

# Analysis of expressed sequence tags from a naked foraminiferan *Reticulomyxa filosa*

Fabien Burki, Sergey I. Nikolaev, Ignacio Bolivar, Jackie Guiard, and Jan Pawlowski

**Abstract:** Foraminifers are a major component of modern marine ecosystems and one of the most important oceanic producers of calcium carbonate. They are a key phylogenetic group among amoeboid protists, but our knowledge of their genome is still mostly limited to a few conserved genes. Here, we report the first study of expressed genes by means of expressed sequence tag (EST) from the freshwater naked foraminiferan *Reticulomyxa filosa*. Cluster analysis of 1630 valid ESTs enabled the identification of 178 groups of related sequences and 871 singlets. Approximately 50% of the putative unique 1059 ESTs could be annotated using Blast searches against the protein database SwissProt + TrEMBL. The EST database described here is the first step towards gene discovery in Foraminifera and should provide the basis for new insights into the genomic and transcriptomic characteristics of these interesting but poorly understood protists.

**Key words:** Rhizaria, Foraminifera, cDNA library, annotation.

**Résumé :** Les Foraminifères sont un des composants majeurs des écosystèmes marins actuels de même qu'ils font parties des plus importants producteurs océaniques de carbonate de calcium. Ils représentent un groupe d'organismes phylogénétiquement très important parmi les protistes amoeboïdes mais notre connaissance de leur génome est pour l'instant limitée à quelques gènes très conservés. Dans cet article, nous rapportons la première étude de gènes exprimés du foraminifère d'eau douce et sans test *Reticulomyxa filosa*. Le regroupement de 1630 étiquettes de gènes exprimés (ESTs) a permis l'identification de 178 groupes de séquences similaires ainsi que de 871 séquences uniques. Environ 50% de ces 1059 séquences, supposées uniques, ont pu être annotées grâce à des recherches de similarité dans la banque de donnée protéique SwissProt + Trembl. La base de donnée décrite ici est une première étape dans la découverte de nouveaux gènes chez les foraminifères et devrait ouvrir de nouvelles perspectives dans la caractérisation du génome et du transcriptome de ces protistes intéressants mais peu connus.

**Mots clés :** Rhizaria, Foraminifera, librairie d'ADNc, annotation.

[Traduit par la Rédaction]

## Introduction

Foraminifers are a major component of modern marine ecosystems (Lee and Anderson 1991). Their cosmopolitan distribution ranges from polar shelves to tropical coral reefs. Planktonic foraminifers are an important and ubiquitous group of marine zooplankton (de Vargas et al. 1997). Benthic foraminifera are reported from marine habitats reaching from supralittoral sands and intertidal mudflats to the deepest abyssal trenches (Gooday 2002; Todo et al. 2005). They are highly diverse even in the most extreme polar environments (Habura et al. 2004; Pawlowski et al. 2002). They are also present in freshwater (Holzmann et al. 2003; Holzmann

and Pawlowski 2002), as well as in terrestrial habitats (Meisterfeld et al. 2001).

Foraminifers play an important role in biogeochemical cycles of inorganic and organic compounds (Sen Gupta 1999). Together with coral reef communities and pelagic microorganisms such as coccolithophores, foraminifers belong to the major oceanic producers of calcium carbonate. The production and burial of both organic carbonate and biogenic carbonate within the marine system provide a potential sink for carbon and therefore play a big part in the global carbon cycle (Barker et al. 2003).

Although the diversity and ecology of modern foraminifers are quite well characterized (Sen Gupta 1999), most as-

Received 16 January 2006. Accepted 31 March 2006 Published on the NRC Research Press Web site at <http://genome.nrc.ca> on 7 September 2006.

Corresponding Editor: B. Golding.

**F. Burki,<sup>1</sup> S.I. Nikolaev,<sup>2</sup> I. Bolivar, J. Guiard and J. Pawlowski.** University of Geneva, Department of Zoology and animal biology, 30 quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland.

<sup>1</sup>Corresponding author (e-mail: [Fabien.Burki@zoo.unige.ch](mailto:Fabien.Burki@zoo.unige.ch)).

