Aalto University
School of Chemical
Technology

Synthetic biology

Course CHEM-E8125, spring 2023

Synthetic genomes

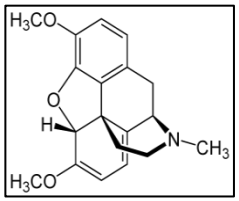
Yeast 2.0 + other

Prof. Merja Penttilä

Synthetic yeast genome *Saccharomyces cerevisiae* 2.0

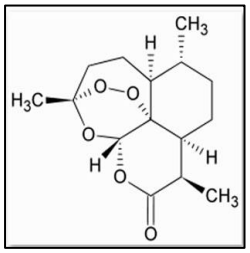
Aalto University
School of Chemical
Technology

Why baker's yeast *Saccharomyces cerevisiae* ?



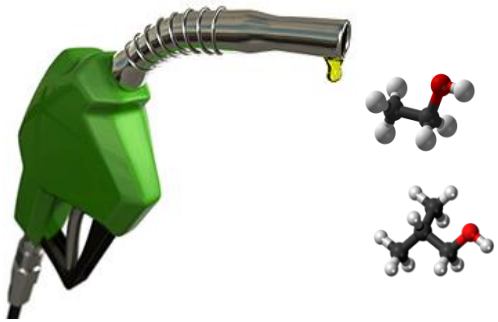
Opiate thebaine

- A companion of man over millennia
- A reliable workhorse



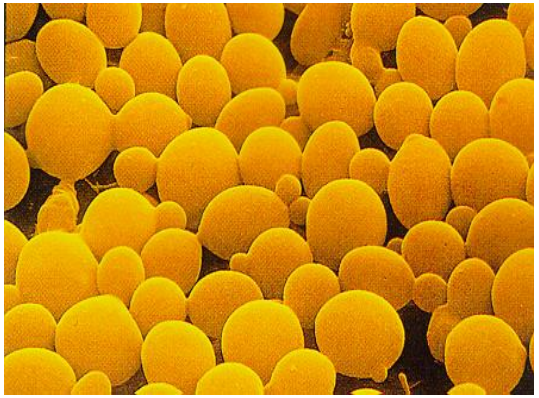
Malaria drug artemisinin

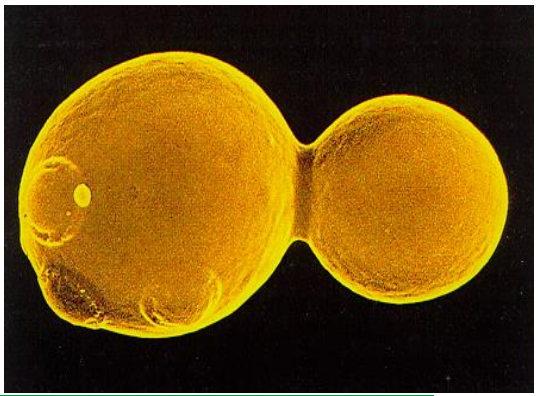
Human insulin



Bioreactors for ethanol production can be over 1000 m^3 (10^6 litres) large

Saccharomyces cerevisiae is a robust process organism



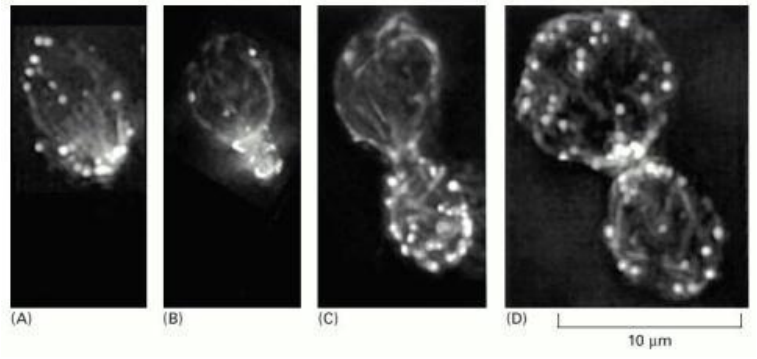


Over 10^8 small cell factories fit in one litre

4 - 5 micrometers

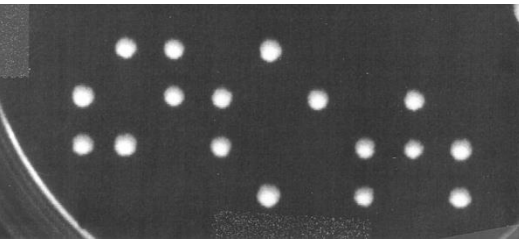
Saccharomyces cerevisiae is a eukaryotic model organism

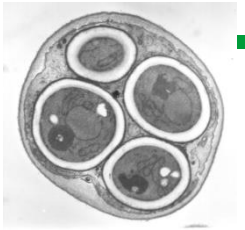
Actin patches and filaments



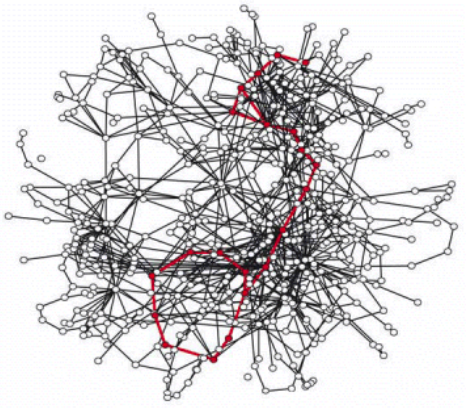
For

- **physiology (anaerobic, aerobic)**
- **genetics (haploid, diploid, polyploid)**
- **cell biology (Nobel prize 2001 to Hartwell for cell cycle regulation)**
- **molecular biology (plasmid transformation (1978) and targeted gene deletion 1983))**
- **genomics (first eukaryotic genome sequenced 1996)**
- **systems biology (-omics methods, metabolic model 2003)**
- **synthetic biology**





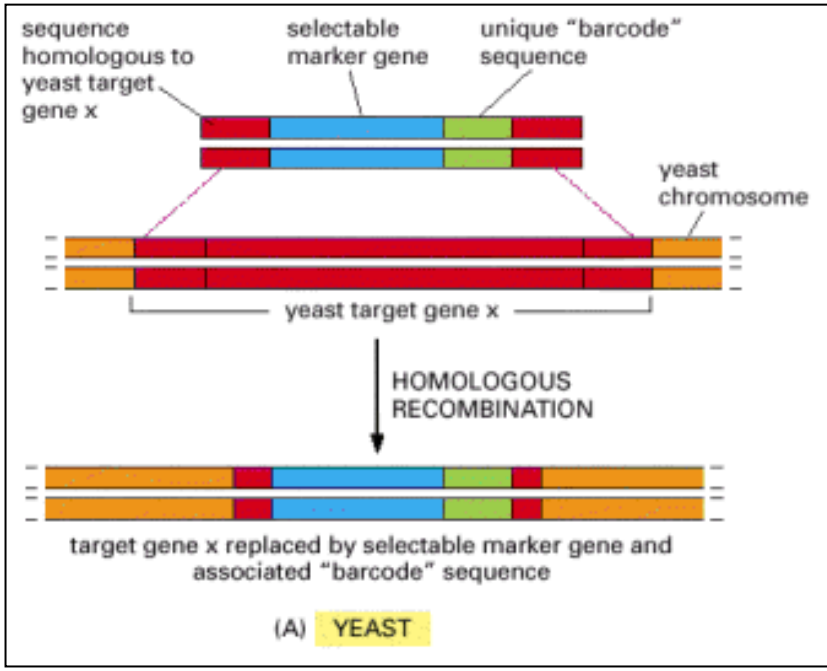
Four spores
In an ascus



Saccharomyces cerevisiae – Molecular tools

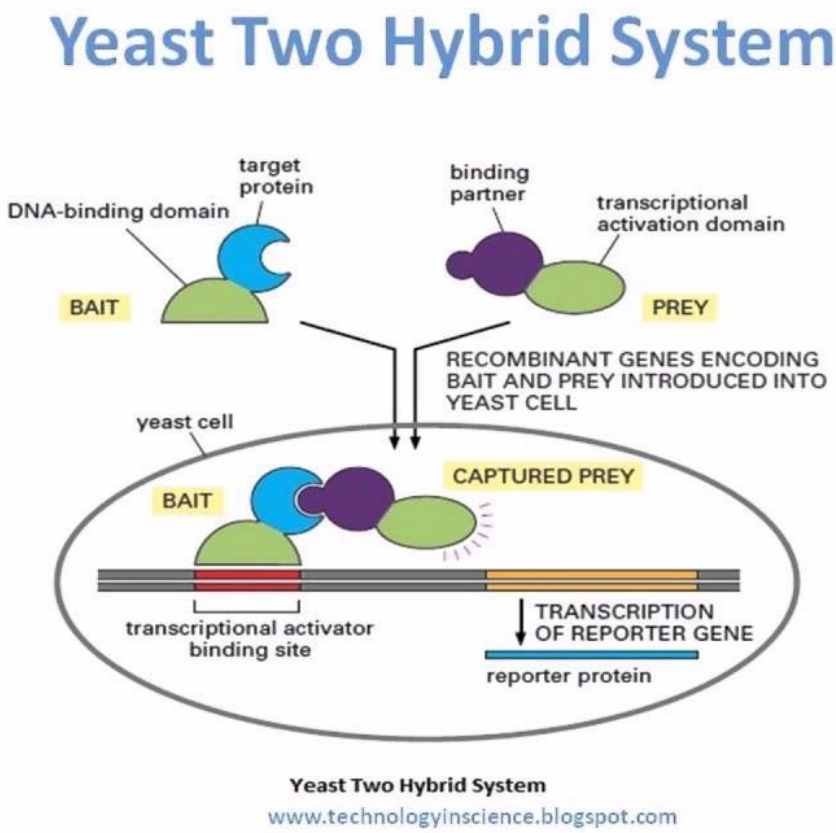
- High transformation frequencies, >> 10^6 /ug DNA
- Versatility of vectors (high/lower copy number plasmids, integrative vectors, single copy telomere vectors, YAC (yeast artificial chromosome)
- High frequency of homologous recombination
- Gene deletion libraries available (in both haploid mating types -> heterozygous diploids)
- Tagged versions of genes for localisation studies
- Two-hybrid libraries for protein-protein interaction studies
- Promoter arrays for e.g. chromatin immuno-precipitation experiments

