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

Milena MDP **Schettini**, Raony GCCL **Cardenas**, Marcella AA **Detoni**, Mauro F **Rebelo**.

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

KEYWORDS: rRNA 18S, rRNA 28S, COI, metagenomics,

ABSTRACT

Biodiversity is currently assessed for environmental characterizations and monitoring through a laborious and time-consuming process of morphological taxonomy. We used rRNA 18S, rRNA 28S and COI, together with NGS 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. 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. By meta-barcoding we identified phyla that are traditionally found in samples of marine benthos, such as Annelida, Arthropoda, Mollusca and Chordata, as well as rare phyla like Entoprocta and Gastrotricha. Taxa identified with meta-barcoding were compared to morphology data from previous studies in the area (REVIZEE, Habitats Project) and geo-validated with the database *Global Biodiversity Information Facility*. For several taxa, this is the first evidence of occurrence in Campos the area and the number of OTU identified suggests an enormous unveiled benthic biodiversity in Campos Basin. Our study supports the application of Meta-Barcoding for environmental characterization and monitoring programs, reducing from years to few months the time currently required for species identification and biodiversity determination.

INTRODUCTION

Sediment fauna characterization and monitoring are mandatory requirements for obtaining oil and gas (O&G) environmental permits 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, for example the Equatorial Margin and Santos Pre-salt Basin in Brazil

Biodiversity is currently assessed for environmental characterizations and monitoring through a laborious and time-consuming process of morphological taxonomy. 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. Expert judgment is never 100% accurate, with evidence of only 50% rate of identification consistency being shared among taxonomists (Culverhouse et al., 2003). Pseudo-absence is frequent in cases, for example, of fragile organisms that require special fixation (Costa-Paiva, Paiva & Klautau, 2007). As a result, invertebrate morphological identification efforts are often limited to few groups, including Mollusca, Crustacea and Polychaeta (Scaramuzza, 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 in sedimentary basins as large as 300.000 km2, this number can grow to tenths of thousands of samples for baseline environmental characterization. The lack of experts is a major bottleneck in the process of identifying biodiversity (Hebert et al., 2003; Mora, Rollo & Tittensor, 2013) and as a result, taxonomists are constantly failing to meet the demands for biodiversity assessment required in monitoring programs, delaying the development of economical activities and the discovery of new species. .

According to the latest Report of the Convention on Biological Diversity (Diversity, 2016), Brazil is the most biologically-diverse country in the world, with more than 100,000 animal species been accounted for. However, only 184 marine invertebrates had their conservation status accessed (Scaramuzza, 2015). It is possible that current risk estimates of environmental impact

are based on underestimated biodiversity inventories, representing a threat to species conservation (Wu, 1982). Developing new technologies and approaches that accelerate species discovery and reveal hidden biodiversity is crucial for setting conservation priorities and efforts. Meta-barcoding uses big data about genetic markers generated through high-throughput new generation sequencing (NGS/HTS) of bulk environmental samples (Leray & Knowlton, 2015), to greatly accelerates species discovery and unveil biodiversity.

Since 2010, more than 600 papers have been published on the use of DNA-based identification methods for species conservation (Goldberg, Strickler & Pilliod, 2015; Bergman et al., 2016), biodiversity inventory determination (Drummond et al., 2015); environmental monitoring (Bohmann et al., 2014; Chariton et al., 2015; Leray & Knowlton, 2015; Brown et al., 2015), DNA extraction/detection (Pedersen et al., 2014; Eichmiller, Bajer & Sorensen, 2014; Ficetola, Taberlet & Coissac, 2016) and the technique has been considered a major tool for Ocean's sustainability in the 21st century (Aricò, 2015). This approach is particularly useful because of its sensitivity to identify minute organisms and of species from debris (Wang et al., 2014). For eukaryote organisms that have not yet had their genetic markers sequenced or have not yet been described morphologically, the concept of Operational Taxonomic Unit (OTU) can be applied (Stackebrandt & Goebel, 1994; Pedersen et al., 2014).

In this study, we combined three different phylogenetic markers (rRNA 18S, rRNA 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 in a survey 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 in at Campos Basin.. 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 -20ºC for 4 years until our analysis was done in 2013.

For each station, we manually homogenized 200 g 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 previous studies: data from the Brazilian program of characterization of the Economical Exclusive Zone (REVIZEE) (Lavrado & 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, Priapulida, Protozoa and Rodophyta were identified by the Habitats Project and provided by CENPES/PETROBRAS (unpublished data). We also used the database *Global Biodiversity Information Facility* (www.gbif.org) for organism geo-localization.

PCR and high-throughput sequencing:

Information on PCR of COI, rRNA 18S and rRNA 28S genes is presented in Supplementary 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.

Bioinformatics and Taxonomic Name Attribution:

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

We obtained an average of 4.83 µg of DNA from each of the 15 samples. Sequencing generated approximately 4.8 million sequences with an average size of 155.1 bp. Over 3.6 million sequences (75.35%) passed quality control and of these; around 3.3 million were clustered in OTU by CD-HIT. Table 2 shows the total number of OTU and the number of OTU with No Hits in

BLAST, Non-attributed to any taxa by LCA and with taxonomic name attributed.. More than 1.6 million sequences (about 50% of all reads) were assigned to 957 prokaryotes and 577 eukaryotes by the LCA algorithm in MEGAN using hits produced by the similarity algorithm BLAST with any of the 3 molecular markers (rRNA18S, rRNA28S, COI), divided by sampling station. LCA further identified 23 phyla, 60 classes, 62 orders, 70 families, 67 genus and 46 species of eukaryotes...Figure 1A shows the distribution of the 13 invertebrate phyla OTU identified by Meta-barcoding for each of the 5 stations and Figure 1B the same for the 38 invertebrate families OTU identified. All other Prokaryote and Eukaryote observed in this study, with any of the 3 molecular markers, to the taxonomic depth of family, are listed in the cladograms available in Supplementary material 2 for each of the 5 sampling stations.

Our analysis identified 38 families of invertebrates in the 15 samples from the 5 sampling stations in Campos Basin. Figure 2 shows a comparison of the spatial distribution of families identified by Meta-Barcoding from phyla with most abundant frequencies: Annelida (9 families, figure. 2A), Arthropoda (10 families, figure. 2B) and Mollusca (7 families, figure. 2C) in relation to previously published morphology taxonomy results in stations B3, B4, C2, F5 and G2.

At first, the LCA algorithm identified 46 species, of which 27 were invertebrates not previously described in the region. A text search of the list of BLAST hits allowed for more 45 species of invertebrates previously identified in Campos Basin to be identified. The full list of species identified in this study is in Supplementary material 3.

DISCUSSION

In this study we report the first meta-barcoding description of the Eukaryote biodiversity in the deep-sea Brazilian continental shelf. More than 1.6 million OTU were assigned to 957 prokaryotes and 577 eukaryotes. Even though the relation between OTU and species must be made with extreme caution, the remaining 1.6 million OTU that could not be identified at this time with the current genetic markers available in Genbank suggests the benthic biodiversity of

Campos Basin could be orders of magnitude higher than anticipated by previous morphological taxonomy studies.

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 and where several previous morphological taxonomic studies were performed. Either by the oil companies interested in obtaining their environmental permits or those involved in conservational programs (such as the Habitats Project) or by the scientific community (specially the REVIZEE program).

The approximately 4.8 million sequences we found are within the expected range expected for the 318 v2 chip, and even though the average size of 155.1 bp was bellow the expected value for the OT2 400 kit, it did not compromise our analysis.

When further analyzing the OTU distributed in the 23 phyla, we found that 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 specific genetic markers. Our discussion will focus on the 13 invertebrate phyla that were identified because of their significance for the legal environmental characterization and monitoring in offshore areas and these artifact findings on chordate will be no longer addressed here.

Our meta-Barcoding 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 (Figure 1 and Supplementary material 2).

The great number of OTUs for Annelida, Arthropoda and Mollusca found by metagenomics agrees with previous results for Campos Basin (Lavrado & Ignacio, 2006)during the REVIZEE project and also by those of the Habitats Project. Recent meta-barcoding study (Leray & Knowlton, 2015) also identified Annelida and Arthropoda as the phyla with more OTUs among

