Title: Refining the taxonomy of the order *Hyphomicrobiales* (*Rhizobiales*) based on whole genome comparisons of over 130 genus type strains

Authors: George C. diCenzo[1*], Yuqi Yang[1], J. Peter W. Young[2], and Nemanja Kuzmanović[3*]

Affiliations: [1] Department of Biology, Queen's University, Kingston, ON, K7P 0S7, Canada

[2] Department of Biology, University of York, York, YO10 5DD, United Kingdom

[3] Julius Kühn Institute (JKI) - Federal Research Centre for Cultivated Plants, Institute for Plant Protection in Horticulture and Urban Green, Braunschweig, 38104, Germany

*** Corresponding authors:** George diCenzo (george.dicenzo@queensu.ca) and Nemanja Kuzmanović (kuzmanovic1306@gmail.com; nemanja.kuzmanovic@julius-kuehn.de)

Keywords: *Hyphomicrobiales*, *Rhizobiales*, *Phyllobacteriaceae*, *Bartonellaceae*, family boundaries

ABSTRACT

The alphaproteobacterial order *Hyphomicrobiales* consists of 38 families comprising 155 validly published genera as of June 2023. The order *Hyphomicrobiales* was first described in 1957 and underwent important revisions in 2020. However, several inconsistencies in the taxonomy of this order remain, and there is a need for a consistent framework for defining families within the order. We propose a common genome-based framework for defining families within the order *Hyphomicrobiales*, suggesting that families represent monophyletic groups in core-genome phylogenies that share pairwise average amino acid identity values above ~75% when calculated from a core set of 59 proteins. Applying this framework, we propose the formation of four new families and to reassign the genera *Salaquimonas*, *Rhodoblastus*, and *Rhodoligotrophos* into *Salaquimonadaceae* fam. nov., *Rhodoblastaceae* fam. nov., and *Rhodoligotrophaceae* fam. nov., respectively, and the genera *Albibacter*, *Chenggangzhangella*, *Hansschlegelia*, and *Methylopila* into *Methylopilaceae* fam. nov.. We further propose to unify the families *Bartonellaceae*, *Brucellaceae*, *Phyllobacteriaceae*, and *Notoacmeibacteraceae* as *Bartonellaceae*; the families *Segnochrobactraceae* and *Pseudoxanthobacteraceae* as *Segnochrobactraceae*; the families *Lichenihabitantaceae* and *Lichenibacteriaceae* as *Lichenihabitantaceae*; and the families *Breoghaniaceae* and *Stappiaceae* as *Stappiaceae*. Lastly, we propose to reassign several genera to existing families. Specifically, we propose to reassign the genus *Pseudohoeflea* to the family *Rhizobiaceae*; the genera *Oricola*, *Roseitalea*, and *Oceaniradius* to the family *Ahrensiaceae*; the genus *Limoniibacter* to the emended family *Bartonellaceae*; the genus *Faunimonas* to the family *Afifellaceae*; and the genus *Pseudochelatococcus* to the family *Chelatococcaceae*. Our data also support the recent proposal to reassign the genus *Prosthecomicrobium* to the family *Kaistiaceae*.

INTRODUCTION

The order *Hyphomicrobiales* of the class *Alphaproteobacteria* consists of 38 families encompassing 155 valid genera (List of Prokaryotic names with Standing in Nomenclature; LPSN; accessed 25 June 2023). It was known by the name *Rhizobiales* Kuykendall 2006 [1] from 2005 until 2020, when Hördt and colleagues [2] pointed out that this was an illegitimate later synonym of *Hyphomicrobiales* Douglas 1957 (Approved List 1980) [3]. This order contains phenotypically diverse organisms, including terrestrial and aquatic bacteria, plant mutualists, and plant, animal, and human pathogens. Moreover, all known alpha-rhizobia (alphaproteobacterial nitrogen-fixing legume symbionts) are found within the order *Hyphomicrobiales* and are spread across seven families (*Rhizobiaceae*, *Phyllobacteriaceae*, *Brucellaceae*, *Nitrobacteriaceae*, *Methylobacteriaceae*, *Xanthobacteriaceae*, and *Devosiaceae*).

We recently proposed a genome-based framework for defining genera within the family *Rhizobiaceae* [4], which is the largest family within the order *Hyphomicrobiales* with 20 validly published genera as of 25 June 2023. While applying our framework to refine the family *Rhizobiaceae* [4], we wondered whether any genera currently assigned to the family *Rhizobiaceae* should be reassigned to novel families. This prompted us to evaluate family-level classifications within the order *Hyphomicrobiales*. Here, we propose a genome-based framework for defining families within the order *Hyphomicrobiales*. We then apply this framework and propose the unification of several families, the formation of four new families, and the reassignment of multiple genera.

METHODS

Datasets

Core-proteome phylogenies and overall genomic relatedness indices (OGRIs) were calculated using the genomes of 138 *Hyphomicrobiales* strains (**Dataset S1**). Where applicable, five *Caulobacterales* strains were included as an outgroup (**Dataset S2**). Most strains were genus type strains. The

61 exceptions were the species type strains "*Pararhizobium mangrovi*" BGMRC 6574^T, which was included
62 as preliminary analyses suggested that this strain may belong to a new genus potentially in a novel
63 family, and *Hyphomicrobium denitrificans* ATCC 51888^T, which was included as the genome sequence
64 of the *Hyphomicrobium* genus type strain was unavailable. Where noted, we included an additional
65 four species type strains from the genus *Bartonella* and the recently proposed *"Flavimaribacter*
66 *sediminis*" WL0058^T (**Dataset S3**). All genomes were downloaded from the National Center for
67 Biotechnology Information (NCBI) Genome database.

68 For construction of a 16S rRNA gene phylogeny, full-length 16S rRNA gene sequences were
69 extracted from the genome sequences of the 143 strains used in the genome-based analyses, where
70 available. If the genome did not contain a full length 16S rRNA gene sequence, the strain's 16S rRNA
71 gene sequence was downloaded from NCBI GenBank via links embedded in the LPSN database [5]. In
72 cases where more than one distinct 16S rRNA gene sequence was extracted from a whole genome
73 sequence, all unique sequences were kept. This dataset was supplemented with the 16S rRNA gene
74 sequences of nine genus type strains that lacked publicly available whole genome sequences but were
75 related to strains that were candidates for reclassification. These 16S rRNA gene sequences were
76 downloaded from the LPSN database (**Dataset S4**).

77

78 **Identification of core gene sets**

79 Initially, the GET_HOMOLOGUES software package version 05052023 [6] and GET_PHYLOMARKERS
80 software package version 2.4.5_17nov2022 [7] were used to identify non-recombining single-copy
81 marker genes present in all 143 genomes, as described previously [8]. This led to the identification of
82 19 marker genes, which we termed "core_143". Next, we used GET_HOMOLOGUES and custom scripts
83 to identify single-copy marker genes present in at least 95% of the target genomes; the stringent
84 filtering of GET_PHYLOMARKERS was not used. This led to the identification of 256 marker genes,

85 which we termed "perc95_143". During preliminary investigations of the perc95_143 gene set, we
86 observed that five strains (*Chenggangzhangella methanolivorans* CHL1^T, *Methylobrevis pamukkalensis*
87 VKM B-2849^T, *Nitratireductor aquibiodomus* JCM 21793^T, *Liberibacter crescens* BT-1^T, and *Methyloligella*
88 *halotolerans* VKM B-2706^T) lacked between 12% and 28% of these genes (see **Text S1**). We therefore
89 repeated the above analyses using a reduced set of 138 strains that excluded those five strains. Two
90 additional strains, *Segnochrobactrum spirostomi* Sp-1^T and *Ahrensia kielensis* DSM 5890^T, lacked 9% and
91 7% of the perc95_143 genes, respectively; however, these two strains were kept in the dataset as they
92 were the only representatives of their respective families. All other strains lacked less than 5% of the
93 perc95_143 genes. Using the reduced dataset of 138 strains, we identified a core set of 59 non-
94 recombining, single-copy marker genes present in all 138 strains (termed "core_138") and 267 single-
95 copy genes present in at least 95% of the target genomes (termed "perc95_138"). A summary of the
96 four gene sets is provided in **Table 1**.
97

98 **Calculation of overall genome relatedness indices**

99 Core-proteome average amino acid identity (cpAAI) was computed as the proportion of differences
100 (including gaps) in pairwise comparisons of protein sequences encoded by a given core gene set.
101 These calculations were performed as described previously [4], and were dependent on the 'ape'
102 package version 5.7-1 in R version 4.3.0 [9]. Whole-proteome average amino acid identity (wpAAI;
103 usually simply known as AAI) was computed using EzAAI version 1.2.2 [10], with default parameters
104 and the dependency Prodigal version 2.6.3 [11].
105

