Molecular Phylogenetics and Evolution 139 (2019) 106546

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Current phylogeny of Rhodospirillaceae: A multi-approach study

Mauro Degli Esposti*, Luis Lozano, Esperanza Martínez-Romero

*Center for Genomic Sciences, UNAM Campus de Cuernavaca, Cuernavaca 62130, Morelos, Mexico*

ARTICLE INFO

*Keywords:*
Alphaproteobacteria
Phylogenetic analysis
Rhodospirillaceae
Functional traits

ABSTRACT

Rhodospirillaceae represents a major family of the class alphaproteobacteria that includes an increasing number of functionally diverse taxa. The aim of this work is to evaluate the present phylogenetic diversity of the Rhodospirillaceae, which includes several metagenome-assembled genomes of uncultivated bacteria, as well as cultivated bacteria that were previously classified in different families. Various methodological approaches have been followed to discern the phylogenetic diversity of the taxa associated with the Rhodospirillaceae, which are grouped in three major sub-divisions and several other taxonomic entities that are currently confined to the genus rank. These genera include *Tistrella*, *Elstera*, *Dongia* and *Ferrovibrio* among cultivated organisms and alphaproteobacteria bacterium 41–28 among uncultivated bacteria. Overall, this study adds at least 11 genera and over 40 species to the current set of taxa belonging to the Rhodospirillaceae, a taxonomic term that clearly requires amendment. We propose to re-classify all taxa associated with the Rhodospirillaceae family under the new order, Diaforabacterales ord. nov. (from the Greek word for diversity, διάφορα). This study also uncovers the likely root of Rhodospirillaceae among recently reported metagenome-assembled genomes of uncultivated marine and groundwater bacteria.

## 1. Introduction

Rhodospirillaceae is a large family of the class of alphaproteobacteria (Garrity et al., 2005). Originally, this family contained predominantly phototrophic bacteria such as the type genus *Rhodospirillum rubrum* (Garrity et al., 2005; Imhoff et al., 1998). Since 2005, the genera affiliated to Rhodospirillaceae have grown from 13 (Garrity et al., 2005) to 51 according to the latest data in NCBI resources (https://www.ncbi.nlm.nih.gov/Tamy/Bxonorowser/wwwtax.cgi, accessed 10 April 2019). However, only one-fourth of these genera have the photosynthetic reaction center and therefore can be considered phototrophic like the type genus *R. rubrum*. The taxonomic richness of the Rhodospirillaceae has been increased considerably by recent works on the metagenome of human gut (Nielsen et al., 2014; Brown et al., 2016), subterranean aquifers (Anantharaman et al., 2016; Probst et al., 2017) and marine environments (Tully et al., 2017; Parks et al., 2017). The latest taxonomy appraisals of the family Rhodospirillaceae have been reported several years ago (Wisniewski-Dyé et al., 2011; Kwan et al., 2012), before the recent addition of new cultivated and metagenomic-derived taxa (Table 1). Consequently, the current taxonomic diversity of the family is not updated.

The functional diversity of the Rhodospirillaceae has also been enhanced by the genomes of recently described bacteria from metagenomic studies, which show extreme or flexible adaptation to anaerobic environments without phototrophic properties (Probst et al., 2017; Degli Esposti et al., 2016; Probst et al., 2018). Genomic analysis has shown that Rhodospirillaceae taxa, either classified and cultivated or uncultivated, contain probably the largest variety of the traits that modulate adaptation to anoxic conditions among alphaproteobacteria (Degli Esposti and Martinez-Romero, 2017; Degli Esposti et al., 2018; Louca et al., 2016). Moreover, the earliest biosynthetic pathways of the membrane quinone involved in bioenergetics, ubiquinone, have been traced to members of the Rhodospirillaceae family (Degli Esposti, 2017), some of which appear to have ancestral forms of terminal oxidases (Degli Esposti et al., 2018; Degli Esposti, 2017; Degli Esposti et al., 2019). These recent findings suggest that Rhodospirillaceae include deep branching lineages of alphaproteobacteria that are related to ancestral prokaryotes which evolved in Proterozoic oceans where oxygen concentrations were low (Degli Esposti, 2017; Degli Esposti et al., 2019; Müller et al., 2012). Proto-mitochondria likely evolved in the same oceans and during the same period of earth geochemical history. Therefore, they might have been related to ancestral lineages of

*Abbreviations:* ANI, Average Nucleotide Identity; GTDB, genome taxonomy database; LGT, lateral gene transfer; MAG, metagenome assembled genome; L, maximum likelihood; NJ, neighbor joining; NR, non redundant NCBI database for proteins; SD, Standard deviation

* Corresponding author.
*E-mail address:* mauro1italia@gmail.com (M. Degli Esposti).

https://doi.org/10.1016/j.ympev.2019.106546
Received 17 October 2018; Received in revised form 27 June 2019; Accepted 27 June 2019
Available online 04 July 2019
1055-7903/ © 2019 Elsevier Inc. All rights reserved.

**Table 1**
List of current cultivated genera of the Rhodospirillaceae family and related taxa. The list includes, in alphabetical order, all the genera currently classified within the Rhodospirillaceae family with nearly complete genome. Genera highlighted in gray have not been previously classified with the family.

| Taxa with sequenced genome | Genomes | Reported taxonomic grouping | References | Phenotype |
|---|---|---|---|---|
| *Aestuariispira insulae* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] and close to Terasakiella (Park et al., 2014) | |
| ***Azospirillum*** | 8 | Rhodospirillaceae; Azospirillum group; Rhodospirillaceae 2 clade | (Garrity et al., 2005; Martijn et al., 2018) | two strains photosynthetic |
| *Caenispirillum* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| ***Candidatus Endolissoclinum*** | 5 | Rhodospirillaceae; pan-thalassic; SAR116 cluster clade | NCBI taxonomy website[#] and (Martijn et al., 2018; Degli Esposti et al., 2014) | endosymbiont of tunicates |
| *Dongia* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Elstera* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Ferrovibrio* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Fodinicurvata* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] and (Martijn et al., 2018; Degli Esposti et al., 2014) | |
| *Haematospirillum* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Indioceanicola* | 1 | Rhodospirillaceae | NCBI taxonomy website[#], close to Rhodocista (Chen et al., 2018) | |
| *Inquilinus* | 2 | Rhodospirillaceae | (Garrity et al., 2005; Martijn et al., 2018) | pathogen |
| *Insolitispirillum* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Limimonas* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Magnetospira* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | magnetotactic |
| ***Magnetospirillum*** | 7 | Rhodospirillaceae; Rhodospirillaceae 1 clade | (Garrity et al., 2005; Martijn et al., 2018) | magnetotactic |
| *Magnetovibrio* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | magnetotactic |
| *Nisaea* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Nitrospirillum* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Niveispirillum* | 3 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Novispirillum* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| *Oceanibaculum* | 3 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Pararhodospirillum* | 1 | Rhodospirillaceae | (Garrity et al., 2005) | photosynthetic |
| *Phaeospirillum* | 2 | Rhodospirillaceae | (Garrity et al., 2005) | photosynthetic |
| *Rhodospira* | 1 | Rhodospirillaceae | (Garrity et al., 2005) | photosynthetic |
| *Rhodospirillum centenum or Rhodocista centenaria* | 1 | separate genus from *R.rubrum*; Rhodospirillaceae 2 clade | (Garrity et al., 2005; Martijn et al., 2018; Tindall, 2001) | photosynthetic |
| ***Rhodospirillum rubrum*** | 3 | type genus of Rhodospirillaceae; Rhodospirillaceae 1 clade | (Garrity et al., 2005; Martijn et al., 2018) | photosynthetic |
| *Rhodovibrio* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Skermanella* | 2 | Rhodospirillaceae | (Garrity et al., 2005) | photosynthetic |
| *Telmatospirillum* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Thalassobaculum* | 2 | Rhodospirillaceae | NCBI taxonomy website[#] | photosynthetic |
| *Thalassospira* | 14 | Rhodospirillaceae; Rhodospirillaceae 1 clade | (Garrity et al., 2005; Martijn et al., 2018) | |
| *Tistlia* | 1 | Rhodospirillaceae | NCBI taxonomy website[#] | |
| ***Tistrella*** | 2 | Rhodospirillaceae; unclassified Alphaproteobacteria | NCBI taxonomy website[#] | |
| *Varunaivibrio* | 1 | Rhodospirillaceae | NCBI taxonomy website[#], close to Magnetovibrio (Patwardhan and Vetriani, 2016) (Wisniewski-Dyé et al., 2011; Martijn et al., 2018) | magnetotactic |
| ***Reyranella*** | 1 | **unclassified Rhodospirillales** | (Kim et al., 2013) | |
| ***Rhodospirillaceae bacterium NORP116*** | 1 | unclassified Rhodospirillaceae | (Tully et al., 2018) | MAG |
| ***Rhodospirillaceae bacterium NORP128*** | 1 | unclassified Rhodospirillaceae | (Tully et al., 2018) | MAG |
| ***Micavibrio*** | 3 | **unclassified Alphaproteobacteria**; Rhodospirillaceae 2 clade | NCBI taxonomy website[#] and (Martijn et al., 2018) | ectoparasite |
| ***Terasakiella*** | 2 | **Methylocystaceae, Rhizobiales** | NCBI taxonomy website[#] and (Degli Esposti et al., 2018) | magnetotactic |
| ***SAR116 cluster*** | 19 | **unclassified Alphaproteobacteria**; SAR116 cluster clade | (Martijn et al., 2018; Oh et al., 2010) | magnetotactic |
| ***alpha proteobacterium BAL199*** | 1 | **unclassified Alphaproteobacteria;** pan-thalassic; SAR116 cluster clade | NCBI taxonomy website[#] and (Martijn et al., 2018; Degli Esposti et al., 2014) | |
| ***Kiloniella*** | 5 | **Kilonelliaceae, Kilonelliales;** unclassified Rhodospirillaceae | NCBI taxonomy website[#] and (Martijn et al., 2018; Richter and Rossello-Mora, 2009; Imhoff and Wiese, 2014) | |

[#] https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=41295 accessed 9 April 2019.

Rhodospirillaceae, as functional and genomic characteristics suggest (Degli Esposti et al., 2018; Degli Esposti et al., 2019; Trubitsyn et al., 2016). Indeed, *R. rubrum* has been found to contain many close relatives of mtDNA- and nuclear-encoded proteins in separate phylogenetic studies (Esser et al., 2004; Atteia et al., 2009; Abhishek et al., 2011; Thiergart et al., 2012).

This study presents an in depth analysis of the current phylogenetic diversity of the Rhodospirillaceae, which has a taxonomic richness much larger than previously considered, encompassing diverse taxa that have been so far classified in other families. We also uncover the roots of Rhodospirillaceae among metagenome-assembled genomes (MAGs) that show similarities with the most ancient lineages of alphaproteobacteria. Our results show that the ancestral lineages of Rhodospirillaceae are close to the earliest alphaproteobacteria. Therefore, this study bears relevance also to the origin of the mitochondrial organelle, an origin which is intimately linked to the phylogenesis of alphaproteobacteria (Degli Esposti et al., 2018; Degli Esposti et al., 2019; Thiergart et al., 2012; Martijn et al., 2018).

## 2. Methodological approaches

This work has developed and used diverse methodological approaches that need to be explained in detail, because their application to study the phylogeny of alphaproteobacteria is novel, or does not follow routine methods of phylogenetic analysis. Therefore, the section of Material and Methods has been expanded to provide a detailed description of the major methodological approaches introduced here.

### 2.1. Phylogenetic analysis using concatenated ribosomal proteins

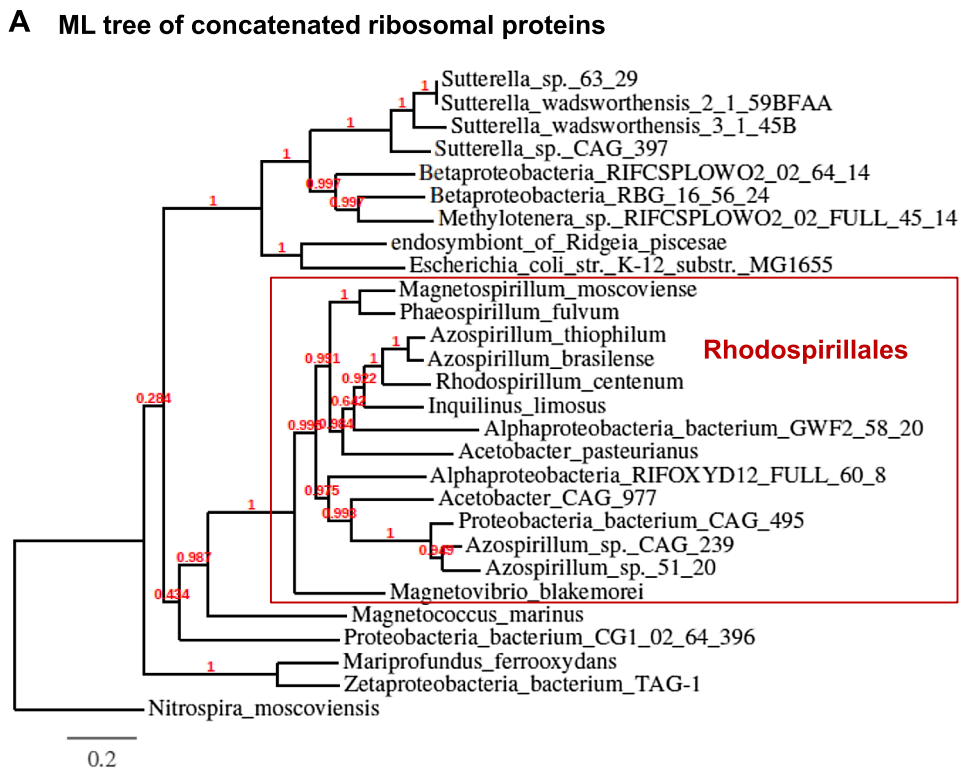
Although 16S rRNA is generally considered the gold standard of current bacterial taxonomy (Woese, 1987; Segata et al., 2013; Schulz et al., 2017), we have used alternative approaches to assess the phylogenetic relationship of alphaproteobacterial taxa with the family of Rhodospirillaceae. The first approach has followed one of the recently introduced methods that are based upon the concatenation of predominantly ribosomal proteins (Probst et al., 2017), as initiated by Ciccarelli et al. (2006). The latter method was used for the original taxonomic assignment of MAGs from human gut microbiomes (Nielsen et al., 2014) that were later discovered to harbour functional traits typical of anaerobic eukaryotes (Degli Esposti et al., 2016). The method of Ciccarelli et al. (2006) has been subsequently expanded to larger sets of protein markers (Anantharaman et al., 2016; Probst et al., 2017; Parks et al., 2017; Louca et al., 2016), which generally achieve higher taxonomic accuracy (Segata et al., 2013). We have chosen to use one of such methods, that of Probst et al. (2017), which is based upon the concatenated alignment of the largest possible number of ribosomal proteins shared among the taxa analysed. The results obtained with this method produced phylogenetic trees showing strong support for the affiliation of *Acetobacter* sp. CAG:977 (Nielsen et al., 2014) with the Rhodospirillaceae family (Fig. 1). However, the methodology adopted (Probst et al., 2017), similarly to other methods that are based upon the concatenation of ribosomal (and other) proteins, could not be extended to all currently available taxa that are listed as 'unclassified alphaproteobacteria', because they cumulatively share very few of these proteins due their scattered distribution in reported genomes, which are often incomplete. Consequently, we undertook subsequent studies using phylogenetic trees of single marker proteins, while using methods predominantly based upon the presence of ribosomal proteins to evaluate the genome completeness of taxa (Probst et al., 2018; Degli Esposti, 2017).

