Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments

Béatrice Lecroq[a,b,1], Franck Lejzerowicz[a,1], Dipankar Bachar[c,d], Richard Christen[c,d], Philippe Esling[e], Loïc Baerlocher[f], Magne Østerås[f], Laurent Farinelli[f], and Jan Pawlowski[a,2]

[a]Department of Genetics and Evolution, University of Geneva, CH-1211 Geneva 4, Switzerland; [b]Institute of Biogeosciences, Japan Agency for Marine-Earth Science and Technology, Yokosuka 237-0061, Japan; [c]Centre National de la Recherche Scientifique, UMR 6543 and [d]Université de Nice-Sophia-Antipolis, Unité Mixte de Recherche 6543, Centre de Biochimie, Faculté des Sciences, F06108 Nice, France; [e]Institut de Recherche et Coordination Acoustique/Musique, 75004 Paris, France; and [f]FASTERIS SA, 1228 Plan-les-Ouates, Switzerland

Edited* by James P. Kennett, University of California, Santa Barbara, CA, and approved June 20, 2011 (received for review December 8, 2010)

Deep-sea floors represent one of the largest and most complex ecosystems on Earth but remain essentially unexplored. The vastness and remoteness of this ecosystem make deep-sea sampling difficult, hampering traditional taxonomic observations and diversity assessment. This problem is particularly true in the case of the deep-sea meiofauna, which largely comprises small-sized, fragile, and difficult-to-identify metazoans and protists. Here, we introduce an ultra-deep sequencing-based metagenetic approach to examine the richness of benthic foraminifera, a principal component of deep-sea meiofauna. We used Illumina sequencing technology to assess foraminiferal richness in 31 unsieved deep-sea sediment samples from five distinct oceanic regions. We sequenced an extremely short fragment (36 bases) of the small subunit ribosomal DNA hypervariable region 37f, which has been shown to accurately distinguish foraminiferal species. In total, we obtained 495,978 unique sequences that were grouped into 1,643 operational taxonomic units, of which about half (841) could be reliably assigned to foraminifera. The vast majority of the operational taxonomic units (nearly 90%) were either assigned to early (ancient) lineages of soft-walled, single-chambered (monothalamous) foraminifera or remained undetermined and yet possibly belong to unknown early lineages. Contrasting with the classical view of multichambered taxa dominating foraminiferal assemblages, our work reflects an unexpected diversity of monothalamous lineages that are as yet unknown using conventional micropaleontological observations. Although we can only speculate about their morphology, the immense richness of deep-sea phylotypes revealed by this study suggests that ultra-deep sequencing can improve understanding of deep-sea benthic diversity considered until now as unknowable based on a traditional taxonomic approach.

DNA barcoding | next-generation sequencing | small subunit ribosomal RNA | microbial eukaryote | cosmopolitanism

Deep-sea sediments are home for a wide range of small-sized metazoan and protistan taxa. The diversity of this meiofaunal community is difficult to estimate because its study suffers from undersampling, difficult access, and the problems involved in culturing deep-sea organisms. Additionally, most of the deep-sea species are tiny, fragile, and difficult to identify. Benthic foraminifera form one of the most abundant and diverse groups of deep-sea meiofauna, found even in the deepest ocean trenches (1). Particularly in abyssal areas, a large proportion of deep-sea foraminifera belongs to early lineages characterized by simple, single-chambered (monothalamous), organic-walled or agglutinated tests, which are poorly preserved in the fossil record (2). These early monothalamous lineages traditionally classified in orders Allogromiida and Astrorhiza have been proposed to form a large radiation in the Neoproterozoic, well before the first multichambered foraminifera appeared (3). However, the assessment of their diversity is hampered by the fragmentation of their delicate tests, a lack of distinctive morphological characters, and their unfamiliarity to meiofaunal workers, which means that they are often overlooked.

During the past decade, molecular studies revealed an astonishing diversity of early foraminifera (4), along with numerous descriptions of new deep-sea monothalamous species and genera (5). The sequences of early lineages were particularly abundant in environmental DNA surveys of marine (6), freshwater (7), and soil (8) ecosystems. Eight new family-rank clades branching at the base of the foraminiferal tree were distinguished in DNA extracts of deep Southern Ocean sediments (9). However, all these studies relied on clone libraries, limiting the number of sequences available for analysis.

By reducing cloning limitations, next-generation sequencing (NGS) methods profoundly altered our perception of microbial ecosystems (10), but only few studies focused on eukaryotes (11, 12). Until now, most environmental DNA sequence data were generated by using 454 technology (13). It is only recently that Illumina ultra-deep sequencing technology was used for environmental microbial diversity assessment (14, 15).

Here, we present a unique application of Illumina technology for the assessment of eukaryotic diversity. Taking advantage of exceptionally high divergence of some short hypervariable regions of foraminiferal small subunit (SSU) ribosomal RNA (rRNA) genes (16), we used the Illumina platform to examine foraminiferal species richness in deep-sea sediments. We massively sequenced the 36-bp-long fragment situated at foraminiferal specific helix 37f of the SSU ribosomal DNA (rDNA) for a set of 31 samples from the Arctic, North Atlantic, Southern, and Pacific Oceans and the Caribbean Sea, with the cultured species *Reticulomyxa filosa* used as a control. Using ultra-deep sequencing, the targeted diversity was thoroughly covered with the majority of obtained sequences assigned to early foraminifera, including monothalamids and undetermined basal lineages. Our study suggests that these inconspicuous simple foraminifera by far outnumbered the well-known multichambered species, challenging our current view of foraminiferal diversity.

