals (Fig. 1, Aurahs et al., 2009a). In this study we used the digital images of 104 Type Pink and 43 Type IIa specimens. In addition to our samples from the Atlantic, twelve shells of *G. ruber* "white" specimens from the Arabian Sea (METEOR 74/1b in 2007) were provided by Kate Darling. These specimens were attributed to the genetic type Ib (Fig. 1; Kate Darling, pers. comm. 2009).

As individuals of planktonic foraminifera in the water column are found in different developmental stages, we compared the plankton samples with adult shells of *G. ruber* from recent sediment. For this we chose a location where the genetic variability of the *G. ruber* community in the water column has been characterised. Aurahs et al. (2009a) have shown that the southwest Mediterranean and the adjacent Gulf of Cadiz were dominated by types Pink (~74%) and IIa (~24%) whereas Type Ia (~2%) were rare. Thus, sedimentary shells of *G. ruber* were sampled from the top 0.5 cm of the sediment recovered from multicorer station 339-2 taken during the cruise M 69/1, in the Mediterranean Sea close to the Strait of Gibraltar (36°18.34'N, 3°8.37'W, 850m water depth). The core top sediment sample was freeze dried and then soaked in distilled water for 30 min before washing over a sieve with 63 µm mesh size. Clay remains from foraminifera shells were removed by agitating the residues in an ultrasonic bath for 15 s and then washed again. The final residues were transferred onto filter paper and dried at 40 °C for 24 h. The dried fraction was collected from the filters and dry sieved for the size fractions of 63–150 µm and ≥150 µm. For the morphometric analyses, the ≥150 µm fraction was splitted with a microsplitter. Specimens of *G. ruber* were quantitatively picked from splits containing a representative aliquot of the sample and separated by their colour into *G. ruber* "pink" (n=88) or "white" (n=145), irrespective of their morphotype.

In order to assess the congruence of historical taxonomic practice with the observed morphological variability in the plankton and sediment samples, we have collected light-microscope images of a range of specimens from the collections of the National Museum of Natural History in Washington DC. Specifically, we selected at random specimens in the collections identified as *G. ruber* or *G. rubra* (irrespective of shell coloration; n= 41) and *G. elongatus* (n=68). A complete sampling of the collection would be impossible due to the large number of specimens identified as *G. ruber* and difficulty in searching the entire collection for non-type material. The specimens originated from a range of locations collected over decades, mostly from recent material (see online supplement).

In order to connect and validate our data with the dominant phenotypic concept for *G. ruber* morphotypes used in the recent literature, we performed measurements on SEM images of specimens defined as *G. ruber* s.l. taken from publications of Wang (2000), Kawahata (2005), Steinke et al. (2005), Löwenstein et al. (2005) and Kuroyanagi et al. (2008).

Digital imaging and morphometric measurements

The range of sources of our samples inevitably resulted in a number of different microscopic and digital imaging setups. The live specimens of *G. ruber* collected in the Atlantic Ocean and Mediterranean Sea were photographed on board using a digital camera mounted on a stereomicroscope, before being processed further for genetic analyses (Aurahs et al., 2009a). Shells from the sediment sample were mounted on a glass slide using a double-sided adhesive tape. The individuals of *G. ruber* white of the Ib genotype, collected in the Arabian Sea were separated in multi-cavity microscope slides. Both collections were digitally photographed with a QIMAGING MICROPUBLISHER 5.0 RTL digital camera mounted on a LEICA Z16 APO stereomicroscope. Specimens from the NMNH in Washington DC were photographed using a Zeiss Axiocam camera mounted on a Zeiss SteREO Discovery V12 stereomicroscope. The set-up allowed multiple images of the same specimen with changing focus, afterwards merged into a single image using Z-STACK provided by the Zeiss AxioVision software. The resulting images have an artificially extended depth of field and a resolution that is not achieved with any other stereomicroscopic set-up we used. The SEM images from the literature were taken in a digital form from the PDF versions of the respective publications.

All measurements were performed by a single researcher (Y.T.) using the software IMAGE ProPlus 6.0. The choice of the morphometric characters to be measured on the specimens was guided by prior observation that the chambers of the last whorl in *G. ruber* individuals with the IIa genotype showed a stronger lateral compression than specimens that yielded the Types Pink and Ia.

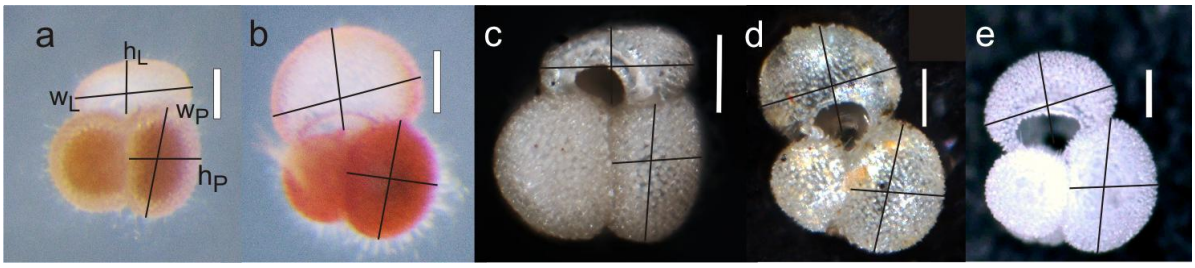

Figure 2 Morphometric parameters used for the characterisation of the compression of the last two chambers in specimens assigned to the *G. ruber* morphospecies. The images mages show representative individuals of genotyped plankton samples, identified as a) Type IIa, b) Type Pink and e) Type Ib; museum collection specimens labelled as c) *G. elongatus* and d) *G. ruber*; w_L*: width of last chamber;* h_L*: height of last chamber;* w_P*: width of penultimate chamber; hP: height of penultimate chamber.*

We aimed at finding a character that could be easily replicated, that could be applied to the various types of images analysed in this study and that could be linked to the existing definition of *G. ruber* morphotypes. Thus, following the definition of *G. ruber* s.l. by Wang (2000), who used the more compressed final chambers in the last whorl of *G. ruber* s.l. as the main difference between *G. ruber* s.l. and *G .ruber* s. s., we defined four parameters to be measured on the ultimate (h_L =height of last chamber; w_L = width of last chamber) and the penultimate chamber (h_P = height of penultimate chamber; w_P = width of penultimate chamber) (Fig. 2).

The measurements were converted into ratios (width against height), describing the degree of chamber compression or "ellipsity" of the ultimate and penultimate chamber ($e_L = h_L/w_L$ and $e_P = h_P/w_P$). We tested the resolving strength of the measured features in a discriminate analyses using Statistica 8 (StatSoft). Descriptive statistics and plots were performed in OriginPro 8.0 (OriginLab).

Molecular clock

For each published SSU rDNA sequence type derived from foraminiferal specimens identified as *Globigerinoides ruber* and *G. conglobatus* we chose a representative sequence from the international gene-bank (http://www.ncbi.nlm.nih.gov/Genbank/; Table 1). Sequences of *Globigerinoides sacculifer* and of one genetic type *Orbulina universa* were chosen as outgroup, based on earlier results from phylogenetic reconstructions (e.g. Darling et al., 1999; Aurahs et al. 2009b). The sequences were automatically aligned using the online available, up-to-date versions of CLUSTALW2 (Larkin et al., 2007), KALIGN version 2.03 (Lassmann and Sonnhammer, 2005), MUSCLE (Edgar, 2004), and MAFFT version 6.24 (Katoh et al., 2005). The algorithms were used under their default settings. Unlike the usual procedure of cutting the highly variable sites in order to compensate for the homology problem in these regions we here did not discard any part of the sequence alignment, following an earlier extensive investigation, which demonstrated the advantages of this method (e.g. Aurahs et al., 2009b). This approach includes as many of the informative sites into a clock calculations as possible, instead of dispensing any sites *a priori*. By covering a considerable amount of alignment space, we aimed to quantify the uncertainty in the molecular clock estimates resulting from alignment ambiguity in addition to the degree of uncertainty resulting from the uncertainty associated with the calibration data derived from the fossil record.

Estimates of divergence time and substitution rate were performed using Bayesian methods as implemented in BEAST 1.4.8. (Drummond and Rambaut, 2007). The four different alignments were tested under various assumptions of differently fixed dates of divergence (time of most recent common ancestor = tmra) and clock models (strict, uncorrelated lognormal and exponential) (Table 2). We hypothesised three nodes in the phylogenetic reconstruction to be well supported by the fossil record.

Table 1 Sequences of *G. ruber* genotypes, *G. conglobatus*, *G. sacculifer* and *O. universa* used in the molecular dating approach. The genetic types IIa to IIb, found to be more related to *G. conglobatus* (Fig. 1) are labelled as *G. ruber**.

