Genetics and Molecular Biology, 35, 1 (suppl), 292-303 (2012)

www.sbg.org.br

Research Article

Metatranscriptomic analysis of small RNAs present in soybean deep sequencing libraries

Lorrayne Gomes Molina[1,2], Guilherme Cordenonsi da Fonseca[1], Guilherme Loss de Morais[1],
Luiz Felipe Valter de Oliveira[1,2], Joseane Biso de Carvalho[1], Franceli Rodrigues Kulcheski[1]
and Rogerio Margis[1,2,3]

[1]Centro de Biotecnologia e PPGBCM, Laboratório de Genomas e Populações de Plantas,
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.
[2]Programa de Pós-Graduação em Genética e Biologia Molecular,
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.
[3]Departamento de Biofísica, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

Abstract

A large number of small RNAs unrelated to the soybean genome were identified after deep sequencing of soybean small RNA libraries. A metatranscriptomic analysis was carried out to identify the origin of these sequences. Comparative analyses of small interference RNAs (siRNAs) present in samples collected in open areas corresponding to soybean field plantations and samples from soybean cultivated in greenhouses under a controlled environment were made. Different pathogenic, symbiotic and free-living organisms were identified from samples of both growth systems. They included viruses, bacteria and different groups of fungi. This approach can be useful not only to identify potentially unknown pathogens and pests, but also to understand the relations that soybean plants establish with microorganisms that may affect, directly or indirectly, plant health and crop production.

Key words: next generation sequencing, small RNA, siRNA, molecular markers.

Introduction

Until recently, analysis of the microbial diversity in environmental samples was conducted only after isolation, culture and identification of microorganisms and subsequent sequencing of cloned libraries (Cardenas and Tiedje, 2008). However, these conventional methods are limited to the minority of species that can be cultured (Chistoserdova, 2010). New culture-independent methods, such as metagenomics and metatranscriptomics, have been developed (Xu, 2006; Cardenas and Tiedje, 2008; Adams *et al.*, 2009; Warnecke and Hess, 2009; Chistoserdova, 2010). These methods refer to studies of the collective set of genomes and transcriptomes of mixed microbial communities and may be applied to the exploration of all microorganisms that reside in marine environments, soils, human and animal clinical samples, sludge, polluted environment, and plants (Kent *et al.*, 2007; Zoetendal *et al.*, 2008; Adams *et al.*, 2009; Al Rwahnih *et al.*, 2009; Poretsky *et al.*, 2009; Shi *et al.*, 2009; Desai *et al.*, 2010; Gifford *et al.*, 2010; Roossinck *et al.*, 2010). The metagenomic approaches, including metatranscriptomics, involve the sequencing of random DNA or RNA-derived complementary DNA (cDNA) profiles and subsequent determination of taxonomic diversity and prospective genes related to response to environmental conditions (Rosen *et al.*, 2009). With the advent of new sequencing technologies (high-throughput sequencing), more data can be generated in a relatively short time, in a practical and cost-effective way (Creer *et al.*, 2010). Moreover, this approach allows the direct sequencing of DNA or cDNA, avoiding any cloning bias and leading to large-scale studies (Adams *et al.*, 2009).

Metagenomic analysis of RNA deep-sequencing data has been used in plant disease diagnostics, such as in grapevine (Al Rwahnih *et al.*, 2009; Coetzee *et al.*, 2010), sweet potato (Kreuze *et al.*, 2009), tomato and *Liatris spicata* (Adams *et al.*, 2009). Coupling of metagenomics with pyrosequencing in these studies has also allowed the detection of bacterial and fungal RNA, suggesting that this approach can contribute to massive identification of crop-associated microbiota, including pathogenic, symbiotic, and free-living organisms.

Send correspondence to Rogerio Margis. Centro de Biotecnologia e PPGBCM, Laboratório de Genomas e Populações de Plantas, Prédio 43431, Universidade Federal do Rio Grande do Sul, Caixa Postal 15005, 91501-970 Porto Alegre, RS, Brazil. E-mail: rogerio.margis@ufrgs.br.

In addition to other sequencing methodologies, deep sequencing libraries of small RNA (sRNA) can contribute to microbial identification studies as they may include non-coding RNAs (*e.g.* rRNA), small regulatory RNAs, such as microbial sRNAs (Shi *et al.*, 2009) and host small interfering RNAs (Kreuze *et al.*, 2009), as well as mRNA fragments that normally would not be represented in libraries enriched for polyA-tailed mRNA commonly used in metatranscriptomic studies. As an example of the application of high-throughput sequencing of plant sRNAs, Kreuze *et al.* (2009), using short read sequences of approximately 24 base pairs (bp), successfully identified viruses infecting sweet potato, even those present in extremely low titer symptomless infections.

With this in mind, the current study describes the use of high-throughput sequencing of sRNAs to identify potential pathogens and other microorganisms from samples of soybeans grown in the field and in controlled-environment conditions.

Material and Methods

Plant material

The sRNA sequences analyzed in this study were obtained from deep-sequencing libraries from different projects related to the soybean transcriptome (Genosoja and Genoprot). The libraries were constructed from root samples of soybean cultivars grown under greenhouse conditions, and from flower, seed and pod samples from soybean plants grown under field conditions.

Root samples were obtained from 'Embrapa 48' and 'BR 16' soybean cultivars grown in a greenhouse at Embrapa-Soja in Londrina, Brazil. Plants were cultivated using a hydroponic system composed of plastic containers (30 liters) and an aerated 6.6 pH-balanced nutrient solution. Seeds were pre-germinated on moist filter paper in the dark at 25 °C ± 1 °C and 65% ± 5% rhr. Plantlets were then placed in polystyrene supports with the roots of the seedlings completely immersed in the nutrient solution. Each tray containing the seedlings was maintained in a greenhouse at 25 °C ± 2 °C and 60% ± 5% rh under natural daylight (photosynthetic photon flux density (PPFD) = 1.5 x 10^3 µmoles m^{-2} s^{-1}, equivalent to 8.93 x 10^4 lux) and a 12:12 h photoperiod. Roots of seedlings with the first trifoliate leaf fully developed (V2 developmental stage) were frozen in liquid nitrogen and stored at -80 °C until RNA extraction.

Three stages of seed germination, pods, and mature seeds from the soybean cultivar 'Conquista' were also used in RNA extraction. In a growth chamber, seeds were incubated for 3, 5 and 7 days on rolls of moistened filter paper at 26 °C. Pods (R3-R5) were collected from field plants grown at the Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, Brazil.

Flower samples (R2-R3) were collected from the soybean cultivar 'Urano' grown at the experimental field of the University of Passo Fundo (UPF), in Passo Fundo, Brazil. Collected flowers were immediately powdered in Trizol (Invitrogen, CA, USA) and stored until RNA extraction.

RNA extraction and sequencing

Total RNA was isolated from roots, seeds, seedlings, pods and flowers using Trizol (Invitrogen, CA, USA), and following the manufacturer's instructions. RNA quality was evaluated by electrophoresis in 1.0% agarose gels, and the amount checked using a Quibit fluorometer and Quant-iT RNA assay kit (Invitrogen, CA, USA) according to the manufacturer's instructions. Approximately 10 µg of total RNA were sent to Fasteris Life Sciences SA (Plan-les-Ouates, Switzerland) for processing and sequencing using Solexa technology on an Illumina Genome Analyzer GAII. Briefly, the processing by Illumina consisted of the following successive steps: acrylamide gel purification of RNA bands corresponding to the size range 20-30 nucleotides (nt), ligation of 3' and 5' adapters to the RNA in two separate subsequent steps, each followed by acrylamide gel purification, cDNA synthesis, and a final step of PCR amplification to generate a DNA colony template library for Illumina sequencing. After removing vector sequences, reads of 19 to 24 nt were used for further analysis.

Sequence analysis

The detailed analysis of endogenous sRNAs, including microRNAs, obtained from the soybean libraries described above is the topic of a separate study and will not be discussed here. However, since in addition a large number of sRNAs unrelated to the soybean genome were identified, a metagenomic analysis was indicated to identify the origin of these sequences. To this end, all reads were assembled into contigs using the Velvet 0.7.31 *de novo* assembly algorithm (Zerbino and Birney, 2008) with the following parameters: hash length of 23, coverage cut-off of 50, expected coverage of 1,000, and a minimum scaffold length of 100.

Assembled contigs matching the soybean genome were eliminated from further analysis using BLAST (BLASTn) "stand alone" version 2.2.20 against the soybean genome database available at Phytozome with the following parameters: expectation value (-e) of $1e^{-10}$, cost to open a gap (-G) of -6, cost to extend a gap (-E) of -6, penalty for a nucleotide mismatch (-q) of -5. The remaining contigs were used to search the NCBI database using BLASTn (nucleotide blast) with default parameters and an expectation value of 10^{-5}. The contigs were classified according to the sequence with the highest hit score found with BLASTn. The number of reads aligning to the contigs, coverage and average depth were determined with the SOAP tool (Li *et al.*, 2009). Default parameters were used and only filtered data (reads aligning to the references with a high confidence) are reported.

Results

Sequence analysis

The sRNA libraries analyzed in this study contained 5,627,802 reads (each consisting of 19-24 bp) from root, 8,610,347 from seeds and pods, and 9,314,206 from flower (Figure 1). The sRNA sequences were assembled into contigs ranging from 40 to 300 nucleotides, approximately. Sequence assembly produced 2,646, 15,521 and 28,382 contigs from root, seed and pod, and flower, respectively. After elimination of the soybean sequences, 253 (root), 2,574 (flower) and 1,959 (seed and pod) contigs remained for further analysis. These contigs were used in BLASTn searches against the NCBI database.

After BLASTn annotation, a large number of sequences remained unidentified (Figure 1), accounting for 73.4% of the field samples (Figure 2A) and 37.2% of the controlled environmental samples (Figure 3A). Contigs that corresponded to soybean sequences, but could not be filtered by the local BLASTn, represented 17.4% of the total contigs from field samples (Figure 2A) and 5.1% from controlled environment samples (Figure 3A). There were also contigs corresponding to sequences from other plant species deposited in NCBI (Figures 2A and 3A).