106 **Core-proteome phylogenies**

107 For the core_143 and core_138 gene sets, GET_PHYLOMARKERS was used to align and concatenate
108 the encoded proteins and to remove non-informative sites; alignment was performed using Clustal

109 Omega version 1.2.4 [12]. For the perc95_143 and perc95_138 gene sets, the encoded proteins were
110 aligned with MAFFT version 7.453 [13], after which the protein alignments were trimmed using trimAl
111 version 1.4.rev22 [14] and the automated1 option, and then concatenated. For all four gene sets, the
112 concatenated protein alignments were used as input for ModelFinder [15] as implemented in IQ-TREE
113 version 2.2.2.4 [16], and the best scoring model was identified based on Bayesian information
114 criterion (BIC). IQ-TREE was then used to create maximum-likelihood (ML) phylogenies from the
115 concatenated alignments using the best-scoring model for each protein set (core_143: LG+F+R8;
116 core_138: LG+F+R9; perc95_143: LG+F+I+R10; perc95_138: LG+F+I+R10). Branch supports were
117 assessed in IQ-TREE using the Shimodaira-Hasegawa-like approximate likelihood ratio test (SH-aLRT)
118 [17] and ultrafast jackknife analysis with a subsampling proportion of 40%, with both metrics
119 calculated from 1000 replicates. Phylogenies were visualized using iTOL [18].
120

121 **16S rRNA gene phylogenies**

122 MAFFT was used to align the 16S rRNA gene sequences, after which trimAl was used to trim the
123 alignment. We additionally aligned the 16S rRNA gene sequences using Clustal Omega with the --full
124 and --full-iter options, and then trimmed the alignment with trimAl using the automated1 option. The
125 trimmed nucleotide alignments were used as input for ModelFinder as implemented in IQ-TREE, which
126 for both alignments, identified GTR+F+I+R6 as the best-scoring model based on BIC. IQ-TREE was then
127 used to create ML phylogenies from both trimmed alignments, using the GTR+F+I+R6 model. Branch
128 supports were assessed in IQ-TREE using SH-aLRT and an ultrafast bootstrap analysis, with both
129 metrics calculated from 1000 replicates.
130

131 **Data availability**

132 All genome sequences used in this work were previously published, and the assembly accessions are

133 provided in **Datasets S1-S3**. Likewise, all 16S rRNA gene sequences used in this study were previously
134 published, and the corresponding GenBank accessions for those not extracted directly from the whole
135 genome sequences are provided in **Dataset S4**. Twelve supplementary figures, four supplementary
136 datasets, and supplementary text are included in the online version of this article. All raw data (cpAAI
137 data, wpAAI data, and Newick formatted phylogenies) used to generate the figures presented in this
138 manuscript are available through FigShare
139 (figshare.com/articles/online_resource/diCenzo_et_al_2023_Hyphomicrobiales_taxonomy/24417334)
140 . All code to repeat the analyses in this study is available through GitHub (github.com/diCenzo-
141 Lab/012_2023_Hyphomicrobiales_taxonomy). Protein marker sets to allow for the identification of the
142 marker proteins in other organisms using the cpAAI_Rhizobiaceae pipeline
143 (github.com/flass/cpAAI_Rhizobiaceae) [4] are available through GitHub (github.com/diCenzo-
144 Lab/012_2023_Hyphomicrobiales_taxonomy).
145

146 **RESULTS AND DISCUSSION**

147 **Genome-based metrics for family-level delimitation in the order *Hyphomicrobiales***

148 To evaluate family assignments in the order *Hyphomicrobiales*, we began with a dataset of 138
149 *Hyphomicrobiales* strains, including 136 genus type strains, including at least one representative from
150 each of the 38 families. Five *Caulobacterales* genus type strains were included as an outgroup. As in
151 our recent study of the family *Rhizobiaceae* [4], we reasoned that genome sequence-based family
152 delineation should consider both phylogenetic relatedness and genome similarity based on one or
153 more overall genome relatedness indices (OGRIs). We chose to work with cpAAI and wpAAI as these
154 appeared to be the most appropriate OGRIs based on our previous work [4]. Therefore, we first
155 identified a set of 19 marker genes present in 100% of the 143 strains (termed core_143) and a set of
156 256 marker genes present in ≥95% of the strains (termed perc95_143). Both datasets were used to

157 construct ML phylogenies (**Figures S1 and S2**), and core_143 was additionally used to calculate
158 pairwise cpAAI values between all 143 strains (**Figure S1**, **Dataset S5**). Pairwise wpAAI was also
159 calculated between all 143 strains (**Figures S3**, **Dataset S6**). Early analysis of these datasets revealed
160 that five strains lacked >10% of the perc95_143 genes and that removal of these five strains
161 significantly increased the size of the core genome (see **Text S1**). We therefore repeated all analyses
162 using a dataset lacking these five strains. Using this reduced dataset, we identified a set of 59 marker
163 genes present in 100% of the strains (termed core_138) and a set of 267 marker genes present in
164 ≥95% of the strains (termed perc95_138). ML phylogenies were constructed from both core_138
165 (**Figure S4**) and perc95_138 (**Figure 1**), and pairwise cpAAI values were calculated between all 138
166 strains using core_138 (**Figure 1**, **Dataset S7**). A summary of the four gene sets is provided in **Table**
167 **1**.

168 We wondered whether we could identify biologically relevant cpAAI or wpAAI thresholds for
169 delineating families in the order *Hyphomicrobiales*. We therefore plotted the distribution of cpAAI
170 values (calculated from core_138) as a histogram, plotting separate distributions for within-family,
171 between-family but within-order, and between-order comparisons (**Figures 2A** and **2C**). In addition to
172 using all cpAAI values calculated from core_138, we plotted the results for a reduced dataset lacking
173 the family *Rhizobiaceae* (**Figure 2B** and **2D**). This was done for two reasons. The family *Rhizobiaceae*
174 was over-represented in our dataset, accounting for over 15% of the strains in our analysis, and thus
175 its inclusion had the potential to bias the distribution. In addition, the family *Rhizobiaceae* was not well
176 separated from the related *Bartonellaceae – Phyllobacteriaceae – Brucellaceae - Notoacmeibacteraceae*
177 clade (as discussed below), and thus had the potential to mask patterns that might otherwise be
178 detected across the remaining 33 *Hyphomicrobiales* families.

179 The within-family, between-family, and between-order distributions of cpAAI were clearly distinct,
180 with mean values of 78.0%, 69.4%, and 62.6%, respectively, for the dataset lacking the family

181 *Rhizobiaceae*. Notably, excluding the family *Rhizobiaceae*, 97.1% of the between-family comparisons
182 fell below 75%, while 80.1% of the within-family comparisons were above 75% (**Figure 2B**). When
183 considering the taxonomic changes proposed below, this increases to 98.3% of between-family
184 comparisons being below 75% and 88.5% (93.8% when excluding *Bartonella bacilliformis* KC583^T that
185 has a highly reduced genome) of within-family comparisons being above 75% (**Figure 2D**). These
186 results, together with the shapes of the distributions, suggested to us that a cpAAI value of ~75%,
187 when calculated from core_138, represented a natural threshold for defining families within the order
188 *Hyphomicrobiales*. Performing similar analyses using cpAAI values calculated from core_143 (**Figure**
189 **S5**), or the wpAAI data (**Figure S6**), also allowed for the identification of potential thresholds (~78%
190 and ~63%, respectively) for defining families within the order *Hyphomicrobiales*. However, these
191 thresholds were less well defined compared to the threshold identified for core_138 and thus we focus
192 our analyses primarily on the cpAAI data generated with core_138.

193 We attempted to supplement our genome-based analyses with 16S rRNA gene phylogenetic
194 studies (**Figures S7 and S8**). However, the 16S rRNA gene phylogenies we built were non-concordant
195 with the genome-based phylogenies and failed to accurately resolve multiple families, consistent with
196 past studies [19]. We therefore only made use of the 16S rRNA gene phylogenies for taxa lacking whole
197 genome sequencing data.

198 In the following sections, we propose several taxonomic revisions in the order *Hyphomicrobiales*
199 based on overall genomic similarity and the need for families to be monophyletic. In general, we
200 aimed to be conservative and limit the number of changes that we propose, and therefore have not
201 proposed to split monophyletic families based solely on low cpAAI values. We do, however, propose
202 to unify sister families where appropriate and only when clearly supported by the cpAAI data.