The deletion libraries are barcoded, 20bp

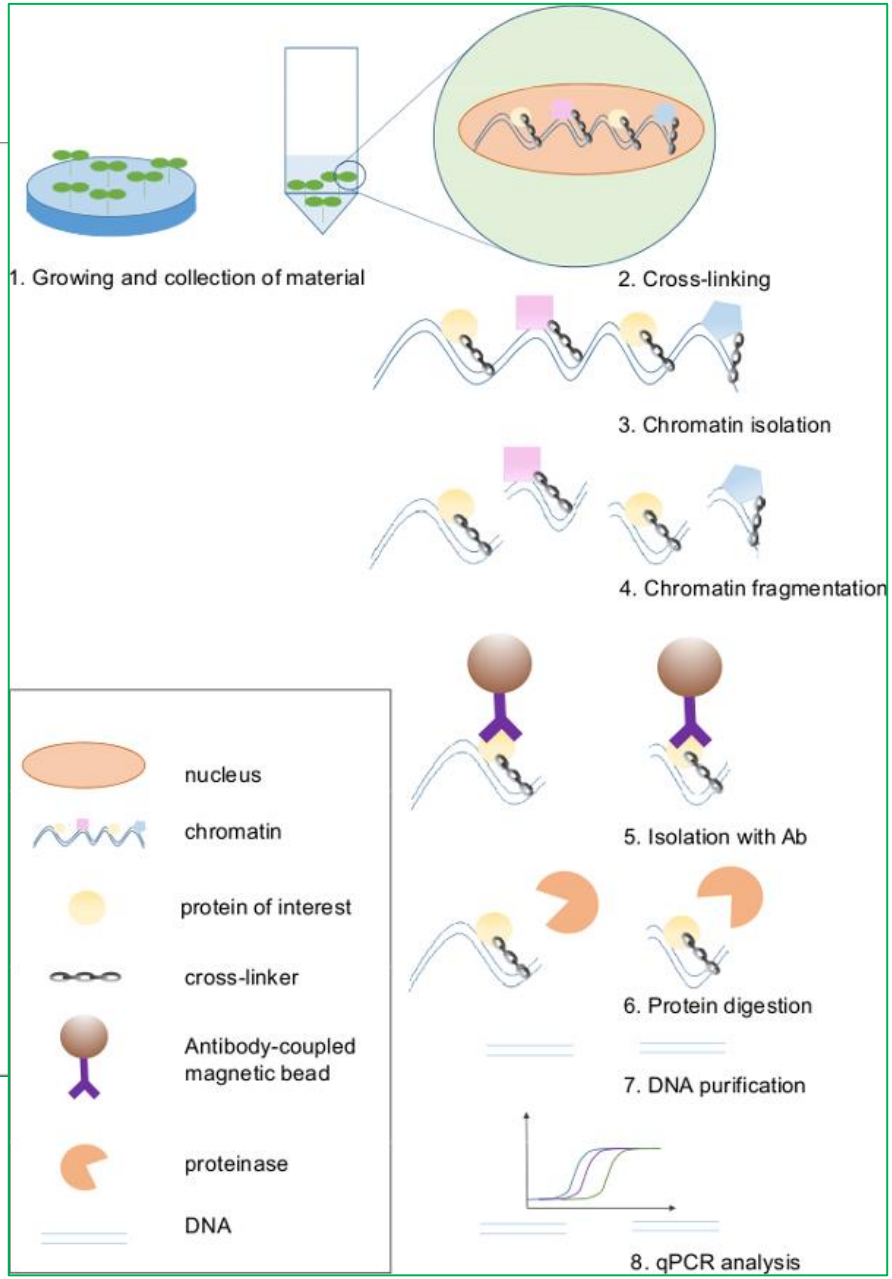


- High through put screens & methods (e.g. disease genes; 70% of mammalian genes with similarity to yeast are functional in yeast)
- Various databases

Yeast two-hybrid system to find all proteins in the cell interacting with the target protein of interest



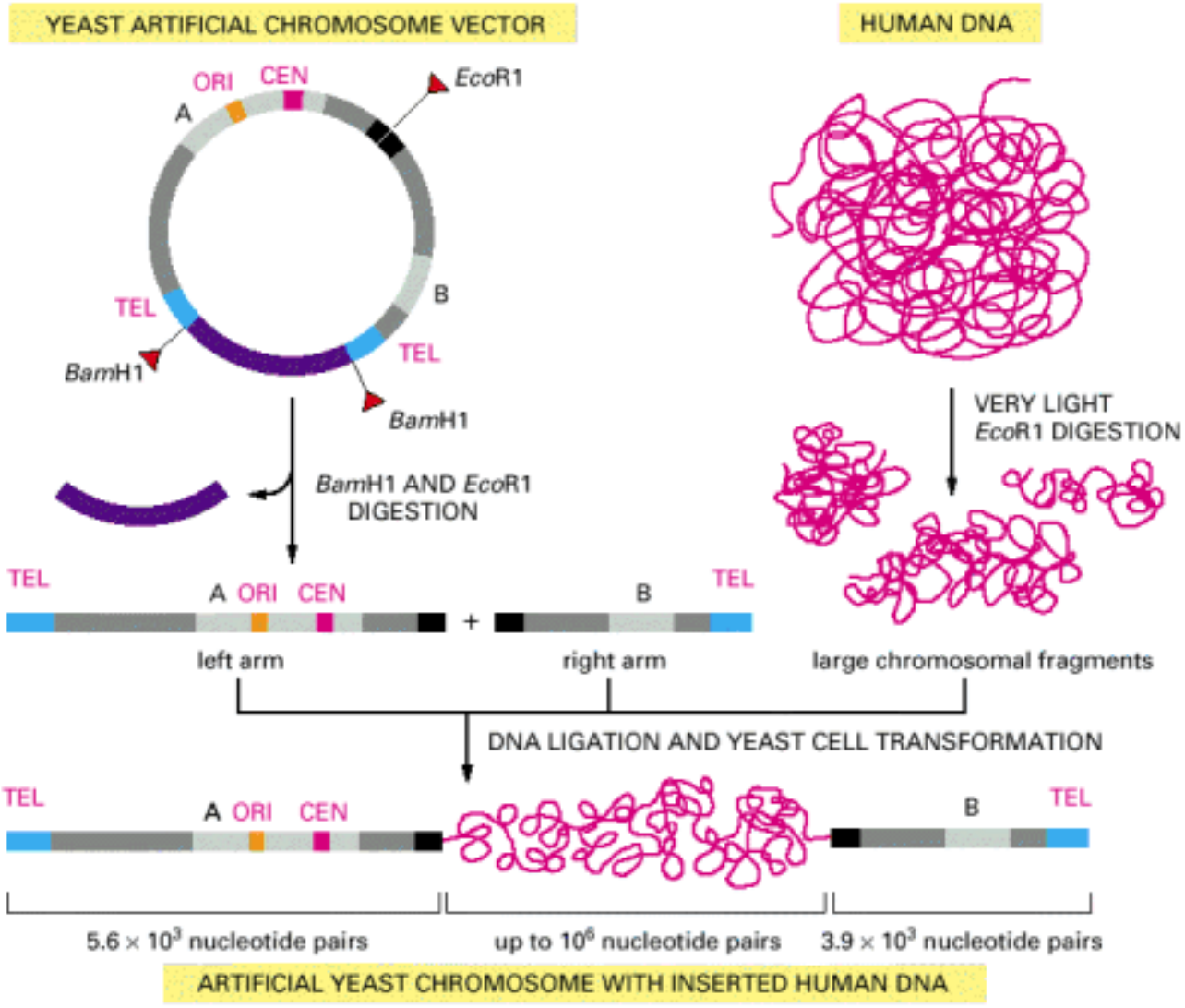
Chromatine immunoprecipitation for finding genome regions binding a particular protein