Apart from the contigs mentioned above, 134 contigs from the controlled environment (root) and 335 from the field (flower, seed and pod) had hits to previously sequenced microorganisms and viruses (Figure 1) and provided the results shown in the following topics. These contigs were distributed in different taxonomic groups based on their best BLASTn hits (Figures 2B and 3B). Contigs showing similarities to sequences from different taxonomic groups (multiple affiliations) were classified at the taxonomic level immediately above.

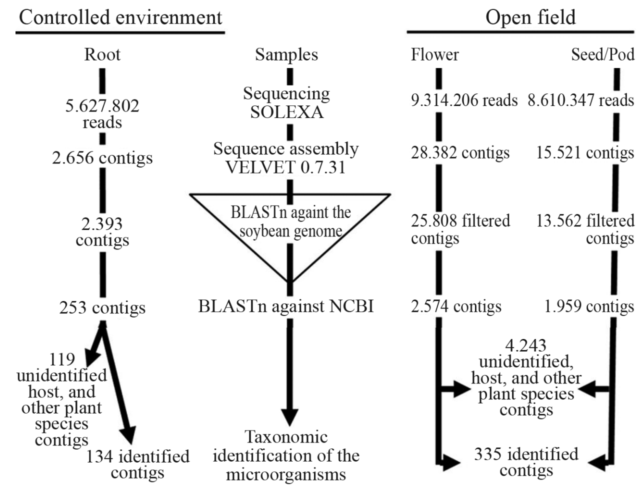

Figure 1 - Chart flow of sequence analysis. The central column identifies the methods. The number of reads obtained at each step is indicated for samples obtained from a controlled environment (left) and from field samples (right).

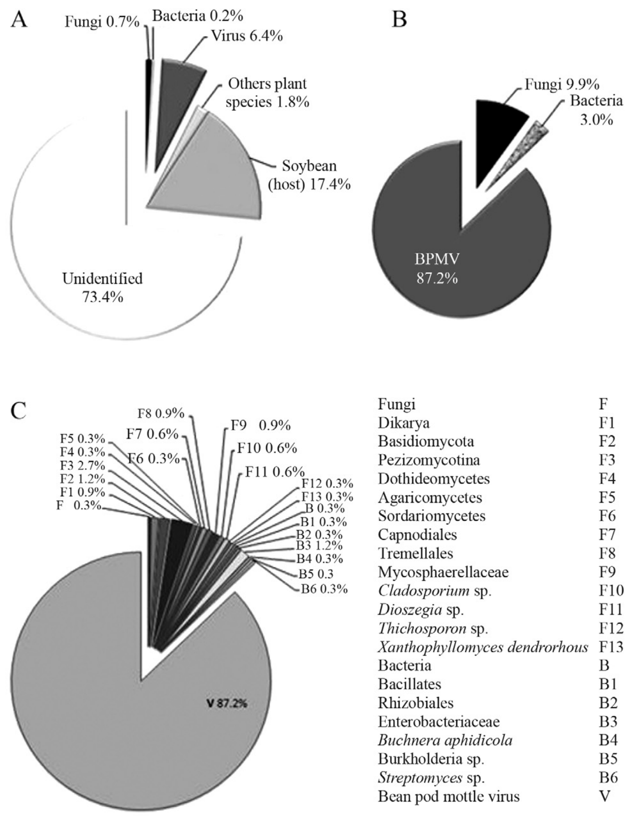

Figure 2 - Comparative percentages of the contigs assembled from sRNA (19-24 nt) libraries of the field (crop soybean) for each sequence classification according to best hit with BLASTn. (A) All groups of organisms, (B) all groups excluding unidentified and plants and (C) best identified taxonomic groups. Bean Pod Mottle Virus (BPMV).

Taxonomic origin of sRNAs from controlled environmental samples

The most represented taxon in the controlled environment samples belongs to the domain Eukaryota (49.3%), including unicellular eukaryotes (46.3%) and fungi (3.0%), followed by the domain Bacteria (32.8%). In addition, 17.9% of the microbial contigs could only be classified to the taxonomic level of the domain Eukaryota (Figure 3B).

Unicellular eukaryote organisms were found only in controlled environment samples. The amoeba genus *Naegleria* was well represented (12 contigs) within the kingdom Excavata, with one contig classified to species level, viz. *Naegleria fowleri* (Table 1, Figures 3C and 4C). Contigs corresponding to the phylum Kinetoplastida were distributed at different levels of taxonomic classification through the family Bodonidae, including *Neobodo designis* and *Rhynchomonas nasuta.* The family Trypanosomatidae was also identified within this phylum, including the genus *Trypanosoma.* Within the kingdom Chromalveolata (group Heterokontophyta), the class Chrysophyceae was identi-

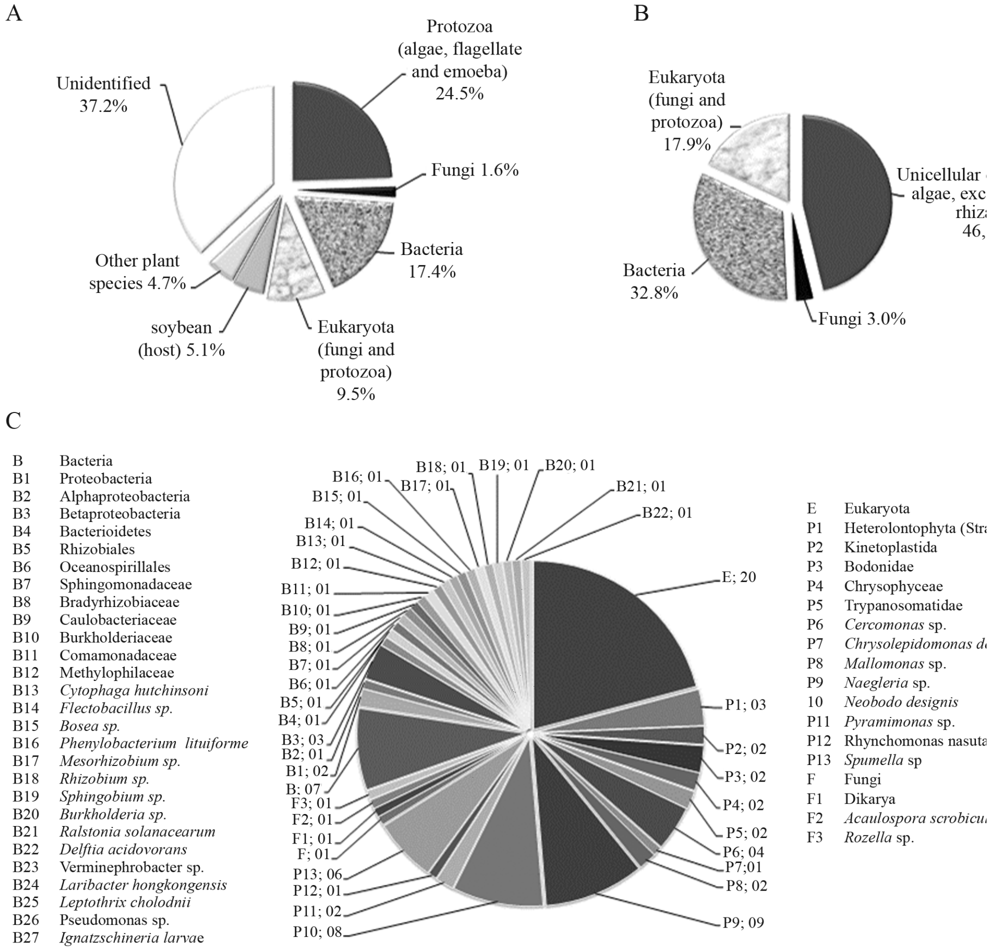

Figure 3 - Comparative percentages of the contigs assembled from sRNA (19-24 nt) libraries of the soybean cultivated in a controlled environment for each sequence classification according to best hit with BLASTn. (A) All groups of organisms, (B) all groups excluding unidentified and plants and (C) best identified taxonomic groups.

fied through two contigs. The genera *Chrysolepidomonas* and *Spumella* (class Chrysophyceae) were represented by one and seven contigs, respectively. One contig was assigned and classified as pertaining to *Spumella elongata*. The genus *Mallomonas* was also identified within the kingdom Chromalveolata (group Heterokontophyta).

Microalgae belonging to kingdom Plantae were represented by the genera *Chlorococcum* and *Pyramimonas*. The phylum Cercozoa belonging to kingdom Rhizaria was represented by six contigs, with four pertaining to the genus *Cercomonas* and one to the genus *Gymnophrys*.

The kingdom Fungi was poorly represented in the controlled environment libraries (Figures 3B, C). One contig was assigned to the genus *Rozella* and another one to the genus *Acaulospora* (Figure 4B). We also found a contig with 94% of identity to an uncultured soil fungus and one with multiple affiliations within the subkingdom Dikarya (Table 1).

The domain Bacteria was the most diverse taxon in controlled environment samples (Figures 3B,C). Contigs affiliated with multiple taxa could be classified only to the domain Bacteria, phyla Bacterioidetes and Proteobacteria, class Alphaproteobacteria and families Sphingomonadaceae, Bradyrhizobiaceae, Caulobacteriaceae, Burkholderiaceae and Comamonadaceae.

The classes Alpha- and Gammaproteobacteria were equally represented, followed by Betaproteobacteria. Within the class Alphaproteobacteria, contigs corresponded to the order Rhizobiales, including the genera *Bosea*, *Rhizobium* and *Mesorhizobium*. There were contigs assigned to families Caulobacteriaceae and Sphingomonadaceae, including the genera *Phenylobacterium* and *Sphingobium*, respectively (Figure 4A).

Within the class Gammaproteobacteria, contigs were derived from the order Burkholderiales, families Burkholderiaceae (including *Burkholderia* and *Ralstonia solanacearum*) and Comamonadaceae (including the genus *Verminephrobacter* and the species *Delftia acidovorans*) and genus *Leptothrix*. The group Oceanospirillales was represented by a single contig. Within the group Pseudomonadales, the genus *Pseudomonas* was the only one to be identified with a single contig. The group Xanthomona-

Table 1 - Taxonomic affiliation of the 134 contigs assembled from libraries obtained from plants grown in a controlled environment (root samples).

