Metagenomics Accelerates Species Discovery and Unravel Great Biodiversity of Benthic Invertebrates in Marine Sediment in Campos Basin, Brazil

Milena MDP Schettini, Raony G C C L Cardenas, Marcella A A Detoni, Mauro F Rebelo.

Instituto de Biofísica Carlos Chagas Filho. Universidade Federal do Rio de Janeiro. Rio de Janeiro, Rio de Janeiro. Brasil.

ABSTRACT

Sediment fauna characterization and monitoring are mandatory requirements for obtaining oil and gas (O&G) environmental licensing for exploration and production (E&P) activities. Currently, for environmental characterizations and monitoring, biodiversity is assessed through morphological taxonomy, a time-consuming process. Taxonomists are constantly failing to meet the demands for biodiversity assessment required in monitoring programs. Thus, we combined three different phylogenetic markers (rRNA 18S, rRNA 28S and COI), HTS and Bioinformatics to identify benthic invertebrate organisms from sediment samples collected in five stations in the Campos Basin in southeast Brazil, an important oil extraction area and one of the best-studied marine biota in Brazil. Our results obtained with metagenomics were compared to morphology data provided by the Habitats Project whereas the database *Global Biodiversity Information Facility* was used for organism localization. We obtained around 4.83 µg of DNA from 15 samples. A total of 3.3 million sequences were clustered in Operational Taxonomic Units and more than 1.6 million sequences (about 50% of all reads) were assigned to 957 prokaryotes and 577 eukaryotes. BLAST identified 23 phyla, 60 classes, 62 orders, 70 families, 67 genus and 46 species of eukaryotes. Our metagenomic analysis identified phyla that are traditionally found in samples of marine benthos, such as Annelida, Arthropoda, Mollusca and Chordata, as well as more rarely found phyla such as Bryozoa, Cnidaria, Echinodermata, Nematoda, Nemertea, Platyhelminthes, Porifera and Priapulida; and even more rare phyla like Entoprocta and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v1 | CC-BY 4.0 Open Access | rec: 3 Jun 2016, publ: 3 Jun 2016

Gastrotricha. The low availability of genetic markers for Brazilian species in Genbank impaired our ability to compare our findings with those obtained morphologically for which no sequences were found in Genbank. Our study shows that metagenomics can be applied for environmental characterization and monitoring programs and, with the possibility of automating the method, may reduce from years to few months the time currently required for species identification and biodiversity determination, which will certainly accelerate species discovery.

INTRODUCTION

Sediment fauna characterization and monitoring are mandatory requirements for obtaining oil and gas (O&G) environmental licensing for exploration and production (E&P) activities. This requirement is expected to remain a key element of environmental management in the future, particularly in the frontiers of deep-sea offshore oil exploration areas: the Equatorial Margin and Santos Pre-salt Basin in Brazil, or the Barents and Siberia seas in the Arctic Ocean.

Currently, for environmental characterizations and monitoring, biodiversity is assessed through morphological taxonomy, a time-consuming process. As a general rule, taxonomic resolution at species level is expected and for some fauna groups, the expertise required is so unique that only a hand full of individuals in the world is fit for the task. Still, expert judgment is never 100% accurate, with only 50% rate of identification success being shared among taxonomists (Culverhouse et al., 2003). At last, fragile organisms that require special fixation procedures may not be properly represented in the samples (Costa-Paiva; Paiva e Kautau, 2007). As a result, invertebrate morphological identification efforts are often limited to few groups, including Mollusca, Crustacea and Polichaeta (MMA, 2015) and some estimates suggest that more than 90% of all marine species have never been named (SCHEFFERS et al, 2012).

The typical number of sediment samples in a monitoring campaign is in the range of tenths, but new areas to be explored can be as large as 300.000 km2, which can result in tenths of thousands of samples for baseline environmental characterization. Taxonomists are constantly failing to meet the demands for biodiversity assessment required in monitoring programs. The lack of experts is a major bottleneck in the process of identifying biodiversity (HEBERT et al, 2003; MORA; ROLLO; TITTENSOR, 2013), which delays operators execution of E&P projects to reach 'first oil' and keep species from being identified.

In Brazil, which, according to the latest Report of the Convention on Biological Diversity (CBD, 2016), is the most biologically-diverse country, with more than 100,000 animal species been accounted for, only 184 marine invertebrates had their conservation status accessed (MMA, 2015). It is possible that current risk estimates of environmental impact are based on underestimated biodiversity inventories, representing a threat to species conservation.

Developing new technologies and approaches that accelerate species discovery and reveal hidden biodiversity is crucial for setting conservation priorities and efforts.

Molecular methods use big data generated through high-throughput sequencing (HTS), which greatly accelerates species discovery. This approach is particularly useful for marine sediment analyses because the higher possibility of identifying minute organisms belonging to groups such as Nematoda, Copepoda, Ostracoda, Rotifera, Kinohyncha, Loricifera, Tardigrada and of species from debris and other sorts of environmental DNA (WANG et al, 2014), if compared to morphology. To classify eukaryote organisms using DNA-based approaches, and which have not yet been described morphologically, the concept of Operational Taxonomic Unit (OTU) can be applied (Schmidt; Mafias Rodrigues; Von Mering, 2014; Stackebrandt; Goebel, 1994).

Since 2010, more than 600 papers have been published on the use of DNA-based identification methods for species conservation (Bergman et al., 2016; Goldberg et al., 2014), biodiversity inventory determination (Drummond et al. 2015); environmental monitoring (Bowman et al., 2014; Brown et al., 2015; Chariton et al., 2015; Leray et al., 2015), DNA extraction/detection (Eichmiller et al., 2014; Pedersen MW et al., 2015; Ficetola et al., 2016) and the technique has been considered a major tool for Ocean's sustainability in the 21st century (Aricó, 2015).

In this study, we combined three different phylogenetic markers (rDNA 18S, rDNA 28S and COI), HTS and Bioinformatics to identify benthic invertebrate organisms with metagenomes from sediment samples collected in Campos Basin in southeast Brazil, an important oil extraction area and one of the best-studied marine biota in Brazil (MILOSLAVICH et al, 2011).

Material and Methods

Sample collection and processing:

Samples were collected at Campos Basin in 2009 as part of 'Habitats Project – Campos Basin Environmental Heterogeneity' coordinated by CENPES/PETROBRAS. Table 1 presents information (collection date, geographic coordinates and depth) on the five sampling stations: B3, B4, C2, G2 and F5. Sediment samples were collected in triplicate, descending a Van Veen grab in three different points around (150 m radius) each of the five stations, totaling 15 sediment samples. At the time these samples were collected, no plans to have them genetically analyzed had been set. Thus, they were kept at -20°C for 4 years until our analysis was done in 2013. For each station, we manually homogenized 200 cm^3 of the muddy sediments and weighted 5g for DNA extraction that was performed using the PowerMax Soil DNA Isolation (MoBio Inc), according to manufacturer's instructions. DNA integrity was accessed by means of agarose gel 1.2%. Quantification was performed in Qibit 2.0 Fluorometer (Life Technologies).

Biogeography data:

Data on the organisms identified in this study were extracted from two main sources: the book entitled "Biodiversidade bentônica da região central da Zona Econômica Exclusiva brasileira" by Lavrado and Ignacio (2006) for the Cnidaria Crustacea, Echinodermata, Mollusca, Nematoda, Polychaeta and Porifera groups, whereas the dada for organisms of the phyla Annelida, Arthropoda, Brachiopoda, Bryozoa, Cnidaria, Echinodermata, Echiura, Foraminifera, Haptophyte, Mollusca, Nematoda, Nemertea, Porifera, Priapula, Protozoa, Rodophyta were identified by the Habitats Project and provided by Petrobras S.A. (unpublished data). We also used the database *Global Biodiversity Information Facility* (www.gbif.org) for organism localization.

In this study, we chose family as the taxonomic group to be used as reference in cladograms in order to be able to compare our findings with those provided by morphological taxonomy. Whenever species descriptions were available for both metagenomic and morphological approach, they were also discussed.

PCR and high-throughput sequencing:

Information on PCR of COI, rDNA 18S and rDNA 28S genes is presented in Supplemented material 1. We used the kit *Ion Xpress*™ Plus *Fragment Library* (Life Technologies) for preparing the libraries for sequencing according to manufacturer's instructions of *Ion Xpress*™ *Plus gDNA Fragment Library Preparation.* Template preparation and sequencing were done using the kit Ion PGM™ Template OT2 400. Sequencing was done using the *Ion Personal Genome Machine (PGM™) System* at the Life Technologies laboratories (São Paulo, SP), using *Chip* 318 v2. Sequencing adapters were removed from reads using Torrent Suite *software* version 4.0.2 (Life Technologies) and assigned to samples based on the combination primer tail-Ion Xpress barcode. Prinseq version 0.20.4 (SCHMIEDER; EDWARDS, 2011) was used to remove either A/T photopolymers bigger than 5 bases, reads with unidentified (N) bases, small length (<80bp) or bad quality reads (Q<20). Remaining reads were clustered in OTUs using CD-HIT-EST version 4.6 (LI; GODZIK, 2006) (up to 97% identity under 100% coverage within a bigger read, word size of 10 and 20 penalty points for gaps).

High quality and low redundancy sequences were compared to NCBI non-redundant nucleotide repositories (NR) (http://www.ncbi.nlm.nih.gov/genbank/) using *Basic Local Alignment Search Tool nucleotides* (BLASTn) version 2.3.0+ (Zhang et al, 2000). Max *e-value* was of 10^{-5} and the number of events per query was limited to 100 (here called as *hits*).

Taxonomic names were attributed to each *read,* based on the reads group of BLAST hits, using the 'Lowest Common Ancestor Assignment – LCA' algorithm in software MEGAN (MEta Genome Analyzer; version 5.10.3; Huson et al., 2007) according to different parameters (Huson et al., 2011). Cladograms and rarefaction curves at family taxonomic level for each station were also built using MEGAN.

The BLAST step was performed using the Elastic Compute Cloud (EC2) service of Amazon (aws.amazon.com). The BLAST for each of the 15 sets of reads correspondent to the 15 samples, run in a parallel scheme using eight threads on up to 96 AWS instances with 8 processors and 16 Gb of RAM each.

RESULTS & DISCUSSION

We obtained an average of 4.83 µg of DNA from each of the 15 samples. Sequencing generated approximately 4.8 million sequences, which is within the expected values for the 318 v2 chip, but with an average size of 155.1 bp, which is bellow the expected value for the OT2 400 kit. Over 3.6 million sequences (75.35%) passed quality control and of these, around 3.3 million were clustered in Operational Taxonomic Units by CD-HIT (Table 2). More than 1.6 million sequences (about 50% of all reads) were assigned to 957 prokaryotes and 577 eukaryotes using BLAST (Table 2). BLAST identified 23 phyla, 60 classes, 62 orders, 70 families, 67 genus and 46 species of eukaryotes (Supplementary Material 2 – Cladograms and Supplementary Material 3 – list of species identified). Figure 1 shows the rate of OTU observed by metagenomics in each of the stations distributed over the 13 invertebrate phyla (Figure 1A) and 38 invertebrate families (Figure 2B). All other Prokaryota and non-invertebrate Eukaryota phyla observed in this study are listed in the cladograms available in the supplementary material. A considerable number of reads were assigned to the families Hominidea and Bovidae, increasing the number of reads belonging to the Chordate phylum. However, these were read alignments generated against the whole human and bovine genomes or chromosomes, as opposed to the three genetic markers that we used in this study. Our results and discussion are focused on invertebrate families belonging to marine benthos and no artifact findings on chordate will be further addressed.

