

## RESEARCH ARTICLE

## SYNTHETIC BIOLOGY

## Design of a synthetic yeast genome

Sarah M. Richardson,<sup>1,2\*</sup> Leslie A. Mitchell,<sup>2,3</sup> Giovanni Stracquadanio,<sup>1,2,4</sup> Kun Yang,<sup>1,2</sup> Jessica S. Dymond,<sup>2†</sup> James E. DiCarlo,<sup>2‡</sup> Dongwon Lee,<sup>1§</sup> Cheng Lai Victor Huang,<sup>2</sup> Srinivasan Chandrasegaran,<sup>5</sup> Yizhi Cai,<sup>2,6</sup> Jef D. Boeke,<sup>2,3#</sup> Joel S. Bader<sup>1,2#</sup>

We describe complete design of a synthetic eukaryotic genome, Sc2.0, a highly modified *Saccharomyces cerevisiae* genome reduced in size by nearly 8%, with 1.1 megabases of the synthetic genome deleted, inserted, or altered. Sc2.0 chromosome design was implemented with BioStudio, an open-source framework developed for eukaryotic genome design, which coordinates design modifications from nucleotide to genome scales and enforces version control to systematically track edits. To achieve complete Sc2.0 genome synthesis, individual synthetic chromosomes built by Sc2.0 Consortium teams around the world will be consolidated into a single strain by “endoreduplication intercross.” Chemically synthesized genomes like Sc2.0 are fully customizable and allow experimentalists to ask otherwise intractable questions about chromosome structure, function, and evolution with a bottom-up design strategy.

The goal of the Sc2.0 project is the complete synthesis of a custom-designed genome for a eukaryotic model organism to serve as a platform for systematic studies of eukaryotic chromosomes. The global Sc2.0 effort to build chromosomes is distributed across many locations. This unique aspect of Sc2.0 motivated the design of aspects of BioStudio software that enable a common assembly strategy and enforce a shared language to describe both intermediate designed chromosomes and living strains.

The starting point for the Sc2.0 genome sequence is the highly curated *Saccharomyces cerevisiae* sequence (1, 2). The principles guiding Sc2.0 genome design balance a desire to maintain a “wild-type” phenotype while introducing inducible genetic flexibility and minimizing sources of genomic instability resulting from the repetitive nature

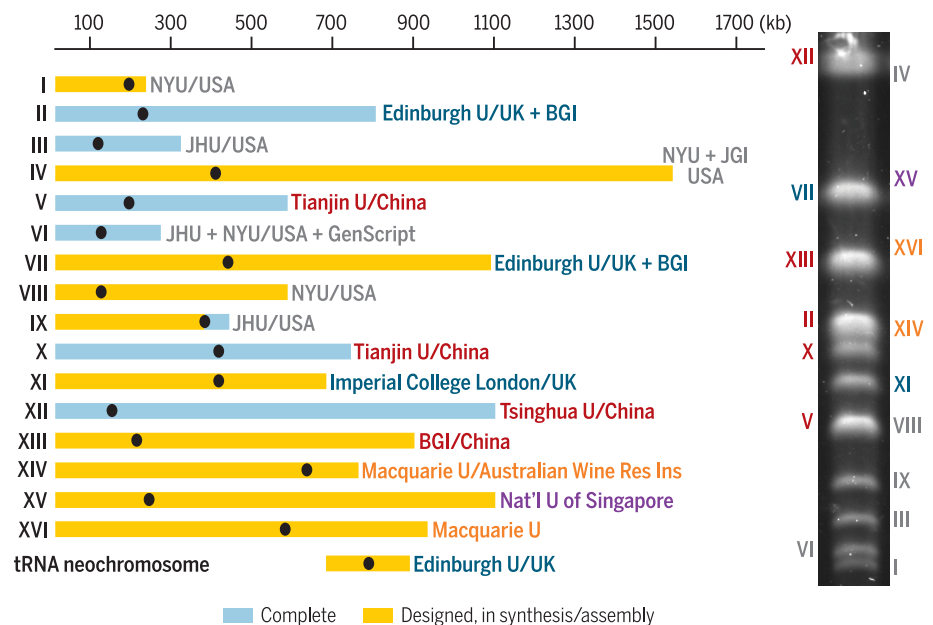
of native yeast DNA. The Sc2.0 chromosomes are therefore designed to encode a slightly modified genetic code in which all TAG stop codons are changed to TAA (3); to include loxP sites that undergird the inducible evolution system SCRAMBLE (4, 5); and to lack repeat elements, transfer RNA (tRNA) genes (relocated to a “neochromo-

some”), and many introns. Further, short recoded sequences within open reading frames (ORFs), called PCRTags, facilitate a polymerase chain reaction (PCR)-based assay to distinguish wild-type from synthetic DNA (5–7). Finally, base substitutions within ORFs introduce or remove enzyme recognition sites to facilitate assembly of synthetic chromosomes.

