NRC Research Press

NOTE

Rare unclassified 16S rRNA operational taxonomic units from the uncharted Engaño Bay (Argentinean Patagonia)

Andrea Y. Calvo, Julieta M. Manrique, and Leandro R. Jones

Abstract: Rare microbes make up most of the diversity of marine microbiomes, and recent works have highlighted their importance for microbial community dynamics and in fragmented habitats. Rare taxa have been infrequently studied in comparison with abundant groups, and rare unclassified sequences are common in culture-independent studies. Here, we describe a detailed analysis of nonclassifiable sequences from the Chubut river estuary at the Argentinean Patagonia. Standard taxonomic assignments of environmental 16S rRNA sequences resulted in about 13% unclassified operational taxonomic units (OTUs). The potential affiliations of these OTUs could be narrowed by mapping the classification software assignments on a phylogeny obtained directly from our environmental sequence data. Customized BLAST analyses were remarkably consistent with these phylogenetic assignments, especially when the unclassified OTUs were blasted against sequences from cultured and type microorganisms. In addition, our BLAST analyses revealed significant similarities between several unclassified OTUs and a plethora of unclassified sequences from around the world. Further phylogenetic comparisons with 6194 carefully selected reference sequences showed that these unclassified sequences may correspond to 5 unnamed groups, possibly encompassing ranks from subclass to family inside the *Alphaproteobacteria*, and to an unknown *Gracilibacteria* lineage. Overall, these results demonstrate the value of straight phylogenetic analysis, customized BLAST searches, and comparisons with sequences from type material, for the systematic study of rare unclassified sequences.

Key words: 16S, environmental DNA, taxonomy, picoplankton, Patagonia.

Résumé : Les microbes rares composent la plus grande partie de la diversité du microbiome marin et des travaux récents ont souligné leur importance dans la dynamique de la communauté microbienne et dans les habitats fragmentés. Les taxons rares ont été peu étudiés comparativement aux groupes abondants et les séquences rares non classées sont fréquentes dans les études ne reposant pas sur la culture des souches. Les auteurs décrivent ici une analyse détaillée de séquences qui ne peuvent être classées provenant de l'estuaire du Rio Chubut en Patagonie Argentine. Les attributions taxonomiques standard de séquences d'ARNr 16S de l'environnement produisaient environ 13 % d'*UTO* qui ne pouvaient être classées. Les affiliations potentielles de ces *UTO* pouvaient être resserrées en cartographiant les attributions obtenues avec un logiciel de classification dans une phylogénie obtenue directement des données de séquences environnementales. Les analyses BLAST sur mesure étaient remarquablement cohérentes avec ces attributions phylogénétiques, particulièrement lorsque les *UTO* non classées étaient alignées avec des séquences de microorganismes cultivés et des microorganismes types. De plus, les analyses BLAST révélaient des similarités significatives entre plusieurs *UTO* non classées et un grand nombre de séquences non classées à travers le monde. D'autres comparaisons phylogénétiques de 6194 séquences de référence soigneusement choisies ont montré que ces séquences non classées pourraient correspondre à cinq groupes non identifiés, englobant possiblement des rangs allant des sous-classes aux familles à l'intérieur des *Alphaproteobacteria* et un lignage inconnu des *Gracilibacteria*. Globalement, ces résultats démontrent la valeur d'une analyse phylogénétique directe, de recherches BLAST sur mesure et de comparaisons de séquences de matériel type pour l'analyse systématique de séquences rares qui ne peuvent être classées. [Traduit par la Rédaction]

Mots-clés : 16S, ADN environnemental, taxonomie, picoplancton, Patagonie.

Received 1 June 2017. Revision received 23 October 2017. Accepted 26 October 2017.

A.Y. Calvo, J.M. Manrique, and L.R. Jones. Laboratorio de Virología y Genética Molecular, Facultad de Ciencias Naturales y Ciencias de la Salud (FCNyCS), Universidad Nacional de la Patagonia San Juan Bosco (UNPSJB), 9 de Julio y Belgrano s/n (9100), Trelew, Chubut, Argentina; Consejo Nacional de Investigaciones Científicas y Técnicas, Avenida Rivadavia 1917 (C1083ACA) Buenos Aires, Argentina.