<sup>2</sup>Current address: University of Geneva, Department of Genetic Medicine and Development, 1 rue Michel-Servet, CH-1211 Geneva, Switzerland.

pects of foraminiferan molecular biology are still poorly studied. Similarly, very little is known about other related amoeboid protists, which form together with foraminiferans the supergroup of Rhizaria (Nikolaev et al. 2004). To accelerate gene discovery and investigate evolution in Foraminifera as well as within other Rhizaria, we have conducted an expressed sequence tag (EST) project on the freshwater naked foraminiferan *Reticulomyxa filosa*.

Unlike other foraminifers, *R. filosa* cells grow rapidly in the laboratory, and simple culture conditions allow one to obtain adequate amounts of pure DNA and RNA. First isolated from decaying leaves in the New York City area (Nauss 1949), *R. filosa* was later rediscovered in a freshwater pond in Bochum, Germany (Hülsmann 1984) and in a freshwater fish tank in a laboratory in Berkeley, Calif. (Koonce and Schliwa 1985). However, it was only recently that this naked species, traditionally considered a group of testate protists (Pawlowski et al. 1999), was clearly identified as belonging to Foraminifera.

Prior to the initiation of this survey, the only molecular data available for the Foraminifera were LSU rRNA, SSU rRNA, and 6 protein coding genes (namely actin, RNA polymerase II,  $\alpha$ -,  $\beta$ -, and  $\gamma$ -tubulin, and ubiquitin) (Archibald and Keeling 2004; Flakowski et al. 2005; Habura et al. 2005; Longet et al. 2003; Pawlowski et al. 1996; Pawlowski et al. 1994). We sequenced about 1900 cDNA clones, resulting in over 1050 unique sequences and increasing the available molecular data for the group more than 100-fold, thus creating the first extensive high-throughput data set for the Foraminifera.

## Materials and methods

### Cells and culture conditions

The strain of *R. filosa*, obtained from Dr. R. Breuker (University of Bochum), was maintained using Volvic table water as a culture medium and fed with prewetted wheat germ flakes as food (Breuker 1997). Cells were inspected for purity by light microscopy, collected by low-speed centrifugation, resuspended into 5 volumes of TriReagent (Invitrogen, Carlsbad, Calif.), and broken using manual pestles and adapted microtubes.

### cDNA construction and ESTs sequencing

Following cell breakage, total RNA was extracted using the TriReagent manufacturer's protocol. Of this total RNA, 250 ng was used in a reverse transcription reaction with an oligo(dT) primer bearing a T7 promoter and in vitro transcription of the resulting DNA with T7 RNA polymerase (MessageAmp aRNA Kit, Ambion Inc., Austin, Tex.). The resulting antisense RNA (aRNA) copies were then converted into cDNA with the SuperScript Choice System for cDNA Synthesis (Invitrogen), generating double-stranded EcoRI-ended cDNAs. These were next amplified by PCR using the adapter sequence as primers and randomly ligated into the pCR 2.1-TOPO vector (Invitrogen). These constructs were directly used to transform bacteria, and this was followed by verification PCRs to check whether the clones were positive. EST sequencing was carried out with the ABI-PRISM Big Dye Terminator Cycle Sequencing Kit and analysed with an ABI-3100 DNA sequencer (Perkin-Elmer Inc., Wellesley, Mass.), all according to the manufacturer's instructions.

### Sequence processing and analysis

The ABI formatted chromatogram ESTs were processed automatically using a custom pipeline. This pipeline included base calling and quality control by PHRED (Ewing et al. 1998), followed by vector and adapter region trimming by CrossMatch (Green 1996). A final manual check eliminated remaining sequences shorter than 200 bp and removed poly(A) tails. The EST sequences were clustered and assembled into contigs using the Paracel Clustering Package (PCP) (Paracel Inc, Pasadena, Calif.). Clusters that contained only 1 sequence were classified as singlets. Altogether, contigs and singlets made up the dataset of putative unique ESTs (uniseqs).

All uniseqs were used to search the SwissProt + TrEMBL protein sequence dataset (swisstrem1) using the blastx algorithm with a cut off at  $1 \times 10^{-5}$ . For ESTs with no similarity found, putative open reading frames (ORF) were predicted with ESTScan (Iseli et al. 1999). To assign functions, uniseqs were also subjected to automatic annotation using AutoFACT (Koski et al. 2005). PHRED, CrossMatch, PCP, blast searches, and ESTScan were all run on the remote server Ludwig-sun2 at the Swiss Institute of Bioinformatics (Lausanne, Switzerland) (Falquet et al. 2003). AutoFACT was executed on the Vital-IT computational facilities at the Swiss Institute of Bioinformatics (Lausanne, Switzerland).

