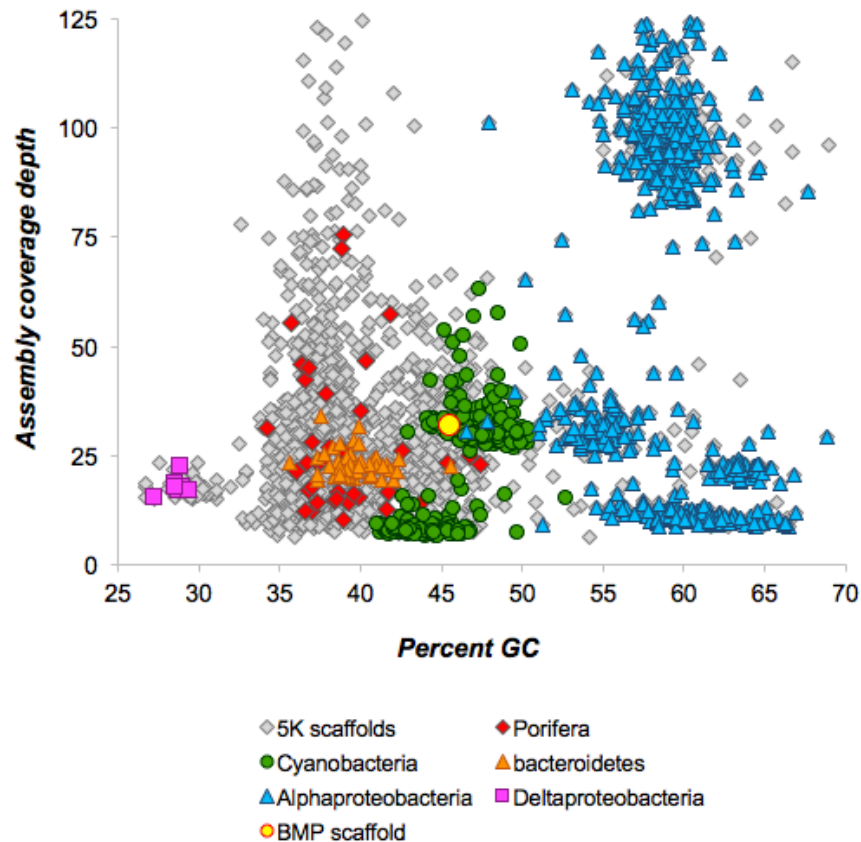


# Simple methods for phylogenetic binning of assembled metagenomic scaffolds



## When/why?

- Cluster/classify hundreds to thousands of small contigs or scaffolds (5K-100K)
- Complex sample mixture (Proks, Eukes, Viruses, plasmids, weird taxa w/ no relatives)
- Flexible granularity choices (contig sizes, axis scales, phylogenetic levels)
- Ability to pinpoint & retrieve individual sequences

## How?

- Input = contig or scf fasta
- Length, %GC tabfile by perl script
- Coverage depth: bowtie/bam/sam (or parse Spades contig name)
- Orf calling: EMBOSS transeq
- Filter: transeq\_to\_multifasta perl script
- Match taxa: Diamond vs Genbank + DarkHorse
- Tally matches in DarkHorse table by perl script
- Plot output table: Excel, R, Matlab, or whatever