### 2.2. Phylogenetic analysis using conserved marker proteins

Any phylogenetic study applied to many diverse taxa needs to strike a balance between taxonomic sampling with an insightful choice of proteins markers (Degli Esposti et al., 2018; Imhoff, 2016; Martin, 2018; Degli Esposti, 2016). We minimized the all too common problem of arbitrary choice in taxonomic sampling by running DeltaBLAST (Boratyn et al., 2012) searches (using the standard BLOSUM62 substitution matrix) which were routinely extended to *all* classified and uncultured taxa of the Rhodospirillaceae, Rhodospirillales and unclassified alphaproteobacteria present in the NR database (versions from January to August 2018). Conserved proteins reported to have a strong phylogenetic signal such as cytochrome *b* and the NuoL subunit of complex I were used as queries of DeltaBLAST searches, which sometimes were extended to 10,000 taxa in order to cover the whole class of alphaproteobacteria. Such extended searches produced NJ trees with robust topologies indicating the deep-branching position of the Rhodospirillaceae family with respect to other members of the parent Rhodospirillales order and the rest of alphaproteobacteria (Degli Esposti et al., 2018; Degli Esposti, 2016) (Fig. 2). Subsequently, about 100 sequences were downloaded and aligned into the program MEGA5.2 (Tamura et al., 2011) using ClustalW followed by iterative manual refinements, which were guided by the 3D structure of the protein as described earlier (Degli Esposti, 2016; Degli Esposti et al., 2014). The sequences were chosen so that the preliminary NJ trees of their alignment could match as close as possible the topology of the NJ tree obtained with the DeltaBLAST routine of all results (Degli Esposti, 2016). Sequences that showed unusual insertions, whether due to possible sequencng errors or not, were removed, and the N- and C-terminus were trimmed to a minimal extent in order to limit the loss in sequence information. The refined alignments were then used to build phylogenetic trees with either the NJ or the ML approach, with 200–500 bootstraps using the MEGA5.2 program as described previously (Degli Esposti et al., 2018; Degli Esposti, 2016) (see the legend of Fig. 3 for other details).

### 2.3. Analysis of topology and clade placement in trees obtained with diverse marker proteins

The major problem of phylogenetic trees obtained with single protein markers is that any program to construct such trees does not yield strong support (high % values of bootstrap replicates) for various internal nodes separating deep branches of Rhodospirillaceae (see Fig. 3). As discussed elsewhere (Thiergart et al., 2012; Martijn et al., 2018; Segata et al., 2013; Martin, 2018; Thiergart et al., 2014), there are several technical reasons that may account for low support values, which are not due to the phylogenetic signal of the cytochrome *b* protein itself, reported to be rather good (Degli Esposti et al., 2018; Degli Esposti, 2016; Degli Esposti et al., 2014). We have verified this by analysing portions of the whole alignment of cytochrome *b* and also its concatenated alignment with the cytochrome $c_1$ (Supplementary Fig. S1A and data not shown). To obtain alternative statistical support for tree topologies we then followed an approach based upon the detailed inspection of the tree branching specific for a given taxon with respect to two major clades within the Rhodospirillaceae family, which have emerged from different phylogenetic trees (Figs. 1–4 cf. (Wisniewski-Dyé et al., 2011; Kwan et al., 2012; Degli Esposti et al., 2018). One clade includes all the *Magnetospirillum* strains (Figs. 1–3, and also Fig. S1]). The other clade pivots on the genus *Azospirillum*, that nowadays includes at least 7 genomes (Table 1) and frequently clusters with the genus *Skermanella* (Fig. 2 cf. (Wisniewski-Dyé et al., 2011; Degli Esposti et al., 2018). These two clades correspond to the recently described clades of Rhodospirillaceae 1 and 2 (Martijn et al., 2018), which include *Magnetospirillum* and *Azospirillum*, respectively (Table 1); their topological relationship with other taxa was analysed within various phylogenetic trees obtained from either DeltaBLAST searches (as in Fig. 2) or from refined alignment of different proteins markers (as in Fig. 3). Some taxa were found in sister position to one or the other clade mentioned above, or clustered in branches that were clearly separate from either clade. In the latter cases, test taxa could lie in a branch that did not show sisterhood to the above taxonomic groupings and were then categorized as part of a **separate clade**. Conversely, when taxa were found to be in sister position to clades that did not include more than one member of the Rhodospirillaceae family, then they were deemed to be in a **sister of another clade** (Fig. 5). This approach is basically similar to previous analyses that dissected phylogenetic trees following topology configurations (Rochette et al., 2014; Williams et al., 2007). To simplify our topology analysis, other possible configurations were not considered. Hence, test taxa were placed into four mutually exclusive topology categories, corresponding to three degrees of freedom in probabilistic terms, as shown in Fig. 5. Random distribution of a given taxon within a single topology category was considered the null hypothesis and expected to be around 25% of the total number of independent trees examined. Significant deviations in the distribution frequencies were examined with the Chi squared ($\chi^2$) test using R (Core Team, 2018); in practice, **p** values lower than 0.05 were obtained whenever the distribution of test taxa within a single topology category exceeded 50% of the total number of trees examined. Strong statistical support for a specific set of frequencies deviating from the



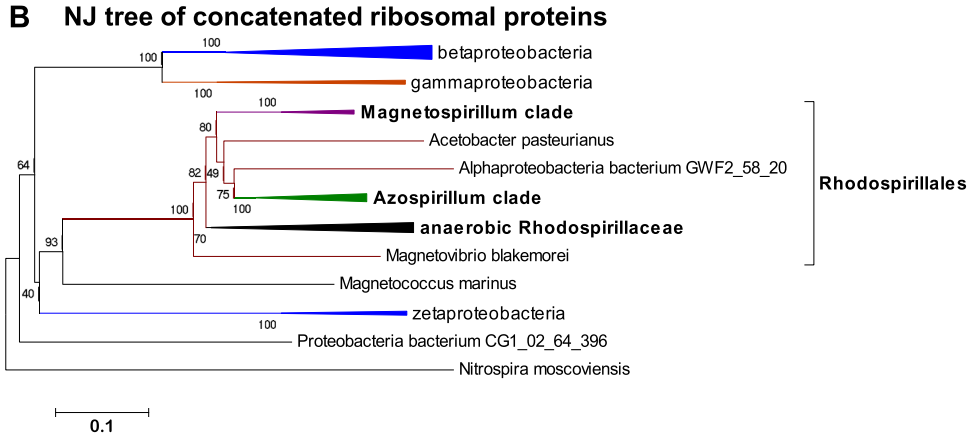




**Fig. 1. A.** ML tree of concatenated ribosomal proteins (Probst et al., 2017) for 28 proteobacterial taxa. The tree was obtained with the RaxML program as described earlier (Degli Esposti et al., 2016) using a concatenated alignment of 23 conserved ribosomal proteins (Probst et al., 2017). Evolutionary model (LG + I + G) was selected using ProtTest v3.4.2 (Darrida et al., 2011). SH-like branch supports values are shown in red aside the internal branches. The dashed box groups the Rhodospirillales taxa, which belong to the Rhodospirillaceae family except *for Acetobacter pausterianus*. **B.** NJ tree obtained with the alignment of 28 concatenated ribosomal proteins as described by Banfield and co-workers (Probst et al., 2017). The percentage value of the bootstraps (500) is shown by each node. Zetaproteobacteria include *Mariprofundus ferrooxydans* and zetaproteobacerium TAG-1 (*Ghiorsea bivora*), gammaproteobacteria include endosymbiont of *Ridgea picesae* and *E. coli* K-12, while betaproteobacteria include betaproteobacterium RGB_16_56_24 and *Methylotenera* RIFCSPLOWO2_02_FULL_45_14. The dashed box groups the Rhodospirillales taxa as in A. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expected random distribution was considered when **p** values of the $\chi^2$ test were < 0.001 (Fig. 5). The same analysis was applied to the various trees reported in the Supplementary Information of Martijn et al. (2018) (Supplementary Fig. S4), which had been previously analysed with statistical tools (Blanc et al., 2015).

### 2.4. Taxonomic distribution of top Blast hits to quantify phylogenetic diversity

Because the integration of the results obtained by the three approaches described so far did not yield a clear picture for the possible affiliation of some taxa, we developed a new approach to quantify the phylogenetic diversity of Rhodospirillaceae, based upon the distribution of top Blast hits obtained against the entire NR database for a selection of 120 proteins. In principle, such an approach would not be novel, because it follows the routine Blast searches that are run to verify the closest relatives of a given taxon. However, the novelty consists in the systematic application to a flexible selection of proteins that collectively define the energy conservation capacity of a bacterium, and consequently its inferred core physiology (see Supplementary Table S1 for the cumulative list of the proteins used). As mentioned in Section 2.1, the protein sets that are normally used to define the taxonomic affiliation of MAGs overwhelmingly contain ribosomal and nucleotide-reacting proteins (Probst et al., 2017; Probst et al., 2018; Martijn et al., 2018; Segata et al., 2013; Schulz et al., 2017; Ciccarelli et al., 2006; Imhoff, 2016; Martin, 2018), thereby overlooking the functional diversity coded into bacterial genomes (Degli Esposti et al., 2018; Segata et al., 2013; Martin, 2018). Functional diversity is much less conserved than ribosomes in prokaryotic *phyla* (Degli Esposti and Martinez-Romero, 2017; Martin, 2018), but tends to follow common traits at the taxonomic level of family and genus (Degli Esposti and Martinez-Romero, 2017; Louca et al., 2016). We found that the following functional traits are present in nearly all members of the Rhodospirillaceae family and their possible relatives (Table 1): the operon of F0F1 ATP synthase (Martin, 2018), the operon of NADH-quinone oxidoreductase or complex I (Degli Esposti and Martinez-Romero, 2017; Degli Esposti, 2017; Degli Esposti et al., 2019), the operon of the bc1 complex (Degli Esposti, 2016), at least one operon of terminal oxidases (Degli Esposti

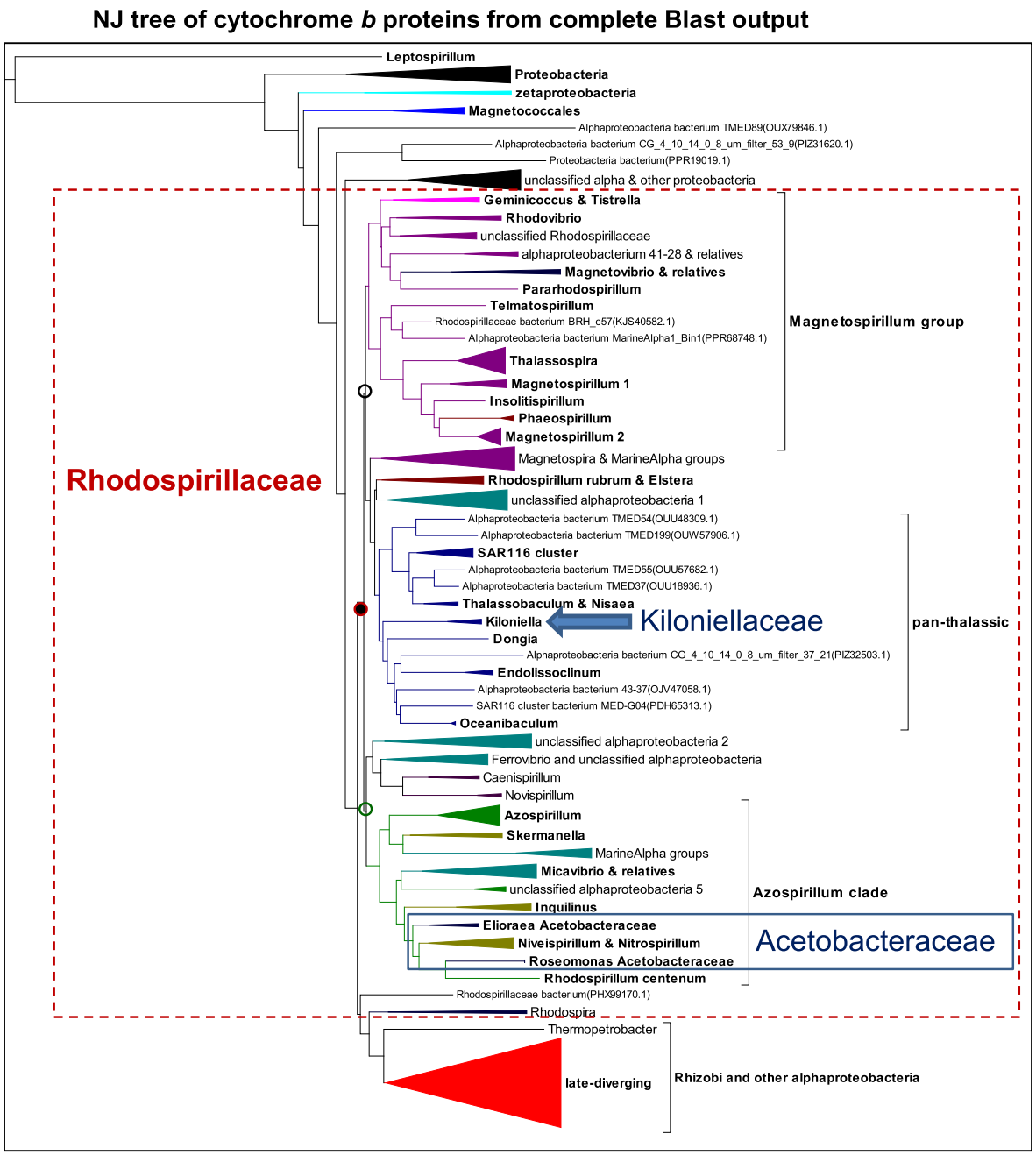



**Fig. 2.** A reference phylogenetic tree was obtained directly from a wide Blast search of cytochrome *b*. The NJ tree was obtained from a search of the cytochrome *b* protein extended to all the Rhodospirillaceae taxa that are known today, classified and unclassified, together with unclassified alphaproteobacteria. The original DeltaBLAST tree was downloaded in Newick and elaborated with the MEGA5.2 program, leaving the layout in descending order of phylogenetic distance. The branch of Kiloniellaceae is indicated by the blue arrow and includes also Rhodospirillales bacterium 47_12_T64, which clusters with *Kiloniella* also in all other phylogenetic trees we have produced. Note that the proteins from late diverging taxa (Rhizobiales and Rhodobacterales) cluster together in the bottom branch labelled 'late-diverging', subtended by the cytochrome *b* of *Thermopetrobacter*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and Martinez-Romero, 2017; Degli Esposti et al., 2018; Degli Esposti et al., 2019), the multigene systems for the assembly of FeS proteins, the multigene system of cell division (Adams and Errington, 2009), the multigene pathways of Q biosynthesis (Degli Esposti, 2017), the multigene system of outer membrane assembly and regulation (Konovalova et al., 2017), the functionally integrated gene sets for central metabolism (the citric and other metabolic cycles) and multiple genes for lipid metabolism. Strictly anaerobic MAGs from human gut microbiomes (Nielsen et al., 2014; Brown et al., 2016), which will be hereafter called anaerobic Rhodospirillaceae (Degli Esposti et al., 2019), possess other multigene traits such as the Rnf complex (Degli Esposti et al., 2018), pyruvate:ferredoxin oxidoreductase (PFO), as well as [FeFe]-hydrogenases and their maturases (Degli Esposti et al., 2016). The proteins defining such traits could thus compensate for those missing due to the absence of common traits, together with conserved proteins having other function, ranging from redox (for instance, superoxide dismutase) to nucleotide biosynthesis (for instance, CTP synthetase). These proteins were added to the list specific for each taxon until reaching a common total of 120. However, the cumulative number of proteins considered for all taxa exceeded 300 (Supplementary Table S1), since several of these proteins were specific to some taxa, functional traits varied among taxa and their genome completeness also differed.