To implement Sc2.0 redesign on a genome-wide scale, it was necessary to establish not only a generic DNA assembly pipeline for 200 kb– to >1 Mb-sized synthetic chromosomes but also flexible computational tools. We have now completed design of the Sc2.0 genome and describe the effort here. To date, 6.5 Sc2.0 designer chromosomes have been constructed in discrete strains (5, 8–13), and here, we show consolidation of 2.5 synthetic chromosomes into a single strain. Near-wild-type phenotypes of the synthetic strains are consistent with widespread tolerance of designer features and indicates overall robustness of *S. cerevisiae* to genetic manipulation.

## Sc2.0 genome design

The design specification stage of a genome-engineering project is crucial to ensure that the novel sequence robustly supports the intended function in vivo and also enables efficient and scalable assembly that can be implemented at multiple facilities in parallel. The Sc2.0 design was specified relative to the *S. cerevisiae* reference sequence on the basis of derivatives of the S288C strain [sequence last updated by the *Saccharomyces* Genome Database (SGD) on 3 February 2011] as



**Fig. 1. Sc2.0 Consortium chromosome assignments.** The chromosomes (roman numerals) and their approximate sizes are indicated, along with the lead institution(s) involved in performing chromosome synthesis, assembly, and debugging, as well as raising the needed funds. The gel on the right shows a typical pulsed-field gel of a wild-type strain, with the chromosome-sized DNA molecules identified. Black dots indicate chromosome centromeres. BGI, Beijing Genomics Institute; JGI, Joint Genome Institute of the U.S. Department of Energy; JHU, The Johns Hopkins University; NYU, New York University; U, University; Res Inst, Research Institute.

<sup>1</sup>Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205, USA. <sup>2</sup>High Throughput Biology Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. <sup>3</sup>Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, New York University Langone Medical Center, New York, NY 10016, USA. <sup>4</sup>School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK. <sup>5</sup>Department of Environmental Health Science, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA. <sup>6</sup>University of Edinburgh, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK.

\*Present address: Ignition Genomics, 12240 Roundwood Road, Unit 703, Timonium, MD 21093, USA. †Present address: Research and Exploratory Development Department, Biological Sciences, Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Mail Stop 21-N352, Laurel, MD 20723, USA. ‡Present address: Department of Ophthalmology, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA. §Present address: McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA. ||These authors contributed equally to this work. #Corresponding author. Email: jef.boeke@nyumc.org (J.D.B.); joel.bader@jhu.edu (J.S.B.)

a series of edits involving deletions, insertions, and base substitutions. Edits are densely spaced, with a mean distance between clusters of edits of 400 base pairs (bp) when mapped back to the reference genome (Table 1). For each chromosome, the design process involved collaboration between a yeast genetics and/or genomics specialist and a computational specialist; both parties used the BioStudio design platform to communicate and track all changes made to the native sequence.

### Modularity for chromosome and genome assembly

Replacing individual *S. cerevisiae* chromosomes with synthetic versions in a single step is challenging. Our assembly strategy exploits the endogenous homologous recombination machinery to replace individual 30- to 60-kb segments of each wild-type chromosome with the corresponding synthetic sequence. The fitness of the resulting recombinant semisynthetic strains is assessed, and any substitution that proves lethal or leads to a measurable fitness defect can be corrected, typically by reverting the sequence to wild type (“debugging”). The hierarchical nature of the assembly scheme facilitates debugging, as specific designer features can be corrected and fixed once bugs are identified (11). This facilitates a “design-build-assemble-test-learn” cycle used in the final stage of production of synthetic chromosomes. The bottom-up assembly strategy introduces constraints for Sc2.0 that are enforced by complementary top-down design requirements; two examples are requirements for genome-wide uniqueness of PCRTags and requirements for hierarchical segmentation of chromosomes for SwAP-In (switching auxotrophies progressively for integration) (5).

To accelerate completion of Sc2.0, we decentralized the project by parceling out assembly of individual synthetic chromosomes to different teams around the world (Fig. 1). As assembly of the various synthetic chromosomes is completed, we have developed an efficient meiotic strategy to combine them, shown here for the consolidation of synIII (8), synVI (9), and synIXR (5). To ensure consistency, teams are required to adhere to the design articulated here; they are, however, permitted to develop alternate segmentation (10) and assembly strategies, provided that the final sequence is not altered from the specified sequence.