203

204 **Taxonomic implications for the families *Rhizobiaceae*, *Ahrensiaceae*, *Bartonellaceae*,**

205 ***Brucellaceae*, *Phyllobacteriaceae*, and *Notoacmeibacteraceae***

206 The six families *Rhizobiaceae*, *Ahrensiaceae*, *Bartonellaceae*, *Brucellaceae*, *Phyllobacteriaceae*, and
207 *Notoacmeibacteraceae* collectively represent over a third of the *Hyphomicrobiales* strains included in
208 our analyses. Of note was the family *Phyllobacteriaceae*, which was not monophyletic in any of the
209 phylogenies and whose genus type strain (*Phyllobacterium myrsinacearum* DSM 5892^T) consistently
210 formed its own lineage separate from all other members of the family (**Figures 1, S1, S2, and S4**).
211 *Pseudohoeflea suaedae* YC6898^T was consistently nested within the family *Rhizobiaceae* in all
212 phylogenies (**Figures 1, S1, S2, and S4**). We therefore propose that the genus *Pseudohoeflea* be
213 reassigned to the family *Rhizobiaceae*. In addition, *Salaquimonas pukyongi* RR3-28^T consistently formed
214 its own lineage (**Figures 1, S1, S2, and S4**), and when calculated from core_138, displayed pairwise
215 cpAAI values <74% against all other strains (**Figure 1**). We therefore propose that the genus
216 *Salaquimonas* be reassigned to the family *Salaquimonadaceae* fam. nov.. Moreover, *Oricola*
217 *cellulosilytica* JCM 19534^T, *Roseitalea porphyridii* MA7-20^T, and *Oceaniradius stylonematis* StC1^T
218 consistently grouped separately from the rest of the family *Phyllobacteriaceae* and instead formed a
219 clade with *Ahrensia kielensis* DSM 5890^T (**Figures 1, S1, S2, and S4**), the genus type strain for the type
220 genus of the family *Ahrensiaceae*. When calculated from core_138, all pairwise cpAAI values between
221 the four strains were >78% (**Figure 1**). We therefore propose that the genera *Oricola*, *Roseitalea*, and
222 *Oceaniradius* be reassigned to the family *Ahrensiaceae*.

223 The remaining 11 members of the family *Phyllobacteriaceae* included in this analysis were
224 paraphyletic in all four phylogenies (**Figures 1, S1, S2, and S4**). The minimal monophyletic group
225 containing all 11 *Phyllobacteriaceae* species also includes the families *Brucellaceae*, *Bartonellaceae*, and
226 *Notoacmeibacteraceae*; the inclusion of the genus *Bartonella* within this monophyletic group was
227 further supported by supplemental phylogenetic analyses (see **Text S2**; **Figures S9-S12**). As a result,
228 it was necessary either to split the family *Phyllobacteriaceae* into three families or to unify the families

229 *Phyllobacteriaceae*, *Brucellaceae*, *Bartonellaceae*, and *Notoacmeibacteraceae*. The cpAAI and wpAAI
230 measurements favour the latter option; when calculated from core_138, all pairwise cpAAI values
231 between members of these families were above 76.5%, except for comparisons involving *B.*
232 *bacilliformis* KC583^T, which has a highly reduced genome (cpAAI values involving this strain ranged
233 from 69% to 74.5%), or *Notoacmeibacter marinus* XMTR2A4^T, which displayed unexpectedly low cpAAI
234 values against all strains (cpAAI values involving this strain ranged from 69% to 76%). Likewise, most
235 cpAAI comparisons calculated with core_143 and wpAAI values were above the proposed thresholds
236 of 78% and 63%, respectively. Based on these results, we propose that the families *Bartonellaceae*,
237 *Brucellaceae*, *Notoacmeibacteraceae* and *Phyllobacteriaceae* be unified with the name *Bartonellaceae* as
238 it takes priority. In addition, the strain *Limoniibacter endophyticus* KCTC 42097^T, currently assigned to
239 the family *Hyphomicrobiaceae*, was consistently nested within the emended family *Bartonellaceae* in all
240 phylogenies. We therefore propose that the genus *Limoniibacter* be reassigned to the family
241 *Bartonellaceae*.

242 In contrast to our initial expectations [4], the data suggested that the family *Rhizobiaceae* did not
243 need to be split into two or more families, while our supplemental phylogenetic analyses (see **Text**
244 **S2**; **Figures S9-S12**) supported the inclusion of the recently proposed genus "*Flavimaribacter*" in the
245 family *Rhizobiaceae* [20]. On the other hand, when calculated from core_138, 82.5% of the cpAAI
246 comparisons between members of the family *Rhizobiaceae* and the emended family *Bartonellaceae*
247 were above our proposed threshold of 75% (**Figure 1**). Similarly, 71.2% of the wpAAI comparisons
248 between members of the family *Rhizobiaceae* and the emended family *Bartonellaceae* were above our
249 proposed threshold of 63% (**Figure S3**), although only 36.6% of the cpAAI comparisons were above
250 our proposed threshold of 78% when calculated from core_143 (**Figure S1**). The possibility of uniting
251 these two families therefore merits consideration. However, we do not favour this approach for two
252 reasons. First, the between-family cpAAI values (calculated from core_138) tended to be lower than

253 the within-family cpAAI values (**Figure 1**), with average values of 75.9% and 80.5%, respectively;
254 qualitatively similar results are seen with cpAAI calculated from core_143 and the wpAAI data (**Figure**
255 **S1 and S3**). In addition, in three of the four phylogenies, the families *Rhizobiaceae* and *Bartonellaceae*
256 were not sister taxa; instead, these families only formed a monophyletic group when the emended
257 family *Ahrensiaceae* was also included. The family *Ahrensiaceae* was better delineated from the family
258 *Rhizobiaceae* based on cpAAI and wpAAI values (**Figures 1, S1, and S3**); the average cpAAI value (when
259 calculated from core_138) for comparisons between species of the families *Ahrensiaceae* and
260 *Rhizobiaceae* was 74.1%, with 84% of comparisons below the proposed threshold of ~75% for family
261 delineation. Likewise, in the cases of cpAAI calculated from core_143 and wpAAI, 99% and 86%,
262 respectively, of the comparisons between species of the families *Ahrensiaceae* and *Rhizobiaceae* were
263 below the proposed thresholds for family delineation. These results suggest it would not be
264 appropriate to unify the families *Ahrensiaceae* and *Rhizobiaceae*. Considering the weight of the
265 evidence, we have therefore chosen to leave the *Rhizobiaceae*, the emended *Ahrensiaceae*, and the
266 emended *Bartonellaceae* as distinct families.
267
268 **Taxonomic implications: formation of new families**
269 Several families in the order *Hyphomicrobiales* were paraphyletic in all four of our phylogenies. To fix
270 the paraphyly and ensure all families are monophyletic, we propose the formation of three new
271 families and the reassignment of three genera to other existing families (see the following section).
272 The genus type strain *Rhodoblastus acidophilus* DSM 137^T is currently assigned to the family
273 *Roseiarcaceae*. However, *R. acidophilus* DSM 137^T did not form a monophyletic group with the genus
274 type strain *Roseiarcus fermentans* DSM 24875^T that represents the type genus of the family
275 *Roseiarcaceae*. Instead, *R. acidophilus* DSM 137^T formed its own lineage as a sister taxon to the family
276 *Methylocystaceae* in all four phylogenies (**Figures 1, S1, S2, and S4**). When calculated from core_138,

277 the pairwise cpAAI values between *R. acidophilus* DSM 137^T and the two included genus type strains
278 of the family *Methylocystaceae* were 74.8% and 74.5% (**Figure 1**), which are below our proposed
279 threshold of 75% for family deliminatation. We therefore propose to reassign the genus *Rhodoblastus*
280 to the family *Rhodoblastaceae* fam. nov..

281 The genus *Methylopila* was not assigned to a family when it was proposed [21], but was
282 subsequently assigned to the family *Methylocystaceae* [22]. However, the genus type strain *Methylopila*
283 *capsulata* DSM 6130^T did not cluster with the other members of the family *Methylocystaceae*, and
284 instead formed its own distinct lineage in the two phylogenies based on core_138 or perc95_138
285 (**Figures 1 and S3**). When calculated from core_138, all comparisons involving *M. capsulata* DSM 6130^T
286 gave cpAAI values less than 74% (**Figure 1**). In the phylogenies based on core_143 or perc95_143,
287 which include five additional genus type strains, *M. capsulata* DSM 6130^T formed a clade with
288 *Chenggangzhangella methanolivorans* CHL1^T also currently assigned to the family *Methylocystaceae*
289 (**Figures S1 and S2**). When calculated from core_143, these two strains had a pairwise cpAAI value of
290 86.5% (**Figure S1**), which is above the proposed threshold of 78% for family delineation. The 16S rRNA
291 gene phylogenies (**Figures S7 and S8**) included two additional genus type strains, *Hansschlegelia*
292 *plantiphila* VKM B-2347^T and *Albibacter methylovorans* DM10^T, of genera currently assigned to the
293 family *Methylocystaceae* but that lack publicly available whole genome sequence. In both phylogenies,
294 *H. plantiphila* VKM B-2347^T and *A. methylovorans* DM10^T formed a monophyletic clade with *M. capsulata*
295 DSM 6130^T and *C. methanolivorans* CHL1^T, rather than with *Methylocystis parvus* OBBP^T, the genus type
296 strain of the type genus of the family *Methylocystaceae*. Based on these results, we propose to reassign
297 the genera *Methylopila*, *Chenggangzhangella*, *Hansschlegelia*, and *Albibacter* to the family
298 *Methylopilaceae* fam. nov..