One of the differentials of our study was that it was done using samples collected from the actual areas were E&P activities are usually carried out. Several previous morphological taxonomic studies were performed in these areas, either by the oil companies interested in obtaining their licenses, or those involved in conservational programs (such as the Habitats Project) or by the scientific community (the REVIZEE program).

The huge taxonomic effort of the Habitats Project generated a databank of 49,289 specimens. A total of 17 phylum, 27 classes, 63 orders, 354 families, 768 genus and 749 species were identified.

Out of the 1,773 macroinvertebrates *taxa* identified by morphological taxonomy, 1,211 or 68% did not have any entry in Genbank found for any of the three markers (COI, rRNA 18S e 28S) used in this study, indicating that Brazilian marine species are underrepresented in Genbank. Thus, there is a need to increase efforts to have sequences from these three molecular markers from more Brazilian species deposited in Genbank, as the limited number of sequences impairs any parallel to be done between the findings obtained with molecular and those obtained with morphological taxonomies.

Our metagenomic analysis identified phyla that are traditionally found in samples of marine benthos, such as Annelida, Arthropoda, Mollusca and Chordata, as well as more rarely found phyla such as Bryozoa, Cnidaria, Echinodermata, Nematoda, Nemertea, Platyhelminthes, Porifera and Priapulida; and more rare phyla like, Entoprocta and Gastrotricha (Supplementary material and Figure 1).

The great number of OTUs for Annelida, Arthropoda and Mollusca found by metagenomics agrees with previous results for Campos Basin found by LAVRADO; IGNACIO, 2006 during the REVIZEE project and also by those of the Habitats Project. Recent metagenomics study carried out by Leray and Knowlton (2015) also identified Annelida and Arthropoda as the phyla with more OTUs among the 22 phyla identified from approximately 0.09 m^3 sediments from coral reef regions in Virginia and Florida, in the United States.

The Entoprocta (or Kamptozoa) phylum comprises about 170 aquatic and sessile species of sizes between 0.5 and 5.0 mm and are mostly marine (Zhang, 2011). Until 2011, only 18 species of Entoprocta were known on the Brazilian coast (Vieira; Migotto, 2011). In this study, all OTUs (6 in the C2 station and 24 in the G2 station) were attributed to the genus *Loxosomella* through the marker rDNA 28S, with over 86% of sequence similarity. This result expands the distribution of the genus that was previously limited to six species collected off the coast of São Paulo (VIEIRA; MIGOTTO, 2011).

As for the cosmopolitan Gastrotricha phylum that comprises about 790 species of aquatic organisms up to 1 mm in length (Zhang, 2011), all 22 OTUs assigned to the phylum (C2 station) were in the *Tetranchyroderma* genus, with over 81% similarity with COI sequences found in the Genbank. This occurrence also expands the distribution that had been previously limited to São Paulo beaches (reported but not formally described – Garraffoni; ARAUJO, 2010), almost a 1000km away from the Campos Basin.

This is a pioneer study in which metagenomics results could be compared to those from a recent comprehensive morphological taxonomy effort that worked with the same samples than those used in our study. However, comparing results between studies should be taken with caution because of the uncertainty on how much DNA is still available considering that samples have been preserved at -20ºC for 40 years and the lack of available genetic markers for the Brazilian marine species in the Genbank. It should also be noted that we analyzed 5g out of 200 gr of the surface (0 to 2 cm) sediment for each of the 15 samples, while the morphological study worked with 1000 cm^3 of sediment from each sample, comprising slices from 0 to 10 cm. Finally, for many species, the sequences of the markers available in Genbank were partial and thus we cannot ensure they properly aligned with the reads to attribute a taxonomic name. However, these restrictions applies mostly to the families that we did not found and we believe that observations made about the families that we actually found are valid.

Our analysis identified 38 families of invertebrates in the 15 samples from the 5 sampling stations in Campos Basin. Figure 2 compares between the families from Annelida (9 families, fig. 2A), Arthropoda (10 families, fig. 2B) and Mollusca (7 families, fig. 2C) phyla identified by metagenomics and morphology taxonomy in stations B3, B4, C2, F5 and G2.

Annelida families Amphinomidae, Enchytraeidae, Glyceridae, Orbiniidae, Serpulidae and Spionidae were previously identified in Campos Basin by the Habitats Project while up to 28 annelida families previously reported by the Habitat project could not be identified by metagenomics. Family Hormogastridae found in our study is most likely a false positive since it is not marine. The Arthropoda families Solenoceridae, Cylindroleberididae and Mysidae have been previously identified in Campos Basin and in the Southeast of Brazil by other authors (CARDOSO; SEREJO, 2007; GBIF, 2016; SEREJO et al, 2007; TÂMEGA; OLIVEIRA; FIGUEIREDO, 2013) while up to 29 arthropoda families previously reported by the Habitat Project were not be identified by metagenomics. Families Miridae, Chalcididae and Formicidae found in our study are most likely false positive since they are not marine insects. All Mollusca families identified by metagenomics in Campos basin, except for Mytilidae, have not been identified by the Habitat Projects, even though they have been previously found in the region (DORNELLAS; SIMONE, 2011; LAVRADO; IGNACIO, 2006; TÂMEGA; OLIVEIRA; FIGUEIREDO, 2013). Up to 15 mollusca families previously reported by the Habitat Project could not be identified by metagenomics. Moreover, metagenomics was also able to find several families not previously reported by morphological taxonomy for a given station, suggesting that the family distribution could be broader than anticipated. That is the case for Echiuridae, Hormogastridae and Pectinariidae among the Annelidae; Desmosomatidae and Hippolytidae for Arthropoda and Arcidae, Mactridae and Pectinidae for Mollusca.

The Habitat Projects identified 749 organisms to the species level but only 64 had at least one sequence of one of the three genetic markers (COI gene, 18S rRNA and 28S here studied) deposited in Genbank and thus were 'eligible' for molecular identification. At first, metagenomics identified 46 species. However, none of the 64 species previously identified by morphological taxonomy by the Habitat Project were found by metagenomics. We believe that these are false negative results that can be explained by the fact that samples were preserved at -20^{o}C for 4 years, by the low sample volume and the fact that the genetic markers here studied were missing in Genbank for a number of organisms that were identified by morphology.

However, we noticed that even after calibration of the parameters for the LCA algorithm (data not shown), some incongruence in the attribution of the name of the species had happened. To overcome that limitation, we manually searched for the 64 species names of found by the Habitat Project among the names of the organisms generated by the BLAST hits for a given read. We were able to identify more 45 species that had been previously described by morphological taxonomy but were not picked by the LCA algorithm. The full list of species identified by molecular and morphological taxonomies, together with the genetic markers available in Genbank are listed in supplementary material 3 (or table). Other false negative results could have been generated by the occurrence of synonymous names at the species level. For instance, according to recent estimates, more than 80% of the algae of some genus and 38% of mollusca have synonymous names. For marine species, this figure would reach 40% (COSTELLO; MAY; STORK, 2013). An ongoing effort is dedicated to resolve synonymous names found in the GBIF database.

Of the 46 invertebrate present in cladograms leaves (most specific possible position) that we identified by the molecular taxonomy, 27 were invertebrates not previously described in the region. These could represent new occurrence in Campos Basin. Because description in the main biogeographic databases that we used (Habitats Project, Revizee and GBIF) usually goes no further than the family taxonomic level, it is not possible to either claim or rule out that the finding corresponds to the first description of the species in the region.

However, we wanted to calculate the likelihood that those events truly represented false positive results, as oppose of being descriptions of new species. False positive results could happen as an artifact due to similarities of genetic sequences shared among species belonging to the same genus and the low representativeness of Brazilian species in Genbank. The high similarity could have led BLAST to relate, with very low error probability, a read from one species not present in the databank to another present in the Genbank and from the same genus (phylogenetic similarity) but belonging to a completely different habitat. By using metadata on the distribution of the species selected by BLAST, we managed to sort out at least one case among our results. *Haliotis diversicolor*, which was identified in our study, is a small (25-85 mm) gastropod form the Indo-Pacific Ocean, with georeferenced records on the coast of Japan, Thailand, Australia, among other countries in the region (GBIF, 2016). Despite the geographical distance, *Haliotis diversicolor* shares high sequence similarity with *H. aurantium*, which has been identified in the Campos basin, and also with three other records corresponding to species found in the Brazilian coast. The lack of genetic markers for these Brazilian species in Genbank may have misled BLAST searches, which in this case erroneously classified the sequence of *H. aurantium* as of *Haliotis diversicolor.*

To further remove false positive results, we wanted to find redundant identification done by each of the three genetic markers for each of the 46 species found by molecular taxonomy, hoping that a doubtful identification by one marker could be resolved by a positive confirmation by the other two. Unfortunately, that was not the case. Out of the 64 species identified by the Habitat Project, 16 species had sequences of all three markers available in Genbank and still were not positively identified by metagenomics. Out of the 46 species identified by molecular metagenomics, other 16 had sequences of all three genetic markers available in Genbank, but by metagenomics they were identified only by one of the three markers and never by two or three. We noticed that many times, even though the sequence for a genetic marker for a specific organism was available in the Genbank, multiple names were attributed to the gene, only partial sequences were available, or sequences were not validated experimentally. Genbank is the best repository for genetic sequences yet available but still does not offer a high level of confidence when it comes to the names attributed to genetic sequences. Our research team is currently working on developing new algorithms to help overcome this limitation.

The problems related with having false positive and false negative results and with the occurrence of synonymous names could be solved if we work only at the level of OTU to compare taxa profile among samples seasonally. The frequency and abundance of OTUs could then be related to environmental changes and could accelerate species discovery by showing that genetic sequences vary according to environmental conditions. Further studies should be done in which such strategy is adopted, as working with OTUs allows us to unravel the hidden biodiversity of the thousands of 'no hit' OTUs and to relate their distribution to environmental changes and activities.

CONCLUSION

Brazil has one of the strictest environmental laws and regulations for E&P activities of the O&G sector in the world. Recent changes were made under resolution CONAMA 422/11 that minimized bureaucracy required by the application process, increased transparency by sharing information online and reduced liability for the O&G operators. In Brazil, the environmental authority IBAMA (Brazilian Institute of the Environment and Renewable Natural Resources) is responsible for issuing 'reference terms' that establish the guidelines and best practices for the environmental licensing and monitoring.

Metagenomics can be applied for environmental characterization and monitoring programs and, with the possibility of automating the method, may reduce from years to few months the time currently required for species identification and biodiversity determination, which will certainly accelerate species discovery.

Nevertheless, the fact that 68% of the organisms identified by morphology did not have sequences of at least one of the three markers used in this study (COI, 18S and 28SrDNA) deposited in the Genbank illustrates how low is the representation of molecular markers of Brazilian marine species in the Genbank. Further studies should focus on sequencing organisms and have their sequences deposited in the Genbank and other international databases.

We believe that metagenomic identification based on species' DNA overcomes several of the limitations associated to morphological methodology. We have shown, as well as the studies done by others, that metagenomics is a reliable approach for the identification of biodiversity, that can be improved by adding more sequences of native species in public and proprietary databanks. It is our opinion that metagenomics consists of the best available technique for generating biodiversity inventories in marine sediments and should be acknowledged as such by oil operators, environmental authorities and the scientific community at large.

REFERÊNCIAS

Aricò S. 2015. "Ocean Sustainability in the 21st Century" report. UNESCO Publishing / Cambridge University Press . 324 pages.