188 the 22 phyla identified from approximately 0.09 m^3 sediments from coral reef regions in Virginia
189 and Florida, in the United States.
190 The Entoprocta (or Kamptozoa) phylum comprises about 170 aquatic and sessile species of sizes
191 between 0.5 and 5.0 mm and are mostly marine (Zhang, 2011). Until 2011, only 18 species of
192 Entoprocta were known on the Brazilian coast (Vieira & Migotto, 2011). In this study, all OTUs
193 (6 in the C2 station and 24 in the G2 station) were attributed to the genus *Loxosomella* through
194 the marker rRNA 28S, with over 86% of sequence similarity. This result expands the distribution
195 of the genus that was previously limited to six species collected off the coast of São Paulo (Vieira
196 & Migotto, 2011). As for the cosmopolitan Gastrotricha phylum that comprises about 790
197 species of aquatic organisms up to 1 mm in length (Zhang, 2011), all 22 OTUs assigned to the
198 phylum (C2 station) were in the *Tetranchyroderma* genus, with over 81% similarity with COI
199 sequences found in the Genbank. This occurrence also expands the limited distribution that had
200 been previously reported but not formally described to São Paulo beaches (Garraffoni & Araújo,
201 2010), almost a 1000km away from the Campos Basin.
202 This is a pioneer study in which meta-barcoding results could be compared to those from a
203 recent comprehensive morphological taxonomy effort that worked with the same samples than
204 those used in our study: the Habitats Project coordinated by CENPES/PETROBRAS. Their huge
205 morphological taxonomy effort generated a databank of almost 50.000 specimens, with
206 identification of 17 phyla, 27 classes, 63 orders, 354 families, 768 genus and 749 species.
207 The comparison between the findings obtained with molecular and morphological taxonomies
208 however is limited since 1,211 (68%) of the 1,773 macro invertebrate *taxa* identified by
209 morphological taxonomy, did not have any entry in Genbank found for any of the three markers
210 (rRNA18S, rRNA28S or COI) used in this study. This also indicates a huge underrepresentation
211 of Brazilian marine species Genbank and the a need to increase efforts to have sequences from
212 these three molecular markers from more Brazilian species deposited in Genbank,
213 The uncertainty on how much DNA was still available in the samples that have been preserved at
214 -20ºC for 4 years as well as the limited amount of sample analyzed in each station (5 g out of 200

g of the 0 to 2 cm slice of sediment, compared to 4 L, of the 0 to 10 cm slices for the morphological study). 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 implies only that absent families may be pseudo absent and do not limit conclusions drawn from the current observations.

Out f the 70 families identified by the Meta-Barcoding, 21 were invertebrate.

Families Amphinomidae, Enchytraeidae, Glyceridae, Orbiniidae, Serpulidae and Spionidae belonging to Annelida phyla were previously identified in Campos Basin by the Habitats Project that also identified other 28 Annelida families not found by meta-barcoding. Hormogastridae found in our study is most likely a false positive since it is not marine family.

Families Solenoceridae, Cylindroleberididae and Mysidae belonging to Arthropoda phyla have been previously identified in Campos Basin and in the Southeast of Brazil by other authors (Cardoso, 2007; Serejo et al., 2007; Tâmega, Oliveira & Figueiredo, 2013) while 29 arthropoda families previously reported by the Habitat Project were not be identified by meta-barcoding.

Families Miridae, Chalcididae and Formicidae found in our study are most likely false positive since they are non-marine insects.

All Mollusca families identified by metagenomics in Campos basin, except for Mytilidae have been previously found in the region (Lavrado & Ignacio, 2006; Dornellas & Simone, 2011; Tâmega, Oliveira & Figueiredo, 2013) although not by the Habitat Projects, that also identified 15 Mollusca families not identified by meta-barcoding.

Meta-barcoding was also able to find, for every sampling station, families not previously reported by the Habitat's Project , suggesting that their distribution could be broader than anticipated estimated by morphological taxonomy. That is the case for Echiuridae, Hormogastridae and Pectinariidae among the Annelidae; Desmosomatidae and Hippolytidae in Arthropoda and Arcidae, Mactridae and Pectinidae among Mollusca.

Out of the 46 species found by meta-barcoding, none of the 21-benthic invertebrates had been previously described by the Habitat project and could represent new occurrences for the project.

We must remember that even though species level is expected, taxonomic penetration to family level is accepted and most specimens in previous studies have been identified only to this level. When searching records from the Habitats projects as well as those of the REVIZEE and GBIF, we found record of the arthropod *Eurythenes gryllus* and all the families of all other newly observed species. However, we cannot discard the possibility of false positive.

The comparison with data from the Habitat Projects was extremely limited by the availability of the three genetic markers (rRNA18S, rRNA28S and COI) deposited in Genbank. Only 64 out of the 749 organisms identified to the species level by in the Habitat's Project had at least one genetic marker sequence and thus were 'eligible' for molecular identification.

However, none of the 64 species were found by meta-barcoding. We believe these to be pseudo absence that could be explained, mainly, by the samples preserved at -20^{o}C for 4 years. But there could be another explanation. We noticed during the analysis of the data that, even after calibration of the parameters for the LCA algorithm (data not shown), some incongruence in the attribution of the taxonomic name to a species could happen due to the selection of a unlikely BLAST read to name the query OTU. To overcome fix this problem, we manually searched the text of the names of the organisms generated by all BLAST hits for a given read, for the names of the 64 species found by the Habitat Project. We then 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.

Other pseudo-absence 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 percentage could reach 40% (Costello, May & Stork, 2013). An ongoing effort is dedicated to resolve synonymous names found in the GBIF database.

The use of biogeographic databases (Habitats Project, REVIZEE and GBIF) to verify and adjust the meta-barcode observations has proven to be a good strategy. False positive results could

happen as an artifact of the low representativeness of Brazilian species in Genbank. Due to similarities of genetic sequences shared among species belonging to the same genera BLAST could relate, with very low error probability, a read from one species not present in the databank to another from the same genus (phylogenetic similarity) present in the Genbank 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. The small (25-85 mm) gastropod *Haliotis diversicolor* identified in our study is native of the Indo-Pacific Ocean, with geo-referenced records on the coast of Japan, Thailand and Australia (GBIF, 2016). We have then to decide if it is a new occurrence of this species in a completely new environment, or a false positive. But there is a third option. Another small gastropod from the same genera, *Haliotis aurantium*, has been previously identified, not only in the Brazilian coast, but specifically in Campos basin. At light of this information, we believe that the lack of genetic markers for this Brazilian species in Genbank may have misled BLAST to erroneously classify an OTU from *H. aurantium* as of *H. diversicolor.* A system that can sort such incongruences could greatly help meta-barcoding analysis.

To further remove false positive results, we tried to verify the occurrence of one species with one genetic marker by the redundant identification of the same species with another genetic marker. This way we were 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 46 species identified by molecular meta-barcoding, 16 had sequences of all three genetic markers available in Genbank, but were always 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 pseudo absences could be solved if we work in a taxonomic free context, looking only at OTU to compare biodiversity profiles among samples. The frequency and abundance of OTUs could then be related to environmental changes, either spatial or seasonal, and species discovery would be accelerate by identification of which OTU vary according to environmental conditions. Were such strategy to be adopted, OTU profiles would allows us to work with the hidden biodiversity of the thousands of 'no hit' OTU and let them and their distribution to tell us about environmental changes.

Species name are a fundamental piece of ecology and in spite of all the uncertainty that this definition may bring with it, a lot of the accumulated knowledge in biology is associated with these units. Even if we may never give up the idea of naming a species, the easiness to gather OTU data is unprecedented and makes scenery of taxonomic free ecology complementary to the traditional one, more and more likely to exist in the years to come.

CONCLUSION

This study contributes with relevant evidence that Meta-Barcoding methods can be a high reliability, fast speed and low cost tool for environmental characterization and monitoring. It may be the only alternative to produce valuable information for decision making about vast and unknown areas in a short time in a way that safe-guard the environment without delaying economical activities. It may also accelerate species discovery and contribute to ecology in ways that have not been fully understood yet.

The methodology can be improved by adding more sequences of native species in public and proprietary databanks, but it is our opinion that meta-barcoding can already be considered an 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.

Brazil has one of the strictest environmental laws and regulations for the O&G sector in the world that is constantly being improved. Recent changes made under resolution CONAMA 422/11 minimized bureaucracy in the application process, increased transparency by sharing information online and reduced liability for the O&G operators. The Brazilian environmental authority IBAMA (Brazilian Institute of the Environment and Renewable Natural Resources) establishes the guidelines and best practices for the environmental licensing and monitoring by means of 'reference terms'. By becoming an early adopter, IBAMA could have a leading role in the implementation of this innovative methodology that can greatly contribute to the conservation of deep-sea environments worldwide

ACKNOWLEDGEMENTS

Authors want to acknowledge the financial support from the Rio de Janeiro State Research Council FAPERJ (grant: E26/111.403/2012) and thank CENPES/PETROBRAS for the donation of sediments samples and access to biodiversity results from the Habitats Project. Finally, we are thankful to Dr. Marcia Triunfol for revision of the manuscript.

REFERENCES

Aricò S. 2015. *Ocean Sustainability in the 21st Century*. UNESCO Publishing / Cambridge University Press.

Bergman PS., Schumer G., Blankenship S., Campbell E. 2016. Detection of Adult Green Sturgeon Using Environmental DNA Analysis. *PLOS ONE* 11:e0153500. DOI: 10.1371/journal.pone.0153500.

Bohmann K., Evans A., Gilbert MTP., Carvalho GR., Creer S., Knapp M., Yu DW., de Bruyn M. 2014. Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in ecology & evolution* 29:358–367. DOI: 10.1016/j.tree.2014.04.003.

Brown EA., Chain FJJ., Crease TJ., MacIsaac HJ., Cristescu ME. 2015. Divergence thresholds and divergent biodiversity estimates: can metabarcoding reliably describe zooplankton

communities? *Ecology and evolution* 5:2234–51. DOI: 10.1002/ece3.1485.

Cardoso IA. 2007. Deep Sea Caridea (Crustacea, Decapoda) From Campos Basin, Rj, Brazil*. *Brazilian Journal of Oceanography* 55:39–50. DOI: 10.1590/S1679-87592007000100005.