Author contributions: B.L. and J.P. designed research; B.L., M.Ø., L.F., and J.P. performed research; ; P.E., M. Ø., and L.F. contributed new reagents/analytic tools; B.L., F.L., D.B., R.C., P.E., L.B., and J.P. analyzed data; and B.L., F.L., R.C., and J.P. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Data deposition: The sequence data reported in this paper have been deposited at the NCBI Sequence Read Archive.

[1]B.L. and F.L. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: jan.pawlowski@unige.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1018426108/-/DCSupplemental.

Results

Short Tags Analysis. In total we analyzed 78,613,888 reads of 36 nucleotides (nt) for 31 samples of surficial deep-sea sediment (0.5- or 5-mL volume; Table S1). After quality filtration, 41,534,552 reads were retained, with the percentage of filtered sequences ranging from 16.3% (SFA31) to 97.8% (SFA04). All identical reads were combined into unique sequence tags, of which the number ranged from 5,065 (SFA33) to 39,647 (SFA03). After removing singletons, the number of tags per sample averaged 6,549, with a number of reads per sample ranging from 404,160 (SFA31) to 3,392,323 (SFA03). For each sample, we reduced Illumina sequencing errors with an original approach combining a filter based on a second strict dereplication of tags and a one-pass clustering as explained in *Materials and Methods*. The number of OTUs per sample averaged 116 and ranged from 17 (SFA33) to 211 (SFA09), with a maximum number of reads for an OTU ranging from 35,207 (SFA22) to 1,244,352 (SFA17). Saturation curves (Fig. S1) clearly showed that most samples were thoroughly sequenced. In general, 20,000 filtered reads were sufficient to cover >95% of the OTUs' diversity present in a single sample.

One of the samples (SFA06) comprised only the cultured species *R. filosa*. The sequencing of this sample produced 2,416,756 reads, corresponding to 1,689 dereplicated tags with at least 2 reads per tag. After filtering and clustering, we recovered one OTU, which was identical to the 37f hypervariable sequence of *R. filosa* previously obtained by using classical Sanger technology, confirming the efficiency of our filtering approach.

Foraminiferal Richness. Each of the 3,609 OTUs obtained for the 31 samples was taxonomically assigned based on our refined database containing 1,048 reference sequences. The undetermined OTUs (UNK) that could not be placed confidently in any high-level taxonomic groups represented 47% of all OTUs, ranging from 23.5% (SFA33) to 57% (SFA18; Table S2). The remaining 1,611 OTUs were successively assigned to three taxonomic levels (*Materials and Methods*). At least one-third of the OTUs assigned at the highest taxonomic level (corresponding to the order) were placed in the group of monothalamous foraminifera (MON; Fig. 1*A*). Together with UNK, they formed >80% of OTUs in most of the samples. The multichambered orders were much less represented, with numbers of OTUs assigned to Textulariida (TEX) ranging from 0 (SFA33) to 15 (SFA29) and those assigned to Rotaliida (ROT) ranging from 1 (SFA33) to 24 (SFA29). The OTUs belonging to Miliolida (MIL), Spirillinida (SPI), Lagenida (LAG), and Robertinida (ROB) were represented by a single OTU at most. In many samples (18 of 31), we also found OTUs assigned to the planktonic order Globigerinida (GLO), but their number never exceeded 4 OTUs. The proportion of major groups was similar between the samples, except for the Pacific Ocean where GLO were particularly diverse.

In total, after combining all samples and average linkage clustering, the MON and UNK reached nearly 90% of all 1,643 OTUs, whereas the proportion of ROT, TEX, and GLO remained relatively low, at 6.5%, 4.2%, and 0.6%, respectively (Fig. 1*B*). In terms of reads, the proportion of UNK diminished (32%), whereas that of ROT and TEX increased. The proportion of reads assigned to MON remained close to 40% of all reads (Fig. 1*B*).

The proportion of OTUs assigned to the second taxonomic level was high, with up to 74.2% for MON, 57.5% for ROT, 62% for TEX, 66% for ROB, 90% for GLO, and 100% for SPI and LAG, although the last four orders were only represented by few OTUs (Fig. 1*C*). Conversely, the proportion of OTUs assigned to the third taxonomic level, or species level, was low, except for GLO (80%). In other groups, this depth of identification applied to 25% of the OTUs for both MON and ROT and 18% of the OTUs for TEX. Unresolved identification conflicts were re-