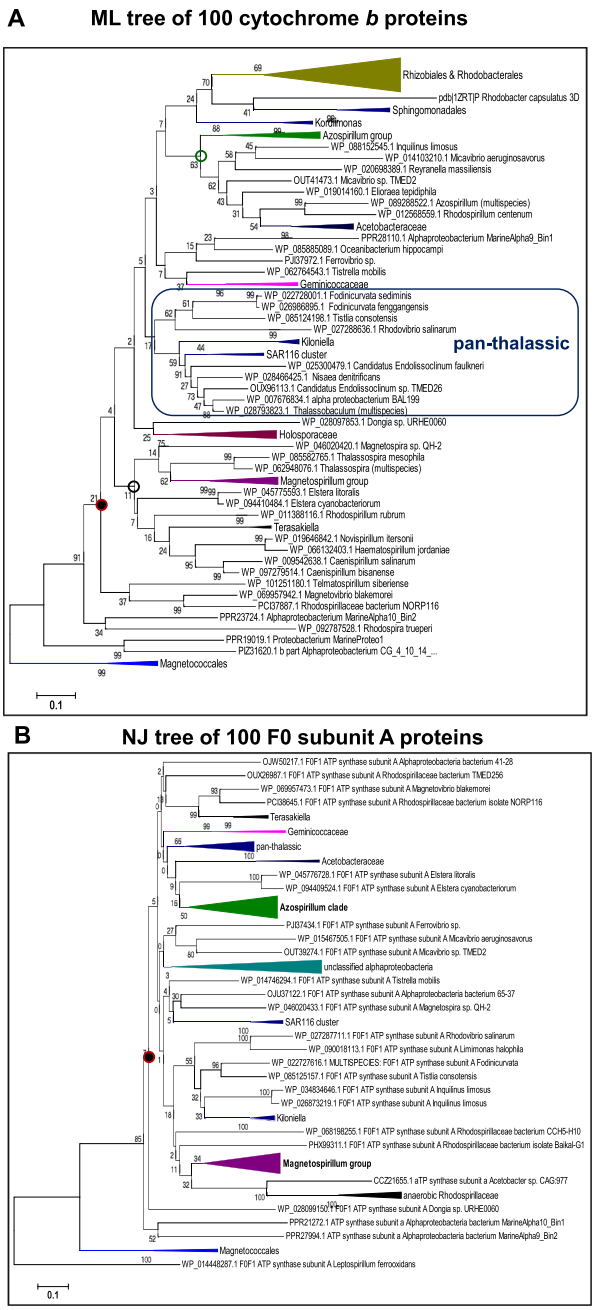



**Fig. 3. The taxonomic diversity of Rhodospirillaceae is expanded by phylogenetic trees obtained with single marker proteins. A.** The ML tree was obtained from the manually curated alignment of 100 cytochrome *b* sequences as described in Section 2.2. The various subdivisions are colour coded as in Fig. 2. **B.** The NJ tree was obtained from the manually curated alignment of 100 sequences of subunit A of the F0 membrane part of the F0F1 ATP synthase. Note that the tree includes the proteins from anaerobic Rhodospirillaceae, namely anaerobic MAGs from human gut metagenomes including *Acetobacter* sp. CAG:977 in deep branching position. See Supplementary Fig. S1B for a detailed tree of these MAGs.

Notably, only 8 of these proteins have been considered in the method recently described to define the taxonomic affiliation of 8000 MAGs (Parks et al., 2017), which is also based upon a total of 120 proteins. In the latter method (Parks et al., 2017), 37% of the proteins are nucleotide-reacting and 56% ribosomal, while our cumulative list included only 9% of nucleotide-reacting proteins and no ribosomal protein (Supplementary Table S1).

The proteins listed for each taxon were then inserted in a Blast search pipeline against the most recent version of the whole NR database and the top hits were retrieved in dedicated excel files as shown for *Magnetovibrio blackemori* (Supplementary Table S2). Although initial searches were extended to the top 10 hits produced by the Blast program, subsequent analysis was limited to the top five hits to reduce the 'taxonomic noise', defined here as the extent of hits that belong to bacterial *phyla* distantly related to that of the test taxon. This concept reflects the inherent mosaicism of bacterial genomes (Trubitsyn et al., 2016; Esser et al., 2007; Ji et al., 2017), which varies around 10% of genomic Blast hits in most classified Rhodospirillaceae, as we have verified in preliminary studies that confirmed previous reports (Esser et al., 2007; Ji et al., 2017). The results of the top five hits were filtered for the hits belonging to the same test taxon (or closely related taxa) and then evaluated for their taxonomic distribution following the major alphaproteobacterial classes and the standardized grouping of alphaproteobacteria shown in Supplementary Table S2. The frequency distribution of the top hits among the Rhodospirillaceae was analysed in granular detail following the latest comprehensive list of their taxa (Table 1, see also Supplementary Table S2), which was partially modified for several test taxa in order to account for bacterial strains showing the largest number of top hits. The frequency distribution of the top hits was computed using the estimated number of filtered results as 100% - the cumulative number of counted hits closely approached this reference value (Supplementary Table S2), the difference deriving from typing errors or other taxa that were not included in the standardized taxonomy scheme. The frequency distribution was subsequently simplified into major groupings as shown in Tables 2 and 3. The frequency values for individual taxonomic groupings were evaluated using non parametric tests such as Mann-Whitney and were considered highly significant when higher or lower than the median plus three Standard Deviations (SD) of a set of values obtained with at least seven different Rhodospirillaceae (see Table 3 for additional details). The selection of 120 proteins produced frequency results that were highly correlated with those obtained using all the proteins coded by the genome of a classified Rhodospirillaceae such as *Magnetovibrio* (Supplementary Fig. S2A). However, strong increases in the number of searched proteins enhanced the taxonomic noise in the results (see above), due to the different evolution and degree of conservation of the diverse proteins considered, and interferences from phenomena of Lateral Gene Transfer (LGT), as noted previously (Thiergart et al., 2012). Conversely, narrowing the Blast searches to the top three hits resulted in loss of phylogenetic diversity in the results, due the large number of taxa considered in our analysis (Supplementary Fig. S2B). Ultimately, the information obtained with these and other calibration studies indicated that a matrix of 120 proteins per five top hits provided enough information to strike a balance between limited taxonomic noise and broad taxonomic diversity in evaluating the affiliation of test taxa to the Rhodospirillaceae family. Clearly, the taxonomic evaluation could be skewed when a test taxon has close relatives with genomes deposited in the NR database (Table 1). In such cases, the search matrix was extended to six or more top hits maintaining the same set of proteins, so as to obtain a minimum of 150 hit results after eliminating those from taxa closely related to the test taxon, as shown in the right part of Table 2.

### 2.5. Other methods

The evaluation of Average Nucleotide Identity (ANI) among genomes was undertaken as previously described (Richter and Rossello-Mora, 2009) using various options of a software program downloaded in Phyton from: https://github.com/widdowquinn/pyani. The percentage values of ANI across aligned genomes was the most frequently used option.

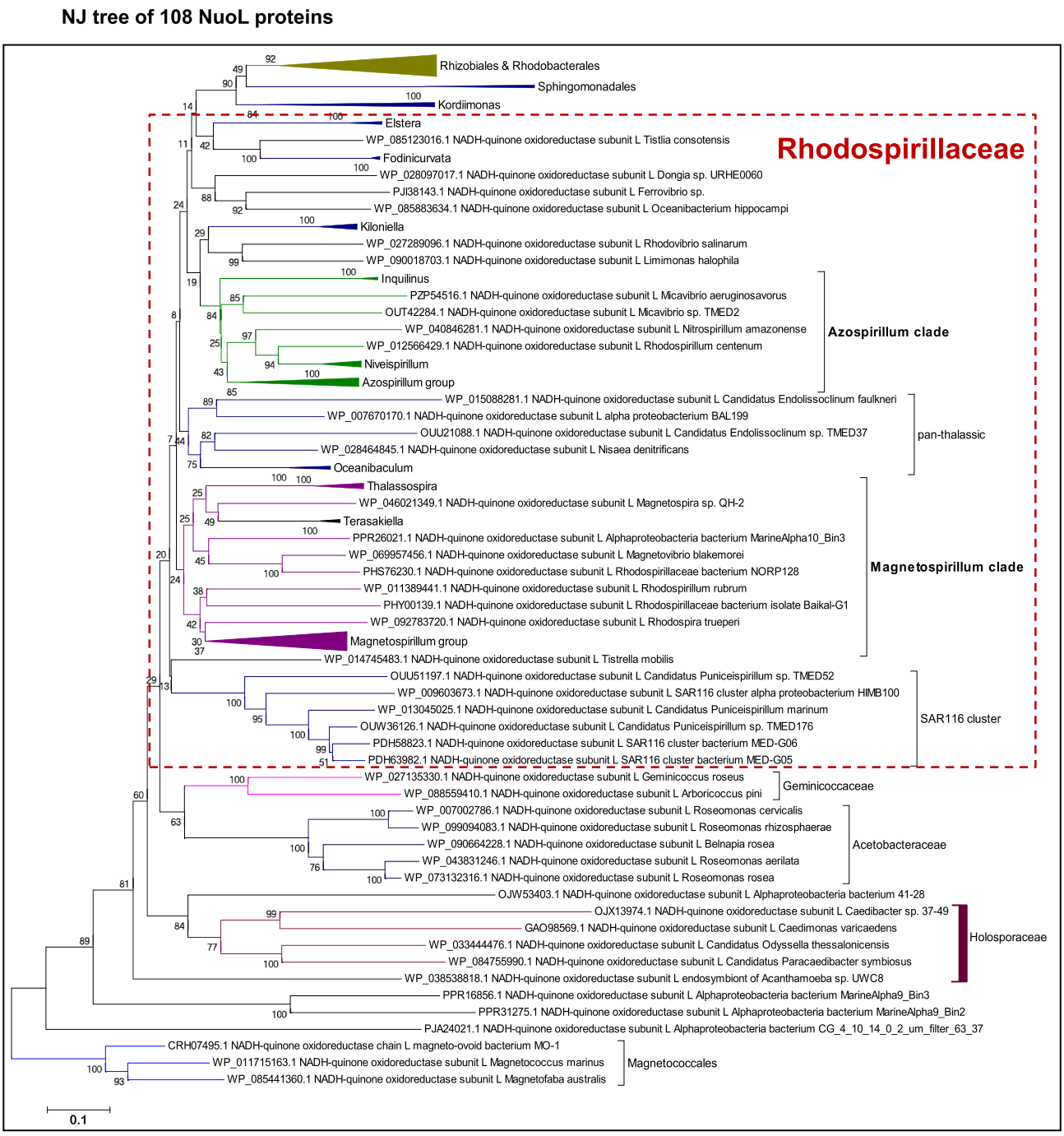



**Fig. 4.** Phylogenetic tree of the NuoL subunit of complex 1. The NJ tree was obtained from the manually curated alignment of 108 sequences of the NuoL (ND5) subunit of complex I, using the same colour code as that in Fig. 2 to differentiate the various subdivisions. Note the clustering of *Magnetovibrio* and closely related taxa within the *Magnetospirixlum* clade.

## 3. Results and discussion

### 3.1. Phylogenetic study of the Rhodospirillaceae using trees based on concatenated ribosomal proteins

The current phylogenetic diversity of the Rhodospirillaceae family is poorly known, but it is likely to be significantly larger than that considered in previous taxonomic studies (Wisniewski-Dyé et al., 2011; Kwan et al., 2012) due to the recent explosion in metagenomic sequencing (Anantharaman et al., 2016; Tully et al., 2017; Parks et al., 2017; Degli Esposti et al., 2018; Martijn et al., 2018). We have excluded that ANI evaluation procedures applied to available genomes could provide new information to assess the phylogenetic diversity of the Rhodospirillaceae family (Supplementary Fig. S3 and data not shown). Therefore, we resorted to approaches based upon protein sequences to evaluate such diversity, starting with phylogenetic trees derived from the concatenation of multiple ribosomal proteins (Probst et al., 2017) (Fig. 1).