Morphospecies	**Genetic type**	**Gene Bank Accession number**	**Nucleotide size (bp)**
G. ruber	Pink	U65634	993
G. ruber	Ia	U80789	981
G. ruber	Ib	Z69599	1.005
G. ruber	Ib	Z83965	1.003
G. ruber	Ib2	EU012470	1.016
G. ruber *	IIa	AF102230	996
*G. ruber**	IIa1	FM866194	820
*G. ruber**	IIa2	FM866181	902
*G. ruber**	IIb	FM866139	863
G. conglobatus	-	AB263465	1.047
G. conglobatus	-	U80790	1.027
G. conglobatus	-	Z83967	1.046
G. sacculifer	-	U65633	1.016
Orbulina universa	Ia	Z83962	988

Table 2 Calibration ages (in Ma) assumed for the five different molecular clock trials for each of the four automated alignments. Each trial was tested under strict clock, uncorrelated lognormal and uncorrelated exponential relaxed clocks. All trials were run using the Yule process for speciation, GTR + invariant sites for substitution model and a UPGMA starting tree. The distributions of the fixed node age priors were considered normal, with a standard deviation of 0.5. All other priors adjustable in BEAST were used in their default settings.

Trials	**1**	**2**	**3**	**4**	**5**
FAD *G. trilobus*	tree prior	24.0	24.0	24.0	24.0
FAD *P. sicana*	17.0	17.0	tree prior	17.0	17.0
FAD *G. conglobatus*	tree prior	tree prior	tree prior	6.2	8.3

First, we considered that the divergence of the *G. ruber/G. conglobatus* lineage from the *G. trilobus/O. universa* lineage must have taken place by the time of the FAD of *G. trilobus*. The exact phylogenetic topology of the divergence in the latest Oligocene remains uncertain, but if a sister status of both lineages is assumed, as indicated by earlier phylogenetic analyses bases on molecular genetic data (e.g., Aurahs et al., 2009b), then it is reasonable to assume that the divergence occurred from *G. primordius* as the common ancestor. The commonly cited age of the FAD of *G. trilobus* is within the Zone M1, defining the base of the Neogene at 23 Ma; (Berggren et al., 1995). This date was confirmed by a taxonomic search of the CHRONOS database (http://chronos.org; search generated by MK using CHRONOS XML searches of the Janus database on 22th October 2009). Obvious outliers and records indicating taxonomic uncertainty (s.l., subspecies or synonyms of unclear significance) have been manually removed. Even after the removal of obvious outliers, the database contained a large number of records of occurrences of this species from the latest Oligocene and we have thus decided to use an older date of 24 Ma for the calibration.

The second node with a well defined age in the literature is the split between *G. trilobus* and *O. universa*. We here use the well constrained FAD of *Praeorbulina sicana* (synonymous to *G. bisphericus*) considered to have taken place around 16.4 Ma (Berggren et al., 1995). Considering the observations on the synonymy between *P. sicana* and *G. bisphericus* by Pearson and Chaisson (1997) and the significant number of slightly older occurrences found in the CHRONOS data, we have decided to date this node at 17 Ma.

For a third node, the split between The Type II genotypes (IIa and IIb) and the sequences attributed to *G. conglobatus*, we used the FAD of *G. conglobatus* as a calibration point. Yet, depending on the source, *G. conglobatus* is first reported to have occurred abundantly in the fossil record either at 6.2 (Chaisson and Pearson, 1997) or 8.3 Ma (based on CHRONOS data, see also Kucera and Schönfeld, 2007). We thus decided to use both ages in two separate trials and monitor their effect on the ages of the other nodes. In a similar manner, we decided to monitor the effect of the choice of calibration points on the resulting molecular clock estimates by applying five combinations of priors (Table 2) to the four alignments tested under the three molecular clock assumptions. The distributions of the fixed node age priors were considered normal, with a standard deviation of 0.5 Ma.

As substitution model we choose GTR + invariant sites, which is considered to be most adequate when dealing with highly variable and conservative regions between sequences (BEAST manual). Speciation was assumed under the Yule process (which assumes a constant speciation rate) and a UPGMA tree was calculated as a starting tree in all trials. A Markov-Chain-Monte-Carlo (MCMC) was performed for 10.000.000 steps, saving every 1000th step, resulting in 9001 tree topologies. The maximum credible tree with mean node heights was determined using TREEAnnotator from the

BEAST package, discarding the first 100 trees under a posterior probability limit of 0.5. The final trees were analysed using FigTree 1.2.2.

Results

Morphometric Analysis

Measurements of h_L (height of last chamber), l_L (length of last chamber), h_P (height of penultimate chamber) and l_P (length of penultimate chamber) were performed on 513 images. The resulting ratios e_L (ellipsity of last chamber) and e_P (ellipsity of penultimate chamber) vary between 0.51 - 0.72 for e_L and 0.57 - 0.78 for e_P (Table 3). The data can be interpreted as showing a tendency towards more compressed last chambers in some groups. Only *G. elongatus* from the museum collection, *G. ruber* s.l. from the literature, *G. ruber* 'white' from the sediment sample and the individuals of genotype IIa show mean values for e_L that are <0.6, the other groups have eL values >0.6 (Table 3; Fig. 3). The most compressed chambers are measured in the individuals of *G. elongatus* from the museum collection, the least compressed chambers are found in the specimens of the Type Pink genotype (Fig 3). As can be seen from Fig. 3, the mean values of Type IIa, *G. ruber* from the museum collection and *G. ruber* s.l. from the literature take an intermediate position within the dataset. These three groups overlap with most other groups in their degree of compression of either their last or penultimate chamber (Fig. 3; Table 4). Type Ib has a similarly intermediate mean value of e_P (~0.7) with a relatively large confidence interval (Fig. 3; Table 4). Even though this artefact cannot be expressed in a definite value, its existence within the data can be deduced from Fig. 3. In a pairwise comparison of image collections obtained under the same conditions i.e. Type IIa and Type Pink, *G. elongatus* and *G. ruber* s.l., as well as *G. ruber* 'white' and 'pink', the groups are separated by a similar distance and direction in the bivariate space, but shifted with respect to each other.

The in-group variance of the last chamber compression in almost all groups is at least twice smaller than the in-group variance of the penultimate chamber compression (Table 3). All groups except Type Ib have a significantly more compressed last chamber than penultimate chamber ($e_L < e_P$; $p < 0.05$; Table 4, Fig. 3). In a combined discriminate analysis, only 42.6 % of all specimens could be correctly classified. This is indicatory for the large overlap in the e_L and e_P ratios between the groups. Only the individuals of the genetic Type Pink and the sediment samples from *G. ruber* white were identified in reasonably high numbers (78.8 and 68.9 % correct grouping). When compared pairwise, the groups' percentages of correct classification varied between 56.5 and 100 % (Table 5). The two genotypes IIa and Type Pink differ significantly in their e_L/e_P ratio and their specimens can be correctly classified to ~75% (p<0.0001; Table 5).