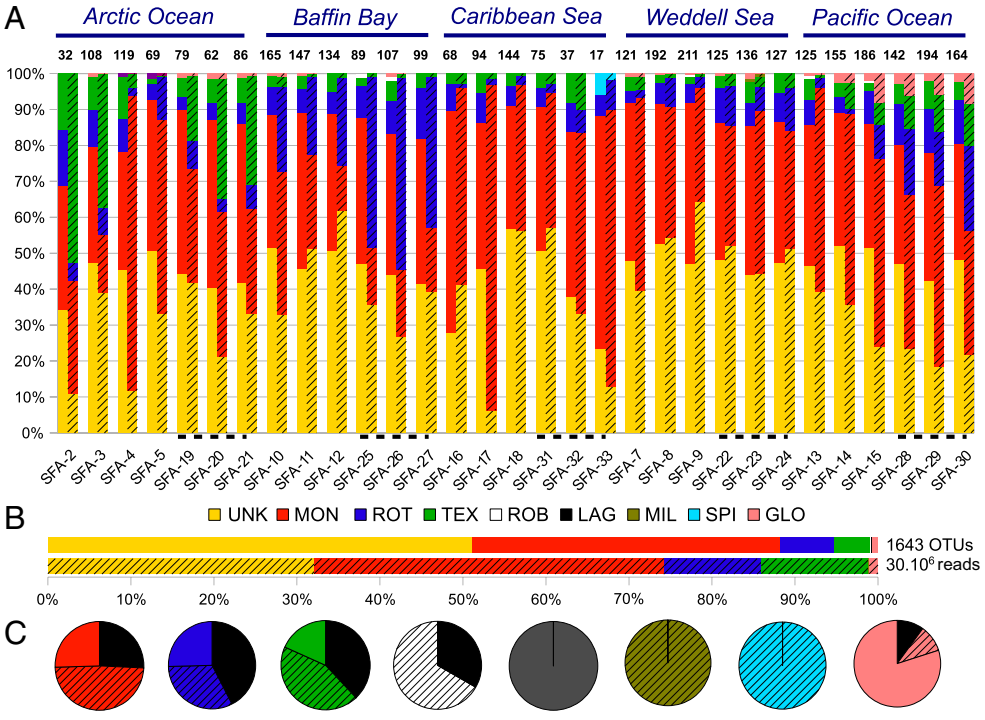

Fig. 1. Taxonomic composition of deep-sea Foraminifera assemblages based on microbarcode sequences. (*A*) Proportion of OTUs (solid bars) and reads (hatched bars) for deep-sea samples grouped according to their geographic origins; replicate samples are grouped above dashed lines. (*B*) Proportion of OTUs (solid bars) and reads (hatched bars) for the whole data set. (*C*) Proportion of OTUs identified at species (solid colored areas), family-clade (hatched colored areas), and order (solid black areas) level for each foraminiferal order. The numbers above or aside bars indicate the number of OTUs (*A* and *B*) or million reads (*B*). Colors correspond to foraminiferal orders: red, MON; green, ROT; dark blue, TEX; white, ROB; dark green, MIL; blue, SPI; pink, GLO.

sponsible for up to 2.2% (SFA 25) of OTUs assigned to the UNK category (0.6% on average). The OTUs that were downgraded to the first level because of conflicts at the second level and to the second level because of conflicts at the third level averaged 1.6% and 1.7%, respectively. Only 4 OTUs across all samples were identified at the order level because of conflicts at the third level.

As shown in the maximum likelihood (ML) tree based on sequences identified by microbarcodes (Fig. 2), the phylogenetic diversity of assigned OTUs is impressive. In addition to rotaliids and textulariids, almost all previously defined clades of monothalamous foraminifera, including well-supported environmental foraminifera (ENFOR) clades are represented in our samples. The number of OTUs assigned to some of these clades (clade G, ENFOR 5) reached the number of OTUs found in the order-level clades of TEX or ROT. The OTUs that could not be properly assigned to any monothalamous clade were combined in a group of "unresolved lineages." The large size of this group (200 OTUs) indicates that the diversity of monothalamiids is much higher than represented in the tree (Fig. 2).

Geographic Ranges. To test the possible extent of cosmopolitanism in the deep sea, we analyzed reads and OTUs found in the