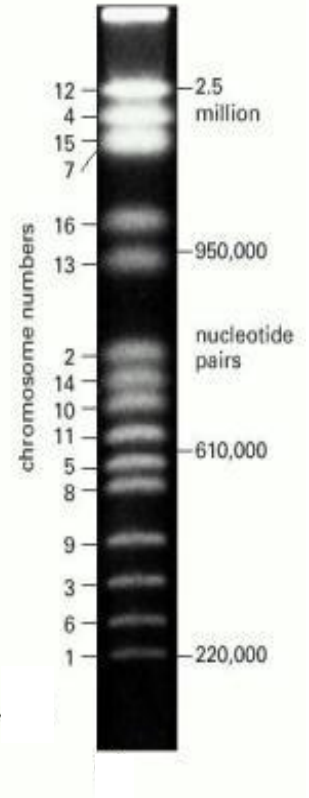
Saccharomyces cerevisiae – artificial chromosomes

- ORI; origin of replication
- CEN; centromere
- TEL; telomere

- Large pieces of DNA can be inserted (>> 50 genes)
- One way to build synthetic genomes
- *Mycoplasma genitalium* genome (0.6Mb) was assembled in yeast



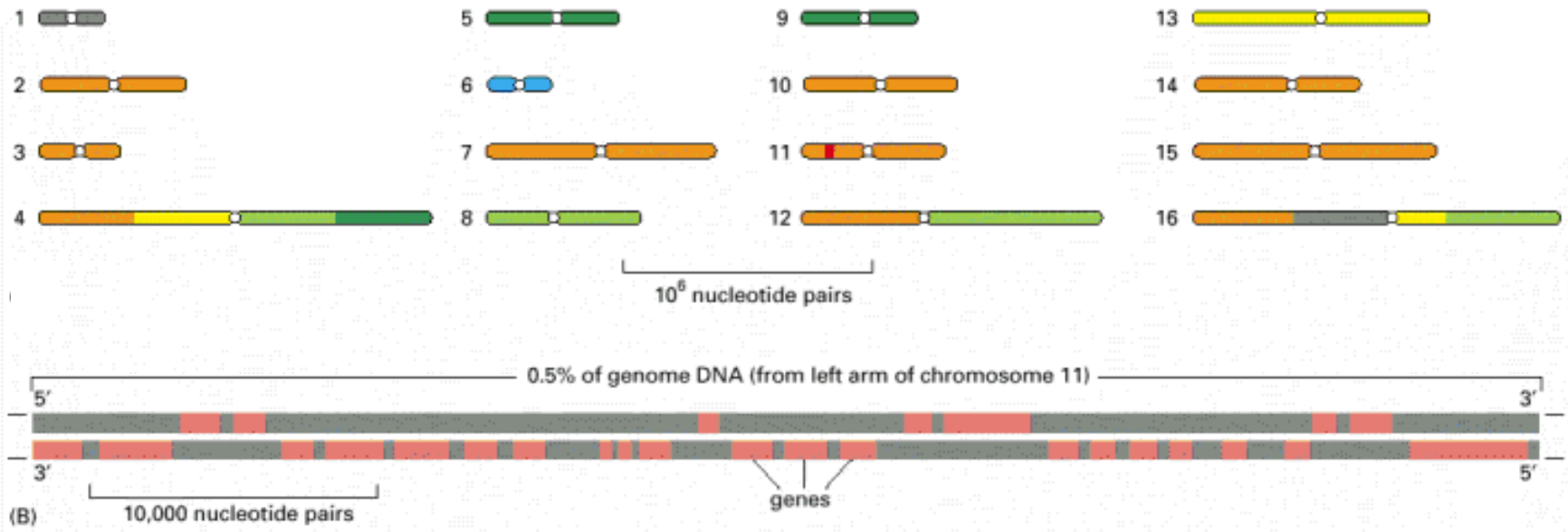
One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. Gibson DG et al. . Proc Natl Acad Sci U S A. 2008 Dec 23;105(51):20404-9



S.cerevisiae chromosomes on a CHEF gel

The *Saccharomyces cerevisiae* genome

- 13.4 Mb in 16 linear chromosomes + mitochondrial DNA
- About 6000 genes
- Compact genome
- Very few introns, 5% of genome (actin gene intron ~500bp, large)
- ~20% ?? of genes without function still



Sequencing of the *Saccharomyces cerevisiae* genome was a global effort

The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. H. Bussey, *et al*., 1995, *PNAS USA* 92:3809-13

Complete DNA sequence of yeast chromosome II. H. Feldmann, *et al*., 1994, *EMBO J* 13(24):5795-809

The complete DNA sequence of yeast chromosome III. S. G. Oliver, *et al*., 1992, *Nature* 357:38-46

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IV. C. Jacq, *et al*. (1997) *Nature* 387(6632 Suppl):75-78

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome V. F. S. Dietrich, *et al.* (1997) *Nature* 387(6632 Suppl):78-81

Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*. Y. Murakami, *et al*., 1995, *Nature Genetics* 10:261-8

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome VII. H. Tettelin, *et al*. (1997) *Nature* 387(6632 Suppl):81-84

Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. M. Johnston, *et al.*, 1994, *Science* 265:2077-82

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. Churcher, *et al.* (1997) *Nature* 387(6632 Suppl):84-87

Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome X. F. Galibert, *et al*., 1996, *EMBO J* 15(9):2031-49

Complete DNA sequence of yeast chromosome XI. B. Dujon, *et al.*, 1994, Nature 369:371-8

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII. M Johnston, *et al*. (1997) *Nature* 387(6632 Suppl):87-90

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIII. S. Bowman, *et al*. (1997) *Nature* 387(6632 Suppl):90-93

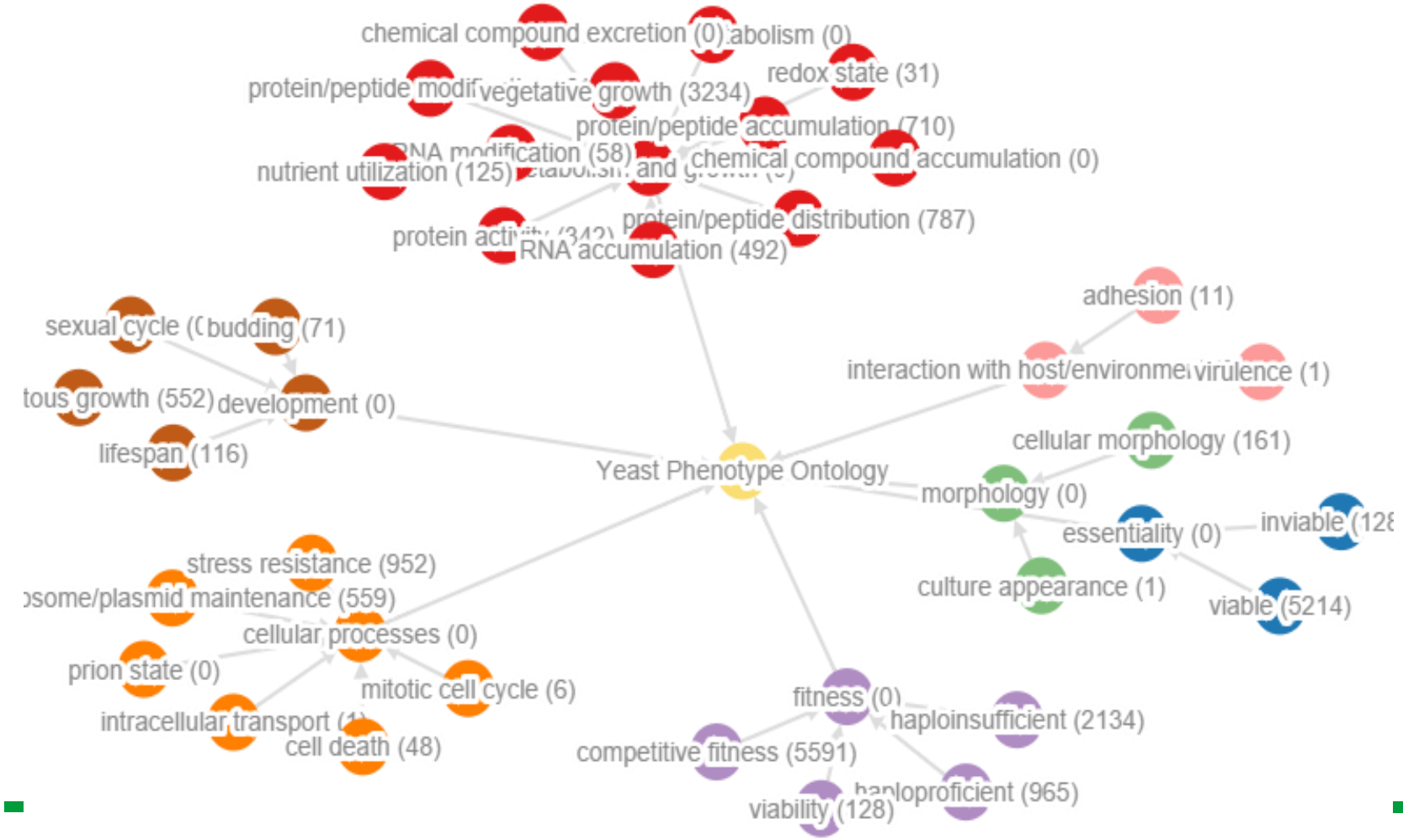
The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV and its evolutionary implications. P. Philippsen, *et al.* (1997) *Nature* 387(6632 Suppl):93-98

