

H. pseudoalbidus report [May 2014]

[ASSEMBLERS]

Testing Different Assemblers

For 2 strains (CBS122503 and CBS122505), several assemblers were tested.

Table 1: Read length data

Strain	Mean	Median	SD
CBS122503	150.6	151	50.7
CBS122505	153.3	152	50.7

CLC

Version: 4.21.104315

Command

```
clc_assembler -o [strain].contigs.fa -q [strain]/all_reads.fq
```

SOAPdenovo

Version: 2.04-r240

SOAP Config file

```
max_rd_len=290  
[LIB]  
q=[strain]/all_reads.fq
```

Command

```
SOAPdenovo-127mer all -s soap.conf -K 75 -R -o [strain] \  
1>[strain].log 2>[strain].err
```

ABySS

Version: 1.3.7

Command

```
abyss-pe np=8 k=75 name=[strain] se='[strain]/all_reads.fq' \
>[strain].log 2>[strain].err
```

Velvet

Version: 1.2.10

Command

```
velveth K75 75 -short -fastq [strain]/all_reads.fq \
>[strain].log 2>[strain].log
```

```
velvetg K75 -cov_cutoff auto -exp_cov auto \
>>CBS122505.log 2>>CBS122505.err
```

SPAdes

Version: v.3.0.0

Command

```
spades.py -o spades/ -s [strain]/all_reads.fq -t 8 \
>spades/[strain].log 2>[strain]/CBS122503.err
```

MIRA

Version: 4.0.2

Change Read names

```
miraconvert -R [strain] -f fastq [strain]/all_reads.fq \
all_reads.renamed.fq
```

MIRA Config file

```
project = Mira_[strain]
job = genome,denovo,accurate
readgroup = Miseq
data= [strain]_mira/all_reads.renamed.fq.fastq
technology = solexa
```

Command

```
mira mira_manifest -t 8 >[strain].log 2>[strain].err
```

Table 2: Assembler statistics for CBS122503

Assemblers	No. contigs	Span	L. contig	N50	GC(%)	N's span
velvet	14,267	58,597,690	87,562	11,905	41.3	2,396
soap	13,951	59,342,370	80,113	12,273	41.1	0
abyss	12,882	55,260,763	88,623	13,012	42.2	0
clc	6,658	53,301,203	139,344	21,365	42.8	2,897
mira	5,250	63,931,855	182,949	29,846	40.1	0
spades	3,709	61,468,046	267,719	47,440	40.6	819