Several MAGs reported from human microbiomes and environmental sources were compared together with representatives of classified Rhodospirillaceae, unclassified betaproteobacteria, gammaproteobacteria including *E. coli* and zetaproteobacteria. A representative of the *phylum* Nitrospirae, which is considered to be sister to the *phylum* proteobacteria (Degli Esposti et al., 2018), was used for rooting the

## Tree topology analysis for groups of alphaproteobacteria

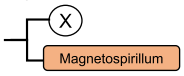
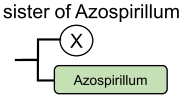
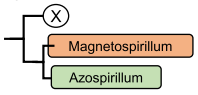
| Tree topology of test taxon (X) | SAR116 & pan-thalassic | Tistrella & Geminicoccus | Aceto bacteraceae | Terasakiella | Anaerobic Rhodospirillaceae | Holosporaceae |
|---|---|---|---|---|---|---|
| sister of Magnetospirillum   | 11<br>25.0% | 11<br>34.4% | 1<br>5.6% | **14**<br>**93.3%** | **7**<br>**58.3%** | 1<br>8.3% |
| sister of Azospirillum   | 13<br>29.5% | 0 | **13**<br>**72.2%** | 0 | 1<br>8.3% | 0 |
| separate clade   | 12<br>27.3% | 4<br>12.5% | 0 | 0 | 3<br>25.0% | 2<br>16.7% |
| sister of another clade   | 8<br>18. 2% | **17**<br>**53.1%** | 4<br>22.2% | 1<br>6.7% | 1<br>8.3% | **9**<br>**75.0%** |
| Total number of tree topologies examined (100%) | 44 | 32 | 18 | 15 | 12 | 12 |
| **p** value | 0.7356 | **0.0001** | **0.0000** | **0.0000** | 0.0460 | **0.0008** |

**Fig. 5. Analysis of tree topology configurations for selected groups of alphaproteobacteria.** The analysis was performed by visual inspection of various phylogenetic trees obtained with different phylogenetic markers such as those in Figs. 2–4 and Supplementary Fig. S1. The results of the Chi squared ($\chi^2$) test of the frequency values are reported in the bottom row of the list and are in bold when their **p** value is below 0.001.

phylogenetic trees of concatenated ribosomal proteins. With such a broad phylogenetic span, the branches containing Rhodospirillaceae (comprised within the red square in Fig. 1A) were relatively short, but their nodes generally had strong statistical support with both methods used (Fig. 1A,B), thereby defining solid phylogenetic features. The family appears to contain a deep branch with *Magnetovibrio*, followed by three major clades: one clade containing *Magnetospirillum*, another clade containing classified *Azospirillum* strains and a third clade comprising unclassified *Azospirillum* strains that cluster together with other anaerobic MAGs. Some of such MAGs have been previously classified under the *Acetobacter* genus (Nielsen et al., 2014), while other MAGs have been categorized under broader taxonomic levels such as 'unclassified proteobacteria', for example Proteobacteria bacterium sp. CAG:495 (Nielsen et al., 2014). Conversely, the environmental anaerobic MAG, alphaproteobacteria bacterium GWF2_58_20 (Anantharaman et al., 2016; Degli Esposti et al., 2019), clusters with the *Azospirillum* clade, which in the tree of Fig. 1 includes also *Acetobacter pasteurianus*, a representative of the other major family of the Rhodospirillales order, the Acetobacteraceae (Williams et al., 2007; Proença et al., 2018).

The phylogenetic features deduced from the trees in Fig. 1 follow trends that have been observed previously, but not discussed in detail. In particular, phylogenetic trees based upon the 16S rRNA marker have shown that the Rhodospirillaceae family is split into two major sister

**Table 2**
Taxonomic distribution of Blast hits frequencies for taxa associated with the Rhodospirillaceae. Frequency values are in bold whenever higher than 20% and colour coded in pale red or pale blue when they are significantly higher or lower, respectively, than the mean obtained with several reference taxa (see Table 3 for details).

| taxonomic group of top hits | *Magnetovibrio blackemori* | *Terasakiella* sp. PR1 | *Roseospirillum parvum* | alpha. GWF2_58_20 | *Kiloniella laminariae* | alpha. RIFOXYD12_FULL_60_8 | Acetobacter sp. CAG:977 | Azospirillum sp. CAG:239 | Azospirillum sp. CAG:239* | Acetobacter sp. 46_36* | Proteobacterium sp. CAG:495 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Azospirillum** clade | 4.40 | 6.17 | 5.41 | **35.99** | 9.46 | 6.83 | 3.49 | 1.86 | 6.96 | 6.36 | 1.84 |
| **Magnetospirillum** clade | **46.65** | **46.36** | **67.99** | 11.88 | 10.41 | **64.00** | **23.42** | 5.74 | **21.52** | **24.29** | 5.03 |
| **pan-thalassic** group | 2.46 | 6.17 | 2.71 | 5.09 | **21.14** | 3.93 | 2.82 | 0.68 | 2.53 | 1.16 | 0.67 |
| **anaerobic Rhodospirillaceae** | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 11.13 | **73.31** | *0.00* | *0.00* | **75.54** |
| separate Rhodospirillaceae | 0.53 | 0.95 | 0.32 | 1.59 | 1.89 | 1.33 | 1.16 | 0.34 | 0 | 0.00 | 0.17 |
| unclassified Rhodospirillaceae | 3.35 | 7.91 | 8.76 | 0.68 | 0.32 | 2.50 | 1.66 | 0.84 | 3.16 | 1.16 | 0.34 |
| **total Rhodospirillaceae** | **57.39** | **67.56** | **83.76** | **55.18** | **43.21** | **78.33** | **42.19** | **81.76** | **34.18** | **33.53** | **83.58** |
| Rhodospirillales | **66.37** | **76.58** | **85.67** | **57.06** | **68.14** | **82.67** | **45.18** | **82.77** | **36.71** | **35.84** | **84.09** |
| other classified alphaproteobacteria | 6.87 | 9.18 | 5.57 | 10.70 | 16.40 | 3.00 | 12.96 | 3.38 | 12.56 | 10.98 | 4.36 |
| unclassified alphaproteobacteria^ | 8.63 | 6.65 | 2.23 | 4.07 | 3.47 | 1.83 | 1.99 | 1.35 | 5.06 | 6.94 | 0.34 |
| **total alphaproteobacteria** | **81.87** | **89.40** | **91.24** | **71.82** | **88.01** | **87.50** | **60.00** | **88.34** | **54.43** | **53.76** | **88.78** |
| **total betaproteobacteria** | 2.64 | 2.06 | 1.59 | 4.41 | 0.00 | 4.33 | 2.67 | 0.34 | 1.27 | 0.00 | 0.34 |
| **total gammaproteobacteria** | 13.03 | 3.96 | 6.37 | 9.51 | 7.57 | 3.17 | 3.67 | 0.84 | 3.16 | 11.56 | 0.84 |
| **total delta/epsilon proteobacteria** | 0.18 | 1.42 | 0.00 | 3.40 | 0.95 | 0.17 | 8.14 | 2.53 | 9.49 | 5.78 | 1.51 |
| **total non-proteobacterial prokaryotes** | 0.88 | 2.53 | 0.80 | 8.15 | 1.26 | 2.83 | **19.77** | 7.26 | **27.22** | **25.43** | 8.04 |
| total estimated hits, 100% | 568 | 632 | 628 | 589 | 317 | 600 | 602 | 592 | 158* | 173* | 597 |

*Results obtained after removing the hits of closely related taxa.
^Taxa currently classified as unclassified that are not clustering with groups of Rhodospirillaceae.

**Table 3**
Taxonomic distribution of Blast hits frequencies for various alphaproteobacterial taxa. Frequency values are in bold whenever higher than 20% and colour coded in pale red or pale blue when they are significantly higher or lower, respectively, than the mean obtained with several reference taxa; the mean values are reported in the right of the table together with the SD and the limits considered for highly significant differences (below or above the mean by three Standard Deviations, 3SD). Note that the results obtained with Alphaproteobacteria bacterium RIFCSPHIGHO2_12_42_13 are very similar to the data shown for Alphaproteobacteria bacterium 41–28.

| taxonomic group of top hits | *Magnetovibrio blackemori* | uncultured α isolate 86 | *Magnetococcus marinus** | *Magnetofaba australis** | MarineProteo 1 _Bin1&2 | Alphaproteo. CG_4_...._53_ 9 | Alphaproteo bacterium 41-28 | Alphaproteo bacterium 40-19 | endosymbiont Acanthamoeb a sp. UWC8 | mean | SD | mean ±3SD |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Azospirillum clade | 4.40 | 9.72 | 2.72 | 1.46 | 2.60 | 2.02 | 14.00 | 2.02 | 4.89 | 6.17 | 3.57 | over 16.9 significant |
| Magnetospirillum clade | **46.65** | **43.83** | 19.03 | 12.37 | 6.40 | 5.33 | 9.27 | 2.33 | 7.18 | 6.4 | 3.09 | over 15.7 significant |
| pan-thalassic group | 2.46 | 6.94 | 2.10 | 2.32 | 4.27 | 4.54 | 8.43 | 1.55 | 4.23 | 4.19 | 2.63 | over 12.1 significant |
| separate Rhodospirillaceae | 0.53 | 1.08 | 0.60 | 1.16 | 1.04 | 1.68 | 3.71 | 0.93 | 2.94 | 1.16 | 0.59 | over 2.93 significant |
| unclassified Rhodospirillaceae | 3.35 | 2.15 | 1.51 | 1.60 | 0.87 | 2.02 | 5.56 | 1.71 | 1.31 | 1.51 | 1.31 | over 5.44 significant |
| **total Rhodospirillaceae** | **57.39** | **63.73** | **25.68** | 18.90 | 15.17 | 16.50 | **38.28** | 7.29 | **20.55** | **54.98** | 9.55 | < 26.3 significant |
| Rhodospirillales | **66.37** | **65.59** | **27.49** | **21.10** | 18.86 | **22.83** | **42.86** | 11.32 | **24.47** | | | |
| other classified alphaproteobacteria | 6.87 | 11.42 | 5.44 | 8.15 | **26.64** | **27.17** (0.67)# | **22.77 (16.02)#** | **67.44 (51.94)#** | **57.1 (47.96)#** | 7.39 | 3.37 | over 17.5 significant |
| unclassified alphaproteobacteria^ | 8.63 | 2.16 | 2.42 | 3.20 | 11.94 | 11.93 | **23.61** | 5.74 | 8.81 | 2.95 | 2.64 | over 10.9 significant |
| **total alphaproteobacteria** | **81.87** | **79.17** | **35.35** | **32.56** | **57.44** | **62.22** | **89.21** | **84.50** | **90.38** | **88.01** | 6.37 | < 68.8 significant |
| total betaproteobacteria | 2.64 | 4.78 | 6.04 | 7.42 | 2.77 | 6.55 | 0.34 | 1.24 | 1.31 | 2.67 | 1.39 | over 6.84 significant |
| total gammaproteobacteria | 13.03 | 8.18 | **28.10** | **21.54** | 16.44 | 11.50 | 4.38 | 6.67 | 2.77 | 3.67 | 3.25 | over 13.43 significant |
| total delta/epsilon proteobacteria | 0.18 | 2.01 | 12.08 | 14.26 | 3.63 | 3.33 | 0.00 | 1.09 | 0.33 | 1.51 | 1.28 | over 5.35 significant |
| total non-proteobacterial prokaryotes | 0.88 | 1.39 | 5.80 | 13.68 | 14.53 | 12.10 | 4.72 | 6.51 | 3.59 | 5.26 | 2.76 | over 13.6 significant |
| total estimated hits, 100% | 568 | 648 | 331* | 687 | 578 | 600 | 593 | 645 | 613 | | | |

*Results obtained after removing the hits of closely related taxa.
^Taxa currently classified as unclassified that are not clustering with groups of Rhodospirillaceae.
#(within brackets) Values obtained for hits of taxa belonging to the Holosporales and Rickettsiales.

clades, one containing Magnetospirilli and the other containing Azospirilli (Wisniewski-Dyé et al., 2011; Kwan et al., 2012; Proença et al., 2018), consistent with the diverse phenotypic characters that separate Magnetospirilli from Azospirilli (Garrity et al., 2005; Wisniewski-Dyé et al., 2011). The trees in Fig. 1 produce also new phylogenetic information, since many anaerobic MAGs are now found to be strongly associated with the Rhodospirillaceae family, irrespective of their original classification obtained with methods using ribosomal proteins (Ciccarelli et al., 2006; Nielsen et al., 2014; Brown et al., 2016; Anantharaman et al., 2016). We therefore shall refer to these taxa as anaerobic Rhodospirillaceae (Fig. 1B cf. (Trubitsyn et al., 2016). To define their relationship with cultivated members of the Rhodospirillaceae family, and further explore the phylogenetic diversity of the family, we followed the other approaches described in Sections 2.2–2.4.

### *3.2. Trees based upon single protein markers expand the phylogenetic diversity of Rhodospirillaceae.*

The overall phylogeny of alphaproteobacteria can be visualized using NJ trees derived from broad Blast searches of the conserved protein cytochrome *b* (Degli Esposti and Martinez-Romero, 2017; Degli Esposti et al., 2018; Degli Esposti et al., 2014). Although such trees contain tightly packed internal branches with poor statistical support (see later), they avoid a very common problem in producing phylogenetic trees based upon single protein markers: the arbitrary choice of taxonomic sampling (Degli Esposti et al., 2018; Trubitsyn et al., 2016; Degli Esposti, 2016). As discussed elsewhere (Degli Esposti et al., 2018), this problem has complicated previous studies on the phylogeny of alphaproteobacteria and their relationship with mitochondria (Rochette et al., 2014; Williams et al., 2007). Any clever selection of taxa that excludes uncultivated or unclassified taxa, which are now abundantly present in NCBI resources, produces an arbitrary bias of sampling that can seriously affect the topology of phylogenetic trees (Degli Esposti et al., 2018). An effective way to avoid this bias is to extend Blast searches of a marker protein to all alphaproteobacterial taxa currently available in the NR database. Because this database contains an over-representation of taxa important for agriculture or medicine (Schulz et al., 2017) (chiefly, Bradyrhizobiaceae, Rhizobiaceae, Brucellaceae and Rickettsiales among alphaproteobacteria), the broad Blast searches can be restricted to only a few members of such over-represented groups. A similar selection can be extended to Pelagibacterales too, in order to minimize the tree distortions produced by their AT-rich genomes, a feature shared with Rickettsiales (Martijn et al., 2018). To reduce phylogenetic redundancy, additional restrictions can be applied to other alphaproteobacterial orders once the basic tree topology of the class has been defined (Fig. 2 cf. (Degli Esposti et al., 2018). The NJ tree shown in Fig. 2 has been obtained following these criteria and therefore may represent an unbiased blueprint for the phylogenesis of most Rhodospirillaceae and the taxa potentially affiliated with them. Similarly to Fig. 1, the tree in Fig. 2 is rooted with a representative of the *phylum* Nitrospirae, in this case *Leptospirillum*. The following discussion assumes that the tree of cytochrome *b* reflects basic phylogenetic relationships among alphaproteobacterial taxa (Degli Esposti et al., 2018; Degli Esposti, 2016).

In the tree of Fig. 2 there is a single large branch that includes nearly all known Rhodospirillaceae, indicated by the filled dot within the dashed square encompassing the phylogenetic span of the family. This suggests that the Rhodospirillaceae family may be monophyletic, in agreement with some studies (Wisniewski-Dyé et al., 2011) and Fig. 1, but branches into two major clades (indicated by empty dots in Fig. 2) which broadly correspond to the *Azospirillum* clade and the *Magnetospirillum* clade that have been discussed earlier (Fig. 1 cf. (Wisniewski-Dyé et al., 2011; Degli Esposti et al., 2018; Proença et al., 2018). The *Azospirillum* clade is partially intermixed with deep branching members of the Acetobacteraceae family such as *Eliorea*, *Belnapia* and *Roseomonas* (cf. (Proença et al., 2018), confirming that this clade of Rhodospirillaceae has close phylogenetic relationships with the Acetobacteraceae family (cf. Fig. 1), as further sustained by subsequent analysis (Fig. 5).