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XV and its evolutionary implications. B. Dujon, *et al.* (1997) *Nature* 387(6632 Suppl):98-102

The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI and its evolutionary implications. H. Bussey, *et al.* (1997) *Nature* 387(6632 Suppl):103-105

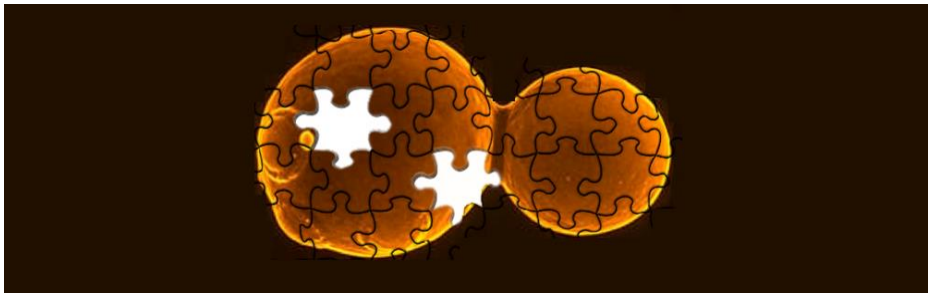
The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. F. Foury, *et al.*, 1998, *FEBS Letters* 440:325-31

Saccharomyces genome database (SGD) contains gene, genome and phenotype information



Current Observable Other Observable

SGD 2015-09-08



Yeast 2.0
"Designer genome"

Jeff Boeke et al, NYU Medical Centre

1. Design principles

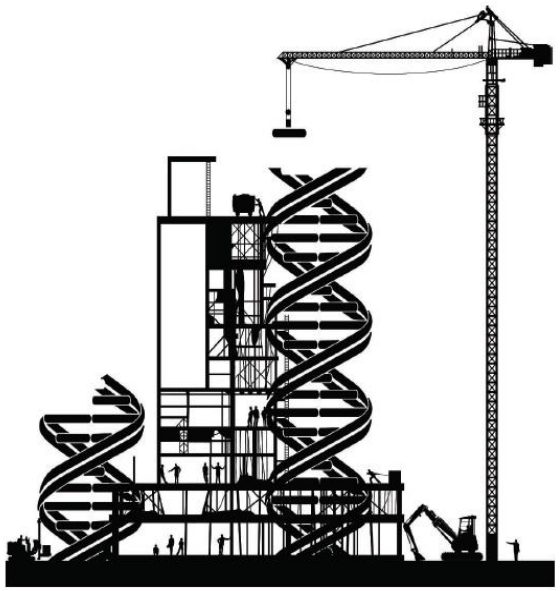
- keep yeastiness
- maintain fitness
- maintain genomic stability
- but increase genetic flexibility; make possible genetic evolution through SCRaMbLE

Questions about species identity and taxonomy

2. Construction principles

- rather conservative
- chromosome by chromosome

A huge construction effort



Building the Ultimate Yeast Genome

(Pennisi 2014)

Design and construction principles of Yeast 2.0

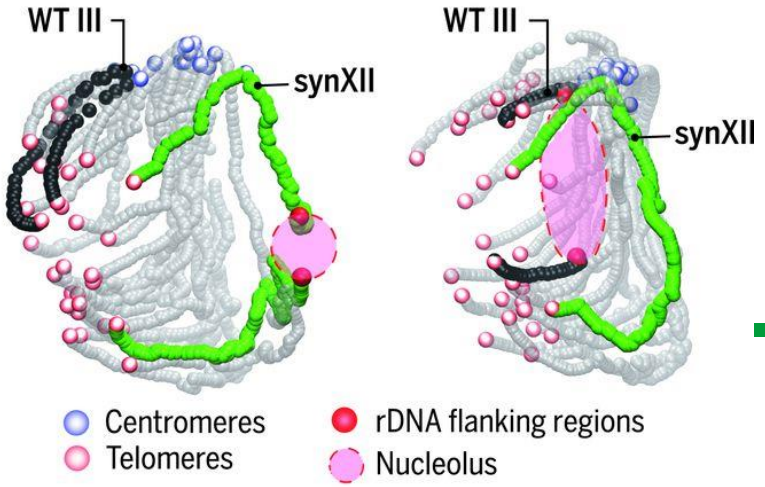
- Rather conservative changes and minimal recoding (= change of DNA sequence) (e.g. due to unknown codon usage or regulatory effects of DNA sequence)
- The synthetic DNA was introduced to the wild-type genome step-wise in 30-60kb pieces replacing the native yeast sequence in the chromosome, “bottom up” approach. This makes it easy to trace back if deleterious changes were introduced.
- Yeast fitness was evaluated in laboratory conditions after each cycle. The cumulative effect of unwanted changes could yield an unviable yeast cell within a relatively low number of cycles, unless monitored.
- Following the success of construction of the first synthetic yeast chromosome, *synIII.*

  N. Annaluru, J. Boeke et al., “Total synthesis of a functional designer eukaryotic chromosome,” *Science*, doi:10.1126/science.1249252, 2014.

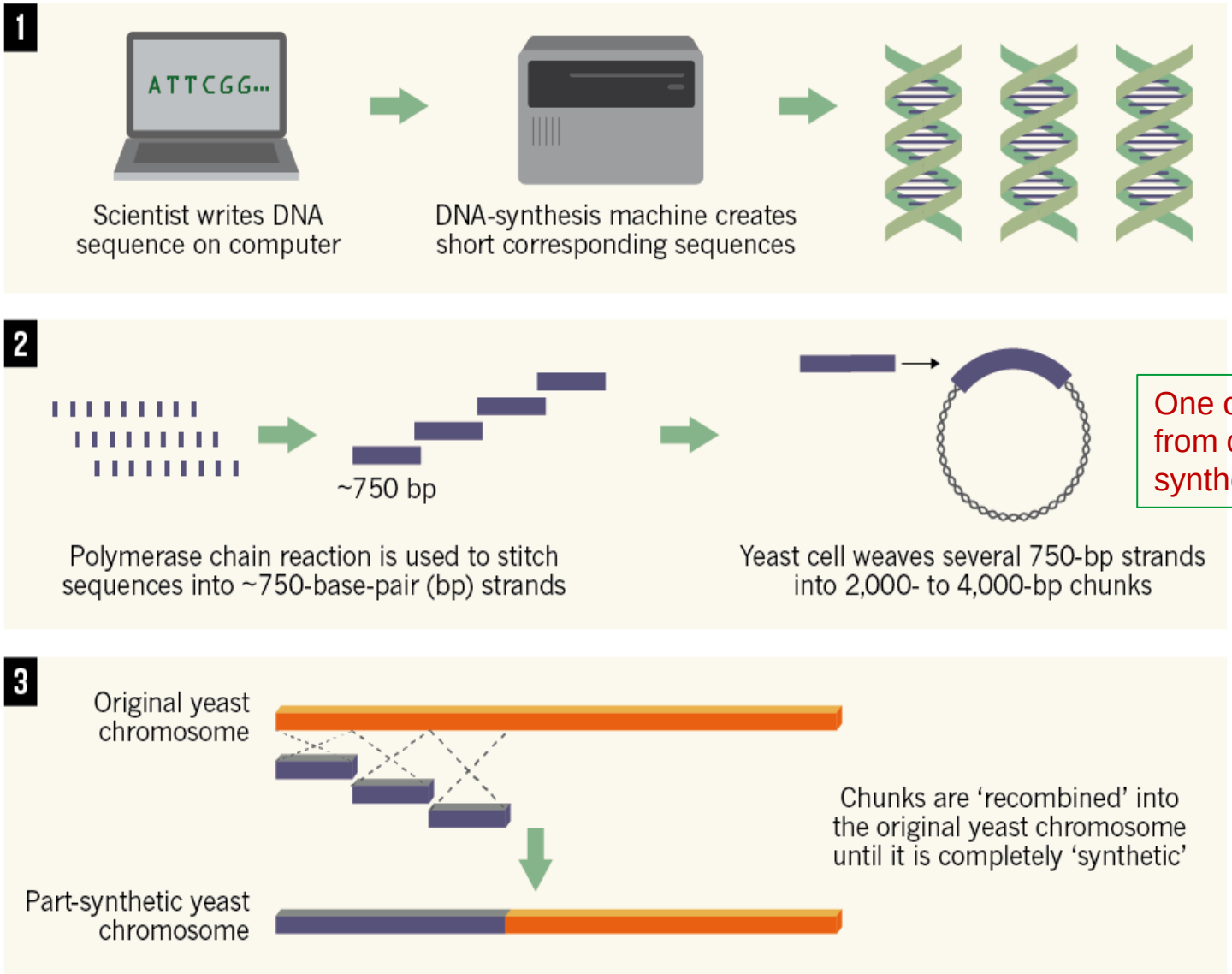
SCIENCE, 10 MARCH 2017, VOL 355, ISSUE 6329