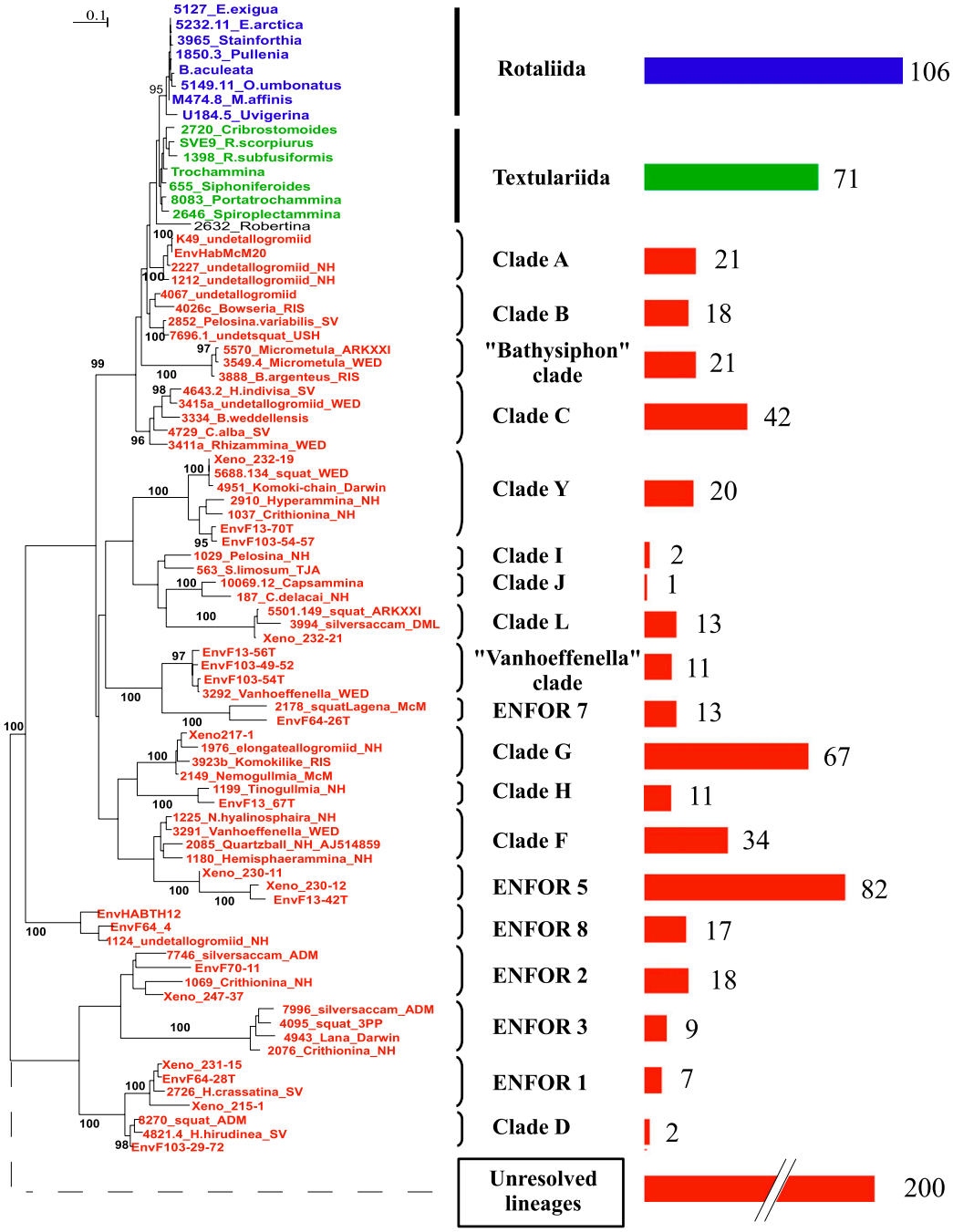

Fig. 2. Phylogenetic diversity of assigned OTUs based on ML analysis of corresponding partial SSU rDNA sequences. The unresolved lineages leaf includes monothalamids sequences that could not be confidently placed in any of illustrated clades. Horizontal bars represent the number of OTUs corresponding to each clade. Sequences identified up to the order-level only (MON) were not included. Bootstraps of 95% or more are indicated. Colors correspond to foraminiferal orders as in Fig. 1.

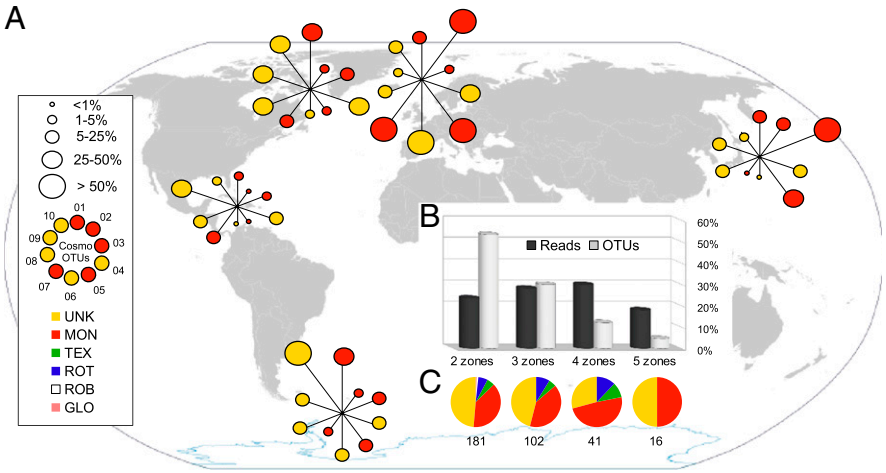

Fig. 3. Cosmopolitanism patterns in deep-sea foraminiferal OTUs. (*A*) Relative abundance of the 10 most sequenced CosmoOTUs across the five geographic zones. Proportion of reads is indicated by circle size. Each circle in a cluster corresponds to a CosmoOTU, in decreasing order of reads abundance (from 1 to 10). (*B*) Proportions of reads and OTUs exclusively recovered in two, three, four, or five different geographic zones. (*C*) Proportions of order-level taxa found in two, three, four, or five different geographic zones. Total numbers of OTUs are indicated. Colors correspond to foraminiferal orders as in Fig. 1.

five distinct geographic zones sampled. In total, 1,643 OTUs and 30,193,893 reads were examined (Table S3). The number of OTUs present in different zones decreased with the number of zones, from 181 occurring in two zones to 16 in five zones (Fig. 3*B*). The proportion of reads in cosmopolitan OTUs (CosmoOTUs) was relatively high (10% of all reads). Half of the CosmoOTUs were assigned to MON, whereas the remaining 8 OTUs could not be assigned (Table S4). Five cosmopolitan monothalamids, including the two most abundant ones, were assigned confidently to the third level, but none of them was characterized morphologically, except CosmoOTU 01, which was assigned to the *Vanhoeffenella* clade, a widely distributed deep-sea genus.

