NEWS AND VIEWS

Whole genome synthesis: from poliovirus to synthetic yeast

Junbiao Dai[1],*, Yizhi Cai[2], Yinjing Yuan[3], Huanming Yang[4] and Jef D. Boeke[5]

[1] Center for Synthetic Biology Engineering Research, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[2] School of Biological Sciences, The King's Buildings, University of Edinburgh, EH9 3BF, United Kingdom
[3] Key Laboratory of Systems Bioengineering (Ministry of Education), SynBio Research Platform, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
[4] BGI-Shenzhen, Shenzhen 518083, China
[5] Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, New York University Langone Medical Center, New York, NY 10011, USA
* Correspondence: jb.dai@siat.edu.cn

About thirty years ago when the human genome project (HGP) was first proposed, there was widespread disagreement on whether it was worth spending $3 billion to sequence a genome with only 3% coding sequences, which was potentially with enormous errors due to technical limitations at the time [1]. However, it turned out to be successful, generating a return on investment to the US economy estimated at more than $796 billion [2]. In addition, it led to a wave of biotechnology revolutions including ultra high-throughput sequencing techniques, long read sequencing, computational methodologies and others, which were not foreseen at the inception of the project. Ultimately it led to a better understanding of the fundamentals of life sciences and promises to provide enormous benefit to human health. Synthesis of a complete genome has encountered similar criticism, especially in the face of emerging techniques that could potentially be used to editing whole genomes [3–6]. We envision the impact of writing genomes will be even larger than that of reading genomes. Also, the societal implications will also be more complicated and need to be handled with greater care [7].

Why synthesize a complete genome? It has been a long interesting question whether a living cell (and not only its genome) could be made from scratch in the lab, a goal which may not be realized anytime soon [8]. Therefore, making an organism's genome, which contains all information a living organism requires, is a good alternative. Second, despite the fact that modifying one genome is facilitated by recent advances in genome-editing technologies, it still takes years to make a small but pervasive change throughout the genome such as eliminating one codon from *E. coli* [9–14]. Third, there are lots of things that may not be feasible, if not impossible, to test without a designer genome bearing these pervasive changes. For example, how else can one create a multitude of randomly rearranged genomes with changed position and copy number of large numbers of genes? What are the functions of the repetitive sequences, which occupy over 50% of human genome and even larger fractions of some plant and other animal genomes?

Efforts to synthesize a whole genome dated back to the time when the HGP was near completion. As early as 2002, by synthesizing a full-length polivirus cDNA, Eckhardt Wimmer and his colleagues were able to, for the first time, assemble a poliovirus genome starting from oligonucleotides [15]. In 2003, a paper from Hamilton O. Smith and Craig Venter reported the synthesis of the 5.6 kb Φx174 bateriophage genome [16], the first organism to get its genome sequenced by Fred Sanger in 1978 [17]. It took two weeks from the design of chemical sequences to the assembly of a living virus. Five years later, the *Mycoplasma genitalia* genome, which is 100 times bigger than that of Φx174, was made from scratch [18]. Unfortunately, the genome failed to boost a living cell due to mutations. It was not until 2010, after synthesis of

the somewhat larger *Mycoplasma mycoides* genome, the group was able to construct the first self-replicating cell powered by a man-made genome [19]. This feat depended critically on the Venter group's critical earlier demonstration that entire genome size DNA molecules could be transferred from *M. mycoides* to a closely related species, *M. capricolum*, a process referred to as genome transplantation [20]. Based on the synthetic *M. mycoides*, the same group recently produced a bacterial cell with a minimal genome by reducing the size over 50% [21]. However, the aforementioned genomes of goat and cattle pathogens are all relatively small, the sequences constructed were essentially wild-type (i.e., not extensively engineered) and of course the genomes in question are from prokaryotic cells. Has the time arrived to tackle much larger and more complex eukaryotic genomes?