Chariton AA., Stephenson S., Morgan MJ., Steven ADL., Colloff MJ., Court LN., Hardy CM. 2015. Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. *Environmental Pollution* 203:165–174. DOI: http://dx.doi.org/10.1016/j.envpol.2015.03.047.

Costa-Paiva EM., Paiva PC., Klautau M. 2007. Anaesthetization and fixation effects on the morphology of sabellid polychaetes (Annelida: Polychaeta: Sabellidae). *Journal of the Marine Biological Association of the UK* 87:1127–1132. DOI: 10.1017/S002531540705223X.

Costello MJ., May RM., Stork NE. 2013. Can We Name Earth ' s Species Before They Go Extinct ? *Science* 339:413–416. DOI: 10.1126/science.1230318.

Culverhouse PF., Williams R., Reguera B., Herry V., Gonzalez Gil S. 2003. Do experts make mistakes? A comparison of human and machine labeling of dinoflagellates. *Marine Ecology Progress Series* 247:17–25.

Diversity C of B. 2016.Brazil - Overview. *Available at https://www.cbd.int/countries/?country=br*

Dornellas APS., Simone LRL. 2011. Annotated list of type specimens of mollusks deposited in Museu de Zoologia da Universidade de São Paulo , Brazil. *Arquivos de Zoologia* 42:1–81.

Drummond AJ., Newcomb RD., Buckley TR., Xie D., Dopheide A., Potter BC., Heled J., Ross HA., Tooman L., Grosser S., Park D., Demetras NJ., Stevens MI., Russell JC., Anderson SH., Carter A., Nelson N. 2015. Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience* 4:46. DOI: 10.1186/s13742-015-0086-1.

Eichmiller JJ., Bajer PG., Sorensen PW. 2014. The Relationship between the Distribution of Common Carp and Their Environmental DNA in a Small Lake. *PLoS ONE* 9:e112611. DOI: 10.1371/journal.pone.0112611.

Ficetola GF., Taberlet P., Coissac E. 2016. How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources* 16:604–607. DOI: 10.1111/1755-

0998.12508.

Garraffoni ARS., Araújo TQ. 2010. Chave de Identificação de Gastrotricha de Águas Continentais e Marinhas do Brasil. *Papeis Avulsos de Zoologia* 50:535–552. DOI: 10.1590/S0031-10492010003300001.

Goldberg CS., Strickler KM., Pilliod DS. 2015. Moving environmental \{DNA\} methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation* 183:1–3. DOI: http://dx.doi.org/10.1016/j.biocon.2014.11.040.

Hebert PDN., Cywinska A., Ball SL., deWaard JR. 2003. Biological identifications through DNA barcodes. *Proceedings. Biological sciences / The Royal Society* 270:313–21. DOI: 10.1098/rspb.2002.2218.

Huson D., Auch A., Qi J., Schuster S. 2007. MEGAN analysis of metagenome data. *Gennome Res.* 17:377–386. DOI: 10.1101/gr.5969107.

Huson DH., Mitra S., Ruscheweyh HJ., Weber N., Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Research* 21:1552–1560. DOI: 10.1101/gr.120618.111.

Lavrado HP., Ignacio BL. 2006. *Biodiversidade bentônica da região central da zona econômica exclusiva brasileira*. Rio de Janeiro: Museu Nacional.

Leray M., Knowlton N. 2015. DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences* 2014:201424997. DOI: 10.1073/pnas.1424997112.

Li W., Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659. DOI: 10.1093/bioinformatics/btl158.

Miloslavich P., Klein E., D??az JM., Hern??ndez CE., Bigatti G., Campos L., Artigas F., Castillo J., Penchaszadeh PE., Neill PE., Carranza A., Retana M V., D??az de Astarloa JM., Lewis M., Yorio P., Piriz ML., Rodr??guez D., Valentin YY., Gamboa L., Mart??n A. 2011. Marine biodiversity in the Atlantic and Pacific coasts of South America: Knowledge and gaps. *PLoS ONE* 6. DOI:

10.1371/journal.pone.0014631.

Mora C., Rollo A., Tittensor DP. 2013. Comment on “Can We Name Earth’s Species Before They Go Extinct?” *Science* 341 :237. DOI: 10.1126/science.1237254.

Pedersen MW., Overballe-Petersen S., Ermini L., Sarkissian C Der., Haile J., Hellstrom M., Spens J., Thomsen PF., Bohmann K., Cappellini E., Schnell IB., Wales NA., Carøe C., Campos PF., Schmidt AMZ., Gilbert MTP., Hansen AJ., Orlando L., Willerslev E. 2014. Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 370. DOI: 10.1098/rstb.2013.0383.

Scaramuzza CA de M. 2015. *Brazil: Fifth National Report to the CDB*. Brasília. DOI: 10.1044/leader.PPL.20012015.20.

Scheffers BR., Joppa LN., Pimm SL., Laurance WF. 2012. What we know and don’t know about Earth's missing biodiversity. *Trends in Ecology and Evolution* 27:501–510. DOI: 10.1016/j.tree.2012.05.008.

Schmieder R., Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. DOI: 10.1093/bioinformatics/btr026.

Serejo CS., Secchin P., Cardoso I., Almeida TC. 2007. 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 RAS, OLAVO G, MARTINS AS eds. *Biodiversidade da fauna marinha profunda na costa central brasileira.* Rio de Janeiro: Museu Nacional, 133–162.

Stackebrandt E., Goebel BM. 1994. 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* 44:846–849. DOI: 10.1099/00207713-44-4-846.

Tâmega FTS., Oliveira PS., Figueiredo MAO. 2013. *Catalogue of the Benthic Marine Life from Peregrino Oil Field, Campos Basin, Brazil.* Rio de Janeiro.

Vieira LM., Migotto AE. 2011. Checklist dos Entoprocta do Estado de São Paulo , Brasil Checklist dos Entoprocta do Estado de São Paulo , Brasil. *Biota Neotropica* 11:0–5. DOI:

10.1590/S1676-06032011000500018.

Wang Y., Tian RM., Gao ZM., Bougouffa S., Qian P-Y. 2014. Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis. *PLoS ONE* 9:e90053. DOI: 10.1371/journal.pone.0090053.

Wu RSS. 1982. Effects of taxonomic uncertainty on species diversity indices. *Marine Environmental Research* 6:215–225. DOI: 10.1016/0141-1136(82)90055-1.

Zhang Z., Schwartz S., Wagner L., Miller W. 2000. A Greedy Algorithm for Aligning DNA Sequences. *JOURNAL OF COMPUTATIONAL BIOLOGY* 7:203–214. DOI: 10.1089/10665270050081478.

Zhang ZQ. 2011. Animal biodiversity: An introduction to higher-level classification and taxonomic richness. *Zootaxa*:7–12. DOI: http://www.mapress.com/zootaxa/list/2011/3148.html.

.

443 **TABLES AND FIGURES**

444 **Table 1 – Survey information.** Date, location and depth of sampling stations B3, B4, C2, F5 and
445 G2 in Campos Basin, southeast Brazil. Samples were collected as part of the Habitats Project
446 coordinated by CENPES/PETROBRAS. Coordinates are based on SIRGAS2000.

447

	Sampling date	Latitude	Longitude	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