- *synII, synV, synVI, synX, and synXII*
- *30% of yeast DNA synthetic*

1. Design of a synthetic yeast genome
2. Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome
3. “Perfect” designer chromosome V and behavior of a ring derivative
4. Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond
5. Bug mapping and fitness testing of chemically synthesized chromosome X
6. Engineering the ribosomal DNA in a megabase synthetic chromosome
7. 3D organization of synthetic and scrambled chromosomes



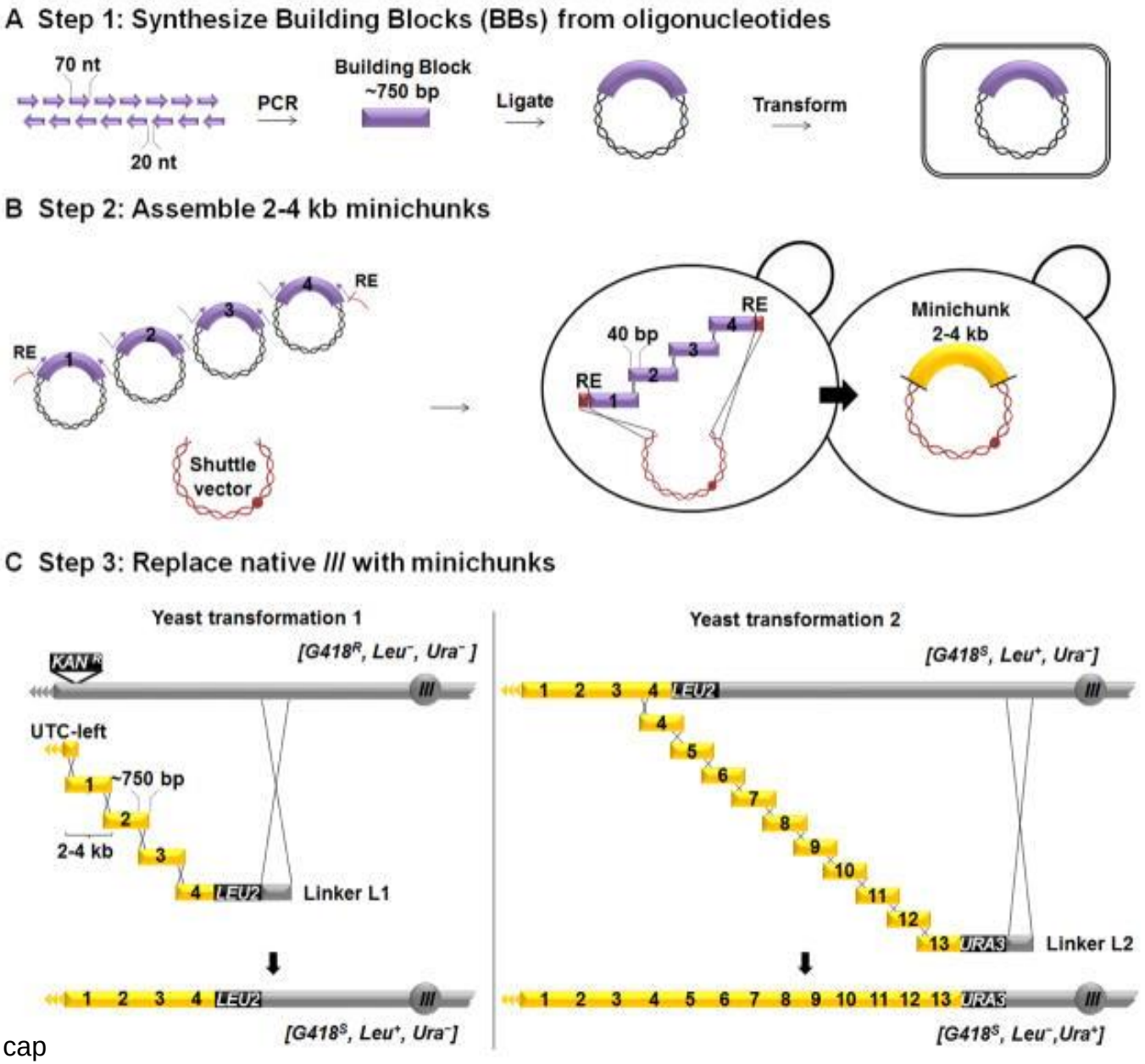
Design and construction principles of Yeast 2.0



One can also order from companies ~10kb synthetic DNA pieces

Building synthetic yeast chromosome III (*synIII*)

Fig. 1 *synIII* construction (**A**) Building block (BB) synthesis. 750 bp BBs (purple) were synthesized from oligonucleotides at the Build-AGenome course. (**B**) Assembly of minichunks. 2–4 kb minichunks (yellow) were assembled by homologous recombination in *S. cerevisiae*. Adjacent minichunks were designed to encode overlap of one BB to facilitate downstream assembly steps. Minichunks were flanked by a rare cutting restriction enzyme (RE) site, *Xma*I or *Not*I. (**C**) Direct replacement of native yeast chromosome *III* with pools of synthetic minichunks. Eleven iterative one-step assemblies and replacements of native genomic segments of yeast chromosome *III* were carried out using pools of overlapping synthetic DNA minichunks, encoding alternating genetic markers (*LEU2* or *URA3*), which enabled complete replacement of native *III* with *synIII* in yeast.



UTC, universal telomere cap

Design and construction principles of Yeast 2.0

Hierarchical assembly plan

- Building blocks (~750bp) are assembled into minichunks (~3kb), which are assembled into chunks (~10kb), which are subsequently assembled into megachunks (30-50kb). The termini of each 10kb chunk encode restriction enzyme sites that enable directional assembly of chunks into megachunks by *in vitro* ligation.
- The megachunks of ~30-60kb (or 3-6 x 10kb chunks) of synthetic DNA are integrated into the yeast genome at a time to replace the corresponding native chromosomal DNA, exploiting homologous recombination that occurs in yeast.
- Current DNA synthesis technologies can readily produce synthetic DNA fragments in the minichunk and chunk range, and DNA synthesis can be done at companies.

Selection of correct integrations in yeast

- The integration of the leftmost megachunk overwrites a kanMX cassette previously introduced into the native chromosome sequence for negative selection purposes.
- As each subsequent megachunk is incorporated, the auxotrophic marker used in the prior round is eliminated by recombination, and selection is imposed for a second selectable marker (e.g. *LEU2*). In this way, the remainder of the synthetic chromosome can be assembled *in vivo*, with alternating selection for *URA3* and *LEU2* markers.
- In addition to monitoring the loss and gain of auxotrophic markers, the integration of synthetic DNA and loss of native sequence is confirmed with PCRtags.

See Figure next page.

Alterations to the Yeast 2.0 genome

PCRTags are incorporated into most open reading frames (ORFs) by recoding a ~20bp segments of the coding regions to a different DNA sequence but encoding the same amino acid sequence. This enables later a quick scan to ascertain that the synthetic substitution of the genome has occurred, and to monitor any changes to the genome later, e.g. after SCRaMbLE (see below).

Synthetic telomere sequences of a simple sequence repeat $(TG_{1\text{-}3})_n$, to reduce chromosome length. These universal telomere caps (UTCs) will be introduced to all chromosome termini.

Removal of introns. There are only about 250-300 introns in the native *S. cerevisiae* genome, and those likely to have no significant impact will be removed. If e.g. small RNA coding regions are present these will be relocated to non-intronic regions.

Removal of non-essential genes for viability/fitness in laboratory conditions, such as mating type genes.

Replacing all TAG stop codons with TAA, allowing for subsequent introduction of new coding schemes with unnatural amino acids (using TAG).

