**ASVs to functional predictions**

Some possible pathways for assigning amplicon sequence variants (ASVs) to functional groups:

**1. PICRUSt2**

PICRUSt2 places ASV sequences in a phylogenetic tree and predicts gene content and metabolic pathway abundance based on this placement. Pro: easy to run with QIIME2 outputs. Con: many layers of inference to go from ASV to function.

Paper: <https://www.nature.com/articles/s41587-020-0548-6>

User Manual and Tutorial: <https://huttenhower.sph.harvard.edu/picrust/>

There is also a QIIME2 Plugin, though it does not produce all the output tables that the original software does: <https://library.qiime2.org/plugins/q2-picrust2/13/>

In the output, there should be a pathways\_out folder containing a path\_abun\_predictions.tsv. This will include some predicted abundances for several BioCyc metabolic pathways in each genome predicted for an ASV. Two pathways that may be of interest are:

METHANOGENESIS-PWY (methanogenesis from H2 and CO2)

<https://metacyc.org/pathway?orgid=META&id=METHANOGENESIS-PWY>

METH-ACETATE-PWY (methanogenesis from acetate)

<https://biocyc.org/pathway?orgid=META&id=METH-ACETATE-PWY>

The output should also have a table EC\_predicted.tsv, which has a count for each predicted Enzyme Commission number (<https://www.rcsb.org/search/browse/ec>) per ASV. If you know the EC numbers of key enzymes for your pathways of interest, then this output could be useful.

**2. Taxonomy-based filter for conserved traits**

For some traits that are well-conserved within a taxonomic group, it may be reasonable to filter based on the taxonomy table for ASVs and select those belonging to a genera, family, etc. known to have the function of interest.

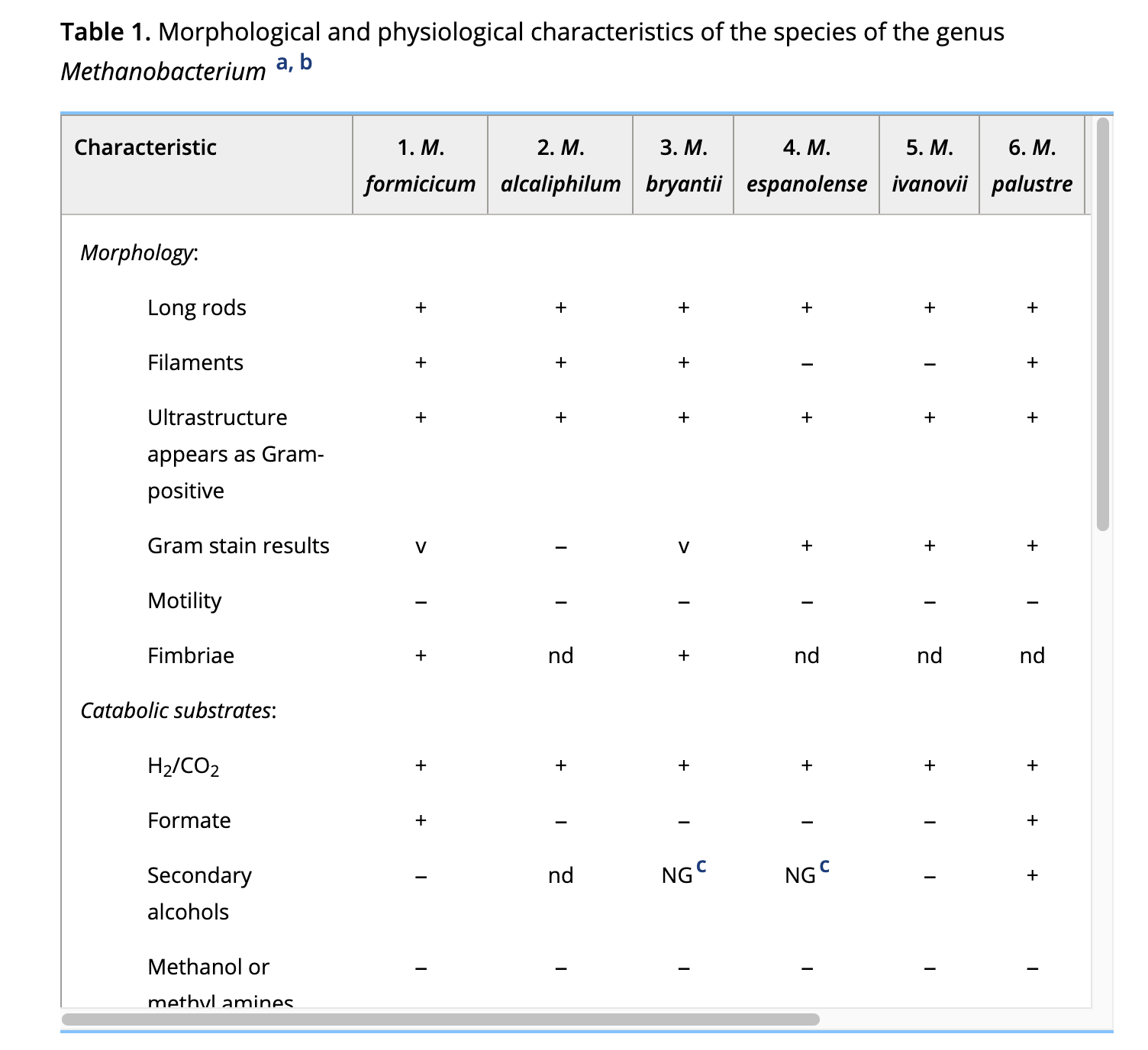
To create a custom filter for taxa with conserved metabolisms, I use a resource like The Prokaryotes (<https://www.springer.com/series/11652?srsltid=AfmBOorB1Y86W-wahywRoe_1V99bbeo4qbJyXaT1f1PAqhq6yahEGStc>) or its online equivalent, Bergey’s Manual of Systematics of Archae and Bacteria. The page for Methanobacterium, for example, will include a summary of its general metabolic features and a table like the one below showing the catabolic substrates for different species in the genus. Further below, there will be species-level entries that sometimes get into finer detail about the metabolism.

