

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

## Mg-traits.bash

0 - Parse parameters  
1 - Database connection  
2 - Download file from MG URL  
3 - Validate fasta file  
4 - Create job directory  
5 - Check for utilities and directories  
6 - Check for duplicates  
7 - Calculate sequence statistics: Number of bases, GC content, GC variance

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-raw.fasta

02-unique.fasta

04-mg\_stats

Split data

05-part-N.fasta

fgs

smrna

06-part-N.fasta

sina

05-part.faa

06-part-N.classify

Functional annotation:  
UProC and Pfam

Check point

mg\_traits