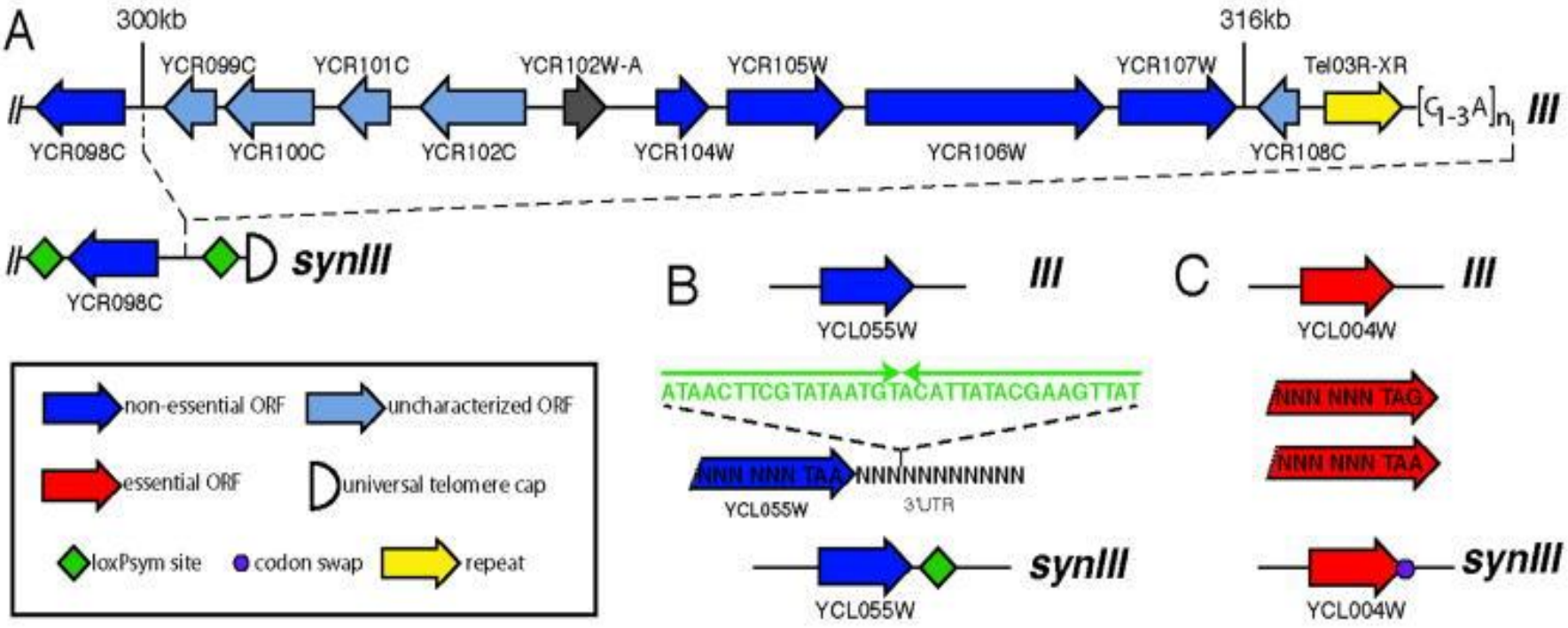
Alterations to the Yeast 2.0 genome

Deletion of transposons to increase the stability of the genome, as well as deletion of *SPT3* gene, required for Ty1 transposon transcription.

Relocation of tRNA genes that are hotspots for genomic instability onto a *separate* centromeric plasmid, a “neochromosome”.

SCRaMbLE. Site-specific recombination sequences, loxP sites, allowing subsequent evolution of the yeast strains will be introduced in all non-essential genes at the 3′ UTRs, as well as points where deletions are made (e.g. tRNA genes, transposons), ultimately resulting in ~5000 sites genome-wide. The loxP spesific recombinase will be expresssed conditionally, e.g. from galactose inducible promoter. This is a powerful tool for evolution and addressing what could constitute a minimal yeast genome (in particular conditions).

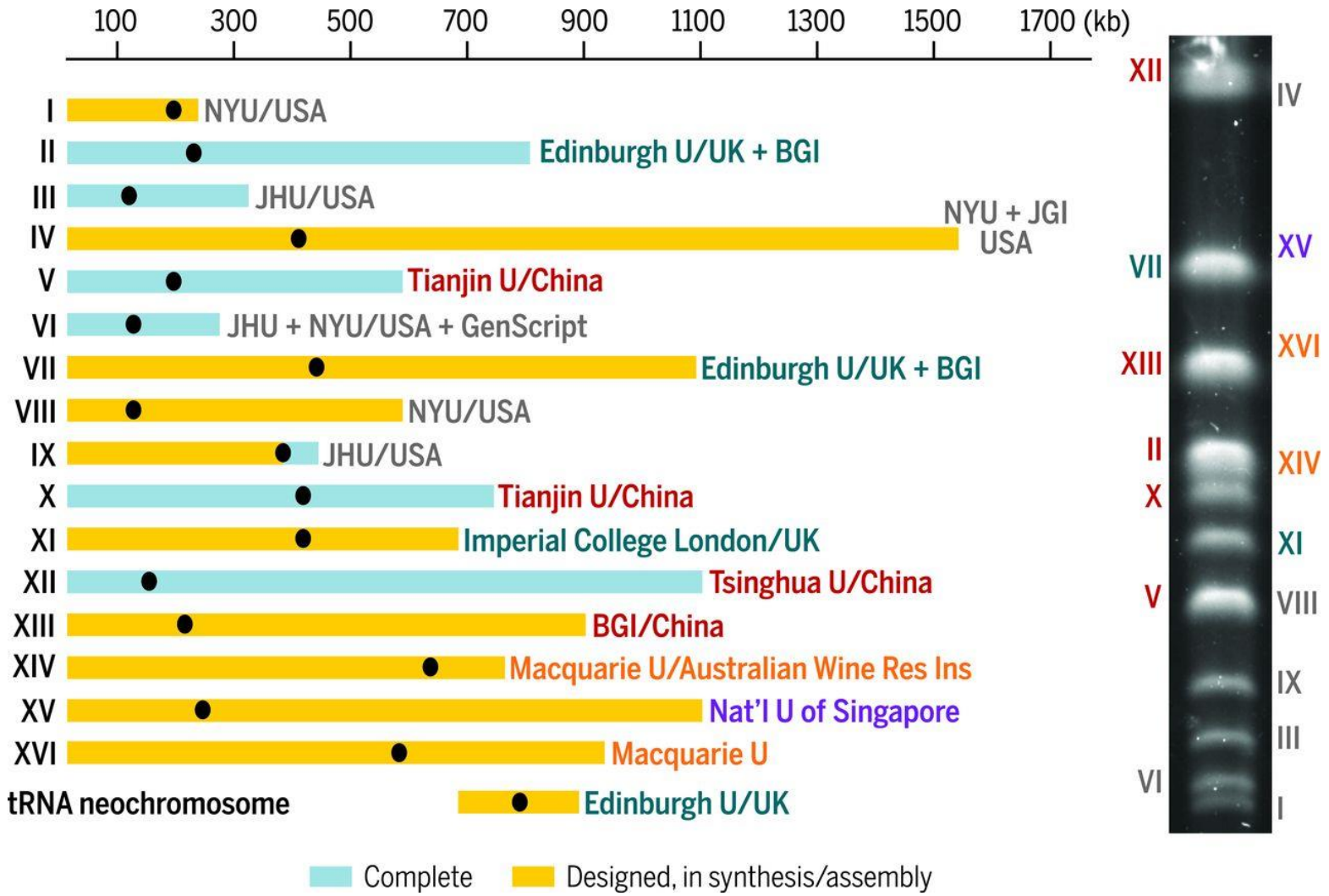
Building synthetic yeast chromosome III (*synIII*)



Representative *synIII* design segments for loxPsym site insertion (A & B) and stop codon TAG to TAA editing (C) are shown. Green diamonds represent loxPsym sites embedded in the 3′ UTR of non-essential genes and at several other landmarks. Fuchsia circles indicate synthetic stop codons (TAG recoded to TAA).

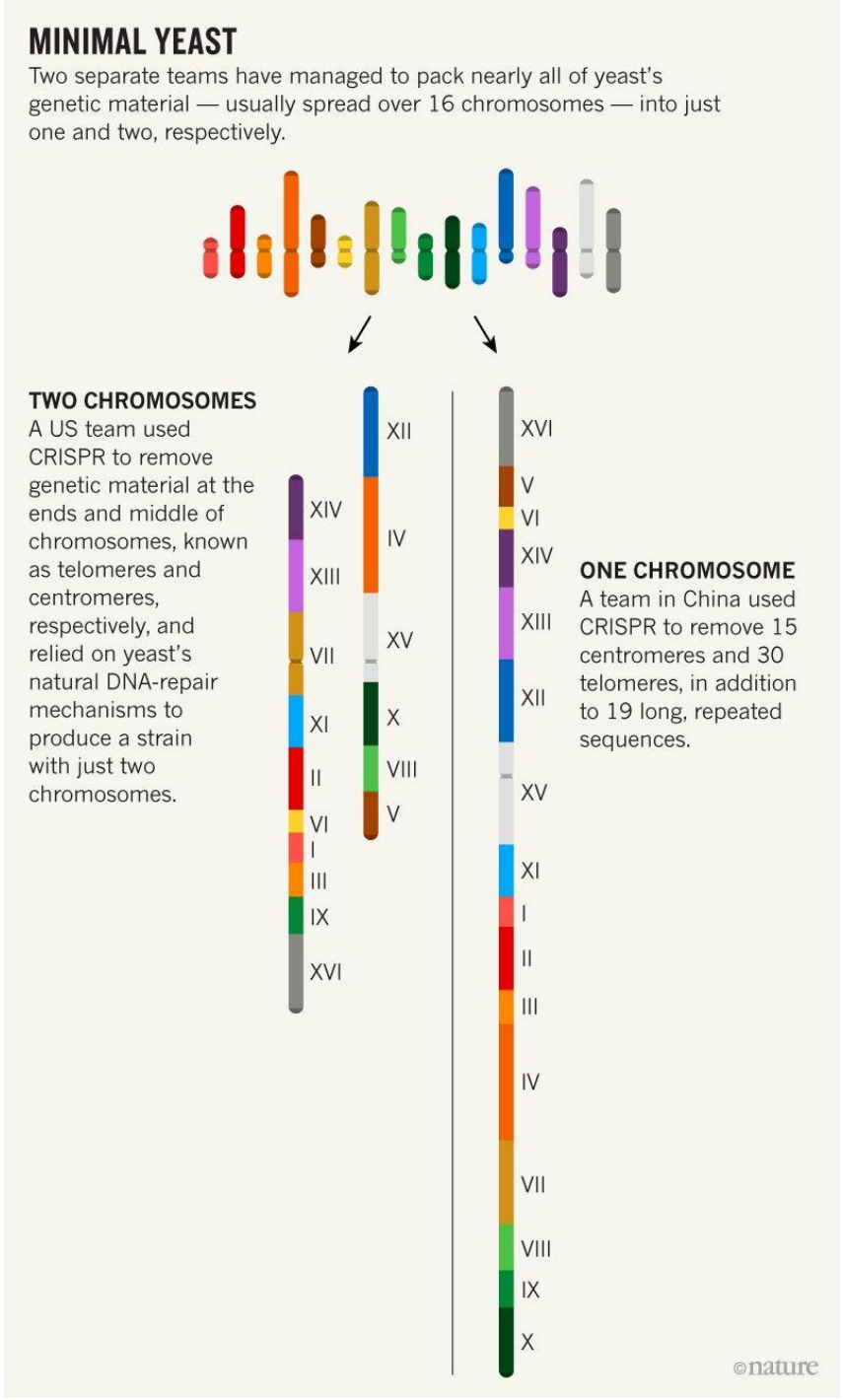
YCR099C: Y (yeast), C (chr.III), R (right from centromere), ORF 99, C/W (Crick/Watson coding strand)

