How to Run MetaBAT

Workflow

- 1. Run MetaBAT and get bins
- 2. Use CheckM to evaluate those bins
-
-

dataset

- CAMI Low Complexity

- 19499 contigs

- Average 7972

- Standard deviation 45485

- Max 1684314

- Min 150

Install on Proteus

- Copy file 'metabat2' to Proteus

- or

- $ bash install_on_proteus.sh

- This will create a 'MetaBAT' folder in home directory.

Install on other environments

- https://bitbucket.org/berkeleylab/metabat/src/master/
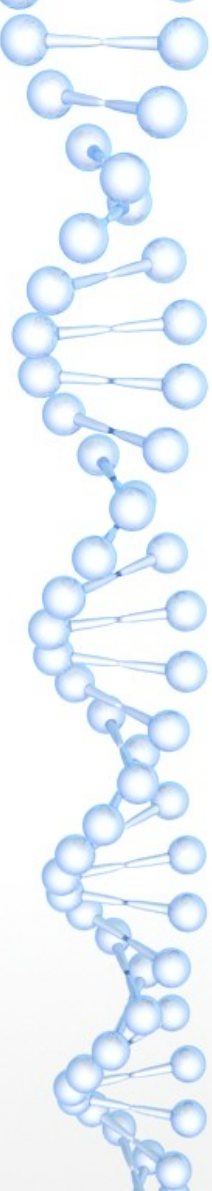
./metabat2

- -i   <input file> (required) .fasta or .fasta.gz
- -o *<bin folder>*/<bin header> (required)
- -m (=2500) minimum size of a contig for binning
- --maxP arg (=95)   Percentage of 'good' contigs considered for binning The greater, the more sensitive.
- --minS arg (=60)   Minimum score of a edge for binning (should be between 1 and 99). The greater, the

more

- --maxEdges arg (=200)   Maximum number of edges per node. The greater, the more sensitive.

• Bins in .fasta or .fa format
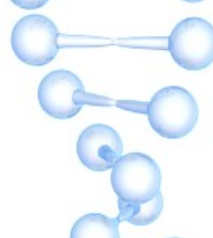
0.fa		59.4 kB	Yesterday
1.fa		4.6 MB	Yesterday
2.fa		571.0 kB	Yesterday
3.fa		3.5 MB	Yesterday
4.fa		2.0 MB	Yesterday
5.fa		18.5 MB	Yesterday
6.fa		3.2 MB	Yesterday
7.fa		6.7 MB	Yesterday
8.fa		19.6 kB	Yesterday
9.fa		142.4 kB	Yesterday
10.fa		11.1 kB	Yesterday

Use CheckM to evaluate

```
#$ -S /bin/bash
#$ -cwd
#$ -M qz85@drexel.edu
#$ -P rosenclassPrj
#$ -l h_rt=48:00:00
#$ -l h_vmem=10G
#$ -l m_mem_free=10G
#$ -q all.q
. /etc/profile.d/modules.sh
module load proteus
module load gcc/4.8.1
module load sge/univa
module load shared
module load perl/5.20.0
module load oracle/jdk/1.7.0_current
module load samtools/1.2
module load bwa/master
module load groopm/0.3.5
checkm lineage_wf ./bins_2500/ ./bins_2500/SCG -x fa -t 8 -f ../bins_2500/Check
M.txt
```

