

The following updates were tested on a dataset generated by Illumina MiSeq from an epishelf lake in the high Arctic (Thaler, unpublished data). The revisions used as their starting point the June 2015 version of the Eukarya database and the January 2015 version of the Bacteria database.

Bacteria: updates to Comamonadaceae (Betaproteobacteria) and Rhodobacteraceae (Alphaproteobacteria)

Comamonadaceae

Revisions focused on the related genera *Rhodofera*, *Albidifera*, *Polaromonas*, *Variovorax*, *Limnohabitans* and *Curvibacter*. To guide revisions of the database, a 16S rRNA gene alignment was constructed using MUSCLE (Edgar 2004). All sequences were checked manually for chimeras by examining alignments, and a number of suspect sequences were removed from all genera. The final alignment contained 135 cultured and environmental sequences, plus 18 other Comamonadaceae sequences. Outgroups were betaproteobacteria *Parapusillimonas granuli*, *Gallionella ferruginea* and *Thiobacillus thioparus*. The alignment had 1582 characters. A maximum-likelihood tree was constructed using RAxML v7.2.7 (Stamatakis 2006), with 100 bootstraps (tree is available on request from Mary Thaler).

Published sequences from cultured described species, following Willems (2014) improved the taxonomic resolution of the database to species-level (Table 1)

Table 1. Described Comamonadaceae species added to the reference database

<i>Caenimonas terrae</i>	<i>Polaromonas aquatica</i>	<i>Rhodofera saidenbachensis</i>
<i>Curvibacter delicatus</i>	<i>Polaromonas cryoconiti</i>	<i>Variovorax boronicumulans</i>
<i>Curvibacter fontanus</i>	<i>Polaromonas glacialis</i>	<i>Variovorax defluvii</i>
<i>Curvibacter gracilis</i>	<i>Polaromonas jejuensis</i>	<i>Variovorax dokdonensis</i>
<i>Limnohabitans australis</i>	<i>Polaromonas rhizosphaerae</i>	<i>Variovorax ginsengisoli</i>
<i>Limnohabitans curvus</i>	<i>Polaromonas vacuolata</i>	<i>Variovorax soli</i>
<i>Limnohabitans parvus</i>	<i>Rhodofera antarcticus</i>	
<i>Limnohabitans planktonicus</i>	<i>Rhodofera fermentans</i>	

The database was also expanded to include the genus *Pseudorhodofera*, a genus of soil bacteria (Bruland et al. 2009). Additional *Limnohabitans* sequences were added to ensure good coverage of the sub-genus level clades found by Kasalický et al. (2013).

The species *Rhodofera ferrireducens* had been moved into the newly created genus *Albidifera* by Ramana and Sasikala (2009) based on the lack of photosynthetic pigments and its basal position in a Maximum Likelihood tree relative to the other two species *R. fermentans* and *R. antarcticus*. The situation was complicated by the description of a new *Rhodofera* species with a very divergent 16S rRNA gene, *R. saidenbachensis* (Kaden et al. 2014), which has rendered *Rhodofera* and *Albidifera* polyphyletic. Note that this situation has been worsened by researchers who have labeled their environmental sequences in GenBank with genus names apparently determined using a BLAST search of the query sequence, resulting in sequences spuriously labeled *Albidifera* which are actually more

closely related to the original two *Rhodiferax* species. This practise should be discouraged. Therefore, pending further support for *Albidiferax* as a valid genus, this version of the database continues to use the original name *Rhodiferax ferrireducens*.

Two genera, *Curvibacter* and *Variovorax*, were not recovered as monophyletic in the RAxML tree. Therefore, *Curvibacter lanceolatus* and *C. gracilis* were given the genus name “Curvibacter1” in the database, while *C. delicatulus* and *C. fontanus* were given the genus name “Curvibacter2”. *Variovorax* was similarly split into “Variovorax1” (*V. paradoxus*, *V. boronicumulans* and *V. ginsengisoli*), “Variovorax2” (*V. dokdonensis* and environmental sequence GQ009710) and “Variovorax3” (*V. soli*). This splitting was tested and found to allow a higher number of sequences to be identified to genus-level. Better taxonomic coverage may be necessary to completely resolve these two genera.