## Results and discussion

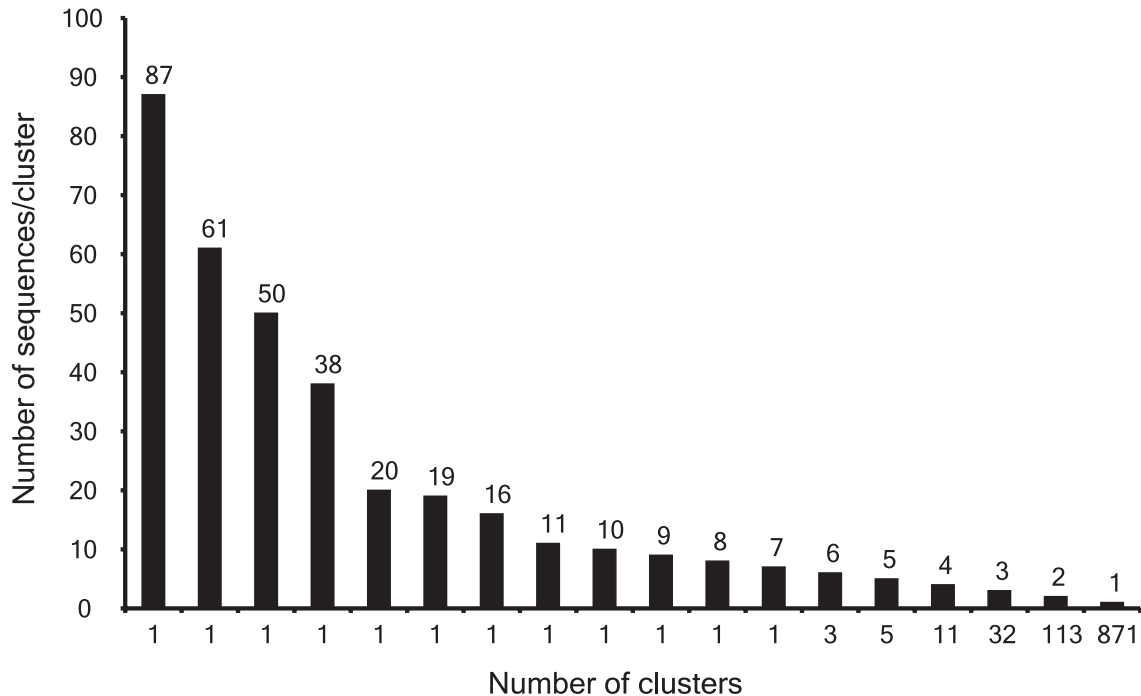
### Sequencing and clustering

A total RNA extraction from *R. filosa* was used in an mRNA amplification procedure to generate antisense RNA copies of each mRNA in the sample. This method was chosen over other nucleic acid amplification methods because it does not significantly distort the relative abundance of individual mRNA sequences within an RNA population. The antisense RNA copies were then converted into cDNA by priming random hexamers, which, in theory, are capable of binding throughout virtually any RNA template and have been shown to contain more 5' information than those primed with oligo(dT) (Ozkaynak et al. 1990).

After ligation and transformation of the cDNA, a total of 1908 clones were isolated and sequenced, leading to 1630 high-quality sequences passing quality checks and vector-trimming (see Material and Methods for more details). Within this EST collection, 759 sequences were assembled into 178 clusters, whereas 871 sequences did not find any homolog after pairwise comparisons (thus stayed as singlets). At the end of the clustering procedure, the CAP4 engine (available in the PCP, see Material and Methods) performed assembly on the clustered sequences by taking all the ESTs that are in a cluster and attempting to assemble them into a single contig representing a single transcript. As expected with ESTs binned according to local similarity (some sequences are occasionally not similar enough to be assembled together), 6 clusters generated more than 1 contig (from 2 to 4). Overall, singlets plus contigs gave a total of 1059 putative uniseqs.