Bin Id	Marker lineage	# genomes	# markers	# marker sets	0	1	2	3	4	5+	Completeness	Contamination	Strain heterogeneity
bins.8	root (UID1)	5656	56	24	0	0	56	0	0	0	100.00	100.00	98.21
bins.6	o__Actinomycetales (UID1663)	488	309	185	0	308	1	0	0	0	100.00	0.18	0.00
bins.20	k__Bacteria (UID203)	5449	104	58	0	52	51	1	0	0	100.00	69.36	96.30
bins.14	root (UID1)	5656	56	24	0	0	0	56	0	0	100.00	200.00	100.00
bins.9	o__Actinomycetales (UID1814)	148	572	276	5	566	1	0	0	0	99.35	0.36	0.00
bins.16	c__Betaproteobacteria (UID3888)	323	387	234	19	365	3	0	0	0	99.15	0.85	0.00
bins.17	o__Pseudomonadales (UID4488)	185	813	308	22	790	1	0	0	0	99.02	0.11	0.00
bins.13	f__Rhodobacteraceae (UID3340)	84	568	330	24	543	1	0	0	0	98.79	0.30	0.00
bins.2	k__Bacteria (UID1452)	924	151	101	2	130	19	0	0	0	98.68	10.89	0.00
bins.26	c__Alphaproteobacteria (UID3337)	468	388	250	22	366	0	0	0	0	98.27	0.00	0.00
bins.22	o__Burkholderiales (UID4000)	193	427	214	11	413	3	0	0	0	97.83	1.17	0.00
bins.24	c__Deltaproteobacteria (UID3216)	83	247	155	23	224	0	0	0	0	96.13	0.00	0.00
bins.11	f__Xanthomonadaceae (UID4214)	55	659	290	80	575	4	0	0	0	95.17	0.86	0.00
bins.4	k__Bacteria (UID203)	5449	104	58	4	6	38	44	9	3	94.83	143.50	89.37
bins.25	k__Bacteria (UID203)	5449	103	58	18	25	14	23	17	6	93.86	204.31	17.58
bins.3	k__Bacteria (UID203)	5449	104	58	5	6	51	29	12	1	92.24	153.21	89.09
bins.5	c__Gammaproteobacteria (UID4202)	67	481	276	50	431	0	0	0	0	91.98	0.00	0.00
bins.12	p__Firmicutes (UID241)	930	213	118	19	186	7	1	0	0	91.40	3.25	0.00
bins.27	k__Bacteria (UID203)	5449	103	57	23	40	33	7	0	0	80.70	46.54	94.44
bins.21	k__Bacteria (UID203)	5449	103	57	50	5	41	6	0	1	80.70	86.84	98.55
bins.19	k__Bacteria (UID2328)	3167	126	75	36	58	28	2	1	1	80.52	42.67	98.00
bins.15	p__Firmicutes (UID1022)	100	295	158	60	235	0	0	0	0	76.58	0.00	0.00
bins.10	o__Clostridiales (UID1212)	172	263	149	79	177	7	0	0	0	74.22	1.31	0.00
bins.23	root (UID1)	5656	56	24	42	1	1	11	1	0	58.33	108.33	100.00
bins.1	root (UID1)	5656	56	24	14	0	1	0	2	39	41.67	159.03	97.77
bins.7	s__algicola (UID2847)	33	496	263	314	180	2	0	0	0	37.39	0.32	0.00
bins.18	k__Bacteria (UID203)	5449	104	58	89	15	0	0	0	0	22.41	0.00	0.00

	Recall									
Precision	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0.6	16	16	15	14	14	14	14	12	12	10
0.7	16	16	15	14	14	14	14	12	12	10
0.8	16	16	15	14	14	14	14	12	12	10
0.9	15	15	14	13	13	13	13	11	11	9
0.95	15	15	14	13	13	13	13	11	11	9
0.99	12	12	11	10	10	10	10	9	9	8

m = 2500(Default)   removeStrain = F
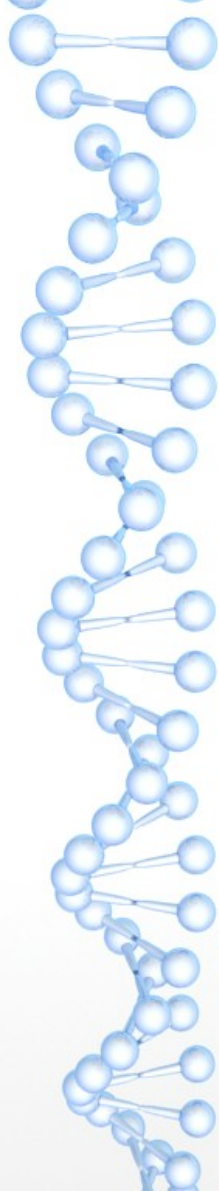
$$\text{Precision} = \frac{\text{True Positive}}{\text{Actual Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$
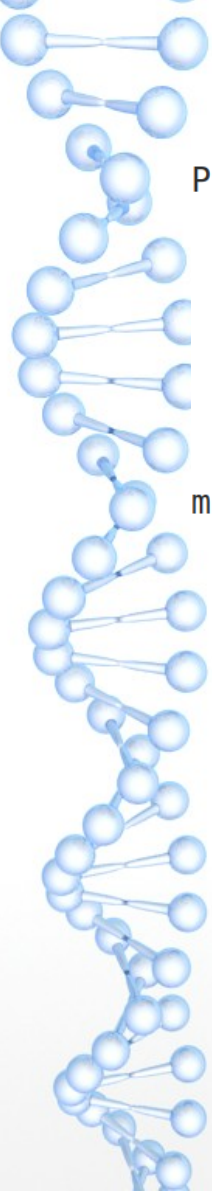
$$\text{Recall} = \frac{\text{True Positive}}{\text{Predicted Results}} \quad \text{or} \quad \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{Total}}$$