Corresponding author: Leandro R. Jones (emails: lrj000@gmail.com; ljones@conicet.gov.ar).

Marine bacteria constitute a large and complex assemblage of taxa. This huge diversity is in a discovery stage, and unclassified operational taxonomic units (uOTUs) are common in environmental DNA studies. This problem especially affects rare taxa, which despite contributing a large share of marine microbial diversity, remain untamed compared with abundant groups (Schloss et al. 2016). Recent works have shown that rare taxa can be important to community dynamics (Shade et al. 2014) and act as a biodiversity reservoir; hence, they are potentially relevant for future environmental change (Choi et al. 2017). Furthermore, it has been observed that some groups that are rare in seawater can be important in particular marine habitats (Sunagawa et al. 2010).

In this work, we set out to shed light on the systematics of rare uOTUs identified among 16S ribosomal gene (rRNA) environmental sequences (eSEQs) from the Argentinean Patagonia. We were particularly interested in identifying unclassified sequences (uSEQs) presenting similarities to sequences from elsewhere because cosmopolitan lineages are more likely to represent conditionally rare groups (Shade et al. 2014), to be important in fragmentary habitats (Sunagawa et al. 2010), and to potentially impact the future microbial biodiversity (Shade et al. 2014; Choi et al. 2017). We combined phylogenies obtained directly from our 16S rDNA sequence data with assignments obtained by a standard classification software and performed customized BLAST analyses and further phylogenetic comparisons with carefully selected reference sequences. The workflow implemented in this study is outlined in supplemental Fig. S1[1]. Several uOTUs were closely related to previously described uSEQs from around the world. We provide evidence that these cosmopolitan uSEQs may belong to 5 undescribed *Alphaproteobacteria* lineages and an unknown *Gracilibacteria* clade.

A seawater sample (N, supplemental Table S1[1]) collected in the uncharted Engaño Bay at about 2 km seawards from the mouth of the Chubut River (43.34°S, 65.03°W) was used to generate high-quality 16S rRNA sequences by high-throughput sequencing (HTS; supplemental Materials and Methods[1]). These sequences were grouped in 97% similarity OTUs, and the obtained OTUs were classified up to the Order rank with Wang's Naïve Bayesian Classifier (WNBC) implemented in Mothur (Schloss et al. 2011) using an 80% cutoff (supplemental Materials and Methods[1]). For comparison, 3 independent samples (F, J, A; supplemental Table S1[1]) were sequenced by molecular cloning and Sanger sequencing and analyzed in an equivalent manner (supplemental Materials and Methods[1]).

About 13% of the Patagonian OTUs could not be classified

The WNBC analyses reflected a microbial diversity similar to that of environments akin from elsewhere. A few bacterial groups dominated the bacterioplankton and were accompanied by several taxa present at relatively low frequencies (supplemental Table S1[1]). We also observed significant amounts of 16S sequences from picoalgae plastids. Out of 250 OTUs, 33 could not be classified (Table 1; supplemental Table S2[1]). Four OTUs could not be classified in any bacterial phylum and 6 were assigned to *Proteobacteria*. Additionally, 7 OTUs could not be assigned beyond *Gammaproteobacteria* and 16 OTUs seemingly corresponded to unknown *Alphaproteobacteria*. There were no significant differences between the number of unclassified Sanger and HTS sequences (supplemental Table S3[1]). However, the proportion of unclassified Sanger OTUs was slightly lower than the proportion of HTS ones (supplemental Table S3[1]), probably due to the higher resolution of Sanger sequences (longer sequences).