The relative abundance of the 10 most sequenced CosmoOTUs among geographical zones is shown in Fig. 3*A*. Six OTUs were abundantly (>50%) sequenced in one geographic zone (Table S4). Among them, 4 were most abundant in the Arctic Ocean, ranging from 60% to 84.3%. Two OTUs (CosmoOTU 08 and 11) were evenly distributed with no less than 10% and no more than one-third of the reads in a zone. The most abundantly sequenced cosmopolitan OTU (CosmoOTU 01), identified as *Vanhoeffenella*, was found in at least four samples per zone. Overall, Baffin Bay appeared as the most representative area for cosmopolitanism because reads belonging to the 16 cosmopolitan OTUs occurred in four samples of this zone on average.

There was no obvious trend between geographic dispersal and taxonomic composition at the order level (Fig. 3*C*), except for the absence of multichambered orders (ROT and TEX) among five zones' taxa. Within each geographic zone, station replicates displayed heterogeneous taxonomic compositions. Indeed, the most sequenced cosmopolitan OTU of a given geographic area was not necessarily the most sequenced OTU in the samples, except for *Vanhoeffenella* in Baffin Bay. Moreover, the number of CosmoOTUs detected in each sample varied greatly, as evidenced by SFA33, where no CosmoOTU was found.

Discussion

DNA Microbarcodes Offer High Taxonomic Resolution. By using a short hypervariable region of the SSU rRNA gene as foraminiferal barcode, we could reduce the size of a fragment necessary for species identification down to 36 nt, corresponding to the Illumina read length at the beginning of this study. Use of such microbarcodes was possible because foraminiferal rRNA genes evolve rapidly (17) and possess specific hypervariable regions that make it possible to distinguish species in the majority of taxonomic groups (16). Although the SSU rRNA genes evolve more slowly in other eukaryotic groups, for each of them it should be possible to find similar hypervariable regions in the internal transcribed spacer and large subunit rDNA, and therefore this approach can be applied more generally to assess eukaryotic diversity in group-specific studies.

Despite their short length, the taxonomic resolution of foraminiferal microbarcodes was relatively good. Approximately 50% of sequences could be assigned to the order-level taxa, and almost all of them could be reliably placed in the family or clade (second level in our analyses). Conversely, the number of sequences assigned to the species level (third level in our analyses) was relatively low. This result can be explained by the limited size of our reference database, which currently contains 1,048 unique microbarcode sequences. Indeed, the high proportion (80%) of sequences assigned to the species level in GLO is related to the fact that many modern globigerinid species were sequenced already (18). The remaining globigerinid sequences assigned only to the first or second level corresponded either to cryptic genetic types or tiny species that were not sequenced yet or to intragenomic polymorphisms that were not detected using the clonal approach. Because the taxonomic resolution of the analyzed hypervariable region is limited for some genera (16), our analyses most likely underestimate the true foraminiferal richness.

By using the Illumina system to sequence foraminiferal microbarcodes, we have reached a sequencing depth rarely approached before. More than 1 million reads were obtained for each sample, dramatically increasing the number of distinct foraminiferal tag sequences detected in deep-sea sediments. However, the origin of these sequences may be multiple. Some may correspond to the rare taxa, escaping detection by morphological or molecular cloning approaches. Some others may be produced by the amplification of extracellular DNA, abundant in the deep-sea sediments (19). Finally, some tags may result from the intragenomic diversity of rDNA copies reported in some foraminifera (20). To take into account such natural polymorphisms, we developed a very stringent filtering system removing the less abundant tags that could represent rare copies of rRNA genes as well as Illumina sequencing errors. This filter and its associated thresholds have been successfully ground truthed on our control sequencing (SFA6).

Cosmopolitan Taxa Are Widespread in the Deep Sea. Several deep-sea foraminiferal morphospecies are known to have wide geo-

graphic ranges (21), and some of them have been shown to be genetically identical across the global ocean (22, 23). Our study confirms this observation, showing that some OTUs were present in all sampling areas (Table S3). Many ubiquitous species could have been overlooked because of the small volume of analyzed samples. Nevertheless, the proportion of CosmoOTUs is probably not much higher, although this assumption should be tested by increased coverage of sampling area.

The CosmoOTUs did not always occur in all samples of a given area, which can be explained by patchiness of deep-sea species or the small sizes of our samples. High abundance of reads belonging to CosmoOTUs could be explained by the widespread occurrence of generalist species having large populations being more likely to be sampled. Such species are flourishing wherever the conditions are favorable but can be found virtually everywhere because of the global dispersal of their numerous specimens, propagules, or resting stages (24). However, the correlation between the number of reads and the amount of DNA in the sample must be carefully addressed because of the variable number of ribosomal copies in eukaryotic cells (25). As far as we are aware, this possibility still awaits experimental testing under NGS conditions.