Overall, these improvements allowed an improvement of 68% more Comamonadaceae reads identified from the epishelf lake dataset

Rhodobacteraceae

An additional taxonomic level was added between family and genus to be able to distinguish “*Roseobacter*-group”, “*Amaricoccus*-group”, “*Rhodobacter*-group”, “*Rhodovulum*-group”, and “*Paracoccus*-group,” following Pujalte et al. (2014). Additionally, representative sequences were added for the following genera: *Yangia*, *Tropicibacter*, *Salinihabitans*, *Primorskybacter*, *Pelagimonas*, *Lentibacter*, and *Epibacterium*.

Eukarya: updates to Chlorophytes, Ciliates, Dinoflagellates, Cryptophytes, Fungi and Metazoans

Chlorophytes

The previous version of the database did not attempt to resolve genera within the Chlorophyceae, a family which contains many examples of polyphyletic genera, most notoriously *Chlamydomonas*. For the current version we added a number of new sequences to the database to cover the 19 lineages proposed in the system of Nakada et al. (2008), as well as some divergent Chlorophyceae sequences that have resisted classification. Additionally, the distinction between two major clades based on the type of flagellar insertion (“CW”, clockwise versus “DO”, directly opposite), is accepted as phylogenetically valid. It is used as a taxonomic rank in between family and lineage.

Ciliates

Liu et al. (2014) have argued that the genus *Askenasia*, which has previously been placed in either Litostomatea or Prostomatea, should in fact be placed in its own class based on evidence of both SSU and LSU ribosomal sequences. Therefore, in this version of the database, *Askenasia*, along with sister genera *Paraspathidium* and *Cyclotrichium* are grouped in the same class-level taxon, provisionally called “AskenasiaClade”.

The previous version of this database did not include ranks between class and genus level for Spirotrichea, and also treated Spirotrichea and Oligotrichea as taxa of equivalent rank, which is incorrect (Foissner et al. 2004). This resulted in a high number of “Unclassified Spirotrichea” in analyses datasets. To correct this problem, taxonomic structure was added at lower ranks, following the phylogenies of Foissner et al. (2004) and Agatha and Strüder-Kypke (2014) (Table 2). In this revised classification, Spirotrichea has the rank of class, and contains two sub-classes, Oligotrichea and Hypotrichea. Although it has been proposed to split Hypotrichea into two classes, Hypotrichea *sensu*

stricto and Stichotrichea, this classification has not been consistently adopted by workers. Therefore, to avoid confusion, in this database version we use only Hypotrichea. Some spirotrich genera belong to intensively studied groups, e.g. Euplotida or the Tintinnids, while others still require more phylogenetic analysis to place them even at sub-class level. Users are cautioned that because of this uneven level of research effort, ciliate entries in the database may have varying number of taxonomic ranks.

Table 2. Classification of Spirotrichea (Ciliophora) used in current database.

Spirotrichea

Caryotricha

Kiitricha

Lyncnophora

Oligotrichea

Halteriidae

Halteria

Meseres

Choreotrichida

Parastrombidinopsis

Pelagostrobilidium

Pseudoamphisiella

Rimostrombidium

Strobilidium

Strombidinopsis

Tintinnids

Codonellopsis

Eutintinnus

Favella

Metacylis

Tintinnidium

Tintinnopsis

Oligotrichida

Tontoniidae

Pseudotontonia

Strombidiidae

Novistrombidium

Strombidium

Varistrombidium

Hypotrichea

Amphisiella

Certesia

Engelmanniella

Gonostomum

Orthamphisiella

Parabirojimia

Paraurostyla

Phacodinium

Prodiscocephalus

Protocruzia

Euplotida

Aspidisca
Diophrys
Euplotes
Euplotidium
Gastrocirrhus
Moneuplotes

Sporadotrichida

Cyrtohymena
Gastrostyla
Hemiurosoma
Laurentiella
Onychodromopsis
Onychodromus
Oxytricha
Pattersoniella
Pleurotricha
Steinia
Sterkiella
Stylonychia
Tetmemena
Trachelostyla