Straight phylogenetic analysis (SPA) shed light on the affiliations of the uOTUs

We reassessed the affiliation of our uOTUs by establishing their phylogenetic relations with OTUs that could be classified by the WNBC analyses. To this aim, we inferred a phylogeny, and the WNBC taxonomic assignments were mapped on the obtained tree (Fig. 1). To align the sequences, we used a multistep, taxonomy-guided approach, in which we fed the sequence alignment algorithm with taxonomic information to facilitate homology detection, after which, the obtained alignment was submitted to maximum-likelihood phylogenetic analysis (supplemental Materials and Methods[1]). The obtained results indicated that some uOTUs may belong or be related to known groups. The uOTUs 0081, 0098, 0136, 0147, and 0165 were clustered in the SAR11 clade (*Pelagibacterales*), which was supported by a relatively high bootstrap support (Fig. 1*a*). Two uOTUs (0022 and 0138) were distantly related to the SAR406 clade of *Deferribacterales* (*Marinimicrobia*) (Fig. 1*b*), and likewise, the uOTU 0170 was distantly related to *Gracilibacteria* sequences (Fig. 1*b*). Three extra uOTUs formed monophyletic groups with the *Flavobacteriales* (uOTU 0159), OCS116 (*Rhodobacterales*; uOTU 0067) and SAR86 (*Oceanospirillales*) (uOTU 050 (Fig. 1*a*). Several uOTUs were close to sequences classified as *Oceanospirillales* (Fig. 1*a*). Likewise, many uOTUs were located close to sequences classified as *Rickettsiales* and *Rhodobacterales* (Fig. 1*a*). Two uOTUs (0103 and 0146) did not cluster monophyletically with other sequences and were very divergent with respect to the rest of sequences (Figs. 1*a* and 1*c*). These results are summarized in Table 1 and supplemental Table S2[1].

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjm-2017-0342.

Table 1. Operational taxonomic units that could not be classified by WNBC analyses (uOTUs), and summary of SPA and BLAST results.

uOTU	WNBC[a]	SPA[b]	BLAST hit[c]	Taxon[d]	Homology[e]
034	*Alphaproteobacteria*	*Rhodobacterales*	JN591806.1	PS1 clade	93%
043	*Proteobacteria*	*Rhodobacterales*	EU802859.1	*Bacteria*	90%
050	*Gammaproteobacteria*	SAR86	EU802840.1	*Bacteria*	91%
0022	*Bacteria*	*Deferribacterales*	JQ199557.1	*Bacteria*	**100%**
0058	*Alphaproteobacteria*	*Rhodobacterales*	EU799247.1	*Bacteria*	94%
0067	*Alphaproteobacteria*	OCS116	JN591913.1	PS1 clade	94%
0076	*Gammaproteobacteria*	*Oceanospirillales*	FR686176.1	*Bacteria*	96%
0078	*Alphaproteobacteria*	*Rhodobacterales*	FR685064.1	*Bacteria*	**99%**
0081	*Alphaproteobacteria*	SAR11	GQ346819.1	*Alphaproteobacteria*	**97%**
0098	*Alphaproteobacteria*	SAR11	EU802479.1	*Bacteria*	95%
0099	*Gammaproteobacteria*	*Oceanospirillales*	GU234924.1	*Bacteria*	96%
0101	*Alphaproteobacteria*	*Rhodobacterales*	FR683303.1	*Bacteria*	93%
0102	*Gammaproteobacteria*	*Oceanospirillales*	KX935466.1	*Bacteria*	95%
0103	*Gammaproteobacteria*	—	HQ163109.1	*Gammaproteobacteria*	95%
0105	*Proteobacteria*	*Oceanospirillales*	FR684004.1	*Bacteria*	93%
0112	*Alphaproteobacteria*	*Rhodobacterales*	DQ234144.2	*Rhodobacteraceae*	92%
0113	*Alphaproteobacteria*	*Rhodobacterales*	EU799247.1	*Bacteria*	95%
0118	*Alphaproteobacteria*	*Rhodobacterales*	FJ825915.1	*Bacteria*	93%
0127	*Alphaproteobacteria*	*Rickettsiales*	JF488534.1	*Alphaproteobacteria*	**98%**
0128	*Gammaproteobacteria*	*Oceanospirillales*	JQ269288.1	*Bacteria*	92%
0135	*Alphaproteobacteria*	*Rhodobacterales*	FR685064.1	*Bacteria*	96%
0136	*Alphaproteobacteria*	SAR11	HQ672177.1	*Bacteria*	**97%**
0138	*Bacteria*	*Deferribacterales*	KU034616.1	*Bacteria*	**97%**
0145	*Proteobacteria*	*Rickettsiales*	EU802860.1	*Bacteria*	90%
0146	*Proteobacteria*	—	JN986470.1	*Bacteria*	96%
0147	*Alphaproteobacteria*	SAR11	NR_074224.1	SAR11	95%
0148	*Alphaproteobacteria*	*Rickettsiales*	JF488580.1	*Alphaproteobacteria*	**98%**
0155	*Proteobacteria*	*Oceanospirillales*	HQ163230.1	*Actinobacteria*	94%
0159	*Bacteria*	*Flavobacteriales*	KC336640.1	*Bacteria*	89%
0165	*Proteobacteria*	SAR11	NR_074224.1	SAR11	95%
0168	*Gammaproteobacteria*	*Oceanospirillales*	JQ269279.1	*Bacteria*	94%
0170	*Bacteria*	*Gracilibacteria*	KU688616.1	*Bacteria*	**97%**
0176	*Alphaproteobacteria*	*Rickettsiales*	JQ199248.1	*Bacteria*	93%