Bergman PS, Schumer G, Blankenship S, Campbell E (2016) Detection of Adult Green Sturgeon Using Environmental DNA Analysis. PLoS ONE 11(4): e0153500. doi:10.1371/journal.pone.0153500

Bohmann et al., Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. 29, 358–367 (2014).

Brown, Emily A; Chain, Frédéric J J; Crease, Teresa J; MacIsaac, Hugh J; Cristescu, Melania E. 2015. Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton communities? Ecology and evolution vol. 5 (11) p. 2234-51

Cardoso, I. A.; Serejo, C. S. Deep Sea Caridea (Crustacea, Decapoda) from Campos Basin, Rj, Brazil. **Brazilian Journal of Oceanography**, v. 55, n. 1, p. 39-50, 2007.

CBD, 2016 - https://www.cbd.int/countries/?country=br

Chariton et al., Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environ. Pollut. 203, 165–174 (2015).

Costa-Paiva EE, Paiva PC and Klautau M. Anaesthetization and fixation effects on the morphology of sabellid polychaetes (Annelida: Polychaeta: Sabellidae). J. Mar. Biol. Ass. U.K. (2007), 87, 1127–1132.

Costello, M. J.; May, R. M.; Stork, N. E. Can We Name Earth's Species Before They Go Extinct? **Science**, v. 339, n. 6118, p. 413-416, jan. 2013.

Culverhouse et al.: Do experts make mistakes? A comparison of human and machine identification of dinoflagellates. Mar Ecol Prog Ser 247: 17–25, 2003.

Drummond et al. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. DOI: 10.1186/s13742-015-0086-1).

Eichmiller et al., 2014 The Relationship between the Distribution of Common Carp and Their Environmental DNA in a Small Lake. DOI: 10.1371/journal.pone.0112611)

Ficetola G. F., P. Taberlet, E. Coissac, How to limit false positives in environmental DNA and metabarcoding? Mol. Ecol. Resour. 16, 604–607 (2016).

Garraffoni, A. R. S.; Araújo, T. Q. Chave de identificação de Gastrotricha de águas continentais e marinhas do Brasil. **Papéis Avulsos de Zoologia**, v. 50, n. 33, p. 535-552, 2010.

GBIF - Global Biodiversity Information Facility. Disponível em: <http://www.gbif.org/>. Acesso em: 15 fev. 2016.

Goldberg C. S., K. M. Strickler, D. S. Pilliod, Moving environmental \{DNA\} methods from concept to practice for monitoring aquatic macroorganisms. Biol. Conserv. 183, 1–3 (2015).

Hebert, P. D. N. et al. Biological identifications through DNA barcodes. **Proceedings of the Royal Society of London Series B: Biological Sciences**, v. 270, p. 313-321, jan. 2003.

Huson, D. H. et al. Integrative analysis of environmental sequences using MEGAN4. **Genome Research**, v. 21, n. 9, p. 1552-1560, set. 2011.

Huson, D. H. Et Al. MEGAN analysis of metagenomic data. **Genoma Research**, p. 1-10, jan. 2007.

Lavrado, H. P.; Ignacio, B. L. (Eds.). **Biodiversidade bentônica da região central da Zona Econômica Exclusiva brasileira.** Rio de Janeiro: Museu Nacional - UFRJ, 2006.

Leray, M.; Knowlton, N. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. **Proc. Natl. Acad. Sci. USA**, v. 112, n. 7, p. 2076-2081, fev. 2015.

Li, W.; Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**, v. 22, p. 1658-1659, 2006.

Miloslavich - Patricia Miloslavich1*, Eduardo Klein1, Juan M. Dı ´az2, Cristia ´n E. Herna ´ndez3, Gregorio Bigatti4, Lucia Campos5, Felipe Artigas6, Julio Castillo1, Pablo E. Penchaszadeh7, Paula E. Neill8, Alvar Carranza9, Marı ´a V. Retana4, Juan M. Dı ´az de Astarloa10, Mirtha Lewis4, Pablo Yorio4,11, Marı ´a L. Piriz4, Diego Rodrı ´guez10, Yocie Yoneshigue-Valentin5, Luiz Gamboa12, Alberto Martı ´n1. Marine Biodiversity in the Atlantic and Pacific Coasts of South America: Knowledge and Gaps. PLoS ONE, January 2011 | Volume 6 | Issue 1 | e14631.

MMA - Ministério do Meio Ambiente. Fifth National Report to the Convention on Biological Diversity: Brazil. Brasília: Ministry of the Environment, 2015.

Mora, C.; Rollo, A.; Tittensor, D. P. Comment on "Can We Name Earth's Species Before They Go Extinct?" **Science**, v. 341, n. 6143, p. 237, 2013.

Pedersen M. W. et al., Ancient and modern environmental DNA. Philos. Trans. R. Soc. London B Biol. Sci. 370 (2014), doi:10.1098/rstb.2013.0383.

Scheffers, B. R. et al. What we know and don't know about Earth's missing biodiversity. **Trends in Ecology and Evolution**, v. 27, n. 9, p. 501-510, set. 2012a.

Scheffers, B. R. et al. Erratum to: "What we know and don't know about Earth's missing biodiversity". **Trends in Ecology and Evolution**, v. 27, n. 12, p. 712-713, dez. 2012b.

Schmidt T. S., Matias Rodrigues J. F., Von Mering C. Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. **PLOS Computational Biology**, v. 10, n. 4, p. 1-10, abr. 2014.

Schmieder, R.; Edwards, R. Quality control and preprocessing of metagenomic datasets. **Bioinformatics**, v. 27, p. 863-864, 2011.

Serejo, C. S. et al. Abundância, diversidade e zonação dos crustáceos no talude da costa central do Brasil (11° - 22° S) coletados pelo Programa REVIZEE/Score Central: prospecção pesqueira. In: COSTA, R. A. S.; OLAVO. G.; MARTINS. A.S. (Eds.) **Biodiversidade da fauna marinha profunda na costa central brasileira**. Rio de Janeiro: Museu Nacional, 2007, p. 133-162.

Stackebrandt, E.; Goebel, B. M. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. **International Journal of Systematic Bacteriology**, v. 44, n. 4, p. 846-849, out. 1994.

Tâmega, F. T. S.; Oliveira, P. S.; Figueiredo, M. A. O. (Eds.) **Catalogue of the Benthic Marine Life from Peregrino Oil Field, Campos Basin, Brazil**. Rio de Janeiro: Instituto Biodiversidade Marinha, 2013.

Vieira, L. M.; Migotto, A. E. Entoprocta Checklist of the State of São Paulo. **Biota Neotrop**., v. 11, suplemento 1, p. 497-501, 2011.

Wang, Y., et al. Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis. **PLoS ONE**, v. 9, n. 3, 2014.

Zhang, Z.-Q. (ed.) Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. **Zootaxa**, v. 3148, p. 1–237, 2011.

Zhang Z. et al. A greedy algorithm for aligning DNA sequences. **J Comput Biol.**, v. 7, n. 1-2, p. 203-214, 2000.

Table 1 – Sampling date, location and depth Location of sampling stations B3, B4, C2, F5 and G2 in Campos Basin, southeast Brazil.

	Sampling date	Latitude (SIRGAS2000)	Longitude (SIRGAS2000)	Depth (m)
Station B3	02/20/2009	-22,997011	-41,352583	77
Station B4	02/21/2009	-23,16851	-41,052264	107
Station C2	07/16/2009	-22,625989	-41,365082	54
Station F5	02/24/2009	-22,290999	-40,110584	143
Station G2	02/25/2009	-21,98502	-40,419918	56

Suppelementar Material 1 – PCR primers and conditions. 1-5 µL of DNA template, 1 µL (5µM) of primers Forward and reverse), 5 µl of 10X buffer, 2 µl of $MgCl_2$ (25 mM), 1 µl of dNTP 10 µM (Fermentas), 0.2 µl de Platinum® Taq DNA Polymerase High Fidelity 5 U.µL-1 (Thermo Scientific) and ultra pure destilaed water (Invitrogen) to complete 50 µl final reaction volume.

Target	Primer (F – Forward; R – reverse)	Denaturation	cycles	denaturation	anealing	Extension	Final extension	References
COI	TITCIAAYCAYAARGAYATTGG (F – jLCO1490); TAIACYTCIGGRTGICCRAARAAYCA (R – jHCO2198)	1' @94oC	10+30	30"@94oC	1'30"@61-52oC (-1oC per cycle) + 1'30"@61-52oC	1'@72oC	5'@72oC	Geller et al., 2013
rRNA 18S	ATGGTTGCAAAGCTGAAC (F – a2.0); GATCCTTCCGCAGGTTCACCTAC (R- 9R)	2' @94oC	40	30"@94oC	30'@55oC	1'@72oC	5'@72oC	Whiting et al., 1997; Whiting, 2002
rRNA 28S	ACCCGCTGAATTTAAGCAT (F – C1'); TGAACTCTCTCTTCAAAGTTCTTTTC (R- C2)	2' @94oC	40	30"@94oC	30'@55oC	1'@72oC	5'@72oC	Van Le et al., 1993; Chen et al., 2003

Table 2 – OTU per sample. OTU without a similar sequence on Genbank NR are under 'No Hits' fragments . OTU that did not comply with established LCA parameters (e.g. score bellow 100) or do not add up to a node are under 'non attributed reads'. Also under 'non-attributed' are Prokaryots attributed by rRNA16S, taxa attributed by genes other than the 3 targets and taxa defined at Genbank as 'undefined'. They were also disabled at the cladograms.

Sample	Total OTU	No Hits	Non attributed	Attributed
St. B3 rep. #1	101,966	20,505	73,653	7,808
St. B3 rep. #2	379,812	65,557	97,849	222,406
St. B3 rep. #3	84,180	12,167	57,290	14,723
St. B4 rep. #1	103,053	25,721	57,290	14,723
St. B4 rep. #2	332,953	35,384	64,066	236,503
St. B4 rep. #3	302,290	50,143	65,134	187,013
St. C2 rep. #1	245,233	34,452	40,687	170,094
St. C2 rep. #2	307,780	59,289	60,866	187,625
St. C2 rep. #3	249,969	56,247	81,114	112,608
St. F5 rep. #1	139,992	50,900	35,349	53,743
St. F5 rep. #2	105,435	32,435	47,684	25,316
St. F5 rep. #3	83,962	43,377	34,877	5,708
St. G2 rep. #1	173,740	71,230	60,632	41,780
St. G2 rep. #2	312,446	88,627	79,156	144,663
St. G2 rep. #3	347,494	32,832	120,519	194,143
TOTAL	3,270,206	678,866	959,986	1,631,453

Figure 1 – OTU occurence in each station. Percentage of OTU for phyla (A) and Family (B) in each station.