Because of the use of random hexamers and the non-directional cloning strategy, this number of uniseqs is likely to be overrepresentative of the actual number of unique genes. Indeed, one expects to find the same gene in 2 or

**Fig. 1.** Frequency of occurrence of the different cluster sizes. The 871 clusters containing only 1 sequence correspond to singlets.



more clusters or singlets if 2 or more nonoverlapping ESTs belonging to the same transcript have been sequenced. To get a rough idea to what extent this may affect the clustering, it is worth mentioning that we observed only 15 groups of uniseqs, representing a total of 33 contigs or singlets (1 group with 4 contigs belonging to the same cluster, 1 group with 3 uniseqs, and 13 groups with 2 uniseqs whose 2 were 2 contigs from the same cluster), harboring the same best hit after a blastx search against the swisstrembl protein database. Although this result does not mean that some other ESTs with different hits may represent the same gene too, it is still a good indication of the number of unique genes in our dataset (likely above 1000).

From the distribution of the EST number in each cluster (Fig. 1), it is apparent that the majority of ESTs are present in low copy number. Most of the reads were singlets (871 sequences), and the biggest number of clusters (113) contained only 2 sequences. On the other hand, the larger cluster comprised 87 ESTs, and this was followed by the next largest clusters of 61, 50, and 38, respectively. The biggest cluster thus contained more than twice the number of sequences as the fourth cluster. Overall, the distribution of EST number per cluster shows that there was a low redundancy in our data, which makes our approach favourable for gene discovery.

#### Comparisons with databases

Each uniseq was searched (blastx) against the swisstrembl amino acid database. With the use of a threshold of  $1 \times 10^{-5}$ , 519 sequences (49%) matched a known protein in the database, whereas 540 (51%) lacked a similarity. This proportion of no hits is slightly above those recently published in other EST studies of various organisms, but it is still within the range typically expected when a large sample is acquired from a eukaryotic genome (Hackett et al. 2005; Jouannic et

al. 2005; Keon et al. 2005; Ribichich et al. 2005). The high percentage of sequences with no match probably reflects both the high genomic divergence of foraminifera (Pawlowski et al. 1997) and the early stage of rhizarian genome exploration (and thus the lack of sequences belonging to organisms related to *R. filosa*). To date, the only genome information available from another rhizarian is an EST dataset of *Bigelowiella natans* comprising about 3500 sequences (Keeling 2001). Sequence comparisons against this dataset using tblastx have revealed that only 233 of our 1059 uniseqs had a significant match (threshold  $\leq 1 \times 10^{-5}$ ), far fewer than the 519 hits obtained against swisstrembl (only 4 hits out of these 233 were new, i.e., corresponded to no similarity when the swisstrembl database was searched). This comparison should be interpreted with some caution since the *B. natans* dataset is very small and not exhaustive at all compared with swisstrembl. However, there are likely to be substantial differences between *R. filosa* and *B. natans* with respect to gene content and gene expression levels, despite their close evolutionary relationship. The phylogenetic analysis of ESTs from both species compared with other eukaryotes is published elsewhere (Burki and Pawlowski 2006). Interestingly, a tblastx search using the 540 sequences lacking a similarity against a homemade EST dataset (all EMBL + GenBank ESTs but without organisms from which whole genome has been made available) identified 31 more matches for those sequences that could not be assigned a similarity through searches of the swisstrembl database.

To further characterize these unknown sequences, we checked for putative ORFs to test whether they do not just represent untranslated regions of a transcript, rRNA contamination, or anything else of a noncoding nature that would prevent homology from being found in a protein database. Out of the 540 sequences that lacked a similarity only 39 had a predicted ORF shorter than 50 amino acids, and 111

