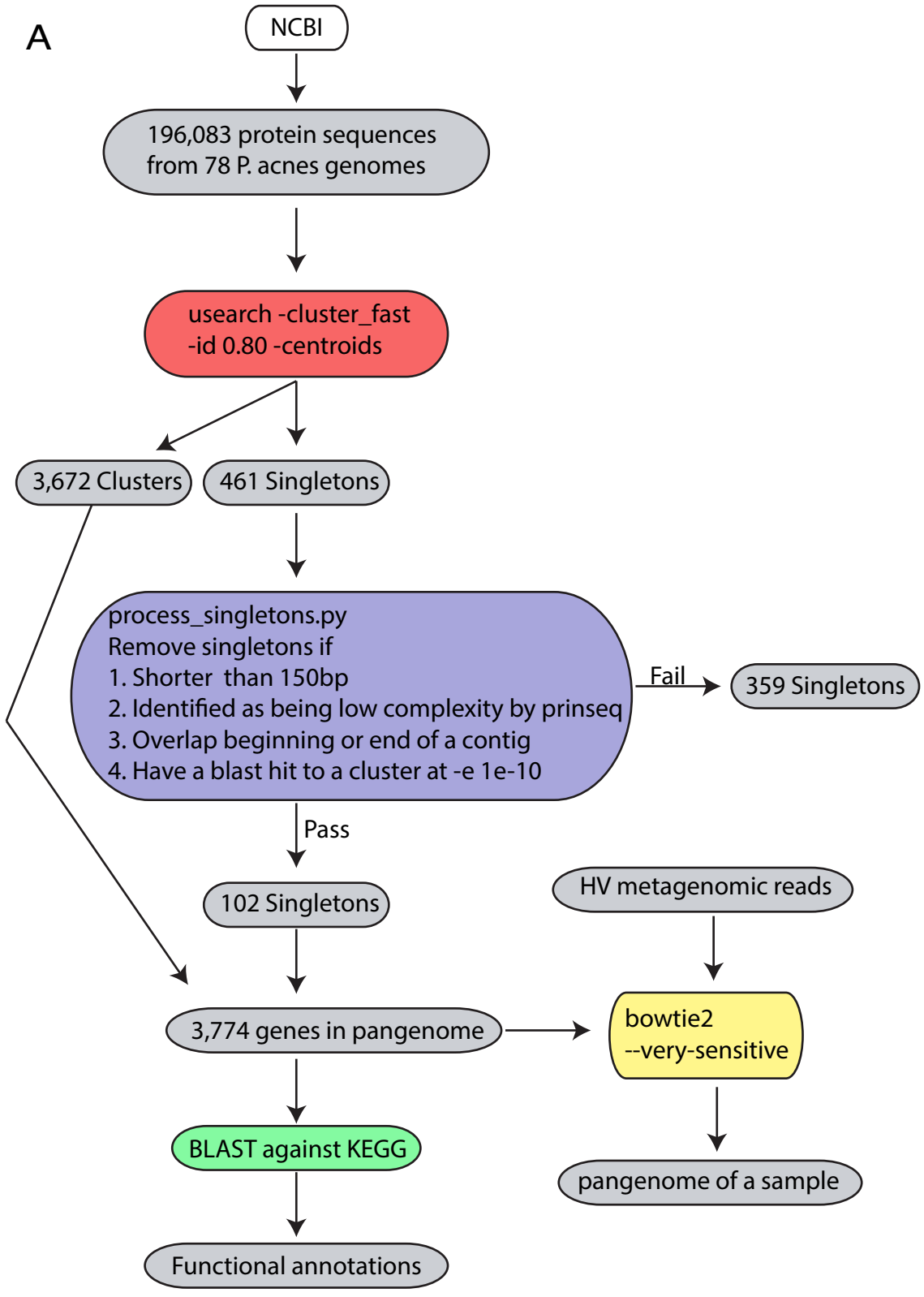
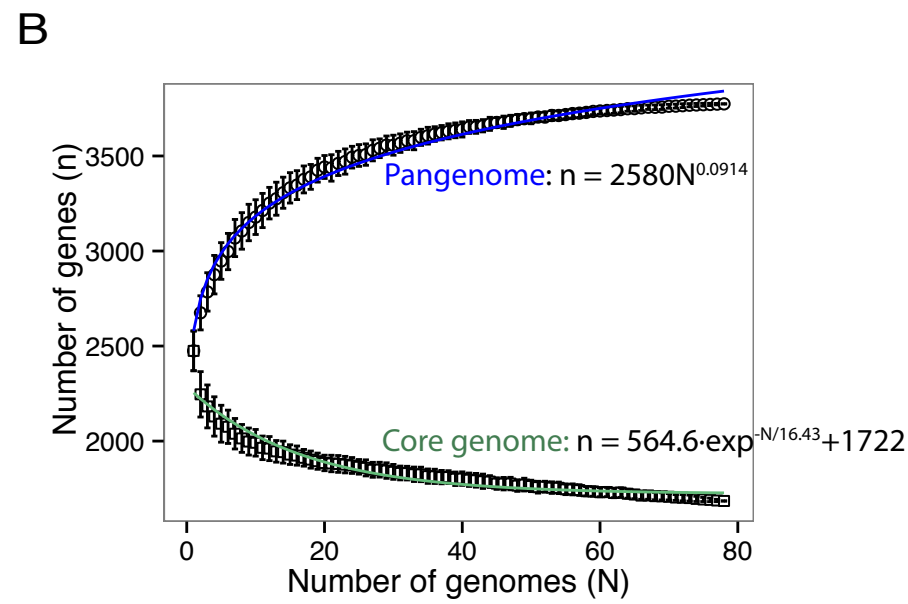
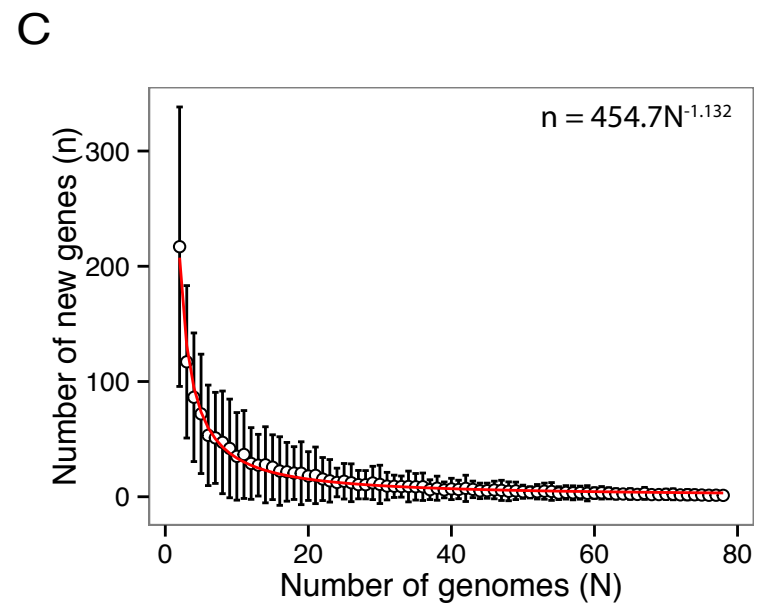
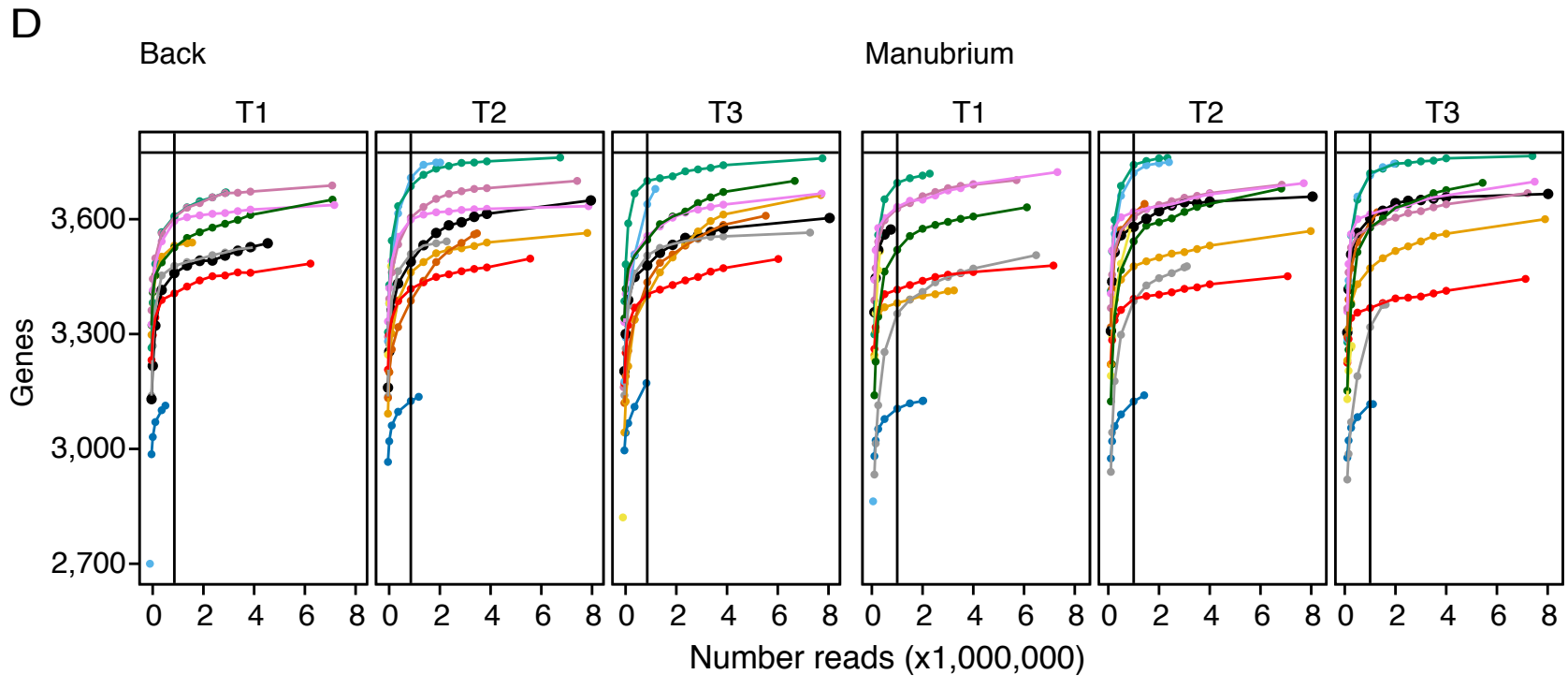

A
NCBI
196,083 protein sequences from 78 P. acnes genomes
usearch -cluster_fast -id 0.80 -centroids
3,672 Clusters
461 Singletons
process_singletons.py
Remove singletons if
1. Shorter than 150bp
2. Identified as being low complexity by prinseq
3. Overlap beginning or end of a contig
4. Have a blast hit to a cluster at -e 1e-10
Fail
359 Singletons
Pass
102 Singletons
HV metagenomic reads
3,774 genes in pangenome
bowtie2 --very-sensitive
BLAST against KEGG
pangenome of a sample
Functional annotations


B
Pangenome: n = 2580N^0.0914
Core genome: n = 564.6·exp^-N/16.43+1722
Number of genes (n)
Number of genomes (N)


C
n = 454.7N^-1.132
Number of new genes (n)
Number of genomes (N)


D
Back
Manubrium
T1
T2
T3
Genes
Number reads (x1,000,000)