Supplementary File 4 - Phylogenomic Results

**Table of content**

* File S4 - Table 2. Summary table of CeMbio taxonomy
* Details result of Phylogenomic analysis
* File S4 - Table 2. Number of genomes/average genes and taxonomic level used for phylogenomic analysis
* Code used to create Phylogenomic tree
* Additional files: Full length phylogenomic tree
  + FileS4-Figure 5 – Sphingobacteria tree (BIGb0170)
  + FileS4-Figure 6 – Comamonas tree (BIGb0172)
  + FileS4-Figure 7 – Enterobacteria tree (JUb66, BIGb0393 and CeEnt1)
  + FileS4-Figure 8 – Stenotrophomonas tree (JUb19)
  + FileS4-Figure 9 – Chryseobacterium tree (JUb44)
  + FileS4-Figure 10 – Sphingomonas tree (JUb134)
  + FileS4-Figure 11 – Acinetobacter tree (MYb10)
  + FileS4-Figure 12 – Ochrobactrum tree (MYb71)
  + FileS4-Figure 13 – Pseudomonas tree (MSPm1 and MYb11)

**File S4 - Table 1. Summary CeMbio bacteria phylogenetic placement.**

|  |  |  |
| --- | --- | --- |
| **Bacteria** | **Verdict** | **Change compared to**  **16S rRNA phylogeny** |
| MYb71 | **NEW SPECIES** | **YES** |
| MYb10 | *Acinetobacter guillouiae* | No |
| MYb11 | *Pseudomonas lurida* | No |
| JUb19 | *Stenotrophomonas indicatrix* | **YES** |
| JUb44 | *Chryseobacterium scophtalnum* | **YES** |
| JUb66 | *Lelliottia amnigena* | No |
| JUb134 | *Sphingomonas molluscorum* | No |
| BIGb0393 | **NEW SPECIES** | **YES** |
| BIGb0170 | *Sphingobacteria multivorum* | **YES** |
| BIGb0172 | *Comamonas piscis* | No |
| CEent1 | *Enterobacter homarchei* | **YES** |
| MSPm1 | **NEW SPECIES** | **YES** |

**Details result of Phylogenomic analysis**

The Phylogenomic tree reconstruction were performed using GTOtree, using default parameters.

For 2 CeMbio bacteria, **BIGb0172** and **JUb134**, not enough published genomes were available to reconstruct a reliable phylogenomic tree. We then used the 16S rRNA phylogenies to classify those organisms as *Comamonas piscis* for BIGb0172 and *Sphingomonas molluscorum* for JUb134.

A screenshot of a cell phone

Description automatically generated

**File S4 - Figure1**: Partial approximately-maximum-likelihood phylogenetic trees from concatenated alignments of nucleotide of representative marker genes, using FastTree 2. Value on branches represent “Local support values” to estimate the reliability of each split in the tree (0 to 1, 1 being a high reliability). The label in red represent the CeMbio bacteria. In A, JUb134 is present twice, the label with the accession number GCF 004341505.1 is an earlier version of the assembly based on illumina sequencing technology.

For the bacteria MYb11, Myb10 and JUb66, phylogenomic reconstruction yielded similar trees as their respective 16S rRNA phylogenetic counterparts.

A screenshot of a cell phone

Description automatically generated

**File S4 – Figure 2:** Partial approximately-maximum-likelihood phylogenetic trees from concatenated alignments of nucleotide of representative marker genes, using FastTree 2. Value on branches represent “Local support values” to estimate the reliability of each split in the tree (0 to 1, 1 being a high reliability). The label in red represent the CeMbio bacteria. In B, MYb11 is present twice, the label with the accession number GCF 002966835.1 is an earlier version of the assembly based on illumina sequencing technology.