299 The genus type strain *Rhodoligotrophos appendicifer* 120-1^T is currently assigned to the family
300 *Parvibaculaceae*. However, instead of clustering with the other members of the family *Parvibaculaceae*,

301 it formed its own lineage as a sister taxon to the genus type strain *Aestuariivirga litoralis* KCTC 52945^T
302 of the family *Aestuariivirgaceae* in all four phylogenies (**Figures 1, S1, S2, and S4**). When calculated
303 from core_138, the pairwise cpAAI value between *R. appendicifer* 120-1^T and *A. litoralis* KCTC 52945^T
304 was 71.2% (**Figure 1**), which is below our proposed threshold of 75% for family delineation. We
305 therefore propose to reassign the genus *Rhodoligotrophos* to the family *Rhodoligotrophaceae* fam. nov..
306

307 **Taxonomic implications: reassignment of genera**

308 The genus type strain *Faunimonas pinastri* A52C2^T is currently assigned to the family
309 *Pleomorphomonadaceae* but did not cluster with other members of this family. Instead, in all four
310 phylogenies, *F. pinastri* A52C2^T formed a clade with the genus type strain *Afifella marina* DSM2698^T
311 that represents the only genus of the family *Afifellaceae* (**Figures 1, S1, S2, and S4**). When calculated
312 from core_138, the cpAAI value between *F. pinastri* A52C2^T and *A. marina* DSM2698^T was 75.6% (**Figure**
313 **1**), which is above our proposed threshold of 75% for family delimitation. We therefore propose to
314 reassign the genus *Faunimonas* to the family *Afifellaceae*.

315 The genus *Pseudochelatococcus* was not assigned to a family when it was proposed [23], but was
316 subsequently associated with the family *Beijerinckiaceae* [24]. However, the genus type strain
317 *Pseudochelatococcus lubricantis* DSM 103870^T did not cluster with the other members of the family
318 *Beijerinckiaceae*, and instead formed a clade with *Chelatococcus asaccharovorans* DSM 6462^T and
319 *Camelimonas lactis* DSM22958^T from the family *Chelatococcaceae* in all four phylogenies (**Figures 1, S1,**
320 **S2, and S4**). Moreover, the family *Chelatococcaceae* was not monophyletic without the inclusion of *P.*
321 *lubricantis* DSM 103870^T (**Figures 1, S1, S2, and S4**). When calculated from core_138, all pairwise cpAAI
322 values in this clade were >75% (**Figure 1**). We therefore propose to assign the genus
323 *Pseudochelatococcus* to the family *Chelatococcaceae*.

324 The genus type strain *Prosthecomicrobium pneumaticum* DSM 16268^T was previously assigned to

325 the family *Hyphomicrobiaceae* [25], although it was recently proposed that the genus
326 *Prosthecomicrobium* be reassigned to the family *Kaistiaceae* [26]. In all four phylogenies, *P.*
327 *pneumaticum* DSM 16268^T formed a clade with *Kaistia adipata* DSM 17808^T and *Bauldia litoralis* ATCC
328 35022^T of the family *Kaistiaceae* rather than with species of the family *Hyphomicrobiaceae* (**Figures 1,**
329 **S1, S2, and S4**). Moreover, the family *Kaistiaceae* was not monophyletic without the inclusion of *P.*
330 *pneumaticum* DSM 16268^T (**Figures 1, S1, S2, and S4**). When calculated from core_138, all pairwise
331 cpAAI values in this clade were >75% (**Figure 1**). Our data are therefore consistent with the proposal
332 to reassign the genus *Prosthecomicrobium* to the family *Kaistiaceae*.
333

334 **Taxonomic implications: unification of families**

335 The families *Segnochrobactraceae* and *Pseudoxanthobacteraceae* each contain a single validly published
336 genus. The genus type strains of these two genera formed a monophyletic group in all four
337 phylogenies (**Figures 1, S1, S2, and S4**). When comparing the genomes of these two strains, the
338 pairwise cpAAI value was 82.9% when calculated from core_138 (**Figure 1**) and 85.9% when calculated
339 from core_143 (**Figure S1**), while the wpAAI value was 70.0% (**Figure S3**). Given that these values are
340 all well above the proposed thresholds for family delimitation, we propose that the families
341 *Segnochrobactraceae* and *Pseudoxanthobacteraceae* be unified with the name *Segnochrobactraceae* as
342 it takes priority.

343 The families *Lichenihabitantaceae* and *Lichenibacteriaceae* each contain a single validly published
344 genus, based on strains that were originally isolated from lichens in the Antarctic and the subarctic
345 zone of the northern hemisphere, respectively [27, 28]. The genus type strains of these two genera
346 formed a monophyletic group in all four phylogenies (**Figures 1, S1, S2, and S4**). When comparing the
347 genomes of these two strains, the pairwise cpAAI value was 77.9% when calculated from core_138
348 (**Figure 1**) and 78.9% when calculated from core_143 (**Figure S1**), while the wpAAI value was 67%

349 (**Figure S3**). Given that these values are all above the proposed thresholds for family delimitation, we
350 propose that the families *Lichenihabitantaceae* and *Lichenibacteriaceae* be unified with the name
351 *Lichenihabitantaceae* as it takes priority.

352 The family *Breoghaniaceae* contains a single genus while the family *Stappiaceae* contains five
353 genera. The six genus type strains from these two families formed a monophyletic group in all four
354 phylogenies (**Figures 1, S1, S2, and S4**). When calculated from core_138, the pairwise cpAAI values
355 between members of the family *Stappiaceae* ranged between 77.0% and 81.6%, which overlapped the
356 cpAAI range of 76.2% to 78.9% for comparisons between members of the family *Stappiaceae* and
357 *Breoghania corrubedonensis* DSM 23382^{T} (**Figure 1**). Similar results are seen when using cpAAI
358 calculated from core_143 and wpAAI values (**Figures S1 and S3**). Considering these results, we
359 propose that the families *Stappiaceae* and *Breoghaniaceae* be unified with the name *Stappiaceae*. Both
360 family names were published in the same article [2] but *Stappiaceae* is our preferred choice because
361 *Breoghaniaceae* includes only a single genus.

362

363 **Other taxonomic implications**

364 In this study, we refrained from splitting monophyletic families based solely on cpAAI data.
365 Nevertheless, multiple families merit further study as candidates for splitting, namely the families
366 *Amorphaceae*, *Devosiaceae*, *Hyphomicrobiaceae*, and *Parvibaculaceae* (**Figure 1**). Additionally, a couple
367 of clades were not well-resolved based on the cpAAI data and the proposed thresholds, and thus also
368 merit further study. The family *Rhodobiaceae* and the emended family *Stappiaceae* are sister taxa that
369 are not well separated in the cpAAI data calculated from core_138 (**Figure 1**), with four of the six
370 between-family comparisons above 75%, although better separation is observed in the cpAAI data
371 calculated from core_143 and the wpAAI data (**Figure S1 and S3**). Likewise, the families
372 *Methylobacteriaceae* and *Salinarimonadaceae* were not well separated by the cpAAI or wpAAI data; six

373 of the eight cpAAI values calculated from core_138 were above 75% with the other two above 74.7%
374 (**Figure 1**), four of the eight cpAAI values calculated from core_143 were above 78% with the other
375 four above 77.5% (**Figure S1**), and all wpAAI values were above the proposed threshold of 63% (**Figure**
376 **S3**). As there was not overwhelming support for unification of the families *Rhodobiaceae* and
377 *Stappiaceae*, or the families *Methylobacteriaceae* and *Salinarimonadaceae*, we chose to leave all four as
378 distinct families. However, these taxa merit further consideration in future work.

