

Genome sequence of the recombinant protein production host *Pichia pastoris*



Kristof De Schutter^{1*}, Yao-Cheng Lin^{2*}, Petra Tiels^{1*}, Annelies Van Hecke¹, Lieven Sterck², Thomas Abeel², Sascha Glinka³, Jacqueline Weber-Lehmann³, Pierre Rouz  ², Yves Van de Peer² & Nico Callewaert¹
e-mail: yalin@psb.vib-ugent.be

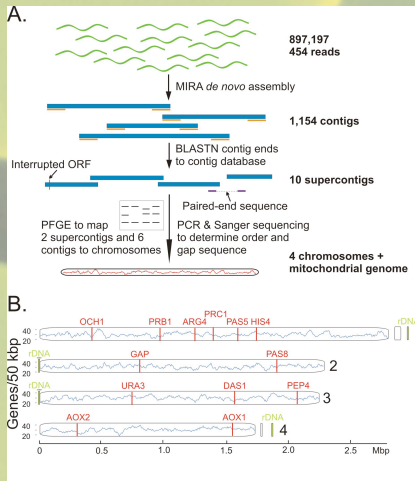


1 Unit for Molecular Glycobiology, VIB Department for Molecular Biomedical Research, UGhent, Belgium.
2 Division of Bioinformatics and Systems Biology, VIB Department of Plant Systems Biology, UGhent, Belgium.
3 Eurofins MWG Operon, Ebersberg, Germany. (* These authors contributed equally to this work)

Introduction

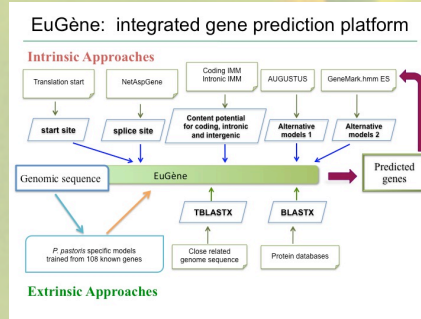
The methylotrophic yeast *Pichia pastoris* is widely used for the production of proteins and as a model organism for studying peroxisomal biogenesis and methanol assimilation. *P. pastoris* strains capable of human-type N-glycosylation are now available, which increases the utility of this organism for biopharmaceutical production. Despite its biotechnological importance, relatively few genetic tools or engineered strains have been generated for *P. pastoris*. To facilitate progress in these areas, we present the 9.43 Mbp genomic sequence of the GS115 strain of *P. pastoris*. We also provide manually curated annotation for its 5,313 protein-coding genes. (Genome portal <http://bioinformatics.psb.ugent.be/webtools/bogas/>)

Genome Sequencing and Assembly



Pichia pastoris genome sequencing and overview. (a) Genome sequencing and assembly strategy. (b) *P. pastoris* gene density and known markers position. Gene density is plotted as a histogram, showing a uniform distribution of genes across each chromosome. The gene density is calculated in a window size of 50 kbp with 5 kbp sliding window. Genes that had been previously mapped to the chromosomes through PFGE are indicated in red, and rDNA repeats in green.

Automatic Gene Prediction



A schematic representation of the EuGene integrated gene prediction platform. The species specific prediction parameters could be trained based on the known gene structure of 108 *P. pastoris* genes to improve the gene prediction accuracy.

Genome Contents Overview

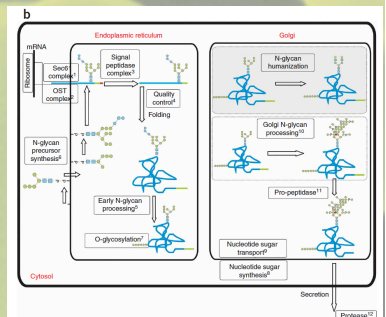
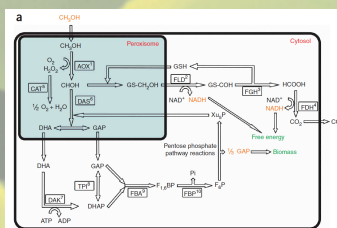
General information	Coding genes	RNA genes	Mitochondrial genome
Size (Mbp): 9.33 (not including rDNA loci, estimated at 0.12 Mbp)	Coding genes: 5,313	tRNA genes: 123	Size (bp): 36, 119
Genome GC content (%): 41.1	Coding (%): 79.6	5s rRNA genes: 21	Genome GC content (%): 22
Assembled chromosomes: 4	Coding GC (%): 41.6	Mean gene length (bp): 1,442	Coding genes: 16
	Single exon genes: 4,680		tRNA genes: 31

Genome Portal and Next-Gen Tool

The BioinformaticsGent Online Genome Annotation System (BOGAS) provides a workspace of gene structure and functional annotation for expert annotators. This system is the backbone for tracking the manual annotation process and is an entry point to access the genome information. The GenomeView with genome browsing and annotation functionality enables the single-base resolution of short read data.

Pichia pastoris Pathways

Methanol utilization pathway (a) and Protein secretion pathway (b) in *Pichia pastoris*



Conclusion & Perspectives

- The pyrosequencing technique and a proper assembly strategy can deliver a reliable eukaryotic genome assembly.
- The development of the assembly, annotation and genome visualization tools will improve the usefulness of new sequencing data.
- The wealth of information provided by a full genome sequence will enable a more rapid development of *P. pastoris* as a protein expression host, building on its exceptional natural capacity for heterologous protein production.