Domain	Kingdom	Organism	Environment	No. of contigs	Identity (%)	Sequence type
Eukaryota		Fungi/Unicellular eucaryotes	Soil, plants, water, humans and animals	25	81-100	18S rRNA, 26S rRNA, 28S rRNA and Genomic DNA
Eukaryota	Chromalveolata	Stramenopiles	Water and soil	5	98-100	28S rRNA and 18S rRNA
Eukaryota	Chromalveolata	*Mallomonas* sp. (algae)	Water	2	94-100	ITS1-5.8 rRNA-ITS2-28S rRNA
Eukaryota	Chromalveolata	Chrysophyceae (phytoplankton)	Marine, lake and river water	2	100	28S rRNA
Eukaryota	Chromalveolata	*Spumella* sp.(flagellate)	Water	6	98-100	18S rRNA
Eukaryota	Chromalveolata	*Spumella elongate* (flagellate)	Water	1	100	18S rRNA
Eukaryota	Chromalveolata	*Chrysolepidomonas* sp. (flagellate)	Water	1	98	Genomic DNA
Eukaryota	Excavata	Kinetoplastida	Water	2	100	28S rRNA
Eukaryota	Excavata	Trypanosomatidae	Marine, lake and river sediments	3	90-100	24S α rRNA and genomic DNA
Eukaryota	Excavata	*Trypanosoma* sp.	Marine, lake and river sediments	1	100	18S rRNA, ITS1, 5.8S
Eukaryota	Excavata	Bodonidae	Marine, lake and river sediments	3	100	18S rRNA
Eukaryota	Excavata	*Rhynchomonas nasuta* (flagellate)	Water and soil	1	100	18S rRNA
Eukaryota	Excavata	*Neobodo* sp. (flagellate)	Marine, lake and river sediments	2	95-98	18S rRNA
Eukaryota	Excavata	*Neobodo designis* (flagellate)	Marine, lake and river sediments	8	100	18S rRNA
Eukaryota	Excavata	*Naegleria* sp. (amoeba)	Water and soil	12	91-100	18S rRNA, 28S rRNA, extrachromosomal rRNA plasmid DNA
Eukaryota	Excavata	*Naegleria fowleri* (amoeba)	Water and soil	1	100	18S rRNA
Eukaryota	Rhizaria	Cercozoa (ameboflagellates)	Soil and water sediments	2	89-100	18S rRNA and 28S rRNA
Eukaryota	Rhizaria	*Cercomonas* sp. (ameboflagellates)	Soil and water sediments	4	100	28S rRNA
Eukaryota	Rhizaria	*Gymnophrys* sp. (ameboflagellates)	Soil and water sediments	1	100	28S rRNA
Eukaryota	Viridiplantae	*Pyramimonas* sp. (microalgae)	Marine, lake and river water	3	95-96	28S rRNA
Eukaryota	Viridiplantae	*Chlorococcum* sp. (microalgae)	Marine, lake and river water	1	100	28S rRNA
Eukaryota	Fungi	Fungi	Soil, plant, human, animal, insect	1	94	18S rRNA
Eukaryota	Fungi	*Acaulospora* sp. (arbuscular mycorrhiza)	Plant roots	1	87	18S rRNA
Eukaryota	Fungi	*Rozella* sp.	Endoparasite of other fungi	1	92	28S rRNA
Eukaryota	Fungi	Dikarya	Soil, plant, human, animal, insect	1	98	26S rRNA
Bacteria		Bacteria	Soil, plants, water, humans and animals	10	92-100	16S rRNA and 23S rRNA
Bacteria		Proteobacteria	Soil, plants, water	2	97-100	16S rRNA and 23S rRNA
Bacteria		Alphaproteobacteria	Soil, plants, water	2	100	16S rRNA
Bacteria		Rhizobiales	Soil and plant root nodules	1	96	Genomic DNA
Bacteria		*Mesorhizobium* sp.	Soil and plant root nodules	1	98	23S rRNA
Bacteria		*Rhizobium* sp.	Soil and plant root nodules	1	100	16S rRNA

Table 1 (cont.)

Domain	Kingdom	Organism	Environment	No. of contigs	Identity (%)	Sequence type
Bacteria		Bradyrhizobiaceae	Soil and plant root nodules	1	100	16S rRNA
Bacteria		*Bosea* sp.	Soil and plants	1	100	16S rRNA
Bacteria		Sphingomonadaceae	Soil	1	100	16S rRNA
Bacteria		*Sphingobium* sp.	Soil	1	100	23S rRNA
Bacteria		Caulobacteriaceae	Water	1	100	23S rRNA
Bacteria		*Phenylobacterium* sp.	Water	1	98	23S rRNA
Bacteria		Betaproteobacteria	Soil, plants, water	4	100	16S rRNA and genomic DNA
Bacteria		Methylophilaceae	Soil	1	98	23S rRNA
Bacteria		*Laribacter* sp.	Water	1	95	Genomic DNA
Bacteria		Oceanospirillales	Lakes, rivers and sea hypersaline and highly alkaline.	1	97	Genomic DNA
Bacteria		Burkholderiaceae	Soil and plants	1	100	23S rRNA
Bacteria		*Burkholderia* sp.	Soil and plants	1	96	Genomic DNA
Bacteria		*Ralstonia solanacearum*	Soil and water	1	100	Genomic DNA
Bacteria		Comamonadaceae	Water	1	100	16S rRNA and 23S rRNA
Bacteria		*Verminephrobacter* sp.	Lumbricid earthworms (*Lumbricidae nephridia*)	1	96	Genomic DNA
Bacteria		*Delftia acidovorans*	Soil and water	1	100	16S rRNA
Bacteria		*Leptothrix* sp.	Aquatic environments with sufficient organic matter	1	90	23S rRNA
Bacteria		*Pseudomonas* sp.	Plant leaves	1	100	16S rRNA
Bacteria		*Ignatzschineria larvae*	Insect mid gut (adult flesh fly)	1	100	23S rRNA
Bacteria		Bacterioidetes	Soil, plants, water	1	100	Genomic DNA
Bacteria		*Flectobacillus* sp.	Water	1	98	16S rRNA
Bacteria		*Cytophaga hutchinsoni*	Soil and plant debris (cellulose degradation)	3	100	Genomic DNA
Total				134		

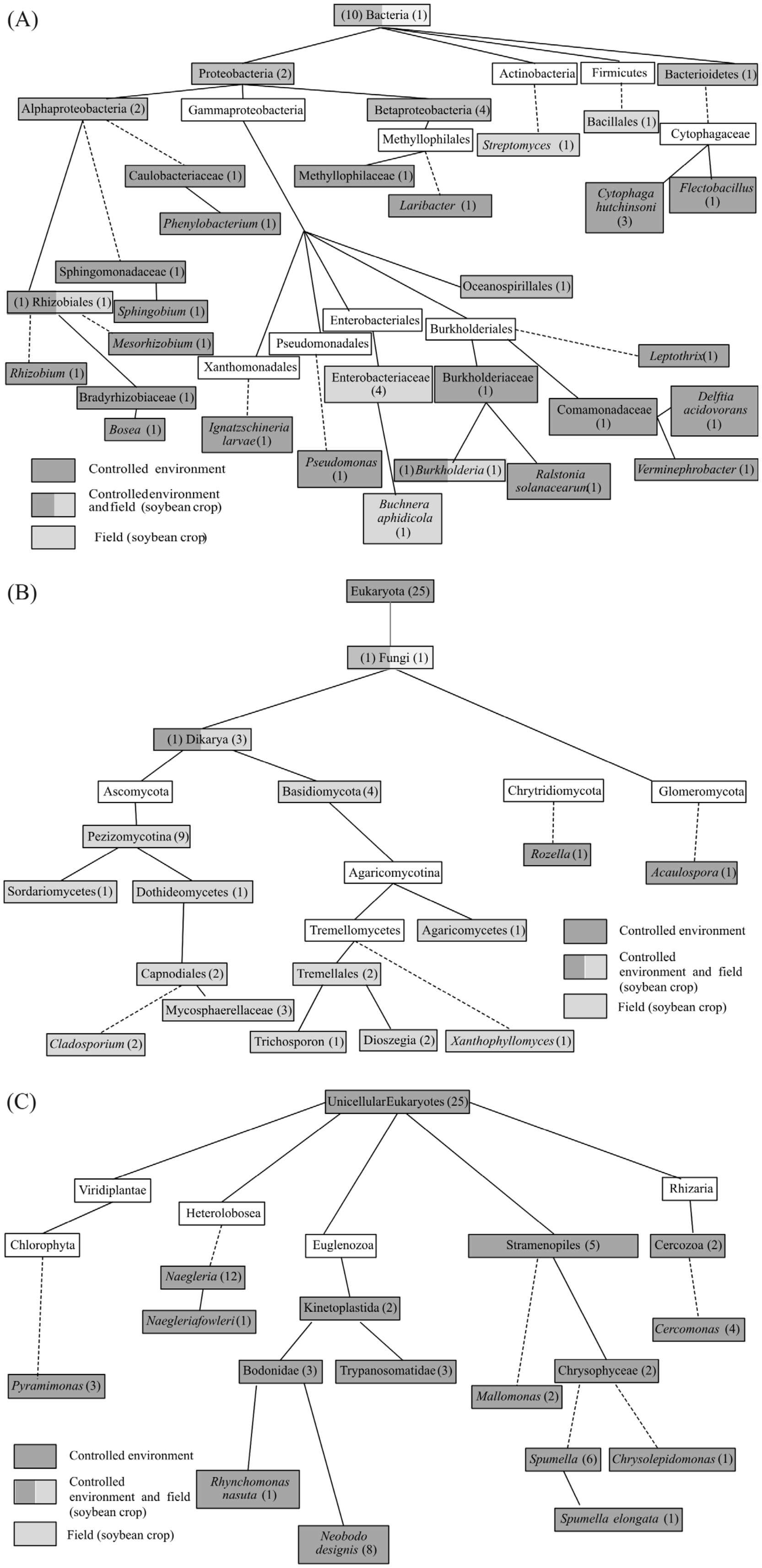

Figure 4 - Taxonomic levels of microorganisms distributed among the libraries of small RNAs in soybean cultivated in a controlled environment and in the field. (A) Bacteria (B) Fungi and (C) other Eukaryota.

dales was represented by the species *Ignatzschineria larvae* (Figure 4), and the phylum Bacterioidetes by the species *Cytophaga hutchinsoni* and genus *Flectobacillus*, both members of the family Cytophagaceae (Figure 4).