379 Although it is not a genus type strain, we included the species type strain "*Pararhizobium*
380 *mangrovi*" BGMRC 6574^T in our dataset as preliminary investigations of the family *Rhizobiaceae*
381 suggested it belonged to a novel genus and potentially a novel family. Our full analysis did not support
382 the reassignment of "*P. mangrovi*" BGMRC 6574^T to a new family but it did support the recent proposal
383 to reclassify this strain as "*Allopararhizobium mangrovi*" BGMRC 6574^T once validly published [29].
384

385 **Emended description of *Hyphomicrobiales* Douglas et al. 1957 (Approved Lists 1980) emend.**
386 **Hördt et al. 2020**

387 The description is as given before [1, 2], with the following modification. The order currently comprises
388 the families *Aestuariivirgaceae*, *Afifellaceae*, *Ahrensiaceae*, *Alsobacteraceae*, *Amorphaceae*,
389 *Ancalomicrobiaceae*, *Aurantimonadaceae*, *Bartonellaceae*, *Beijerinckiaceae*, *Blastochloridaceae*,
390 *Boseaceae*, *Chelatococcaceae*, *Cohaesibacteraceae*, *Devosiaceae*, *Hyphomicrobiaceae*, *Kaistiaceae*,
391 *Lichenihabitantaceae*, *Methylobacteriaceae*, *Methylocystaceae*, *Methylopilaceae* fam. nov.,
392 *Nitrobacteraceae*, *Parvibaculaceae*, *Phreatobacteraceae*, *Pleomorphomonadaceae*, *Rhabdaerophilaceae*,
393 *Rhizobiaceae*, *Rhodobiaceae*, *Rhodoblastaceae* fam. nov., *Rhodoligotrophaceae* fam. nov., *Roseiarcaceae*,
394 *Salaquimonadaceae* fam. nov., *Salinarimonadaceae*, *Segnochrobactraceae*, *Stappiaceae*,
395 *Tepidamorphaceae*, and *Xanthobacteraceae*. The type genus is *Hyphomicrobium*. The families
396 *Breoghaniaceae*, *Brucellaceae*, *Lichenibacteriaceae*, *Notoacmeibacteraceae*, *Phyllobacteriaceae*, and

397 *Pseudoxantobacteraceae* were removed from the order and their genera transferred to other families
398 within the order.
399
400 **Description of *Methylopilaceae* fam. nov.**
401 Me.thyl.o.pil.a'ce.ae (N.L. fem. n. *Methylopila*, type genus of the family; -*aceae*, ending to denote a
402 family; N.L. fem. pl. n. *Methylopilaceae*, the *Methylopila* family).
403 Cells are aerobic, Gram-negative, rod-shaped or cocci, and motile or non-motile. The G+C content
404 as calculated from genome sequences is 68.2-69.7%, while the range provided in the literature is 64.4-
405 70.4 mol% [30, 31]. The family currently comprises the genera *Albibacter*, *Chenggangzhangella*,
406 *Hansschlegelia*, and *Methylopila* (the type genus). This family can be differentiated from other families
407 based on phylogenetic analyses of core proteome and 16S rRNA gene sequences, as well as OGRI
408 calculations (cpAAI and wpAAI).
409
410 **Description of *Rhodoblastaceae* fam. nov.**
411 Rho.do.blast.a'ceae (N.L. masc. n. *Rhodoblastus*, type genus of the family; -*aceae*, ending to denote a
412 family; N.L. fem. pl. n. *Rhodoblastaceae*, the *Rhodoblastus* family).
413 The description is as given for *Rhodoblastus* [32], which is the type and currently the sole genus
414 of the family. This family can be differentiated from other families based on phylogenetic analyses of
415 core proteome sequences and OGRI calculations (cpAAI and wpAAI).
416
417 **Description of *Rhodoligotrophaceae* fam. nov.**
418 Rho.do.li.go.tro'ph.a'ce.ae (N.L. masc. n. *Rhodoligotrophos*, type genus of the family; -*aceae*, ending to
419 denote a family; N.L. fem. pl. n. *Rhodoligotrophaceae*, the *Rhodoligotrophos* family).
420 The description is as given for *Rhodoligotrophos* [33], which is the type and currently the sole

421 genus of the family. This family can be differentiated from other families based on phylogenetic
422 analyses of core proteome sequences and OGRI calculations (cpAAI and wpAAI).
423
424 **Description of *Salaquimonadaceae* fam. nov.**
425 Sal.a.qui.mo'nad.a'ce.ae (N.L. fem. n. *Salaquimonas*, type genus of the family; -*aceae*, ending to denote
426 a family; N.L. fem. pl. n. *Salaquimonadaceae*, the *Salaquimonas* family).
427 The description is as given for *Salaquimonas* [34], which is the type and currently the sole genus
428 of the family. This family can be differentiated from other families based on phylogenetic analyses of
429 core proteome sequences and OGRI calculations (cpAAI and wpAAI).
430
431 **Emended description of *Afifellaceae* Hördt et al. 2020**
432 The description is as given before [2], with the following modification. The family currently comprises
433 the genera *Afifella* (the type genus), and *Faunimonas*.
434
435 **Emended description of *Ahrensiaceae* Hördt et al. 2020**
436 The description is as given before [2], with the following modification. The family currently comprises
437 the genera *Ahrensia* (the type genus), *Pseudahrensia*, *Oricola*, *Roseitalea*, and *Oceaniradius*.
438
439 **Emended description of *Bartonellaceae* Gieszczykiewicz 1939 (Approved Lists 1980) emend.**
440 **Brenner et al. 1993**
441 The description is based on descriptions provided previously [2, 35–39]. Cells are Gram-negative,
442 aerobic, with variable morphology (rod, ovoid, coccoid, coccobacilli, or ring or disk shaped). Cells can
443 be motile by means of flagella, or non-motile. Generally catalase and oxidase positive. Predominantly
444 aerobic or facultatively anaerobic heterotrophs, with some species able to utilize carbohydrates while

445 others cannot. Generally non-spore-forming. The optimum growth temperature ranges between 20˚C
446 and 37˚C. The predominant respiratory quinone is generally Q-10. Some species are human
447 pathogens. The G+C content as calculated from genome sequences of genus type strains is 38.2-65.1%
448 The family currently comprises the genera *Aminobacter*, *Aquamicrobium*, *Aquibium*, *Bartonella* (the
449 type genus), *Brucella*, *Chelativorans*, *Corticibacterium*, *Daeguia*, *Falsochrobactrum*, *Limoniibacter*,
450 *Mesorhizobium*, *Nitratireductor*, *Notoacmeibacter*, *Ochrobactrum*, *Paenochrobactrum*, *Phyllobacterium*,
451 *Pseudaminobacter*, *Pseudochrobactrum*, *Tianweitania*, and *Zhengella*.
452
453 **Emended description of *Beijerinckiaceae* Grarity et al. 2006 emend. Hördt et al. 2020**
454 The description is as given before [2]. The family currently comprises the genera *Beijerinckia* (the type
455 genus), *Methylocapsa*, *Methylocella*, *Methyloferula*, *Methylorosula*, and *Methylovirgula*. The genus
456 *Pseudochelatococcus* was removed from the family and transferred to the family *Chelatococcaceae*.
457
458 **Emended description of *Chelatococcaceae* Dedysh et al. 2016**
459 The description is as given before [40], with the following modification. The family currently comprises
460 the genera *Camelimonas*, *Chelatococcus* (the type genus), and *Pseudochelatococcus*.
461
462 **Emended description of *Hyphomicrobiaceae* Babudieri 1950 (Approved Lists 1980) emend. Hördt**
463 **et al. 2020**
464 The description is as given before [2] with the following modifications. The family currently comprises
465 the genera *Dichotomicrobium*, *Filomicrobium*, *Hyphomicrobium* (the type genus), *Methyloceanibacter*,
466 *Methyloligella*, *Pedomicrobium*, *Rhodomicrobium*, and *Seliberia*. The genus *Limoniibacter* was removed
467 from the family and transferred to the emended family *Bartonellaceae*. The family also does not
468 encompass the genus *Caenibius* (*Erythrobacteraceae, Sphingomonadales*), which appears to have been

469 mistakenly included in the description of the family *Hyphomicrobiaceae* by Hördt et al. (2020) [2].

470

471 **Emended description of *Lichenihabitantaceae* Noh et al. 2019**

472 The description is as given before [27] with the following modifications. Cells are aerobic or facultative
473 anaerobic. The family currently comprises the genera *Lichenibacterium* and *Lichenihabitans* (the type
474 genus).

475

476 **Emended description of *Methylocystaceae* Bowman 2006 emend. Hördt et al. 2020**

477 The description is as given before [2]. The family currently comprises the genera *Methylocystis* (the
478 type genus) and *Methylosinus*. The genera *Albibacter*, *Chenggangzhangella*, *Hansschlegelia*, and
479 *Methylopila* were removed from the family and transferred to *Methylopilaceae* fam. nov..

480

481 **Emended description of *Parvibaculaceae* Hördt et al. 2020**

482 The description is as given before [2], with the following modification. The family currently comprises
483 the genera *Parvibaculum* (the type genus) and *Tepidicaulis*. The genera *Anderseniella* and *Pyruvatibacter*
484 are tentatively assigned to this family as well. The genus *Rhodoligotrophos* was removed from the
485 family and transferred to *Rhodoligotrophaceae* fam. nov..

486

487 **Emended description of *Pleomorphomonadaceae* Hördt et al. 2020**

488 The description is as given before [2]. The family currently comprises the genera *Chthonobacter*,
489 *Hartmannibacter*, *Methylobrevis*, *Mongoliimonas*, *Oharaeibacter*, and *Pleomorphomonas* (the type
490 genus). The genus *Faunimonas* assigned to this family by Proença et al. (2022) [41] was removed from
491 the family and transferred to the family *Afifellaceae*.