**Table 1.** Top 20 hits of the *R. filosa* ESTs with the swisstrembl protein database.

Rank	Uniseq ID	Cut-off value	Accession No.	Protein description	Organism
1	cl136	0	Q820D0	Glutamine amidotransferase class-II;phosphoribosyl transferase	<i>Nitrosomonas europaea</i>
2	cl047	0	Q84VE1	Adenosylhomocysteinase-like protein	<i>Oryza sativa</i>
3	cl007	0	Q7XAS6	Pollen 2-phosphoglycerate dehydrogenase 2	<i>Cynodon dactylon</i>
4	cl039	0	Q8AVH9	Phosphoprotein phosphatase 2A-alpha catalytic chain	<i>Xenopus laevis</i>
5	cl003	0	Q9Y796	Glyceraldehyde-3-phosphate dehydrogenase	<i>Cryptococcus curvatus</i>
6	cl109	0	Q9ZSE4	Serine/threonine protein phosphatase PP2A catalytic subunit	<i>Hevea brasiliensis</i>
7	cl015	0	Q9Y018	Actin 1	<b><i>Reticulomyxa filosa</i>*</b>
8	cl006	0	Q9Y019	Actin 2	<b><i>Reticulomyxa filosa</i></b>
9	cl013	0	Q7KQK2	Polyubiquitin	<i>Plasmodium falciparum</i>
10	cl063	0	Q26233	Alpha-tubulin	<b><i>Reticulomyxa filosa</i></b>
11	jaR174	0	Q26235	Beta-tubulin	<b><i>Reticulomyxa filosa</i></b>
12	cl051	0	Q5DFR4	Hypothetical protein	<i>Schistosoma japonicum</i>
13	cl052	0	Q26236	Beta-tubulin	<b><i>Reticulomyxa filosa</i></b>
14	cl009	0	O44024	Alpha-tubulin 3	<b><i>Reticulomyxa filosa</i></b>
15	cl010	0	P56839	Phosphoenolpyruvate phosphomutase	<i>Mytilus edulis</i>
16	cl113	1e-94	Q6ZLZ9	Alpha tubulin	<i>Plasmodium falciparum</i>
17	cl025	2e-91	Q9ZSW1	Tubulin beta-1 chain	<i>Cyanophora paradoxa</i>
18	re407	1e-90	Q75JR8	Protein phosphatase 6 catalytic subunit	<i>Dictyostelium discoideum</i>
19	cl117	5e-89	Q7PY97	ENSANGP00000018457	<i>Anopheles gambiae s</i>
20	re549	3e-88	Q6WE52	Actin	<i>Thecamoeba similis</i>

\*Bolding indicates 6 matches corresponding to *R. filosa* TrEMBL entries.

(including the 39 ORFs just mentioned) had a predicted ORF shorter than 100 amino acids. Considering that this cut-off is often used to discriminate real genes from random ORFs, we can suppose that the remaining 429 ESTs are indeed coding regions that do not have similar known proteins available.

The top 20 most significant blastx hits against swisstrembl are shown in Table 1. We found here some classical, highly expressed housekeeping proteins (actins, polyubiquitin, tubulins) accounting for half of the 20 best hits. Importantly, 6 matches within the top 20 corresponded to *R. filosa* TrEMBL entries. In fact, we obtained an *R. filosa* hit in most of the cases for these very few *R. filosa* proteins already deposited in the public databases. When they are available, matching these few proteins belonging to the same organism is a good sign of the quality of our dataset.

Table 2 displays the identity of clusters made up of 10 or more ESTs. The largest cluster (87 ESTs) represents 1 of the 2 paralogues of foraminiferan actin gene (ACT2). The other most common transcripts (Table 2) were those encoding polyubiquitin (61 ESTs), actin 1 (50 ESTs), and alpha-tubulin (38 ESTs), suggesting that these are the most expressed genes in foraminifers. Note that no ribosomal protein appears in these most abundant ESTs. This is in contrast to what has been frequently reported for other analyses of this type, in which this particular class of gene can be seen to dominate frequency tables (Abu et al. 2004; Jouannic et al. 2005; Ribichich et al. 2005). Furthermore, 1 out of the 10 most abundant ESTs do not have any match to known proteins. Obviously, we cannot speculate on the function of the protein encoded by this unclassified cluster, but because of the high frequency of the ESTs belonging to the cluster they might be highly expressed genes that are either specific to foraminifers or are evolving sufficiently rapidly to be be-

