##Process and assemble all samples separate, pool contigs and bin with BinSanity

srun cat contigs/* > all.contigs.fa

##Rename so no double contig names

srun anvi-script-reformat-fasta all.contigs.fa -o all.BSF.contigs.renamed.fa -l 0 --simplify-names --report-file all.BSF.contig_names_map.tsv &

##Run BinSanity Commands

srun get-ids -f ./ -l all.BSF.contigs.renamed.fa -o Contigs.ids

srun -p highmem --mem 240G --error binProfile.err --output binProfile.out Binsanity-profile -i all.BSF.contigs.renamed.fa -s all.BSF.mapped.sorted.filt.bam --ids Contigs.ids -c Contigs.x100.lognorm --transform scale &

Run Binsanity-lc (relatively low-memory version of binsanity)

srun -p highmem --mem 240G --cpus-per-task 12 --output binLC.out --error binLC.err Binsanity-lc --threads 20 -x 3000 -f ./ -l all.BSF.contigs.renamed.fa -c Contigs.x100.lognorm.cov -o binsanity_all_BSF &

srun -p highmem --mem 100G --ntasks 1 --cpus-per-task 12 checkm lineage_wf -f checkm_table_BSF_all.tsv --tab_table -t 12 --pplacer_threads 12 -x fna ./ checkm_results_BSF_all 2> checkm.err.log &

##Predict proteins using Prodigal Bash Script

prodigal.sh

```
#! /bin/sh

#SBATCH --output prodigal.out
#SBATCH --error prodigal.err
#SBATCH --ntasks 1
#SBATCH --cpus-per-task 1
#SBATCH --partition highmem
#SBATCH --mem 100G

bins=("BinSanityLC-kmean-bin_13-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_14-bin_0-
```

refined_6.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_9.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_10.fna" "BinSanityLC-kmean-bin_61-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_18-bin_0-refined_0.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_15.fna" "BinSanityLC-kmean-bin_68-bin_0-refined_18.fna" "BinSanityLC-kmean-bin_24-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_73-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_24-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_85-bin_0.fna" "BinSanityLC-kmean-bin_28-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_86-bin_0-refined_11.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_55-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_20.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_36-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_6.fna")
for b in "${bins[@]}"; do
    prodigal -q -p meta -f gff -i $b -o $b.gff -a $b.faa
done

Annotate Bins using Diamond and KEGG T10000 db
annotate.T10000.sh

#! /bin/sh

#SBATCH --output diamond.T10000out
#SBATCH --error diamond.T10000err
#SBATCH --ntasks 1
#SBATCH --cpus-per-task 5
#SBATCH --partition batch
#SBATCH --mem-per-cpu 10G

bins=("BinSanityLC-kmean-bin_13-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_9.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_10.fna" "BinSanityLC-kmean-bin_61-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_18-bin_0-refined_0.fna" "BinSanityLC-kmean-

bin_45-bin_0-refined_15.fna" "BinSanityLC-kmean-bin_68-bin_0-refined_18.fna"
"BinSanityLC-kmean-bin_24-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_45-bin_0-
refined_6.fna" "BinSanityLC-kmean-bin_73-bin_0-refined_4.fna" "BinSanityLC-kmean-
bin_24-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_1.fna"
"BinSanityLC-kmean-bin_85-bin_0.fna" "BinSanityLC-kmean-bin_28-bin_0-
refined_2.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_2.fna" "BinSanityLC-kmean-
bin_86-bin_0-refined_11.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_2.fna"
"BinSanityLC-kmean-bin_55-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_92-bin_0-
refined_3.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_4.fna" "BinSanityLC-kmean-
bin_57-bin_0-refined_20.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_5.fna"
"BinSanityLC-kmean-bin_36-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_57-bin_0-
refined_3.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_6.fna")
for b in "${bins[@]}"; do
    diamond blastp --threads 5 --tmpdir /tmp --sensitive --evalue 0.000001 --outfmt 6 --out
$b.KO.b6 --db /srv/databases/internal/diamond/KEGG-T10000.dmnd --query
faas/$b.faa
done

Annotate Bins using Diamond and KEGG Prokaryotes db
annotate.proks.sh

#! /bin/sh

#SBATCH --output diamond.proks.out
#SBATCH --error diamond.proks.err
#SBATCH --ntasks 1
#SBATCH --cpus-per-task 5
#SBATCH --partition batch
#SBATCH --mem-per-cpu 5G

bins=("BinSanityLC-kmean-bin_13-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_44-
bin_0-refined_1.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_4.fna" "BinSanityLC-
kmean-bin_14-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_5.fna"
"BinSanityLC-kmean-bin_57-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_14-bin_0-
refined_6.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_9.fna" "BinSanityLC-kmean-
bin_57-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_8.fna"
"BinSanityLC-kmean-bin_45-bin_0-refined_10.fna" "BinSanityLC-kmean-bin_61-bin_0-
refined_12.fna" "BinSanityLC-kmean-bin_18-bin_0-refined_0.fna" "BinSanityLC-kmean-
bin_45-bin_0-refined_15.fna" "BinSanityLC-kmean-bin_68-bin_0-refined_18.fna"
"BinSanityLC-kmean-bin_24-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_45-bin_0-