Further contigs were classified at the genus leve and shared high identity (at least 87%) to nucleotide sequences of species deposited in the NCBI database (Supplementary Material, Table S1). These may belong to the same or a closely related species, viz. the unicellular eukaryotes *Naegleria gruberi*, *Mallomonas asmundae* and *Chrysolepidomonas dendrolepidota*, the bacteria *Phenylobacterium lituiforme*, *Laribacter hongkongensis* and *Leptothrix cholodnii,* and the fungus *Acaulospora scrobiculata.*

Taxonomic origin of sRNAs from field samples

The BLAST analysis revealed that 87.2% of the contigs assembled from field samples were related to BPMV (Bean Pod Mottle Virus) (Figure 2B). These contigs were assembled from seed and pod samples and showed identities ranging from 94 to 100% for both RNA1 and RNA2 sequences from the virus (Table 2). The remaining sequences were fungi (9.9%) and bacteria (3.0%).

Within the kingdom Fungi, contigs were classified according to the subkingdom Dikarya and phylum Basidiomycota. One contig was assigned within the phylum Basidiomycota, subphylum Agaricomycetes, and two contigs were assigned to the order Tremellales. Within the order Tremellales, two contigs represented the genus *Dioszegia* and one contig the genus *Trichosporon* (Figure 4B). In addition, one contig was not identical but showed high identity (94%) to the species *Xanthophyllomyces dendrorhous* and could be classified only in the genus *Xanthophyllomyces* (Table 2) which belongs to the class Tremellomycetes.

The phylum Ascomycota subphylum Pezizomycotina was the most represented fungal taxon (Figure 2C). It was identified through nine contigs that had multiple affiliations within this taxon, these being with the classes Sordariomycetes and Dothideomycetes. Within this latter there were contigs affiliated to the order Capnodiales, including the family Mycosphaerellaceae and the genus *Cladosporium* (Figure 4B).

A single contig was classified only at the domain level Bacteria. There were contigs to the class Gammaproteobacteria, one to the genus *Burkholderia* and four to the family Enterobacteriaceae, including one to *Buchnera aphidicola.* There were also contigs associated with the class Alphaproteobacteria (order Rhizobiales) and the phyla Firmicutes (order Bacillales) and Actinobacteria from the genus *Streptomyces* (Figures 2 and 4A, Table 2).

Discussion

High-throughput sequencing allows the direct sequencing of DNA or cDNA from the environment, avoiding any cloning bias and also being less costly and faster than the Sanger method (Cardenas and Tiedje, 2008). The Illumina (Solexa) technology is capable of generating 36 million reads with average length of 35 bp within 4 days, which is several times higher than the output of traditional cloning libraries. The short read lengths are compensated by massive output, speed, simplicity and coverage, including regions recalcitrant to cloning.

To obtain ribosomal sequences (commonly used in the identification of species) from complex microbial communities, classical polymerase chain reaction (PCR) has been used with primers complementary to highly conserved regions of rRNA (Rosen *et al.* 2009), representing the bias of being a targeted approach (Bailly *et al.*, 2007).

Because the new sequencing technologies have the potential to sequence technically difficult regions, such as those that form firm secondary structures, they are useful in the analysis of rRNA (Cardenas and Tiedje, 2008).

The metagenomic analysis of sRNA high-throughput libraries, as done in this study, is a non-targeted approach that avoids amplification steps and the design of the 'universal primers'. In this way, the present study provided several rRNA sequences that could be used in taxonomic identification (Tables 1 and 2).

Here, we demonstrate the feasibility of metatranscriptomics/metagenomics to study the microbial communities present in soybean plants cultivated in field and controlled environment. In this work, high-throughput sequencing generated several records for bacteria, fungi, unicellular eukaryotes, and viruses at different taxonomic levels (species, genus, family, order, class, phylum, kingdom or domain).

In root tissues of soybean plants cultivated in a controlled environment, three groups of organisms were detected, these being unicellular eukaryotes, bacteria and fungi. The unicellular eukaryotes were the most abundant group of organisms in root tissues of soybean cultivated in a controlled environment. Unicellular eukaryotes included some subgroups of eukaryote microorganisms, such as microalgae (genera *Pyramimonas* and *Chlorococcum,* these belonging to the division Chlorophyta, kingdom Plantae, and the genus *Mallomonas* belonging to the phylum Heterokontophyta or Stramenopiles, kingdom Chromalveolata), flagellates (including *Neobodo designis, Rhynchomonas nasuta* and the genus *Trypanosoma*, belonging to the class Kinetoplastida, kingdom Excavata, and the genus *Spumella* belonging to the phylum Heterokontophyta or Stramenopiles, kingdom Chromalveolata), amoeba (genus *Naegleria*, including *N. loweri*, belonging to the class Heterolobosea, kingdom Excavata), and ameboflagellates (genera *Cercomonas* and *Gymnophrys* belonging to the phylum Cercozoa, kingdom Rhizaria), which are free-living organisms in freshwater, soil and marine habitats (Bailly *et al.*, 2007). The fact that plants grown in a controlled environment were cultivated in a hydroponic system with nutritive solution could explain the high percentage of

Table 2 - Taxonomic affiliation of the 335 contigs assembled from libraries obtained from soybean plants grown in the field (flowers, pods, mature seeds and samples of germinating seeds).

Domain	Kingdom	Organism	Environment	No. of contigs	Identity (%)	Sequence type
Eukaryota	Fungi		Soil, dung, plant, decaying wood, fungal, insect, human	1	100	18S rRNA
Eukaryota	Fungi	Dikarya	Soil, dung, plant, rock, decaying wood, fungal, insect, human	3	100	18S, ITS1, 5.8S rRNA, ITS 2, 28S rRNA
Eukaryota	Fungi	Basidiomycota	Plant, humans, animals, water, soil	4	100	18S rRNA and 28S rRNA
Eukaryota	Fungi	Agaricomycetes	Plant, plant debris	1	100	28S rRNA
Eukaryota	Fungi	*Xanthophyllomyces* sp. (yeast)	Plant	1	94	28S rRNA
Eukaryota	Fungi	Tremellales	Human, parasites of other fungi	2	100	5.8S rRNA gene, ITS2, 26S rRNA
Eukaryota	Fungi	*Trichosporon* sp. (yeasts)	Soil, human	1	100	18S rRNA
Eukaryota	Fungi	*Dioszegia* sp. (yeasts)	Plant	2	100	26S rRNA, 18S rRNA
Eukaryota	Fungi	Pezizomycotina	Plants, soil, debris plant, rock	9	100	18S rRNA gene, ITS1, 5.8S rRNA, ITS 2, 28S rRNA
Eukaryota	Fungi	Sordariomycetes	Soil, dung, plant, decaying wood, fungal, insect, human	1	100	28S rRNA
Eukaryota	Fungi	Dothideomycetes	Plants, soil, debris plant, rock	1	100	28S rRNA
Eukaryota	Fungi	Capnodiales	Plants, soil, debris plant, rock	2	97-100	18S rRNA and 28S rRNA
Eukaryota	Fungi	Mycosphaerellaceae	Plants, soil, debris plant	3	100	18S rRNA and 28S rRNA
Eukaryota	Fungi	*Cladosporium* sp.	Plants, soil, debris plant	2	98-100	18S rRNA
Bacteria		Bacteria	Plant, humans, animals, water, soil	1	100	16S rRNA
Bacteria		Rhizobiales	Soil and root plant nodules	1	100	16S rRNA
Bacteria		Enterobacteriaceae	Plant, humans, animals, water, soil	4	100	16S rRNA, 23S rRNA and genomic DNA
Bacteria		*Buchnera aphidicola*	Insect	1	100	Genomic DNA
Bacteria		*Burkholderia* sp.	Soil, plant, human	1	100	Chromosome 2
Bacteria		*Streptomyces* sp.	Plant	1	100	16S rRNA
Bacteria		Bacillales	Plant, humans, animals, water, soil	1	100	23S rRNA
Aphanabionta		Bean pod mottle virus	Plant	292	94-100	RNA1, soy RNA1-b polyprotein, RNA2, putative defective coat protein, capsid polyprotein
Total				335		

unicellular eukaryotes found in this habitat, since these organisms feed on the sediment of organic matter present in water.

The domain Bacteria was the second highest group in terms of occurrence in root samples, including endophytic and epiphytic bacteria. Species, genera, families, and orders related to growth promotion and nodulation through endosymbiosis with rhizobia (Juteau *et al.*, 2004; Kuklinsky-Sobral *et al.*, 2004; Delmotte *et al.*, 2009; Ikeda *et al.*, 2009, 2010; Okubo *et al.*, 2009) were found in this work (Table S1). They include the genera *Rhizobium, Mesorhizobium, Sphingobium, Burkholderia, Bosea and Pseudomonas*, the families Bradyrhizobiaceae, Sphingomonadaceae and Rhizobiaceae, and the order Rhizobiales. We also found some species and genera involved in the degradation of cellulose in plant debris and that of organic matter in humid and freshwater habitats, such as the species/genera *Cytophaga hutchinsoni, Delftia acidovorans, Ignatzschineria larvae*, *Laribacter* (Woo *et al.*, 2009), *Leptothrix* and *Phenylobacterium.*

Ralstonia solanacearum is considered a bacterial pathogen. It causes wilt in important crops (including soybean) in other countries. Nevertheless, populations of *Ralstonia solanacearum* may occur as free-living microorganisms in watercourses or in a latent form in plants without causing disease (Grey and Steck, 2001; Mole *et al.*, 2007). The genus *Pseudomonas* identified in this study includes members that are pathogenic to soybean, such as *P. syringae* pv. *glycinea*, or non-pathogenic ones, surviving as saprophytes, epiphytes or endophytes (Kuklinsky-Sobral *et al.*, 2004).