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Station B3
Station B4
Station C2
Station F5
Station G2
Priapulida
Porifera
Platyhelminthes
Nemertea
Nematoda
Mollusca
Gastrotricha
Entoprocta
Echinodermata
Cnidaria
Bryozoa
Arthropoda
Annelida

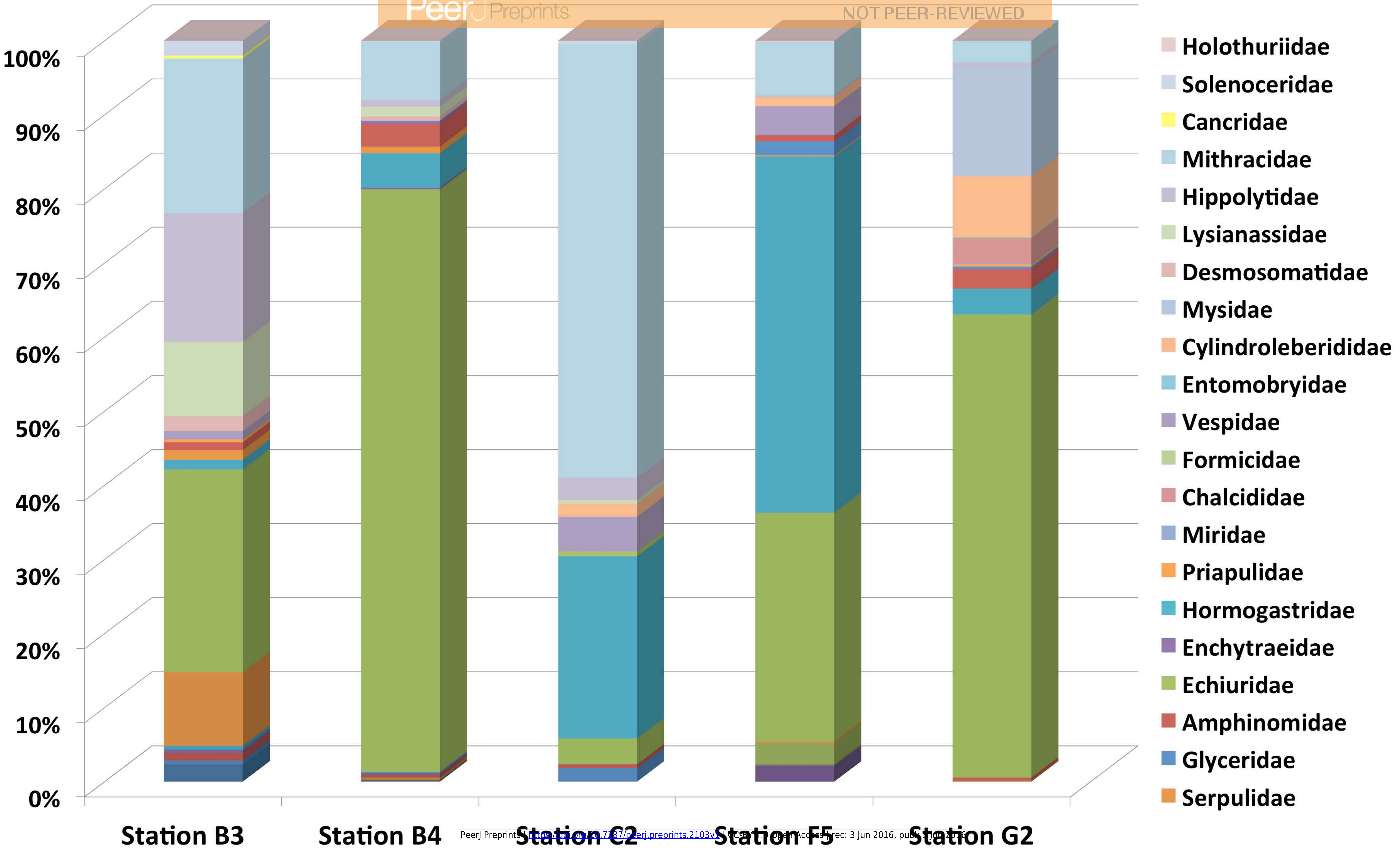
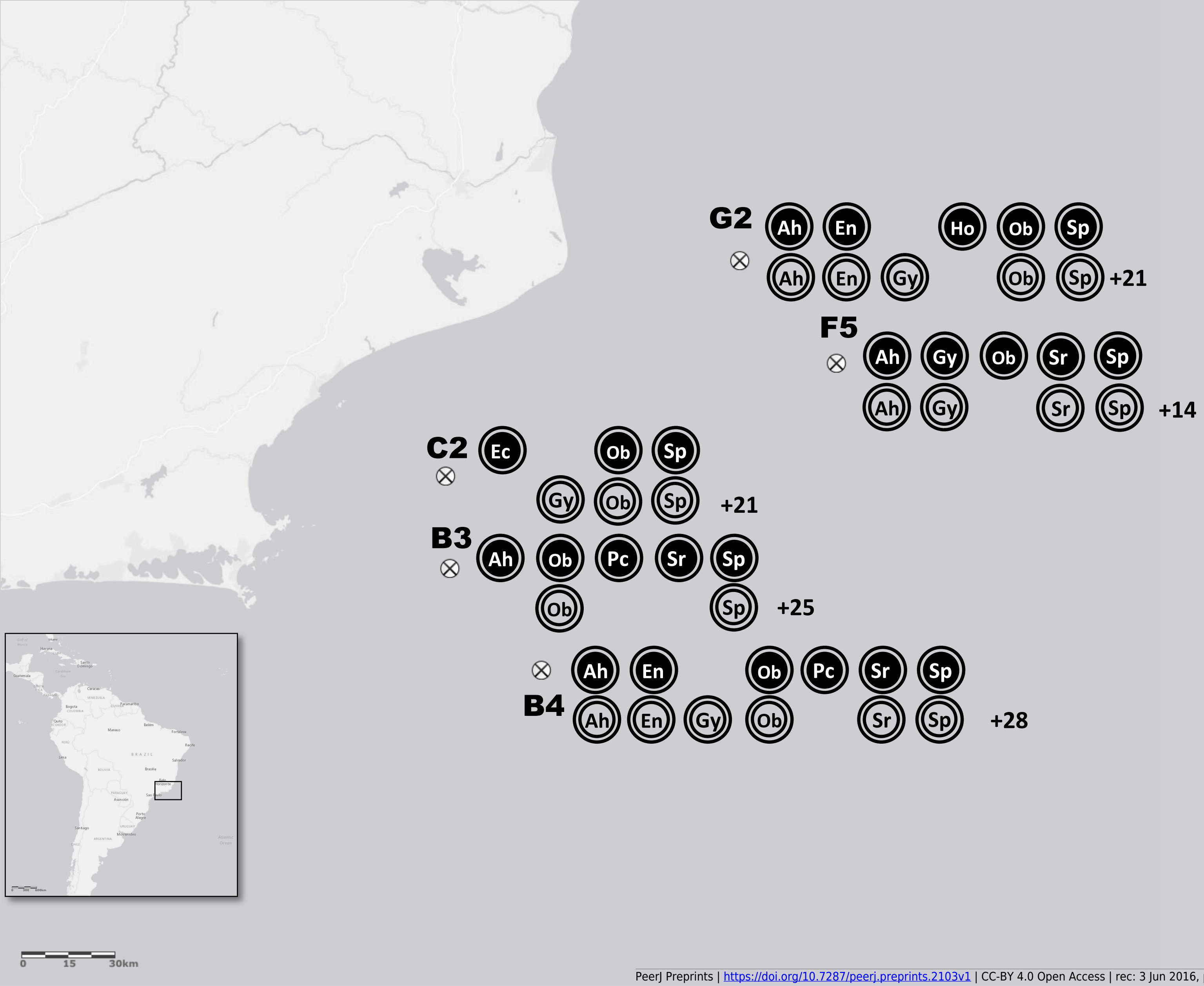

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Station B3
Station B4
Station C2
Station F5
Station G2
Holothuriidae
Solenoceridae
Cancridae
Mithracidae
Hippolytidae
Lysianassidae
Desmosomatidae
Mysidae
Cylindroleberididae
Entomobryidae
Vespidae
Formicidae
Chalcididae
Miridae
Priapulidae
Hormogastridae
Enchytraeidae
Echiuridae
Amphinomidae
Glyceridae
Serpulidae

Figure 2 – Distribution of the main invertebrate phylum identifyed by molecular and morphological taxonomy in Campos Basin. A) annelida distribution, b) arthropoda distribution, C) mollusca distribution.

Annelida distribution

Molecular | **Morphological**



Ah Ah **Amphinomidae**

En En **Enchytraeidae**

Ec Ec **Echiuridae***

Gy Gy **Glyceridae**

Ho Ho **Hormogastridae****

Ob Ob **Orbiniidae**

Pc Pc **Pectinariidae***

Sr Sr **Serpulidae**

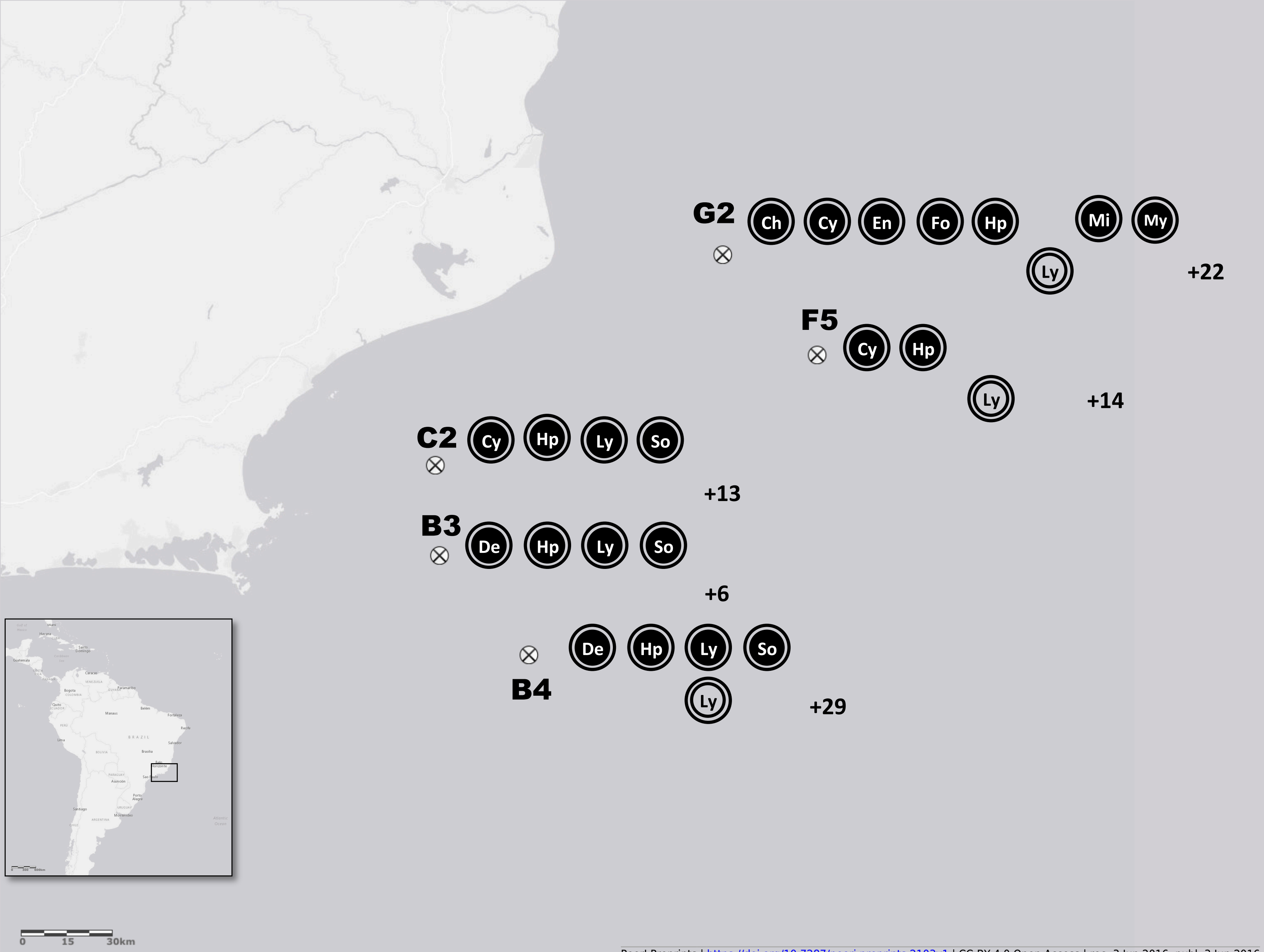
Sp Sp **Spionidae**

* Present in other stations of Habitats

** Non-marine family

Arthropoda distribution

Molecular **Morphological**



Ch	Ch	Chalcididae***
Cy	Cy	Cylindroleberididae**
De	De	Desmosomatidae*
En	En	Entomobryidae
Fo	Fo	Formicidae***
Hp	Hp	Hippolytidae*
Ly	Ly	Lysianassidae
Mi	Mi	Miridae***
My	My	Mysidae**
So	So	Solenoceridae**