Diverse groups of unclassified alphaproteobacteria are nested within each major clade of the Rhodospirillaceae (Fig. 2 cf. (Degli Esposti et al., 2018), thereby expanding the phylogenetic diversity of the family beyond the taxonomic boundaries considered so far (Table 1). The *Azospirillum* clade includes the following groups of previously unclassified taxa: *Micavibrio*, a genus of ectoparasites (Davidov et al., 2006), and its MAG relatives, for instance *Micavibrio* sp. TMED 27 (Tully et al., 2017), plus other MAGs such as alphaproteobacteria bacterium isolate NORP106 (Tully et al., 2018); MAGs of the MarineAlpha5 group (Martijn et al., 2018) and; a group of three taxa, labelled 'unclassified alphaproteobacteria 5′ at the bottom of Fig. 2, containing Alphaproteobacteria bacterium 65–37 and the unclassified Rhodospirillaceae strain CCH5-H10, the genome of which has been deposited directly into the NR database (https://www.ncbi.nlm.nih.

gov/Taxonomy/Browser/wwwtax.cgi?id=1768806, accessed 18 June 2019).

Unclassified Rhodospirillaceae and alphaproteobacteria are inserted also within the *Magnetospirillum* clade, either in groups forming distinct branches or as isolated taxa (Fig. 2, top to middle). Some MAGs are closely associated instead with classified members of the Rhodospirillaceae. For instance, Rhodospirillaceae bacterium isolate NORP116 and NORP128 (Tully et al., 2018) tightly cluster with *Magnetovibrio* in the branch labelled 'Magnetovibrio & relatives' in Fig. 2, top. Because this tight association has been consistently found in trees obtained with other marker proteins (Fig. 3 and results not shown), it is quite likely that isolate NORP116 and NORP128 represent new strains of the genus *Magnetovibrio*.

Interestingly, a well-defined group contains the SAR116 cluster of oceanic organisms (Oh et al., 2010), which are classified as 'unclassified alphaproteobacteria' (https://www.ncbi.nlm.nih.gov/Taxonomy/, accessed 15 October 2018). Recently, however, members of the SAR116 cluster have been proposed to form a clade with *Endolissoclinum* (Martijn et al., 2018), an endosymbiont of marine invertebrates (Kwan et al., 2012), and its relative alpha proteobacterium BAL199, which have been previously labelled 'pan-thalassic' (Degli Esposti et al., 2014). Our phylogenetic trees clearly sustain this phylogenetic association with Rhodospirillaceae (Figs. 2 and 3), which includes also marine MAGs and classified bacteria such as *Oceanibaculum* and *Thalassobaculum*. We have thus preliminarily labelled the overall group containing all the taxa just mentioned as 'pan-thalassic' for it encompasses only marine bacteria. Remarkably, all the cytochrome *b* proteins of the Kiloniellaceae family (Davidov et al., 2006; Imhoff and Wiese, 2014) cluster right in the middle of this group of Rhodospirillaceae (Figs. 2 and 3). Although a *Kiloniella* species has been recently considered among unclassified Rhodospirillales (Martijn et al., 2018), the family named after the genus *Kiloniella* is formally classified in a separate order from that of Rhodospirillales (Imhoff and Wiese, 2014). We therefore expanded our phylogenetic analysis to verify the novel possibility that the entire family of Kiloniellaceae may represent a subdivision of the Rhodospirillaceae, instead of its traditional classification as a different taxonomic entity (Imhoff and Wiese, 2014). The analysis was also aimed to clarify the phylogenetic relationship of another family of Rhodospirillales that has been proposed recently, the Geminicoccaceae (Proença et al., 2018).

The newly introduced family of Geminicoccaceae is named after its proposed type species, *Geminicoccus roseus*, a previously unclassified alphaproteobacterium (Brown et al., 2016; Foesel et al., 2007), the genome of which shares unique features with that of *Tistrella* (Xu et al., 2012), a deep branching Rhodospirillaceae (Wisniewski-Dyé et al., 2011; Martijn et al., 2018) exhibiting a rich variety of metabolic traits (Degli Esposti et al., 2014). Indeed, cytochrome *b* and other proteins of *Tistrella* and *Geminicoccus* cluster together (Figs. 2 and 3, cf. (Martijn et al., 2018), sometimes with those of *Arboriscoccus pini*, which was previously known as Rhodospirillales bacterium B29T1 and is also part of the Geminicoccaceae family (Proença et al., 2018). However, these species do not consistently form a branch with defined topology with respect to the bulk of the Rhodospirillaceae, and therefore require additional analysis for clarifying their phylogenetic relationship with the family.

### 3.3. Limits of phylogenetic trees based on protein markers and tree topology analysis

To verify the robustness of the new phylogenetic inferences on Rhodospirillaceae deduced from the trees in Figs. 1 and 2, we undertook the standard approach of constructing refined alignments of marker proteins and then compute Maximum Likelihood (ML) analysis with large number of bootstraps (Tamura et al., 2011; Degli Esposti et al., 2014). Approx. 100 sequences of cytochrome *b* proteins were aligned so as their preliminary tree could match the basic tree topology obtained with broad DeltaBLAST searches (Fig. 2); the alignment was subsequently refined by manual adjustments guided by the known 3D structure of the protein (Degli Esposti et al., 2014) and expanded by including sequences from taxa representing diverse orders of alphaproteobacteria besides the core group of Rhodospirillaceae (Fig. 3A and 4, see also Supplementary Fig. S1). Fig. 3A shows a representative ML tree of cytochrome *b* that was obtained by this approach and confirms the previous finding that proteins from the genus *Kiloniella* cluster within the group of 'pan-thalassic' taxa of the Rhodospirillaceae. This group appears to be relatively well defined in the tree of Fig. 3A, and is split in multiple branches when the sequences of cytochrome *b* are concatenated with those of cytochrome *c*1 (Supplementary Fig. S1), or when using other protein markers such as subunit A of F0F1 ATP synthase (Fig. 3B) and the NuoL subunit of complex I (Fig. 4). Conversely, the *Magnetospirillum* clade forms a cohesive group of Rhodospirillaceae taxa in either ML or NJ trees of diverse protein markers and generally includes *Magnetovibrio* and closely related taxa (Figs. 3 A and 4, see also Supplementary Fig. 1). Intriguingly, we found that conserved bioenergetic proteins of *Terasakiella*, a genus classified in the Methylocystaceae family of the order Rhizobiales (Han et al., 2016), frequently cluster with those of the *Magnetospirillum* clade (Figs. 3 and 4, see also Supplementary Fig. 1). Indeed, *Terasakiella* shares with *Magnetovibrio* ancestral operons for cytochrome *c* oxidase (Degli Esposti et al., 2018; Degli Esposti et al., 2019) and most proteins of the magnetosome island typical of magnetotactic Rhodospirillaceae, as recently documented (Monteil et al., 2018). However, the membrane A subunit of F0F1 ATP synthase of *Terasakiella* clustered in a branch containing *Magnetovibrio* that lies far away from the major group containing Magnetospirilli (Fig. 3B), consistent with a recent tree obtained with 16S rRNA (Monteil et al., 2018).

All phylogenetic trees constructed as described above showed two general features that weakened the information they provide, irrespective of the protein marker used and the number of aligned sequences: 1) the Rhodospirillaceae are split in multiples branches rather than forming a monophyletic group; 2) the internal nodes separating these branches show low, and sometimes very low values of bootstrap support. The latter feature is not uncommon also for concatenated proteins (supplementary Fig. S1A, cf. (Martijn et al., 2018) and fundamentally reflects the close phylogenetic relationships among member of the family, which cannot be finely resolved with conserved proteins such as cytochrome *b*. The former feature has been noted before (Williams et al., 2007) and interpreted to signify that the family of Rhodospirillaceae may be polyphyletic, with some taxa segregating in a different phylogenetic space from that occupied by other taxa along the alphaproteobacterial tree. Indeed, our trees consistently show that taxa associated with the *Azospirillum* clade segregate together with other alphaproteobacteria such as *Kordiimonas*, Sphingomonadales and Rhodobacterales, and also with members of the Rhodospirillales order that do not belong to the Rhodospirillaceae, namely Acetobacteraceae and Geminicoccaceae (Figs. 3A and 4 - see also some trees reported in Ref. (Martijn et al., 2018).

Notably, members of a deep-branching group of alphaproteobacteria, the Holosporaceae (single family of the order Holosporales (Szokoli et al., 2016), either occupy a basal branch with good support (Fig. 4) or intermix with clades of the Rhodospirillaceae that have relatively low branch support (Fig. 3A and S1). The incongruent relationship of Holosporaceae vs. Rhodospirillaceae in different phylogenetic trees has been discussed recently, without reaching a firm resolution on the basis of bootstrap analysis (Martijn et al., 2018). This and other phylogenetic inconsistencies prompted us to explore a different approach to evaluate the statistical likelihood that groups of closely related taxa could be associated with Rhodospirillaceae and its major clades. The approach is based upon visual inspection of the tree topology of one taxonomic group vs. the two major clades of Rhodospirillaceae, which can generate statistically robust inferences when the frequency of a specified topology configuration largely exceeds the

expected random variation of 25%, as shown in Fig. 5 (see also supplementary Fig. S4).

The application of our analysis of tree topology configuration has shown that the pan-thalassic group, represented by the SAR116 cluster and closely related taxa such as *Oceanibaculum*, distributes in the four mutually exclusive categories with comparable frequencies close to the expected random distribution (Fig. 5). This finding, confirmed by the analysis of the trees reported in (Martijn et al., 2018) (Supplementary Fig. S4), strongly suggests that the pan-thalassic group forms a phylogenetic subdivision that is independent of the major clades of Rhodospirillaceae; therefore, it could be considered as a separate taxonomic entity. A similar conclusion can be reached for the combination of *Tistrella* and *Geminicoccus*, which most frequently are found in sister position with respect to branches which do not contain Rhodospirillaceae (corresponding to the category of 'sister to another clade' in Fig. 5). This configuration also includes the cases in which *Tistrella* and *Geminicoccus* cluster together with *Arboriscoccus* (Supplementary Fig. S5), the other member of the proposed family of Geminicoccaceae (Proença et al., 2018). The resulting unclear phylogenetic relationship of *Tistrella* with other Rhodospirillaceae has been subsequently considered to arise from a separate taxonomic entity, as confirmed by further analysis (not shown). The same consideration has been applied to other members of the Rhodospirillaceae that did not consistently cluster with one group or the other, for instance *Dongia* (Liu et al., 2010). *Dongia* and the following taxa were found to have no statistically significant association with any major clade of Rhodospirillaceae by the approach of tree topology configuration (cf. Fig. 5): *Elstera* (Rahalkar et al., 2012); *Ferrovibrio*, MAGs of the MarineAlpha group 9&10 (Martijn et al., 2018) and *Enhydrobacter* (Kawamura et al., 2012).

The results in Fig. 5 strongly confirm that *Terasakiella* is affiliated to the Magnetospirillum clade (Fig. 3A cf. Ref. (Monteil et al., 2018); a similar conclusion may be extended to anaerobic Rhodospirillaceae, in agreement with the tree in Fig. 3B, but the statistical support for this possibility is not very strong (Fig. 5). In contrast, members of the Acetobacteraceae such as *Roseomonas* preferentially cluster with (or within) the *Azospirillu*m clade, consistent with previous finding (Figs. 1–3). Conversely, the Holosporaceae most frequently cluster with branches that do not contain Rhodospirillaceae (Fig. 5), which mainly correspond to the Rickettisales when the phylogenetic trees contain these taxa (Supplementary Fig. S4, cf. Ref. (Martijn et al., 2018). Such data indicates that the Holosporaceae occupy a phylogenetic space that is peripheral to that encompassing the majority of Rhodospirillaceae. This would be a much clearer conclusion than the unresolved situation discussed recently (Martijn et al., 2018) and has been tested further in our subsequent investigations (see Table 3 and Supplementary Fig. S5). The analysis of tree topology configuration has additionally indicated that the following genera are likely to be part of the pan-thalassic group: *Thalassobaculum, Tistlia, Fodinicurvata, Nisaea, Rhodovibrio, Limimonas* and also *Kiloniella*. The latter results are clearly consistent with the phylogenetic trees in Figs. 3 and 4 and will be further confirmed by the results of next approach we have followed (Table 2).

### 3.4. Taxonomic distribution of Blast hits of functional proteins refines the phylogenetic diversity of Rhodospirillaceae.

Although the approach of tree topology configuration helped sustaining new taxonomic affiliations to the Rhodospirillaceae, as in the case of *Kiloniella* and *Terasakiella* (Fig. 5), it did not produce conclusive data regarding the affiliation of other taxa, in particular for anaerobic MAGs such as *Acetobacter* sp. CAG:977. Another concern regarded the separation of the pan-thalassic group from other groupings within the Rhodospirillaceae, which so far has been mainly based upon the interpretation of tree topologies (Fig. 5). We thus examined a broad range of functional traits, expanding on those presented recently (Degli Esposti and Martinez-Romero, 2017; Degli Esposti et al., 2018; Degli Esposti et al., 2019) to verify whether the pan-thalassic group may have distinctive functional properties. The results of this analysis indicated that the majority of the taxa proposed to be part of the pan-thalassic group lack the anaerobic traits that are commonly distributed in other members of the Rhodospirillaceae, including those considered to be as separate taxonomic entities such as *Elstera* (Rahalkar et al., 2012). Thus, the pan-thalassic group shows a more pronounced aerobic physiology than the rest of the family, with the low affinity AOX oxidase more common than oxidases having high affinity for oxygen such as the bd oxidases (cf. (Degli Esposti et al., 2018; Degli Esposti et al., 2019).

The analysis of functional traits led to the development of an alternative approach for evaluating the taxonomic affiliation of alphaproteobacterial taxa on the basis of the distribution of Blast hits obtained with 120 functional proteins as described in Section 2.4. This approach produced very interesting results, which are summarized in Tables 2 and 3. The first result of note is that the frequency of top hits is strongly correlated to the phylogenetic closeness of taxa. For example, high frequency values of Rhodospirillaceae isolate NORP116 and NORP128 in the blast analysis of *Magnetovibrio* proteins (Supplementary Table S2) strongly correlate with the clustering of these taxa in phylogenetic trees (Figs. 3 and 4). *Terasakiella* sp. PR1, now renamed *Terasakiella magnetica* (Monteil et al., 2018), shows a statistically significant high frequency of top hits with members of the *Magnetospirillum* clade (Table 2), in clear agreement with phylogenetic trees (Fig. 3A, cf. (Monteil et al., 2018). Similar statistically significant values for the *Magnetospirillum* clade were obtained with two MAGs we have previously analysed, alphaproteobacterium RIFOXYD12_Full_60_8 and *Acetobacter* sp. CAG:977 (Table 2 cf. Fig. 1). These anaerobic Rhodospirillaceae can therefore be firmly assigned to the *Magnetospirillum* clade, in agreement with some phylogenetic trees (Figs. 3B and 5). The same applies to *Roseospirillum* (Table 2), a taxon currently classified within the Rhodobiaceae family of Rhizobiales (Glaeser and Overmann, 1999), but displaying anaerobic functional traits equivalent to those of *R.rubrum* (Degli Esposti et al., 2018; Degli Esposti et al., 2019).