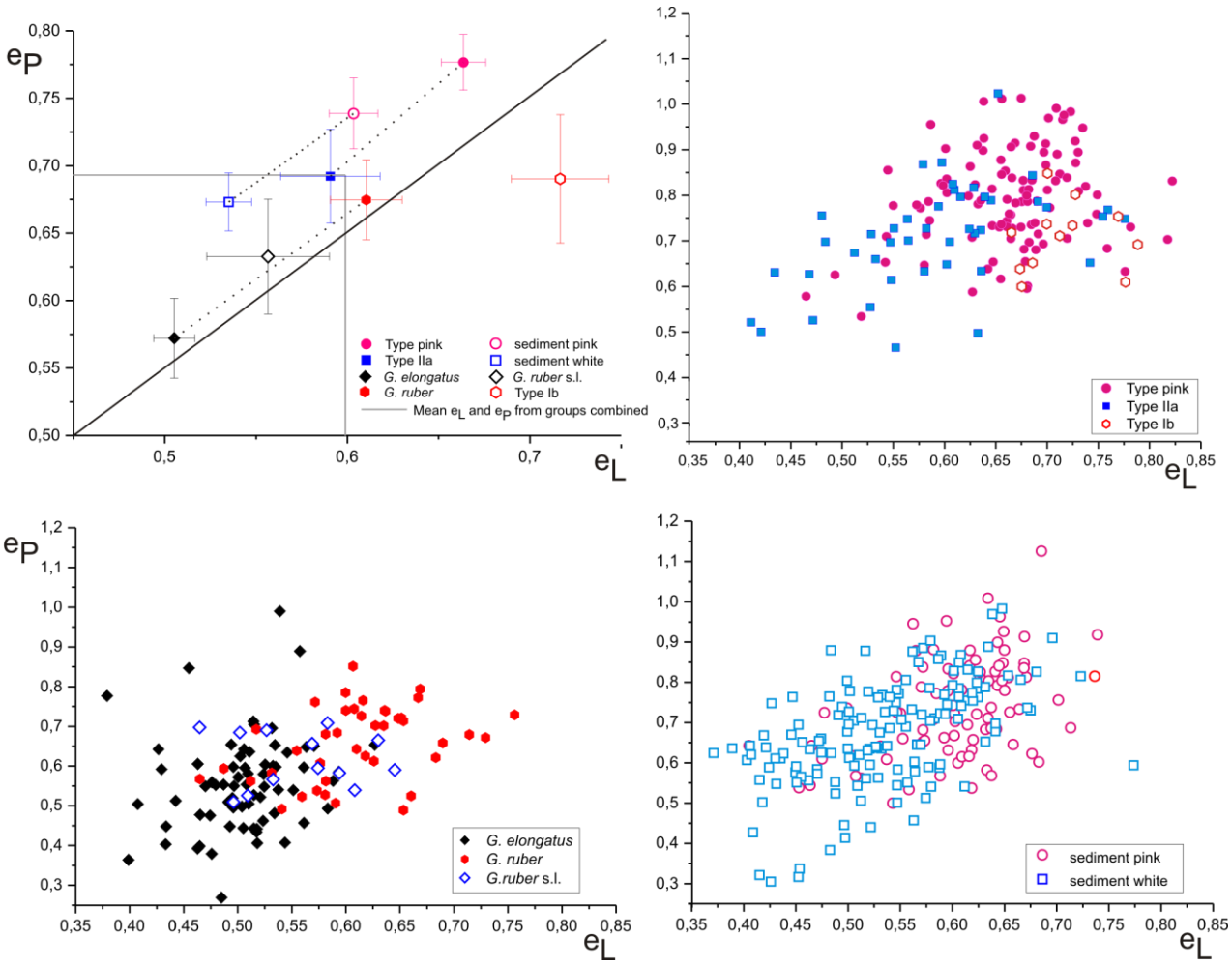

Figure 3 The relationship between mean values of eL and eP from individual image collections, error bars represent the 95% confidence interval of the mean. Doted lines connect data points from the same source, e.g. plankton images of genotyped specimens of Type Pink and Type IIa. Solid dark line marks the 1:1 relation between e_L and e_P. The solid grey lines mark the mean values of e_L and e_P for all groups combined.

Table 3 Descriptive statistics on the chamber compression ratios in all groups**.**

		N total	ratio	mean	variance	min.	max.
Genotyped individuals	Type Pink	**104**	e_L	0.664	0.004	0.465	0.822
			e_P	0.777	0.011	0.516	0.995
	Type IIa	**43**	e_L	0.591	0.008	0.411	0.776
			e_P	0.692	0.013	0.448	1.006
	Type Ib	**12**	e_L	0.717	0.002	0.665	0.789
			e_P	0.690	0.006	0.582	0.831
Museum / literature	*G. elongatus*	**68**	e_L	0.505	0.002	0.379	0.627
			e_P	0.572	0.015	0.285	1.006
	G. ruber	**41**	e_L	0.610	0.004	0.465	0.756
			e_P	0.675	0.009	0.505	0.868
	G. ruber s.l.	**13**	e_L	0.557	0.003	0.465	0.645
			e_P	0.633	0.005	0.526	0.725
Sediment samples	*G. ruber* ‘pink’	**88**	e_L	0.603	0.004	0.405	0.739
			e_P	0.739	0.015	0.500	1.126
	G. ruber ‘white’	**145**	e_L	0.535	0.006	0.371	0.774
			e_P	0.673	0.017	0.305	0.983

Table 4 Differences (row minus column) between the mean values for chamber compression between groups. The in-group difference between e_L and e_P (diagonal) has been tested for significance in paired t-tests. The values above the diagonal show e_P differences, the values below the diagonal show e_L differences between groups, both from two sample *t*-tests; * highlights significant differences ($p<0.05$).

	Δe_L / Δe_P	**Pink**	**IIa**	**Ib**	***G. ruber***	***G. elongatus***	***G. ruber* s.l.**	***G. ruber* 'white'**	***G. ruber* 'pink'**
Geno.ype	**Type Pink**	**-0.113***	-0.085*	-0.086*	-0.086*	-0.205*	-0.144*	-0.104*	-0.038*
	Type IIa	0.073*	**-0.102***	0.002	0.018	0.019	0.060	0.020	-0.047*
	Type Ib	-0.053*	0.126*	**0.026**	0.016	0.118*	0.058	0.017	-0.049
Museum/ Literature	***G. ruber***	0.053*	0.020	-0.106*	**-0.064***	0.103*	0.042	0.002	-0.064*
	G. elongatus	0.159*	-0.086*	-0.211*	-0.105*	**-0.067***	-0.061	-0.101*	-0.167*
	***G. ruber* s.l.**	0.107*	-0.034	-0.160*	-0.054*	0.051*	**-0.076***	-0.040	-0.106*
Sediment	***G. ruber* 'white'**	0.129*	-0.056*	-0.182*	-0.075*	0.030*	-0.021	**-0.138***	-0.066*
	***G. ruber* 'pink'**	0.060*	0.013	-0.113*	-0.007	0.098*	0.047*	0.068*	**-0.135***

Tabel 5 Percentage of correct assigned specimens in a linear discriminate analyses of chamber compression ratios (values below diagonal); values above the diagonal show the result of a pairwise Hotelling's T^2-test; ** indicates $p < 0.001$; * indicates $p < 0.05$

	T^2	**Pink**	**IIa**	**Ib**	***G. ruber***	***G. elongatus***	***G. ruber* s.l.**	***G. ruber* 'white'**	***G. ruber* 'pink'**
Geno.ype	**Discriminat**								
	Type Pink	-	**	**	**	**	**	**	**
	Type IIa	74.2	-	**		**		**	
	Type Ib	75.0	83.6	-	**	**	**	**	**
Museum/ Literature	***G. ruber***	74.5	58.3	84.9	-	**	*	**	*
	G. elongatus	94.2	80.2	100	85.3	-	*	**	**
	***G. ruber* s.l.**	86.3	66.1	98.8	68.5	71.6	-		*
Sediment	***G. ruber* 'white'**	81.5	62.8	93.0	72.0	65.3	64.6	-	**
	***G. ruber* 'pink'**	69.8	56.5	90.0	58.9	85.9	67.3	71.7	-

The Arabian Sea specimens of Type Ib are significantly different from all other groups. The specimens of *G. ruber* 'white' and 'pink' derived from the recent sediment show ~71% correct classification ($p<0.0001$; Table 5). The individuals of *G. elongatus* and *G. ruber* from the museum collection can be classified with an accuracy of ~85%.

However, we believe that our dataset is affected by methodological inconsistency, as a consequence of the different sources of the material and the equipment they were photographed and measured with. Measurements performed on the individuals of Type IIa and *G. ruber* sensu lato from the recent literature cannot be discriminated from one another with any significance (~66% correct classification). Only ~ 64 % of the individuals of *G. ruber* s.l. and the individuals of *G. ruber* 'white' from the sediment can be discriminated from one another. All other data sets are significantly different from the e_L and e_P ratios of *G. ruber* s.l. (Table 5).

The digital images of the measured shells can be requested from the corresponding author. Measurements are available as online supplement.

Molecular Dating and times of divergence

The variety of different molecular clock presets (combination of priors, fixed nodes and clock models) resulted in a total of 60 reconstructions, 15 for each alignment (Table 2). In 54 of the 60 trials, the same tree topology has been recovered as the most parsimonious (Table 6; Fig. 4). Here, *G. trilobus* and *O. universa* were placed as a monophyletic sister clade to the remaining sequences, which had exactly the same phylogeny as shown in Fig.1. For the KALIGN and MUSCLE alignments, BEAST reconstructed alternative tree topologies under the strict and lognormal relaxed clock assumption of trials 1 (no calibration date for the root) and 3 (no calibration date for the split of *G. trilobus* and *O. universa*, and for the split of *G. conglobatus* and the Types II; Table 2), were unable to resolve the relationship of *G. trilobus* and *O. universa*. The six resulting reconstructions had *G. trilobus* singled out as sister to the rest of the tree. Given the strong fossil support for the sister relationship between *G. trilobus* and *O. universa* confirmed by molecular phylogenies (e.g. Aurahs et al., 2009b), the time estimates from these six reconstructions were excluded from further interpretation. The results of all molecular age estimates are provided in the online supplement.

In general, the node age estimates varied more strongly between the different clock models then they did between the four automated alignments. Moreover, the differences between the different clock assumptions were unevenly distributed within the reconstructions, whereas the separate alignments

resulted in a far more similar age offset for all the nodes. Only under the assumption of a strict clock model, a single fixed node resulted in a relatively large deviation between alignments. KALIGN and MUSCLE alignments were not able to resolve the relationship of *G. trilobus* and *O. universa* as direct sisters in this particular case. Interestingly, under the relaxed clock models, the same alignments were able to produce a phylogeny identical with the one from the other two alignments. The trials where three ages were fixed resulted in the most similar time estimates between the four alignments.