[a]Wang's Naïve Bayesian Classifier (WNBC) assignment (80% cutoff).
[b]SPA, straight phylogenetic analysis. Closer taxon according to phylogenetic analysis.
[c]GenBank match ID.
[d]GenBank taxonomy of the matching sequence.
[e]Percent identity. Bold indicates significant homology.

BLAST analyses revealed similarities to uSEQs from remote regions

The uOTUs were further compared with those in the GenBank database, with the aim of detecting similarities to uSEQs from elsewhere and (or) taxa not represented among our sequences. We performed comparisons with the whole database and with the subsets of sequences corresponding to cultured organisms and type specimens (Federhen 2015). The corresponding results are summarized in Table 1 and detailed in supplemental Tables S4–S7[1]. Eight uOTUs were very similar to sequences from remote places (supplemental Table S7[1]). The uOTU 0022 displayed 100% homology with an eSEQ from California (USA), and it was 99% identical to a sequence from an isolate from the Jiulong River estuary in China. This uOTU also displayed similarities above 97% to many sequences from several places worldwide (supplemental Table S7[1]). The uOTUs 0078 and 0081 were very similar to sequences amplified from fjords in Norway and Vancouver Island, respectively. In addition, the uOTU 0081 was significantly similar (>97%) to a sequence obtained from the northeast subarctic Pacific Ocean (supplemental Table S7[1]). The uOTUs 0136 and 0138 were 97% homologous with sequences from the North Pacific and Austria, respectively. The uOTUs 0127 and 0148 were very similar to *Alphaproteobacteria* isolates from the Gulf of Maine (USA) and to 65 eSEQs from diverse geographic origins (Table S7[1]). The uOTU 0170 displayed a 97% homology with eSEQs from Italy (Table S7[1]).

As expected, the similarities to type material sequences were rather low, but interestingly, there was quite a good concordance with our SPA assignments. For example, phylogenetic and BLAST comparisons with both cultured and type material suggested a relationship with *Rhodobacterales* for the uOTU 0058 (supplemental

