Whole genome sequencing, assembly and annotation of the *Saccharomyces cerevisiae* strain CEN.PK 113-7D

Jurgen F. Nijkamp[1,3], Marcel A. van den Broek[2,3], Jean-Marc Daran[2,3], Marcel J.T. Reinders[1,4] and Dick de Ridder[1,3]

[1]The Delft Bioinformatics Lab, Department of Mediamatics, Delft University of Technology, Mekelweg 4, 2628 CD Delft [2]Industrial Microbiology Group, Department of Biotechnology, Delft University of Technology, Julianalaan 67, 2628 BC Delft. [3]Kluyver Centre for Genomics of Industrial Fermentation, P.O. Box 5057, 2600 GA Delft. [4]Netherlands Bioinformatics Center, 260 NBIC, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

E: J.F.Nijkamp@tudelft.nl

The baker's yeast *S. cerevisiae* CEN.PK strain is extensively used for systems biology and fermentation studies, as it is robust and highly relevant to industry. In biotechnology, metabolic engineering efforts are enhanced by evolutionary engineering to improve strain performance by selecting for certain phenotypes. Mutations arising during the evolutionary process can be deduced using next-generation sequencing and then linked to the observed phenotypes, in order to reverse-engineer the resulting strains. The basis for such re-sequencing of evolved strains of CEN.PK is to have a high-quality reference genome. To construct this, two libraries with sequencing data were obtained and assembled: a 454 shotgun and an Illumina paired-end dataset. The resulting contig set was scaffolded using the well-known *S. cerevisae* S288c genome.

The obtained genome was annotated using a combination of tools. ORFs were identified using three *ab initio* and two comparative gene model predictors. The predicted gene models were validated with an RNAseq dataset. Nine RNA samples were taken from chemostat fermentations under three different conditions: glucose-limited anaerobic, glucose limited aerobic and nitrogen limited anaerobic. Almost all genes will be expressed in at least one of these conditions. The annotated CENPK genome was then compared to the known yeast genomes S288c, RM11-1A and YJM789, and a number of CEN.PK-specific genes and gene variants were catalogued.

The resulting annotated genome sequence forms a solid basis not only for our work in evolutionary engineering, but also serves as a reference point for further genomic studies in CEN.PK-derived strains.