492

493 **Emended description of *Rhizobiaceae* Conn 1938 (Approved Lists 1980) emend. Hördt et al. 2020**

494 The description is as given before [2, 42], with the following modification. The family currently
495 comprises the genera *Agrobacterium*, *Allorhizobium*, *Ciceribacter*, *Endobacterium*, *Ensifer*,
496 *Ferranicluibacter*, *Gellertiella*, *Georhizobium*, *Hoeflea*, *Lentilitoribacter*, *Liberibacter*, *Martelella*, *Mycoplana*,
497 *Neorhizobium*, *Onobrychidicola*, *Pararhizobium*, *Peteryoungia*, *Pseudohoeflea*, *Pseudorhizobium*,
498 *Rhizobium* (the type genus), *Shinella*, *Sinorhizobium*, and *Xaviernesmea*.

499

500 **Emended description of *Roseiarcaceae* Kulichevskaya et al. 2014 emend. Hördt et al. 2020**

501 The description is as given before [2, 43], with the following modification. The family comprises the
502 genus *Roseiarcus*, which is the type and currently the sole genus of the family. The genus *Rhodoblastus*
503 was removed from the family and transferred to *Rhodoblastaceae* fam. nov..

504

505 **Emended description of *Segnochrobactraceae* Akter et al. 2020**

506 The description is as given before [44], with the following modification. The family currently comprises
507 the genera *Pseudoxanthobacter* and *Segnochrobactrum* (the type genus).

508

509 **Emended description of *Stappiaceae* Hördt et al. 2020**

510 The description is as given before [2], with the following modification. The family currently comprises
511 the genera *Breoghania*, *Hongsoonwoonella*, *Pannonibacter*, *Pseudovibrio*, *Roseibium*, and *Stappia* (the
512 type genus).

513

514 **ACKNOWLEDGEMENTS**

515 This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC)
516 grant to G.C.D.

517 **CONFLICT OF INTEREST**

518 The authors declare that they have no conflict of interest.

519

520 **REFERENCES**

521 1. **Kuykendall LD**. Order VI, *Rhizobiales* ord. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM
522 (editors). *Bergey's Manual of Systematic Bacteriology: The Proteobacteria, (Part C)*. New York, NY:
523 Springer; 2005. p. 324.

524 2. **Hördt A, López MG, Meier-Kolthoff JP, Schleuning M, Weinhold L-M, *et al.*** Analysis of 1,000+
525 type-strain genomes substantially improves taxonomic classification of *Alphaproteobacteria*. *Front*
526 *Microbiol* 2020;11:468.

527 3. **Douglas HC**. Order III. *Hyphomicrobiales* Douglas, ord. nov. In: Breed RS, Murray EGD, Smith NR
528 (editors). *Bergey's Manual of Determinative Bacteriology*. Baltimore, USA: The Williams and Wilkins
529 Co.; 1957. p. 276.

530 4. **Kuzmanović N, Fagorzi C, Mengoni A, Lassalle F, diCenzo GC**. Taxonomy of *Rhizobiaceae*
531 revisited: proposal of a new framework for genus delimitation. *Int J Syst Evol Microbiol*
532 2022;72:005243.

533 5. **Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M**. List of Prokaryotic names
534 with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–
535 5612.

536 6. **Contreras-Moreira B, Vinuesa P**. GET_HOMOLOGUES, a versatile software package for scalable
537 and robust microbial pangenome analysis. *Appl Environ Microbiol* 2013;79:7696–7701.

538 7. **Vinuesa P, Ochoa-Sánchez LE, Contreras-Moreira B**. GET_PHYLOMARKERS, a software package
539 to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies,
540 used for a critical geno-taxonomic revision of the genus *Stenotrophomonas*. *Front Microbiol*

541 2018;9:771.

542 8. **Kuzmanović N, Biondi E, Overmann J, Puławska J, Verbarg S, *et al.*** Genomic analysis provides
543 novel insights into diversification and taxonomy of *Allorhizobium vitis* (i.e. *Agrobacterium vitis*). *BMC*
544 *Genomics* 2022;23:462.

545 9. **Paradis E, Schliep K**. ape 5.0: an environment for modern phylogenetics and evolutionary
546 analyses in R. *Bioinformatics* 2019;35:526–528.

547 10. **Kim D, Park S, Chun J**. Introducing EzAAI: a pipeline for high throughput calculations of
548 prokaryotic average amino acid identity. *J Microbiol* 2021;59:476–480.

549 11. **Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, *et al.*** Prodigal: prokaryotic gene
550 recognition and translation initiation site identification. *BMC Bioinform* 2010;11:119.

551 12. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, *et al.*** Fast, scalable generation of high-quality
552 protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;7:539.

553 13. **Katoh K, Standley DM**. MAFFT multiple sequence alignment software version 7: improvements
554 in performance and usability. *Mol Biol Evol* 2013;30:772–780.

555 14. **Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T**. trimAl: a tool for automated alignment
556 trimming in large-scale phylogenetic analyses. *Bioinformatics* 2009;25:1972–1973.

557 15. **Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS**. ModelFinder: fast model
558 selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.

559 16. **Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al.*** IQ-TREE 2: new
560 models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*
561 2020;37:1530–1534.

562 17. **Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, *et al.*** New algorithms and methods
563 to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*
564 2010;59:307–321.

565 18. **Letunic I, Bork P**. Interactive tree of life (iTOL) v3: an online tool for the display and annotation
566 of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
567 19. **Hassler HB, Probert B, Moore C, Lawson E, Jackson RW, *et al.*** Phylogenies of the 16S rRNA
568 gene and its hypervariable regions lack concordance with core genome phylogenies. *Microbiome*
569 2022;10:104.
570 20. **Wang X-N, Wang L, He W, Yang Q, Zhang D-F**. Description of *Flavimaribacter sediminis* gen. nov.,
571 sp. nov., a new member of the family *Rhizobiaceae* isolated from marine sediment. *Curr Microbiol*
572 2023;80:301.
573 21. **Doronina NV, Trotsenko YA, Krausova VI, Boulygina ES, Tourova TP**. *Methylopila capsulata* gen.
574 nov., sp. nov., a novel non-pigmented aerobic facultatively methylotrophic bacterium. *Int J Syst*
575 *Bacteriol* 1998;48:1313–1321.
576 22. **Bowman JP**. Family V. *Methylocystaceae* fam. nov. In: Brenner DJ, Krieg NR, Staley JT, Garrity GM
577 (editors). *Bergey's Manual of Systematic Bacteriology: The Proteobacteria, (Part C)*. New York, NY:
578 Springer; 2005. p. 411.
579 23. **Kämpfer P, Glaeser SP, Gräber M, Rabenstein A, Kuever J, *et al.*** *Pseudochelatococcus lubricantis*
580 gen. nov., sp. nov. and *Pseudochelatococcus contaminans* sp. nov. from coolant lubricants. *Int J Syst*
581 *Evol Microbiol* 2015;65:147–153.
582 24. **Chen X, Li Q-Y, Li G-D, Lei H, Jiang Y, *et al.*** *Enterovirga rhinocerotis* gen. nov., sp. nov., isolated
583 from *Rhinoceros unicornis* faeces. *Antonie Van Leeuwenhoek* 2017;110:553–562.
584 25. **Garrity GM, Bell JA, Lilburn T**. Class I. *Alphaproteobacteria* class. nov. In: Brenner DJ, Krieg NR,
585 Staley JT, Garrity GM (editors). *Bergey's Manual of Systematic Bacteriology: The Proteobacteria, (Part*
586 *C)*. New York, NY: Springer; 2005. p. 1.
587 26. **Vasilyeva L, Grouzdev D, Koziaeva V, Berestovskaya Y, Novikov A, *et al.*** *Prosthecodimorpha*
588 *staleyi* gen. nov., sp. nov., novel prosthecate bacteria within the family *Ancalomicrobiaceae* and

589 reclassification of the polyphyletic genus *Prosthecomicrobium*. *Microbiology* 2022;91:479–488.

590 27. **Noh H-J, Baek K, Hwang CY, Shin SC, Hong SG, *et al.*** *Lichenihabitans psoromatis* gen. nov., sp.
591 nov., a member of a novel lineage (*Lichenihabitantaceae* fam. nov.) within the order of *Rhizobiales*
592 isolated from Antarctic lichen. *Int J Syst Evol Microbiol* 2019;69:3837–3842.