FAD G. trilobus: The split of the *G. ruber*/*G. conglobatus* in-group and of the *G. trilobus/O. universa* out-group was considered to correspond with the FAD of *G. trilobus* around 24.0 Ma in the fossil record. In the trial where this node had not been fixed, the reconstructed ages of the split varied between ~33 Ma (CLUSTALW alignment) and ~20 Ma (MUSCLE AND MAFFT alignment). The mean age for the node in this trial (combined from all twelve combinations of alignments and clock models) was 25.5 Ma, close to the age (24 Ma) we used in all the other trials.

FAD Preorbulina: We hypothesised the age of the split between *G. trilobus* and *O. universa* to be consistent with the FAD of *Preorbulina* dated at ~17 Ma. The trial where this node had not been fixed and only the root was fixed at 24.0 Ma resulted in age estimates for the FAD of *Preorbulina* between ~19 Ma (from CLUSTALW2 alignment) and ~10 Ma (from MUSCLE alignment), with a mean age of all estimates for this trial at ~14 Ma.

Split of the G. ruber s.str. and G. ruber s.l/G. conglobatus lineages: The dating of the split between the *G. ruber s.str.* lineage and the *G. conglobatus*/Type II cluster varied slightly between alignments, ranging from 14.4 to 16.0 Ma (clock models and trials integrated). In all alignments (trials integrated), the different clock models had a slightly larger offset, ranging from 13.3 (strict clock) to 16.4 Ma (uncorrelated lognormal). The mean age of this split, different clock models and alignments combined, is 15.1 Ma. As in all the higher nodes, the relaxed clock models tended to produce older date estimates than the strict clock approach.

Split of G. conglobatus and Type IIa + b: We considered the FAD of *G. conglobatus* in the fossil record as the split between this species and the last common ancestor of the genetic types IIa and IIb. The two dates we tested as fixed ages for this node, 6.2 and 8.3 Ma, showed little differences in their effect on the other nodes. Only the dating of the split between Type IIa and IIb was affected by being between 0.5-2 Ma older under the assumption of an 8.3 Ma FAD of *G. conglobatus*. In the three trials where the node had no calibration date, the age estimates from the strict clock model were again younger (~ 5 Ma) than the relaxed clock models (~9 Ma).

Table 6 Divergence age estimates in the *G. ruber* phylogeny from the molecular dating approach; all values are integrated from the four different automated alignments and all trials. * highlights the three nodes that have been linked with fixed ages in at least one of the trials (Table 2). Two alignments could not resolve the phylogenetic relationship under the strict clock assumption properly. Therefore, only two uncalibrated, dates exist for these nodes, and no mean age under the strict clock assumption was calculated.

		Mean/Min./Max. age (in Ma)			
Node	Total mean	Strict clock	uncorrelated lognormal	uncorrelated exponential	FAD in the fossil record
*Root; Split in-group – out-group	25.5	- /19.8 / 30.6	24.5 / 19.9 / 30.4	26.4 / 21.6 / 33.5	FAD *G. trilobus* ~24 Ma
*Split *G. trilobus*-*O.universa*	13.9	- /13.2 / 19.8	13.9 / 12.6 / 16.2	12.6 / 10.3 / 14.1	FAD *Praeorbulina* ~17 Ma
Split *G. ruber* – *G. conglobatus*	15.1	13.3 / 12.1 / 14.7	15.4 / 12.0 / 18.9	16.4 / 12.7 / 20.2	~12 Ma FAD *G. extremus*
*Split *G. conglobatus* – Types II	8.3	5.6 / 5.0 / 6.4	7.3/ 5.8 / 9.5	10.3 / 8.6 / 13.9	FAD *G. conglobatus* ~8 Ma
Split Type Pink – Types I	6.4	4.2/ 3.0/ 5.0	6.1/ 4.5/ 8.2	8.4/ 7.4/ 10.6	no fossil evidence
Split Type II – Type IIb	3.4	2.6/ 2.0/ 3.2	3.3/ 2.6/ 4.0	4.0/ 3.0/ 5.7	~4 Ma (FAD *G. elongatus*?)
Split Type I – Type Ib	2.7	1.5/ 1.4/ 1,7	2.4/ 1.8/ 3.3	4.0/ 3.5/ 5.2	no fossil evidence

Split of G. ruber Type Pink and Type Ia and Ib: The last common ancestor of the *G. ruber* pink genotype and the *G. ruber* white genotypes Ia and Ib is dated in all our trials between 4.2 Ma (strict clock) and 8.4 Ma (uncorrelated exponential), regardless of the number of fixed nodes. The overall mean age of this node was 6.4 Ma. This is considerably older than the reports of the first shells of *G. ruber* pink in the sediment record (< 750 ka, Thompson et al. 1979)

Split of Type IIa and Type IIb: The dating of the split of the two genotypes ranges from 2.6 Ma (strict clock model) to 4.0 Ma (uncorrelated exponential) and a mean node age of 3.3 Ma. The age estimate for the split is therewith close to the FAD of *G. elongatus* (~4 Ma, Perconig 1969)

Split of G. ruber Type Ia and Ib: The divergence age estimated for the two *G. ruber* white genotype Ia and Ib ranges from ~2 to 5 Ma, the mean node age being 2.7 Ma.

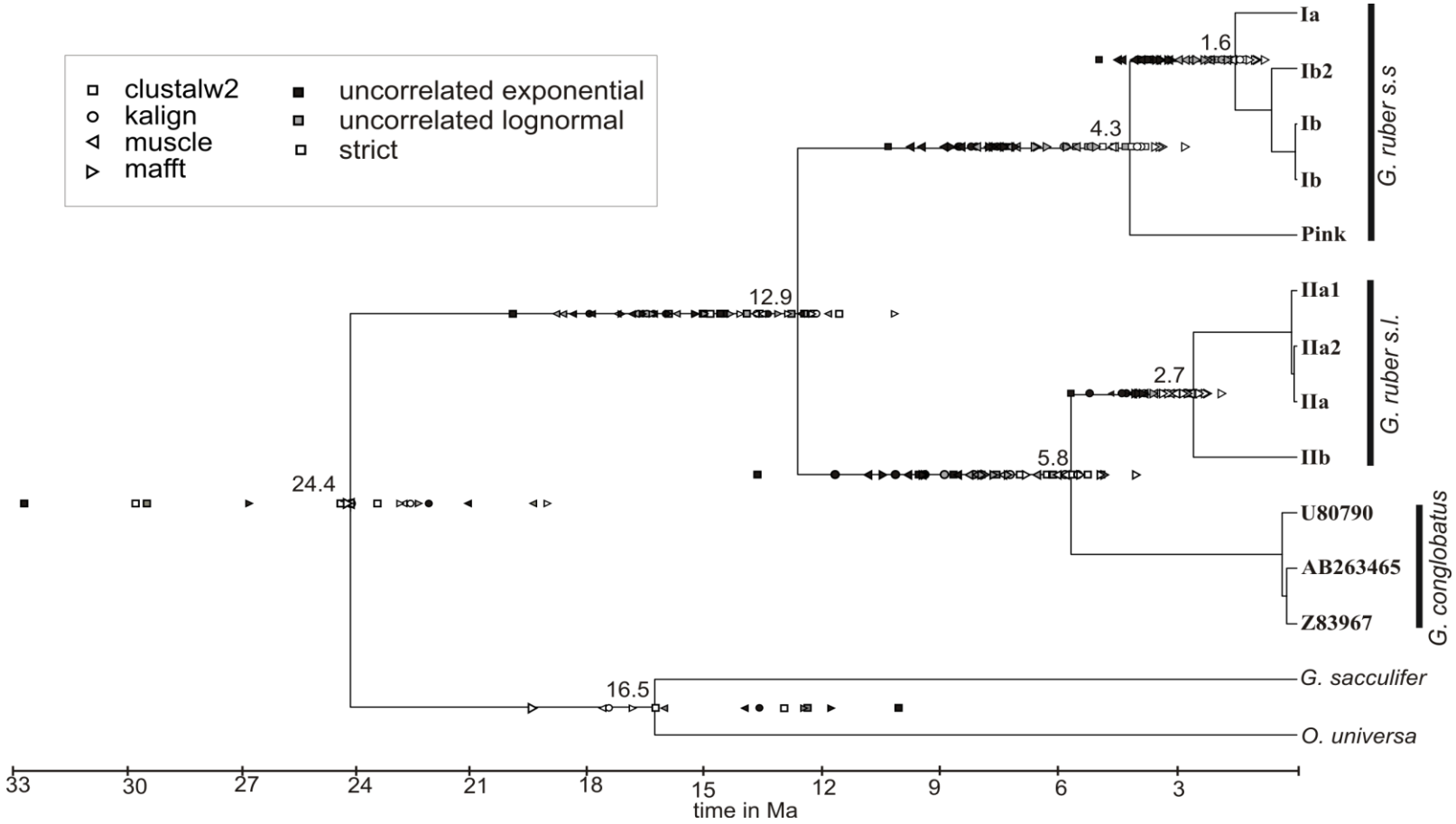

Figure 4 Molecular phylogeny of the genus *Globigerinoides* with time estimate ranges from the different molecular clock presets. The tree shown resulted from the CLUSTALW alignment and trial 4 under the strict clock assumption. Numbers at each node indicate the divergence ages estimated from this particular trial. Symbols indicate individual time estimates under the various assumptions of node ages, alignments and clock models. Time estimates, confidence intervals and substitution rates are presented in detail in the online supplement.