Conversely, alphaproteobacterium GWF2_58_20 shows statistically significant high frequency of top hits with members of the *Azospirillum* clade (Table 2), again in agreement with phylogenetic trees (Fig. 1). Therefore, this anaerobic organism can be confidently assigned to the *Azospirillum* clade of the Rhodospirillaceae family. Such a clade does not include the anaerobic Rhodospirillaceae currently classified as *Azospirillum* sp. because the top hits for their proteins predominantly include homologs of closely related anaerobic MAGs that have been originally classified as *Acetobacter* sp. and also as Proteobacteria bacterium sp. CAG:495 (Nielsen et al., 2014; Brown et al., 2016) (Table 2 and results not shown). When these hits are removed from the results, as in the case of *Azospirillum* sp. CAG:239 and *Acetobacter* sp. 46_36 in Table 2, the *Magnetospirillum* clade shows a statistically significant enrichment in hit frequency compared with other subdivisions of the Rhodospirillaceae (Table 2). It is thus concluded that anaerobic Rhodospirillaceae formsa tight phylogenetic cluster that is associated with the *Magnetospirillum* clade via an ancestral branch containing *Acetobacter* sp. CAG:977, as indicated by phylogenetic trees of ATP synthase subunits (Figs. 3B, 5 and also Supplementary Fig. S1B and graphical summary). We propose to rename these anaerobic bacteria as strains of a new genus, Candidatus *Anaerospirillum* gen.nov. (Table 4).

The recent publication on the genome of thermophilic MAGs has reported that two previously unclassified bacteria, bacterium HR39 and Bacterium HR40, are phylogenetically close to *Tistrella* and *Geminicoccus* (Kato et al., 2018). We examined the nearly complete genome of bacterium HR40 and found that the distribution of Blast hits of its functional proteins strongly indicates a close affiliation to separate taxa of Rhodospirillaceae, as well as to *Geminicoccus* and *Arboriscoccus* (data not shown). We therefore propose that bacterium HR40 can be considered another member of the Geminicoccaceae family of Rhodospirillales (Table 4).

**Table 4**
Proposed re-classification of the Rhodospirillaceae family and its parent order. The initial part of the Table shows the current classification of Rhodospirillaceae as listed in the NCBI taxonomy website, https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id = 41295 accessed 9 April 2019. The bottom part shows the new classification we propose in this work. New proposed taxa are in red. See text and Fig. 7 for further details.

| **Current taxonomy in NCBI website** | | | |
|---|---|---|---|
| **Order:** | **Other subdivision** | **Family:** | ***Genera:*** |
| Rhodospirillales | | Rhodospirillaceae | all those in Table 1 |
| | | Geminicoccaceae | *Geminicocus, Arboriscoccus* |
| | | Acetobacteraceae | *Acetobacter*, *Gluconobacter, Roseomonas, Belnapia, Acidiphilium* etc. |
| **New proposed taxonomy** | | | |
| **Order:** | **Other subdivision** | **Family:** | ***Genera:*** |
| **Diaforabacterales** | | anaerobic Rhodospirillaceae | *Azospirillum* sp. CAG:239 & sp. CAG:260; *Azospirillum* sp. 51_20 & 47_25; *Acetobacter* (Ca. ***Anaerospirillum***) sp. CAG: 977 &CAG:267; *Acetobacter* (Ca. ***Anaerospirillum***) sp. 46_36 ; Proteobacteria bacterium (Ca. ***Anaerospirillum***) sp. CAG:495 |
| | *Magnetospirillum* clade | photosynthetic Rhodospirillaceae | *Rhodospirillum rubrum, Pararhodospirillum, Phaerospirillum, Rhodospira, Caenispirillum, Haematospirillum, Insolitospirillum, Novispirillum, Rhodospirillaceae* bacterium BRH_c57 |
| | | **Magnetospirillaceae** | *Magnetospirillum, Magnetovibrio, Rhodospirillaceae* bacterium NORP116, *Rhodospirillaceae* bacterium NORP28, *Varunaivibrio, Magnetospira, Terasakiella Thalassospira Roseospirillum* |
| | *Azospirillum* clade | | *Azospirillum, Nitrospirillum, Inquilinus, Rhodospirillum centenum (Rhodocista), Indioceanicola, Niveispirillum, Skermanella, Micavibrio, Reyranella,* Alphaproteobacteria bacterium GWF2_58_20 |
| | pan-thalassic group | | Ca. *Endolissoclinum,* SAR116 cluster, Ca. *Puniceispirillum, Nisaea,* alpha proteobacteria BAL199, *Oceanibaculum, Thalassobaculum, Fodinicurvata, Rhodovibrio, Tistlia, Limimonas, Kiloniella,* Rhodospirillales bacterium 47_12_T64, |
| | separate group | | Alphaproteobacteria bacterium 41–28 |
| | separate group | | *Dongia* |
| | separate group | | *Ferrovibrio* |
| | separate group | | *Enhydrobacter* |
| | separate group | | Rhodospirillales bacteriusm URHD008 |
| | separate group | | *Elstera* |
| | separate group | | *Tistrella* |
| | | Geminicoccaceae | *Geminicocus, Arboriscoccus*, Bacterium HR40 |
| | | Acetobacteraceae | *Acetobacter*, *Gluconobacter, Roseomonas, Belnapia, Acidiphilium* etc. |

### 3.5. Exploring the roots of the Rhodospirillaceae family

The results obtained with the approach of quantitative analysis of top Blast hits (Table 2) prompted us to explore the phylogenetic boundaries of the Rhodospirillaceae family. In particular, we investigated the genome properties of unclassified alphaproteobacteria and Proteobacteria that frequently lie in deep branches between the most ancestral group of alphaproteobacteria, the Magnetococcales (Degli Esposti et al., 2018), and the branches with recognized Rhodospirillaceae. Such taxa included members of deep-branching clades of marine MAGs such as MarineProteo1_Bin1 (Martijn et al., 2018) and also Alphaproteobacteria bacterium CG_4_10_14_0_8_um_filter_53_9, which has been recently found in a groundwater metagenome (Probst et al., 2018) (Figs. 2 and 3). Given that the genome of the latter MAG is nearly complete (Probst et al., 2018), while that of MarineProteo1_Bin1, albeit far from complete, can be complemented with proteins of its close relative MarineProteo1_Bin2 (Martijn et al., 2018), we applied the approach of Blast hit analysis of 120 functional proteins to identify whether such taxa may be related to Rhodospirillaceae. Expecting a strong degree of genomic chimaerism as in Magnetococcales (Esser et al., 2004; Ji et al., 2017), we compared the results of this analysis with those obtained with two prototypic species of Magnetococcales, *Magnetococcus marinus* and *Magnatofaba australis* (Lin et al., 2018) (Table 3).

In agreement with previous studies, approximately one third of the top hits for Magnetococcales proteins are with alphaproteobacterial homologs (Degli Esposti et al., 2018; Esser et al., 2004; Ji et al., 2017), while Rhodospirillaceae contribute more hits than any other family of the alphaproteobacteria class (Table 3 and results not shown). However, the situation is different for the proteins of MarineProteo1 and Alphaproteobacteria bacterium CG_4_10_14_0_8_um_filter_53_9, which have many top hits with homologs from Rhizobiales, Rhodobacterales and unclassified alphaproteobacteria (Table 3 and results not shown). The frequency for these hits is statistically different from those of Magnetococcales but comparable with those found for the proteins of Alphaproteobacteria bacterium 41–28, which lies in deep branches of trees obtained with single protein markers (Figs. 2, 6 and also Supplementary Fig. S5). Therefore, the relative proportion of top hits with Rhodospirillaceae vs. those with other alphaproteobacterial families, or unclassified taxa, provides a quantitative signature to distinguish deep branching alphaproteobacteria. The largest divarication in this relative proportion is shown by members of the Holosporaceae family such as endosymbiont of *Acanthamoeba* sp. UWC8, the proteins of which show ca. 50% of top hits with those of the Rickettsiales and Holosporales orders (Table 3, right). Intriguingly, Alphaproteobacteria bacterium 40–19 shows a similar taxonomic distribution of top hits, indicating that it may be closely affiliated to the Holosporaceae family, while Alphaproteobacteria bacterium 41–28 has intermediate values of top hits with Holosporaceae (Table 3). Nevertheless, previously unclassified Alphaproteobacteria bacterium 41–28 can be considered to form another separate sub-division of the Rhodospirillaceae, together with its close relative Alphaproteobacteria bacterium RIFCSPHIGHO2_12_42_13 (Fig. 6 and Table 3).

On the opposite boundary of deep branching alphaproteobacteria, we found that uncultured alpha proteobacterium isolate 86 shows a taxonomic distribution of top hits that is very similar to that of *Magnetovibrio* (Table 3). Alpha proteobacterium isolate 86 was chosen for this comparative analysis because it shares an archaic molecular feature with several other taxa presented in Table 3, namely the fusion of cytochrome *c*1 to the C-terminus of the cytochrome *b* protein (Fig. 6). This rare fusion was initially discovered in Bradyrhizobiaceae (Degli Esposti, 2014) and is present also in other Rhizobiales such as

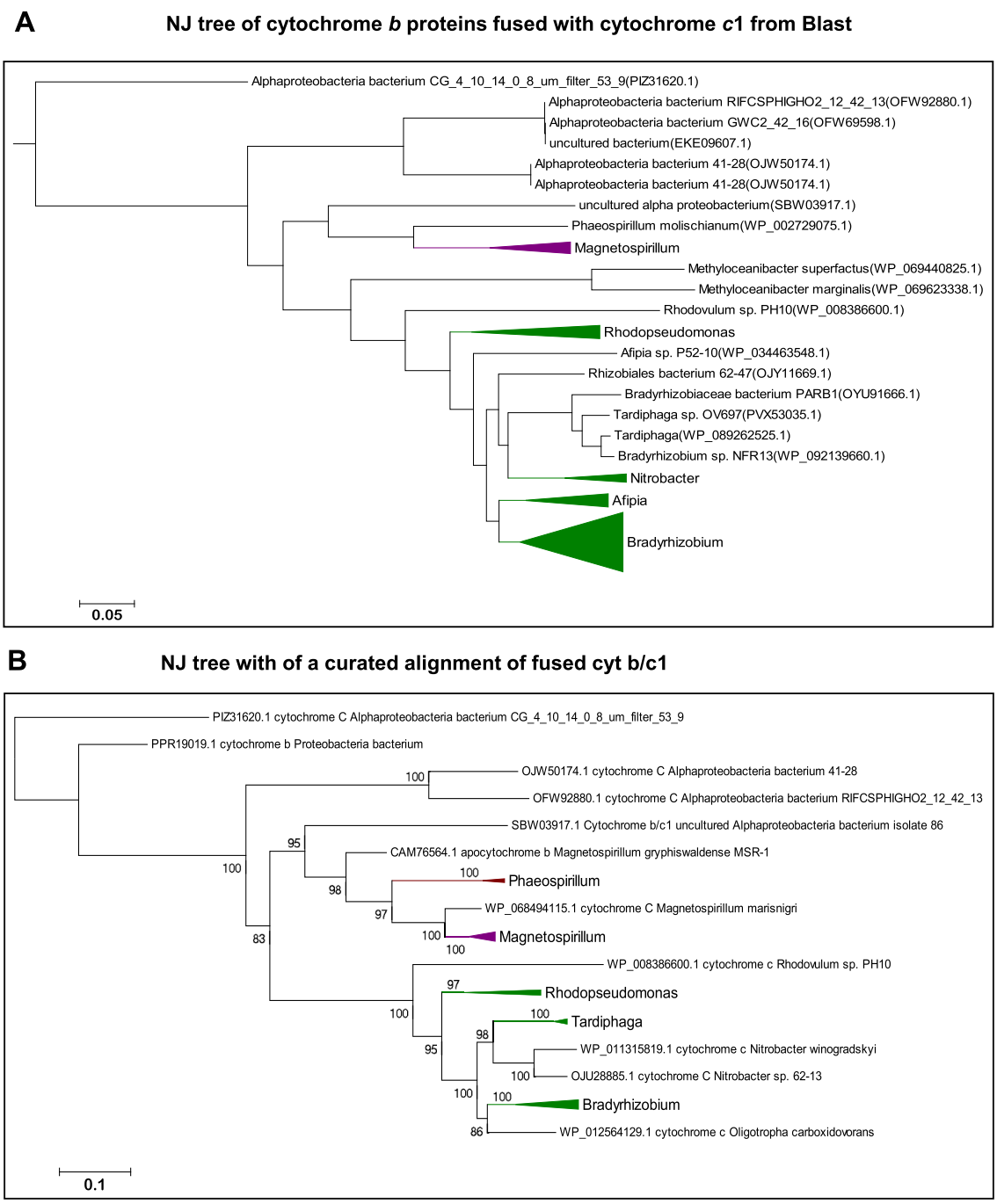



**Fig. 6.** Phylogenetic trees of the ancestral form of cytochrome *b* fused with cytochrome *c*1 The layout of the NJ trees is in descending order of phylogenetic distance as in Fig. 2. **A.** The tree was obtained with a DeltaBLAST search extended to all classified and unclassified alphaproteobacteria. **B.** The tree was obtained with the MEGA5 program and 200 bootstraps from the manually curated alignment of a selection of fused cytochrome *b*/c1 proteins in A, as described in the legend of Fig. 3A. Note the strong support values for most of the internal nodes. Similar results were obtained using the ML approach. Analysis of tree topology configuration has confirmed the close taxonomic relationship of Alphaproteobacteria bacterium 41–28 with Alphaproteobacteria bacterium RIFCSPHIGHO2_12_42_13 (cf. Table 2).

*Methyloceanibacter* (Fig. 6A). Among Rhodospirillaceae, it is present only in taxa belonging to the *Magnetospirillum* clade, which now includes also alpha proteobacterium isolate 86 on the basis of its top hits distribution (Table 3).