refined_6.fna" "BinSanityLC-kmean-bin_73-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_24-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_85-bin_0.fna" "BinSanityLC-kmean-bin_28-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_86-bin_0-refined_11.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_55-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_20.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_36-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_6.fna")
for b in "${bins[@]}"; do

    diamond blastp --threads 5 --tmpdir /tmp --sensitive --evalue 0.000001 --outfmt 6 --out $b.KO.prok.b6 --db /srv/databases/internal/diamond/KEGG-prokaryotes.dmnd --query faas/$b.faa
done

Clean gff for annotate_features
bins=("BinSanityLC-kmean-bin_13-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_9.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_10.fna" "BinSanityLC-kmean-bin_61-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_18-bin_0-refined_0.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_15.fna" "BinSanityLC-kmean-bin_68-bin_0-refined_18.fna" "BinSanityLC-kmean-bin_24-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_45-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_73-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_24-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_85-bin_0.fna" "BinSanityLC-kmean-bin_28-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_86-bin_0-refined_11.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_2.fna" "BinSanityLC-kmean-bin_55-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_4.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_20.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_36-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_6.fna")

for b in "${bins[@]}"; do

```
    awk -F'\t' '/^#/{print $0; next} {sub(/ID\=[^_]*_/, "ID="$1"_", $9); print
$1"\t"$2"\t"$3"\t"$4"\t"$5"\t"$6"\t"$7"\t"$8"\t"$9}' $b.gff > $b.clean.gff
done
```

##Combine Diamond b6 files

```
for b in "${bins[@]}"; do
    srun cat ${b}.KO.b6 ${b}.KO.prok.b6 > ${b}.all.b6
done
```

##Annotate the .gff files using annotate_features.py

```
bins=("BinSanityLC-kmean-bin_13-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_44-
bin_0-refined_1.fna" "BinSanityLC-kmean-bin_57-bin_0-refined_4.fna" "BinSanityLC-
kmean-bin_14-bin_0-refined_12.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_5.fna"
"BinSanityLC-kmean-bin_57-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_14-bin_0-
refined_6.fna" "BinSanityLC-kmean-bin_44-bin_0-refined_9.fna" "BinSanityLC-kmean-
bin_57-bin_0-refined_8.fna" "BinSanityLC-kmean-bin_14-bin_0-refined_8.fna"
"BinSanityLC-kmean-bin_45-bin_0-refined_10.fna" "BinSanityLC-kmean-bin_61-bin_0-
refined_12.fna" "BinSanityLC-kmean-bin_18-bin_0-refined_0.fna" "BinSanityLC-kmean-
bin_45-bin_0-refined_15.fna" "BinSanityLC-kmean-bin_68-bin_0-refined_18.fna"
"BinSanityLC-kmean-bin_24-bin_0-refined_1.fna" "BinSanityLC-kmean-bin_45-bin_0-
refined_6.fna" "BinSanityLC-kmean-bin_73-bin_0-refined_4.fna" "BinSanityLC-kmean-
bin_24-bin_0-refined_5.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_1.fna"
"BinSanityLC-kmean-bin_85-bin_0.fna" "BinSanityLC-kmean-bin_28-bin_0-
refined_2.fna" "BinSanityLC-kmean-bin_53-bin_0-refined_2.fna" "BinSanityLC-kmean-
bin_86-bin_0-refined_11.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_2.fna"
"BinSanityLC-kmean-bin_55-bin_0-refined_6.fna" "BinSanityLC-kmean-bin_92-bin_0-
refined_3.fna" "BinSanityLC-kmean-bin_35-bin_0-refined_4.fna" "BinSanityLC-kmean-
bin_57-bin_0-refined_20.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_5.fna"
"BinSanityLC-kmean-bin_36-bin_0-refined_3.fna" "BinSanityLC-kmean-bin_57-bin_0-
refined_3.fna" "BinSanityLC-kmean-bin_92-bin_0-refined_6.fna")

for b in "${bins[@]}"; do
    srun -p highmem --mem 40G annotate_features --type CDS --mapping
/srv/databases/internal/json/KEGG-
T10000.mapping.json,/srv/databases/internal/json/KEGG-prokaryotes.mapping.json --
fields gene,gene_family,product,organism,database,md5,Ontology_term,dummy --
conflict quality --out ${b}.annotated.gff -b diamond.out/combined/${b}.all.b6
gffs/${b}.clean.gff > ${b}.annotate.log &
done
```

##Explore bins with kegg mapper, make input files

```
for b in "${bins[@]}"; do
    srun ~/myScripts/genesKO_from_gff.py ${b}.annotated.gff ${b}.forMapping.txt &
done
```