593 28. **Pankratov TA, Grouzdev DS, Patutina EO, Kolganova TV, Suzina NE, *et al.*** *Lichenibacterium*
594 *ramalinae* gen. nov, sp. nov., *Lichenibacterium minor* sp. nov., the first endophytic, beta-carotene
595 producing bacterial representatives from lichen thalli and the proposal of the new family
596 *Lichenibacteriaceae* within the order *Rhizobiales*. *Antonie Van Leeuwenhoek* 2020;113:477–489.

597 29. **Ma T, Xue H, Piao C, Jiang N, Li Y**. Phylogenomic reappraisal of the family *Rhizobiaceae* at the
598 genus and species levels, including the description of *Ectorhizobium quercum* gen. nov., sp. nov.
599 *Front Microbiol* 2023;14:1207256.

600 30. **Zou X, Li X, Wang X-M, Chen Q, Gao M, *et al.*** *Hansschlegelia beijingensis* sp. nov., an aerobic, pink-
601 pigmented, facultatively methylotrophic bacterium isolated from watermelon rhizosphere soil. *Int*
602 *J Syst Bacteriol* 2013;63:3715–3719.

603 31. **Yang L-Q, Liu L, Salam N, Xiao M, Kim C-J, *et al.*** *Chenggangzhangella methanolivorans* gen. nov.,
604 sp. nov., a member of the family *Methylocystaceae*, transfer of *Methylopila helvetica* Doronina et al.
605 2000 to *Albibacter helveticus* comb. nov. and emended description of the genus *Albibacter*. *Int J Syst*
606 *Bacteriol* 2016;66:2825–2830.

607 32. **Imhoff JF**. Transfer of *Rhodopseudomonas acidophila* to the new genus *Rhodoblastus* as
608 *Rhodoblastus acidophilus* gen. nov., comb. nov. *Int J Syst Evol Microbiol* 2001;51:1863–1866.

609 33. **Fukuda W, Yamada K, Miyoshi Y, Okuno H, Atomi H, *et al.*** *Rhodoligotrophos appendicifer* gen.
610 nov., sp. nov., an appendaged bacterium isolated from a freshwater Antarctic lake. *Int J Syst Evol*
611 *Microbiol* 2012;62:1945–1950.

612 34. **Kim Y-S, Jeon YJ, Kim K-H**. *Salaquimonas pukyongi* gen. nov., sp. nov., a novel bacterium within

613 the family *Phyllobacteriaceae*. *Int J Syst Evol Microbiol* 2019;69:3751–3756.

614 35. **Alton GG, Forsyth JRL**. *Brucella*. In: Baron S (editor). *Medical Microbiology*. Galveston (TX):
615 University of Texas Medical Branch at Galveston. http://www.ncbi.nlm.nih.gov/books/NBK8572/
616 (1996, accessed 21 October 2023).

617 36. **Holmes B, Popoff M, Kiredjian M, Kersters K**. *Ochrobactrum anthropi* gen. nov., sp. nov. from
618 human clinical specimens and previously known as group Vd. *Int J Syst Bacteriol* 1988;38:406–416.

619 37. **Brenner DJ, O'Connor SP, Winkler HH, Steigerwalt AG**. Proposals to unify the genera *Bartonella*
620 and *Rochalimaea*, with descriptions of *Bartonella quintana* comb. nov., *Bartonella vinsonii* comb.
621 nov., *Bartonella henselae* comb. nov., and *Bartonella elizabethae* comb. nov., and to remove the
622 family *Bartonellaceae* from the order *Rickettsiales*. *Int J Syst Bacteriol* 1993;43:777–786.

623 38. **Huang Z, Guo F, Lai Q, Shao Z**. *Notoacmeibacter marinus* gen. nov., sp. nov., isolated from the gut
624 of a limpet and proposal of *Notoacmeibacteraceae* fam. nov. in the order *Rhizobiales* of the class
625 *Alphaproteobacteria*. *Int J Syst Bacteriol* 2017;67:2527–2531.

626 39. **Willems A**. The Family *Phyllobacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E,
627 Thompson F (editors). *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp. 355–
628 418.

629 40. **Dedysh SN, Haupt ES, Dunfield PF**. Emended description of the family *Beijerinckiaceae* and
630 transfer of the genera *Chelatococcus* and *Camelimonas* to the family *Chelatococcaceae* fam. nov. *Int*
631 *J Syst Evol Microbiol* 2016;66:3177–3182.

632 41. **Proença DN, Whitman WB, Shapiro N, Woyke T, Kyrpides NC, *et al.*** *Faunimonas pinastri* gen.
633 nov., sp. nov., an endophyte from a pine tree of the family *Pleomorphomonadaceae*, class
634 *Alphaproteobacteria*. *Int J Syst Evol Microbiol*;72. Epub ahead of print December 2022. DOI:
635 10.1099/ijsem.0.005623.

636 42. **Conn HJ**. Taxonomic relationships of certain non-spore-forming rods in soil. *J Bacteriol*

637 1938;36:320–321.

638 43. **Kulichevskaya IS, Danilova OV, Tereshina VM, Kevbrin VV, Dedysh SN**. Descriptions of
639 *Roseiarcus fermentans* gen. nov., sp. nov., a bacteriochlorophyll *a*-containing fermentative
640 bacterium related phylogenetically to alphaproteobacterial methanotrophs, and of the family
641 *Roseiarcaceae* fam. nov. *Int J Syst Evol Microbiol* 2014;64:2558–2565.

642 44. **Akter S, Shazib SUA, Shin MK**. *Segnochrobactrum spirostomi* gen. nov., sp. nov., isolated from the
643 ciliate *Spirostomum yagiui* and description of a novel family, *Segnochrobactraceae* fam. nov. within
644 the order *Rhizobiales* of the class *Alphaproteobacteria*. *Int J Syst Evol Microbiol* 2020;70:1250–1258.

645

646 **FIGURE LEGENDS**

647 **Figure 1. Phylogenetic and core-proteome AAI (cpAA) analyses of the order *Hyphomicrobiales*.**
648 On the left, a maximum likelihood phylogeny of 133 *Hyphomicrobiales* type strains is shown, built using
649 the concatenated protein alignments encoded by the perc95_138 gene set (267 genes present in at
650 least 95% of the strains). The phylogeny was rooted using five *Caulobacterales* type strains as the
651 outgroup. The numbers on the nodes indicate the ultra-fast jackknife values using a 40% resampling
652 rate (top numbers) and the SH-aLRT support values (bottom numbers), both calculated from 1000
653 replicates. Values are only shown at nodes where at least one value is below 100. The scale bar
654 represents the average number of amino acid substitutions per site. To the right of the phylogeny is
655 the current family assignment of each of the 133 *Hyphomicrobiales* type strains, followed to the right
656 by the proposed family assignment of each strain. On the righthand side, a matrix is provided showing
657 the cpAAI values between each pair of strains calculated using the proteins encoded by the core_138
658 gene set (59 genes present in 100% of the strains). Values less than 75% are in white while all values
659 greater than 85% are the same shade of blue. Black boxes indicate the proposed families.

660

661 **Figure 2. Distribution of core-proteome AAI (cpAAI) comparisons of the order *Hyphomicrobiales*.**
662 Pairwise cpAAI values were calculated based on 59 nonrecombinant loci from the core genome of 133
663 members of the order *Hyphomicrobiales* and five members of the order *Caulobacterales*. Results are
664 summarized as histograms with a bin width of 1%. The cpAAI values calculated between two strains
665 belonging to different orders (yellow), different families but same order (blue), or the same family
666 (red) are summarized separately. Dashed vertical lines represent the mean value of each distribution.
667 In all plots, cpAAI values where both strains belong to the order *Caulobacterales* were excluded. (**A**)
668 The distribution of all pairwise cpAAI values with the classification (i.e., between order, between family,
669 or within family) based on existing taxonomic assignments. (**B**) The distribution of all pairwise cpAAI

670 values except for those including at least one strain from the family *Rhizobiaceae*, with the
671 classification based on existing taxonomic assignments. (**C**) The distribution of all pairwise cpAAI
672 values with the classification based on the proposed taxonomic assignments. (**D**) The distribution of
673 all pairwise cpAAI values except for those including at least one strain from the family *Rhizobiaceae*,
674 with the classification based on the proposed taxonomic assignments.

675 **Table 1.** Properties of the gene sets used for phylogenetic and overall genome similarity analyses.