Phylogenetic trees of the fused b/c1 protein are monophyletic and show much stronger bootstrap support for their internal nodes than those of the cytochrome *b* protein alone (Fig. 6B cf. Fig. 3A), even after concatenation with the separate cytochrome *c*1 protein (Supplementary Fig. S1A). Consequently, these trees can provide valuable phylogenetic information for a fundamental lineage of Rhodospirillaceae that has retained the archaic fused b/c1 protein (Degli Esposti, 2014). Alphaproteobacteria bacterium CG_4_10_14_0_8_um_filter_53_9 can be considered the most ancestral bacterium of such a lineage, since it possesses the deepest branching b/c1 protein among alphaproteobacteria (Fig. 6 and data not shown). Its closest homolog is the cytochrome *b* of MarineProteo1_Bin 1, despite lacking the C-terminal fusion of cytochrome *c*1 (Fig. 6B). The next branch contains the fused b/c1 protein of alphaproteobacteria bacterium 41–28 and its close relative Alphaproteobacteria bacterium RIFCSPHIGHO2_12_42_13, which apparently has multiple names in NCBI resources (Fig. 6). These bacteria appear to lie in an intermediate taxonomic position between Rhodospirillaceae and Holosporaceae, as indicated by the taxonomic

## Rhodospirillales taxonomy

| order NCBI | family NCBI | order GTDB | family GTDB | NCBI currently defined taxa and number of genomes | this work |
|---|---|---|---|---|---|
| | | | | | |
| Rhodospirillales | Rhodospirillaceae | Rhodospirillales | Rhodospirillaceae | **Rhodospirillum rubrum** 2 | **phosynthetic Rhodospirillaceae** |
| | | | | Pararhodospirillum photometricum 2 | |
| | | | | Phaeospirillum molishianum & fulvum | |
| | | | | Caenispirillum salinarum & bisanense | |
| | | | | Novispirillum itersonii, Haematospirillum and Insolitispirillum | |
| | | | | Rhodospira trueperi | |
| | | | | Rhodospirillaceae bacterium BRH_c57 | |
| Rhodospirillales | Rhodospirillaceae | Rhodospirillales | Magnetospirillaceae | **Magnetospirillum magnetotacticum** and other 11 Magnetospirilli | **Magnetospirillaceae Magnetospirillum clade** |
| | | | | Telmatospirillum siberiense | |
| | | | Magnetovibrionaceae | **Magnetovibrio blackemori** & 9 Magnetovibrio sp. | |
| | | | | Rhodospirillaceae bacterium isolate NORP116 & 128 | |
| | | | Magnetospiraceae | **Magnetospira sp. QH-2** | |
| unclassified Alpha. | | **Rhodospirillales** | Rhodospirillaceae | Alphaproteobacteria bacterium RIFOXYD12_FULL_60_8 | |
| Rhodospirillales | Rhodospirillaceae | | Thalassospiraceae | **Thalassospira xiamenensis** & other 47 Thalassospira taxa | |
| Rhizobiales | Rhodobiaceae | **Rhodospirillales** | Rhodospirillaceae | **Roseospirillum parvum** | |
| Rhizobiales | Methylocystaceae | **Rhodospirillales** | Terasakiellaceae | **Terasakiella pusilla** and magnetica (sp. PR1) | |
| Rhodospirillales | Rhodospirillaceae | **RNF32** | CAG-239 | **Azospirillum sp. CAG:239** | **anaerobic Rhodospirillaceae Magnetospirillum clade** |
| | | | | Azospirillum sp. CAG:260 & sp. isolate CIM:MAG 289 | |
| | | | | Azospirillum sp. 51_20 & sp. 47_25 | |
| Rhodospirillales | Acetobacteraceae | **RNF32** | CAG-239 | Acetobacter sp. CAG:267 & Acetobacter sp. 46_36 | |
| | Acetobacteraceae | **RNF32** | CAG-977 | **Acetobacter sp. CAG:977** | |
| unclassified Proteo. | | **RNF32** | CAG-239 | **Proteobacteria bacterium CAG:495** | |
| Rhodospirillales | Rhodospirillaceae | **Azospirillales** | Azospirillaceae | **Azospirillum lipoferum** | **Azospirillum clade** |
| | | | | Azospirillum brasilense 4 strains | |
| | | | | Azospirillum sp. B510, B4 & B506 | |
| | | | | Azospirillum halopraeferens | |
| | | | | Azospirillum thiophilum, humicireducens & oryzae | |
| | | | | Azospirillum sp. RU37A & RU38E | |
| | | | | Skermanella stibiiresistens & aerolata | |
| | | | | Nitrospirillum amazonense | |
| | | | | Niveispirillum irakense | |
| | | | | Rhodospirillum centenum SW & Indioceanicola | |
| | | | Inquilinaceae | **Inquilinus limosus** DSM 16000 & 2 other strains | |
| Rhodospirillales | unclassified | **Reyranellales** | Reyranellaceae | **Reyranella massiliensis** | |
| | | | | Rhodospirillaceae bacterium CCH5-H10 | |
| unclassified Alpha. | | **Micavibrionales** | Micavibrionaceae | **Micavibrio aeruginosavorus** ARL-13 & EPB | |
| | | | | Micavibrio sp. TMED27 & other 8 Micavibrio sp. | |
| Rhodospirillales | Rhodospirillaceae | **TMED26** | TMED26 | **Candidatus Endolissoclinum sp. TMED26** | **pan-thalassic group** |
| | | | | Nisaea sp. 2 | |
| unclassified Alpha. | | **Puniceispirillales** | Puniceispirillaceae | **SAR116 cluster alpha proteobacterium HIMB100** | |
| unclassified Alpha. | | | | SAR116 cluster bacterium REDSEA-S10_B10N8 | |
| unclassified Alpha. | | | | SAR116 cluster bacterium 14 strains | |
| unclassified Alpha. | | | | Candidatus Puniceispirillum marinum IMCC1322 | |
| unclassified Alpha. | | | | Candidatus Puniceispirillum sp. TMED213 | |
| Rhodospirillales | Rhodospirillaceae | **Thalassobaculales** | Thalassobaculaceae | **Thalassobaculum salexigens** & litoreum | |
| | | | | Candidatus Endolissoclinum faulkneri L2 & L5 | |
| unclassified Alpha. | | **Thalassobaculales** | Thalassobaculaceae | alpha proteobacterium BAL199 | |
| Rhodospirillales | Rhodospirillaceae | **Oceanibaculales** | Oceanibaculaceae | **Oceanibaculum** 3 species | |
| Kiloniellales | Kiloniellaceae | **Kiloniellales** | Kiloniellaceae | **Kiloniella laminariae** & other 5 Kiloniella strains | |
| Rhodospirillales | | **Kiloniellales** | Kiloniellaceae | Rhodospirillales bacterium 47_12_T64 | |
| Rhodospirillales | Rhodospirillaceae | **Kiloniellales** | Fodinicurvataceae | **Fodinicurvata sediminis** & fenggangensis DSM 21160 | |
| | | | | Tistlia consotensis | |
| | | | | Rhodovibrio salinarum | |
| | | | | Limimonas halophila | |
| Rhodospirillales | Rhodospirillaceae | **Dongiales** | Dongiaceae | **Dongia sp. URHE0060** | separate |
| | | **Sneathiellales** | Ferrovibrionaceae | **Ferrovibrio sp.** | separate |
| | | **Tistrellales** | Tistrellaceae | **Tistrella mobilis** & 4 Tistrella sp. | separate |
| | | **Elsterales** | Elsteraceae | **Elstera litoralis** & cyanobacteriorum | separate |
| Rhodospirillales | Geminicoccaceae | **Geminicoccales** | Geminicoccaceae | **Geminicoccus roseus** DSM 18922 | separate |
| | | | | **Arboriscoccus pini** or Rhodospirillales bacterium B29T1 | separate |
| unclassified Bacteria | | not present | not present | **Bacterium HR40** | Geminicoccaceae, separate |
| Rhodospirillales | unclassified | **Pseudomonadales** | Moraxellaceae | **Enhydrobacter aerosaccus** | separate |
| unclassified Alpha. | | **UBA9655** | UBA9655 | **Alphaproteobacteria bacterium 41-28** | separate, intermediate ancestral |
| | | | | Alphaproteobacteria bacterium RIFCSPHIGHO2_12_42_13 | separate, intermediate ancestral |
| unclassified Alpha. | | **UBA1280** | UBA6156 | Alphaproteobacteria bacterium CG_4_10_14_0_8_um_filter_53_9 | **ancestral alphaproteobacteria** |
| Rhodospirillales | Acetobacteraceae | **Acetobacterales** | Acetobacteraceae | **Acetobacter** | **Acetobacteraceae** |
| | | | | Elioraea, Roseomonas, Belnapia | |
| | | | | Acidiphilium, Acidocella & many other genera | |
| Rhodospirillales | | | | Rhodospirillales bacterium URHD0017 | ? |

**Fig. 7.** Comparison of the diverse taxonomic subdivisions proposed for Rhodospirillales. The current classification available in the NCBI website is compared for 230 different taxa encompassing the great majority of the bacterial genomes that are listed under the family of Rhodospirillaceae and the parent order of Rhodospirillales. The central columns list the new orders and families introduced in the GTDB taxonomy (Parks et al., 2018) for the indicated taxa. These taxonomic subdivisions were exported from the early version of the GTDB website (http://gtdb.ecogenomic.org/, accessed 30 August 2018) and are highlighted in pale yellow when they differ from the established NCBI taxonomy (https://www.ncbi.nlm.nih.gov/taxonomy, accesses on 30 August 2018). The rightmost column lists the various subdivisions which are introduced and discussed in this work (Table 4). Question marks indicate taxa for which we do not have enough data for evaluating their placement in the subdivisions introduced here. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

distribution of top Blast hits (Table 3) and some phylogenetic trees (Supplementary Fig. S5 and data not shown). In the tree of fused b/c1 proteins, their branch is followed by that containing alpha proteobacterium isolate 86 and various *Magnetospirillum* taxa (Fig. 6), which represent *bona fide* Rhodospirillaceae (Table 2). Fused b/c1 proteins are common in Nitrospirae but very rare in other classes of proteobacteria (Degli Esposti et al., 2018; Kato et al., 2018), thereby indicating that their bicistronic gene has followed a scattered pattern of gene split from Nitrospirae to proteobacteria, while maintaining a monophyletic ascendance in alphaproteobacteria (Fig. 6). We have identified here the earliest alphaproteobacteria of such ancestral lineage, namely Alphaproteobacteria bacterium CG_4_10_14_0_8_um_filter_53_9 and MarineProteo1 MAGs (Fig. 6 and Table 3). The latter taxa have been previously considered to represent a branch of proteobacteria diverging earlier than alphaproteobacteria, from which proto-mitochondria might have originated (Martijn et al., 2018). Our results suggest instead that this and related MAGs are part of an ancestral branch of alphaproteobacteria (Table 3 and Figs. 2, 3 and 6), thereby confirming the concept of alphaproteobacterial origin of proto-mitochondria (Degli Esposti et al., 2018).

## 4. Conclusions

This work considerably expands the phylogenetic diversity of the family Rhodospirillaceae, which is clearly polyphyletic when analysed in depth. Organisms currently classified among Rhodospirillaceae belong to two major clades pivoting on *Magnetospirillum* and *Azospirillum*, plus another group labelled pan-thalassic (Table 4). In addition to these major groupings, Rhodospirillaceae include a set of separate taxonomic entities that may require higher taxonomic ranks than the genus level into which they are currently classified. These include *Tistrella, Elstera, Dongia* and *Ferrovibrio*, plus a group of marine MAGs currently named MarineAlpha9 and 10 (Martijn et al., 2018) (Supplementary Table S2; Table 4 and Fig. 7). Finally, the family of Rhodospirillaceae embraces taxa previously classified in orders different from the Rhodospirillales, in particular Kiloniellales and Rhizobiales (Table 1), but also unclassified alphaproteobacteria such as those of the SAR116 cluster and the *Micavibrio* genus (Fig. 5 and Table 4).

In sum, the family Rhodospirillaceae has a taxonomic richness much larger than previously considered, which would be typical for a taxonomic rank higher than a family. We propose here to re-classify all the previously established members of the Rhodospirilaceae family, plus those affiliated with the same family that we have reported here, under a new order of alphaproteobacteria called **Diaforabacterales** ord. nov. (from the Greek word for diversity, διάφορα). This order would replace the current order of Rhodospirillales, which was originally established for purple photosynthetic bacteria such as the type genus *Rhodospirillum* (Garrity et al., 2005), nowadays representing only a small part of the phylogenetic diversity of the order (Tables 1 and 4). These purple bacteria would remain classified in the Rhodospirillaceae family containing only photosynthetic organisms, re-defined as 'photosynthetic Rhodospirillaceae' (Table 4). The newly introduced genus of Ca. *Anaerospirillum* would instead be classified in a separate family, that of 'anaerobic Rhodospirillaceae' – a term we have introduced here to identify anaerobic MAGs (Nielsen et al., 2014). This family is sister to the proposed new family of Magnetospirillaceae fam.nov. (cf. (Parks et al., 2018), which contains all the magnetotactic members of the Rhodospirillaceae plus *Terasakiella* and some marine MAGs (Table 4). We propose that the new order of Diaforabacterales includes other taxa of the current Rhodospirillaceae family under the subdivision of the *Azospirillum* clade and the pan-thalassic group, or as separate groups, as shown in Table 4. Some of such separate groups might form new families once more genomes will become available. Finally, Diaforabacterales would incorporate the families of Geminicoccaceae and Acetobacteraceae as for the current order of Rhodospirillales (Table 4).

Our proposed re-classification of Rhodospirillales appears to be congruent with the new standardized bacterial taxonomy GTDB (Parks et al., 2018), which was published online while we were completing this work. The GTDB taxonomy (Parks et al., 2018) (http://gtdb.ecogenomic.org/, accessed 30 September 2018) distributes the organisms currently classified in the Rhodospirillaceae family among 15 new orders and at least 26 new families, as shown in Fig. 7. Interestingly, separate taxonomic entities we have uncovered here - namely *Tistrella, Elstera, Dongia* and *Ferrovibrio* - correspond to new orders and families in the GTDB taxonomy (Fig. 7). Moreover, the genera *Terasakiella* and *Roseospirillum* are associated with Rhodospirillales families and not with the Rhizobiales families in which they have been classified so far (Fig. 7), again in agreement with our phylogenetic analysis (Figs. 5 and 7, Table 4). However, the GTDB taxonomy still maintains Kiloniellales as a separate order from Rhodospirillales, with the addition of a new family, Fodinicurvataceae, which includes Rhodospirillaceae taxa that cluster with the pan-thalassic group (Figs. 5 and 7, Table 4). This appears to be, at present, the major discrepancy between our proposed reclassification of Rhodospirillaceae and Rhodospirillales (Table 4) and that deduced from the GTDB taxonomy (Parks et al., 2018). Further work is clearly needed to validate the new, largely expanded set of taxonomic subdivisions embracing the bacteria traditionally classified in the Rhodospirillaceae family. Taxa affiliated with this polyphyletic family may represent a crucial node in the taxonomic space from which early alphaproteobacteria, and consequently proto-mitochondria have evolved (Degli Esposti et al., 2019). Future studies will explore this possibility, taking advantage of the ever increasing number of MAGs that can fill this space (Parks et al., 2017; Degli Esposti et al., 2019; Martijn et al., 2018; Kato et al., 2018; Parks et al., 2018).