### SwAP-In

SwAP-In and its variations enable a chromosome to be segmentally assembled. A chromosome is divided computationally into “megachunks” (30 to 60 kb long), each comprising a set of “chunks,” typically  $\leq 10$  kb in length. Chunks are in turn bounded by restriction enzyme (RE) recognition sites otherwise absent from the 10-kb segment. Chunks can be assembled into megachunks by restriction enzyme cutting and ligation *in vitro* (Fig. 2A), and the megachunks are subsequently integrated into the host genome, replacing the corresponding wild-type segment. Megachunks are introduced sequentially from left to right by using the endogenous homologous recombina-

tion (HR) machinery and termini that consist of either (i) terminal “UTC” (universal telomere cap) sequences, for the first and last megachunk extremities, or (ii) terminal sequences of  $\sim 500$  bp facilitating integration into a partially synthetic, partially native chromosome (Fig. 2B). The rightmost chunk in each megachunk contains a selectable marker. It is possible to incorporate a series of megachunks into the yeast genome by alternating between just two selectable markers; we chose *URA3* and *LEU2*. As each megachunk is introduced, the previously used marker is overwritten as a consequence of HR with the incoming megachunk (Fig. 2B). Thus, if megachunk  $m$  is tagged with *URA3*, megachunk  $m + 1$  will be tagged with *LEU2*,  $m + 2$  with *URA3*, and so on. Alternatively, chunks as originally designed can be provided as a series of “minichunks” that overlap each other by one building block and are recombined with each other and into the genome simultaneously, as was done with synIII (8) and synX (11), by using the auxotrophic marker switching specified by SwAP-In.

The first and last megachunks of a synthetic chromosome require special treatment; for these, one end is provided by a telomere seed sequence (TeSS) within the larger UTC fragment, and the

other end encodes terminal homology targeting the resident chromosome. The TeSS end is designed to grow a new telomere rather than participate in homologous recombination. The megachunk at the rightmost end of a synthetic chromosome may contain a selectable marker, but it is more convenient to introduce the very last megachunk in a “markerless” format provided that the second-to-last megachunk is integrated using *URA3*. In this case, selection is provided by the expected 5-fluoroorotic acid (FOA) resistance phenotype conferred by the terminal megachunk as it overwrites the resident *URA3* marker from the penultimate megachunk (Fig. 2B).

### Recoding of RE sites for SwAP-In

SwAP-In requires rare-cutting RE sites to be present approximately every 10 kb and, moreover, that the RE sites leave nonpalindromic overhangs to enforce directional assembly of chunks (14) (Fig. 2A). Additionally, two distinct RE sites must flank the selectable marker at the “right end” of each megachunk (Fig. 2C). The RE site to the right of the selectable marker is used to produce the right end of megachunk  $m$ , and the RE site to the left is used to generate the left end of megachunk  $m + 1$ . The required sites either occur

**Table 1. Design challenges and policies adopted.** CDS, gene coding sequence; snoRNA, small nucleolar RNA.

Design challenge or amendment	Policy adopted by design team
Subtelomeric repeats of varying copy number on multiple chromosomes	Delete and monitor for phenotypes as chromosomes are combined. Exception: vitamin biosynthesis genes retain one copy.
Dispersed repeated genes of high copy number, as well as high-copy COS and seripauperin genes	Delete and monitor for phenotypes as chromosomes are combined.
loxPsym sites <300 bp apart when inserted algorithmically (not especially useful and more difficult to synthesize)	loxPsym thinning to eliminate the loxPsym site closer to the centromere.
Stop codon overlaps a second CDS; insertion of loxPsym site would disrupt second CDS; also TAG recoding to TAA could disrupt CDS	Favor preservation of “verified ORFs” over “dubious ORFs” and “uncharacterized ORFs”; always add loxPsym site to a verified ORF in this case
Tandem repeats inside CDSs (34)	Use GeneDesign’s RepeatSmasher module to recode such genes to minimize DNA level repetitiveness, making DNA easier to synthesize and assemble. In synthesis phase, permit 10% length variation for homopolymer
Homopolymer tracts, including frequent A and T tracts, are difficult to synthesize	tracts >10 bp provided they are in a noncoding region.
Introns	Delete pre-mRNA introns precisely, except from genes with evidence of a fitness defect caused by intron deletion (35, 36). The <i>HAC1</i> intron, which uses separate splicing machinery and is known to play a critical role in regulation of the unfolded protein response, was not deleted (9). Delete all tRNA introns precisely. These are individually nonessential and were deleted with their host introns.
Intronically embedded snoRNAs	They could be “refactored” by insertion into the array of snoRNAs on chr II.