**Table 2.** Transcript abundance as measured by EST redundancy.

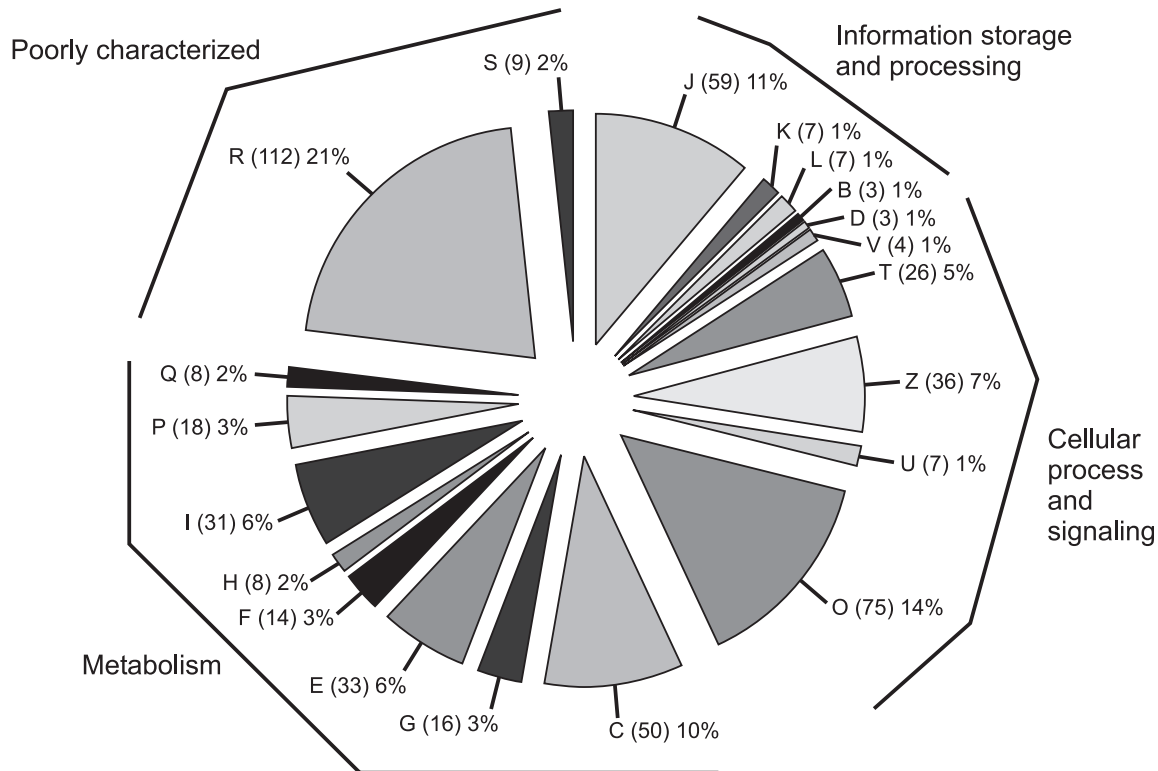
Rank	EST count	Identity of clusters
1	87	Actin 2
2	61	Polyubiquitin
3	50	Actin 1
4	38	Alpha-tubulin 3
5	20	Glyceraldehyde-3-phosphate dehydrogenase
6	19	No similarity
7	16	Putative cysteine protease
8	11	Putative cysteine protease
9	10	SFN protein

yond detection. Therefore, they would be interesting candidates for further investigation.

### Functional annotation

As mentioned, 49% (519/1059) of the total nonredundant ESTs (uniseqs) share significant similarities with SwissProt or TrEMBL entries. Putative function was examined by classifying uniseqs by cluster of orthologous group (COG) categories (Tatusov et al. 2003). They could be assigned for approximately the same number of ESTs (526) (Fig. 2), according to the automatic annotation protocol as implemented in AutoFACT (Koski et al. 2005). The ESTs were unequally distributed between the different functional categories. A large proportion were represented by a poorly annotated category, as 21% of ESTs fell into the class *General function prediction only*. The second largest class of proteins was related to protein modification and turnover (14%), and this was followed by translation (11%) and energy production (10%) categories. Many other classes are represented by fewer uniseqs. Globally, our dataset covers a broad range of

**Fig. 2.** Functional classification of the *R. filosa* ESTs based on COG categories. For each fraction, the label includes the COG category name (left), the number of clusters (centre), and the corresponding percentage of clusters. J, translation, ribosomal structure, and biogenesis; K, transcription; L, replication, recombination and repair; B, chromatin structure and dynamics; D, cell cycle control, cell division, and chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; Z, cytoskeleton; U, intracellular trafficking, secretion, and vesicular transport; O, Post-translational modification, protein turnover, and chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolic biosynthesis, transport and metabolism; R, general function prediction only; S, function unknown.



functionality since the annotation procedure classified the ESTs into 20 out of the 26 COG functional categories.

