## Meeting 22nd of January, 2016 (13:00)

**Points of previous meeting**

Are Mahesh assemblies good enough?

- Map reads to its own assembly.
- Check Mahesh report.
- SOAPdenovo tested with some strains, not improved.
- Check read coverage for all of them.

Hybrid strains (1009, 1011)

- Make a kraken database with them, not possible.
- 1011 the worst set of reads, probably should be sequenced again to analyze further.

Unknown species (1006, 1010, 1012)

- Map them in pairs: 1006 and 1012 probably same specie, 1010 different one.
- Check regions shared with CBS767 reference genome.

**Include hybrid strains in Kraken database**

Not possible, it needs a GI number to work, it needs to be included on the taxonomy.

**Map raw reads to its own assembly to check assemblies**

Overall alignment rate - Bowtie2

| Assembly | AH reads | BC reads |
|---|---|---|
| 1001 | 98.25% | 98.13% |
| 1002 | 97.87% | 97.54% |
| 1003 | 98.42% | 98.30% |
| 1004 | 96.19% | 96.59% |
| 1005 | 97.70% | 97.57% |
| 1006 | 98.26% | 98.12% |
| 1007 | 97.99% | 97.83% |
| 1008 | 97.52% | 97.23% |
| 1009 | 91.33% | 91.26% |
| 1010 | 93.66% | 93.53% |
| 1011 | 85.94% | 85.46% |
| 1012 | 98.09% | 97.92% |
| 1013 | 98.22% | 98.15% |
| 1014 | 98.10% | 97.88% |
| 1015 | 98.13% | 97.90% |
| 1016 | 98.39% | 98.25% |
| 1017 | 89.81% | 89.58% |

| Assembly | AH reads | BC reads |
|---|---|---|
| 1018 | 92.05% | 91.86% |
| 1019 | 98.04% | 98.32% |

Most of them quite good except from 1011 and 1017.

**Map weird strains between each other to check how close they are.**

Overall alignment rate - Bowtie2

Columns: assembly / Rows: Reads (Both sets of reads)

| Strain | 1006 | 1010 | 1012 | 1009 | 1011 |
|---|---|---|---|---|---|
| 1006 | | 18.34% | 97.88% | 12.24% | 11.69% |
| | | 18.40% | 97.73% | 12.29% | 11.70% |
| 1010 | 20.29% | | 20.35% | 7.68% | 7.84% |
| | 20.21% | | 20.28% | 7.73% | 7.81% |
| 1012 | 97.27% | 18.87% | | 12.59% | 12.13% |
| | 97.11% | 18.85% | | 12.55% | 12.07% |
| 1009 | 12.60% | 6.59% | 12.58% | | 76.23% |
| | 12.61% | 6.62% | 12.58% | | 76.18% |
| 1011 | 14.71% | 6.90% | 14.77% | 71.37% | |
| | 14.65% | 6.87% | 14.71% | 70.95% | |

1006, 1010 and 1012 – Not Debaryomyces hansenii strains

1006 and 1012 same specie

1010 different specie

1009 and 1011 hybrid/double size genome strains

They don't seem to be the same, but sequences of 1011 cannot be completely reliable, bad raw reads and not a good assembly due to that.

**PreQC on 1006 and 1012 to prepare for an improvement of the assembly and SOAPdenovo assembly**

Duplicates removal.

Check that coverages don't change that much after removing duplicates.

Assembly not improved. Probably Mahesh assemblies are the best we can get.

**Coverage of all the raw data of all the strains.**

**Coverage of raw reads**

| Strain | Coverage |
|---|---|
| 1001ah | 20 |
| 1001bc | 29 |
| 1002ah | 9 |

| Strain | Coverage |
|---|---|
| 1002bc | 15 |
| 1003ah | 18 |
| 1003bc | 28 |
| 1004ah | 10 |
| 1004bc | 17 |
| 1005ah | 17 |
| 1005bc | 24 |
| 1006ah | 16 |
| 1006bc | 24 |
| 1007ah | 19 |
| 1007bc | 29 |
| 1008ah | 10 |
| 1008bc | 16 |
| 1009ah | 8 |
| 1009bc | 13 |
| 1010ah | 10 |
| 1010bc | 15 |
| 1011ah | 3 |
| 1011bc | 5 |
| 1012ah | 18 |
| 1012bc | 28 |
| 1013ah | 14 |
| 1013bc | 22 |
| 1014ah | 18 |
| 1014bc | 29 |
| 1015ah | 15 |
| 1015bc | 26 |
| 1016ah | 21 |
| 1016bc | 32 |
| 1017ah | 15 |
| 1017bc | 22 |
| 1018ah | 21 |
| 1018bc | 30 |
| 1019ah | 16 |
| 1019bc | 22 |

Some of the coverages are really low, probably that is why Mahesh assemblies are as best as we can get with these raw sequences. New sequencing should probably be ordered to continue studying these strains.

**Compare weird strains with CBS767 to check regions in common**

Will be explained in the meeting.