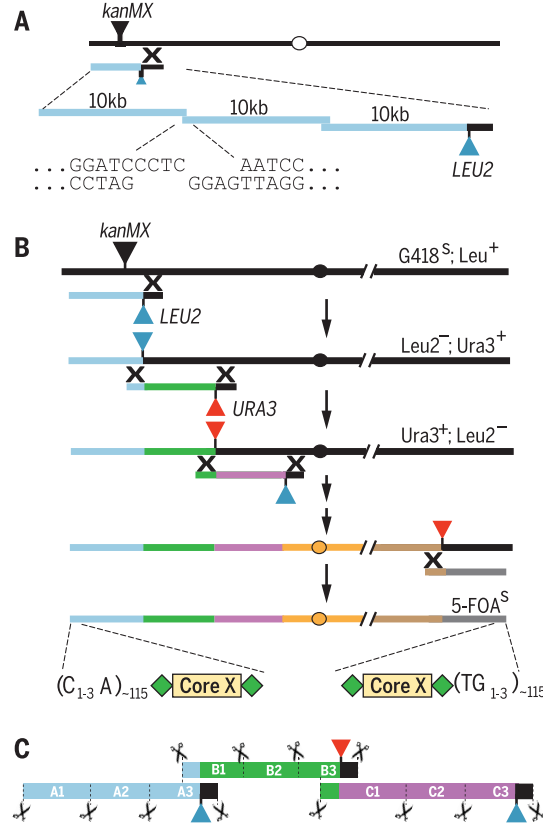
naturally within the synthetic megachunk or are introduced by synonymous recoding of ORFs. Synonymous recoding constrains site placement but reflects our design choice to avoid altering noncoding regions with possible regulatory function, especially promoters. Furthermore, two additional rules are imposed: (i) the terminal homologies defined by the sites flanking the selectable marker must be at least 500 bp long; and (ii) the selectable marker is always added by inserting it into the coding sequence of a nonessential ORF, tempora-

rily knocking out function of that one gene. The ORF sequence is subsequently restored in the next round of megachunk incorporation when the marker is overwritten by SwAP-In (Fig. 2B).

Each chunk is typically assembled from ~750-bp “building blocks” or 2- to 4-kb “minichunks,” and these may be further decomposed into overlapping oligonucleotides. The oligonucleotides are assembled by polymerase chain assembly (15, 16) into building blocks or minichunks, which can be subsequently combined into chunks by various assembly methods,

## Fig. 2. Megachunk assembly.

(A) Megachunks are assembled from three to six chunks by RE (restriction enzyme) cutting and ligation. RE sites have distinct non-palindromic overhangs to assure proper assembly. (B) Sequential SwAP-In HR steps to complete synthetic chromosome (bottom). Colored segments, megachunks; black, native sequence; blue and red triangles, *LEU2* and *URA3*, respectively; X, sites of HR; 5-FOA<sup>S</sup>, 5-fluoroorotic acid (FOA)-sensitive allele. Inserts between dashed lines are UTCs (universal telomere caps). (C) Detail of right end of each megachunk. Each chunk begins and ends with a RE site (scissors) designed to ligate specifically to the adjacent chunk(s). However, the rightmost chunk in each megachunk contains three types of sequence: synthetic (colors), marker (colored triangle), and native (black). The marker is inserted in an endogenous non-essential gene during assembly of each megachunk.



**Table 2. Versions of synV.** These designed versions (12) include those created by computational specialists (CS), yeast specialists (YS), and live strains synthesized (Syn).

Version	Description	Workflow or associated strain(s)
5_0_00	Wild type	
5_1_00	PCRTags added	BioStudio PCR tagger (CS)
5_2_00	Stop codons swapped	BioStudio codon juggler (CS)
5_3_00	loxP sites inserted	BioStudio chromosome splicer (CS)
5_3_01 ~ 5_3_38	Intermediate editing steps	BioStudio graphical user interface (YS)
5_3_39	Segmentation	BioStudio chromosome splitter (CS)
5_3_40 ~ 5_3_43	Intermediate sequence/annotation corrections	BioStudio reports and expert review (CS, YS)
5_9_01 ~ 5_9_21	Draft synV strains with mutations compared to the final design version 5_3_43	yXZX345, yXZX446, yXZX473, yXZX512, yXZX538, yXZX633, yXZX651, yXZX682, yXZX699, yXZX702, yXZX713, yXZX723, yXZX732, yXZX743, yXZX757, yXZX762, yXZX780, yXZX796, yXZX809, yXZX822, yXZX829 (Syn)
5_9_22	Final synV strain, same as designed	yXZX846 (Syn)

in vitro or in yeast (8, 10, 13, 17, 18). The overall hierarchy of different-sized Sc2.0 DNA intermediates is shown in fig. S1.

A computational challenge occurs at the chunk level in placing RE sites at regular 10-kb intervals and at special sites required for SwAP-In, an optimization problem that we term “segmentation” (see the supplementary materials for additional descriptions of the methods and algorithms) (14).

## PCRTags

Given that synthetic chromosomes are assembled iteratively in 30- to 60-kb megachunk steps, we must be able to verify and quantify the synthetic content of the genome. We developed the “PCRTagging” watermark system (5) to satisfy this need by introducing slight nucleotide sequence alterations through synonymous recoding within ORFs to specify pairs of primers specific to either the wild-type or synthetic version of that gene (fig. S2). Ultimately, all Sc2.0 synthetic chromosomes are validated by whole-genome sequencing at least once; “semisynthetic” strains are recommended to be sequenced at major intervals during assembly (e.g., 300 to 500 kb integrated) in order to identify major structural variants that occur at about that frequency and to eliminate them early in assembly (10–13).