Discussion

Morphometric signals in G. ruber phenotypes

Phenotypic plasticity is a well documented feature in planktonic foraminiferal morphospecies, with deviations from typical morphological characteristics of each species being commonly attributed to the influence of habitat parameters on its shell growth (e.g. Kennett, 1976; Kahn, 1981). Consequently, the colour dimorphism in *Globigerinoides ruber*, as well as its morphological variants have all been regarded to be caused by environmental influences (Parker, 1962; Tolderlund and Bé, 1971; Hecht, 1974). With the discovery of distinct SSU rDNA genotypes in *G. ruber* and many other planktonic foraminiferal species, this assumption became, at least in theory, verifiable. In *Globigerinella siphonifera*, two morphologically and ecologically differing types were reported to correlate with the genetic divergence in the species (Huber et al., 1997; de Vargas et al., 2002). Similar correlations were reported from *Globorotalia truncatulinoides* (de Vargas et al., 2001) and *Orbulina universa* (Morard et al., 2009). In this study, we combined morphometric molecular and fossil data, to test whether such a correlation between genetic distinction and shell morphology also exists in the *G. ruber* morphospecies.

The comparison of images from genotyped individuals of *G. ruber* (white and pink) from plankton samples indicate the possibility to separate the SSU rDNA genotypes IIa and Pink by differences in their shell morphology. Besides the fact that Type IIa was only found in individuals of *G. ruber* 'white' and only specimens of *G. ruber* 'pink' yielded the Pink genotype (Aurahs et al., 2009a), the form of the chambers of the last whorl in the shells of Type IIa was strikingly different from the 'classic' inflated chambers seen in the individuals of Type Pink (Figs 2, 3). This observation is consistent with molecular phylogenetic reconstructions (Fig. 1; Darling et al. 1999; Aurahs et al., 2009a), which place Type IIa outside of the *G. ruber* sensu stricto clade, as a sister of *G. conglobatus*, which, too, is characterised by strong compression of the chambers in the last whorl.

As the results from the discriminate analysis show, the individuals of Type IIa and Type Pink are significantly different in terms of the compression of their last and penultimate chamber (Table 5). Regardless of their colouration, 75 % of the individuals could be correctly classified. This is a surprisingly good resolution, taken into account that the images of the specimens were all taken aboard a moving research vessel, the shells emerged in water and not all ideally orientated (i.e. tilting positions, primary aperture not fully facing the camera). Moreover, plankton samples contain a range of ontogenetic stages of planktonic foraminifera which do not always show the taxonomically important adult characters (Brummer et al., 1986). Thus, preadult stages of Type IIa in our collections might not have developed the compression of their final chambers yet, preventing a better separation of the two genotypes. This was one of the reasons for the attempt to validate the morphometric signal we found in the living plankton

with shells of *G. ruber* ‘white’ and ‘pink’ from recent sediment. This exercise showed that in a region where the two genetic types were expected to dominate, the same degree of morphometric separation can be found in shells of the foraminifera that accumulated in the surface sediment over decades to centuries. This is remarkable, considering that in the sedimentary foraminifera, it was not possible to ensure that specimens of the white chromotype all belonged to Type IIa genotype. The observed similarity in the degree of separation thus implies that Type IIa dominated the *G. ruber* white specimens throughout the time of deposition of the surface sample. Alternatively, if the other genetic types of *G. ruber* white were frequently represented in the sedimentary material, they too must have shown a morphological divergence in the ellipsity of the last chambers from the Pink genotype. The possibility to separate morphologically Type IIa and Pink in the plankton material from the same region and season as well as from the time-integrated signal in the sediment suggests that the observed morphological separation is not an ecophenotypic signal (Fig. 5). A clear genetic component in the morphological variability in *O. universa* was also reported by Morard et al. (2009), although both theirs and our study suffer from insufficient numbers of specimens collected from the same plankton haul.

If we assume that the observed morphological difference between Type IIa and Pink reflects their genetic separation, then the broad taxonomic concept of *G. ruber* as originally introduced by Parker (1962) must be abandoned. The few data available from Type Ib seem to suggest that these genetic types are not only genetically, but also possibly morphologically closer to their Pink Type sister than the Type IIa. However, the data available for the genetic types within the *G. ruber* sensu stricto clade are too few to warrant a taxonomic revision. The large genetic and morphological separation of Type IIa, on the other hand, could be projected onto a new taxonomic concept. Our analyses reveal that individuals labelled as *G. ruber* s.l. in the literature cannot be separated in the analysed variables from the specimens of Type IIa and *G. ruber* white from the sediment sample. All other groups are significantly different from the *G. ruber* s.l. images (Fig. 3). Thus, the genetically defined Type IIa seems to correspond to the morphological concept of *G. ruber* s.l., as also suggested in the pilot study by Kuroyanagi et al. (2008).

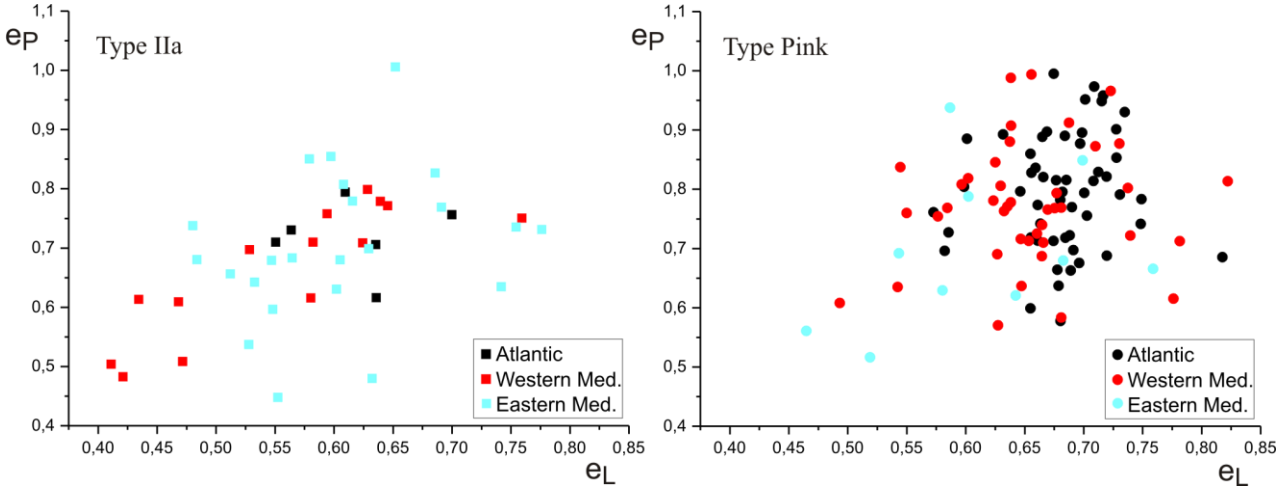

Figure 5 Chamber compression ratios of individuals of Type IIa and Type Pink.

Individuals are separated by the provinces they were sampled from, Eastern Atlantic Ocean, Western and Eastern Mediterranean Sea. The direct comparison of Type IIa and Type Pink in a combined graph is shown in Fig. 3.

Of all the various species and subspecies of *Globigerinoides* described in the literature, the d'Orbigny's species *G. elongatus* is the only one allied with the *G. ruber* - *G. conglobatus* clade, which is defined clearly by the compression of the last chambers (Banner and Blow, 1960; Cordey, 1967; Perconig, 1969). We have thus tested the hypothesis that specimens of Type IIa – *G. ruber* s.l. are consistent with the species concept of *G. elongatus* by analysing museum material from the NMNH in Washington, D.C. This analysis (Fig. 3) reveals that the species concept of *G. elongatus* was indeed consistently used for specimens with compressed chambers that overlap with *G. ruber* s.l. and are distinct from *G. ruber* (both pink and white chromotype). It is important to underline that the museum material was selected randomly, irrespective of what our current taxonomic opinion would have been. Thus, some of the specimens labelled as *G. ruber* in the collections did show a degree of chamber compression more consistent with what we have observed in Type IIa (i.e., *G. elongatus*) and vice versa. Nevertheless, the clear separation seen in Fig. 3 bears witness to the fact that the dominant usage of the taxonomy has been consistent with the critical role of chamber compression in the species concept.