Urostylida

Diaxonella
Holosticha
Paruroleptus
Pseudokeronopsis
Pseudourostyla
Rigidothrix
Uroleptus
Urostyla

Dinoflagellates

Former versions of the database treated “GPP” as an equivalent taxonomic rank with Gymnodiniales, Peridinales and Prorocentrales, whereas it should be a higher rank which contains them. Also, sequence EF058238.1 appears to have been labeled in error:

Eukaryota;Alveolata;Dinoflagellates;Dinophyceae_Gloedinium;Thoracosphaeraceae_Scripsiella;Gloeodinium_montanum_strain_CCAC0066;

Note that elsewhere, “Thoracosphaeraceae_Scripsiella” is treated as a taxonomic rank within Thoracosphaerales. This entry has been changed to:

Eukaryota;Alveolata;Dinoflagellates;Dinophyceae_GPP;Dinophyceae_Peridinales;Glenodiniaceae;Gloeodinium;Gloeodinium_montanum_strain_CCAC0066;

The genus *Katodinium* has undergone important taxonomic revisions in the past decades. The sequence AF274267.1, identified as *Katodinium rotundatum* in the previous database version, has been transferred to *Heterocapsa* (Hansen 1995), and this new taxonomic position has been confirmed by LSU rRNA phylogeny. The name of the genus is changed in the taxonomy file and moved into Peridinales. In addition, a sequence has been added for *Heterocapsa arctica*. A new sequence of

Katodinium glaucum has been added, following the phylogenetic analysis of Reñé et al. (2015).

Cryptophytes

Laza-Martínez et al. (2012) have shown that the commonly detected cryptophyte family Geminigeraceae stands in need of taxonomic revision. Five clades are detectable: 1) *Teleaulax amphioxeia* and *P. prolonga*, 2) the species *T. minuta*, 3) the species *T. acuta*, 4) the genus *Geminigera* and 5) the species *T. gracilis*. One more species, *P. nannoplanctonica* is basal to the other five clades. To reflect the confused state of classification in Geminigeraceae, this version of the database does not use *Teleaulax* as a genus-level rank. Instead, *T. amphioxeia* and *P. prolonga* are grouped together in the genus-level group called “PlagioselmisTeleaulaxComplex”. For other *Teleaulax* species, the classification jumps directly from family to species level, e.g. “Geminigeraceae;Teleaulax_minuta;”. *Plagioselmis* is used as a genus-level rank for *P. nannoplanctonica* only.

The previous database version used “unclassified_Cryptophyta” as a taxonomic name at various rank levels. This practise should be strongly discouraged, as the classification software will assume that different unclassified cryptophytes are related to each other, leading it to fail to classify environmental sequences from known taxa. In the current version, these “unclassified sequences” were handled in various ways. Some were identified to genus level based on a BLAST search (FJ032651.1, GQ375265 and EF195738.1). Sequence GQ375264.1 is culture CCMP 2045, which has since been identified as *Rhodomonas* (<https://ncma.bigelow.org>).

