Assembly: general tools

<u>The assemblathon stats script</u>

The assemblathon (www.assemblathon.org) used a script to obtain standardized metrics for the assemblies that were submitted. Here we use (a slightly modified version of) this script. It takes the size of the genome, and one sequence file as input. The script breaks the sequences into contigs when there are 20 or more N's, and reports all sorts of metrics.

Program	Options	Explanation
assemblathon_stats.pl		Provide basic assembly metrics
	-size	size (in Mbp, million basepairs) of target genome (optional)
	sequencefile	fasta file of contigs or scaffolds to report on

Example, for a 5.2 Mbp genome:

```
assemblathon_stats.pl -s 5.2 scaffolds.fasta
```

OR, save the output to a file with

```
assemblathon_stats.pl -s 5.2 scaffolds.fasta >
metrics.txt
```

Here, '>' (redirect) symbol used to 'redirect' what is written to the screen to a file.

Assembly viewing

'afg' and 'ace' files represent the assembly with contigs, and all read alignments, and depending on the program that produced it, the scaffolding information (newbler will not have scaffolding information in the ace file, for example). This is why these files usually are quite large. More information on the different file formats can be found here
http://www.cbcb.umd.edu/research/contig_representation.shtml

We can view the alignments in these files using the program Tablet. Open the program, click the 'Open assembly' button and find your afg or ace file.