Early Lineages Dominate Deep-Sea Foraminiferal Assemblage. Sequences assigned to early foraminiferal lineages grouped here in a paraphyletic assemblage of monothalamids by far outnumbered those assigned to multichambered rotaliids and textulariids, confirming previous environmental DNA surveys of foraminiferal diversity (6, 26). Moreover, we predict that the proportion of monothalamids is even greater because many UNK probably also belong to this group. This prediction is based on the fact that the rotaliids and textulariids have an easily recognizable, specific signature at the beginning of the 37f region (16), and all sequences having this signature were properly assigned to these groups. Some UNK could belong to the LAG, an order of calcareous foraminiferans known to be present in the deep sea, but the genetic diversity of which has not been probed yet. However, the variability of UNK was so large that it is highly improbable that all of them represent a single group.

If our prediction is correct, >80% of OTUs found in deep-sea sediments represent the early foraminiferal lineages. However, the identification of these lineages is not straightforward. Some of them could be assigned to previously established monothalamous clades (27), which comprise soft-walled allogromiids and agglutinated astrorhizids. Some others were assigned to the environmental clades (ENFOR 1–8; Fig. 2). These clades are composed almost exclusively of the sequences obtained in the previous environmental DNA surveys of foraminiferal diversity (4, 9). They also include the so-called hermit or squatter sequences that belong to species living inside or outside the empty tests of other foraminifera (28). These undetermined sequences are particularly abundant in DNA extractions of komokiaceans, xenophyophores, and some large astrorhizids, the tests of which create an ideal habitat for a rich microbial community (29).

We can only speculate about the possible morphology and ecology of members of these environmental clades. Probably, they are tiny amoeboid cells that remained undetected because of their small size, passing through the 63-μm sieve routinely used to study deep-sea foraminifera. They may be naked and dwelling in the interstitial water or living as parasites inside or outside other foraminiferans. Some of them could be similar to the new cercozoan species isolated from environmental samples (30). Characterization of environmental clades is needed, but we can already assume that our perception of foraminiferal diversity and understanding of their role in functioning of deep-sea ecosystems will profoundly change as result of this study.

Ultra-Deep Sequencing Offers a Powerful Tool for Exploring Deep-Sea Richness. There is an increasing body of information about the richness of macro- and megafaunal species living in and on deep-sea sediments (31, 32), but much less is known about the richness of benthic meiofaunal-sized metazoans (particularly nematodes and harpacticoid copepods) and protists. Our study shows that NGS provides an extremely powerful tool for investigating deep-sea meiofauna. Among the NGS technologies, Illumina ultra-deep sequencing seems particularly well adapted to group-specific studies. Like foraminifera, other meiofaunal taxa most likely also possess a short DNA region that can be used to distinguish between closely related species. The barcodes do not need to be as short as 36 bases, given that the length of Illumina reads is >100 bp in the latest version. It should be easy to adapt the existing barcodes, such as the V9 region of the SSU rRNA gene used for assessment of eukaryotic diversity (12).

The NGS technologies have potential to overcome the intrinsic difficulties of deep-sea research. They offer the capacity to process a higher number of samples, balancing the critical problem of chronic undersampling of deep-sea habitats. Moreover, they generate the minimum amount of sequences necessary for group-specific surveys, recovering a part of diversity considered to be unknowable (33). We believe that further optimization of experimental design and reference database enlargement will lead to broader applications of the NGS for biomonitoring and exploring evolution of the deep-sea environment.

Materials and Methods

Sampling. Thirty-one surface sediment samples (SFA2–5 and SFA7–33), for which depth, coordinates, and volume are presented in Table S5, were collected either with box corer or multicorer during RV *Polarstern* cruises ARK XXII-2 (Arctic Ocean, 2007) and ANTXXIV-2 (Southern Ocean, 2007–2008); RV *Merian* cruise MS MERIAN 09/02 (Baffin Bay, 2008); RV *Tansei Maru* KT07-14 (Pacific Ocean, 2007); and RV *Galathea 3* (Caribbean Sea, 2007). For each geographic region, at least three samples were replicates coming from the same station and from the same gear. One to 5 mL of sediment was taken from the upper layer (0- to 2-cm depth) of the multicore/boxcore samples and frozen at −20 °C immediately after collection.

DNA Extraction, Amplification, and Massive Sequencing. Metagenomic DNA extracts were obtained by using either the MO Bio PowerSoil kit for small deep-sea sediment samples (0.5 mL) or the MO Bio PowerMax Soil kit for bigger volumes (5 mL), both according to the protocol except for cell lysis, which was extended to 40 min. Extraction products were then stored at −20 °C. Additionally, cultured *R. filosa* DNA was extracted and processed as other environmental samples (SFA6). In the first step, a fragment of SSU rDNA (~400 bp) was amplified by PCR (15 cycles, 50 °C for annealing temperature) with a set of foraminiferal-specific primers (30). In the second step, PCR products were reamplified for additional 10 cycles (in a remote laboratory) and attached to the surface of the Solexa flow cell channels by adaptors. After solid-phase bridge amplification, the DNA colonies were sequenced on a Genome Analyzer GAII instrument (Illumina) for 36 cycles by using a Chrysalis 36 Cycles Version 2 kit.