## loxPsym sites enabling SCRaMble

The inducible genome rearrangement system “SCRaMble” is based on a chemically inducible Cre recombinase (4, 5). We seeded the genome with the palindromic recombination site loxPsym (19). These sites were inserted 3 bp downstream of the stop codon of every nonessential ORF, and loxPsym sites were also inserted when features (other than introns) were deleted. We applied a thinning algorithm to remove loxPsym sites that were <300 bp apart. The SCRaMble system was designed to permit on-the-fly genome rearrangements leading to a combinatorially diverse population of cells with a corresponding selectable phenotypic diversity. Consistent with design goals, strains generated by SCRaMble with the circular synIXR chromosome led to inversions and deletions at the designed sites, including minimized versions of synIXR, with no changes to the non-synthetic chromosomes (20). A large number of strains also contained duplications, providing additional useful variation to evolve new phenotypes.

## Stop codon recoding/stop swaps

When one is building an organism’s genome from scratch, systematic elimination of codons in favor of synonymous codons is straightforward (3, 21). Similar to “REcoli,” the designed Sc2.0 genome replaces all UAG stop codons with UAA; de novo synthesis rather than the multiplex genome engineering and related methods will be used to produce the living strain. Eukaryotes can survive with a single stop codon, including several naturally occurring ciliates with variant genetic codes (22) in which conventional stop codons instead encode amino acids. Elimination of TAG therefore seems unlikely to compromise yeast fitness a priori.

We developed a general algorithm to change any codon into any other across an entire chromosome. ORFs that overlap one another may not be free to change without an unintended non-synonymous change in one or the other (fig. S3). In such cases, we refer to the annotations of the overlapping ORFs. We permit nonsynonymous changes to be made automatically to dubious genes on behalf of verified genes, but we require that nonsynonymous changes to verified genes on behalf of dubious or verified ORFs (as defined by SGD) be reviewed by team members, for example, by making comparisons to variations found in closely related strains or species. Typically, verified ORFs are not altered to allow a TAG stop codon of a dubious or uncharacterized ORF to be converted to TAA. Genome-wide, there were 15 instances of verified ORFs overlapping other verified ORFs where a TAG codon swap was required.

### Deletion of repeats, tRNAs, and introns

Virtually all sequenced genomes contain transposons; the *S. cerevisiae* genome has five families (and overall, about 50 copies) of retrotransposons called Ty elements that are bounded by long terminal repeat (LTR) sequences; recombination between the two LTRs has led to formation of hundreds of “solo LTRs” in the genome. Bottom-up design of a synthetic yeast genome allows removal of every base pair of retrotransposon and LTR repeats, producing a potentially more stable genome free of mobile elements. We also removed all tRNA genes from their native genomic loci, relocating them to a specialized neochromosome encoding only tRNAs. Our rationale is that tRNAs lead to genomic instability by at least two mechanisms: (i) replication fork collapse, presumably due to collision between RNA polymerases Pol II and Pol III (23), and (ii) insertion of Ty elements for which tRNA genes are preferred sites (24).

A proof-of-principle experiment showed that 17 tRNA genes “refactored” by flanking *S. cerevisiae* tRNA coding sequences with flanking sequences from a related species lacking transposons could be maintained on a centromeric plasmid and that at least one of the refactored tRNA genes could complement a deletion of an essential tRNA (fig. S4) (8).

Pre-tRNA and pre-mRNA introns were deleted precisely in most cases. For additional details on exceptions, see Table 1.

### BioStudio

The first Sc2.0 chromosome arm to be completed, synIXR, was designed manually using specially developed programs followed by visual verification and minor hand-editing in DNA Strider (25). The synthetic product was incorporated into yeast in the form of an episomal circular chromosome (5). For the first Sc2.0 synthetic chromosome that we completed, synIII, we formalized an editing mode in which an experimental biologist (J.D.B.) worked with a computational biologist (S.M.R.) to design a chromosome with the requested sequence alterations using a series of Perl scripts. Scripts were also used to implement the hierarchical assembly strategy, segmenting the designed synIII sequence into megachunks, chunks, mini-

chunks, building blocks, and finally, oligonucleotides. An example of design steps for synV is shown in Table 2 and a supplemental movie (movie S1) (12).

### Annotation and version control

The yeast genome-sequencing project involved dozens of lab groups and still requires a major database employing experts working with the larger community to maintain its annotation (26). As updates are made to the wild-type reference sequence and annotation, the substantial investment in existing infrastructure, such as the SGD database (www.yeastgenome.org), is critical to success.

To enable participation of multiple genome designers within multiple groups, we introduced a genome version control system. Version control software allows incremental “rollbacks” to previous designs when errors or other problems are encountered. It also permits asynchronous, distributed document manipulation by tracking the person responsible for each version and permitting authorized designers to accept or reject proposed changes to the Sc2.0 genome. We introduced version control functions to BioStudio for genome synthesis projects modeled after version control systems like Concurrent Versions System (CVS), Subversion, and Git (27–29) used for software development projects. BioStudio requires genome editors to annotate each change with a time stamp, editor name, and explanation.