448
450

451

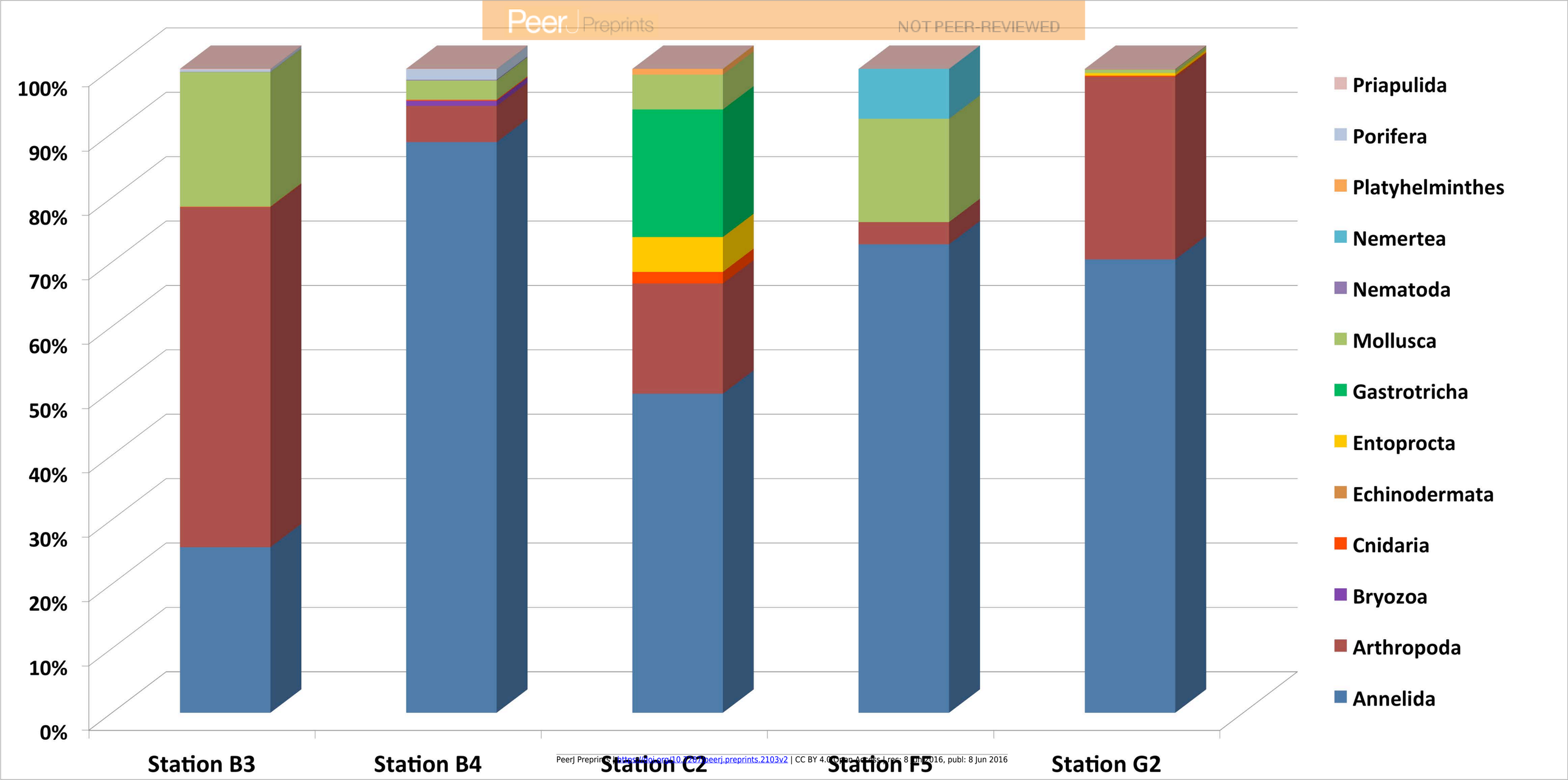
Table 2 – OTU per sample. OTU without a similar sequence on Genbank NR are under 'No Hits' fragments . OTU that did not comply with established LCA parameters (e.g. score bellow 100) or do not add up to a node are under 'non attributed reads'. Also under 'non-attributed' are Prokaryots attributed by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

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 occurrence in each station. Percentage of invertebrate OTU for phyla (A) and Family (B) in each station.

461

462



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v2 | CC BY 4.0 Open Access | rec: 8 Jun 2016, publ: 8 Jun 2016

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%

Station B3
Station B4
Station C2
Station F5
Station G2

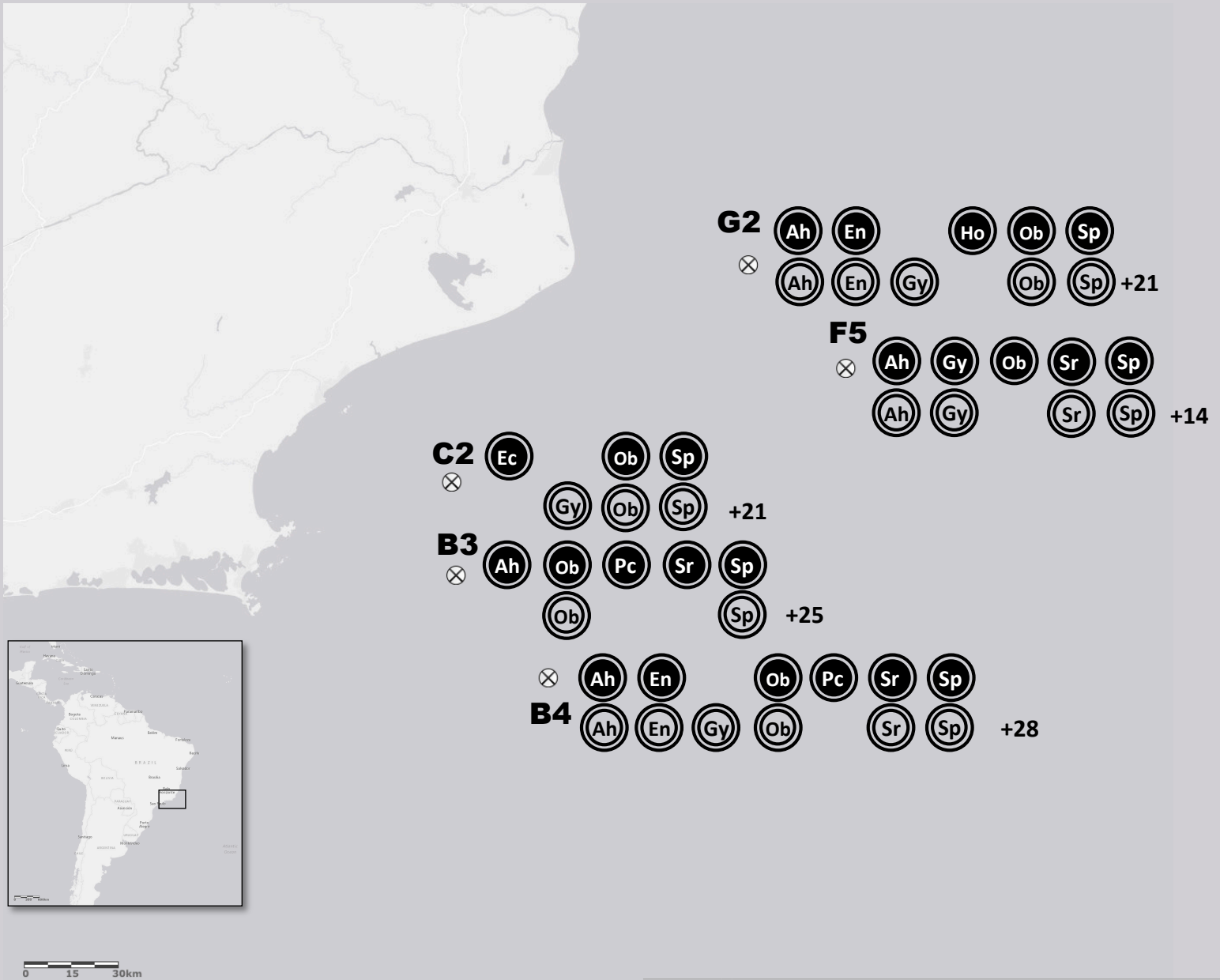
Holothuriidae
Solenoceridae
Cancridae
Mithracidae
Hippolytidae
Lysianassidae
Desmosomatidae
Mysidae
Cylindroleberididae
Entomobryidae
Vespidae
Formicidae
Chalcididae
Miridae
Priapulidae
Hormogastridae
Enchytraeidae
Echiuridae
Amphinomidae
Glyceridae
Serpulidae

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v2 | CC-BY 4.0 Open Access | rec: 8 Jun 2016, pub: 8 Jun 2016

Figure 2 – Geographical distribution of the main invertebrate phyla in Campos Basin. identified by Meta-Barcoding and previous morphological taxonomy studies in Campos Basin. A) Annelida distribution, b) Arthropoda distribution, C) Mollusca distribution. Full circles with initials of the invertebrate family name records the presence of families identified by Meta-barcoding while empty circles families identified by morphological taxonomy. Asterisks indicate the source of the morphological identification.

Annelida distribution

Molecular | Morphological



- Ah Amphinomidae
- En Enchytraeidae
- Ec Echiuridae*
- Gy Glyceridae
- Ho Hormogastridae**
- Ob Orbiniidae
- Pc Pectinariidae*
- Sr Serpulidae
- Sp Spionidae