* Present in other stations of Habitats Project

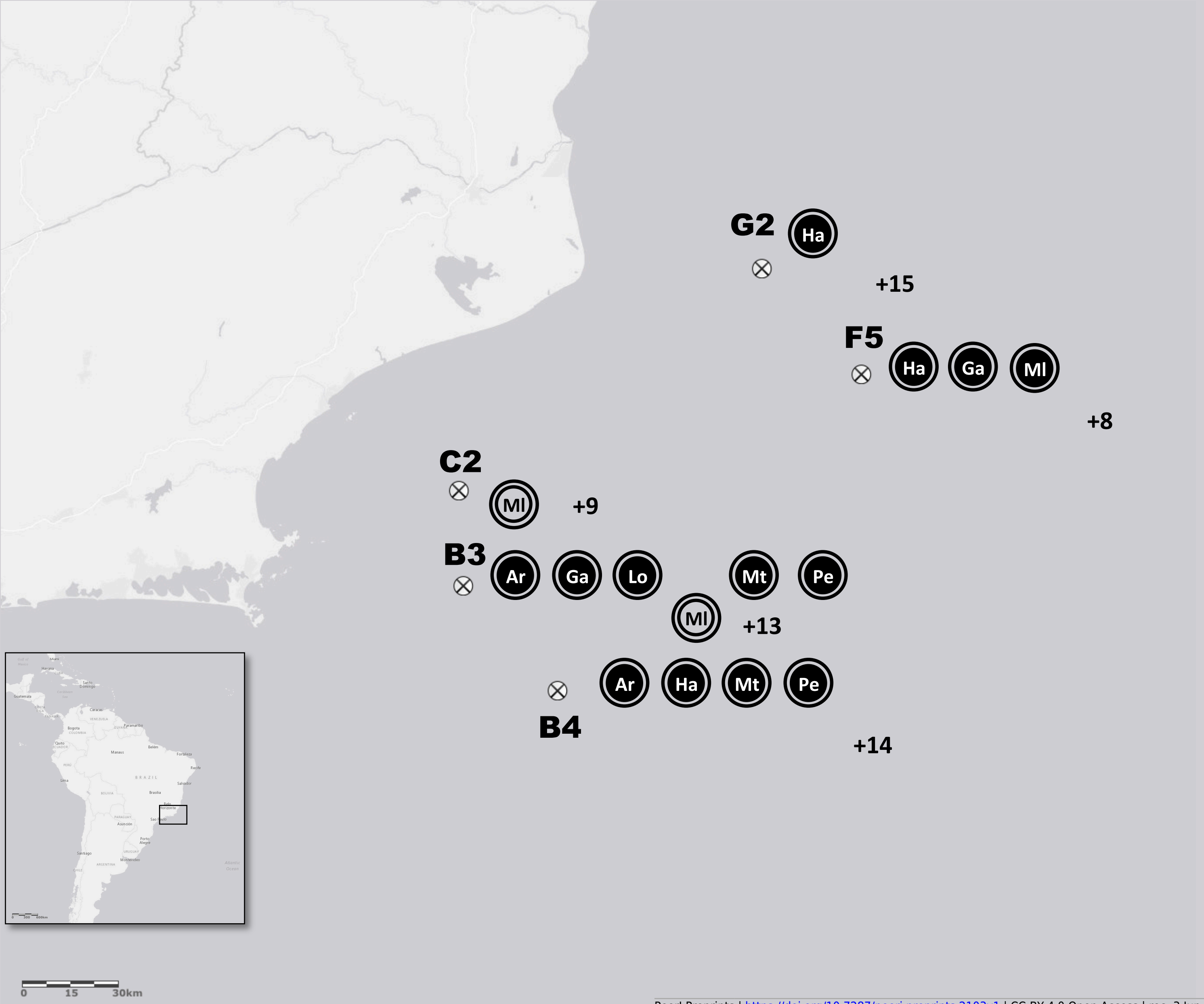
** Previous studies in Campos basin

***Non-marine family

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v1 | CC-BY 4.0 Open Access | rec: 3 Jun 2016, publ: 3 Jun 2016

Mollusca distribution

Molecular | Morphological



Ar Ar Arcidae* **

Ga Ga Galeommatidae**

Ha Ha Haliotidae**

Lo Lo Lottiidae**

Mt Mt Mactridae* **

Ml Ml Mytilidae

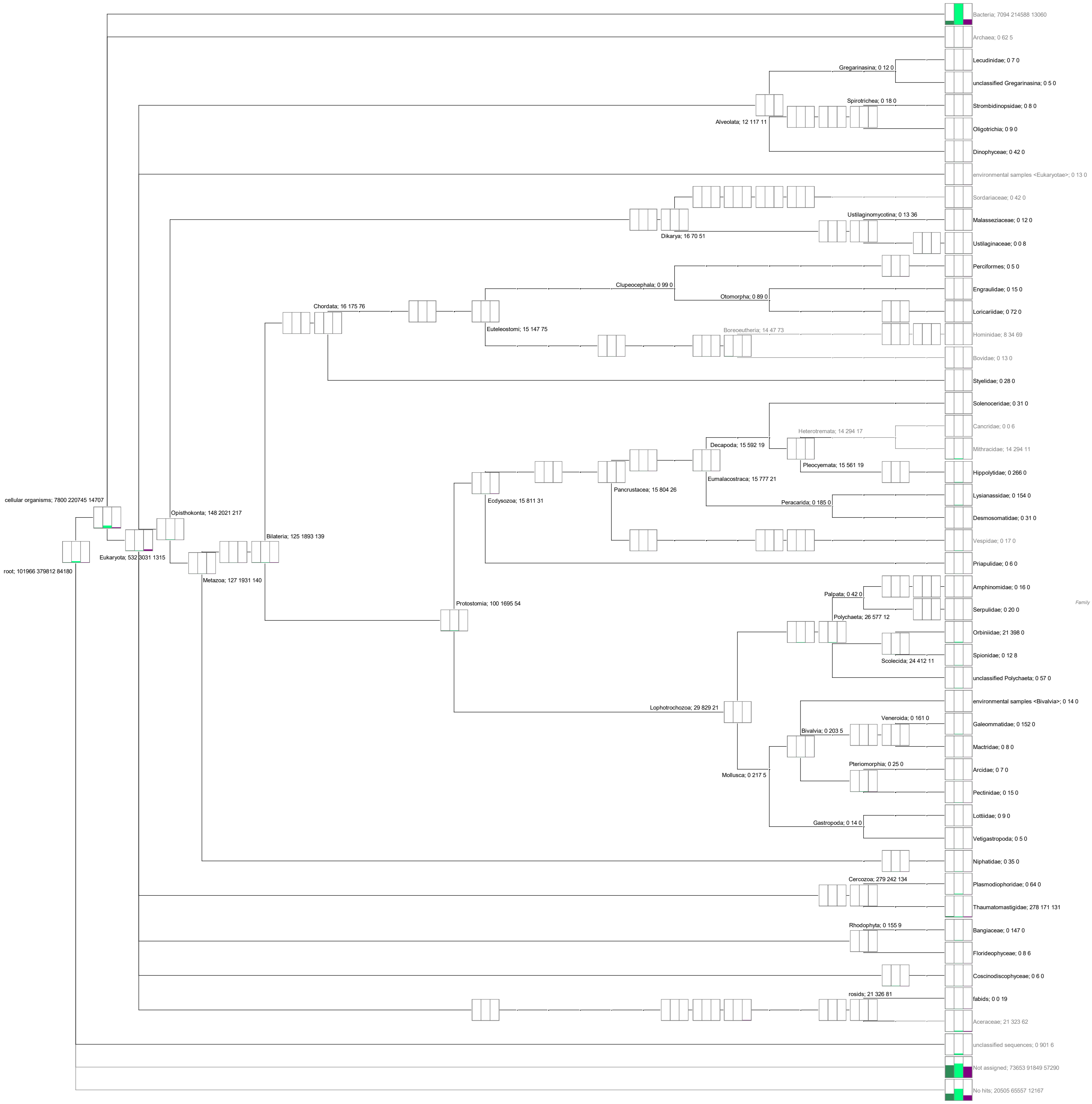
Pe Pe Pectinidae* **

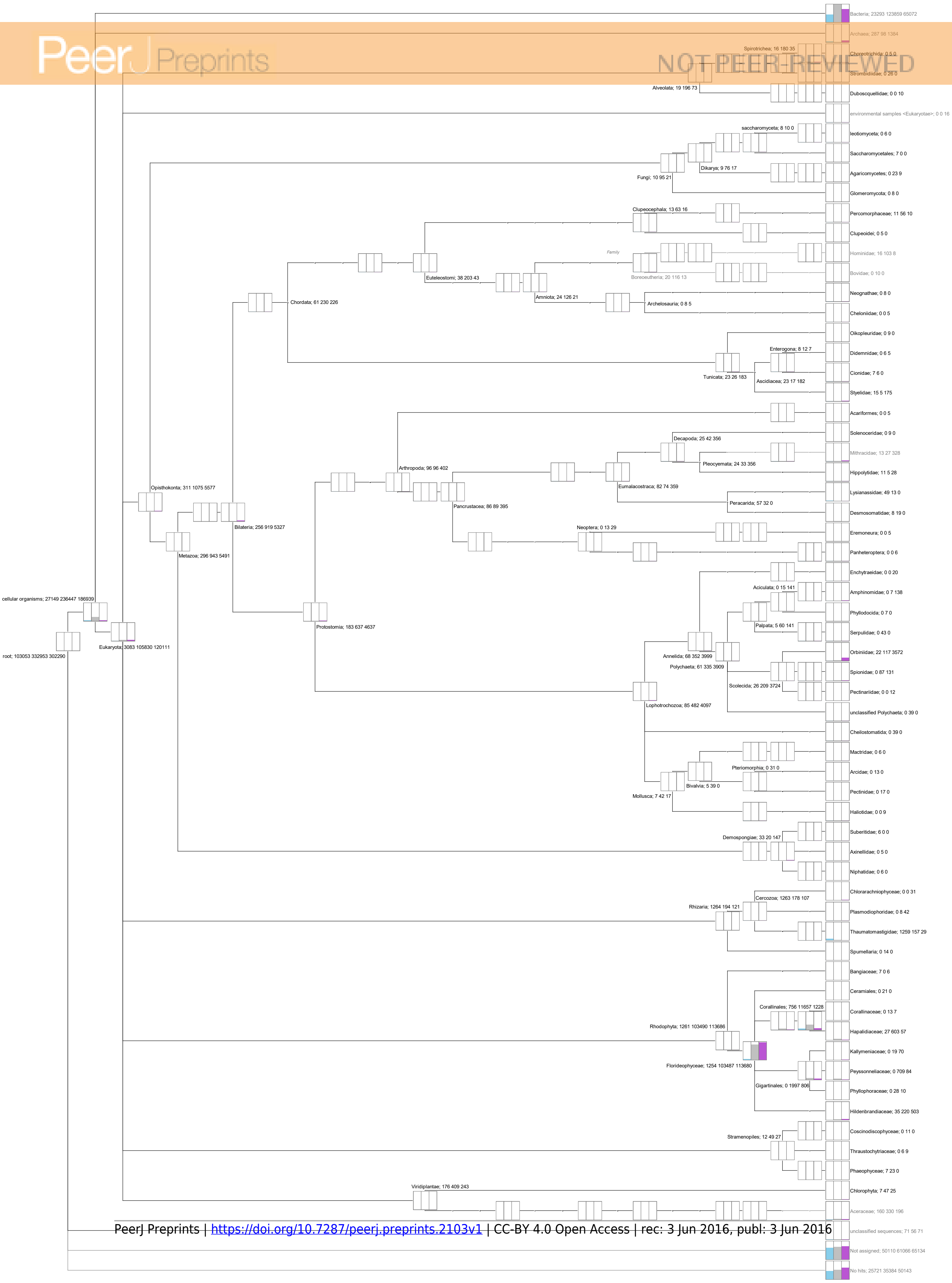
* Present in other stations of Habitats Project