## Acknowledgements

The authors thank José Fahrni for technical assistance and sequencing; Lorenza Bordoli and Vassilios Ioannidis for help with programming; Laurent Falquet for giving permission to access the Ludwig server at the Swiss Institute of Bioinformatics; and Victor Jongeneel and Bruno Nyffeler for constructive discussions and help with the Vital-IT server. This research was supported by the Swiss National Science Foundation, grant 3100A0-100415 (J.P.) The ESTs published in this paper have been submitted to the GeneBank with the accession numbers EE664348-EE665977.

## References

- Abu, S.M., Li, G., and Asiegbu, F.O. 2004. Identification of Heterobasidion annosum (S-type) genes expressed during initial stages of conidiospore germination and under varying culture conditions. *FEMS Microbiol. Lett.* **233**: 205–213.
- Archibald, J.M., and Keeling, P.J. 2004. Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. *J. Eukaryot. Microbiol.* **51**: 113–118.
- Barker, S., Higgins, J., and Elderfield, H. 2003. The future of the

carbon cycle: review, calcification response, ballast and feedback on atmospheric CO<sub>2</sub>. *Phil. Trans. R. Soc. Lond.* **361**: 1977–1999.

- Breuker, R. 1997. Cytoskeletonkomponenten des plasmodialen Rhizopoden *Reticulomyxa filosa*. Dissertation. Ruhr-Universität, Bochum, Germany.
- Burki, F., and Pawlowski, J. 2006. Monophyly of Rhizaria and Multigene Phylogeny of Unicellular Bikonts. *Mol. Biol. Evol.* [E-pub ahead of print.]
- de Vargas, C., Zaninetti, L., Hilbrecht, H., and 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.
- Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**: 175–185.
- Falquet, L., Bordoli, L., Ioannidis, V., Pagni, M., and Jongeneel, C.V. 2003. Swiss EMBnet node web server. *Nucl. Acids Res.* **31**: 3782–3783.
- Flakowski, J., Bolivar, I., Fahrni, J., and Pawlowski, J. 2005. Actin phylogeny of Foraminifera. *J. Foramin. Res.* **35**: 93–102.
- Gooday, A. 2002. Organic-walled Allogromiids: aspects of their occurrence, diversity and ecology in marine habitats. *J. Foramin. Res.* **32**: 384–399.
- Green, P. 1996. PHRAP documentation [online]. Available from <http://www.phrap.org/phrap.docs/phrap.html>.
- Habura, A., Pawlowski, J., Hanes, S., and Bowser, S.S. 2004. Un-