	Recall									
Precision	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0.6	16	16	15	15	14	14	14	12	11	9
0.7	15	15	14	14	13	13	13	11	10	9
0.8	15	15	14	14	13	13	13	11	10	9
0.9	14	14	13	13	12	12	12	10	9	8
0.95	14	14	13	13	12	12	12	10	9	8
0.99	11	11	10	10	9	9	9	8	8	7

$m = 2000$ removeStrain = F

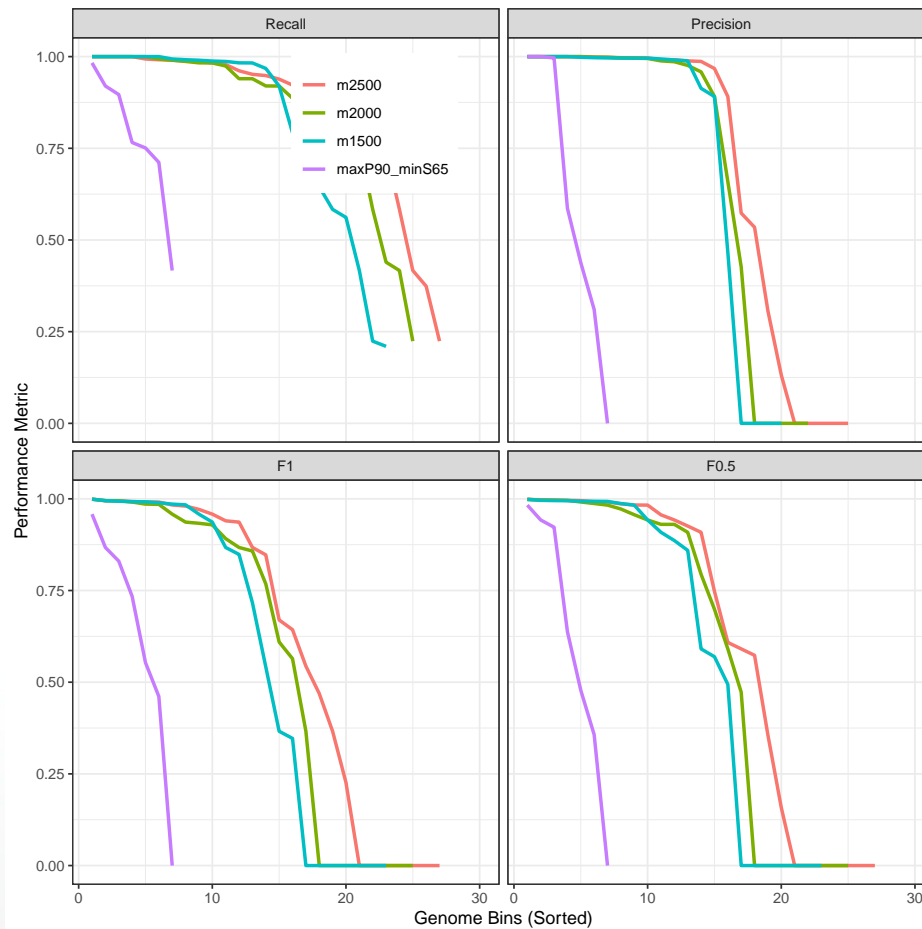|	Recall									
Precision	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
0.6	15	15	13	13	13	12	12	10	10	9
0.7	15	15	13	13	13	12	12	10	10	9
0.8	15	15	13	13	13	12	12	10	10	9
0.9	14	14	12	12	12	11	11	9	9	8
0.95	13	13	11	11	11	10	10	9	9	8
0.99	12	12	10	10	10	9	9	8	8	7