Reads Filtering. Base calling was performed by using GAPipeline (Version 1.0; Illumina). Low-quality reads based both on single base score evaluations (reads with 1 base <10, fastq-solexa scoring scheme) and averaged scores throughout the entire lengths (quality value of <20) were removed. Reads with >30 identical bases as well as reads containing no undetermined base (N) were discarded. After strict dereplication, singletons were excluded. We developed a two-step filtering method to remove sequencing errors and to temper intragenomic polymorphisms. The first step was based on the assumptions that (*i*) Illumina quality scores decrease by the end of the sequencing-by-synthesis reaction and (*ii*) intraspecific variation occurs near the 3′ end rather than the 5′ end of the sequenced fragment (16). After the trimming of the six 3′ terminal bases and a strict dereplication on the 30 remaining bases, sequence tags with an unchanged number of occurrences as well as tags with a number of occurrences of <0.01% of the total number of reads in the sample were removed. The second step consisted of a one-pass clustering designed to reduce the noise due to (*i*) random errors and (*ii*) interoperons variations. Less abundant sequences were clustered with a more abundant sequence from which they derive, and according to the

analysis of both reference sequences and the deep sequencing of a single specimen (SFA6), we allowed up to four differences. Edit distances were calculated by using a global alignments procedure (Needleman–Wunsch algorithm; 3′ terminal gaps not counted as differences). Finally, clusters with number of occurrences of <0.01% of all reads kept after the first step were discarded. The most abundant sequence of a cluster was retained for downstream analyses and assigned the total number of occurrences found in the cluster.

Reads Identification. A curated database comprising 1,048 different 37f hypervariable regions of the foraminiferal SSU rDNA was built. Each sequence was extracted from the 3′ position off the sequencing primer and to a length between the natural minimum length of the region (19 nt) and up to 30 nt, which corresponded to the length of filtered reads. Each entry was annotated to three taxonomic levels. The first level corresponded to the order category of the morphology-based classification of Foraminifera, modified by combining Allogromiida and Astrorhizida into a group of monothalamids according to ref. 34. The second level corresponded to a family or clade defined by previous phylogenetic studies (9, 27). Finally, the third level corresponded either to the genus level or to the species level for well-described voucher specimens. The taxonomic resolution of the barcoding region was assessed, and identical sequences corresponding to different isolates of the same taxa were kept to analyze taxonomy conflicts (Fig. S2)

For each filtered OTU, the best global alignments with a reference sequence were searched by using a penalty of 1 for mismatch, gapopen, and gapextend when calculating the edit distance. We assigned sequences to the consensus of the third taxonomic level using successively alignments found at 0, 1, 2, and 3 differences over the 30 nt. Unassigned sequences were then assigned similarly to the second level but only with up to 2 differences over the first 20 nt. The remaining unassigned sequences were assigned to the first taxonomic level by progressively searching exact matches against the first 19–12 nt of reference sequences. Sequences involved in conflicts at a given level were assigned to the above level or to the UNK category for order-level conflicts.

OTU Delineation. Although OTUs were already delineated after the clustering embedded in the filtering procedure, these OTUs did not allow for a global estimation of the diversity across samples. Thus, a nonsupervised clustering of all filtered OTUs as well as reference sequences was conducted. A distance matrix was build by using pairwise distances calculated as described above, followed by average linkage at percentages of 80, 85, 90, 95, 96, 97, 98, and 99%. Analyses of clusters containing reference sequences involved in conflicts led us to choose the 96% average linkage. Within each cluster, the previous assignation of each sequence was used to assign the cluster to the deepest consensus taxonomic level when no reference sequence was included in a cluster. When a reference sequence was present, no conflict with previous assignments was found.

Rarefaction analyses were computed with a Python script by using the random module to randomize the reads. Curves were drawn for all samples separately or after combining by average linkage when pooling all samples.

Phylogenetic Reconstruction. A set of partial (853–1,317 nt) SSU rDNA sequences commonly used in foraminiferal phylogenies (9) was selected based on microbarcode identification. In total, 82 sequences were analyzed, including 8 rotaliids, 7 textulariids, and up to 5 sequences for each of the 20 well-defined clades of monothalamids. ClustalW alignment (35) was refined manually by using Seaview 4.0 (36). The ML tree was built by using RaxML (37), with the GTR+I+Γ substitution model and 100 bootstraps.

ACKNOWLEDGMENTS. We thank the Captain, officers, crew, and chief scientist of R/V *Polarstern* (ARK XXII/2 and ANTXXIV-2), R/V *Merian* (09/02), R/V *Tansei Maru* (KT07-14), and HDMS *Vædderen* (Galathea 3-Winmargin); Angelika Brandt, Michal Kucera, and Marit-Solveig Seidenkrantz for providing samples; Alexandra Weber, José Fahrni and Jackie Guiard for technical assistance; and Davor Trumbić for help with the analyses. This work was supported by Swiss National Science Foundation Grant 31003A-125372 and by a G. and L. Claraz donation. R. Christen acknowledges support of the Aquaparadox project financed by the Agence National de Recherche programme "Biodiversité" and the Pôle Mer PACA.