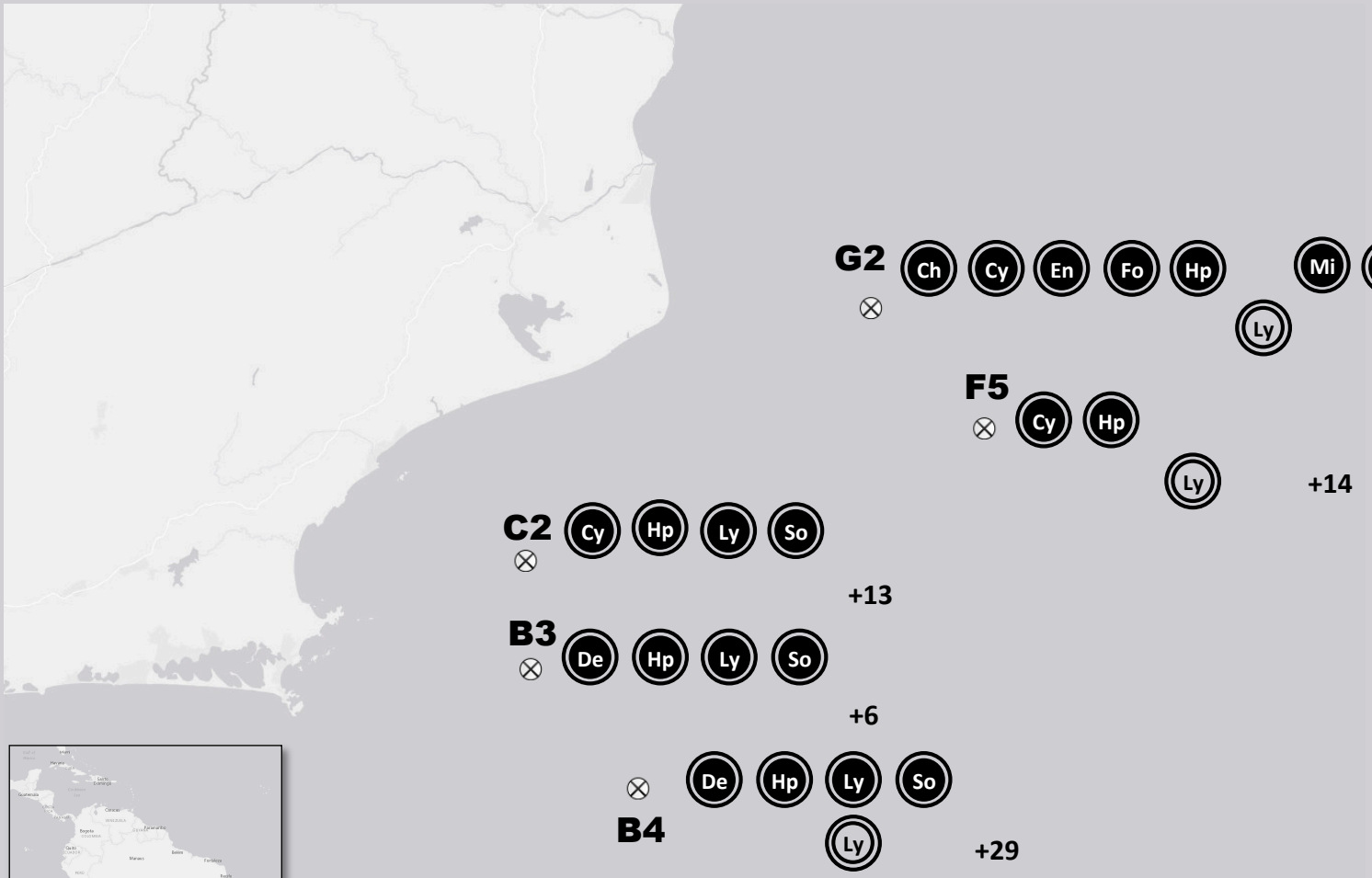
* Present in other stations of Habitats

** Non-marine family

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v2 | CC BY 4.0 Open Access | rec: 8 Jun 2016, publ: 8 Jun 2016

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

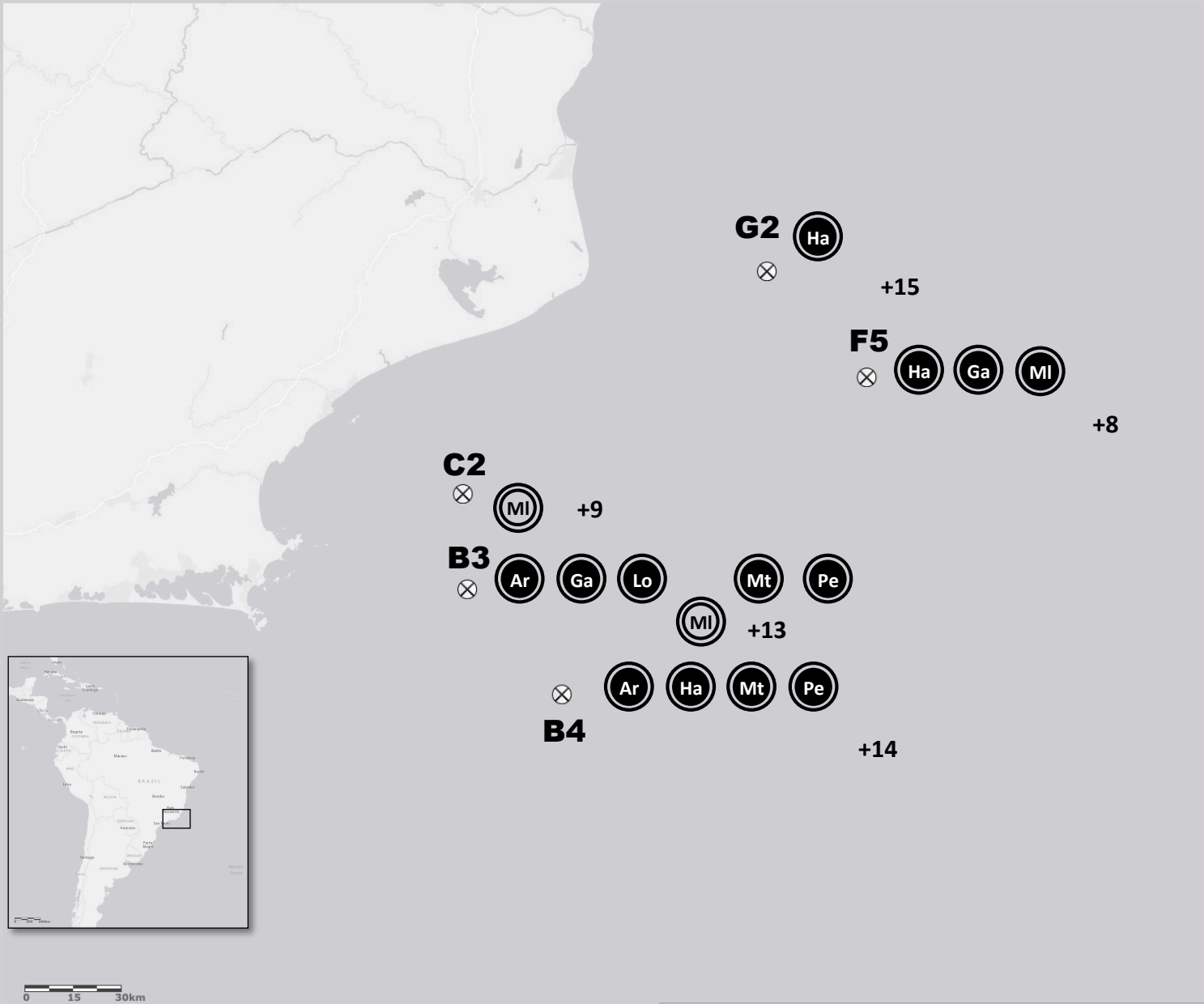
** Previous studies in Campos basin

***Non-marine family

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v2 | CC BY 4.0 Open Access | rec: 8 Jun 2016, publ: 8 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

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2103v2 | CC BY 4.0 Open Access | rec: 8 Jun 2016, publ: 8 Jun 2016

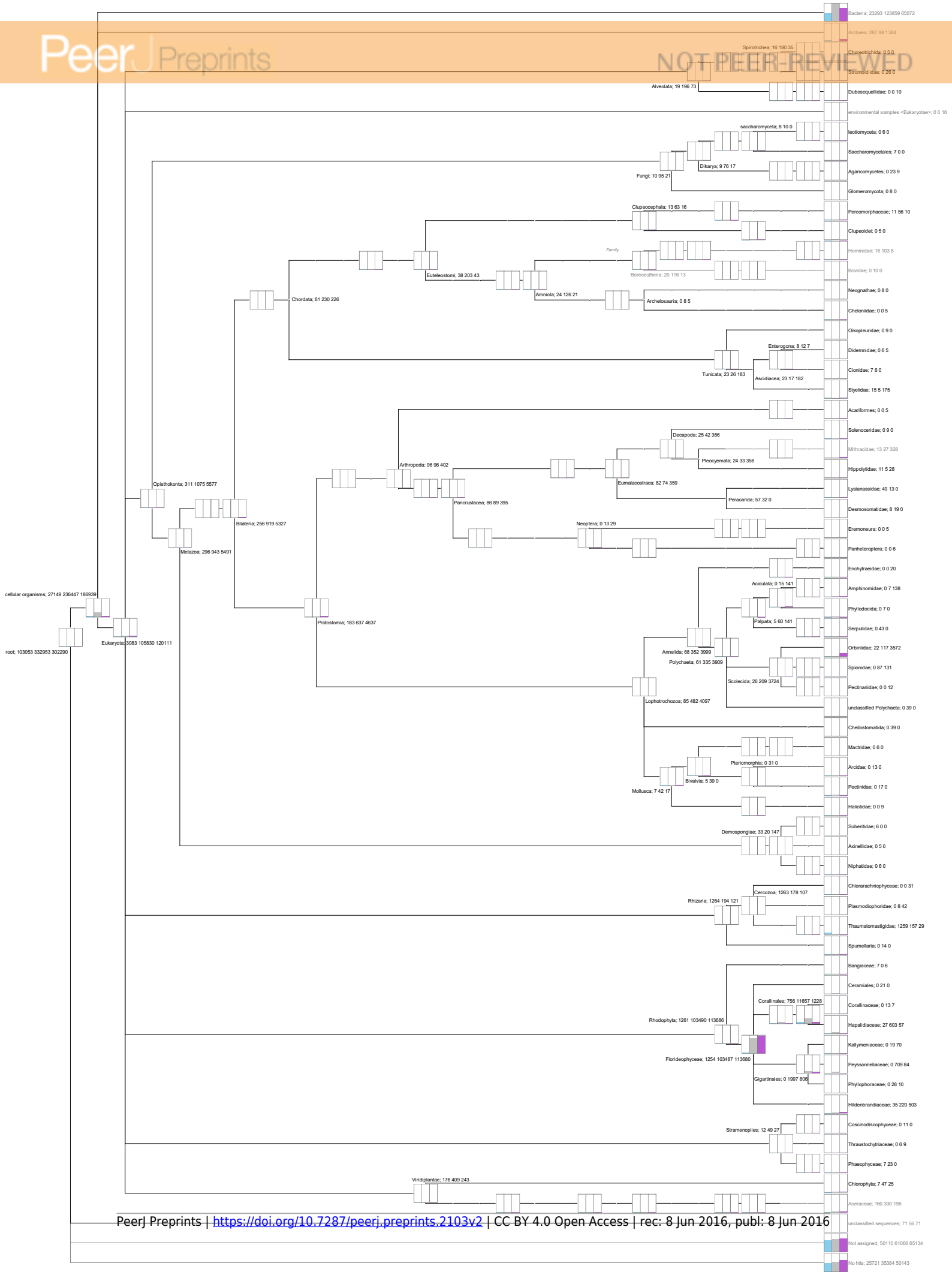
Supplementary Material 1 – PCR primers and conditions. 1-5 µL of DNA template, 1 µL (5µM) of primers Forward and reverse), 5 µl of 10X buffer, 2 µl of $MgCl_2$ (25 mM), 1 µl of dNTP 10 µM (Fermentas), 0.2 µl de Platinum Taq DNA Polymerase High Fidelity 5 U.µL-1 (Thermo Scientific) and ultra pure distilled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