A global consortium is needed for Yeast 2.0



Aalto University
School of Chemical Technology

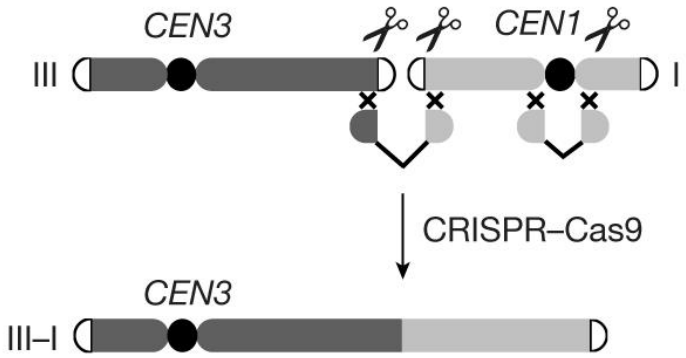
http://syntheticyeast.org/collaborators/



Fusion of yeast chromosomes - Karyotype engineering

Shao et al. Creating a functional single-chromosome yeast. Nature 560, pages 331–335 (2018)

Luo et al. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. Nature 560, pages 392–396 (2018)



Synthetic *Saccharomyces* genome Sc3.0

Dai et al. Genome Biology (2020) 21:205. Sc3.0: revamping and minimizing the yeast genome. https://doi.org/10.1186/s13059-020-02130-z

Sc2.0 had 8% of reduction of genome size and inclusion of a neochromosome, without major effects on the yeast performance.

Sc3.0, building of a more compact genome is starting.

- Start with Sc2.0.
- Relocate all ~ 1000 essential genes, including their regulatory sequences from each chromosome, to a centromeric plasmid. These form essential gene arrays, eArray.
- Build one circular centromere-containing DNA including all eArrays, or a linear chromosome derived from it, synE.
- Make it non-ScRaMbLEable, so it will remain intact.
- Scamble the rest of the genome or do any specific designs.

Synthetic bacterial genomes

Science 329:52- (2010)

Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome

Daniel G. Gibson,[1] John I. Glass,[1] Carole Lartigue,[1] Vladimir N. Noskov,[1] Ray-Yuan Chuang,[1] Mikkel A. Algire,[1] Gwynedd A. Benders,[2] Michael G. Montague,[1] Li Ma,[1] Monzia M. Moodie,[1] Chuck Merryman,[1] Sanjay Vashee,[1] Radha Krishnakumar,[1] Nacyra Assad-Garcia,[1] Cynthia Andrews-Pfannkoch,[1] Evgeniya A. Denisova,[1] Lei Young,[1] Zhi-Qing Qi,[1] Thomas H. Segall-Shapiro,[1] Christopher H. Calvey,[1] Prashanth P. Parmar,[1] Clyde A. Hutchison III,[2] Hamilton O. Smith,[2] J. Craig Venter[1,2]*

We report the design, synthesis, and assembly of the 1.08–mega–base pair *Mycoplasma mycoides* JCVI-syn1.0 genome starting from digitized genome sequence information and its transplantation into a *M. capricolum* recipient cell to create new *M. mycoides* cells that are controlled only by the synthetic chromosome. The only DNA in the cells is the designed synthetic DNA sequence, including "watermark" sequences and other designed gene deletions and polymorphisms, and mutations acquired during the building process. The new cells have expected phenotypic properties and are capable of continuous self-replication.

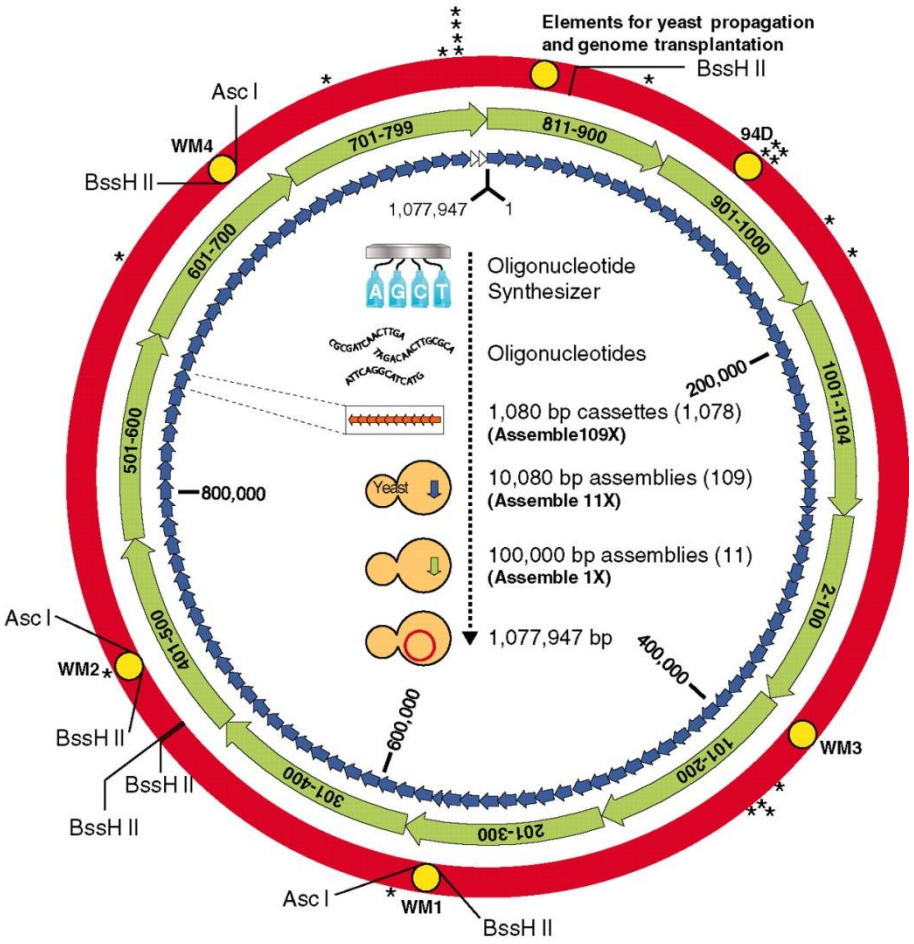
crude *M. mycoides* or *M. capricolum* extracts, or by simply disrupting the recipient cell's restriction system (*8*).

We now have combined all of our previously established procedures and report the synthesis, assembly, cloning, and successful transplantation of the 1.08-Mbp *M. mycoides* JCVI-syn1.0 genome, to create a new cell controlled by this synthetic genome.

Synthetic genome design. Design of the *M. mycoides* JCVI-syn1.0 genome was based on the highly accurate finished genome sequences of two laboratory strains of *M. mycoides* subspecies *capri* GM12 (*8*, *9*, *11*). One was the genome donor used by Lartigue *et al.* [GenBank accession CP001621] (*10*). The other was a strain created by transplantation of a genome that had been cloned and engineered in yeast, YCpMmyc1.1-Δ*typeIIIres* [GenBank accession CP001668] (*8*). This project was critically dependent on the accuracy of these sequences. Although we believe that both finished *M. mycoides* genome sequences are reli-

- Changing *Mycoplasma capricolum* cells to *M. genitalium* JCVI syn1.0 genome.
- A species changed into another with the use of a synthetic genome – "Synthetic cell".
- One single nucleotide mistake (deletion in dnaA-DNA replication) caused a severe problem for the project.

Gibson et al. Science 329:52- (2010)



The assembly of a synthetic M. mycoides genome in yeast. A synthetic M. mycoides genome was assembled from 1078 overlapping DNA cassettes in three steps. In the first step, 1080-bp cassettes (orange arrows), produced from overlapping synthetic oligonucleotides, were recombined in sets of 10 to produce 109 ~10-kb assemblies (blue arrows). These were then recombined in sets of 10 to produce 11 ~100-kb assemblies (green arrows). In the final stage of assembly, these 11 fragments were recombined into the complete genome (red circle). With the exception of two constructs that were enzymatically pieced together in vitro (27) (white arrows), assemblies were carried out by in vivo homologous recombination in yeast. Major variations from the natural genome are shown as yellow circles. These include four watermarked regions (WM1 to WM4), a 4-kb region that was intentionally deleted (94D), and elements for growth in yeast and genome transplantation. In addition, there are 20 locations with nucleotide polymorphisms (asterisks). Coordinates of the genome are relative to the first nucleotide of the natural M. mycoides sequence. The designed sequence is 1,077,947 bp. The locations of the Asc I and BssH II restriction sites are shown. Cassettes 1 and 800-810 were unnecessary and removed from the assembly strategy (11). Cassette 2 overlaps cassette 1104, and cassette 799 overlaps cassette 811.