In the past decade, high-throughput sequencing technology has completely revamped our knowledge about genomic information. Nowadays, we are able to read about 15 petabases per year [22]. The trove of sequencing data allows us to identify new functional proteins and pathways never previously known. In addition, it provides a chance to compare the genomes from related species and guide more rational genome redesign. In 2006 when we first designed the synthetic yeast project, the price for every base pair was around $1 and typical projects were 1 kb in length. Today, the price has dropped substantially (about 20-fold) to obtain DNA fragments of multiple kilobases, enabling more affordable genome building, although price is still a major obstacle. New technologies to synthesize DNA, facilitated by methods exploiting synthesis of oligonucleotides made on microarrays and produced as pools [23,24] and other methods [25] are beginning to come on line. If the price of gene synthesis continues to drop as expected, it is possible to synthesize a 10-million-bp genome for less than 1 million USD in five years. Finally, different technologies to assemble large size DNA fragments from small ones have been invented, enabling "scarless" assembly of designer DNA such as the golden-gate cloning method and *in vitro* as well as *in vivo* recombination-based techniques [26–31]. All of these provide the knowledge and tools to think about designing and synthesizing larger genomes such as the *S. cerevisiae* genome (Sc2.0, Figure 1).

The aim of the Sc2.0 project is to build a designer genome which will allow us not only to test our ability to construct a yeast with multiple synthetic chromosomes, but also to answer a series of important biological questions. Therefore, besides testing our ability to synthesize a ~12-million-base pair yeast genome, we also incorporated a lot of engineered features into the synthetic genome [32], such as i) Genome reduction by systematically removing all retrotransposons and sub-telomeric repeats. In addition, most introns were removed, and the tRNA genes were relocated to a specialize chromosome. ii) Genome recoding to incorporate trackable watermarks (PCRtags) without changing the amino acid sequences. In addition, the least abundant codon, namely the TAG stop codon was replaced by TAA, allowing future genetic code manipulation. iii) Genome expansion by addition of symmetrical loxP sites, leading to the development of a "SCRaMbLE" system to facilitate future genome manipulation [33]. These design features are not "fixed" and new ones could be included in the future. Recently, two papers, one on the design and synthesis of a chromosome arm and the other about the first completely synthetic chromosome were published [32,34]. Excitingly, seven papers which covering five synthetic chromosomes were published in recent issue of Science, setting another milestone in the journey of eukaryotic genome synthesis [35–41]. Work on the synthesis of remaining yeast chromosomes is ongoing around the world.

What can the synthetic yeast be used for? The initial designs of the project allow us not only to generate an organism with a completely synthetic genome, but also to test several important biological questions such as the function of pervasive retrotransposons, repetitive sequences, and introns. At the end, we hope the synthetic yeast can also be used as a platform in biotechnology industry. One potential example is that with the incorporation of the SCRaMbLE system, the synthetic yeast could be used to produce new strains with improved fermentation efficiency and ethanol tolerance, therefore, better for the wine industry or biofuel production. In addition, the strains can also be served as a cell-factory to manufacture valuable products such as fine chemicals, drugs, antibiotics or vaccines by shuffling the exogenous pathways into the synthetic genome. More applications could be developed in the future with the Sc2.0 "open source" project.