	Gene set			
	core_143	perc95_143	core_138	perc95_138
Number of *Hyphomicrobiales* strains	138	138	133	133
Number of *Caulobacterales* strains	5	5	5	5
Genomes containing the genes (%)	100	≥95	100	≥95
Number of genes	19	256	59	267

676

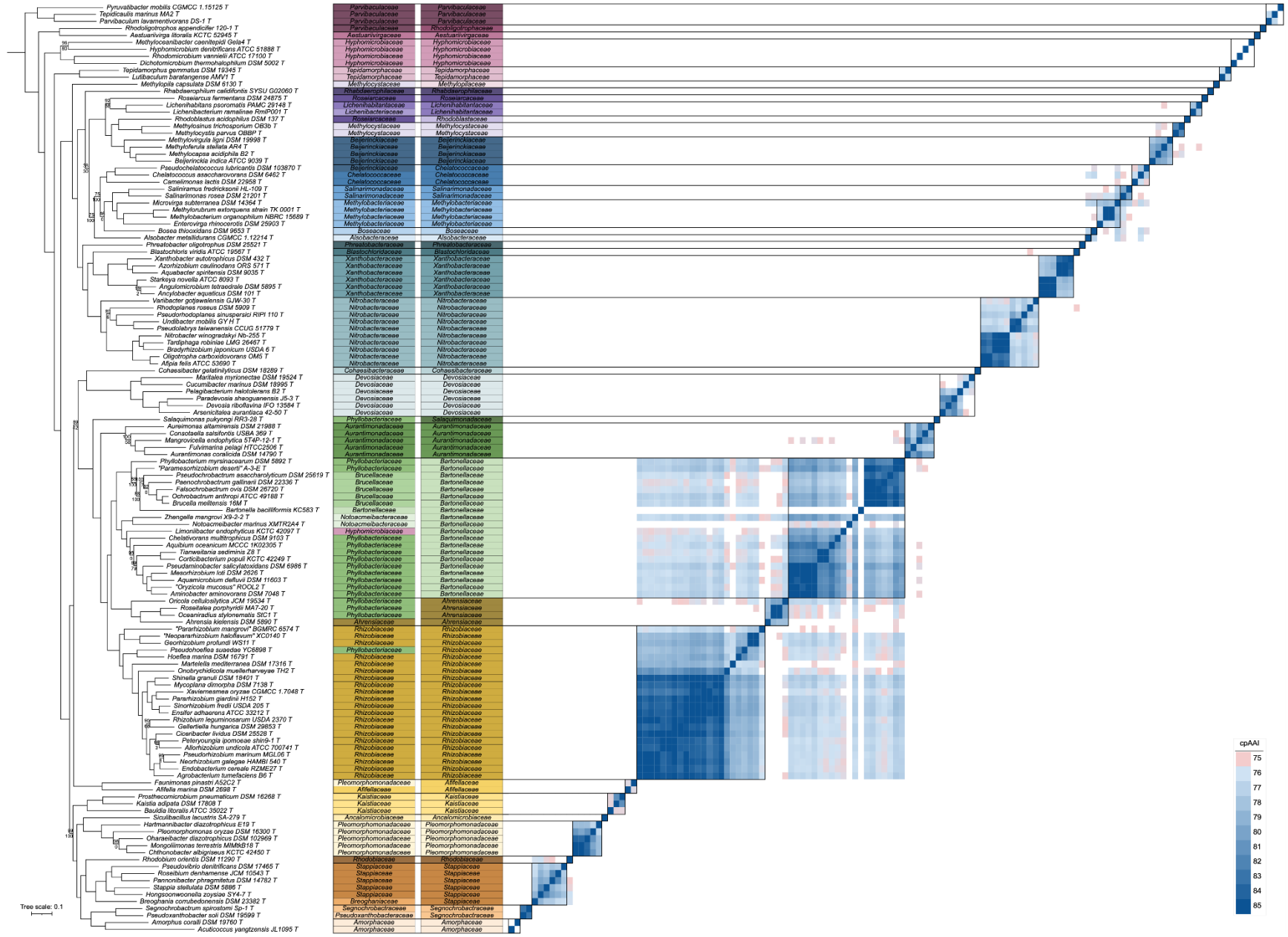

Figure 1. Phylogenetic and core-proteome AAI (cpAA) analyses of the order *Hyphomicrobiales*. On the left, a maximum likelihood phylogeny of 133 *Hyphomicrobiales* type strains is shown, built using the concatenated protein alignments encoded by the perc95_138 gene

680 set (267 genes present in at least 95% of the strains). The phylogeny was rooted using five *Caulobacterales* type strains as the outgroup. The
681 numbers on the nodes indicate the ultra-fast jackknife values using a 40% resampling rate (top numbers) and the SH-aLRT support values
682 (bottom numbers), both calculated from 1000 replicates. Values are only shown at nodes where at least one value is below 100. The scale
683 bar represents the average number of amino acid substitutions per site. To the right of the phylogeny is the current family assignment of
684 each of the 133 *Hyphomicrobiales* type strains, followed to the right by the proposed family assignment of each strain. On the righthand side,
685 a matrix is provided showing the cpAAI values between each pair of strains calculated using the proteins encoded by the core_138 gene set
686 (59 genes present in 100% of the strains). Values less than 75% are in white while all values greater than 85% are the same shade of blue.
687 Black boxes indicate the proposed families.

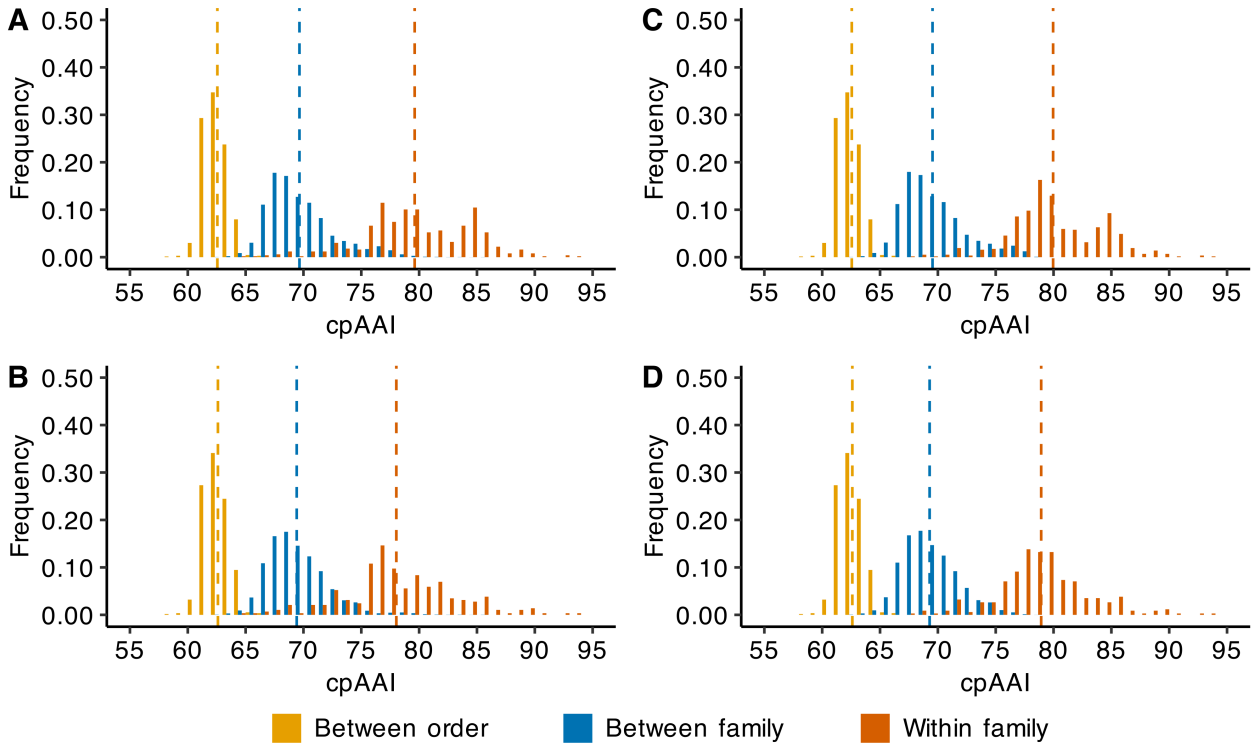

Figure 2. Distribution of core-proteome AAI (cpAAI) comparisons of the order *Hyphomicrobiales*. Pairwise cpAAI values were calculated based on 59 nonrecombinant loci from the core genome of 133 members of the order *Hyphomicrobiales* and five members of the order *Caulobacterales*. Results are summarized as histograms with a bin width of 1%. The cpAAI values calculated between two strains belonging to different orders (yellow), different families but same order (blue), or the same family (red) are summarized separately. Dashed vertical lines represent the mean value of each distribution. In all plots, cpAAI values where both strains belong to the order *Caulobacterales* were excluded. (**A**) The distribution of all pairwise cpAAI values with the classification (i.e., between order, between family, or within family) based on existing taxonomic assignments. (**B**) The distribution of all pairwise cpAAI values except for those including at least one strain from the family *Rhizobiaceae*, with the classification based on existing taxonomic assignments. (**C**) The distribution of all pairwise cpAAI values with the classification based on the proposed taxonomic assignments. (**D**) The distribution of all pairwise cpAAI values except for those including at least one strain from the family *Rhizobiaceae*, with the classification based on the proposed taxonomic assignments.