

pre-process.SR.fasta + .config.bash

Mg-traits.bash

- 0 - Parse parameters
- 1 - Database connection
- 2 - Download file from MG URL
- 3 - Validate file
- 4 - Create job directory
- 5 - Check for utilities and directories
- 6 - Download data files from SVN (pfam, tffile, SLV)
- 7 - Check for duplicates
- 8 - Calculate sequence statistics: Number of bases, GC content, GC variance

qsub finish_runner

- 1-3. Check number of sub-jobs, genes and 16S sequences
4. Functional annotation: UProC and Pfam
5. Create functional table
6. Compute codon usage
7. Create codon and amino acid usage table
- 8-9. Compute nucleotide and dinucleotide composition
10. Create dinucleotide table
- 11-13. Get taxonomic annotation from SINA
- 14-18. Insert aa, codon, dinuc, taxnonmy, pfam to mg_traits_jobs table
19. Insert simple traits mg_traits_jobs table
20. PCAs for aa, codon, dinuc, function and taxonomy
21. Load PCA data in the DB

01-environment + 01-raw.fasta

Split data: 05-part.fasta
qsub fgs_runner

05-part.faa

sortmerna

01-sormerna.fasta

Split data: 06-part.fasta
qsub sina_runner

06-part-16S.classify.fasta

.config.finish.bash

mg_traits_jobs