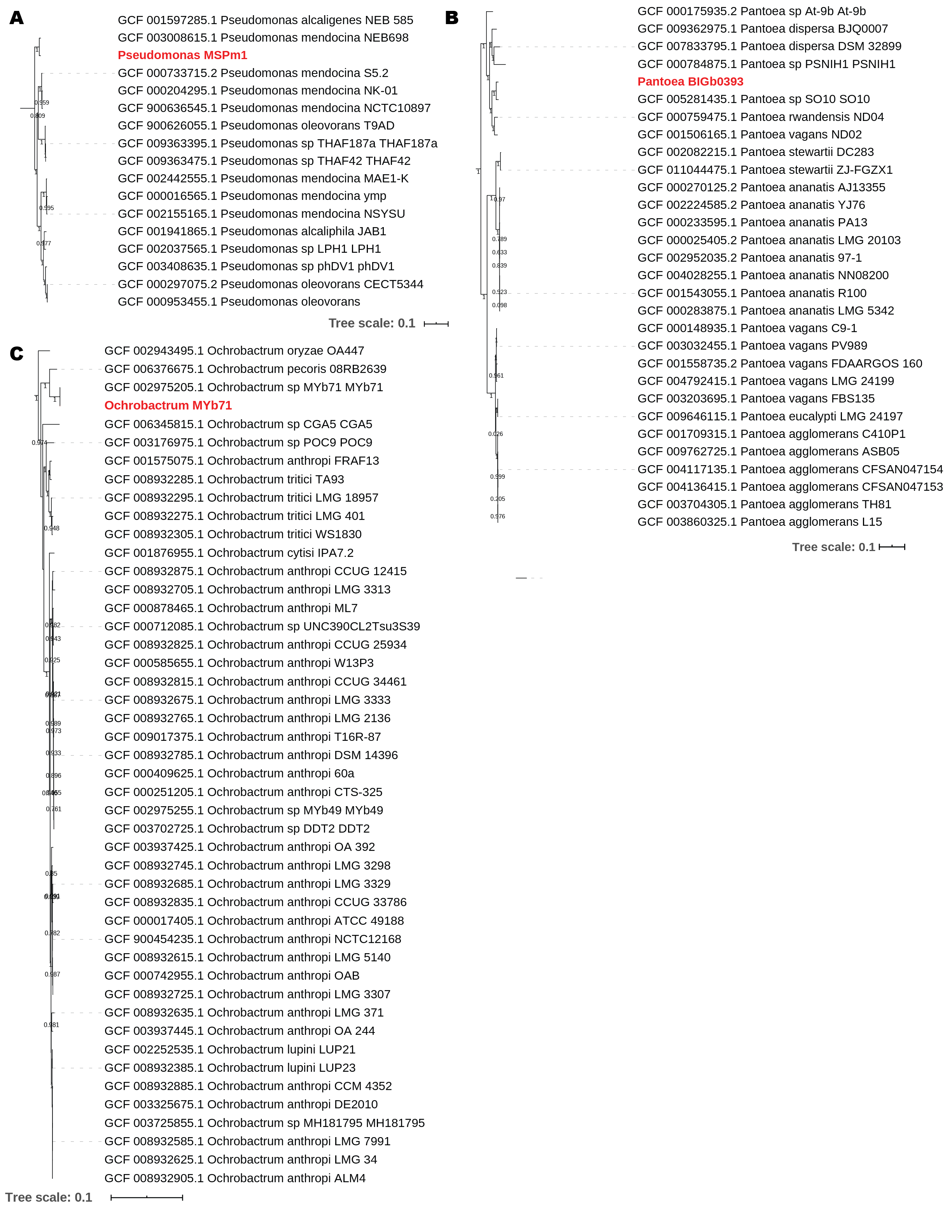
The bacteria CeEnt1, BIGb0170 and JUb19 could be attributed more accurately to a single species. In more details:

* For **CEent1**, 16S rRNA phylogeny indicates that the bacteria belongs to the *E. cloacae* “complex” and the phylogenomic reconstruction places it more specifically within the *E. hormaechei* clade. However, the genomes present in the *E. cloacae* and *E. homerchei* are highly similar, hence, to avoid further confusion we recommend keeping CeEnt1 defined as an *E. cloacae*.
* For **BIGb0170**, phylogenomic reconstruction places the bacteria close to the  *Sphingobacteria multivorum* species. This is confirmed with a 98% ANI value between the BIGb0170 genome and the wB15 reference.
* For **JUb44**, our 16S rRNA phylogeny places the bacteria on a separate *Chryseobacterium* branch, between *C. indoltheticum, C. balustinum* and *C. scophtalnum* and the phylogenomic reconstruction place this our genome close to *C. scophtalnum*. The ANI values between JUb44 and all *C. scophtalnum* is 93.7% placing it within the species border with the previously described *Chryseobacterium*.
* For **JUb19**, 16S rRNA phylogeny indicates that the bacteria belongs to the *S. maltophilia* “complex”. This complex is a polyphyletic group of 23 species-like clades. Phylogenomic reconstruction indicates that JUb19 is 99% similar to *Stenotrophomonas* *indicatrix.* Both JUb19 and S. indicatrix genomes have an 86% ANI with the reference S. maltophilia, hence reinforcing that the two clades are different Stenotrophomonas species. In addition, studies have shown that the *S. maltophilia* complex regroups different clades of bacteria that can be phylogenetically associated to either environmental or clinical samples and both JUb19 and *S. indicatrix* belong to a clade dominated by environmental isolates. Even if unlikely, we can’t exclude the possibility of opportunistic potential of bacteria present in the *S. indicatrix* clade and bacteria should be handled with care.



