

Blastocladiella emersonii Genome Features, Statistics and Comparisons

Features	<i>Blastocladiella</i>	<i>Allomyces</i>¹	<i>Catenaria</i>²
Genome Size (Mbp)	60.40 *	57.06	36.22
G + C (%)	51.92	61.59	50.92
Coverage	58.5x (Illumina+454)	7.31x (ABI)	93.4x (Illumina)
Total Sequenced (bp)	3,535,030,995	417,108,600	3,382,948,000
Total Consensus (bp)	423,386,588	-	-
Scaffolds	772	101	801
Longest	656,966	3,145,054	811,185
Mean	34,463	564,956	10,735
Shortest	2008	2522	1013
N50 (L50)	155,559 (52)	11,114,524 (17)	159,537 (68)
Contigs	14,099	8973	2577
Longest	31,229	312,993	-
Mean	1806	5865	-
Shortest	100	200	-
N50 (L50)	3912 (1751)	35,497 (364)	-
Predicted ORFs	14,785 – 24,097	18,773	14,188
Gene Density			
ESTs (Assembled)	23,370 (2552)	-	41,482
ESTs Mapped (%)	99.53 - 87.42	-	87.7

Key

* = Estimated by gsDenovoAssembler (newbler)

- = Data unavailable

xxxx = Estimated

1. <http://genome.jgi.doe.gov/Catan1/Catan1.info.html>
2. http://www.broadinstitute.org/annotation/genome/multicellularity_project/GenomeStats.html

Workings

Total Number Sequenced Bases

Mitochondria

$$109103 / (60.4 * 1000000) = 0.0018x$$

454 3Kb

$$172107783 + 173315055 / (60.4 * 1000000) = 5.7x$$

454 20Kb

$$67116270 / (60.4 * 1000000) = 1.1x$$

Illumina (Two x 76bp)

$$2(76 * 20541992) = 3122382784 / (60.4 * 1000000) = 51.69x$$

Total Number Bases Sequenced

$$109103 + 172107783 + 173315055 + 67116270 + 3122382784 = 3535030995$$

Coverage

$$3535030995 / (60.4 * 1000000) = 58.5x$$

Total Number Bases in Final Assembly

$$109103 + 149684185 + 150829356 + 59713874 + 63050070 \text{ (pseudoreads)} = 423386588$$

Catenaria

$$93.4 = x / (36.22 * 1000000) \rightarrow 93.4 = x / 36220000 \rightarrow x = 36220000 * 93.4 = 3382948000$$

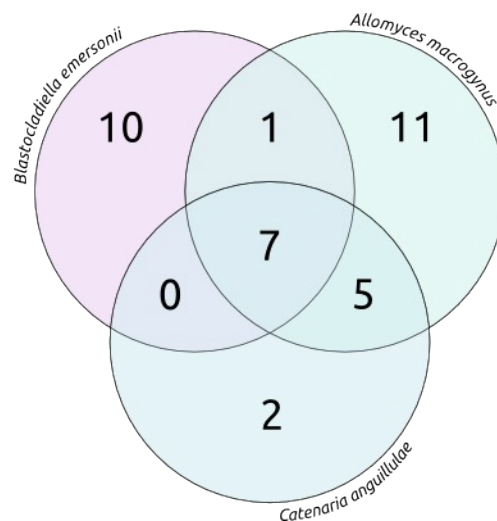
Allomyces

$$(57.06 * 1000000) * 7.31 = \sim 417108600$$

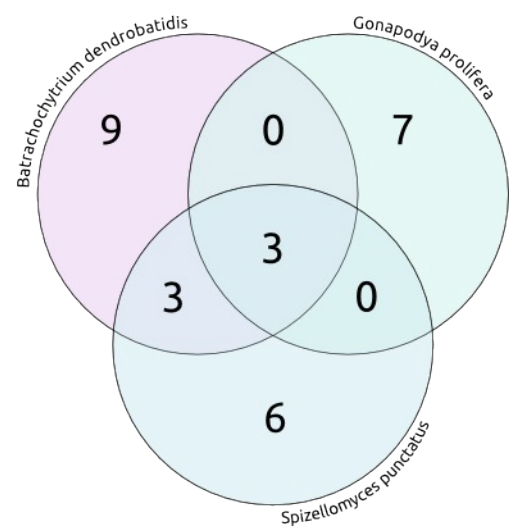
ORF / Gene / Protein Prediction

CEGMA Analysis

KOGs Absent from CEGMA Analysis of Three Blastocladiomycota Taxa



KOGs Absent from CEGMA Analysis of Three Chytridiomycota Taxa



EST Mapping

	CD-Hit 95%	CD-Hit 80%	CD-Hit 60%
Query Seqs	24108	18941	14794
Assembled EST Seqs	2552	2552	2552
Hits	2540	2239	2231
No Hit	12	313	321
HSPH	254288	15568	11672
Total Queries	2552	2552	2552
	99.53%	87.74%	87.42%