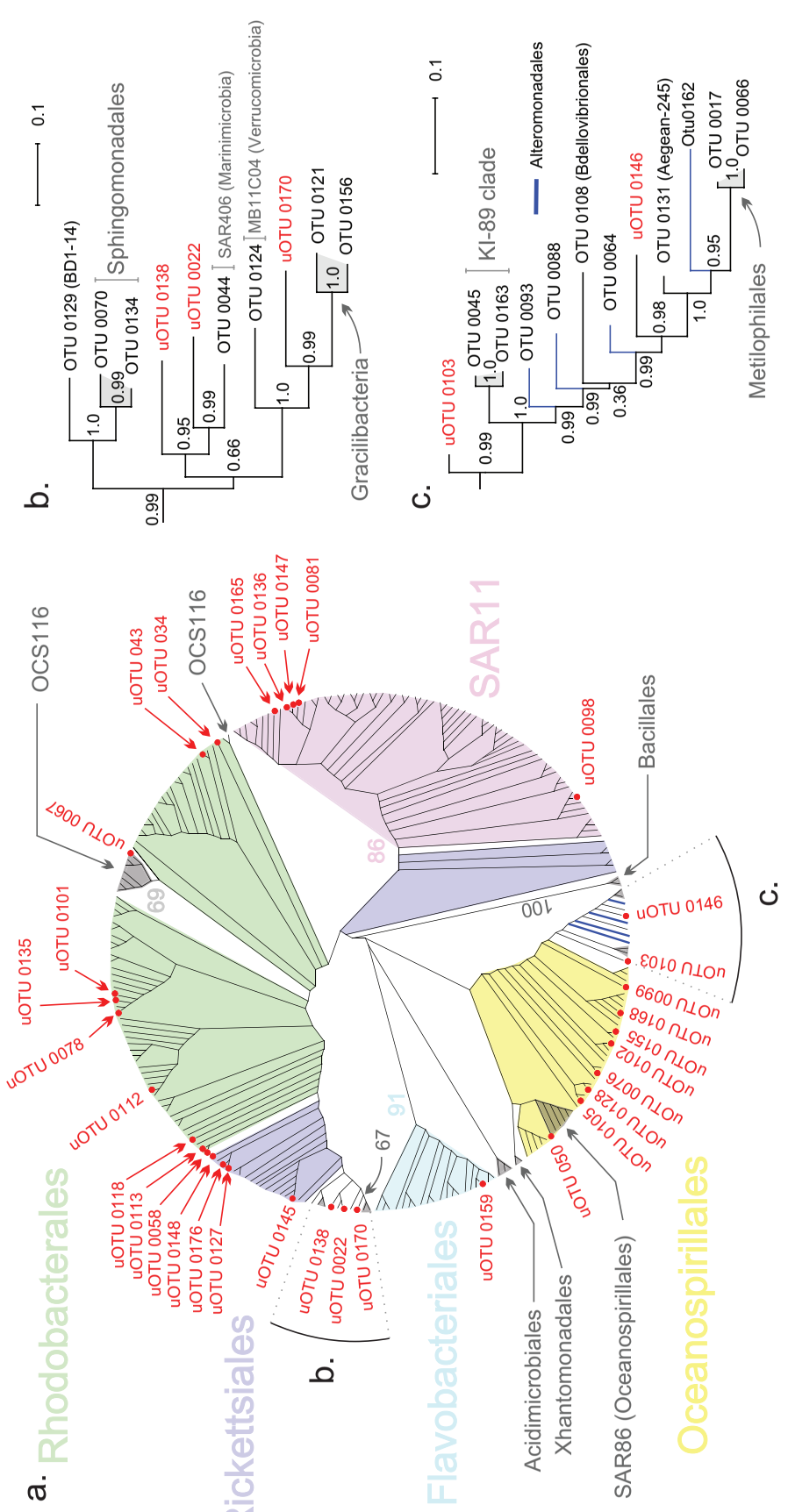

Fig. 1. Straight phylogenetic analysis (SPA) of 250 97% operational taxonomic units (OTUs) from the uncharted Engaño Bay. Panel *a* displays the obtained tree, with Wang's Naïve Bayesian Classifier taxonomy mapped on it. Panels *b* and *c* show detailed views of the clades indicated as *b* and *c*, respectively, in panel *a*. Numbers close to branches indicate bootstrap supports (n = 100, panel *a*) or aBayes local supports (panels *b* and *c*). The scale bars' units in panels *b* and *c* are substitutions per aligned position. [Colour online.]