Endophytic and epiphytic bacteria can contribute to the health, growth and development of plants. Promotion of plant growth by these bacteria may result from indirect effects, such as the biocontrol of soilborne diseases through competition for nutrients, siderophore-mediated competition for iron, antibiosis, or the induction of systemic resistance in the host plant. Direct effects, such as the production of phytohormones, providing the host plant with fixed nitrogen, or solubilization of phosphorus and iron present in soil, may also be of relevance (Kuklinsky-Sobral *et al.*, 2004).

The bacterial families Commamonadaceae and Caulobacteriaceae have members involved in sediment degradation in freshwater. Other groups of bacteria, such as members of the family Methylophilaceae and the order Oceanospirillales can survive in soil and humid habitats.

Fungi present in the roots of soybean plants were found to belong to the genus *Rozella,* these being involved in the biological control of other fungi in plants. The genus *Acaulospora* forms arbuscular mycorhiza in plants, thus increasing nutrient absorption from soil, mainly phosphorus.

In seeds, seedlings, pods and flowers from soybean plants cultivated in field (crop), three groups of organisms were identified: virus, fungi and bacteria. The virus group was represented by a single member only, the *Bean Pod Mottle Virus* (BPMV). This virus is widespread in the major soybean growing areas throughout Brazil and the world (Anjos *et al.*, 2000; Giesler *et al.*, 2002). 0Mottling originates at the hilum and is also referred to as “bleeding hilum” since the hilum color seems to bleed from its normal zone. From then on, the mottling of the seed has similar coloration as the hilum (Giesler *et al.*, 2002). BPMV was detected mainly in the mature seed library of sRNA, and seeds used in the RNA extraction procedure showed symptoms of mottling (data not shown), predicting the occurrence of BPMV infection.

The identified fungi were classified at species, genus, family, order, class, subphylum, and phylum levels. The yeast genera *Xanthophyllomyces*, *Dioszegia* and *Thichosporon* may occur in flower, seed, stem and leaf surfaces, and are essential for biological control of pathogens (Kucsera *et al.*, 1998; Wang *et al.*, 2008; Weber *et al.*, 2008). The genus *Cladosporium* found in this study in flower and seed tissues of soybean is known to contain entomopathogenic species with potential for biocontrol of pest insects (Pimentel *et al.*, 2006). The family Mycosphaerellaceae was detected in this work. It includes various genera, especially *Cercospora, Pseudocercospora, Mycosphaerella, Septoria*, *Ramularia*, etc. that represent more than 10,000 species. The genera *Cercospora*, *Mycosphaerella* and *Septoria* contain some species considered as pathogenic to soybean (Crous *et al.*, 2009). The subkingdom Dikarya, phylum Basidiomycota, subphylum Pezizomycotina, classes Dothideomycetes, Agaricomycetes and Sordariomycetes, and order Tremellales contain both phytopathogenic and nonpathogenic fungi species related to soybean, including species with potential for biocontrol of pest insects and diseases.

Moreover, in the field environment many bacteria were found associated with soybean tissue. The bacterium *Buchnera aphidicola* colonizes insects and may be present on plant surfaces. The genus *Burkholderia* includes endophytic species present in soil and plant (Kuklinsky-Sobral *et al.*, 2005). The family Enterobacteriaceae includes many endophytes, such as the genus *Pantoea,* which occur mainly in leguminous plants involved in biological control of phytopathogens (Delmotte *et al.*, 2009; Ikeda *et al.*, 2010). The orders Rhizobiales and Bacillales contain endophytes and species that colonize the rhizosphere and phyllosphere, and are related to soybean nodulation and biological control of pest insects, fungi and bacteria.

Several contigs could not be classified deeper into the eukaryote domain due to multiple affiliations or hitting to sequences from environmental samples (many from uncultured freshwater eukaryotes) (Table S1).

The high percentage of unknown sequences in this study could correspond to artifacts or to transcript fragments from poorly known taxa. Similarly, in many samples of a metagenomic study of plant viruses (Roossinck *et al.*, 2010), sequences without similarity to GenBank sequences represented more than half of the contigs. Furthermore, in a metatranscriptomic analysis of microbial communities from watercourse, half of the possible protein-encoding sequences from pyrosequencing had no significant hits to previously sequenced genes. The length of the contigs influenced this frequency, as the analysis of larger (> 200 bp) sequences resulted in twice the frequency of annotated sequences when compared to shorter (< 100 bp) reads (Poretsky *et al.*, 2009).

The sequencing of small RNA molecules in the size range of 19 to 25 was chosen as it corresponds to the sizes of small interfering RNAs (siRNAs) and virus derived interfering RNAs (vsiRNAs). These are normally produced when RNA interference mechanisms are activated in order to degrade endogenous or pathogen-derived RNAs. The size range of the sequenced small RNAs could be enlarged up to 100 nucleotides, in order to include other small RNA molecules originated by other processes of degradation. Sizes under 19 nucleotides should be avoided since they would have ambiguous matches in different genomes and would not be useful in species prediction.

The challenges that remain in metagenomics and metatranscriptomics include DNA and RNA extraction, low stability, abundance and proportion of mRNAs in total RNA extracts (Cardenas and Tiedje, 2008). In addition, many of the difficulties encountered in microbial diversity studies are due to the complexity of the microbial community and its unevenness (few populations are of high frequency and many populations of low abundance). These difficulties can be reduced in metatranscriptomics by focusing on the active populations in a sample (Morales and Holben, 2010). When analyzing a soil microbial community, Urich *et al.* (2008) demonstrated that the deep-sequencing data from total RNA is naturally enriched in both functionally (such as mRNA libraries) and taxonomically relevant molecules, *i.e.* mRNA and rRNA, respectively.

Nevertheless, upon using deep-sequencing data, significant differences have been found in the taxonomic distribution of cDNAs derived from total RNA compared to DNA libraries from soil samples (Bailly *et al.*, 2007). Differences were also found when comparing cDNA derived from mRNA-enriched libraries with DNA libraries from marine microbial communities (Gilbert *et al.*, 2008). This suggests that both DNA and RNA sequences should be analyzed complementarily when investigating the diversity of species from environmental samples. In this sense, sRNA sequencing libraries can contribute to microbial identification with other types of sequences, viz. non-coding RNAs and mRNA fragments that would not normally be present in libraries enriched for polyA-tailed mRNA.

Acknowledgments

LGM was sponsored by a M.Sc. grant; GCF, GLM, LFVO and FRK by a PhD grant, JBC by a DTI-RHAE and RM by a Productivity and Research grant from the National Council for Scientific and Technological Development (CNPq, Brazil). This work was financially supported by GenoSoja consortium (CNPq 5527/2007-8) and GenoProt (CNPq 559636/2009-1).

References

Adams IP, Glover RH, Monger WA, Mumford R, Jackeviciene E, Navalinskiene M, Samuitiene M and Boonham N (2009) Next-generation sequencing and metagenomic analysis: A universal diagnostic tool in plant virology. Mol Plant Pathol 10:537-545.

Al Rwahnih M, Daubert S, Golino D and Rowhani A (2009) Deep sequencing analysis of RNAs from a grapevine showing Syrah decline symptoms reveals a multiple virus infection that includes a novel virus. Virology 387:395-401.

Anjos JRN, Charchar MJA and Gomes AC (2000) Identificação do vírus do mosqueado do feijoeiro (“Bean Pod Mottle Virus”) em soja no Brasil. Documentos Embrapa Cerrados 19:1-15.

Bailly J, Fraissinet-Tachet L, Verner MC, Debaud JC, Lemaire M, Wesolowski-Louvel M and Marmeisse R (2007) Soil eukaryotic functional diversity, a metatranscriptomic approach. ISME J 1:632-642.

Cardenas E and Tiedje JM (2008) New tools for discovering and characterizing microbial diversity. Curr Opin Biotechnol 19:544-549.

Chistoserdova L (2010) Recent progress and new challenges in metagenomics for biotechnology. Biotechnol Lett 32:1351-1359.

Coetzee B, Freeborough MJ, Maree HJ, Celton JM, Rees DJ and Burger JT (2010) Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology 400:157-163.

Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, Power DM, Packer M, Carvalho GR, Blaxter ML, Lambshead PJ *et al.* (2010) Ultrasequencing of the meiofaunal biosphere: Practice, pitfalls and promises. Mol Ecol 19(Suppl 1):4-20.

Crous PW, Summerell BA, Carnegie AJ, Wingfield MJ and Groenewald JZ (2009) Novel species of Mycosphaerellaceae and Teratosphaeriaceae. Persoonia 23:119-146.

Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R, von Mering C and Vorholt JA (2009) Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc Natl Acad Sci USA 106:16428-16433.

Desai C, Pathak H and Madamwar D (2010) Advances in molecular and “-omics” technologies to gauge microbial communities and bioremediation at xenobiotic/anthropogen contaminated sites. Bioresour Technol 101:1558-1569.

Giesler LJ, Ghabrial SA, Hunt TE and Hill JH (2002) Bean Pod Mottle Virus: A threat to U.S. soybean production. Plant Disease 86:1280-1289.

Gifford SM, Sharma S, Rinta-Kanto JM and Moran MA (2010) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. ISME J 5:461-472.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P and Joint I (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. PLoS One 3:e3042.

Grey BE and Steck TR (2001) The viable but nonculturable state of *Ralstonia solanacearum* may be involved in long-term survival and plant infection. Appl Environ Microbiol 67:3866-3872.

Ikeda S, Kaneko T, Okubo T, Rallos LE, Eda S, Mitsui H, Sato S, Nakamura Y, Tabata S and Minamisawa K (2009) Development of a bacterial cell enrichment method and its application to the community analysis in soybean stems. Microb Ecol 58:703-714.

Ikeda S, Okubo T, Anda M, Nakashita H, Yasuda M, Sato S, Kaneko T, Tabata S, Eda S, Momiyama A *et al.* (2010) Community- and genome-based views of plant-associated bacteria: Plant-bacterial interactions in soybean and rice. Plant Cell Physiol 51:1398-1410.

Juteau P, Tremblay D, Villemur R, Bisaillon JG and Beaudet R (2004) Analysis of the bacterial community inhabiting an aerobic thermophilic sequencing batch reactor (AT-SBR) treating swine waste. Appl Microbiol Biotechnol 66:115-122.