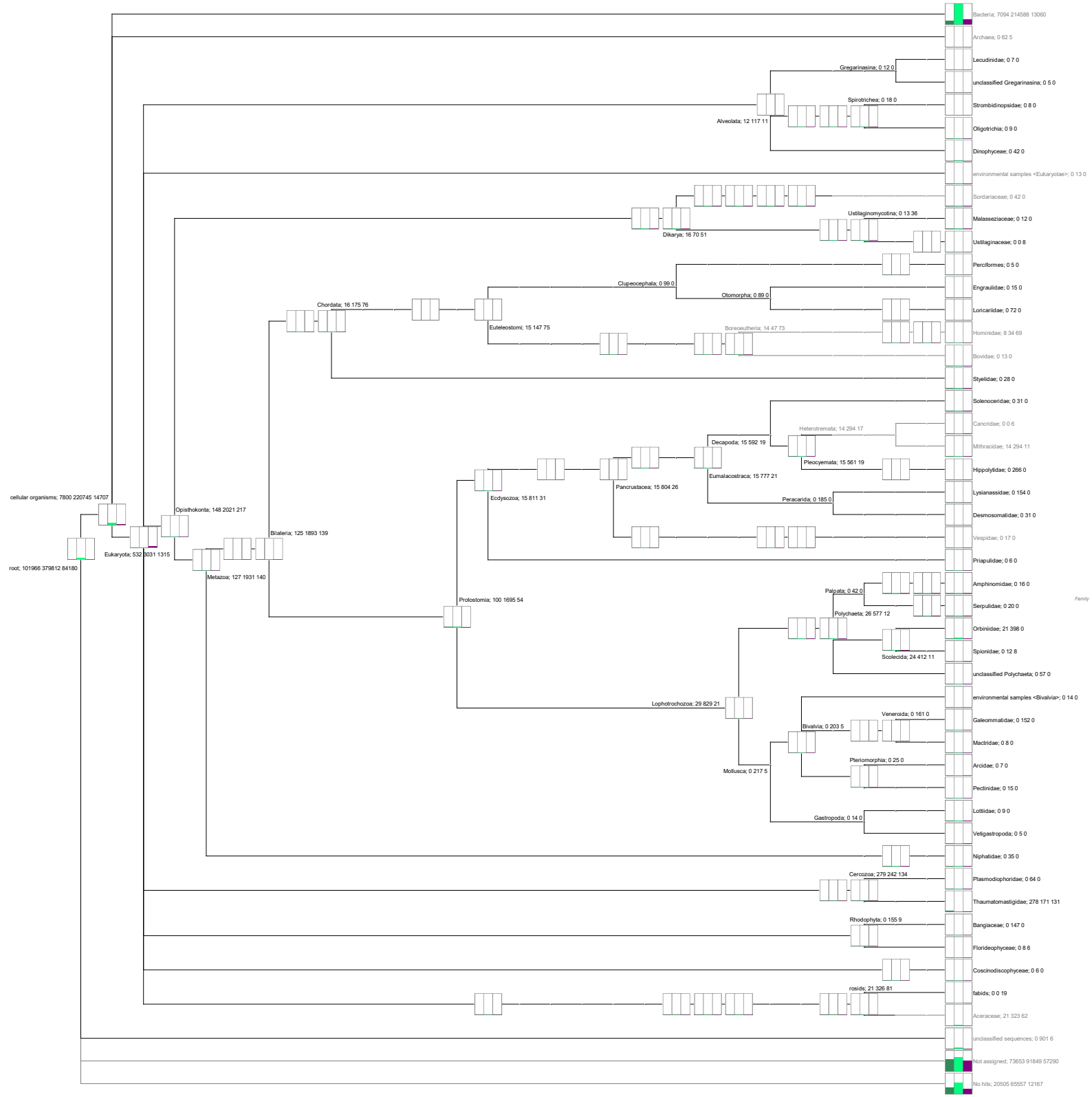
Target	Primer (F – Forward; R – reverse)	Denaturation	cycles	denaturation	annealing	Extension	Final extension	References
COI	TITCIAAYCAYAARGAYATTGG (F – jLCO1490); TAIACYTCIGGRTGICCRAARAAYCA (R – jHCO2198)	1' @94oC	10+30	30"@94oC	1'30"@61-52oC (-1oC per cycle) + 1'30"@61-52oC	1'@72oC	5'@72oC	Geller et 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

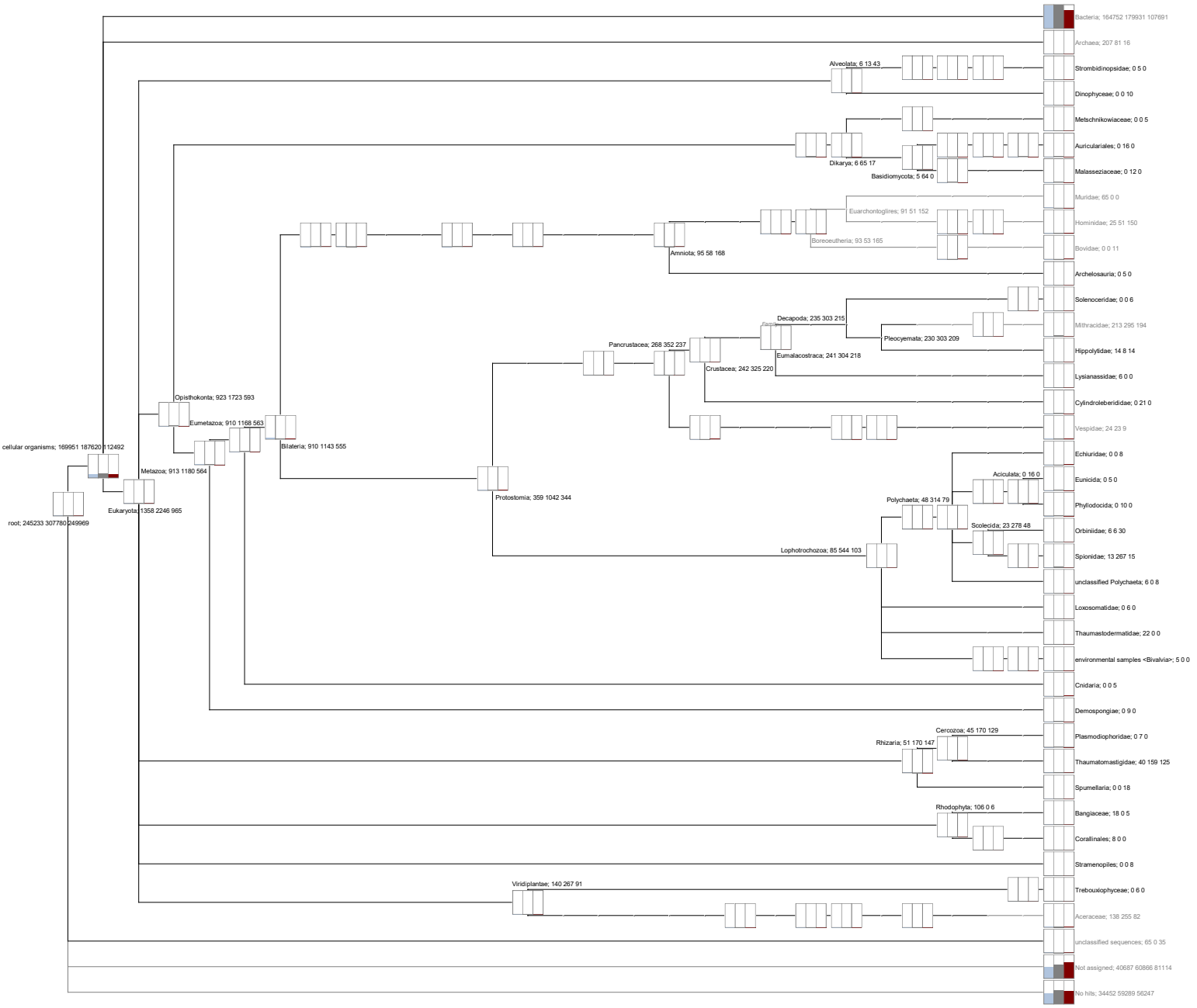
475 **Supplementary material 2 – Family level Cladograms of the 5 sampling stations.**