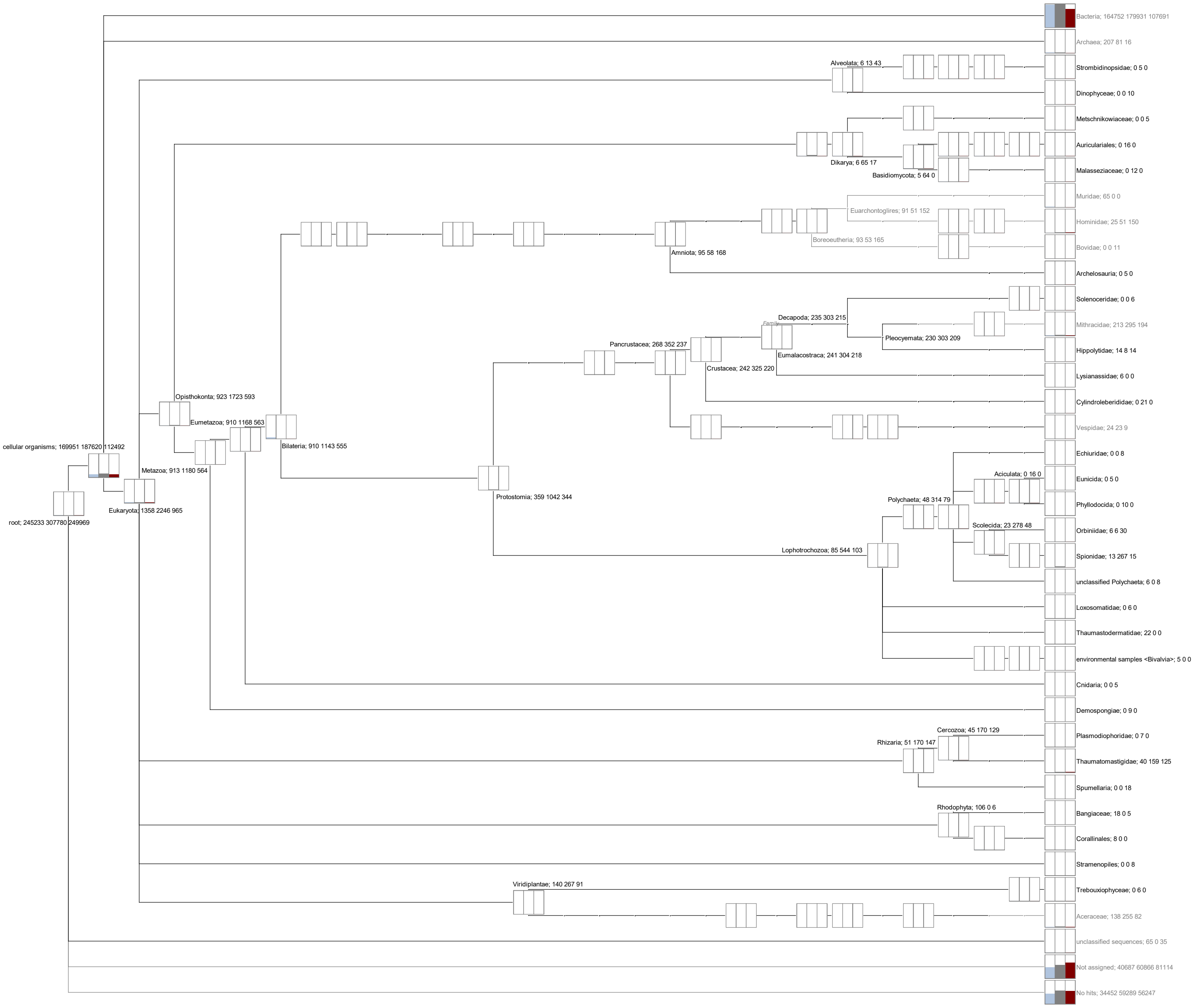
** Previous studies in Campos basin

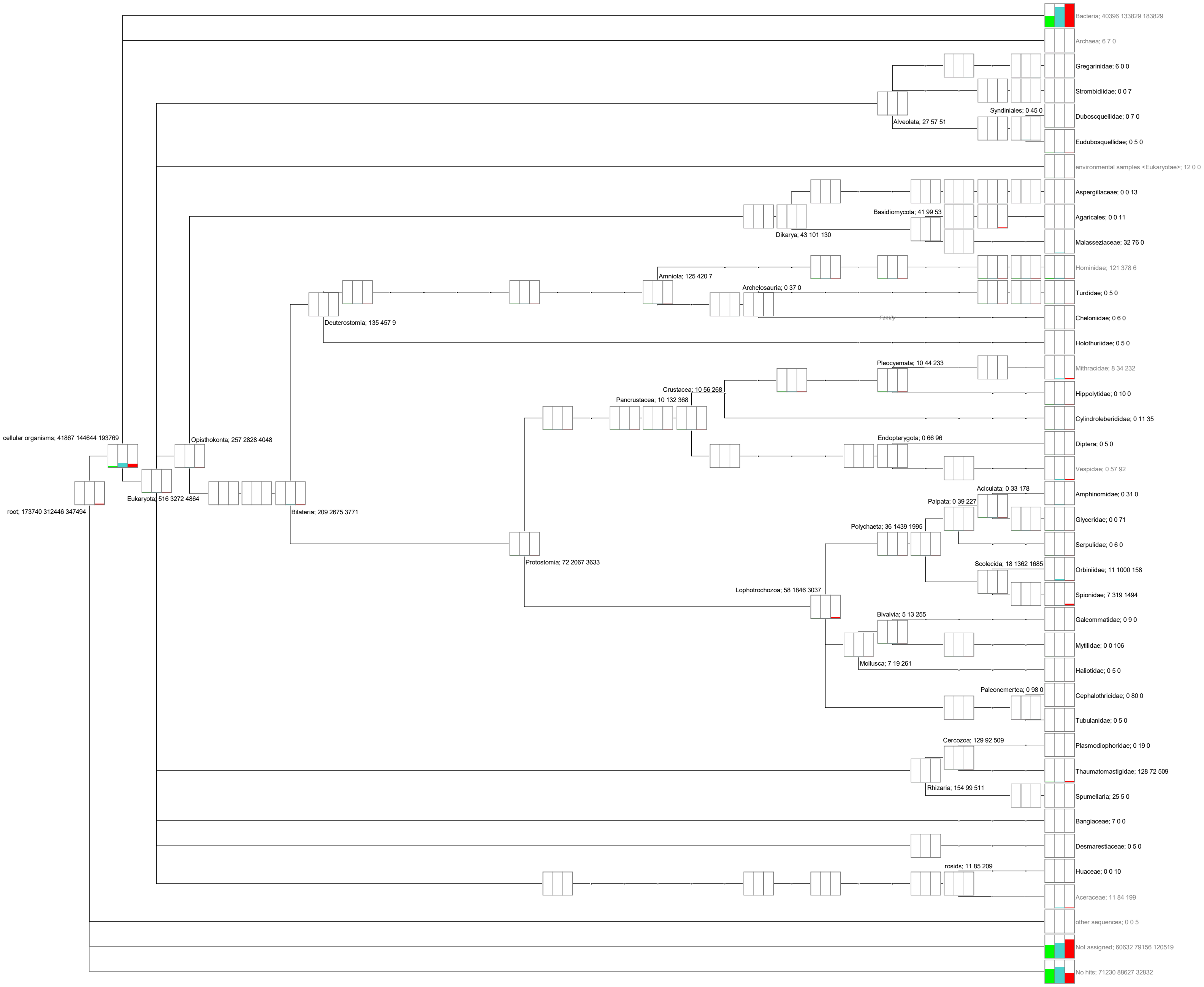
Suplementar material 2 – Family level Cladograms of the 5 sampling stations.

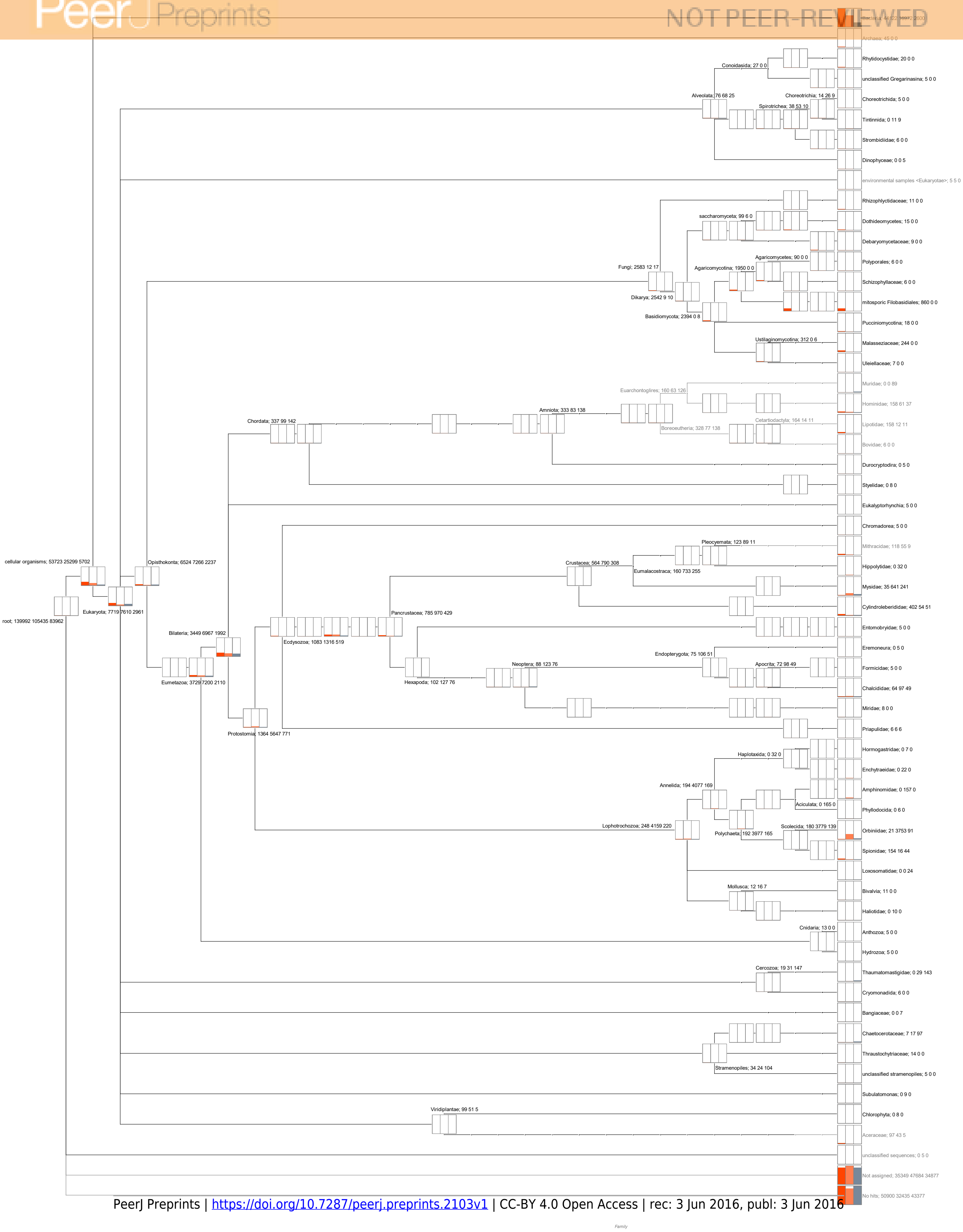
Cladograms were built using speciments identified with any of the 3 target genes. Bar inside the squares represent the number of reads from each gene used to create the node. A) Family cladogram for station B3; b) Family cladogram for station B4; C) Family cladogram for station C2; D) Family cladogram for station G2; E) Family cladogram for station F5.