Kent AD, Yannarell AC, Rusak JA, Triplett EW and McMahon KD (2007) Synchrony in aquatic microbial community dynamics. ISME J 1:38-47.

Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I and Simon R (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. Virology 388:1-7.

Kucsera J, Pfeiffer I and Ferenczy L (1998) Homothallic life cycle in the diploid red yeast *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). Antonie Van Leeuwenhoek 73:163-168.

Kuklinsky-Sobral J, Araujo WL, Mendes R, Geraldi IO, Pizzirani-Kleiner AA and Azevedo JL (2004) Isolation and characterization of soybean-associated bacteria and their potential for plant growth promotion. Environ Microbiol 6:1244-1251.

Kuklinsky-Sobral J, Araújo WL, Mendes R, Pizzirani-Kleiner AA and Azevedo JL (2005) Isolation and characterization of endophytic bacteria from soybean (*Glycine max*) grown in soil treated with glyphosate herbicide. Plant Soil 273:91-99.

Li R, Yu1 Y, Li Y, Lam TW, Yiu SM, Kristiansen K and Wang J (2009) SOAP2: An improved ultrafast tool for short read alignment. Bioinformatics 25:1966-1967.

Mole BM, Baltrus DA, Dangl JL and Grant SR (2007) Global virulence regulation networks in phytopathogenic bacteria. Trends Microbiol 15:363-371.

Morales SE and Holben WE (2010) Linking bacterial identities and ecosystem processes: Can “omic” analyses be more than the sum of their parts? FEMS Microbiol Ecol 75:2-16.

Okubo T, Ikeda S, Kaneko T, Eda S, Mitsui H, Sato S, Tabata S and Minamisawa K (2009) Nodulation-dependent communities of culturable bacterial endophytes from stems of field-grown soybeans. Microbes Environ 24:253-258.

Pimentel IC, Glienke-Blanco C, Gabardo J, Stuart RM and Azevedo JL (2006) Identification and colonization of endophytic fungi from soybean (*Glycine max* (L.) Merril) under different environmental conditions. Braz Arch Biol Technol 49:705-711.

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP and Moran MA (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. Environ Microbiol 11:1358-1375.

Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, Lai H, Chavarria F, Shen G and Roe BA (2010) Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. Mol Ecol 19(Suppl 1):81-88.

Rosen GL, Sokhansanj BA, Polikar R, Bruns MA, Russell J, Garbarine E, Essinger S and Yok N (2009) Signal processing for metagenomics: Extracting information from the soup. Curr Genomics 10:493-510.

Shi Y, Tyson GW and DeLong EF (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean’s water column. Nature 459:266-269.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C and Schuster SC (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the metatranscriptome. PLoS One 3:e2527.

Wang QM, Jia JH and Bai FY (2008) Diversity of basidiomycetous phylloplane yeasts belonging to the genus *Dioszegia* (Tremellales) and description of *Dioszegia athyri* sp. nov., *Dioszegia butyracea* sp. nov. and *Dioszegia xingshanensis* sp. nov. Antonie Van Leeuwenhoek 93:391-399.

Warnecke F and Hess M (2009) A perspective: Metatranscriptomics as a tool for the discovery of novel biocatalysts. J Biotechnol 142:91-95.

Weber RW, Becerra J, Silva MJ and Davoli P (2008) An unusual Xanthophyllomyces strain from leaves of Eucalyptus globulus in Chile. Mycol Res 112:861-867.

Woo PC, Lau SK, Tse H, Teng JL, Curreem SO, Tsang AK, Fan RY, Wong GK, Huang Y, Loman NJ *et al.* (2009) The complete genome and proteome of *Laribacter hongkongensis* reveal potential mechanisms for adaptations to different temperatures and habitats. PLoS Genet 5:e1000416.

Xu J (2006) Microbial ecology in the age of genomics and metagenomics: Concepts, tools, and recent advances. Mol Ecol 15:1713-1731.

Zerbino DR and Birney E (2008) Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821-829.

Zoetendal EG, Rajilic-Stojanovic M and de Vos WM (2008) High-throughput diversity and functionality analysis of the gastrointestinal tract microbiota. Gut 57:1605-1615.

Internet Resources

Phytozome, ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v6.0/Gmax/assembly/sequence/.

Supplementary Material

The following online material is available for this article:

Tables S1 - Relationship among metatranscriptomic sequencing data and species sequences in public data banks

This material is available as part of the online article from http://www.scielo.br/gmb.

Table. S1. Relationship among metatranscriptomic sequencing data and species sequences in public data Banks.

Class	Order	Family	Genus	Specie	Taxon	Gene sequence	Acession Genbank	E-value	Length (bp)	Query (%)	Max Identity (%)
Sphingobacteria	Sphingobacteriales	Cytophagaceae	Cytophaga	*Cytophaga hutchinsoni*	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Cytophagaceae; Cytophaga hutchinsoni	Genomic DNA	CP000383.1	5,00E-019	55	100	100
Sphingobacteria	Sphingobacteriales	Cytophagaceae	Cytophaga	*Cytophaga hutchinsoni*	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Cytophagaceae; Cytophaga hutchinsoni	Genomic DNA	CP000383.1	7,00E-023	62	100	100
Sphingobacteria	Sphingobacteriales	Cytophagaceae	Cytophaga	*Cytophaga hutchinsoni*	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Cytophagaceae; Cytophaga hutchinsoni	Genomic DNA	CP000383.1	3,00E-017	51	100	100
Sphingobacteria	Sphingobacteriales	Cytophagaceae	Flectobacillus	*Flectobacillus sp.*	Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales; Cytophagaceae: Flectobacillus sp. environmental samples.	16S rRNA	HQ111157.1	9,00E-022	63	100	98
					Bacteria; Bacteroidetes; Sphingobacteria; Sphingobacteriales;Sphingobacteriaceae; Sphingobacterium sp. (environmental samples);Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;	Genomic DNA	GU925156.1	1,00E-020	58	100	100
Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Bosea	*Bosea sp.*	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;Bradyrhizobiaceae; Bosea sp.	16S rRNA	HQ260890.1	1,00E-023	63	98	100
Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae			Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae: Bosea sp.; Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Bradyrhizobiaceae; uncultured alphaproteobacterium	16S rRNA	HQ259685.1	7,00E-018	53	100	100
Alphaproteobacteria	Caulobacteriales	Caulobacteriaceae	Phenylobacterium	*Phenylobacterium lituiforme*	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae;	23S rRNA	DQ306696.1	5,00E-021	61	100	98

					Phenylobacterium lituiforme						
Alphaproteobacteria	Caulobacteriales	Caulobacteriaceae			Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Caulobacteraceae: Phenylobacterium sp; Brevundimonas sp.; Caulobacter sp.	23S rRNA	CP000747.1	1,00E-024	65	100	100
Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	Mesorhizobium	*Mesorhizobium sp.*	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Phyllobacteriaceae; Mesorhizobium sp.	23S rRNA	EU697970.1	2,00E-027	73	100	98
Alphaproteobacteria	Rhizobiales	Rhizobiaceae	Rhizobium	*Rhizobium sp.*	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium sp.	16S rRNA	DQ454123.3	3,00E-023	63	98	100
Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium	*Sphingobium sp.*	Bacteria; environmental samples: uncultured bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales;Sphingomonadaceae; Sphingobium sp.	23S rRNA	GU925159.1	5,00E-024	64	100	100
Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae			Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae: Sphingomonas sp; Novosphingobium sp.	16S rRNA	FJ938155.1	2,00E-017	51	100	100
Alphaproteobacteria					Bacteria; Proteobacteria; Alphaproteobacteria; environmental samples: uncultured alphaproteobacteria.	16S rRNA	GU553112.1	1,00E-049	119	100	97
Alphaproteobacteria					Bacteria; Proteobacteria; Alphaproteobacteria;Uncultured bacterium; Bacteria; Proteobacteria; Alphaproteobacteria; Sphingomonadales; Sphingomonadaceae; Novosphingobium; environmental samples	16S rRNA	AB599923.1	7,00E-023	62	100	100
Alphaproteobacteria	Rhizobiales				Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;Methylobacteriaceae; Methylobacterium sp.;Bacteria; Proteobacteria;	Genomic DNA	CP000943.1	2,00E-014	51	96	97

					Alphaproteobacteria; Rhizobiales;Xanthobacteraceae; Xanthobacter autotrophicus						
Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia	*Burkholderia sp.*	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae; Burkholderia sp.	Genomic DNA	CP002014.1	4,00E-014	52	100	96
Betaproteobacteria	Burkholderiales	Burkholderiaceae	Ralstonia	*Ralstonia solanacearum*	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;Burkholderiaceae; Ralstonia solanacearum	Genomic DNA	FP885906.2	7,00E-026	66	100	100
Betaproteobacteria	Burkholderiales	Burkholderiaceae			Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Burkholderiaceae: Ralstonia solanacearum, Burkholderia sp.	23S rRNA	FP885906.2	4,00E-020	60	95	100
Betaproteobacteria	Burkholderiales	Comamonadaceae	Delftia	*Delftia acidovorans*	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Comamonadaceae; Delftia acidovorans	16S rRNA	FJ410384.1	9,00E-017	51	100	100
Betaproteobacteria	Burkholderiales	Comamonadaceae	Verminephrobacter	*Verminephrobacter sp.*	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;Comamonadaceae; Verminephrobacter sp.	Genomic DNA	CP000542.1	7,00E-028	80	100	96
Betaproteobacteria	Burkholderiales	Comamonadaceae			Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;Comamonadaceae: Comamonas sp.; Curvibacter sp.; Variovorax sp.; Acidovorax sp.; Delftia sp.	16S rRNA and 23S rRNA	CP001220.1	1,00E-024	65	100	100
Betaproteobacteria	Methylophilales	Methylophilaceae			Bacteria; Proteobacteria; Betaproteobacteria; Methylophilales; Methylophilaceae: Methylotenera sp. Methylovorus sp.	23S rRNA	CP002056.1	7,00E-028	75	98	98
Betaproteobacteria	Neisseriales	Neisseriaceae	Laribacter	*Laribacter hongkongensis*	Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Laribacter hongkongensis.	Genomic DNA	CP001154.1	1,00E-012	54	88	95
Betaproteobacteria	Burkholderiales		Leptothrix	*Leptothrix cholodnii*	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Leptothrix cholodnii	23S rRNA	CP001013.1	2,00E-042	133	100	90
Betaproteobacteria					Bacteria; environmental samples: uncultured bacteria; Bacteria; Proteobacteria; Betaproteobacteria.	16S rRNA	EU357219.1	2,00E-037	88	100	100