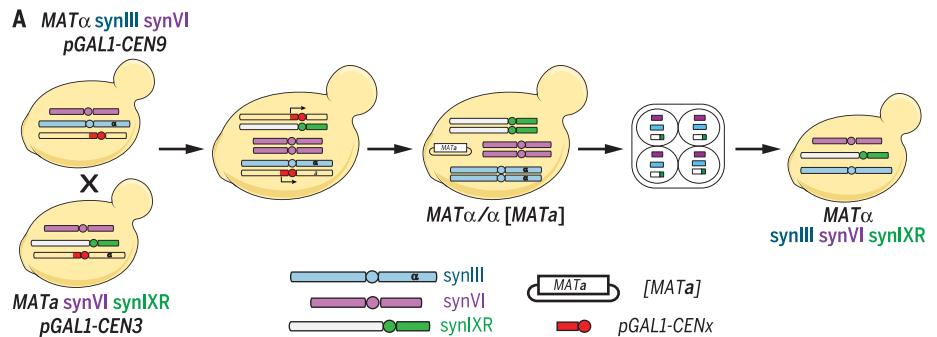
### Visualization and interface

BioStudio can be used on the command line, and it also offers a graphical user interface through the genome annotation viewer GBrowse from Generic Model Organism Database project (30) [now largely

replaced by JBrowse (31)]. GBrowse is highly compatible as a BioStudio graphical user interface because it displays GFF files (generic feature format or gene-finding format) through Web browsers. GBrowse further offers a robust plug-in architecture that lets developers extend five base plug-in types: finders, filters, highlighters, annotators, and dumpers. BioStudio further extends the dumper to appear as a sixth type, an editor, giving users access through GBrowse to the BioStudio algorithms and the underlying DNA sequence.

### Combining Sc2.0 chromosomes by an endoreduplication backcross

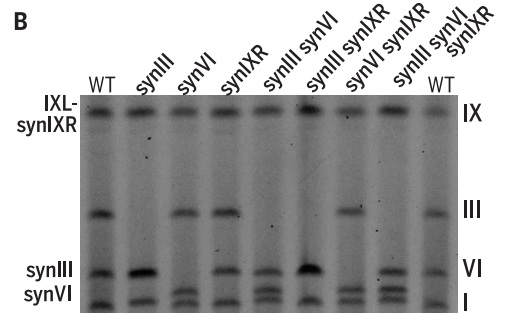
The Sc2.0 project is modularizing genome assembly and constructing each of 16 synthetic chromosomes (synI-synXVI) in discrete strains (5, 8–13). Synthetic chromosomes can be consolidated into a single strain by mating and sporulation. However, multiple meiotic crossovers challenge recovery of nonrecombinant progeny chromosomes. Although PCRTag analysis can track synthetic DNA efficiently (6), as the numbers and lengths of synthetic chromosomes increase, it will become increasingly difficult to find spores containing entirely full-length synthetic chromosomes in the progeny. Here, we establish a conditional chromosome destabilization program to generate Sc2.0 poly-synthetic chromosome strains, called an “endoreduplication intercross.” We simultaneously disrupt centromere function of two specified native chromosomes in a doubly heterozygous diploid synthetic strain (e.g., synIII/III VI/synVI) using the *GALI* promoter in cis (32), generating a “ $2n - 2$ ” strain. In diploids each chromosome can be individually lost, yielding hemizygotes for the destabilized



**Fig. 3. Consolidation of Sc2.0 chromosomes by endoreduplication intercross.**

(A) A pGAL-CENx construct was integrated into the native chromosome in pairs of synthetic chromosome strains, which were then mated. After growth on galactose to induce destabilization of native chromosomes and selection on FOA medium, the  $2n - 2$  state was confirmed by PCRTag analysis. Haploid poly-syn chromosome strains were then generated by sporulation and dissection. An episomal copy of *MATa* was introduced to permit sporulation in synIII-containing strains. Double-syn strains were generated similarly (fig. S6).

(B) Electrophoretic karyotypes of syn strains showing faster migration of synIII and synVI compared with native strains; note that IXL-synIXR and native IX migrate identically.



**Table 3. Summary statistics for design of Sc2.0.** WT, wild type; SYN, synthetic.

	WT size	SYN size	No. of stop codon swaps	No. of loxP sites added	bp of PCRTAG recoded	bp of RE sites recoded	No. of tRNA deleted	bp of tRNA deleted	bp of repeats deleted
chr01	230208	181030	19	62	3535	210	4	372	3987
chr02	813184	770035	93	271	13651	1215	13	993	7030
chr03	316617	272195	44	100	5272	250	10	794	7358
chr04	1531933	1454671	183	479	25398	2298	28	2261	11674
chr05	576874	536024	61	174	8760	813	20	1471	11181
chr06	270148	242745	30	69	4553	369	10	835	9297
chr07	1090940	1028952	126	380	17910	1572	36	2887	13284
chr08	562643	506705	61	186	9980	714	11	878	19019
chr09	439885	405513	54	142	7943	436	10	736	11632
chr10	745751	707459	85	249	12582	1102	24	1853	7523
chr11	666816	659617	68	199	11769	1017	15	1243	4214
chr12	1078177	999406	122	291	15129	1539	19	1646	10843
chr13	924431	883749	100	337	15911	0	21	1691	7673
chr14	784333	753096	96	260	13329	1113	14	1152	5115
chr15	1091291	1048343	147	399	18015	2058	20	1612	9542
chr16	948066	902994	127	334	15493	1374	17	1338	10048
Total	12071297	11352534	1416	3932	199230	16080	272	21762	149420