- expected foraminiferal diversity revealed by small-subunit rDNA analysis of Antarctic sediment. *J. Eukaryot. Microbiol.* **51**: 173–179.
- Habura, A., Wegener, L., Travis, J.L., and Bowser, S.S. 2005. Structural and functional implications of an unusual foraminiferal {beta}-tubulin. *Mol. Biol. Evol.* **22**: 2000–2009.
- Hackett, J.D., Scheetz, T.E., Su Yoon, H., Soares, M.B., Bonaldo, M.F., Casavant, T.L., and Bhattacharya, D. 2005. Insights into a dinoflagellate genome through expressed sequence tag analysis. *BMC Genomics.* **6**: 80.
- Holzmann, M., and Pawlowski, J. 2002. Freshwater foraminiferans from Lake Geneva: past and present. *J. Foramin. Res.* **32**: 344–350.
- Holzmann, M., Habura, A., Giles, H., Bowser, S., and Pawlowski, J. 2003. Freshwater foraminiferans revealed by analysis of environmental DNA samples. *J. Eukaryot. Microbiol.* **50**: 135–139.
- Hülsmann, N. 1984. Biology of the genus *Reticulomyxa* (Rhizopoda). *J. Protozool.* **31**: 55A.
- Iseli, C., Jongeneel, C., and Bucher, P. 1999. ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* pp. 138–148.
- Jouannic, S., Argout, X., Lechavue, F., Fizames, C., Borgel, A., Morcillo, F., et al. 2005. Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*). *FEBS Lett.* **579**: 2709–2714.
- Keeling, P.J. 2001. Foraminifera and Cercozoa are related in actin phylogeny: two orphans find a home? *Mol. Biol. Evol.* **18**: 1551–1557.
- Keon, J., Antoniw, J., Rudd, J., Skinner, W., Hargreaves, J., and Hammond-Kosack, K. 2005. Analysis of expressed sequence tags from the wheat leaf blotch pathogen *Mycosphaerella graminicola* (anamorph *Septoria tritici*). *Fungal Genet. Biol.* **42**: 376–389.
- Koonce, M., and Schliwa, M. 1985. Bidirectional organelle transport can occur in cell processes that contain single microtubules. *J. Cell Biol.* **100**: 322–326.
- Koski, L., Gray, M., Lang, B.F., and Burger, G. 2005. AutoFACT: an automatic functional annotation and classification tool. *BMC Bioinformatics.* **6**: 151.
- Lee, J., and Anderson, O. 1991. *Biology of Foraminifera*. Academic Press, New York.
- Longet, D., Archibald, J.M., Keeling, P.J., and Pawlowski, J. 2003. Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. *Int. J. Syst. Evol. Microbiol.* **53**: 1735–1739.
- Meisterfeld, R., Holzmann, M., and Pawlowski, J. 2001. Morphological and molecular characterization of a new terrestrial allogromiid species: *Edaphoallogromia australica* gen. et spec. nov. (Foraminifera) from Northern Queensland (Australia). *Protist.* **152**: 185–192.
- Nauss, R. 1949. *Reticulomyxa filosa* gen. et sp. nov., a new primitive plasmodium. *Bull. Torrey Bot. Club.* **76**: 161–176.
- Nikolaev, S.I., Berney, C., Fahrni, J.F., Bolivar, I., Polet, S., Mylnikov, A.P., et al. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. *PNAS.* **101**: 8066–8071.
- Ozkaynak, E., Rueger, D., Drier, E., Corbett, C., Ridge, R., Sampath, T., and Oppermann, H. 1990. OP-1 cDNA encodes an osteogenic protein in the TGF-beta family. *EMBO J.* **9**: 2085–2093.
- Pawlowski, J., Bolivar, I., Guiard-Maffia, J., and Gouy, M. 1994. Phylogenetic position of foraminifera inferred from LSU rRNA gene sequences. *Mol. Biol. Evol.* **11**: 929–938.
- Pawlowski, J., Bolivar, I., Fahrni, J., Cavalier-Smith, T., and Gouy, M. 1996. Early origin of foraminifera suggested by SSU rRNA gene sequences. *Mol. Biol. Evol.* **13**: 445–450.
- Pawlowski, J., Bolivar, I., Fahrni, J., de Vargas, C., Gouy, M., and Zaninetti, L. 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.
- Pawlowski, J., Bolivar, I., Fahrni, J., De Vargas, C., and Bowser, S.S. 1999. Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. *J. Eukaryot. Microbiol.* **46**: 612–617.
- Pawlowski, J., Fahrni, J., Brykczynska, U., Habura, A., and Bowser, S.S. 2002. Molecular data reveal high taxonomic diversity of allogromiid foraminifera in Explorers Cove (McMurdo Sound, Antarctica). *Polar Biology.* **25**: 96–105.
- Ribichich, K., Salem-Izacc, S., Georg, R., Vencio, R., Navarro, L., and Gomes, S. 2005. Gene discovery and expression profile analysis through sequencing of expressed sequence tags from different developmental stages of the Chytridiomycete *Blastocladiella emersonii*. *Eukaryot. Cell.* **4**: 455–464.
- Sen Gupta, B. 1999. *Modern Foraminifera*. Kluwer, Dordrecht, The Netherlands.
- Tatusov, R., Fedorova, N., Jackson, J., Jacobs, A., Kiryutin, B., Koonin, E., et al. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics.* **4**: 41.
- Todo, Y., Kitazato, H., Hashimoto, J., and Gooday, A.J. 2005. Simple Foraminifera flourish at the ocean's deepest point. *Science (Washington, D.C.)*, **307**: 689.