1. Todo Y, Kitazato H, Hashimoto J, Gooday AJ (2005) Simple foraminifera flourish at the ocean's deepest point. *Science* 307:689.
2. Gooday AJ (2002) Organic-walled allogromiids: aspects of their occurrence, diversity and ecology in marine habitats. *J Foraminiferal Res* 32:384–399.
3. Pawlowski J, et al. (2003) The evolution of early Foraminifera. *Proc Natl Acad Sci USA* 100:11494–11498.
4. Pawlowski J, Fahrni JF, Brykczynska U, Habura A, Bowser SS (2002) Molecular data reveal high taxonomic diversity of allogromiid Foraminifera in Explorers Cove (McMurdo Sound, Antarctica). *Polar Biol* 25:96–105.
5. Gooday AJ, Holzmann M, Guiard J, Cornelius N, Pawlowski J (2004) A new monothalamous foraminiferan from 1000 to 6300 m water depth in the Weddel Sea: Morphological and molecular characterisation. *Deep Sea Res Part II Top Stud Oceanogr* 51:1603–1616.
6. Habura A, Pawlowski J, Hanes SD, Bowser SS (2004) Unexpected foraminiferal diversity revealed by small-subunit rDNA analysis of Antarctic sediment. *J Eukaryot Microbiol* 51:173–179.
7. Holzmann M, Habura A, Giles H, Bowser SS, Pawlowski J (2003) Freshwater foraminiferans revealed by analysis of environmental DNA samples. *J Eukaryot Microbiol* 50:135–139.
8. Lejzerowicz F, Pawlowski J, Fraissinet-Tachet L, Marmeisse R (2010) Molecular evidence for widespread occurrence of Foraminifera in soils. *Environ Microbiol* 12: 2518–2526.
9. Pawlowski J, Fontaine D, da Silva AA, Guiard J (2010) Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. *Deep-Sea Res* II: in press.
10. Sogin ML, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci USA* 103:12115–12120.
11. Stoeck T, et al. (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol* 7:72–91.
12. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE* 4:e6372.
13. Edgcomb V, et al. (March 10, 2011) Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J*, 10.1038/ismej.2011.6.
14. Lazarevic V, et al. (2009) Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* 79:266–271.
15. Caporaso JG, et al. (June 3, 2010) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci USA* 108(Suppl 1):4516–4522.
16. Pawlowski J, Lecroq B (2010) Short rDNA barcodes for species identification in foraminifera. *J Eukaryot Microbiol* 57:197–205.
17. Pawlowski J, et al. (1997) Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. *Mol Biol Evol* 14:498–505.
18. Darling KF, Kucera M, Wade CM (2007) Global molecular phylogeography reveals persistent Arctic circumpolar isolation in a marine planktonic protist. *Proc Natl Acad Sci USA* 104:5002–5007.
19. Dell'Anno A, Danovaro R (2005) Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* 309:2179.
20. Pawlowski J (2000) Introduction to the molecular systematics of foraminifera. *Micropaleontology* 46:1–12.
21. Douglas RG, Woodruff F (1981) *The Sea, The Oceanic Lithosphere*, ed Emiliani C (Wiley, NY), pp 1233–1327.
22. Pawlowski J, et al. (2007) Bipolar gene flow in deep-sea benthic foraminifera. *Mol Ecol* 16:4089–4096.
23. Lecroq B, Gooday AJ, Pawlowski J (2009) Global genetic homogeneity in deep-sea foraminiferan *Epistominella exigua* (Rotaliida: Pseudoparrellidae). *Zootaxa* 2096:23–32.
24. Alve E, Goldstein ST (2010) Dispersal, survival and delayed growth of benthic foraminiferal propagules. *J Sea Res* 63:36–51.
25. Medinger R, et al. (2010) Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. *Mol Ecol* 19(Suppl 1):32–40.
26. Habura A, Goldstein ST, Broderick S, Bowser SS (2008) A bush, not a tree: The extraordinary diversity of cold-water basal foraminiferans extends to warm-water environments. *Limnol Oceanogr* 53:1339–1351.
27. Pawlowski J, et al. (2002) Phylogeny of allogromiid Foraminifera inferred from SSU rRNA gene sequences. *J Foraminiferal Res* 32:334–343.
28. Grimm GW, et al. (2007) Diversity of rDNA in Chilostomella: Molecular differentiation patterns and putative hermit types. *Mar Micropaleontol* 62:75–90.
29. Lecroq B, Gooday AJ, Cedhagen T, Sabbatini A, Pawlowski J (2010) Molecular analysis reveal high levels of eukaryotic richness associated with enigmatic deep-sea protists (Komokiacea). *Marine Biodiversity* 39:45–55.
30. Bass D, et al. (2009) Phylogeny of novel naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea revised. *Protist* 160:75–109.
31. Brandt A, et al. (2007) First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature* 447:307–311.
32. Ebbe B, et al. (2010) *Life in the World's Oceans: Diversity, Distribution and Abundance*, ed McIntyre AD (Blackwell Publishing, Oxford), pp 139–160.
33. Danovaro R, et al. (2010) Deep-sea biodiversity in the Mediterranean Sea: The known, the unknown, and the unknowable. *PLoS ONE* 5:e11832.
34. Pawlowski J (2009) *Encyclopedia of Microbiology*, ed Schaechter M (Elsevier, Oxford), pp 646–662.
35. Larkin MA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948.
36. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224.
37. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.