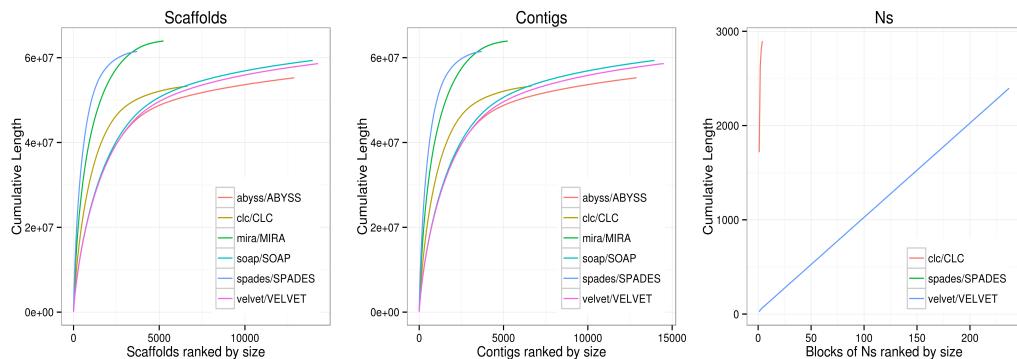


Figure 1: CBS122503 scaffolds plot

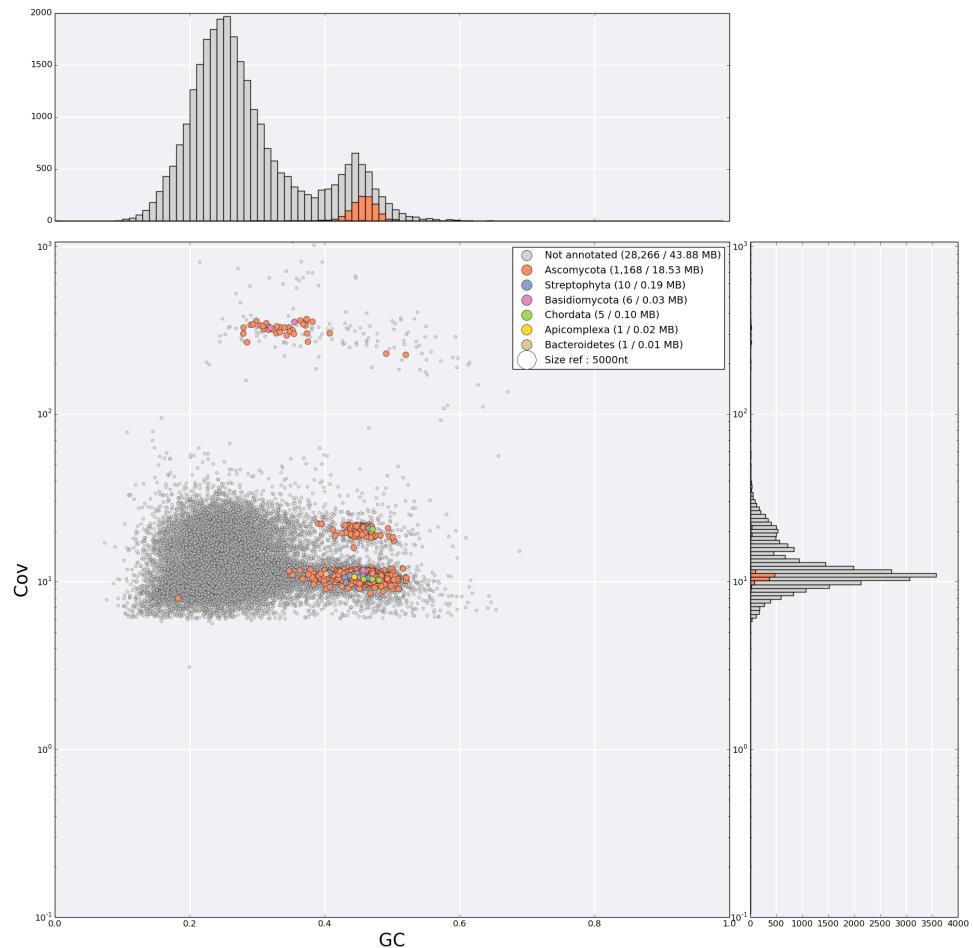


Figure 2: CBS122503 velvet TAGC plot

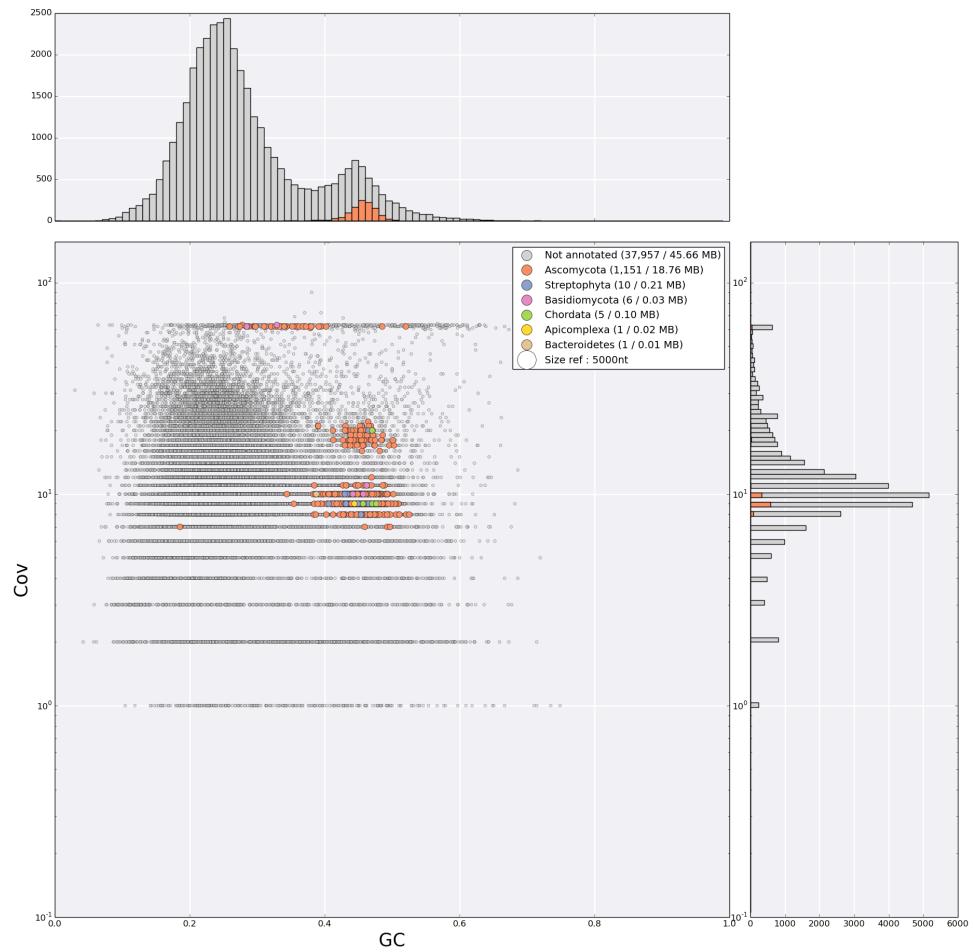


Figure 3: CBS122503 soap TAGC plot