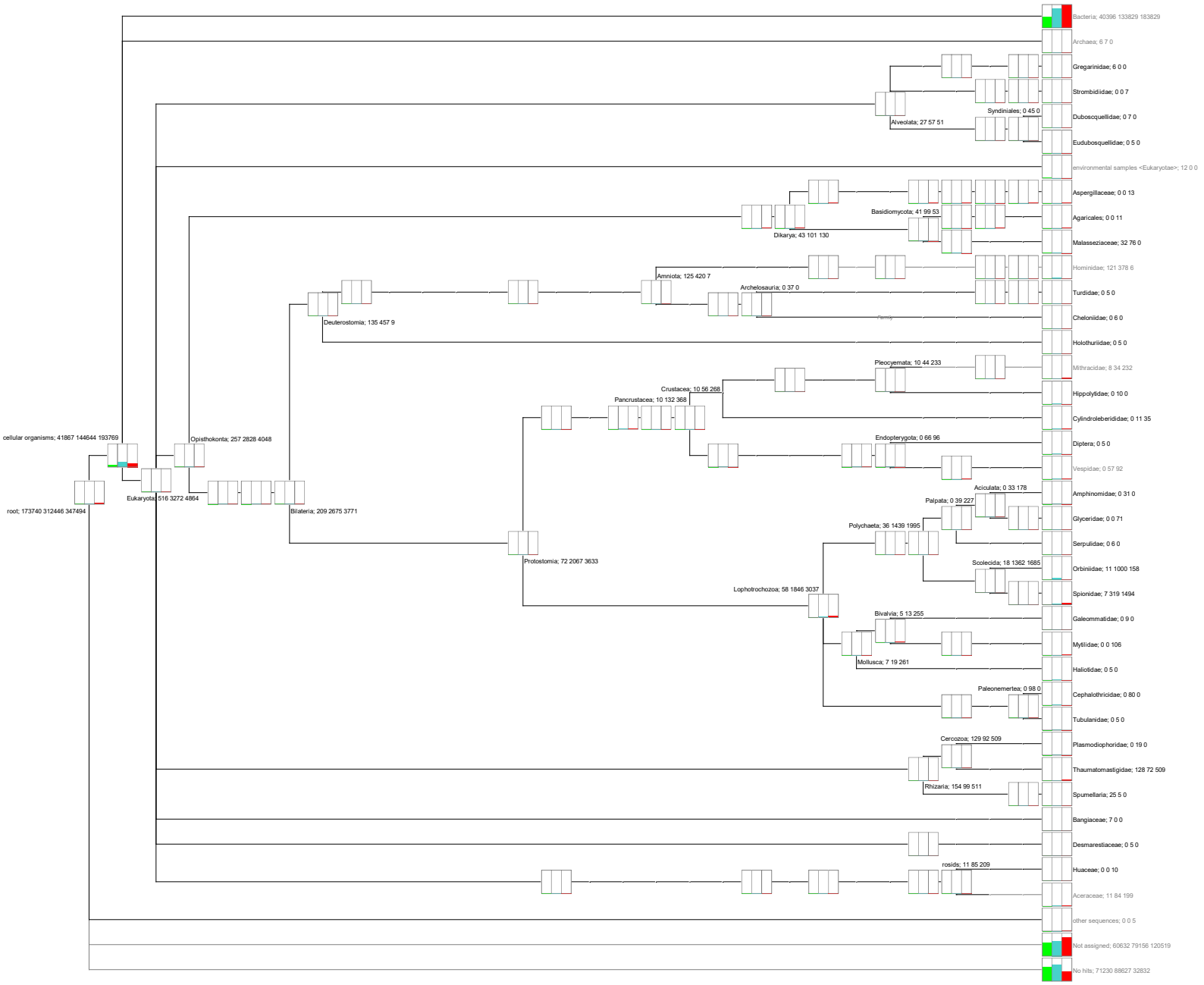
476 Cladograms were built using specimens identified with any of the 3 target genes. Bar
477 inside the squares represent the number of reads from each gene used to create the
478 node. A) Family cladogram for station B3; b) Family cladogram for station B4; C) Family
479 cladogram for station C2; D)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