**Bergey’s Manual of Systematics of Archae and Bacteria**

Origninally published September 2015 and maintained online:

<https://onlinelibrary.wiley.com/doi/book/10.1002/9781118960608>

\*full text of entries is behind paywall



Much of this data has also been transferred to the BacDive resource and some of it may be more easily searchable there.

**BacDive**

<https://bacdive.dsmz.de/>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC8728306/>

If using this kind of approach, note that the names of taxa are under development. For some background, see here:

Sanford, Robert A., Karen G. Lloyd, Konstantinos T. Konstantinidis, and Frank E. Löffler. "Microbial taxonomy run amok." *Trends in Microbiology* 29, no. 5 (2021): 394-404.

<https://www.cell.com/trends/microbiology/abstract/S0966-842X(20)30327-9>

This means you may have to reconcile the names provided by a version of a SILVA taxonomic classifier with the names as they appear in other resources. Some taxa, for which the only isolate material is metagenomes, may not appear in all resources. Other taxa may appear under different families or genera across different databases.

**SILVA**

<https://www.arb-silva.de/search/>

**GTDB**

<https://gtdb.ecogenomic.org/tree>

**SeqCode Registry (for new taxa identified by metagenome sequencing only)**

<https://registry.seqco.de/>

**List of Prokaryotic names with Standing in Nomenclature (includes taxa named based on culture or metagenome)**

<https://lpsn.dsmz.de/>

**Ribosomal Database Project (16S)**

Originally published by Michigan State University

<https://www.canr.msu.edu/cme/resources>

Maintained by the Quensen group

<https://john-quensen.com/tutorials/tutorial-1/>

Classifier here:

<https://sourceforge.net/projects/rdp-classifier/>

**3. RT-qPCR**

One way to get at function is to quantify transcripts of key methanogenesis genes. If this profiling is done on individual isolates in parallel with 16S sequencing, then it may be possible to link ASVs with functional activity. Pro: not much ambiguity. Con: culturing isolates from the system may be hard. I’m not sure if there would be a way to link functional gene counts to ASVs in community-level data.

**4. Metagenome data**

Long-read sequences, or long and short reads together, are generating pretty complete and high-quality metagenomes these days. With adequate sequencing depth, metagenomes could be linked to ASVs via their 16S sequences and functional annotations could be mined for methanogen-related functions.