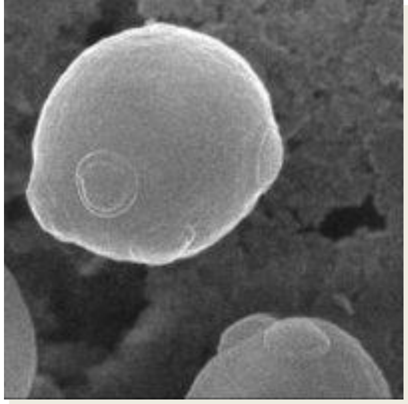
EUKARYOTIC GENOMES
DSE – Bioinformatics Sem V

Size of major eukaryotic genomes

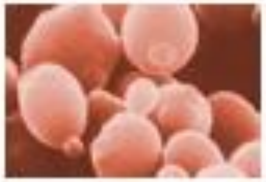
Species	Genome size (Mb)
Fungi	
Saccharomyces cerevisiae	12.1
Aspergillus nidulans	25.4
Protozoa	
Tetrahymena pyriformis	190
Invertebrates	
Caenorhabditis elegans	97
Drosophila melanogaster	180
Bombyx mori (silkworm)	490
Strongylocentrotus purpuratus (sea urchin)	845
Locusta migratoria (locust)	5000
Vertebrates	
Takifugu rubripes (pufferfish)	400
Homo sapiens	3200
Mus musculus (mouse)	3300
Plants	
Arabidopsis thaliana (vetch)	125
Oryza sativa (rice)	430
Zea mays (maize)	2500
Pisum sativum (pea)	4800
Triticum aestivum (wheat)	16 000
Fritillaria assyriaca (fritillary)	120 000

Yeast (*Saccharomyces cerevisiae*) Genome Sequencing

Fungal genomes: *S. cerevisiae*



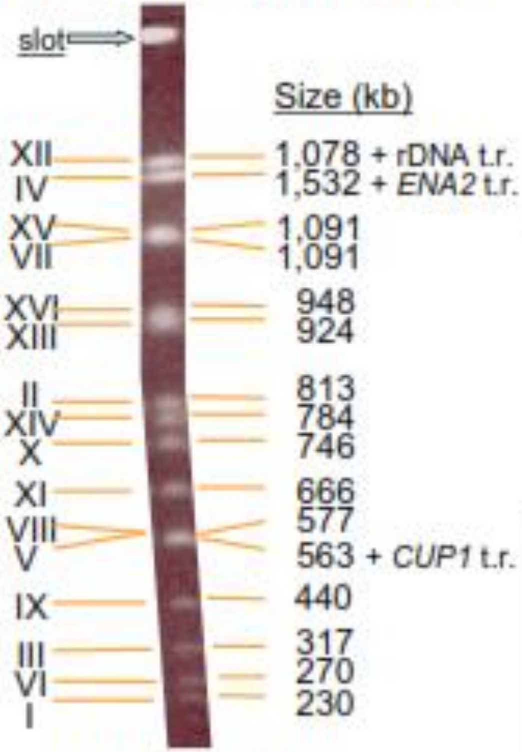
- First completely sequenced eukaryote genome
- Very compact genome:
  - Short intergenic regions
  - Scarcity of introns
  - Lack of repetitive sequences
- Strong evidence of duplication:
  - Chromosome segments
  - Single genes
- Redundancy: non-essential genes provide selective advantage



The exquisite beauty of a eukaryotic genome: *Saccharomyces cerevisiae*

1996 

Nucleus: 16 chromosomes



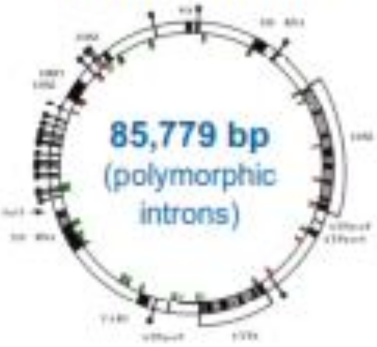
TOTAL 12,071 kb
+ tandem repeats

rDNA 9 kb repeat units
70-120 tandem copies
Total **0.6 - 1.1 Mb**

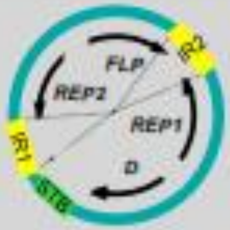
Dispersed repeated sequences

tRNA genes (identical)
Ty elements and LTR (similar)