## Acknowledgements

This work was sponsored by Conacyt grant Basic Science No. 253116. We thank Michelle Degli Esposti for her valuable help in statistical analysis.

## Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106546.

## References

Garrity, G.M., Bell, J.A., Lilburn, T., 2005. Family I. Rhodospirillaceae. Bergey's Manual® of Systematic Bacteriology, Vol. 2C, pp. 1–20. Springer US. 1-574.

Imhoff, J.F., Petri, R., Süling, J., 1998. Reclassification of species of the spiral-shaped phototrophic purple non-sulfur bacteria of the alpha-Proteobacteria: description of the new genera Phaeospirillum gen. nov., Rhodovibrio gen. nov., Rhodothalassium gen. nov. and Roseospira gen. nov. as well as transfer of Rhodospirillum fulvum to Phaeospirillum fulvum comb. nov., of Rhodospirillum molischianum to Phaeospirillum molischianum comb. nov., of Rhodospirillum salinarum to Rhodovibrio salexigens. Int. J. Syst. Bacteriol. 48, 793–798.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., et al., 2014. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. Nat. Biotechnol. 32, 822–828.

Brown, C.T., Olm, M.R., Thomas, B.C., Banfield, J.F., 2016. Measurement of bacterial replication rates in microbial communities. Nat. Biotechnol. 34, 1256–1263.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., et al., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 13219.

Probst, A.J., Castelle, C.J., Singh, A., Brown, C.T., Anantharaman, K., Sharon, I., et al., 2017. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO(2) concentrations. Environ. Microbiol. 19, 459–474.

Tully, B.J., Sachdeva, R., Graham, E.D., Heidelberg, J.F., 2017. 290 metagenome-assembled genomes from the Mediterranean Sea: a resource for marine microbiology. PeerJ. 5, e3558.

Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N., et al., 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat. Microbiol. 2, 1533–1542.

Wisniewski-Dyé, F., Borziak, K., Khalsa-Moyers, G., Alexandre, G., Sukharnikov, L.O., Wuichet, K., et al., 2011. Azospirillum genomes reveal transition of bacteria from aquatic to terrestrial environments. PLoS Genet. 7, e1002430.

Kwan, J.C., Donia, M.S., Han, A.W., Hirose, E., Haygood, M.G., Schmidt, E.W., 2012.

Genome streamlining and chemical defense in a coral reef symbiosis. Proc. Natl. Acad. Sci. USA 109, 20655–20660.
Degli Esposti, M., Cortez, D., Lozano, L., Rasmussen, S., Nielsen, H.B., Martínez-Romero, E., 2016. Alpha proteobacterial ancestry of the [Fe-Fe]-hydrogenases in anaerobic eukaryotes. Biol. Direct. 11, 34.
Probst, A.J., Ladd, B., Jarett, J.K., Geller-McGrath, D.E., Sieber, C.M.K., Emerson, J.B., Anantharaman, K., et al., 2018. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol. 3, 328–336.
Degli Esposti, M., Martinez-Romero, E., 2017. The functional microbiome of arthropods. PLoS One 12, e0176573.
Degli Esposti, M., Martínez, J., Martínez-Romero, E., 2018. From Alphaproteobacteria to Proto-Mitochondria. In: Degli Esposti, M. (Ed.), Phylogeny and evolution of bacteria and mitochondria. CRC Press Taylor & Francis Group, Boca Raton, FL, pp. 166–203.
Louca, S., Parfrey, L.W., Doebeli, M., 2016. Decoupling function and taxonomy in the global ocean microbiome. Science. 353, 1272–1277.
Degli Esposti, M., 2017. A journey across genomes uncovers the origin of ubiquinone in cyanobacteria. Genome Biol. Evol. 9, 3039–3053.
Degli Esposti, M., Mentel, M., Martin, W.F., Sousa, F.L., 2019. Oxygen reductases in alphaproteobacterial genomes: Physiological evolution from low to high oxygen environments. Frontiers in Microbiology. 10, 499.
Müller, M., Mentel, M., van Hellemond, J.J., Henze, K., Woehle, C., Gould, S.B., et al., 2012. Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. Microbiol. Mol. Biol. Rev. 76, 444–495.
Trubitsyn, D., Abreu, F., Ward, F.B., Taylor, T., Hattori, M., Kondo, S., et al., 2016. Draft genome sequence of Magnetovibrio blakemorei Strain MV-1, a marine vibrioid magnetotactic bacterium. Genome Announc. 4 (6) pii: e01330–16.
Esser, C., Ahmadinejad, N., Wiegand, C., Rotte, C., Sebastiani, F., Gelius-Dietrich, G., et al., 2004. A genome phylogeny for mitochondria among alpha-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. Mol. Biol. Evol. 21, 1643–1660.
Atteia, A., Adrait, A., Brugière, S., Tardif, M., van Lis, R., Deusch, O., et al., 2009. A proteomic survey of *Chlamydomonas reinhardtii* mitochondria sheds new light on the metabolic plasticity of the organelle and on the nature of the alpha-proteobacterial mitochondrial ancestor. Mol. Biol. Evol. 26, 1533–1548.
Abhishek, A., Bavishi, A., Bavishi, A., Choudhary, M., 2011. Bacterial genome chimaerism and the origin of mitochondria. Can. J. Microbiol. 57, 49–61.
Thiergart, T., Landan, G., Schenk, M., Dagan, T., Martin, W.F., 2012. An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. Genome Biol. Evol. 4, 466–485.
Martijn, J., Vosseberg, J., Guy, L., Offre, P., Ettema, T.J.G., 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. Nature. 557, 101–105.
Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.
Segata, N., Börnigen, D., Morgan, X.C., Huttenhower, C., 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nat. Commun. 4, 2304.
Schulz, F., Eloe-Fadrosh, E.A., Bowers, R.M., Jarett, J., Nielsen, T., Ivanova, N.N., et al., 2017. Towards a balanced view of the bacterial tree of life. Microbiome. 5, 140. https://doi.org/10.1186/s40168-017-0360-9.
Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. Science. 311, 1283–1287.
Imhoff, J.F., 2016. New Dimensions in Microbial Ecology-Functional Genes in Studies to Unravel the Biodiversity and Role of Functional Microbial Groups in the Environment. Microorganisms. 4. pii: E19. http://doi.org/10.3390/microorganisms4020019.
Martin, W.F., 2018. Early Microbial Evolution from a Physiological Perspective. In: Degli Esposti, M. (Ed.), Phylogeny and evolution of bacteria and mitochondria. CRC Press Taylor & Francis Group, Boca Raton, FL, pp. 1–25.
Degli Esposti, M., 2016. Late Mitochondrial Acquisition, Really? Genome Biol. Evol. 8, 2031–2035.
Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., Madden, T.L., 2012. Domain enhanced lookup time accelerated BLAST. Biol Direct. 7, 12.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.
Degli Esposti, M., Chouaia, B., Comandatore, F., Crotti, E., Sassera, D., Lievens, P.M., et al., 2014. Evolution of mitochondria reconstructed from the energy metabolism of living bacteria. PLoS One 9, e96566.
Thiergart, T., Landan, G., Martin, W.F., 2014. Concatenated alignments and the case of the disappearing tree. BMC Evol. Biol. 14, 266.
Rochette, N.C., Brochier-Armanet, C., Gouy, M., 2014. Phylogenomic test of the hypotheses for the evolutionary origin of eukaryotes. Mol. Biol. Evol. 31, 832–845.
Williams, K.P., Sobral, B.W., Dickerman, A.W., 2007. A robust species tree for the Alphaproteobacteria. J. Bacteriol. 189, 4578–4586.
Core Team, R., 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Blanc, B., Gerez, C., Ollagnier de Choudens, S., 2015. Assembly of Fe/S proteins in bacterial systems: Biochemistry of the bacterial ISC system. Biochim. Biophys. Acta. 1853, 1436–1447.
Adams, D.W., Errington, J., 2009. Bacterial cell division: assembly, maintenance and disassembly of the Z ring. Nat. Rev. Microbiol. 7, 642–653.
Konovalova, A., Kahne, D.E., Silhavy, T.J., 2017. Outer Membrane Biogenesis. Annu. Rev. Microbiol. 71, 539–556.
Esser, C., Martin, W.F., Dagan, T., 2007. The origin of mitochondria in light of a fluid prokaryotic chromosome model. Biol. Lett. 3, 180–184.
Ji, B., Zhang, S.D., Zhang, W.J., Rouy, Z., Alberto, F., Santini, C.L., et al., 2017. The chimeric nature of the genomes of marine magnetotactic coccoid-ovoid bacteria defines a novel group of Proteobacteria. Environ. Microbiol. 19, 1103–1119.
Richter, M., Rossello-Mora, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. USA 106, 19126–19131.
Proença, D.N., Whitman, W.B., Varghese, N., Shapiro, N., Woyke, T., Kyrpides, N.C., Morais, P.V., 2018. Arboriscoccus pini gen. nov., sp. nov., an endophyte from a pine tree of the class Alphaproteobacteria, emended description of Geminicoccus roseus, and proposal of Geminicoccaceae fam. nov. Syst. Appl. Microbiol. 41, 94–100.
Davidov, Y., Huchon, D., Koval, S.F., Jurkevitch, E., 2006. A new alpha-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. Environ. Microbiol. 8, 2179–2188.
Tully, B.J., Wheat, C.G., Glazer, B.T., Huber, J.A., 2018. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J. 12, 1–16.
Oh, H.M., Kwon, K.K., Kang, I., Kang, S.G., Lee, J.H., Kim, S.J., et al., 2010. Complete genome sequence of "Candidatus Puniceispirillum marinum" IMCC1322, a representative of the SAR116 clade in the Alphaproteobacteria. J. Bacteriol. 192, 3240–3241.
Imhoff, J.F., Wiese, J., 2014. The order Kiloniellales. In: The Prokaryotes. Springer, Berlin, pp. 301–306.
Foesel, B.U., Gössner, A.S., Drake, H.L., Schramm, A., 2007. Geminicoccus roseus gen. nov., sp. nov., an aerobic phototrophic Alphaproteobacterium isolated from a marine aquaculture biofilter. Syst. Appl. Microbiol. 30, 581–586.
Xu, Y., Kersten, R.D., Nam, S.J., Lu, L., Al-Suwailem, A.M., Zheng, H., et al., 2012. Bacterial biosynthesis and maturation of the didemnin anti-cancer agents. J. Am. Chem. Soc. 134, 8625–8632.
Han, S.B., Su, Y., Hu, J., Wang, R.J., Sun, C., Wu, D., et al., 2016. *Terasakiella brassicae sp. nov.*, isolated from the wastewater of a pickle-processing factory, and emended descriptions of *Terasakiella pusilla* and the genus *Terasakiella*. Int. J. Syst. Evol. Microbiol. 66, 1807–1812.
Monteil, C.L., Perrière, G., Menguy, N., Ginet, N., Alonso, B., Waisbord, N., et al., 2018. Genomic study of a novel magnetotactic Alphaproteobacteria uncovers the multiple ancestry of magnetotaxis. Environ. Microbiol. 20, 4415–4430.
Szokoli, F., Castelli, M., Sabaneyeva, E., Schrallhammer, M., Krenek, S., Doak, T.G., et al., 2016. Disentangling the taxonomy of rickettsiales and description of two novel symbionts ("Candidatus Bealeia paramacronuclearis" and "Candidatus Fokinia cryptica") sharing the cytoplasm of the ciliate protist Paramecium biaurelia. Appl. Environ. Microbiol. 82, 7236–7247.
Liu, Y., Jin, J.H., Liu, Y.H., Zhou, Y.G., Liu, Z.P., 2010. Dongia mobilis gen. nov., sp. nov., a new member of the family Rhodospirillaceae isolated from a sequencing batch reactor for treatment of malachite green effluent. Int. J. Syst. Evol. Microbiol. 60, 2780–2785.
Rahalkar, M., Bahulikar, R.A., Deutzmann, J.S., Kroth, P.G., Schink, B., 2012. Elstera litoralis gen. nov., sp. nov., isolated from stone biofilms of Lake Constance. Germany. Int. J. Syst. Evol. Microbiol. 62, 1750–1754.
Kawamura, Y., Fujiwara, N., Naka, T., Mitani, A., Kubota, H., Tomida, J., et al., 2012. Genus Enhydrobacter Staley et al. 1987 should be recognized as a member of the family Rhodospirillaceae within the class Alphaproteobacteria. Microbiol. Immunol. 56, 21–26.
Glaeser, J., Overmann, J., 1999. Selective enrichment and characterization of Roseospirillum parvum, gen. nov. and sp. nov., a new purple nonsulfur bacterium with unusual light absorption properties. Arch. Microbiol. 171, 405–416.
Kato, S., Sakai, S., Hirai, M., Tasumi, E., Nishizawa, M., Suzuki, K., Takai, K., 2018. Long-term cultivation and metagenomics reveal ecophysiology of previously uncultivated thermophiles involved in biogeochemical nitrogen cycle. Microbes Environ. 33, 107–110.
Lin, W., Zhang, W., Zhao, X., Roberts, A.P., Paterson, G.A., Bazylinski, D.A., Pan, Y., 2018. Genomic expansion of magnetotactic bacteria reveals an early common origin of magnetotaxis with lineage-specific evolution. ISME J. 12, 1508–1519.
Degli Esposti, M., 2014. Bioenergetic evolution in proteobacteria and mitochondria. Genome Biol. Evol 6, 3238–3251.
Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004.
Darrida, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest3: fast selection of best-fit models of protein evolution. Bioinformatics. 27, 1164–1165.
Tindall, B.J., 2001. Rhodocista centenaria vs Rhodospirillum centenum: a reply to Gest and Favinger. Int. J. Syst. Evol. Microbiol. 51, 711–713.
Kim, S.J., Ahn, J.H., Lee, T.H., Weon, H.Y., Hong, S.B., Seok, S.J., et al., 2013. Reyranella soli sp. nov., isolated from forest soil, and emended description of the genus Reyranella Pagnier et al. 2011. Int. J. Syst. Evol. Microbiol. 63, 3164–3167.
Park, S., Park, J.M., Kang, C.H., Yoon, J.H., 2014. Aestuariispira insulae gen. nov., sp. nov., a lipolytic bacterium isolated from a tidal flat. Int. J. Syst. Evol. Microbiol. 64, 1841–1846.
Chen, R.W., Wang, K.X., Zhou, X.F., Long, C., Tian, X.P., Long, L.J., 2018. Indioceanicola profundi gen. nov., sp. nov., isolated from Indian Ocean sediment. Int. J. Syst. Evol. Microbiol. 68, 3707–3712.
Patwardhan, S., Vetriani, C., 2016. Varunaivibrio sulfuroxidans gen. nov., sp. nov., a facultatively chemolithoautotrophic, mesophilic alphaproteobacterium from a shallow-water gas vent at Tor Caldara, Tyrrhenian Sea. Int. J. Syst. Evol. Microbiol. 66, 3579–3584.