Functional profiling with Picrust2

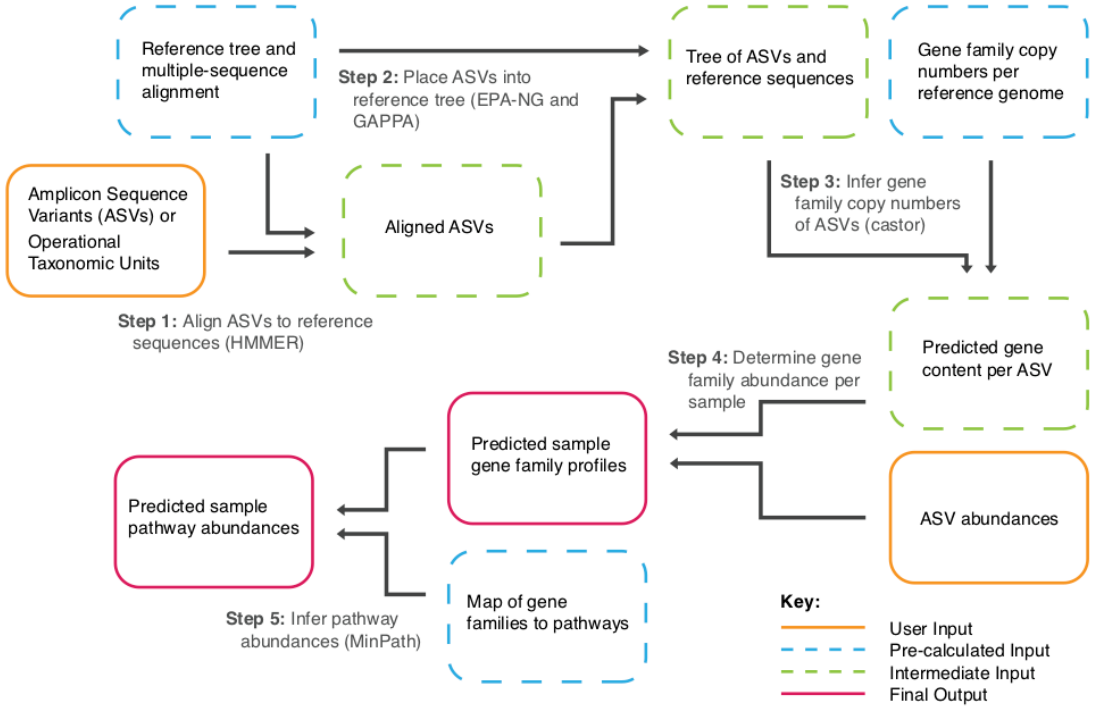
Dr. Thomas Keller

Two large questions with 16s

- Who's there
  - Taxonomic profiling (abundance analysis, broadly speaking)
- What are they doing?
  - Functional profiling – extrapolated from marker gene data
    - Main caveat to always keep in mind is that there are several estimations here,
    - Gold standard would be metagenomics to actually assemble metagenomes to more clearly define pathways and abundances

Broad overview of algorithm

- Takes OTUs/ASVs and maps them to a reference 16s with sequenced genomes
- Predict copy number per ASV/OTU
- Predict abundance per sample
- Map gene families to pathways



Picrust runs on the commandline/cluster

- We don’t have time to delve into linux for this workshop, but I’ll show the commands to run Picrust2 for those that want to try this and have a little familiarity with the command line
- The data files provided are the output files from a Picrust2 run

What does picrust2 actually output?

- Stratified and unstratified data
  - Stratified splits out ASVs by the EC/pathway abundances, much larger tables!
  - Potentially more informative
  - Unstratified rolls the ASVs up into one abundance per EC/pathway/KEGG
- EC (enzyme commission numbers –metacyc)
- Pathways from metacyc
- Kegg ortholog (KO) numbers

Limitations

- Amplicon prediction will hopefully be correlated with shotgun metagenomics, but the functions predicted might vary substantially
- Accuracy of the sample depends on the reference genomes
  - Also, 16s doesn’t generally have resolution to separate strains within species
  - Certain environments better represented (human gut vs ocean)
- Full list:
  - https://github.com/picrust/picrust2/wiki/Key-Limitations

Interpreting Picrust2 data with Aldex2

- The authors of Picrust2 recommend packages like Aldex2 for downstream statistical analysis of the output'
- Gui alternative – STAMP
  - https://beikolab.cs.dal.ca/software/STAMP
- These packages take into account the fact that the data are compositional (relative) based
- Otherwise, the analyses and output very similar to rna-seq differential analyses using DESEQ2 or edgeR