Phylogeny and molecular divergence time estimates

The molecular phylogeny of the *Globigerinoides* clade (Fig. 4), when combined with data from the fossil record on the occurrence of the individual species (Fig. 6), allows a re-interpretation of the origin of both the *G. ruber* s.str. clade and the *G. conglobatus* – *G. elongatus* clade. The critical piece of evidence is here provided by the molecular clock analysis (Fig. 6; Table 6). Divergence time estimates for the *Globigerinoides* cluster have been performed previously by de Vargas et al. (1997) and Darling and al. (1999), resulting in two very different models. While the time estimates generated by de Vargas et al. (1997) are comparable with our results, the node ages from Darling et al. (1999) are much older. The split between the *G. ruber* s.str. lineage (Types Pink, Ia and Ib) and *G. conglobatus* (together with Type IIa), for example, was estimated at 22 Ma (in comparison to 12-15 Ma in our analysis), the split between Type Pink and Type I (a + b) was dated at 11 Ma (in comparison to 3-5 Ma in our analysis). We speculate that this deviation is caused, at least partly, by Darling et al. (1999) calibrating the split among the *O. universa* genotypes to the FAD of *O. universa* at 16.4 Ma. This age can reasonably be considered the maximum age for the split of the genetic types within *Orbulina* and thus the resulting estimates for the *Globigerinoides* events must also be considered maximum estimates. Further, and this applies to the results form de Vargas et al. (1997) as well, the authors used a highly truncated alignment (~ 540 bp from over 1000 bp), resulting in a reduced signal strength in the phylogenetic resolution on the species level (Aurahs et al. 2009b). Moreover, the sets of SSU rDNA sequences used by de Vargas et al. (1997) and Darling et al. (1999) are not identical to ours: SSU rDNA sequence alignment used by de Vargas et al. (1997) is missing Type IIa and IIb (in Darling et al., 1999 "*G. ruber* California Bight"). Darling et al. (1999) excluded the sequence of *G. sacculifer* from their alignment, calibrating instead the split of the genetic types of *O. universa* with the FAD of this species.

The molecular clock analyses in this study suggest that the *G. ruber* s.str. clade diverged from the *G. conglobatus – G. elongatus* clade in the middle Miocene. Most molecular clock dates cluster between 18 and 12 Ma, with an average of 15 Ma (Fig. 6; Table 6). It has been suggested that molecular clocks tend to overestimate the divergence ages (e.g. Rodriguez-Trelles et al., 2002), which would make the younger part of the range for this split more likely. The distribution of *G. ruber* in the fossil record shows a remarkable feature in the late Miocene around 8 Ma, when this species apparently disappears from the fossil record. This phenomenon has been termed "pseudo-extinction" (e.g. Liska, 1985) and is not known from any other planktonic foraminifera. The treatment of this event as "pseudo-extinction" follows from the phylogenetic hypothesis that *G. ruber* originated in the earliest Miocene at 20-22 Ma, consistent with the usage of the species name as recorded in the CHRONOS database (Fig. 6). Alternatively, the early Miocene specimens with a similar morphology have been labelled as *G. subquadratus* (Brönniman, 1954). Analysis of the CHRONOS database indicates that researchers have used either name, with a slight preference for *G. ruber*. The identical shape of the occurrence frequency curves for both species in the early Miocene (Fig. 6) indicates that the usage of these names was arbitrary.

A radically different phylogenetic hypothesis is that of Cordey (1967), who considered the "pseudo-extinction" as a real extinction of a dominantly early Miocene species and derived the post-extinction *G. ruber* from *G. obliquus*. The calculated divergence age clearly supports Cordey's concept. If *G. ruber* s.str. evolved from *G. obliquus*, then the divergence between this clade and the *G. conglobatus* clade would be manifested as the FAD of *G. extremus*, the ancestor of *G. conglobatus* and *G. elongatus* (e.g. Kenneth and Srinivasan, 1983; Fig. 6). This event is commonly dated at ~ 8 Ma (Berggren et al. 1995), although analysis of the CHRONOS database reveals numerous occurrences dated as early as 10-12 Ma and Perconig (1969) also reported an earlier age for this divergence (Fig. 6). Only one of the 64 molecular clock estimates dates this split as older than 20 Ma. In addition, if the *G. ruber* lineage continued into the early Miocene, it, too, would have to diverge from *G. obliquus* to share a common ancestor with the *G. conglobatus* clade, as suggested by all molecular analyses (e.g. de Vargas et al., 1997, Aurahs et al., 2009b). However, the early Miocene form of *G. ruber – G. subquadratus* has been derived in the literature from the unrelated *Globoturborotalita brazieri* (Srinivasan and Kennet, 1981b). We thus conclude that the late Miocene *G. ruber* "pseudo-extinction" marks in fact the first appearance of the present-day *G. ruber* s.str. clade and all records of *G. ruber* older than 8 Ma are referring to an unrelated lineage.

The divergence between *G. conglobatus* and the clade containing type IIa (*G. elongatus*) should correspond to the FAD of *G. conglobatus*, if, as is commonly accepted, both clades originated from *G. extremus* (see Perconig, 1969, Kennett and Srinivasan, 1983). In this phylogeny, the entire clade, including the ancestor *G. extremus* is characterised by chamber compression in the last whorl (Fig. 6).

The estimated divergence times from the molecular clock (trials without FAD *G. conglobatus* as a prior) are consistent with the observed FAD of *G. conglobatus*. Interestingly, the calculated mean divergence age from all trials where the node was not fixed (8.3 Ma) is remarkably close to the older age estimate of 8.3 Ma (Fig. 4). Thus, the fossil phylogeny is supported by the molecular clock analysis. The divergence age estimates between Type IIa and Type IIb range around 3-4 Ma and are remarkably similar with the observed FAD of *G. elongatus* in the CHRONOS database at 4 Ma (Fig. 6). This coincidence could indicate that the divergence between the sister Types IIa and IIb may have been associated with morphological divergence that lead to the erection of the species *G. elongatus*. The temporal coincidence between the molecular clock and the records in the CHRONOS database (extracted from the original species designations) indicate that the species concept of *G. elongatus* has been consistently and meaningfully applied, before it was abandoned.

Finally, the divergence between *G. ruber* Pink and Types I was dated to the late Miocene – early Pliocene. With an average age of 6.4 Ma, this divergence is remarkably close to the first appearance of *G. ruber* (after the "pseudo-extinction"), suggesting that the lineage which gave rise to the *G. ruber* s.str. clade radiated rapidly after its divergence from its ancestor *G. obliquus*. It could be that the extinction of the early-middle Miocene form (*G. subquadratus*) at 8 Ma left an empty niche which permitted a rapid radiation of an unrelated but ecologically similar form. Interestingly, the molecular clock estimate also suggests that the lineage leading to the Pink genetic type was separated from the *G. ruber* s.str. "white" genotypes much earlier than the first occurrence of the pink pigmentation in the sedimentary record would suggest (e.g. Thompson et al., 1979). These results not only justify the separate treatment of the two chromotypes for paleonevironmental reconstructions, they also indicate that paleoenvironmental reconstructions based on *G. ruber* prior to the first occurrence of the pink colouration have been integrating two distinct lineages. The further divergence within the *G. ruber* s.str. clade, dated into the early Quaternary, could be taken to indicate that such diversification has been a common phenomenon in the *Globigerinoides* clade and that an unknown number of more-or-less cryptic types may have existed and became extinct.

Our analysis also delivers some information on the possible age of the last common ancestor of the two main extant lineages of the genus *Globigerinoides*. The mean age of the split between the *G. ruber* and *G. trilobus* clades in the trials where the age of the root was not fixed is 25.5 Ma (Fig. 6, Table 6). We note that there is considerable scatter among the trials, but even the oldest estimate places the divergence into the Oligocene. The molecular clock thus supports our original assumption of their late Oligocene divergence from a common ancestor (possibly *G. primordius*), manifested either as FAD *G. trilobus* or *G. obliquus* (Fig. 6).