**File S4 – Figure 3:** Partial approximately-maximum-likelihood phylogenetic trees from concatenated alignments of nucleotide of representative marker genes, using FastTree 2. Value on branches represent “Local support values” to estimate the reliability of each split in the tree (0 to 1, 1 being a high reliability). The label in red represent the CeMbio bacteria.

Finally, 4 of the CeMbio bacteria can be described as new species: BIGb0393, MYb71, JUb44 and MSPm1. Respectively those bacteria are new species of: *Pantoea, Ochrobactrum, Chryseobacterium* and *Pseudomonas.*

* Although the phylogenomic reconstruction of **MYb71** agrees with the 16S rRNA one and places the bacteria close to Ochrobactrum pecoris, there is a 90% ANI value between the two bacteria. This indicates that MYb71 is likely a different species from *O. pecoris,* and the lack of closely related 16S rRNA sequence suggest that MYb71 is a new *Ochrobactrum* species.
* For **MSPm1**, the phylogeny is similar to the 16S rRNA one, MSPm1 branch out of the same *Pseudomonas* clade, including *P. mendocia* and *P. oleovorans*. The phylogenomic tree shows more precisely that MSPm1 branches out from the *P. mendocina* clade. The original *P. mendocina* referenceis *P. mendocina* NCTC 10897 submitted in 1970. In contrast the closely related genome *P. mendocina* NEB698, has been deposited by New England Bioscience and is now referenced on ATCC has a *P. pseudoalcaligenes*. This change in classification is correct as NEB698 and NCTC 10897 have an ANI of 86%. Similarly, MSPm1 shares an 86% ANI with *P. mendocina* NCTC 10897 and in consequence does not belong to the *mendocina* species*.* It is also not a *P. pseudoalcaligenes* as it has an ANI of 93.7%, just below the 94-96% species value cutoff. On the 16S rRNA level, they all share very similar sequences, ranging from 98 to 99% between MSPm1 and *P. mendocia*, *P. oleovorans*, *P. alaliphila* and *P. sihuiensi.* But all of those genomes share an ANI value below 87%. Thus, MSPm1 could be described as a new species.
* Finally, **BIGb0393**, the closest named relative based on phylogenomic is *P. rwandensis*, followed by *P. vagans* and *P. dispersa*. All ANI values are below 85%, in consequence BIGb0393 do not belong to any of those species. Comparing the phylogenomic to the 16S rRNA phylogeny, the closest 16S rRNA sequence (with 98% similarity) is *P. stewartii* subsp. *stewartii* DC283 which has an ANI value of 79.6%. Taking all those results together BIGb0393 is likely a novel species of *Pantoea*.
* Finally, for 

**File S4 – Figure 4:** Partial approximately-maximum-likelihood phylogenetic trees from concatenated alignments of nucleotide of representative marker genes, using FastTree 2. Value on branches represent “Local support values” to estimate the reliability of each split in the tree (0 to 1, 1 being a high reliability). The label in red represent the CeMbio bacteria. In B, MY71 is present twice, the label with the accession number GCF .0.03975205.1 is an earlier version of the assembly based on illumina sequencing technology.

**File S4 - Table 2. Summary of number of genomes, average genes used and taxonomy level used for comprehensive phylogenomic reconstructions.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Bacteria** | **Genus** | **# of genomes used** | **Average - min/max  of genes used** | **Taxonomy level used for reference** |
| BIGb0393 | Pantoea | 246 | 156 - 21/172 | Genera |
| JUb66 | Lelliottia | 246 | 156 - 21/172 | Genera |
| CEent1 | Enterobacter | 246 | 156 - 21/172 | Genera |
| MSPm1 | Pseudomonas | 594 | 167 - 118/172 | Genera |
| MYb11 | Pseudomonas | 594 | 167 - 118/172 | Genera |
| MYb71 | Ochrobactrum | 895 | 115 - 76/117 | Family |
| MYb10 | Acinetobacter | 326 | 164 - 115/169 | Family |
| BIGb0170 | Sphingobacteria | 271 | 87 - 12/90 | Family |
| BIGb0172 | Comamonas | 498 | 189 - 10/200 | Family |
| JUb134 | Sphingomonas | 322 | 113 - 92/117 | Family |
| JUb19 | Stenotrophomonas | 527 | 164 - 34/169 | Genera |
| JUb44 | Chryseobacter | 275 | 87 - 62/80 | Family |

**Code for Phylogenomic reconstruction**