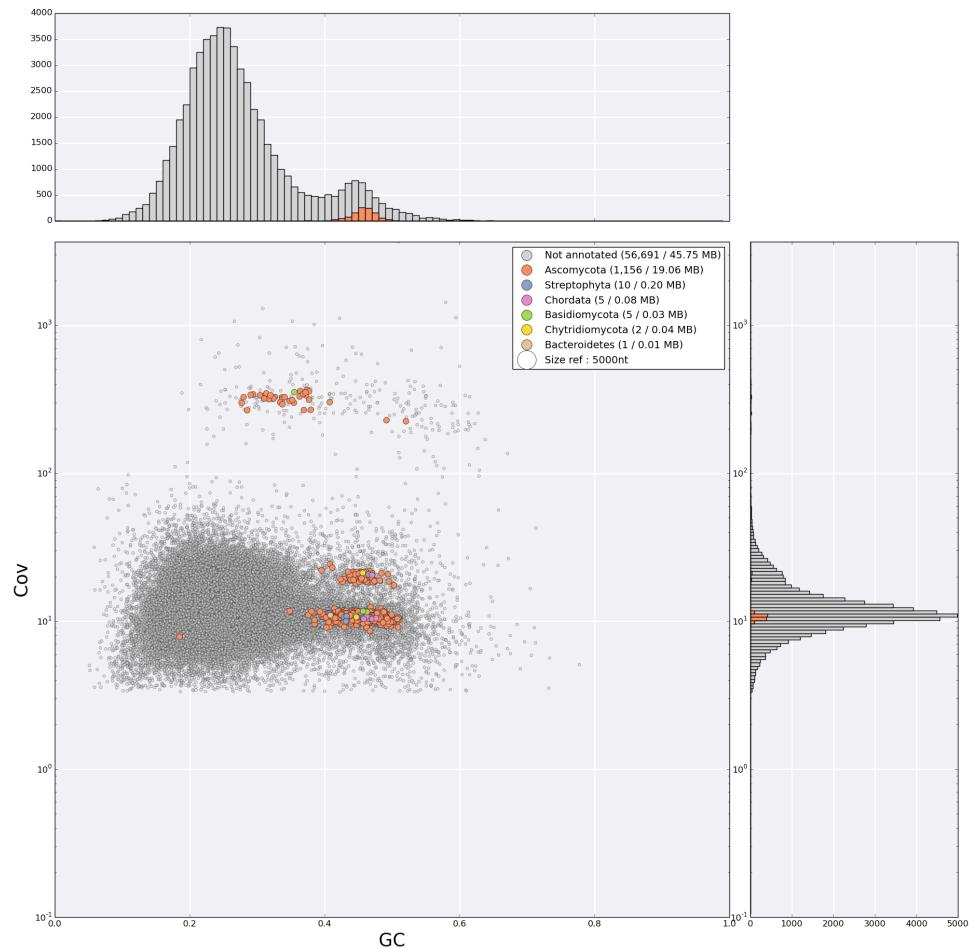


Figure 4: CBS122503 abyss TAGC plot

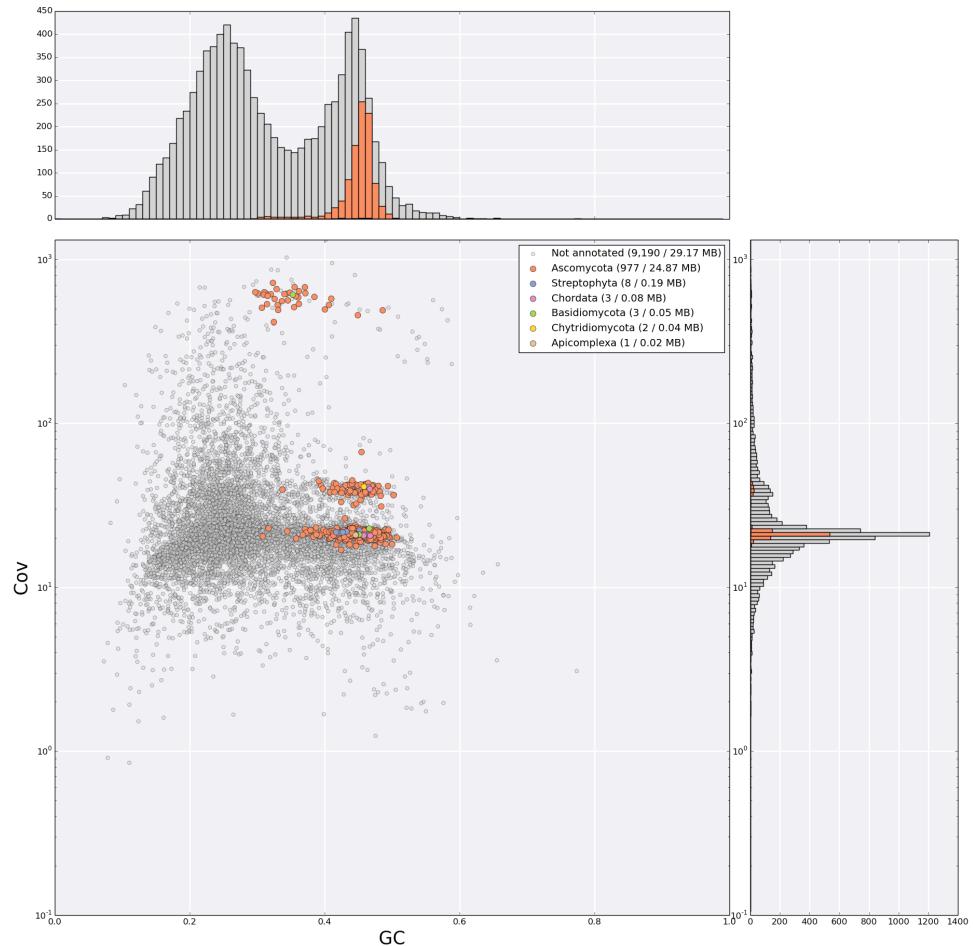


Figure 5: CBS122503 clc TAGC plot

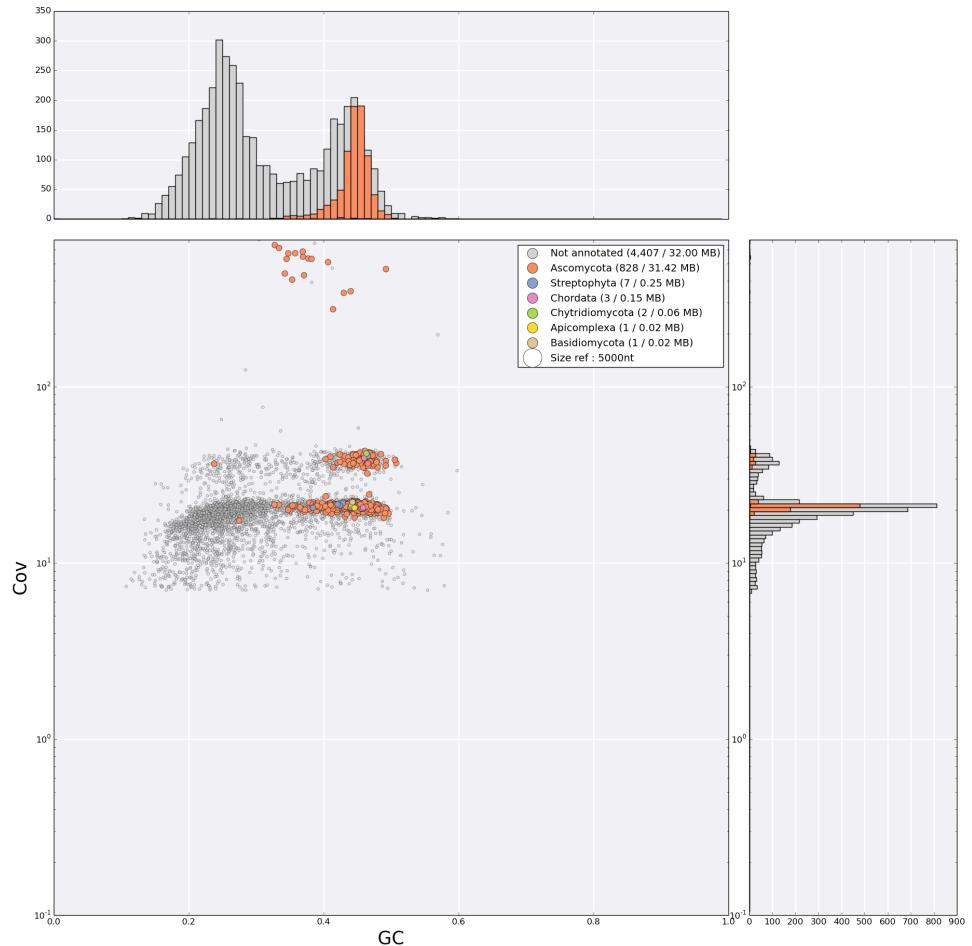


Figure 6: CBS122503 mira TAGC plot

