

Erratum Dohm et al., Genome Res. 2007 17: 1697-1706, Table 2

An error occurred when finalising Table 2. The correct assembly property values for *H. influenzae* and *E. coli* are given below. The assembly properties of the assemblies for *S. cerevisiae* chromosomes 5 and 7 remain unchanged. We apologize for this mistake.

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----------------|-------------|-------------|-------------|-------------|---------|------|-------|-------|-------|---------------|
| Species | GenBank | Seq. Length | Reads (Mio) | Read Length | Contigs | Mean | Max | N50 | N80 | Seq. cov. (%) |
| <i>S.cer 5</i> | NC_001137 | 576869 | 2.5 | 30 | 237 | 2379 | 41960 | 16096 | 5137 | 97.36 |
| <i>S.cer 5</i> | NC_001137 | 576869 | 5 | 30 | 319 | 1793 | 54053 | 27975 | 17848 | 98.23 |
| <i>S.cer 5</i> | NC_001137 | 576869 | 2.5 | 32 | 233 | 2423 | 42420 | 16987 | 5621 | 97.45 |
| <i>S.cer 7</i> | NC_001139 | 1090946 | 10 | 30 | 330 | 3262 | 60516 | 24771 | 10280 | 98.32 |
| <i>S.cer 7</i> | NC_001139 | 1090946 | 5 | 30 | 367 | 2914 | 32709 | 9737 | 3700 | 97.77 |
| <i>S.cer 7</i> | NC_001139 | 1090946 | 5 | 32 | 300 | 3568 | 46493 | 15820 | 6715 | 97.88 |
| <i>H. inf</i> | NC_007146.1 | 1913428 | 10 | 30 | 589 | 3183 | 48963 | 12323 | 5589 | 97.64 |
| <i>H. inf</i> | NC_007146.1 | 1913428 | 10 | 32 | 554 | 3389 | 67333 | 18402 | 7227 | 97.78 |
| <i>E. col</i> | NC_000913.2 | 4639675 | 15 | 30 | 2311 | 1969 | 35437 | 4506 | 2033 | 97.58 |
| <i>E. col</i> | NC_000913.2 | 4639675 | 20 | 30 | 1401 | 3247 | 53973 | 10084 | 4733 | 97.77 |
| <i>E. col</i> | NC_000913.2 | 4639675 | 20 | 32 | 1276 | 3570 | 73047 | 12754 | 6169 | 97.91 |

Table 2: Assembly properties for assemblies with simulated reads of *Saccharomyces cerevisiae* chromosome 5 (coverage 130-, 260-, 139-fold) and chromosome 7 (coverage 275-, 138-, 147-fold) and the complete genomes of *Haemophilus influenzae* (coverage 157- and 167-fold) and *Escherichia coli* (coverage 97-, 129-, 138-fold). Simulation assumptions were realistic, i.e. with missing reads and sequencing error rate of 0.6% per base; coverage range Column labels: 1: Species name (abbreviated). 2: GenBank accession number. 3: Length of GenBank reference sequence. 4: Number of reads used for assembly (in millions). 5: Read length (bases). 6: Number of contigs > 50 bp after assembly. 7: Average contig length. Only contigs >50 bp were counted. 8: Largest contig observed in assembly. 9: N50 length of assembly. 10: N80 length of assembly. 11: Percent coverage of reference sequence with SHARCGS contigs. All contigs could be aligned against the reference sequence error- and gap-free.