chromosome; most such  $2n - 1$  strains endoreuplicate the remaining single chromosomes to regenerate a  $2n$  state (33). Conditional chromosome destabilization is also useful to backcross synthetic strains to wild type, called an “endoreplication backcross,” used to debug synV (12).

To take steps toward building Sc2.0, we built double and triple synthetic chromosome strains from individual synVI (9), synIII (8), and synIXR (5) strains (Fig. 3A and fig. S6). Here, synIXR was first converted from a circular chromosome to a linear version attached to native IXL (IXL-synIXR, yLM461; herein referred to as synIXR) (fig. S7). We successfully destabilized pairs of native chromosomes, demonstrating that yeast tolerates the  $2n - 2$  state in these three cases (III/VI, VI/IXR, and III/IXR). All combinations of synthetic chromosomes were capable of directing growth of diploid yeast cells in the absence of the corresponding native chromosomes. Although meiotic proficiency was not selected for during design, diploids homozygous for synIII, synVI, and synIXR readily underwent meiosis and sporulation, producing genotypes consistent with endoreduplication. The meiotic proficiency of heterozygous diploid synIII synVI synIXR/III VI IXR cells suggests that the extensively modified synthetic chromosome structure did not appreciably inhibit homolog pairing, an observation also made for synV/V strains (12). Karyotypic analysis by pulsed-field gel electrophoresis in the haploid strains generated here enabled visualization of expected mobility shifts of synIII and synVI (Fig. 3B). In principle, any pair of Sc2.0 chromosomes may be consolidated into a single strain without sequence alteration using this strategy.

### The Sc2.0 genome

We have completed the design of a synthetic eukaryotic genome; a summary of the changes made

by design is shown in Table 3. Over one-third of the yeast chromosomes have now been synthesized and assembled according to this standard with only minimal problems encountered, testifying to the soundness of design (5, 8–13). The SwAP-In assembly method has made it relatively straightforward to implement a global strategy for writing the remainder of the genome. We have devised an efficient strategy for synthetic chromosome consolidation and shown its successful implementation to build polysynthetic Sc2.0 strains. Further improvements in both the software and DNA synthesis technology, with current synthesis costs for Sc2.0 averaging approximately US\$0.10 per base pair, mean that genome-wide synthesis projects like this one will become routine. At this price, the overall cost for the Sc2.0 DNA, accounting for required overlaps, the synthesis of URA3 and LEU2 markers that are incorporated and then deleted, and errors in synthesis that require resynthesis of segments, is estimated to be approximately US\$1.25 million. The total costs of the project, including labor for assembly, genotyping, sequencing, evaluating fitness and phenotypes, debugging and correcting bugs, developing and maintaining software and servers, and other activities and associated indirect costs will be, of course, considerably higher. The next design frontier could involve living systems that will be less and less similar to native genomes and more like de novo designs.

### REFERENCES AND NOTES

- J. M. Cherry *et al.*, *Nucleic Acids Res.* **26**, 73–79 (1998).
- A. Goffeau *et al.*, *Science* **274**, 546–567 (1996).
- F. J. Isaacs *et al.*, *Science* **333**, 348–353 (2011).
- J. Dymond, J. Boeke, *Bioeng. Bugs* **3**, 168–171 (2012).
- J. S. Dymond *et al.*, *Nature* **477**, 471–476 (2011).
- L. A. Mitchell *et al.*, *J. Vis. Exp.* **99**, e52941 (2015).
- S. M. Richardson, P. W. Nunley, R. M. Yarrington, J. D. Boeke, J. S. Bader, *Nucleic Acids Res.* **38**, 2603–2606 (2010).