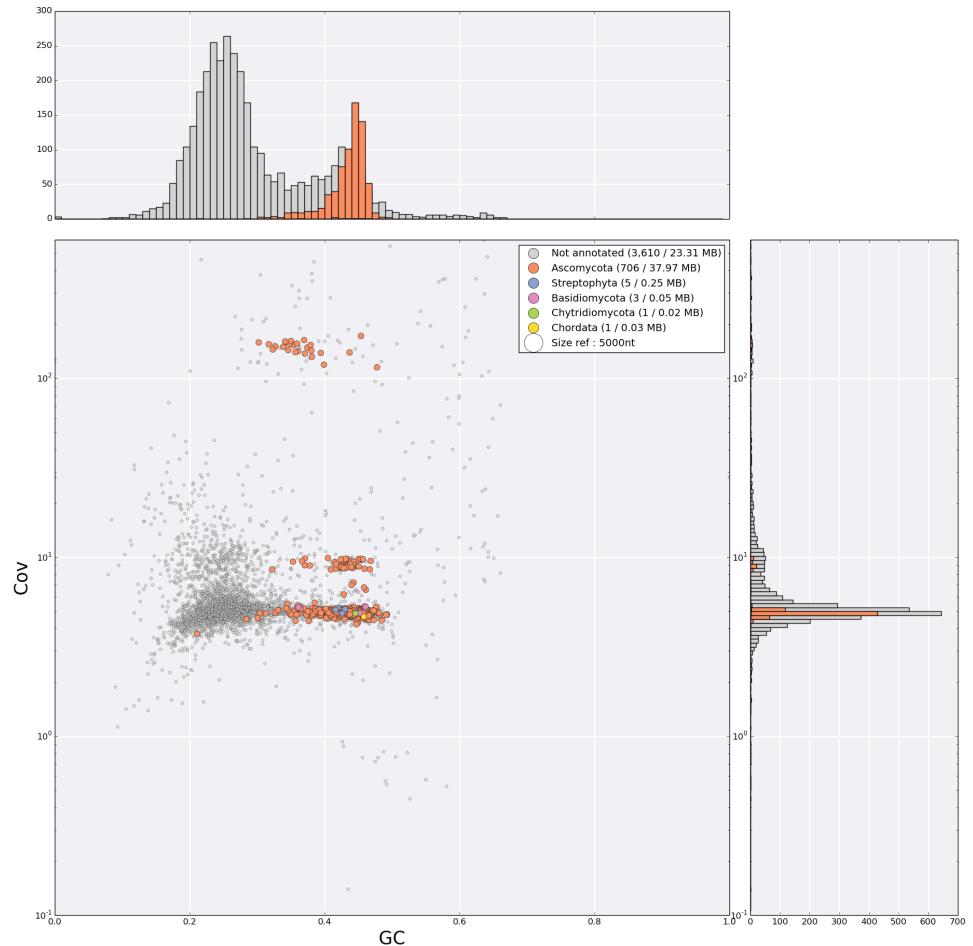


Figure 7: CBS122503 spades TAGC plot

Table 3: Assembler statistics for CBS122505

Assemblers	No. contigs	Span	L. contig	N50	GC(%)	N's span
velvet	13,604	59,032,761	95,064	14,165	41.2	2,344
soap	13,276	59,334,377	93,435	14,417	41.1	0
abyss	12,308	55,298,553	95,736	14,671	42.2	0
clc	6,504	53,508,252	141,329	22,755	42.7	0
mira	4,394	63,745,515	217,598	37,975	40.1	0
spades	3,205	61,498,734	323,268	57,397	40.6	43