Bacteria; 7094 214588 13060
Archaea; 0 62 5
Gregarinasina; 0 12 0
Lecudinidae; 0 7 0
unclassified Gregarinasina; 0 5 0
Spirotrichea; 0 18 0
Strombidinopsidae; 0 8 0
Alveolata; 12 117 11
Oligotrichia; 0 9 0
Dinophyceae; 0 42 0
environmental samples <Eukaryotae>; 0 13 0
Sordariaceae; 0 42 0
Ustilaginomycotina; 0 13 36
Malasseziaceae; 0 12 0
Dikarya; 16 70 51
Ustilaginaceae; 0 0 8
Perciformes; 0 5 0
Clupeocephala; 0 99 0
Engraulidae; 0 15 0
Otomorpha; 0 89 0
Chordata; 16 175 76
Loricariidae; 0 72 0
Euteleostomi; 15 147 75
Boreoeutheria; 14 47 73
Hominidae; 8 34 69
Bovidae; 0 13 0
Styelidae; 0 28 0
Solenoceridae; 0 31 0
Heterotremata; 14 294 17
Cancridae; 0 0 6
Decapoda; 15 592 19
Mithracidae; 14 294 11
Pleocyemata; 15 561 19
Hippolytidae; 0 266 0
Eumalacostraca; 15 777 21
Pancrustacea; 15 804 26
Lysianassidae; 0 154 0
cellular organisms; 7800 220745 14707
Ecdysozoa; 15 811 31
Peracarida; 0 185 0
Desmosomatidae; 0 31 0
Opisthokonta; 148 2021 217
Bilateria; 125 1893 139
Vespidae; 0 17 0
Eukaryota; 532 3031 1315
root; 101966 379812 84180
Priapulidae; 0 6 0
Metazoa; 127 1931 140
Amphinomidae; 0 16 0
Palpata; 0 42 0
Protostomia; 100 1695 54
Serpulidae; 0 20 0
Family
Polychaeta; 26 577 12
Orbiniidae; 21 398 0
Spionidae; 0 12 8
Scolecida; 24 412 11
unclassified Polychaeta; 0 57 0
environmental samples <Bivalvia>; 0 14 0
Lophotrochozoa; 29 829 21
Veneroida; 0 161 0
Galeommatidae; 0 152 0
Bivalvia; 0 203 5
Mactridae; 0 8 0
Pteriomorphia; 0 25 0
Mollusca; 0 217 5
Arcidae; 0 7 0
Pectinidae; 0 15 0
Lottiidae; 0 9 0
Gastropoda; 0 14 0
Vetigastropoda; 0 5 0
Niphatidae; 0 35 0
Plasmodiophoridae; 0 64 0
Cercozoa; 279 242 134
Thaumatomastigidae; 278 171 131
Rhodophyta; 0 155 9
Bangiaceae; 0 147 0
Florideophyceae; 0 8 6
Coscinodiscophyceae; 0 6 0
rosids; 21 326 81
fabids; 0 0 19
Aceraceae; 21 323 62
unclassified sequences; 0 901 6
Not assigned; 73653 91849 57290
No hits; 20505 65557 12167


Bacteria; 23293 123859 65072
Archaea; 287 98 1384
Spirotrichea; 16 180 35
Choreotrichida; 0 5 0
Strombidiidae; 0 26 0
Alveolata; 19 196 73
Duboscquellidae; 0 0 10
environmental samples <Eukaryotae>; 0 0 16
saccharomyceta; 8 10 0
leotiomyceta; 0 6 0
Saccharomycetales; 7 0 0
Dikarya; 9 76 17
Agaricomycetes; 0 23 9
Fungi; 10 95 21
Glomeromycota; 0 8 0
Clupeocephala; 13 63 16
Percomorphaceae; 11 56 10
Clupeoidei; 0 5 0
Family
Hominidae; 16 103 8
Boreoeutheria; 20 116 13
Bovidae; 0 10 0
Euteleostomi; 38 203 43
Amniota; 24 126 21
Neognathae; 0 8 0
Chordata; 61 230 226
Archelosauria; 0 8 5
Cheloniidae; 0 0 5
Oikopleuridae; 0 9 0
Enterogona; 8 12 7
Didemnidae; 0 6 5
Cionidae; 7 6 0
Tunicata; 23 26 183
Ascidiacea; 23 17 182
Styelidae; 15 5 175
Acariformes; 0 0 5
Solenoceridae; 0 9 0
Decapoda; 25 42 356
Mithracidae; 13 27 328
Pleocyemata; 24 33 356
Arthropoda; 96 96 402
Hippolytidae; 11 5 28
Eumalacostraca; 82 74 359
Lysianassidae; 49 13 0
Opisthokonta; 311 1075 5577
Peracarida; 57 32 0
Desmosomatidae; 8 19 0
Pancrustacea; 86 89 395
Bilateria; 256 919 5327
Neoptera; 0 13 29
Eremoneura; 0 0 5
Metazoa; 296 943 5491
Panheteroptera; 0 0 6
Enchytraeidae; 0 0 20
Aciculata; 0 15 141
Amphinomidae; 0 7 138
cellular organisms; 27149 236447 186939
Phyllodocida; 0 7 0
Protostomia; 183 637 4637
Palpata; 5 60 141
Serpulidae; 0 43 0
Eukaryota; 3083 105830 120111
root; 103053 332953 302290
Orbiniidae; 22 117 3572
Annelida; 68 352 3999
Polychaeta; 61 335 3909
Spionidae; 0 87 131
Scolecida; 26 209 3724
Pectinariidae; 0 0 12
Lophotrochozoa; 85 482 4097
unclassified Polychaeta; 0 39 0
Cheilostomatida; 0 39 0
Mactridae; 0 6 0
Pteriomorphia; 0 31 0
Arcidae; 0 13 0
Bivalvia; 5 39 0
Mollusca; 7 42 17
Pectinidae; 0 17 0
Haliotidae; 0 0 9
Demospongiae; 33 20 147
Suberitidae; 6 0 0
Axinellidae; 0 5 0
Niphatidae; 0 6 0
Chlorarachniophyceae; 0 0 31
Cercozoa; 1263 178 107
Rhizaria; 1264 194 121
Plasmodiophoridae; 0 8 42
Thaumatomastigidae; 1259 157 29
Spumellaria; 0 14 0
Bangiaceae; 7 0 6
Ceramiales; 0 21 0
Corallinales; 756 11657 1228
Corallinaceae; 0 13 7
Rhodophyta; 1261 103490 113686
Hapalidiaceae; 27 603 57
Kallymeniaceae; 0 19 70
Florideophyceae; 1254 103487 113680
Peyssonneliaceae; 0 709 84
Gigartinales; 0 1997 806
Phyllophoraceae; 0 28 10
Hildenbrandiaceae; 35 220 503
Coscinodiscophyceae; 0 11 0
Stramenopiles; 12 49 27
Thraustochytriaceae; 0 6 9
Phaeophyceae; 7 23 0
Viridiplantae; 176 409 243
Chlorophyta; 7 47 25
Aceraceae; 160 330 196
unclassified sequences; 71 56 71
Not assigned; 50110 61066 65134
No hits; 25721 35384 50143


Bacteria; 164752 179931 107691
Archaea; 207 81 16
Alveolata; 6 13 43
Strombidinopsidae; 0 5 0
Dinophyceae; 0 0 10
Metschnikowiaceae; 0 0 5
Auriculariales; 0 16 0
Dikarya; 6 65 17
Basidiomycota; 5 64 0
Malasseziaceae; 0 12 0
Muridae; 65 0 0
Euarchontoglires; 91 51 152
Hominidae; 25 51 150
Boreoeutheria; 93 53 165
Bovidae; 0 0 11
Amniota; 95 58 168
Archelosauria; 0 5 0
Solenoceridae; 0 0 6
Family
Decapoda; 235 303 215
Mithracidae; 213 295 194
Pleocyemata; 230 303 209
Pancrustacea; 268 352 237
Hippolytidae; 14 8 14
Eumalacostraca; 241 304 218
Crustacea; 242 325 220
Lysianassidae; 6 0 0
Cylindroleberididae; 0 21 0
Opisthokonta; 923 1723 593
Eumetazoa; 910 1168 563
Vespidae; 24 23 9
cellular organisms; 169951 187620 112492
Bilateria; 910 1143 555
Echiuridae; 0 0 8
Metazoa; 913 1180 564
Aciculata; 0 16 0
Eunicida; 0 5 0
Protostomia; 359 1042 344
Polychaeta; 48 314 79
Phyllodocida; 0 10 0
Eukaryota; 1358 2246 965
root; 245233 307780 249969
Scolecida; 23 278 48
Orbiniidae; 6 6 30
Lophotrochozoa; 85 544 103
Spionidae; 13 267 15
unclassified Polychaeta; 6 0 8
Loxosomatidae; 0 6 0
Thaumastodermatidae; 22 0 0
environmental samples <Bivalvia>; 5 0 0
Cnidaria; 0 0 5
Demospongiae; 0 9 0
Cercozoa; 45 170 129
Plasmodiophoridae; 0 7 0
Rhizaria; 51 170 147
Thaumatomastigidae; 40 159 125
Spumellaria; 0 0 18
Rhodophyta; 106 0 6
Bangiaceae; 18 0 5
Corallinales; 8 0 0
Stramenopiles; 0 0 8
Viridiplantae; 140 267 91
Trebouxiophyceae; 0 6 0
Aceraceae; 138 255 82
unclassified sequences; 65 0 35
Not assigned; 40687 60866 81114
No hits; 34452 59289 56247