$m = 1500$ removeStrain = F

F1 and F-beta

The score combining precision and recall

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall}.$$

The recall that we can accept has a wider range than precision, so it's reasonable to use F0.5

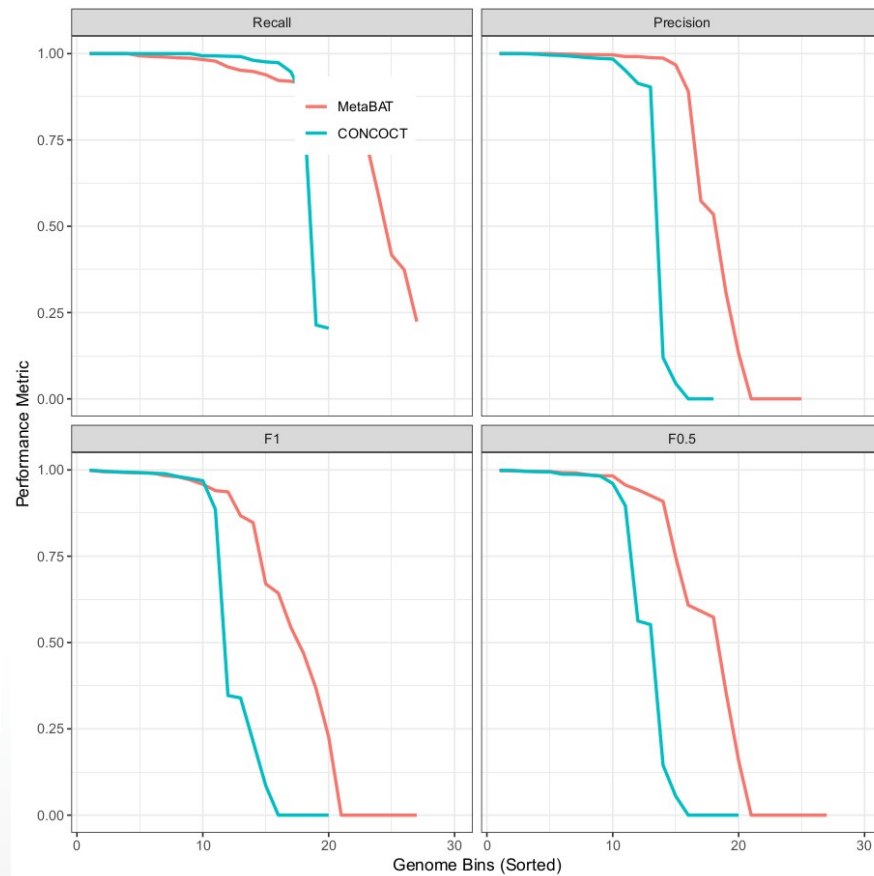# More results

Compare MetaBAT with CONCOCT using default parameters

```
$MetaBAT
          Recall
Precision 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.95
     0.7   15  14  14  14  14  12  12   10
     0.8   15  14  14  14  14  12  12   10
     0.9   14  13  13  13  13  11  11    9
     0.95  14  13  13  13  13  11  11    9
     0.99  11  10  10  10  10   9   9    8

$CONCOCT
          Recall
Precision 0.3 0.4 0.5 0.6 0.7 0.8 0.9 0.95
     0.7   11  11  11  11  11  11  10    9
     0.8   11  11  11  11  11  11  10    9
     0.9   11  11  11  11  11  11  10    9
     0.95  10  10  10  10  10  10  10    9
     0.99   6   6   6   6   6   6   6    5
```


15

References

- https://bitbucket.org/berkeleylab/metabat/src/master/
- https://github.com/Ecogenomics/CheckM/wiki
- https://github.com/BinPro/CONCOCT
-