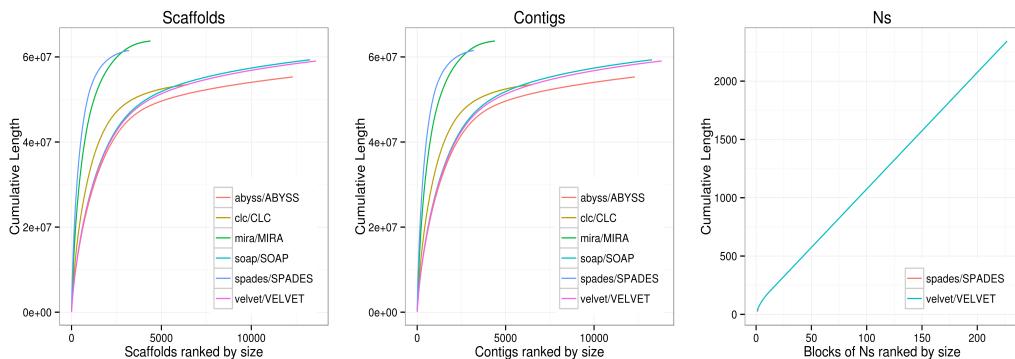


Figure 8: CBS122505 scaffolds plot

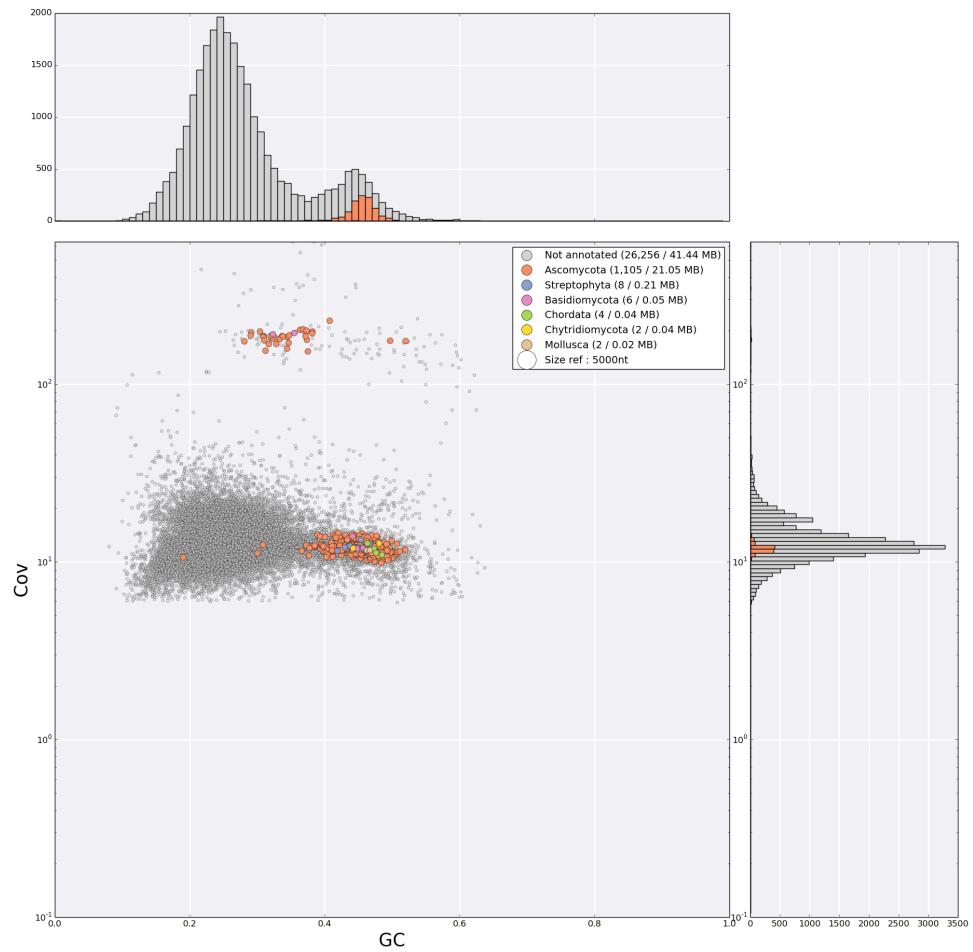


Figure 9: CBS122505 velvet TAGC plot

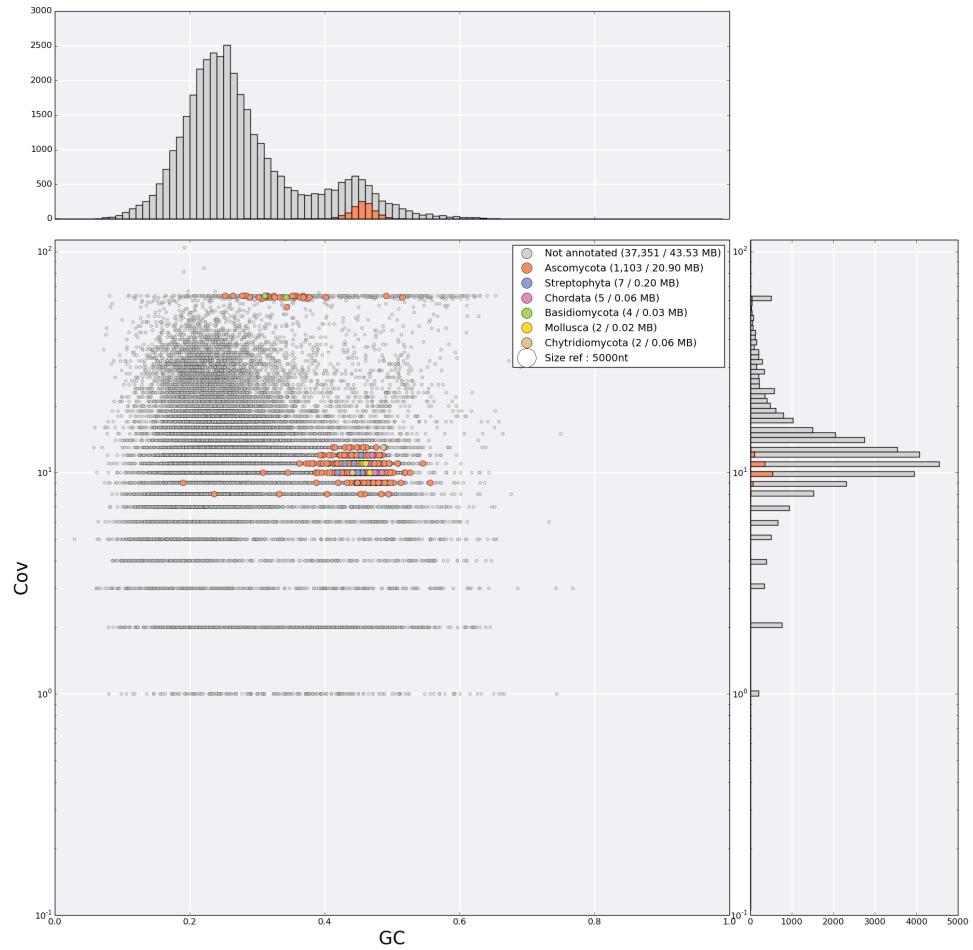