Table S2[1]), and likewise, the uOTUs 0081 and 0136 were monophyletically clustered with SAR11 sequences and more similar to SAR11 type material (supplemental Table S2[1]).

The cosmopolitan uSEQs may belong to unknown *Alphaproteobacteria* and *Gracilibacteria* lineages

Our SPA suggested potential relationships between the cosmopolitan uOTUs and several *Alphaproteobacteria* groups and the phyla *Gracilibacteria*, *Verrucomicrobia*, and *Marinimicrobia* (Fig. 1). Thus, we combined the uOTU sequences and the similar sequences from around the world (supplemental Table S7[1]) with 6194 reference sequences from these taxa, and submitted the entire data to phylogenetic analysis, as detailed in section 2 (Extended uOTU analyses) of supplementary Materials and Methods[1]. In agreement with the SPA, the obtained phylogeny showed that the uOTU 0078 and the similar sequences from elsewhere may correspond to a divergent *Rhodobacterales* lineage and that the uOTUs 0127 and 0148, together with 66 sequences from different places worldwide, may belong to a divergent *Rickettsiales* family (supplemental Fig. S4[1]). In addition, the analysis also corroborated that the uOTUs 0136 and 0081 and 2 unknown sequences from Vancouver Island and the northeast Pacific Ocean were closely related to the SAR11 clade. The uOTU 0138 and 6 uSEQs from elsewhere were clustered in a long branch, close to the Candidatus *Midichloriaceae* (*Rickettsiales*) family. Interestingly, the uOTU 0022 and 159 sequences from diverse places from around the world were included in the *Alphaproteobacteria* clade, but were clearly separated from the 3 previously proposed *Alphaproteobacteria* subclasses (Ferla et al. 2013). The uOTU 0170 and 2 uSEQs from marine epiphytic bacteria obtained in Italy (Mancuso et al. 2016) were clustered forming a divergent *Gracilibacteria* clade.

Discussion

As observed in similar studies performed elsewhere, our study region presented a high abundance of SAR11, *Rhodobacterales*, *Acidimicrobiales*, *Oceanospirillales*, and picoalgae chloroplast sequences (Johnson and Sieburth 1982; Suzuki et al. 2001; Biegala et al. 2003; Lovejoy et al. 2007; Alonso-Gutiérrez et al. 2009; Campbell et al. 2009; Feng et al. 2009; Fortunato et al. 2013; Wu et al. 2014; Liu et al. 2015). Likewise, other taxa, including *Bacillales*, *Xantomonadales*, SAR86, OCS116, *Rhizobiales*, *Rickettsiales*, MB11C04, *Bdellobibrionales*, *Alteromonadales*, *Aegean-245*, *Methylophilales*, KI89A, SAR406, *Sphingomonadales*, and *Gracilibacteria* (also known as BD1-5 and GN02), were detected at low frequencies. About 13% of the OTUs identified here could not be classified. The fact that there was no significant difference between the proportions of Sanger and HTS uSEQs and that there was only a slightly higher proportion of HTS uOTUs, highlight the value of the V1–V3 regions of the 16S gene in taxonomic studies (Liu et al. 2007, 2008). Difficulties in eSEQ classification can be generally attributed to the sketchy knowledge of marine microbial systematics (Schloss et al. 2016). However, methodological issues can also represent a challenge, such as undetected chimaeras, artificial classifications (i.e., para- or polyphyletic taxa) or phenomena not accounted for by classification algorithms, such as gene transformation and deep coalescence. Notwithstanding that, the identification of previously described uSEQs similar to several uOTUs (supplemental Table S7[1]; supplemental Fig. S4[1]) provided strong evidence that these OTUs actually correspond to real 16S lineages. In addition, this observation supports the concept that locally rare taxa could be relevant at a worldwide scale (Sunagawa et al. 2010; Shade et al. 2014; Choi et al. 2017).