Betaproteobacteria					Bacteria; Proteobacteria; Betaproteobacteria:Uncultured beta proteobacterium (environmental samples);Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae; Herbaspirillum seropedicae	Genomic DNA	EF157669.1	2,00E-030	74	100	100
Betaproteobacteria					Bacteria; Proteobacteria; Betaproteobacteria:Uncultured beta proteobacterium (environmental samples);Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;Comamonadaceae; (environmental samples: uncultured Comamonadaceae)	16S rRNA	HQ341778.1	2,00E-019	55	100	100
Betaproteobacteria					Bacteria; uncultured bacterium (environmental samples); Bacteria; Proteobacteria; Betaproteobacteria; environmental Samples.	16S rRNA	EU835409.1	2,00E-019	55	100	100
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas		Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;Pseudomonadaceae; Pseudomonas sp. (environmental samples); Bacteria; environmental samples	16S rRNA	HM980737.1	8,00E-019	54	100	100
Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Ignatzschineria	*Ignatzschineria larvae*	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;Xanthomonadaceae; Ignatzschineria larvae	23S rRNA	HM162936.1	3,00E-015	48	100	100
Gammaproteobacteria	Oceanospirillales				Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Hahellaceae; Hahella chejuensis; Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Halomonadaceae: Carnimonas nigrificans And Zymobacter palmae	Genomic DNA	CP000155.1	5,00E-014	51	96	97
					Bacteria; Proteobacteria; environmental samples:uncultured proteobacteria	23S rRNA	DQ312965.1	6,00E-025	72	100	97
					Bacteria; Proteobacteria; Gammaproteobacteria; environmental samples: uncultured	16S rRNA	HQ331513.1	7,00E-017	51	100	100

					bacteria and gammaproteobacteria; Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;Burkholderiaceae; Burkholderia sp.						
					Bacteria; environmental samples	23S rRNA	GU924062.1	3,00E-026	68	100	100
					Bacteria; environmental samples: uncultured bacteria	16S rRNA	AJ874158.1	2,00E-017	52	100	100
					Bacteria; environmental samples: uncultured bacteria	16S rRNA	GQ480126.1	1,00E-020	58	98	100
					Bacteria; environmental samples: Uncultured bacterium	16S rRNA	GQ480158.1	4,00E-035	84	100	100
					Bacteria; environmental samples: uncultured bacterium.	23S rRNA	DQ313043.1	1,00E-010	52	100	92
					Bacteria; environmental samples: uncultured bacteria	16S rRNA	GU568902.1	2,00E-019	55	100	100
					Bacteria; uncultured bacterium environmental samples.	16S rRNA	EU995773.1	1,00E-028	71	100	100
					Bacteria; Uncultured bacteria environmental samples.	16S rRNA	FN777729.1	6,00E-017	50	100	100
					Bacteria; uncultured bacterium (environmental samples).	16S rRNA	HQ343206.1	1,00E-026	68	100	100
					Bacteria; environmental samples	23S rRNA	GU926606.1	2,00E-006	50	62	100
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales;Chromulinaceae; Spumella; unclassified Spumella sp.	18S rRNA	AJ236857.1	2,00E-028	76	97	98
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; Spumella-like flagellate; uncultured Chrysophyceae	18S rRNA	AY651079.1	4,00E-020	56	100	100
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chromulinaceae; Spumella sp.	28S rRNA	EF681936.1	7,00E-018	53	100	100
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella elongata*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chromulinaceae; Spumella elongata.	28S rRNA	EF681931.1	7,00E-023	62	100	100
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales;Chromulinaceae: Spumella-like flagellate	18S rRNA	EF577181.1	5,00E-020	56	100	100

Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; unclassified Chrysophyceae: Spumella-like flagellate	18S rRNA	AY651088.1	2,00E-019	55	100	100
Chrysophyceae	Chromulinales	Chrysolepidomonadaceae	Chrysolepidomonas	*Chrysolepidomonas dendrolepidota.*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chrysolepidomonadaceae; Chrysolepidomonas dendrolepidota.	Genomic DNA	AF409121.1	6,00E-015	50	100	98
Chrysophyceae					Eukaryota; environmental samples;Eukaryota; stramenopiles; Chrysophyceae; Hydrurales; Hydruraceae; Hydrurus foetidus; Eukaryota; stramenopiles; Chrysophyceae; Chromulinales;Chrysolepidomonadaceae; Chrysolepidomonas dendrolepidota.	28S rRNA	GU928040.1	3,00E-021	59	100	100
Chrysophyceae					Eukaryota; stramenopiles; Chrysophyceae; environmental samples	18S rRNA	AB520733.1	8,00E-057	123	100	100
Synurophyceae	Synurales	Mallomonadaceae	Mallomonas	*Mallomonas sp.*	Eukaryota; stramenopiles; Synurophyceae; Synurales; Mallomonadaceae; Mallomonas sp.	28S rRNA	AF409121.1	7,00E-013	60	71	100
Synurophyceae	Synurales	Mallomonadaceae	Mallomonas	*Mallomonas asmundae*	Eukaryota; stramenopiles; Synurophyceae; Synurales; Mallomonadaceae; Mallomonas asmundae	ITS1-5.8 rRNA-ITS2-28S rRNA	AF409122.1	2,00E-024	76	100	94
Chrysophyceae	Chromulinales	Chromulinaceae	Spumella	*Spumella sp.*	Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chromulinaceae: uncultured Spumella (environmental samples); Eukaryota; Uncultured eukaryote (environmental samples)	18S rRNA	GU067799.1	2,00E-039	91	100	100
					Eukaryota; stramenopiles; Synurophyceae; Synurales;Mallomonadaceae: Mallomonas tonsurata; and Eukaryota; stramenopiles; Eustigmatophyceae; Eustigmatales;Monodopsidaceae; Nannochloropsis gaditana; and Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chrysolepidomonadaceae; Chrysolepidomonas dendrolepidota	28S rRNA	FJ030885.1	1,00E-018	53	100	100

				Eukaryota; stramenopiles; Chrysophyceae; environmental samples:Uncultured chrysophyte; Eukaryota; stramenopiles; Synurophyceae; Ochromonadales; Ochromonadaceae; environmental samples:Uncultured Spumella	18S rRNA	GU067890.1	2,00E-012	43	100	100
				Eukaryota; stramenopiles; Synurophyceae; Synurales; Mallomonadaceae; Mallomonas sp.; Eukaryota; stramenopiles; Chrysophyceae; Chromulinales; Chromulinaceae; Spumella; unclassified Spumella sp.	28S rRNA	FJ030885.1	1,00E-014	50	100	98
				Eukaryota; stramenopiles; Synurophyceae; Synurales; Mallomonadaceae; Mallomonas asmundae; Eukaryota; stramenopiles;Chrysophyceae;Chromulinales;Chrysolepidomonadaceae; Chrysolepidomonas dendrolepidota	Genomic DNA	AF409122.1	1,00E-019	59	100	98
Kinetoplastida	Bodonidae	Rhynchomonas	*Rhynchomonas nasuta*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Rhynchomonas nasuta	18S rRNA	DQ465526.1	6,00E-020	56	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis.	18S rRNA	DQ207585.1	9,00E-027	69	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis.	18S rRNA	DQ207585.1	1,00E-020	58	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	18S rRNA	DQ207585.1	7,00E-016	51	94	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	18S rRNA	DQ207585.1	1,00E-016	50	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis.	28S rRNA	FJ176710.1	1,00E-016	53	100	98
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	18S rRNA	DQ207585.1	5,00E-021	58	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	28S rRNA	FJ176710.1	5,00E-030	86	100	95

Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis;	18S rRNA	DQ207585.1	2,00E-018	53	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	18S rRNA	DQ207585.1	8,00E-023	61	100	100
Kinetoplastida	Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis;	18S rRNA	DQ207585.1	3,00E-022	60	100	100
Kinetoplastida	Bodonidae			Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae: Neobodo designis; Bodo sp.	18S rRNA	DQ207585.1	2,00E-032	79	100	100
Kinetoplastida	Bodonidae			Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae: Bodonidae sp.; Cryptaulax sp.;Rhynchomonas nasuta	18S rRNA	EU106845.1	5,00E-015	50	100	98
Kinetoplastida	Bodonidae			Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Bodonidae sp.; Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Rhynchomonas sp.; Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Cryptaulax sp.	18S rRNA	EU106845.1	6,00E-019	57	100	98
Kinetoplastida	Trypanosomatidae			Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania sp., Crithidia sp.	24S alpha rRNA	AM502245.1	4,00E-025	83	90	96
Kinetoplastida	Trypanosomatidae			Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania major; Trypanosoma sp.	Genomic DNA	CP000079.1	1,00E-033	83	100	98
Kinetoplastida				Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; environmental Samples: uncultured Bodonid; Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Bodo sp.; Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Phytomonas sp.	28S rRNA	GQ902915.1	9,00E-022	60	100	100
Kinetoplastida				Eukaryota; environmental samples: uncultured Eukaryote;Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Rhynchomonas sp.; Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Herpetomonas sp.;Eukaryota; Euglenozoa; Kinetoplastida;	28S rRNA	GU927617.1	9,00E-019	56	100	100