Besides science, the Sc2.0 project has become a great platform for international collaboration (Figure 1). Since the first synthetic yeast genome meeting organized by BGI and Tsinghua University in 2012, which sparked the partnership, the Sc2.0 consortium now has over ten research groups across four continents and also industrial participants. In addition, the BAG (build-a-genome) course [42], which was designed as the education part of Sc2.0, has become a real success. Since it was first taught at Johns Hopkins University in 2011, three more US universities/colleges are now piloting or offering the course. Outside the US, Tianjin University has become the first institute offering the course to both postgraduates and undergraduates, and has finished making two chromosomes in the past two years. The BAG course served as a great teaching vehicle to educate the young
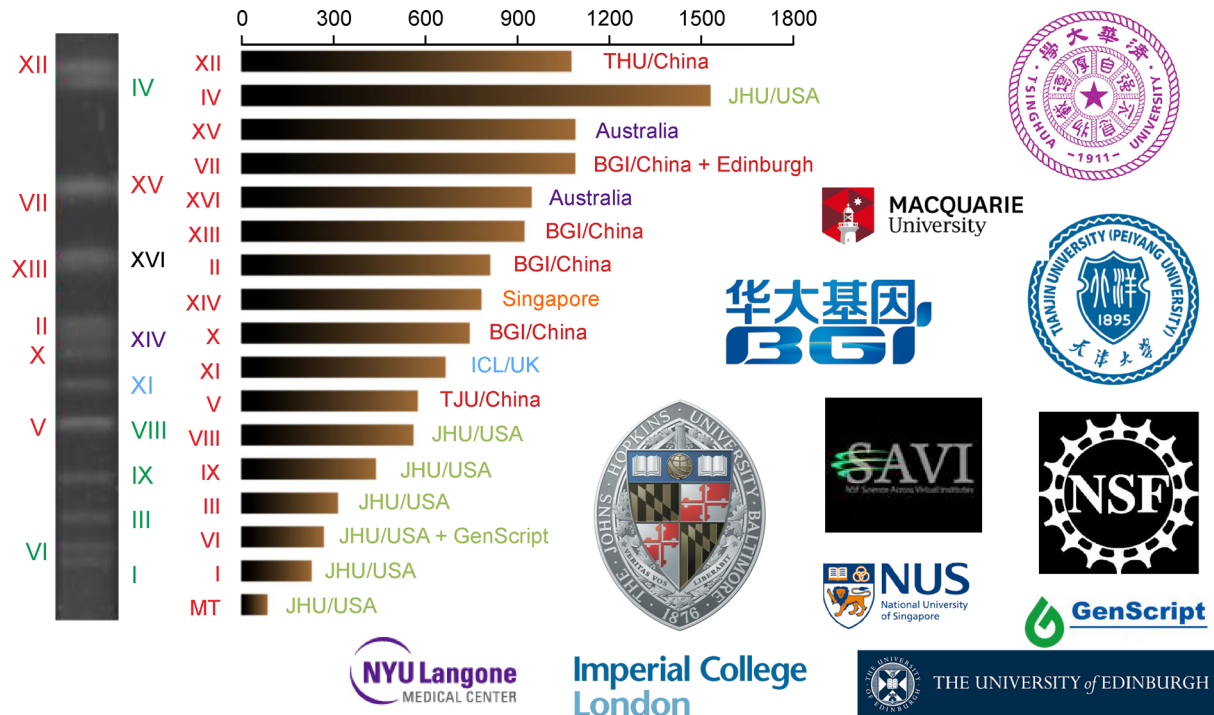
Figure 1. Sc2.0: Design and construct an entirely synthetic version of *S. cerevisiae* genome together.

students, giving them chance to experience real scientific research and synthetic biology. Up to now, through collaboration, the consortium has *ChrII*, *ChrV*, *ChrVI*, *ChrX* and *ChrXII* completely synthesized [36, 38–41] while several other chromosomes are in progress or near completion. In addition, several tools/technologies have been developed through the execution of Sc2.0 [26,43,44]. Finally, it is worth mentioning that in this international collaboration project, the three teams in China will in total contribute five chromosomes, which cover over 40% of the yeast genome, much more than China's contribution to the human genome project (1%).

One concern regarding Sc2.0 and many other engineered organisms is the worry that the artificial organisms could contaminate the environment, endanger the natural species or somehow be used nefariously. It may be hard to turn the synthetic yeast into a bioweapon but it does have the theoretical potential, however slim, to become a threat to the natural species, if beneficial mutations keep accumulating. While no plan exists to release Sc2.0 into natural environments, and ongoing studies suggest that the designer changes tend to make it less fit rather than more so, there is a general need for safeguarding mechanisms for engineered microbes, some of which are intended to be deployed into the natural environment. To address this potential problem, a "genome safeguard" system has been designed and will eventually be incorporated into the synthetic yeast. In this system, two

orthogonal classes of switches were constructed, allowing the synthetic strains to survive only in the presence of small molecules absent from the native environment [45]. Recently, CRISPR-Cas9 gene drive systems were designed to prevent unintended genome editing occurring through the escape of lab strains [46]. Studies in bacterial systems have developed genome safeguards based on the TAG codon-engineered strains (*REcoli*) and orthogonal tRNA synthetase systems [12,13], and on riboregulation of essential genes [47].

The consortium of investigators leading the Sc2.0 project also came together around a "Statement of Ethics and Governance" addressing bioethical and biosafety aspects of the project. This statement, which is posted on the project's web site syntheticyeast.org and the process by which it was collectively drafted, is described in a recently submitted publication [7].