Mitochondria



ca. 30 copies
Total **2.6 Mb**



two-micron **plasmid**
6,318 bp
ca. 50 copies
Total **0.3 Mb**

RNA "virus" 3 - 5 kb
100-1000 copies
Total **0.3 - 5 Mb**

	Total Noyau	Total Mitoch.
Chromosomes	**16**	**1**
Protein-coding genes		
active CDS	**5 769**	**7**
pseudogenes	77	-
RNA-coding genes		
transfer RNAs	275	23
sno RNAs	77	-
sn RNAs	6	-
ribosomal RNAs	3	2
Other RNAs	> 4	1
Mobile genetic elements		
complete (actifve)	52	-
incomplete (traces)	220	-
Introns		
spliceosomal	273	-
group I and group II	-	1 - 10

Yeast genomes are found in a variety of forms

Haploids, Aneuploids
Homozygous diploids, Heterozygous diploids
Interspecific hybrids, Partial hybrids, mosaics

A resource for functional data in eukaryotes
> 85 % of genes functionally characterized

Salient features of yeast

- single-celled (unicellular) , organism (called budding yeast)
- possess a cellular organization similar to that found in higher, multi-cellular organisms such as humans -eukaryotic.
- The similarity in cellular organization between yeast and higher eukaryotes translates to similarities in their fundamental cellular processes so discoveries in yeast frequently provide direct or indirect clues as to how biological processes work in humans.
- It is an excellent model organism for the study of cell-cycle control, mitosis and meiosis, DNA repair and recombination, and the checkpoint controls important for genome stability.
- **first eukaryotic genome** to be completely sequenced
- Also important in industry, where it is used to make bread, beer, wine, enzymes, and pharmaceuticals.

General advantages of *S. cerevisiae*

- Strains have both a stable haploid and diploid state; Recessive mutations are conveniently manifested in haploid strains and complementation tests can be carried out with diploid strains
- Viable with a large number of markers
- The ease of **gene disruptions** and single step gene replacements offers an outstanding advantage for experimentation
- Yeast genes can functionally be expressed when fused to the green fluorescent protein (GFP) thus allowing to localize gene products in the living cell by fluorescence microscopy
- The yeast system has also proven an invaluable tool to clone and to maintain large segments of foreign DNA in yeast artificial chromosomes (YACs) being extremely useful for other genome projects and to search for protein-protein interactions using the two-hybrid approach
- Transformation can be carried out directly with short single-stranded synthetic oligonucleotides, permitting the convenient productions of numerous altered forms of proteins
- Extensively exploited in the analysis of gene regulation, structure–function relationships of proteins, chromosome structure, and other general questions in cell biology

Therapeutic products from yeasts

- Prokaryotic products: Tetanus toxin fragment C; Streptokinase
- Surface antigens of viruses: Hepatitis B; HIV; Foot and mouse disease; Influenza; Polio; Polyoma; Epstein-Barr; Oncogenic retroviruses
- Malaria antigen
- Animal products: Hirudin; porcine interferon; interleukin; trypsin inhibitor
- Human hormones: Insulin; parathyroid hormone; growth hormone, chorionic gonadotropin
- Human growth factors: IGF1; NGF; EGF; tissue factor; CSF; GM-CSF; TNF
- Human blood proteins: Hemoglobin; factors VIII and XIII; alpha-1-antitrypsin; antithrombin III; serum albumin
- Various human enzymes; CFTR; estrogen receptor; INF-alpha; INF-beta1

The yeast genome

- *S. cerevisiae* contains a haploid set of 16 well-characterized chromosomes, ranging in size from 200 to 2,200 kb
- Total sequence of chromosomal DNA is 12.8 Mb
- 6,183 ORFs over 100 amino acids long
- ***Sequenced in year***: 1996

- ***Strain sequenced***: S288c

Sequencing the *S. cerevisiae* genome

The genome was sequenced by a highly cooperative consortium in the early 1990s, chromosome by chromosome (the whole genome shotgun approach was not used).

This involved 600 researchers in > 100 laboratories.

--Physical map created for all XVI chromosomes
--Library of 10 kb inserts constructed in phage
--The inserts were assembled into contigs

The sequence released in 1996, and published in 1997 (Goffeau et al., 1996; Mewes et al., 1997)

Features of the *S. cerevisiae* genome

Sequenced length:	12,068 kb = 12,068,000 base pairs
Length of repeats:	1,321 kb
Total length:	13,389 kb (~ 13 Mb)

Open reading frames (ORFs):	6,275 (see updates below)
Questionable ORFs (qORFs):	390
Hypothetical proteins:	5,885

Introns in ORFs:	220
Introns in UTRs:	15
Intact Ty elements:	52
tRNA genes:	275
snRNA genes:	40

Features of the *S. cerevisiae* genome

A notable feature of the genome is its high gene density (about one gene every 2 kilobases). Most bacteria have about one gene per kb, but most eukaryotes have a much sparser gene density.

Also, only 4% of *S. cerevisiae* genes are interrupted by introns. By contrast, 40% of genes from the fungus *Schizosaccharomyces pombe* have introns.

Karyotype