Raw data

Classified – non classified

1017 and 1018: more than 95% of its reads classified.

1006 and 1012: around 17% of the reads classified.

1010: around 14% of the reads classified.

1009: around 72% of the reads classified.

1011: around 58% of the reads classified.

The rest of strains: between 80% and 98% of the reads classified.

Taxonomy

All of them low values of bacteria, archaea, virus or human contamination.

1017 and 1018: more than 91% of the reads belong to Saccharomyces cerevisiae.

1006 and 1012: around 12% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

1010: around 10% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

1009: Around 68,5% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

1011: Around 53% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

The rest of strains: From 75% to 95% of the reads match with Debaryomyces hansenii.

Assemblies

Classified – non classified

1017 and 1018: around 98% of the reads classified.

1006 and 1012: around 40% of the reads classified.

1010: 75% of the reads classified.

1009 and 1011: around 96% of the reads classified

The rest of strains: Between 52% and 97% of the reads classified. Weird things: 1019 suddenly is 69,89 (raw reads: around 98%), 1005 is 52,66% (raw reads: around 98%), 1013 is 67,16% (raw reads: 96%).

Taxonomy

All of them low values of bacteria, archaea, virus or human contamination, except for 1005 that has 6,14% of Bacteria reads, 1013 that has 3.78% of bacteria reads, and 1019 that has 3,03% of bacteria reads.

1017 and 1018: more than 97% of the reads belong to Saccharomyces cerevisiae.

1006 and 1012: between 25 and 33% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

1010: 25% of the reads match with Debaryomyces hansenii, the rest mostly unclassified.

1009 and 1011: Around 95% of the reads match with Debaryomyces hansenii.

The rest of strains: From 60% to 96% of the reads match with Debaryomyces hansenii.

ITS alignment

It can be differentiated the 18s ribosomal sequence, the ITS1, the 5,8s ribosomal sequence and a bit more difficult the ITS2 and the 28s ribosomal sequence.

Muscle:

ClustalW:

ClustalO:

T-Coffee:

K-align:

Differences between Spades assembled sequences and MaSuRCA assembled sequences.

Spades:

1001, 1002, 1003, 1004, 1005, 1006, 1007, 1008, 1012, 1013, 1014, 1015, 1016, 1019

Easy to find the whole region of ribosomal sequences.

MaSuRCA:

1009, 1010, 1011, 1017, 1018

Almost impossible to find the ribosomal sequences. In 1018 is non-existent at all, and in most of the rest I couldn't even find regions containing ITS sequences, only short fragments of 18s or 28s.

Phylogeny:

Two examples of all the trees that could be done:

Mr. Bayes – JC 69

All the MaSuRCA assemblies align together. Probably not reliable.

1008 NODE 222 c1051-2073 NODE 309 c444-1 NODE 383 c571-357 NODE 334 c336-815 NODE 345 c1-791 NODE 183 c112-65
1002 NODE 162 c848-1 NODE 325 c490-261 NODE 100 c462-2459
1001 NODE 133 c1384-362 NODE 102 c361-2358
1011 jcf7180000062816 c89-529 jcf7180000062901 c1-339 jcf7180000062250 c534-395
1017 jcf7180000031087 c461-1
1009 jcf7180000024027 c479-1
1010 jcf7180000017065 c254612-254220 jcf7180000016991 c94923-94864
1015 NODE 186 c1603-2062 NODE 454 c385-1 NODE 209 c1449-1 NODE 143 c4700-3374
1013 NODE 104 c919-4229 NODE 397 c54-393
1007 NODE 79 c3630-1 NODE 74 c3992-3935
1014 NODE 28 c146647-147436 NODE 136 c1-2895 NODE 144 c2438-2381
1005 NODE 63 c4572-8201 NODE 362 c441-349
1019 NODE 60 c7009-3323
1003 NODE 67 c1030-4716
1016 NODE 91 c2184-1 NODE 90 c1-1558
1004 NODE 413 c2156-5785 NODE 442 c442-350
1006 NODE 89 c1095-1884 NODE 73 c2894-1 NODE 32 c128168-128111
1012 NODE 126 c515-1304 NODE 93 c1-2894 NODE 152 c55-112

2

Neighbor-joining – kimura

This one looks maybe a bit better.

1001 NODE 133 c1384-362 NODE 102 c361-2358
1008 NODE 222 c1051-2073 NODE 309 c444-1 NODE 383 c571-357 NODE 334 c336-815 NODE 345 c1-791 NODE 183 c112-65
1009 jcf7180000024027 c479-1
1011 jcf7180000062816 c89-529 jcf7180000062901 c1-339 jcf7180000062250 c534-395
1006 NODE 89 c1095-1884 NODE 73 c2894-1 NODE 32 c128168-128111
1012 NODE 126 c515-1304 NODE 93 c1-2894 NODE 152 c55-112
1017 jcf7180000031087 c461-1
1010 jcf7180000017065 c254612-254220 jcf7180000016991 c94923-94864
1015 NODE 186 c1603-2062 NODE 454 c385-1 NODE 209 c1449-1 NODE 143 c4700-3374
1013 NODE 104 c919-4229 NODE 397 c54-393
1007 NODE 79 c3630-1 NODE 74 c3992-3935
1014 NODE 28 c146647-147436 NODE 136 c1-2895 NODE 144 c2438-2381
1005 NODE 63 c4572-8201 NODE 362 c441-349
1019 NODE 60 c7009-3323
1003 NODE 67 c1030-4716
1004 NODE 413 c2156-5785 NODE 442 c442-350
1016 NODE 91 c2184-1 NODE 90 c1-1558
1002 NODE 162 c848-1 NODE 325 c490-261 NODE 100 c462-2459