					Trypanosomatidae; Crithidia sp.					
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria sp.*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria lovaniensis and N. fowleri	18S rRNA	U80062.1	1,00E-019	59	94	100
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria fowleri*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria fowleri	18S rRNA	U80059.1	3,00E-016	50	100	100
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi.	Genomic DNA	AC237584.1	8,00E-032	90	100	95
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi.	28S rRNA	AB298288.1	1,00E-019	61	100	96
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi	extrachromosomal rRNA plasmid DNA	AB298288.1	7,00E-023	83	100	91
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi.	extrachromosomal rRNA plasmid DNA	AB298288.1	9,00E-012	64	98	89
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi	rRNA plasmid DNA	AB298288.1	1,00E-021	76	97	93
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria sp.*	Eukaryota; environmental samples;Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria sp.	18S rRNA	FN860554.1	4,00E-017	54	100	98
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria sp.*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria sp.	SSU rRNA	Y10187.1	1,00E-014	59	100	94
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi	rRNA plasmid DNA	AB298288.1	4,00E-022	71	97	95
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria sp.*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria sp.	18S rRNA	AY376151.1	1,00E-017	52	100	100
Chytridiomycetes	Spizellomycetales	Spizellomycetales incertae sedis	Rozella	*Rozella sp.*	Eukaryota; Fungi; Chytridiomycota; Chytridiomycetes; Spizellomycetales; Spizellomycetales incertae sedis; Rozella sp.	28S rRNA	DQ273766.1	9,00E-017	72	91	92

Glomeromycetes	Diversiporales	Acaulosporaceae	Acaulospora	*Acaulospora scrobiculata*	Eukaryota; Fungi; Glomeromycota; Glomeromycetes; Diversisporales;Acaulosporaceae; Acaulospora scrobiculata	18S rRNA	FM876791.1	2,00E-012	83	79	87
					Eukaryota; Fungi; environmental samples: uncultured soil fungus.	18S rRNA	EU489955.1	1,00E-045	123	100	94
					Eukaryota; Fungi; Dikarya; Basidiomycota; Agaricomycotina;Agaricomycetes; Agaricomycetidae; Agaricales; Physalacriaceae; Armillaria sp.; Eukaryota; Fungi; Dikarya; Ascomycota; Saccharomycotina;Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia sp.	26S rRNA	HM036345.1	4,00E-015	51	100	98
Prasinophyceae	Pyramimonadales		Pyramimomas	*Pyramimomas sp.*	Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Pyramimonadales; Pyramimonas sp.	28S rRNA	EU375499.1	5,00E-034	91	100	96
Prasinophyceae	Pyramimonadales		Pyramimomas	*Pyramimomas sp.*	Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Pyramimonadales; Pyramimonas sp.	28S rRNA	EU375499.1	1,00E-024	77	96	95
	Cercomonadida	Cercomonadidae	Cercomonas	*Cercomonas sp.*	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae; Cercomonas sp.	28S rRNA	FJ176707.1	9,00E-027	69	100	100
	Cercomonadida	Cercomonadidae	Cercomonas	*Cercomonas sp.*	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae; Cercomonas sp.	28S rRNA	DQ454123.3	3,00E-024	63	100	100
	Cercomonadida	Cercomonadidae	Cercomonas	*Cercomonas sp.*	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae; Cercomonas sp.	28S rRNA	FJ176707.1	4,00E-021	58	100	100
	Cercomonadida	Cercomonadidae	Cercomonas	*Cercomonas sp.*	Eukaryota; Rhizaria; Cercozoa; environmental samples: uncultured Cercozoan; Eukaryota; Rhizaria; Cercozoa;Cercomonadida; Cercomonadidae; Cercomonas sp.	Genomic DNA	FJ353183.1	3,00E-018	52	100	100
					Eukaryota; Rhizaria; Cercozoa:uncultured freshwater cercozoan (environmental samples)	18S rRNA	EU567255.1	8,00E-028	70	100	100
					Eukaryota; environmental samples:uncultured freshwater eukaryote	18S rRNA	AY919786.1	3,00E-033	80	100	100
					Eukaryota; environmental samples:uncultured freshwater eukaryote and Uncultured Amoebozoa	18S rRNA	AY919722.1	1,00E-034	83	100	100

	Eukaryota; environmental samples.	28S rRNA	GU929097.1	3,00E-021	59	100	100
	Eukaryota; Metazoa; Bryozoa; Gymnolaemata; Cheilostomatida;Flustrina; Buguloidea; Bugulidae; Bugula turrita; and Eukaryota; Apusozoa; Apusomonadidae; Apusomonas sp.; Eukaryota; environmental samples (uncultured Eukaryote); Eukaryota; Apusozoa; Apusomonadidae; Amastigomonas sp.	28S rRNA	AY210457.1	4,00E-014	112	96	81
	Eukaryota; Fungi; environmental samples: Uncultured soil fungus; Eukaryota; stramenopiles; Chrysophyceae; environmental samples	26S rRNA	EU861712.1	7,00E-013	57	92	94
	Eukaryota; environmental samples: uncultured eukaryote and freshwater eukaryote.	18S rRNA	GQ861575.1	4,00E-015	48	100	100
	Eukaryota; environmental samples: Uncultured eukaryote .	18S rRNA	AB521584.1	3,00E-016	50	100	100
	Eukaryota; environmental samples: Uncultured freshwater eukaryote .	18S rRNA	AY919722.1	9,00E-017	51	100	100
	Eukaryota; Fungi; environmental sample: uncultured soil fungus; Eukaryota; stramenopiles; Chrysophyceae; Chromulinales;Chromulinaceae; Spumella; unclassified Spumella sp.;	Genomic DNA	EU489955.1	5,00E-019	55	100	100
	Eukaryota; Alveolata; Ciliophora; environmental samples; Eukaryota; stramenopiles; Synurophyceae; Ochromonadales; Ochromonadaceae; environmental samples:Uncultured Spumella; Uncultured chrysophyte	18S rRNA	GU067790.1	2,00E-018	54	100	100
	Eukaryota; Nucleariidae; Micronuclearia sp.; Eukaryota; Fungi; Dikarya; Ascomycota; Taphrinomycotina; Schizosaccharomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae; Schizosaccharomyces sp.;Eukaryota; Fungi; Dikarya; Ascomycota;	18S rRNA	GU001165.1	1,00E-010	40	100	100

					Taphrinomycotina;Taphrinomycetes; Taphrinales; Taphrinaceae; Taphrina sp.;Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;Sordariomycetes; Sordariomycetidae; Magnaporthales;Magnaporthaceae; mitosporic Magnaporthaceae; Phialophora sp.						
					Eukaryota; environmental samples: uncultured eukaryota;Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo sp.	18S rRNA	FN865918.1	9,00E-017	51	100	100
					Eukaryota; environmental samples: uncultured freshwater eukaryote	18S rRNA	AY919786.1	5,00E-021	57	100	100
					Eukaryota; environmental samples: uncultured eukaryote	18S rRNA	FN867270.1	3,00E-018	73	100	91
					Eukaryota; environmental samples: uncultured eukaryote.	18S rRNA	FN867270.1	4,00E-012	56	94	92
					Eukaryota; environmental samples:Uncultured freshwater eukaryote.	18S rRNA	AY919722.1	2,00E-014	46	100	100
					Eukaryota; environmental samples: uncultured eukaryote.	Genomic DNA	HQ230179.1	1,00E-021	59	100	100
					Eukaryota; environmental samples: uncultured eukaryote.	18S rRNA	FN867270.1	4,00E-017	54	100	98
					Eukaryota; environmental samples: uncultured eukaryote.	18S rRNA	AB273934.1	7,00E-022	67	97	96
					Eukaryota; environmental samples: uncultured eukaryote.	18S rRNA	FN862490.1	6,00E-020	56	100	100
					Eukaryota;environmental samples: uncultured Fungi;Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae;Cercomonas sp.	Genomic DNA	EU516994.1	6,00E-025	68	100	98
					Eukaryota; uncultured freshwater eukaryote environmental samples.	18S rRNA	AY919786.1	2,00E-023	62	100	100
					Eukaryota; uncultured eukaryote and freshwater eukaryote(environmental samples).	18S rRNA	FN389650.1	1,00E-020	57	100	100
					Eukaryota; environmental samples.	18S rRNA	AB273934.1	1,00E-020	66	90	98
Kinetoplastida		Trypanosomatidae	Trypanosoma	*Trypanosoma sp.*	Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;Trypanosoma sp.	18S rRNA, ITS1, 5.8S	FJ712718.1	2,00E-18	54	100	100

					Eukaryota; Rhizaria; Cercozoa; Cercomonadida; unclassified Cercomonadida	28S rRNA	FJ032646.1	1,00E-10	61	100	89
Kinetoplastida		Trypanosomatidae			Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;Leishmania spp.; Trypanossoma spp.; Blastocrithidia sp.	24S alpha rRNA; 18S rRNA; ITS1; 5.8S rRNA; ITS2; 28S rRNA	AM502245.1	5,00E-24	82	100	92
Chlorophyceae	Chlorococcales	Chlorococcaceae	Chlorococcum	*Chlorococcum dsp.*	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Chlorococcales; Chlorococcaceae; Chlorococcum	28S rRNA	AF183470.1	1,00E-004	95	29	100
Prasinophyceae	Pyramimonadales		Pyramimomas	*Pyramimomas sp.*	Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Pyramimonadales; Pyramimonas	28S rRNA	EU375499.1	3,00E-014	51	100	96
					Eukaryota; stramenopiles; Synurophyceae; Synurales; Mallomonadaceae; Mallomonas; Eukaryota; stramenopiles; PX clade; Phaeophyceae; Laminariales; Lessoniaceae; Lessonia sp. ; Eukaryota; stramenopiles; PX clade; Phaeophyceae; Fucales; Seirococcaceae; Cystosphaera sp.	28S rRNA	FJ030885.1	8,00E-011	50	96	93
					Eukaryota; environmental samples	18S rRNA	GU474399.1	2,00E-021	58	100	100
Proteomyxidea	Reticulosida	Gymnophryidae	Gymnophrys	*Gymnophrys sp.*	Eukaryota; Rhizaria; Cercozoa; Gymnophrys sp.	28S rRNA	FJ973379.1	7,00E-006	53	76	100
Heterolobosea	Schizopyrenida	Vahlkampfiidae	Naegleria	*Naegleria gruberi*	Eukaryota; Heterolobosea; Schizopyrenida; Vahlkampfiidae; Naegleria gruberi	rDNA plasmid DNA	AB298288.1	2,00E-015	53	94	98
Kinetoplastida		Bodonidae	Neobodo	*Neobodo designis*	Eukaryota; Euglenozoa; Kinetoplastida; Bodonidae; Neobodo designis	18S rRNA	DQ207585.1	9,00E-020	55	100	100