Total synthesis of *Escherichia coli* with a recoded genome

Fredens et al, Nature 569, pages 514–518 (2019)

E.coli Syn61

Nature uses 64 codons to encode the synthesis of proteins from the genome, and chooses 1 sense codon—out of up to 6 synonyms—to encode each amino acid. Synonymous codon choice has diverse and important roles, and many synonymous substitutions are detrimental. Here we demonstrate that the number of codons used to encode the canonical amino acids can be reduced, through the genome-wide substitution of target codons by defined synonyms. We create a variant of *Escherichia coli* with a four-megabase synthetic genome through a high-fidelity convergent total synthesis. Our synthetic genome implements a defined recoding and refactoring scheme—with simple corrections at just seven positions—to replace every known occurrence of two sense codons and a stop codon in the genome. Thus, we recode 18,214 codons to create an **organism with a 61-codon genome; this organism uses 59 codons to encode the 20 amino acids, and enables the deletion of a previously essential transfer RNA**.

Fig. 1. Design of the synthetic genome, implementing a defined recoding scheme for synonymous codon compression.

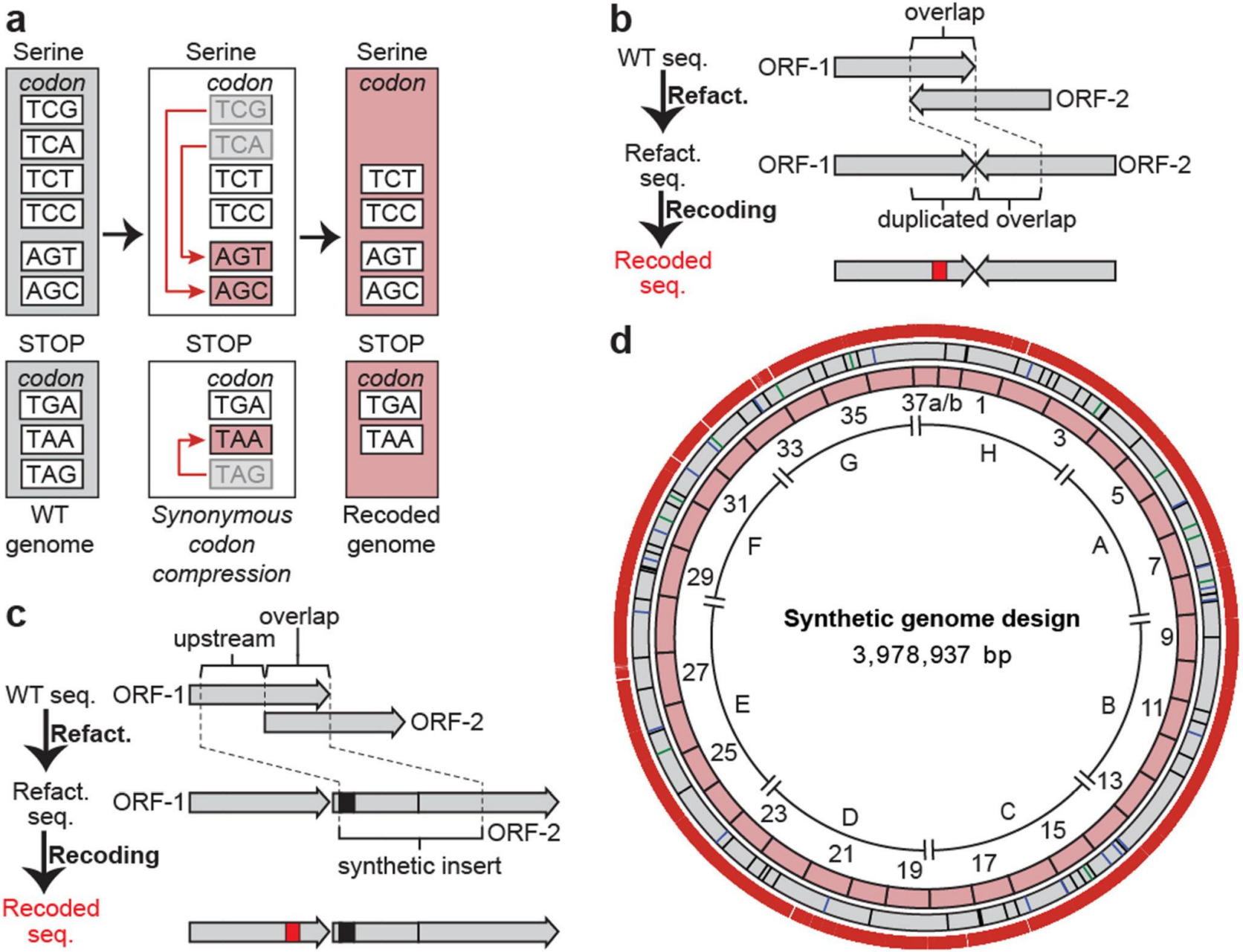


Fig. 1. Design of the synthetic genome, implementing a defined recoding scheme for synonymous codon compression.

a, The defined recoding scheme for synonymous codon compression. Synonymous **serine codons** and three stop codons used in the genome of wild-type E. coli are shown (grey boxes). Systematically implementing a defined recoding scheme for synonymous codon compression (red arrows) recodes target codons to defined synonyms, and **replaces the amber stop codon TAG with the ochre stop codon TAA**. This creates an organism with a **recoded genome that uses a reduced number of serine and termination codons** (pink boxes).

b, Refactoring of 3′, 3′ overlaps enables their independent recoding. **The overlap between two ORFs (ORF1 and ORF2) is duplicated, which enables independent recoding** (red box) of these ORFs.

c, Refactoring 5′, 3′ overlaps. The overlap plus 20 bp upstream is duplicated to generate a synthetic insert. When the overlap is longer than 1 bp at the end of the upstream ORF, an in-frame TAA (black box) is introduced in the beginning of the synthetic insert; this in-frame stop codon ensures the termination of translation from the original ribosome-binding site. Thus, all full-length translation of the downstream ORF is initiated from the **reconstructed ribosome-binding site** in the synthetic insert. This refactoring enables the independent recoding (red box) of ORFs.

d, Map of the **synthetic genome design with all TCG, TCA and TAG codons removed**. Outer ring shows positions (18,218 red bars) of all TCG to AGC, TCA to AGT and TAG to TAA recodings. Grey ring shows positions of designed silent mutations in overlaps (12 green bars), refactoring of 3′, 3′ overlaps (schematic shown in **b,** 21 blue bars) and refactoring of 5′, 3′ overlapping regions (schematic shown in c, 58 black bars). Pink ring shows 37 fragments of approximately 100 kb in size each. Fragment 37 is shown as 37a and 37b to reflect the final assembly. The sections A to H are indicated.

New synthetic genomes to come

Genome Project-write

Launched in 2016
Jeff Boeke, George Church, Andrew Hessel, Nancy J. Kelley et al
https://engineeringbiologycenter.org/

Human
Plant
etc

Aim is to reduce the costs of engineering and testing large genomes, including human genomes, more than 1000-fold within ten years. The group will also seek to develop new technologies, transformative medical applications, and an ethical framework for genome-scale engineering.

Reading

http://syntheticyeast.org/

N. Annaluru, J. Boeke et al., “Total synthesis of a functional designer eukaryotic chromosome,” *Science*, 2014. doi:10.1126/science.1249252,

Richardson et al, Design of a Synthetic Yeast Genome, Science 2017

http://www.yeastgenome.org/ SGD Saccharomyces genome database

YeastBook: An Encyclopedia of the Reference Eukaryotic Cell

Genome engineering: Synthetic genome technology for yeast.
Casci, Tanita // Nature Reviews Genetics; Nov2011, Vol. 12 Issue 11, p742

Group work

Prepare together a **15 min presentation** that contains:

A synthetic design of a ~30-50kb region (”megachunk”) of a selected *Saccharomyces cerevisiae* chromosome, following the same design and construction rules as used for creation of Sc2.0 (slides 18, 19).

- Why did you choose this region?
- Tell what is in the selected region (genes, introns etc). What would you include or omit from the design? (No need to go for a single nucleotide level). Illustrate as in slide 20.
- Which computer programs would you use/need?
- Brief explanation of the wet lab construction procedure

Answers to the questions:

- Explain the Scramble mechanism
- What is the significance and impact of the yeast 2.0?
- What would you use the yeast for or develop further? How?

You may also point out the possible problems you encounted in your group work.

Send the slides to merja.penttila@vtt.fi by noon the 25.3. the latest.
Present as a group. Speak clearly and slowly.