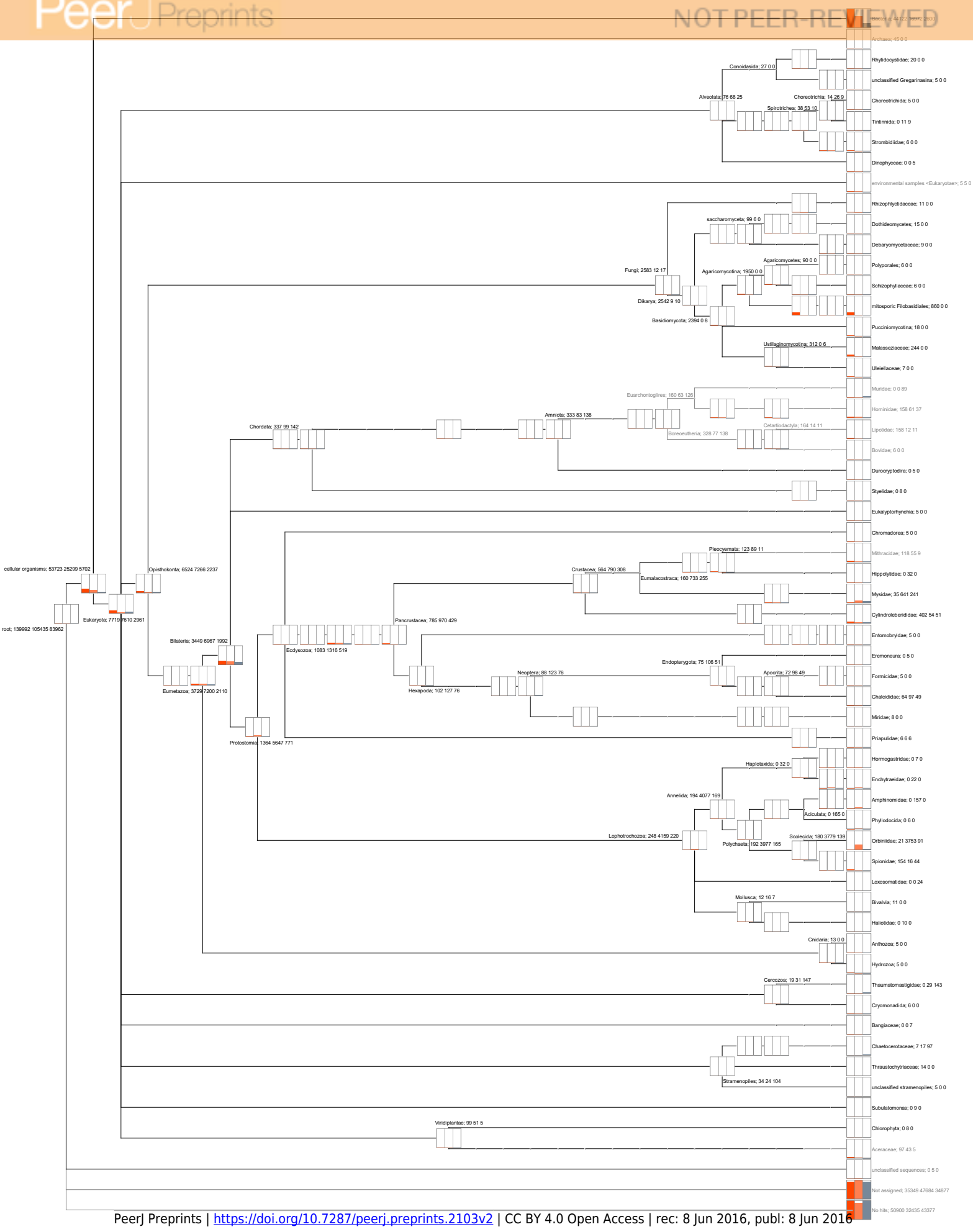
481






cellular organisms; 169951 187620 112492
root; 245233 307780 249969
Eukaryota; 1358 2246 965
Metazoa; 913 1180 564
Opisthokonta; 923 1723 593
Eumetazoa; 910 1168 563
Bilateria; 910 1143 555
Protostomia; 359 1042 344
Lophotrochozoa; 85 544 103
Pancrustacea; 268 352 237
Crustacea; 242 325 220
Eumalacostraca; 241 304 218
Decapoda; 235 303 215
Pleocyemata; 230 303 209
Amniota; 95 58 168
Boreoeutheria; 93 53 165
Euarchontoglires; 91 51 152
Dikarya; 6 65 17
Basidiomycota; 5 64 0
Alveolata; 6 13 43
Viridiplantae; 140 267 91
Rhizaria; 51 170 147
Cercozoa; 45 170 129
Rhodophyta; 106 0 6
Polychaeta; 48 314 79
Scolecida; 23 278 48
Aciculata; 0 16 0
Bacteria; 164752 179931 107691
Archaea; 207 81 16
Strombidinopsidae; 0 5 0
Dinophyceae; 0 0 10
Melschnikowiaceae; 0 0 5
Auriculariales; 0 16 0
Malasseziaceae; 0 12 0
Muridae; 65 0 0
Hominidae; 25 51 150
Bovidae; 0 0 11
Archelosauria; 0 5 0
Solenoceridae; 0 0 6
Mithracidae; 213 295 194
Hippolytidae; 14 8 14
Lysianassidae; 6 0 0
Cylindroleberididae; 0 21 0
Vespidae; 24 23 9
Echiuridae; 0 0 8
Euricida; 0 5 0
Phyllodocida; 0 10 0
Orbiniidae; 6 6 30
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
Plasmodiophoridae; 0 7 0
Thaumatomastigidae; 40 159 125
Spumellaria; 0 0 18
Bangiaceae; 18 0 5
Corallinales; 8 0 0
Stramenopiles; 0 0 8
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
Alveolata; 27 57 51
Dubosquellidae; 0 7 0
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
Cheloniidae; 0 6 0
Holothuriidae; 0 5 0
Pleocyemata; 10 44 233
Mithracidae; 8 34 232
Crustacea; 10 56 268
Pancrustacea; 10 132 368
Hippolytidae; 0 10 0
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
Aciculata; 0 33 178
Amphinomidae; 0 31 0
Palpata; 0 39 227
root; 173740 312446 347494
Bilateria; 209 2675 3771
Glyceridae; 0 0 71
Polychaeta; 36 1439 1995
Serpulidae; 0 6 0
Scolecida; 18 1362 1685
Orbiniidae; 11 1000 158
Protostomia; 72 2067 3633
Spionidae; 7 319 1494
Lophotrochozoa; 58 1846 3037
Galeommatidae; 0 9 0
Bivalvia; 5 13 255
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



Supplementary material 3 – Eligible and identified species by molecular taxonomy. List of the species identified by molecular biology taxonomy in this project in spite of previous records in the literature for Campos Basin ('Identified at': Schettini et al., 2016); non-identified by molecular taxonomy but with previous records for Campos Basin ('Identified at': Habitat) and eligible species identified by molecular taxonomy after manual text search among BLAST hits ('Identified at': Habitats and BLAST hits). The + signal indicates the presence of at least one sequence for the genetic marker in Genbank.

Specie	rRNA18S	rRNA28S	COI	Identified in
Cnemidocarpa verrucosa	+	+	+	Schettini et al., 2016
Desmarestia dudresnayi	+	+	+	Schettini et al., 2016
Erythrophyllum delesserioides	+	+	+	Schettini et al., 2016
Eurythenes gryllus	+	+	+	Schettini et al., 2016
Galeomma turtoni	+	+	+	Schettini et al., 2016
Grifola frondosa	+	+	+	Schettini et al., 2016
Haliotis diversicolor	+	+	+	Schettini et al., 2016
Hormogaster redii	+	+	+	Schettini et al., 2016
Lysmata seticaudata	+	+	+	Schettini et al., 2016
Malassezia globosa	+	+	+	Schettini et al., 2016
Marenzelleria arctia	+	+	+	Schettini et al., 2016
Mimachlamys varia	+	+	+	Schettini et al., 2016
Mysidium columbiae	+	+	+	Schettini et al., 2016
Parotocinclus maculicauda	+	+	+	Schettini et al., 2016
Pinctada imbricata	+	+	+	Habitats and BLAST hits
Platynereis dumerilii	+	+	+	Habitats and BLAST hits
Pontocaris lacazei	+	+	+	Habitats
Praxillella affinis	+	+	+	Habitats
Progoniada regularis	+	+	+	Habitats and BLAST hits
Protodorvillea kefersteini	+	+	+	Habitats
Pteria colymbus	+	+	+	Habitats
Scalibregma inflatum	+	+	+	Habitats and BLAST hits
Scapharca broughtonii	+	+	+	Schettini et al., 2016
Serpula vermicularis	+	+	+	Schettini et al., 2016
Syllis gracilis	+	+	+	Habitats and BLAST hits
Syllis variegata	+	+	+	Habitats and BLAST hits
Travisia brevis	+	+	+	Habitats and BLAST hits
Travisia forbesii	+	+	+	Habitats and BLAST hits
Travisia pupa	+	+	+	Habitats and BLAST hits
Aglaophamus circinata		+	+	Habitats and BLAST hits
Alpheus formosus		+	+	Habitats
Amphipholis squamata		+	+	Habitats
Aricidea wassi		+	+	Habitats and BLAST hits
Chelonia mydas		+	+	Schettini et al., 2016
Praxillella pacifica		+	+	Habitats and BLAST hits
Priapulus caudatus		+	+	Schettini et al., 2016
Scolelepis bonnieri		+	+	Schettini et al., 2016
Scolelepis foliosa		+	+	Schettini et al., 2016
Amphimedon queenslandica	+		+	Schettini et al., 2016
Axiothella rubrocincta	+		+	Habitats and BLAST hits
Bathyarca pectunculoides	+		+	Habitats
Bathyglycinde profunda	+		+	Habitats
Bathyglycinde sibogana	+		+	Habitats
Caprella equilibra	+		+	Habitats and BLAST hits
Ceratocephale abyssorum	+		+	Habitats and BLAST hits
Ciona intestinalis	+		+	Schettini et al., 2016
Clymenella torquata	+		+	Habitats and BLAST hits
Pectinaria granulata	+		+	Schettini et al., 2016
Perna viridis	+		+	Schettini et al., 2016
Protaspis grandis	+		+	Schettini et al., 2016

Specie	rRNA18S	rRNA28S	COI	Identified in
Syllis hyalina	+		+	Habitats and BLAST hits
Didemnum candidum			+	Schettini et al., 2016
Leodamas rubra			+	Habitats and BLAST hits
Leodia sexiesperforata			+	Habitats
Leptochelia dubia			+	Habitats
Leucothoe urospinosa			+	Habitats and BLAST hits
Lumbrineris latreilli			+	Habitats and BLAST hits
Lysidice ninetta			+	Habitats and BLAST hits
Lysmata anchisteus			+	Schettini et al., 2016
Macrochaeta clavicornis			+	Habitats
Marphysa bellii			+	Habitats and BLAST hits
Mendicula ferruginosa			+	Habitats and BLAST hits
Mooreonuphis pallidula			+	Habitats and BLAST hits
Neanthes acuminata			+	Habitats and BLAST hits
Nereimyra punctata			+	Habitats and BLAST hits
Notomastus latericeus			+	Habitats and BLAST hits
Ophelina acuminata			+	Habitats and BLAST hits
Pyropia haitanensis			+	Schettini et al., 2016
Scapharca kagoshimensis			+	Schettini et al., 2016
Scoloplos armiger			+	Schettini et al., 2016
Isolda pulchella			+	Habitats and BLAST hits
Apophlaea lyallii	+	+		Schettini et al., 2016
Chaetoceros curvisetus	+	+		Schettini et al., 2016
Coelomactra antiquata	+	+		Schettini et al., 2016
Crassinella lunulata	+	+		Habitats and BLAST hits
Cryptococcus friedmannii	+	+		Schettini et al., 2016
Cyclaspis alba	+	+		Habitats
Cylichna alba	+	+		Habitats and BLAST hits
Engraulis japonicus	+	+		Schettini et al., 2016
Euclymene oerstedi	+	+		Habitats and BLAST hits
Eulalia viridis	+	+		Habitats and BLAST hits
Eumida sanguinea	+	+		Habitats and BLAST hits
Exogone dispar	+	+		Habitats and BLAST hits
Galathowenia oculata	+	+		Habitats
Glycera americana	+	+		Habitats and BLAST hits
Glycera southeastatlantica	+	+		Habitats and BLAST hits
Goniada emerita	+	+		Habitats
Hesiospina aurantiaca	+	+		Habitats and BLAST hits
Patelloida striata	+	+		Schettini et al., 2016
Scopelocheirus schellenbergi	+	+		Schettini et al., 2016
Subulatomonas tetraspora	+	+		Schettini et al., 2016
Ophelina cylindricaudata		+		Habitats and BLAST hits
Ophiactis lymani		+		Habitats
Trypanosyllis zebra		+		Habitats and BLAST hits
Ahnfeltiopsis leptophylla	+			Schettini et al., 2016
Crucigera zygophora	+			Schettini et al., 2016
Leitoscoloplos pugettensis	+			Schettini et al., 2016
Malassezia nana	+			Schettini et al., 2016
Ophiura ljungmani	+			Habitats
Owenia fusiformis	+			Habitats and BLAST hits
Panthalis oerstedi	+			Habitats and BLAST hits
Paralacydonia paradoxa	+			Habitats and BLAST hits
Paramphinome jeffreysii	+			Habitats and BLAST hits
Pholoe minuta	+			Habitats
Phtisica marina	+			Habitats
Phyllodoce longipes	+			Habitats and BLAST hits
Solenocera crassicornis	+			Schettini et al., 2016
Strombidium paracalkinsi	+			Schettini et al., 2016
Phagomyxa odontellae	+			Schettini et al., 2016

490