Bacteria; 40396 133829 183829
Archaea; 6 7 0
Gregarinidae; 6 0 0
Strombidiidae; 0 0 7
Syndiniales; 0 45 0
Duboscquellidae; 0 7 0
Alveolata; 27 57 51
Eudubosquellidae; 0 5 0
environmental samples <Eukaryotae>; 12 0 0
Aspergillaceae; 0 0 13
Basidiomycota; 41 99 53
Agaricales; 0 0 11
Dikarya; 43 101 130
Malasseziaceae; 32 76 0
Hominidae; 121 378 6
Amniota; 125 420 7
Archelosauria; 0 37 0
Turdidae; 0 5 0
Deuterostomia; 135 457 9
Family
Cheloniidae; 0 6 0
Holothuriidae; 0 5 0
Pleocyemata; 10 44 233
Mithracidae; 8 34 232
Crustacea; 10 56 268
Hippolytidae; 0 10 0
Pancrustacea; 10 132 368
Cylindroleberididae; 0 11 35
cellular organisms; 41867 144644 193769
Opisthokonta; 257 2828 4048
Endopterygota; 0 66 96
Diptera; 0 5 0
Vespidae; 0 57 92
Eukaryota; 516 3272 4864
root; 173740 312446 347494
Bilateria; 209 2675 3771
Aciculata; 0 33 178
Amphinomidae; 0 31 0
Palpata; 0 39 227
Glyceridae; 0 0 71
Polychaeta; 36 1439 1995
Serpulidae; 0 6 0
Protostomia; 72 2067 3633
Scolecida; 18 1362 1685
Orbiniidae; 11 1000 158
Lophotrochozoa; 58 1846 3037
Spionidae; 7 319 1494
Bivalvia; 5 13 255
Galeommatidae; 0 9 0
Mytilidae; 0 0 106
Mollusca; 7 19 261
Haliotidae; 0 5 0
Paleonemertea; 0 98 0
Cephalothricidae; 0 80 0
Tubulanidae; 0 5 0
Cercozoa; 129 92 509
Plasmodiophoridae; 0 19 0
Thaumatomastigidae; 128 72 509
Rhizaria; 154 99 511
Spumellaria; 25 5 0
Bangiaceae; 7 0 0
Desmarestiaceae; 0 5 0
rosids; 11 85 209
Huaceae; 0 0 10
Aceraceae; 11 84 199
other sequences; 0 0 5
Not assigned; 60632 79156 120519
No hits; 71230 88627 32832


Bacteria; 44122 16972 2600
Archaea; 45 0 0
Rhytidocystidae; 20 0 0
unclassified Gregarinasina; 5 0 0
Choreotrichida; 5 0 0
Tintinnida; 0 11 9
Strombidiidae; 6 0 0
Dinophyceae; 0 0 5
environmental samples <Eukaryotae>; 5 5 0
Rhizophlyctidaceae; 11 0 0
Dothideomycetes; 15 0 0
Debaryomycetaceae; 9 0 0
Polyporales; 6 0 0
Schizophyllaceae; 6 0 0
mitosporic Filobasidiales; 860 0 0
Pucciniomycotina; 18 0 0
Malasseziaceae; 244 0 0
Uleiellaceae; 7 0 0
Muridae; 0 0 89
Hominidae; 158 61 37
Lipotidae; 158 12 11
Bovidae; 6 0 0
Durocryptodira; 0 5 0
Styelidae; 0 8 0
Eukalyptorhynchia; 5 0 0
Chromadorea; 5 0 0
Mithracidae; 118 55 9
Hippolytidae; 0 32 0
Mysidae; 35 641 241
Cylindroleberididae; 402 54 51
Entomobryidae; 5 0 0
Eremoneura; 0 5 0
Formicidae; 5 0 0
Chalcididae; 64 97 49
Miridae; 8 0 0
Priapulidae; 6 6 6
Hormogastridae; 0 7 0
Enchytraeidae; 0 22 0
Amphinomidae; 0 157 0
Phyllodocida; 0 6 0
Orbiniidae; 21 3753 91
Spionidae; 154 16 44
Loxosomatidae; 0 0 24
Bivalvia; 11 0 0
Haliotidae; 0 10 0
Anthozoa; 5 0 0
Hydrozoa; 5 0 0
Thaumatomastigidae; 0 29 143
Cryomonadida; 6 0 0
Bangiaceae; 0 0 7
Chaetocerotaceae; 7 17 97
Thraustochytriaceae; 14 0 0
unclassified stramenopiles; 5 0 0
Subulatomonas; 0 9 0
Chlorophyta; 0 8 0
Aceraceae; 97 43 5
unclassified sequences; 0 5 0
Not assigned; 35349 47684 34877
No hits; 50900 32435 43377
Conoidasida; 27 0 0
Alveolata; 76 68 25
Choreotrichia; 14 26 9
Spirotrichea; 38 53 10
saccharomyceta; 99 6 0
Agaricomycetes; 90 0 0
Fungi; 2583 12 17
Agaricomycotina; 1950 0 0
Dikarya; 2542 9 10
Basidiomycota; 2394 0 8
Ustilaginomycotina; 312 0 6
Euarchontoglires; 160 63 126
Amniota; 333 83 138
Chordata; 337 99 142
Cetartiodactyla; 164 14 11
Boreoeutheria; 328 77 138
Pleocyemata; 123 89 11
cellular organisms; 53723 25299 5702
Opisthokonta; 6524 7266 2237
Crustacea; 564 790 308
Eumalacostraca; 160 733 255
Eukaryota; 7719 7610 2961
root; 139992 105435 83962
Pancrustacea; 785 970 429
Bilateria; 3449 6967 1992
Ecdysozoa; 1083 1316 519
Endopterygota; 75 106 51
Neoptera; 88 123 76
Apocrita; 72 98 49
Eumetazoa; 3729 7200 2110
Hexapoda; 102 127 76
Protostomia; 1364 5647 771
Haplotaxida; 0 32 0
Annelida; 194 4077 169
Aciculata; 0 165 0
Lophotrochozoa; 248 4159 220
Scolecida; 180 3779 139
Polychaeta; 192 3977 165
Mollusca; 12 16 7
Cnidaria; 13 0 0
Cercozoa; 19 31 147
Stramenopiles; 34 24 104
Viridiplantae; 99 51 5
Family

Suplementar material 3 – List of species identified by molecular and morphological taxonomy

Specie	18S	20S	COI	Study
Cnemidocarpa verrucosa	+	+	+	Schettini
Desmarestia dudresnayi	+	+	+	Schettini
Erythrophyllum delesserioides	+	+	+	Schettini
Eurythenes gryllus	+	+	+	Schettini
Galeomma turtoni	+	+	+	Schettini
Grifola frondosa	+	+	+	Schettini
Haliotis diversicolor	+	+	+	Schettini
Hormogaster redii	+	+	+	Schettini
Lysmata seticaudata	+	+	+	Schettini
Malassezia globosa	+	+	+	Schettini
Marenzelleria arctia	+	+	+	Schettini
Mimachlamys varia	+	+	+	Schettini
Mysidium columbiae	+	+	+	Schettini
Parotocinclus maculicauda	+	+	+	Schettini
Pinctada imbricata	+	+	+	Habitats and Hits
Platynereis dumerilii	+	+	+	Habitats and Hits
Pontocaris lacazei	+	+	+	Habitats
Praxillella affinis	+	+	+	Habitats
Progoniada regularis	+	+	+	Habitats and Hits
Protodorvillea kefersteini	+	+	+	Habitats
Pteria colymbus	+	+	+	Habitats
Scalibregma inflatum	+	+	+	Habitats and Hits
Scapharca broughtonii	+	+	+	Schettini
Serpula vermicularis	+	+	+	Schettini
Syllis gracilis	+	+	+	Habitats and Hits
Syllis variegata	+	+	+	Habitats and Hits
Travisia brevis	+	+	+	Habitats and Hits
Travisia forbesii	+	+	+	Habitats and Hits
Travisia pupa	+	+	+	Habitats and Hits
Aglaophamus circinata		+	+	Habitats and Hits
Alpheus formosus		+	+	Habitats
Amphipholis squamata		+	+	Habitats
Aricidea wassi		+	+	Habitats and Hits
Chelonia mydas		+	+	Schettini
Praxillella pacifica		+	+	Habitats and Hits
Priapulus caudatus		+	+	Schettini
Scolelepis bonnieri		+	+	Schettini
Scolelepis foliosa		+	+	Schettini
Amphimedon queenslandica	+		+	Schettini
Axiothella rubrocincta	+		+	Habitats and Hits
Bathyarca pectunculoides	+		+	Habitats
Bathyglycinde profunda	+		+	Habitats
Bathyglycinde sibogana	+		+	Habitats
Caprella equilibra	+		+	Habitats and Hits
Ceratocephale abyssorum	+		+	Habitats and Hits

Specie	18S	20S	COI	Study
Ciona intestinalis	+		+	Schettini
Clymenella torquata	+		+	Habitats and Hits
Pectinaria granulata	+		+	Schettini
Perna viridis	+		+	Schettini
Protaspis grandis	+		+	Schettini
Syllis hyalina	+		+	Habitats and Hits
Didemnum candidum			+	Schettini
Leodamas rubra			+	Habitats and Hits
Leodia sexiesperforata			+	Habitats
Leptochelia dubia			+	Habitats
Leucothoe urospinosa			+	Habitats and Hits
Lumbrineris latreilli			+	Habitats and Hits
Lysidice ninetta			+	Habitats and Hits
Lysmata anchisteus			+	Schettini
Macrochaeta clavicornis			+	Habitats
Marphysa bellii			+	Habitats and Hits
Mendicula ferruginosa			+	Habitats and Hits
Mooreonuphis pallidula			+	Habitats and Hits
Neanthes acuminata			+	Habitats and Hits
Nereimyra punctata			+	Habitats and Hits
Notomastus latericeus			+	Habitats and Hits
Ophelina acuminata			+	Habitats and Hits
Pyropia haitanensis			+	Schettini
Scapharca kagoshimensis			+	Schettini
Scoloplos armiger			+	Schettini
Isolda pulchella			++	Habitats and Hits
Apophlaea lyallii	+	+		Schettini
Chaetoceros curvisetus	+	+		Schettini
Coelomactra antiquata	+	+		Schettini
Crassinella lunulata	+	+		Habitats and Hits
Cryptococcus friedmannii	+	+		Schettini
Cyclaspis alba	+	+		Habitats
Cylichna alba	+	+		Habitats and Hits
Engraulis japonicus	+	+		Schettini
Euclymene oerstedi	+	+		Habitats and Hits
Eulalia viridis	+	+		Habitats and Hits
Eumida sanguinea	+	+		Habitats and Hits
Exogone dispar	+	+		Habitats and Hits
Galathowenia oculata	+	+		Habitats
Glycera americana	+	+		Habitats and Hits
Glycera southeastatlantica	+	+		Habitats and Hits
Goniada emerita	+	+		Habitats
Hesiospina aurantiaca	+	+		Habitats and Hits
Patelloida striata	+	+		Schettini
Scopelocheirus schellenbergi	+	+		Schettini
Subulatomonas tetraspora	+	+		Schettini
Ophelina cylindricaudata		+		Habitats and Hits
Ophiactis lymani		+		Habitats
Trypanosyllis zebra		+		Habitats and Hits

Specie	18S	20S	COI	Study
Ahnfeltiopsis leptophylla	+			Schettini
Crucigera zygophora	+			Schettini
Leitoscoloplos pugettensis	+			Schettini
Malassezia nana	+			Schettini
Ophiura ljungmani	+			Habitats
Owenia fusiformis	+			Habitats and Hits
Panthalis oerstedi	+			Habitats and Hits
Paralacydonia paradoxa	+			Habitats and Hits
Paramphinome jeffreysii	+			Habitats and Hits
Pholoe minuta	+			Habitats
Phtisica marina	+			Habitats
Phyllodoce longipes	+			Habitats and Hits
Solenocera crassicornis	+			Schettini
Strombidium paracalkinsi	+			Schettini
Phagomyxa odontellae	+			Schettini