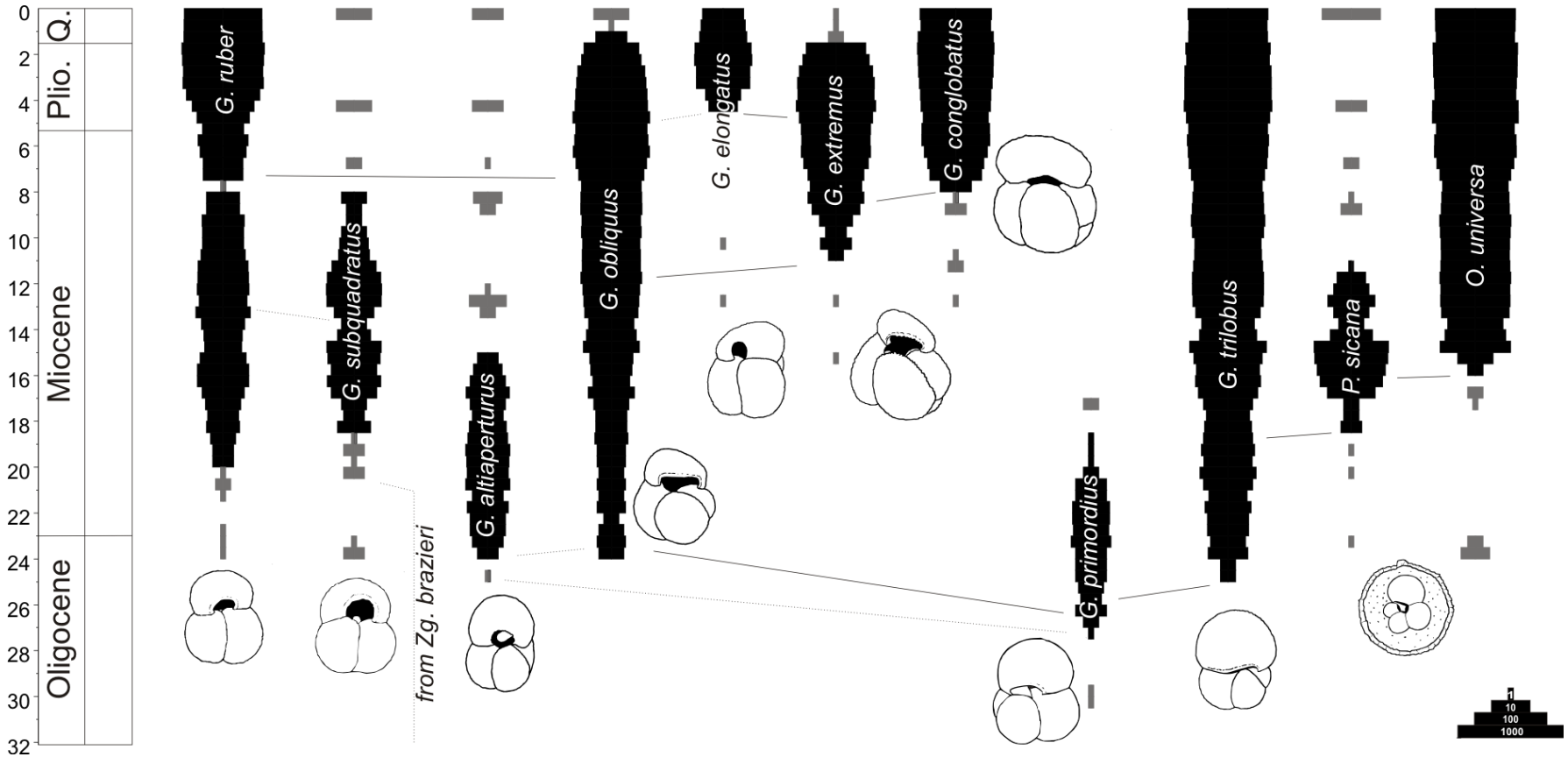

Figure 6 Stratigraphic distribution of species of the genus *Globigerinoidae* from the late Oligocene to recent. Frequency distribution of the species are on logarithmic scale and indicate numbers of reported findings per 0.5 Ma in the CHRONOS database (see material and methods section), black bars indicate continuous findings; grey bars indicate rare and scattered findings. The scheme is a combination of three different hypotheses on fossil phylogeny of *G. ruber* and *G. elongatus*. Kenneth and Srinivasan (1983) see *G. subquadratus* as descendant of *Globoturborotalita brazieri*, and postulate that *G. ruber* originates from *G. subquadratus* in the mid Miocene. Cordey (1967) prefers a late Miocene origin of *G. ruber* from *G. obliquus*. The origin of *G. elongatus* from the *G. obliquus* – *G. extremus* lineage is drawn after Perconig (1969). The continuous lines represent our preferred phylogenetic hypothesis, a combination of Perconig (1969), Cordey (1967) and the results from the molecular clock approach (see Table 6). Alternative relationships are indicated by dotted lines.

Conclusion

This work provides additional evidence for the species status of several SSU rDNA genotypes found in *Globigerinoides ruber*. We show that the genetic Type IIa could be morphologically discriminated from the genetic Type Pink and most likely also the *G. ruber* s.str. 'white' genotypes. The morphology and phylogeny of Type IIa is in all respects consistent with the species concept of *G. elongatus* and its usage in the micropaleontological community and corresponds to the informally defined morphotype *G. ruber* s.l. (Wang, 2000). We thus recommend that *G. elongatus* should be reinstated as a distinct extant species of planktonic foraminifera. By combining data from the fossil record with molecular phylogeny and molecular clock, we were able to show that the *G. ruber* lineage originated and diversified in the late Miocene after the *G. ruber* "pseudo-extinction" and that all earlier records of *G. ruber* refer to a different lineage (*G. subquadratus*).

We further show that the split between the *G. ruber* 'pink' and *G. ruber* 'white' s.str. genotypes is ancient and occurred shortly after the first appearance of *G. ruber* after the pseudo-extinction. Our observations support the current practice of treating the two chromotypes of *G. ruber* separately. They also indicate that proxies based on *G. ruber* in sediments pre-dating the first appearance of the pink pigment in the sediment have been potentially amalgamating specimens belonging to two distinct lineages. The taxonomic concept of *G. ruber* s.str. requires taxonomic revision as well, with the name *G. ruber* reserved to the Pink genotype. This study does not provide sufficient data to allow a morphological separation between the Pink and white Types I of *G. ruber* s.str. A separation based on shell pigmentation alone would lead to the unprecedented situation in the taxonomy of planktonic foraminifera where a species-level character can only be applied for a limited time span of the species.

Literature

Anand, P., Elderfield, H., Conte, M. H. 2003. Calibration of Mg/Ca thermometry in planktonic foraminifera from a sediment trap time series. Paleoceanography, 18, 1050.

Aurahs, R., Grimm, G. W., Hemleben, V., Hemleben, C., Kucera, M. 2009. Geographical distribution of cryptic genetic types in the planktonic foraminifer *Globigerinoides ruber*. Mol. Ecol. 18, 1692 – 1706.

Aurahs, R., Göker, M.,. Grimm, G.W., Hemleben, V., Hemleben C., Schiebel R., Kučera, M. 2009. Using the Multiple Analysis Approach to Reconstruct Phylogenetic Relationships among Planktonic Foraminifera from Highly Divergent and Length-polymorphic SSU rDNA Sequences. Bioinformatics and Biology Insights. 3, 155–177.

Banner, F. T., Blow, W. H., 1960. Some primary types of species belonging to the superfamily Globigerinaceae. Contr. Cushman Found. Foraminiferal Res. 11, 1-41.

Berggren, W. A., Hilgen, F. J., Langereis, C. G., Kent, D. V., Obradovich, J. D., Raffi, I., Raymo M. E., Shackleton N. J. 1995. Late Neogene chronology: New perspectives in high-resolution stratigraphy Geol. Soc. Am. Bul. 107, 1272-1287;

Brummer, G.J.A., Hemleben, C., Spindler, M. 1987. Ontogeny of extant spinose planktonic foraminifera (Globigerinidae) – a concept exemplified by *Globigerinoides sacculifer* (Brady) and G. ruber (D'Orbigny). Mar. Micropaleontol. 12, 357-381.

Chaisson, W.P., Pearson, P.N., 1997. Planktonic foraminifer biostratigraphy at Site 925: Middle Miocene-Pleistocene. In: Shackleton, N.J., Curry, W.B., Richter, C., Bralower, T.J. (Eds.), Proceedings of the Ocean Drilling Program, Scientific Results, pp. 3-31.

Chiessi, C. M., Ulrich, S., Mulitza, S., Pätzold, J., Wefer, G., 2007. Signature of the Brazil-Malvinas Confluence (Argentine Basin) in the isotopic composition of planktonic foraminifera from surface sediments. Mar. Micropaleontol. 64, 52–66.

Cordey, W. G. 1967. The development of *Globigerinoides ruber* (D'Orbigny 1839) from the Miocene to recent. Paleontology 10, 647-659.

Cushman, J. A. 1914. A monograph of the Foraminifera of the North Pacific ocean. Part 4. Chilostomellidae Globigerinidae, Nummulitidae. Washington Smithsonian. Inst. U. S. Nation. Mus. Bull. 71, 1-46.

Cushman, J. A.1927. Some new genera of the Foraminifera. Contr. Cushman Lab. Foraminif. Res. Sharon Mass. 2, 77-81.

d'Orbigny, A.D., 1826. Tableau méthodique de la classe de céphalopodes. Annales Des Sciences Naturelles 14, 1–277.

d'Orbigny, A.D., 1839. Voyage dans l'Amerique Meriodionale. Strasbourg, France.