Figure 10: CBS122505 soap TAGC plot

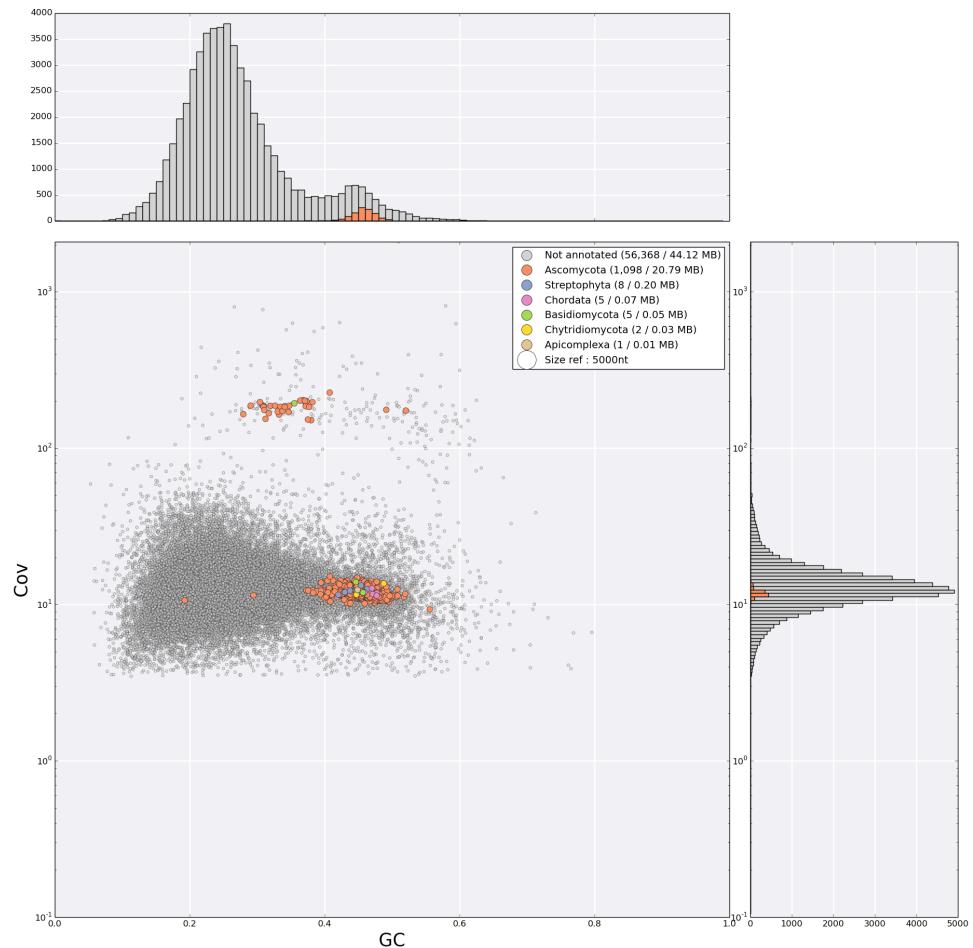


Figure 11: CBS122505 abyss TAGC plot

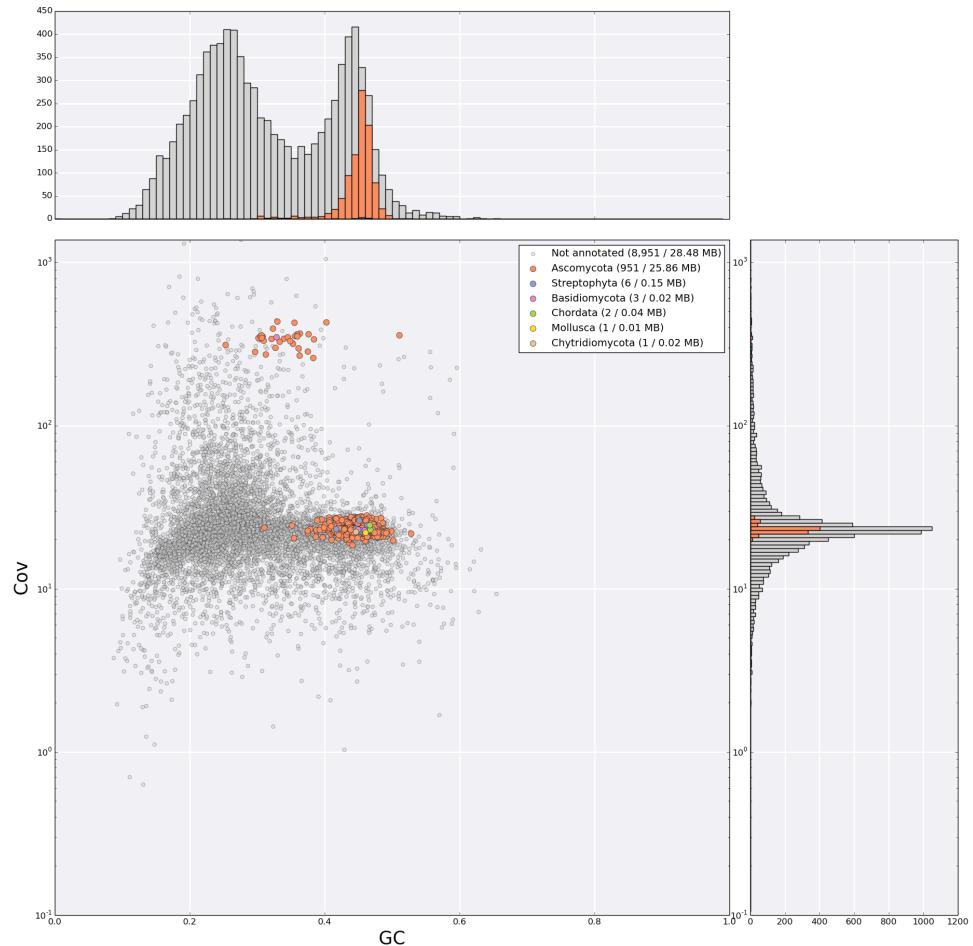