Phylogenetic classification has been used in a limited way in microbial ecology compared with classification algorithms based on sequence similarity and composition (Hiraoka et al. 2016). Here, we show that SPA and (de novo) phylogenetic comparisons with reference sequences can greatly improve the taxonomic assignment of uSEQs and may drive the discovery of new taxa. These findings suggest an untapped potential of phylogenetic evidence in eDNA systematic studies. Some challenges faced by phylogenetic classification could be also foreground. A conspicuous one is that some taxa may not be retrieved by phylogenetic analysis, as observed here for the *Rickettsiales*, *Rhodobacterales*, and *Rhizobiales* (Fig. 1; supplemental Fig. S4[1]). This can be due to some taxa actually not being monophyletic or to methodological or molecular marker issues. The implementation of scoring statistics is also challenging. The use of branch supports is an apparently straightforward solution. However, depending on the used marker and method, some (actual) groups may not be retrieved as supported clades. Furthermore, it must be considered that the probability of recovering a given clade by chance (even with no or a low support) is extremely low, and depends on the number of sequences from the group of interest and the total number of sequences in the data. This problem becomes more complicated if topological uncertainty (i.e., multiple optimal phylogenies) is taken into account. A good amount of research and analysis of empirical data will be required to approach the subject in a reasonably finished form.

Finally, it is also worth highlighting the observed correspondence between our phylogenetic and BLAST analyses (supplemental Table S2[1]). Interestingly, the correspondence was greater for the BLAST comparisons with sequences from type specimens, which we attribute to a more accurate taxonomic determination of the corresponding exemplars. This highlights the value of sequences from type material in eSEQs classification.

Acknowledgements

This work was partially supported by grant PICT-2795 from the National Agency of Scientific and Technological Promotion (ANPCyT) and the Asociación Civil Argentina

Genetics (ArGen). We want to thank the useful comments provided by 2 anonymous reviewers.

References

Alonso-Gutiérrez, J., Lekunberri, I., Teira, E., Gasol, J.M., Figueras, A., and Novoa, B. 2009. Bacterioplankton composition of the coastal upwelling system of 'Ría de Vigo', NW Spain. FEMS Microbiol. Ecol. **70**(3): 493–505. doi:10.1111/j.1574-6941.2009.00766.x. PMID:19780832.

Biegala, I.C., Not, F., Vaulot, D., and Simon, N. 2003. Quantitative assessment of picoeukaryotes in the natural environment by using taxon-specific oligonucleotide probes in association with tyramide signal amplification-fluorescence in situ hybridization and flow cytometry. Appl. Environ. Microbiol. **69**(9): 5519–5529. doi:10.1128/AEM.69.9.5519-5529.2003. PMID:12957941.

Campbell, B.J., Yu, L., Straza, T.R.A., and Kirchman, D.L. 2009. Temporal changes in bacterial rRNA and rRNA genes in Delaware (USA) coastal waters. Aquat. Microb. Ecol. **57**: 123–135. doi:10.3354/ame01335.

Choi, C.J., Bachy, C., Jaeger, G.S., Poirier, C., Sudek, L., Sarma, V.V.S.S., et al. 2017. Newly discovered deep-branching marine plastid lineages are numerically rare but globally distributed. Curr. Biol. **27**(1): R15–R16. doi:10.1016/j.cub.2016.11.032. PMID:28073013.

Federhen, S. 2015. Type material in the NCBI Taxonomy Database. Nucleic Acids Res. **43**: D1086–D1098. doi:10.1093/nar/gku1127. PMID:25398905.

Feng, B.W., Li, X.R., Wang, J.H., Hu, Z.Y., Meng, H., Xiang, L.Y., and Quan, Z.X. 2009. Bacterial diversity of water and sediment in the Changjiang estuary and coastal area of the East China Sea. FEMS Microbiol. Ecol. **70**(2): 236–248. doi:10.1111/j.1574-6941.2009.00772.x.