- N. Annaluru *et al.*, *Science* **344**, 55–58 (2014).
- L. A. Mitchell *et al.*, *Science* **355**, eaaf4831 (2017).
- Y. Shen *et al.*, *Science* **355**, eaaf4791 (2017).
- X. Li *et al.*, *Science* **355**, eaaf4706 (2017).
- Z.-X. Xie *et al.*, *Science* **355**, eaaf4704 (2017).
- W. Zhang *et al.*, *Science* **355**, eaaf3981 (2017).
- S. M. Richardson *et al.*, *Lect. Notes Comput. Sci.* **6068**, 280–289 (2010).
- R. W. Barnett, H. Erfle, *Nucleic Acids Res.* **18**, 3094 (1990).
- R. B. Ciccarelli, P. Gunyuzlu, J. Huang, C. Scott, F. T. Oakes, *Nucleic Acids Res.* **19**, 6007–6013 (1991).
- H. Muller *et al.*, *Methods Mol. Biol.* **852**, 133–150 (2012).
- S. M. Richardson, S. J. Wheelan, R. M. Yarrington, J. D. Boeke, *Genome Res.* **16**, 550–556 (2006).
- R. H. Hoess, A. Wierzbicki, K. Abremski, *Nucleic Acids Res.* **14**, 2287–2300 (1986).
- Y. Shen *et al.*, *Genome Res.* **26**, 36–49 (2016).
- M. J. Lajoie *et al.*, *Science* **342**, 361–363 (2013).
- H. M. Salim, K. L. Ring, A. R. Cavalcanti, *Protist* **159**, 283–298 (2008).
- A. Admire *et al.*, *Genes Dev.* **20**, 159–173 (2006).
- H. Ji *et al.*, *Cell* **73**, 1007–1018 (1993).
- C. March, *Nucleic Acids Res.* **16**, 1829–1836 (1988).
- D. G. Fisk *et al.*, *Yeast* **23**, 857–865 (2006).
- B. Berliner, in *Proceedings of the USENIX Winter 1990 Technical Conference*, Washington, DC, 22 to 26 January 1990 (USENIX Association, Berkeley, CA, 1990), pp. 341–352.
- C. M. Pilato, B. Collins-Sussman, B. W. Fitzpatrick, *Version Control with Subversion* (O’Reilly Media, Sebastopol, CA, ed. 2, 2008).
- S. Chacon, B. Straub, ProGit (SpringerLink Online service, ed. 2, 2014).
- L. D. Stein *et al.*, *Genome Res.* **12**, 1599–1610 (2002).
- M. E. Skinner, A. V. Uzirov, L. D. Stein, C. J. Mungall, I. H. Holmes, *Genome Res.* **19**, 1630–1638 (2009).
- A. Hill, K. Bloom, *Mol. Cell. Biol.* **7**, 2397–2405 (1987).
- R. J. Reid *et al.*, *Genetics* **180**, 1799–1808 (2008).
- K. J. Verstrepen, A. Jansen, F. Lewitter, G. R. Fink, *Nat. Genet.* **37**, 986–990 (2005).
- J. Parenteau *et al.*, *Cell* **147**, 320–331 (2011).
- J. Parenteau *et al.*, *Mol. Biol. Cell* **19**, 1932–1941 (2008).

### ACKNOWLEDGMENTS

This work was supported in part by funding from NSF (grants MCB-0718846 and MCB-1026068 to J.D.B. and J.S.B., MCB-1616111 to J.D.B., MCB-0546446 and MCB-1445545 to J.S.B.), the Department of Energy (grant number DE-FG02097ER25308 to S.M.R.); NSERC Postdoctoral Fellowship to L.A.M.; and Microsoft Research (to J.S.B.). We thank all Sc2.0 Consortium members and members of the *S. cerevisiae* community for their helpful comments during the design phase of this project. We thank the leadership and staff of the SGD for helping make this project possible. J.D.B. and J.S.B. are founders and directors of Neochromosome, Inc. J.D.B. serves as a scientific advisor to Recombinetics, Inc., and Sample6, Inc. These arrangements are reviewed and managed by the committees on conflict of interest at NYU Langone Medical Center (J.D.B.) and Johns Hopkins University (J.S.B.). Additional information including design diagrams, PCRTAG sequences, feature summary tables, and variants in the physical strains corresponding to Sc2.0 synthetic chromosomes can be accessed on the Sc2.0 website, [www.syntheticyeast.org](http://www.syntheticyeast.org).

### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/355/6329/1040/suppl/DC1](http://www.sciencemag.org/content/355/6329/1040/suppl/DC1)  
Materials and Methods  
Figs. S1 to S7  
References (37–45)  
Movie S1

15 February 2016; accepted 26 January 2017  
10.1126/science.aaf4557



**Design of a synthetic yeast genome**

Sarah M. Richardson, Leslie A. Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S. Dymond, James E. DiCarlo, Dongwon Lee, Cheng Lai Victor Huang, Srinivasan Chandrasegaran, Yizhi Cai, Jeffrey D. Boeke and Joel S. Bader (March 9, 2017)  
*Science* **355** (6329), 1040-1044. [doi: 10.1126/science.aaf4557]

Editor's Summary

---

This copy is for your personal, non-commercial use only.

---

- Article Tools** Visit the online version of this article to access the personalization and article tools:  
<http://science.sciencemag.org/content/355/6329/1040>
- Permissions** Obtain information about reproducing this article:  
<http://www.sciencemag.org/about/permissions.dtl>

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.