Figure 12: CBS122505 clc TAGC plot

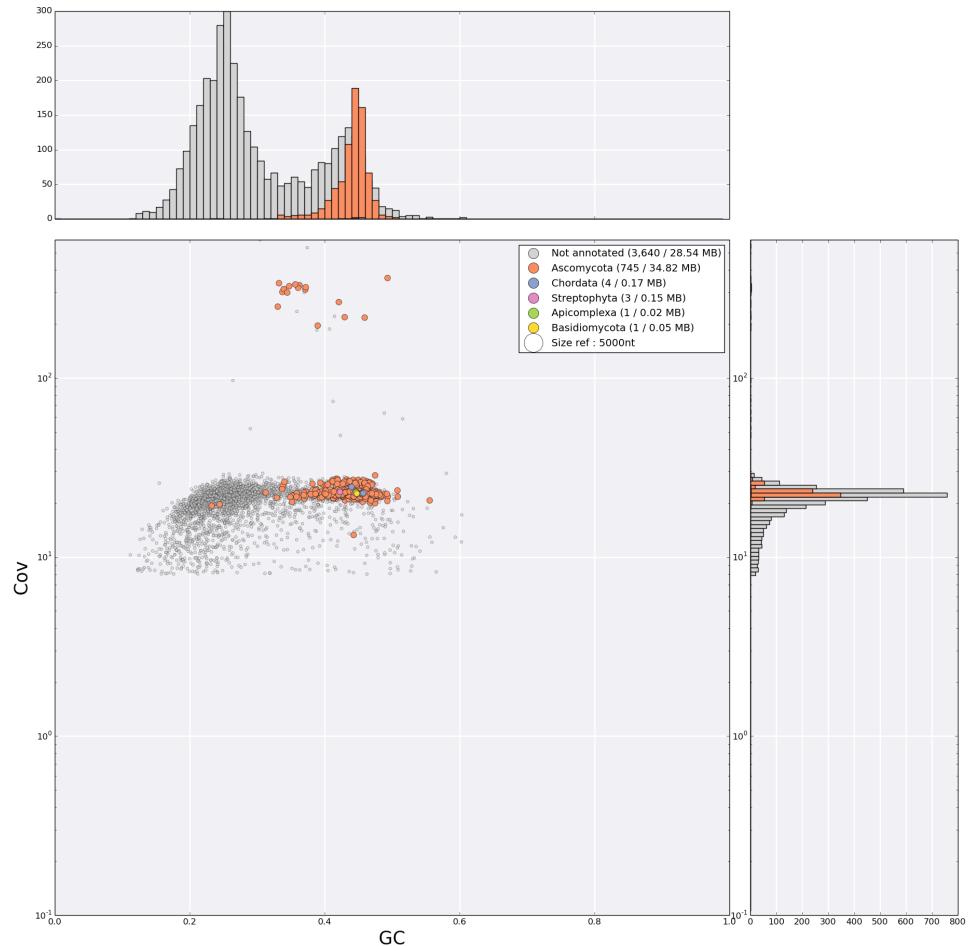


Figure 13: CBS122505 mira TAGC plot

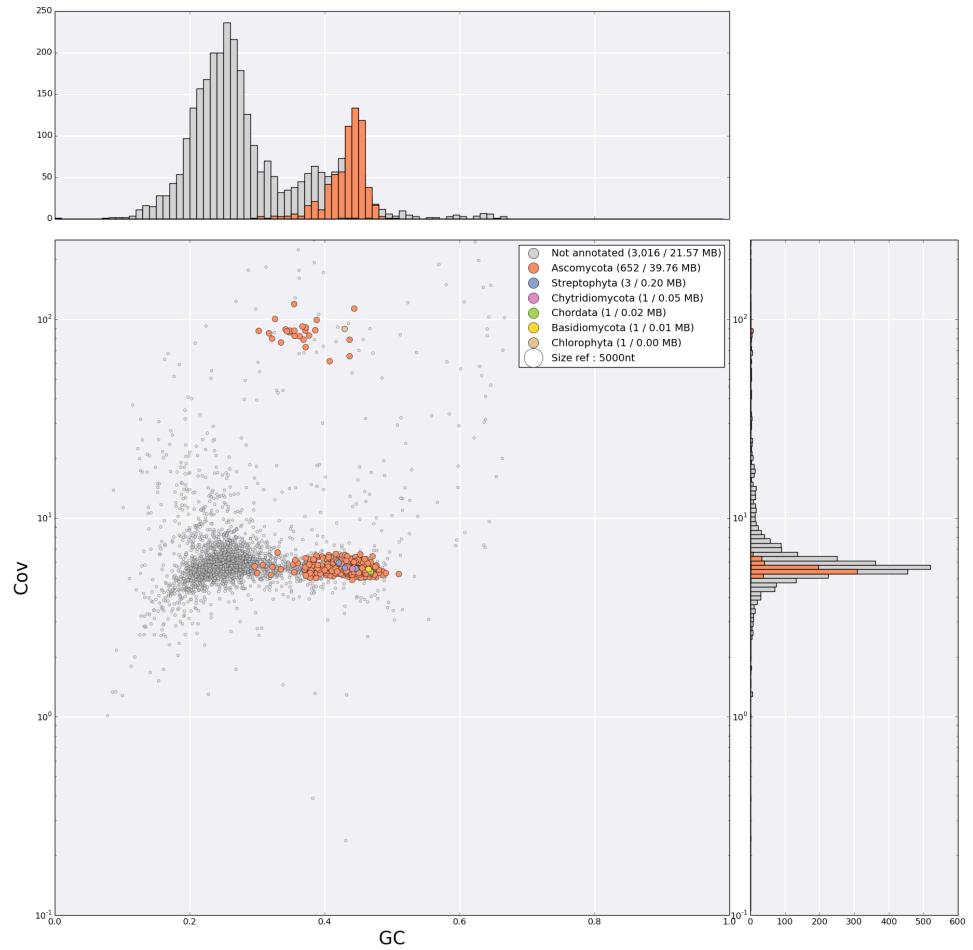


Figure 14: CBS122505 spades TAGC plot