Ferla, M.P., Thrash, J.C., Giovannoni, S.J., and Patrick, W.M. 2013. New rRNA gene-based phylogenies of the Alphaproteobacteria provide perspective on major groups, mitochondrial ancestry and phylogenetic instability. PLoS One, **8**(12): e83383. doi:10.1371/journal.pone.0083383. PMID:24349502.

Fortunato, C.S., Eiler, A., Herfort, L., Needoba, J.A., Peterson, T.D., and Crump, B.C. 2013. Determining indicator taxa across spatial and seasonal gradients in the Columbia River coastal margin. ISME J. **7**(10): 1899–1911. doi:10.1038/ismej.2013.79. PMID:23719153.

Hiraoka, S., Yang, C.C., and Iwasaki, W. 2016. Metagenomics and bioinformatics in microbial ecology: current status and beyond. Microbes Environ. **31**(3): 204–212. doi:10.1264/jsme2.ME16024. PMID:27383682.

Johnson, P.W., and Sieburth, J.M. 1982. In-situ morphology and occurrence of eucaryotic phototrophs of bacterial size in the picoplankton of estuarine and oceanic waters. J. Phycol. **18**(3): 318–327. doi:10.1111/j.1529-8817.1982.tb03190.x.

Liu, J., Fu, B., Yang, H., Zhao, M., He, B., and Zhang, X.H. 2015. Phylogenetic shifts of bacterioplankton community composition along the Pearl Estuary: the potential impact of hypoxia and nutrients. Front. Microbiol. **6**: 64. doi:10.3389/fmicb.2015.00064. PMID:25713564.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., and Knight, R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res. **35**(18): e120. doi:10.1093/nar/gkm541. PMID:17881377.

Liu, Z., DeSantis, T.Z., Andersen, G.L., and Knight, R. 2008. Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Res. **36**(18): e120. doi:10.1093/nar/gkn491. PMID:18723574.

Lovejoy, C., Vincent, W.F., Bonilla, S., Roy, S., Martineau, M.-J., Terrado, R., et al. 2007. Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Arctic Seas. J. Phycol. **43**(1): 78–89. doi:10.1111/j.1529-8817.2006.00310.x.

Mancuso, F.P., D'Hondt, S., Willems, A., Airoldi, L., and De Clerck, O. 2016. Diversity and temporal dynamics of the epiphytic bacterial communities associated with the canopy-forming seaweed *Cystoseira compressa* (Esper) Gerloff and Nizamuddin. Front. Microbiol. **7**: 476. doi:10.3389/fmicb.2016.00476. PMID:27092130.

Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. PLoS One, **6**(12): e27310. doi:10.1371/journal.pone.0027310. PMID:22194782.

Schloss, P.D., Girard, R.A., Martin, T., Edwards, J., and Thrash, J.C. 2016. Status of the archaeal and bacterial census: an update. mBio, **7**(3): e00201–16. doi:10.1128/mBio.00201-16. PMID:27190214.

Shade, A., Jones, S.E., Caporaso, J.G., Handelsman, J., Knight, R., Fierer, N., and Gilbert, J.A. 2014. Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. mBio, **5**(4): e01371–14. doi:10.1128/mBio.01371-14. PMID:25028427.

Sunagawa, S., Woodley, C.M., and Medina, M. 2010. Threatened corals provide underexplored microbial habitats. PLoS One, **5**(3): e9554. doi:10.1371/journal.pone.0009554. PMID:20221265.

Suzuki, M.T., Béjà, O., Taylor, L.T., and DeLong, E.F. 2001. Phylogenetic analysis of ribosomal RNA operons from uncultivated coastal marine bacterioplankton. Environ. Microbiol. **3**(5): 323–331. doi:10.1046/j.1462-2920.2001.00198.x. PMID:11422319.

Wu, W., Huang, B., and Zhong, C. 2014. Photosynthetic picoeukaryote assemblages in the South China Sea from the Pearl River estuary to the SEATS station. Aquat. Microb. Ecol. **71**: 271–284. doi:10.3354/ame01681.