Fungi/Metazoans

This clade has been reorganized following the advice of André Comeau. Additionally, a number of parasitic taxa were added, including *Cryptosporidium* (Apicomplexa, Alveolata), *Entamoeba* (Amoebozoa), *Carpediemonas* (Fornicata), *Trichomonas* (Parabasalia), *Blastocystis* (Stramenopiles), and a number of diplomonads and retortamonads

References

- Agatha S, Strüder-Kypke MC. 2014. What morphology and molecules tell us about the evolution of Oligotrichea (Alveolata, Ciliophora). *Acta Protozool* 53: 77–90. doi:10.4467/16890027AP.14.008.1445
- Bruland N, Bathe S, Willems A, Steinbüchel A. 2009. *Pseudorhodofera soli* gen. nov., sp. nov. and *Pseudorhodofera caeni* sp. nov., two members of the class Betaproteobacteria belonging to the family Comamonadaceae. *Int J Syst Evol Microbiol* 59: 2702–2707. doi: 10.1099/ijs.0.006791-0
- Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797. DOI: 10.1093/nar/gkh340
- Foissner W, Moon-van der Staay SY, van der Staay GWM, Hackstein JHP, Krautgartner W-D, Berger H. 2004. Reconciling classical and molecular phylogenies in the stichotrichines (Ciliophora, Spirotrichea), including new sequences from some rare species. *Eur J Protistol* 40: 265–281. doi: 10.1016/j.ejop.2004.05.004
- Hansen G. 1995. Analysis of the thecal plate pattern in the dinoflagellate *Heterocapsa rotundata* (Lohmann) comb. nov. (= *Katodinium rotundatum* (Lohmann) Loeblich). *Phycologia* 34: 166–170.
- Kaden, R, Spröer C, Beyer D, Krolla-Sidensteirn P. 2014. *Rhodofera saidenbachensis* sp. nov., a psychrotolerant, very slowly growing bacterium within the family *Comamonadaceae*, proposal of appropriate taxonomic position of *Albidifera ferrireducens* strain T118^T in the genus

- Rhodiferax* and emended description of the genus *Rhodiferax*. *Int J Syst Evol Microbiol* 64: 1186–1193. doi: 10.1099/ij.s.0.054031-0
- Kasalický V, Jezbera J, Hahn M, Šimek K. 2013. The diversity of the *Limnohabitans* genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. *PLoS ONE* 8: e58209. doi: 10.1371/journal.pone.0058209
- Laza-Martínez A, Arluzea J, Miguel I, Orive E. 2012. Morphological and molecular characterization of *Teleaulax gracilis* sp. nov. and *T. minuta* sp. nov. (Cryptophyceae). *Phycologia* 51: 649–661. doi: 10.2216/11-044.1
- Liu A, Yi Z, Lin X, Hu Z, Al-Farraj SA, Al-Rashed KAS. 2014. Molecular phylogenetic lineage of *Plagiopogon* and *Askenasia* (Protozoa, Ciliphora) revealed by their gene sequences. *J Ocean Univ China*. 14: 724–730. doi: 10.1007/s11802-015-2559-3
- Nakada T, Misawa K, Nozaki H. 2008. Molecular systematics of Volvocales (Chlorophyceae, Chlorophyta) based on exhaustive 18S rRNA phylogenetic analyses. *Mol Phylogenetics Evol* 48: 281–291. doi: 10.1016/j.ympev.2008.03.016
- Pujalte MF, Lucena T, Ruvira MA, Arahál DR Macián MC. 2014. The family *Rhodobacteraceae*. Rosenberg E et al. (eds.) *The Prokaryotes – Alphaproteobacteria and Betaproteobacteria*. Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-30197-1_377
- Ramana CV, Sasikala. 2009. *Albidoferax*, a new genus of *Comamonadaceae* and reclassification of *Rhodiferax ferrireducens* (Finneran et al. 2003) as *Albidiferax ferrireducens* comb. nov. *J Gen Appl Microbiol* 55: 301–304.
- Reñé A, Camp J, Garcés. 2015. Diversity and phylogeny of Gymnodiniales (Dinophyceae) from the NW Mediterranean Sea revealed by a morphological and molecular approach. *Protist* 166: 234–263. doi: 10.1016/j.protis.2015.03.001
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688–2690. doi: 10.1093/bioinformatics/btl446