

Meeting 26th February, 2016 (13:00)

Points of previous meeting

- Check quality of assemblies and read coverage

- Conclusions about species:

- 1017 and 1018: *Saccharomyces cerevisiae* (used as control until now, probably not needed anymore in further analysis)
- 1001, 1002, 1003, 1004, 1005, 1007, 1008, 1013, 1014, 1015, 1016, 1019: *Debaryomyces hansenii* – will be used for variant calling and phylogeny.
- 1006 and 1012: unknown specie 1. *
- 1010: unknown specie 2. *
- 1009 and 1011: probable hybrids. *

* The study of this strains have been put aside for the moment as the *D. hansenii* strains analysis is more needed now.

Variant calling

Pre-Variant calling

Alignment: mapping done in the first steps of the project - Bowtie2

Remove PCR duplicates: picard tools - MarkDuplicates

AddReadGroups: picard tools – AddOrReplaceReadGroups

Local realignment/BAQ – samtools - calmd

Variant calling (leaving filtering for later)

Freebayes – cohort analysis (bam-list)

Freebayes – individual analysis (bam)

VarScan – cohort analysis (mpileup)

Post-Variant calling

vcffilter

snpEff

SnpSift

(some done, but we should discuss the next steps with this data to know which approach to take is best)

Mapping against *Debaryomyces fabryi*

Genome recently published, it is a specie that has been mistaken before with *Debaryomyces hansenii*, so it existed the possibility that it may be one of our unknown species.

Overall alignment rate (%)

Strain	AH	BC
1001	13.02	13.05
1002	14.07	14.04
1006	38.44	38.51
1009	12.85	12.87
1010	22.33	22.30
1011	13.39	13.41
1012	39.79	39.65

1001 and 1002 used as controls, known to be *Debaryomyces hansenii*.
 1009 and 1011 hybrid species (map around 50% against *Debaryomyces hansenii*)
 1006, 1010 and 1012 Unknown species (1006 and 1012 same specie, 1010 different one)

With this results we could conclude that none of them is *Debaryomyces fabryi*, but at least is the closest specie that we found until now to one of our unknown species, as 1006 and 1012 have an overall alignment of 38-39% with its reference genome (Being 9% against *D. hansenii*), same with the 1010 specie, but a bit less (22% overall alignment rate, being 5% against *D. hansenii*).

So we still have 2 unknown species, but at least we know a little bit more about them.

Annotation using Maker

For now it has been run on *D. hansenii* strains:

Number of transcripts predicted:

	CBS767	1001	1002	1003	1004	1005	1007	1008	1013	1014	1015	1016	1019
Maker	6326	6064	6401	6097		6327	6219	6326	6421	6096	6375	6153	6380
Ref	6290												5313

Non-*D. Hansenii* strains and strain 1004 (very fragmented assembly) still running.