Darling, K.F., Wade, C.M., Kroon, D., Brown, A.J.L., 1997. Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. Mar. Micropaleontol. 30, 251-266.

Darling, K.F., Wade, C.M., Kroon, D., Brown, A.J.L., Bijma, J., 1999. The diversity and distribution of modern planktic foraminiferal small subunit ribosomal RNA genotypes and their potential as tracers of present and past ocean circulations. Paleoceanography 14, 3-12.

Darling, K.F., Kucera, M., Pudsey, C.J.,Wade, C.M., 2004. Molecular evidence links cryptic diversification in polar plankton to Quaternary climate dynamics. Proc. Natl. Acad. Sci. USA. 101, 7657–7662.

Darling, K.F., Kucera, M., Kroon, D., Wade, C.M., 2006. A resolution for the coiling direction paradox in Neogloboquadrina pachyderma. Paleoceanography 21 (2), PA2011.

de Vargas, C., Zaninetti, L., Hilbrecht, H., Pawlowski, J., 1997. Phylogeny and rates of molecular evolution of planktonic foraminifera: SSU rDNA sequences compared to the fossil record. J. Mol. Evol. 45, 285–294.

de Vargas, C., Renaud, S., Hilbrecht, H., Pawlowski, J., 2001. Pleistocene adaptive radiation in *Globorotalia truncatulinoides*: genetic, morphologic, and environmental evidence. Paleobiology 27, 104-125.

Drummond, A.J., Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.BMC Evolutionary Biology 7:214

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucl. Acid. Res. 32, 1792-1797.

Hecht, A.D., 1974. Intraspecific variation in recent populations of Globigerinoides ruber and *Globigerinoides trilobus* and their application to paleoenvironmental analyses J. Paleontolo. 48, 1217-1234.

Huber, B.T., Bijma, J., Darling, K.F., 1997. Cryptic speciation in the living planktonic foraminifer *Globigerinella siphonifera* (d'Orbigny). Paleobiology 23, 33.62.

Hemleben, C., Spindler, M., Anderson, O.R., 1989. Modern Planktonic Foraminifera. Springer Verlag, New York, 363 pp.

Kahn, M.I., 1981. Ecological and paleo-ecological implications of the phenotypic variation in 3 species of living planktonic foraminifera from the northeastern Pacific Ocean 50 degrees North 145 degrees West. J. Foram Res. 11, 203-211.

Katoh, K., Kuma, K., Toh, H., Miyata, T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucl. Acid. Res. 33 511-518.

Kawahata,H., 2005. Stable isotopic composition of two morphotypes of *Globigerinoides ruber* (white) in the subtropical gyre in the north Pacific. Paleontological Research 9 (1), 27–35.

Kennett, J.P., 1976. Phenotypic variation in some Recent and late Cenozoic planktonic foraminifera. In: Hedley, R.H., Adams, C.G. (Eds.), Foraminifera, vol. 2. Academic Press, New York, pp. 111–170.

Kennett, J.P., Srinivasan, M.S., 1983. Neogene Planktonic Foraminifera: A Phylogenetic Atlas. Hutchinson Ross Publishing Co., Stroudsburg, PA, 265 pp.

Kucera, M., Schönfeld, J. 2007. The origin of modern oceanic foraminiferal faunas and Neogene climate change. In Deep-Time Perspectives on Climate Change: Marrying the Signal from Computer Models and Biological Proxies. Edited by Williams M, Haywood AM, Gregory FJ, Schmidt DN. London: The Geological Society. 2, 409–26.

Kuroyanagi, A., Tsuchiya, M., Kawahata, H., Kitazato, H., 2008. The occurrence of two genotypes of the planktonic foraminifer *Globigerinoides ruber* (white) and paleoenvironmental implications. Mar. Micropaleontol. 68, 236–243.

Larkin, M. A., Blackshields, G., Brown N.P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., Higgins, D. G. 2007. ClustalW and ClustalX version 2. Bioinformatics 23, 2947–8.

Lassmann, T., Sonnhammer, E.L. 2005 Kalign—an accurate and fast multiple sequence alignment algorithm. BMC Bioinformatics 6, 298.

Lea, D. W., Pak, D. K., Peterson, L. C., Hughen, K. A. 2003. Synchroneity of tropical and high-latitude Atlantic temperatures over the last glacial termination. Science 301, 1361-1364.

Liska, R. D., 1985. The range of *Globigerinoides ruber* (d'Orbigny) from the Middle to Late Miocene in Trinidad and Jamaica. Micropaleontology, 31, 372-379.

Loewemark, L., Hong, W.L., Yui, T.F., Hung, G.W., 2005. A test of different factors influencing the isotopic signal of planktonic foraminifera in surface sediments from the northern South China Sea. Mar. Micropaleontol. 55/1–2, 49–62.

Morard, R., Quillévéré, F., Escarguel, G., Ujiie, Y., de Garidel-Thoron, T., Norris, R. D., de Vargas, C. 2009. Morphological recognition of cryptic species in the planktonic foraminifer *Orbulina universa*. Mar. Micropaleontol. 71, 148-165.

Numberger, L., Hemleben, C., Hoffmann, R., Mackensen, A., Schulz H., Wunderlich, J.M., Kucera, M. 2009. Habitats, abundance patterns and isotopic signals of morphotypes of the planktonic foraminifer *Globigerinoides ruber* (d'Orbigny) in the eastern Mediterranean Sea since the Marine Isotopic Stage 12. Mar. Micropaleontol. 73, 90–104.

Perconig, E., 1969. Evolucion de los *Globigerinoides amplus*, *obliquus, extremus* y *elongatus* en el Neogeno de Andalucia (Espana). Revta. esp. Micropaleont. 1, 37-43.

Parker, F. L., 1962. Planktonic foraminiferal species in Pacific sediments. Micropaleontology 8, 219-254.

Rodríguez-Trelles, F., Tarrío R., Ayala, F. J. 2002. A methodological bias toward overestimation of molecular evolutionary time scales. Proc. Natl. Acad. Sci. USA. 99, 8112-8115.

Sadekov, A., Eggins, S.M., De Deckker, P., Kroon D., 2008. Uncertainties in seawater thermometry deriving from intratest and intertest Mg/Ca variability in *Globigerinoides ruber*. Paleoceanography. 23, PA1215.

Sadekov, A., Eggins, S. M., De Deckker, P., Ninnemann, U., Kuhnt, W., Bassinot, F. (2009), Surface and subsurface seawater temperature reconstruction using Mg/Ca microanalysis of planktonic

foraminifera *Globigerinoides ruber*, *Globigerinoides sacculifer*, and *Pulleniatina obliquiloculata*, Paleoceanography, 24.

Schmidt, G. A., Mulitza, S., 2002. Global calibration of ecological models for planktic foraminifera from coretop carbonate oxygen-18, Mar. Micropaleontol. 44, 125-140.

Srinivasan, M.S., Kennett, J.P., 1981. Neogene planktonic foraminiferal biostratigraphy and evolution: equatorial to subantarctic, South Pacific. Mar. Micropaleontol. 6, 499 - 533.

Srinivasan, M.S., Kennett, J.P. 1981. A review of Neogene planktonic foraminiferal biostratigraphy: applications in the equatorial and south Pacific. Soc. Eco. Paleontol. Mineral. Spec. Pub. Suppl. 32, 395-432

Steinke, S., Chiu, H.Y., Yu, P.S., Shen, C.C., Loewemark, L., Mii, H.S., Chen, M.T. 2005. Mg/Ca ratios of two *Globigerinoides ruber* (white) morphotypes: implications for reconstructing past tropical/subtropical surface water conditions. Geochemistry, Geophysics, Geosystems 6.

Thompson, P.R., Be, A.W.H., Duplessy, J.C., Shackleton, N.J. 1979. Disappearance of pink-pigmented *Globigerinoides ruber* at 120,000 yr BP in the Indian and Pacific Oceans. Nature 280, 554-558.

Tolderlund D. S., Bé A. W. H., 1971. Seasonal distribution of planktonic foraminifera in the western North Atlantic. Micropaleontology 17, 297-329.

Lin, H.L., Wang, W.C., Hung, G. W. 2004. Seasonal variation of planktonic foraminiferal isotopic composition from sediment traps in the South China Sea. Mar. Micropaleontol. 53, 447-460.

Wang, L.J., 2000. Isotopic signals in two morphotypes of *Globigerinoides ruber* (white) from the South China Sea: implications for monsoon climate change during the last glacial cycle, Palaeogeography, Palaeoclimatology, Palaeoecology 161 (3–4), pp. 381–394.

Zaric, S., Donner, B., Fischer, G., Mulitza, S., Wefer, G. 2005. Sensitivity of planktic foraminifera to sea surface temperature and export production as derived from sediment trap data. Mar. Micropaleontolo. 55, 75–105.