Given the increasing capacity to synthesize larger DNA fragments and deceasing price on gene synthesis, we are optimistic on the completion of the DNA synthesis phase of project Sc2.0 by 2017. It will become the first eukaryotic system with a genome size over 10 million bases completely synthesized in human history. However, we still face possible challenges since despite that the yeast strain bearing a single synthetic chromosome seems to grow normally, how the final strain containing all 16 synthetic chromosomes will behave remains to be determined. It is quite possible that the creeping effect

of accumulation of many tiny fitness defects distributed throughout the chromosomes will conspire to thwart final assembly into a single strain.

In the mean time, it makes sense to think about what could be the next genome in the pipeline after Sc2.0. There are several microbial candidates such as the bacterium *E. coli*, the model prokaryote used widely in metabolic engineering and synthetic biology. Actually, recent work from George Church's lab demonstrated that seven codons could be eliminated from the *E. coli* genome after design and reconstruction [14]. On the eukaryotic side, the worm *C. elegans*, with a relatively small genome (97 million bases) would allow us to test whether we have sufficient knowledge and technology to design a synthetic genome capable of directing normal differentiation. Excitingly and also controversially, the Human Genome Project-Write (HGP-Write) was proposed recently, promoting not only the ability to design and synthesize an ultra large genome, but also the ethical framework [48]. However, many technical challenges remain since none of these organisms have the high intrinsic efficiency of mitotic recombination as that offered by budding yeast, and therefore, it is not practical to perform the step-wise genome replacement methods used in Sc2.0. Recent advances on CRISPR/Cas9-mediated genome editing could potentially provide a good solution [3, 4]. Alternatively, *S. cerevisiae* assembly followed by genome/chromosome transplantation could potentially be used.

ACKNOWLEDGEMENTS

REFERENCES

1. DeLisi, C. (2008) Meetings that changed the world: Santa Fe 1986: Human genome baby-steps. Nature, 455, 876–877

2. Battelle Technology Partnership Practice, (2011) Economic impact of the Human Genome Project

3. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E. and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. Science, 339, 823–826

4. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. Science, 339, 819–823

5. Hsu, P. D., Lander, E. S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157, 1262–1278

6. Sander, J. D. and Joung, J. K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotechnol., 32, 347–355

7. Sliva, A., Yang, H., Boeke, J. D. and Mathews, D. J. (2015) Freedom and responsibility in synthetic genomics: the synthetic yeast project. Genetics, 200, 1021–1028

8. Service, R. F. (2013) The life force. Science, 342, 1032–1034

9. Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D. and Church, G. M. (2012) Genome-scale promoter engineering by coselection MAGE. Nat. Methods, 9, 591–593

10. Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., *et al.* (2011) Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. Science, 333, 348–353

11. Lajoie, M. J., Rovner, A. J., Goodman, D. B., Aerni, H.-R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., *et al.* (2013) Genomically recoded organisms expand biological functions. Science, 342, 357–360

12. Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., Gregg, C. J., Stoddard, B. L. and Church, G. M. (2015) Biocontainment of genetically modified organisms by synthetic protein design. Nature, 518, 55–60

13. Rovner, A. J., Haimovich, A. D., Katz, S. R., Li, Z., Grome, M. W., Gassaway, B. M., Amiram, M., Patel, J. R., Gallagher, R. R., Rinehart, J., *et al.* (2015) Recoded organisms engineered to depend on synthetic amino acids. Nature, 518, 89–93

14. Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M. G., Moosburner, M., *et al.* (2016) Design, synthesis, and testing toward a 57-codon genome. Science, 353, 819–822

15. Cello, J., Paul, A. V. and Wimmer, E. (2002) Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. Science, 297, 1016–1018

16. Smith, H. O., Hutchison, C. A. 3rd, Pfannkoch, C. and Venter, J. C. (2003) Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. Proc. Natl. Acad. Sci. USA, 100, 15440–15445

17. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchison, C. A. III, Slocombe, P. M. and Smith, M. (1978) The nucleotide sequence of bacteriophage phiX174. J. Mol. Biol., 125, 225–246

18. Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., *et al.* (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. Science, 319, 1215–1220

19. Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., *et al.* (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 329, 52–56

20. Lartigue, C., Vashee, S., Algire, M. A., Chuang, R.-Y., Benders, G. A., Ma, L., Noskov, V. N., Denisova, E. A., Gibson, D. G., Assad-Garcia, N., *et al.* (2009) Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science, 325, 1693–1696

21. Hutchison, C. A. III, Chuang, R.-Y., Noskov, V. N., Assad-Garcia, N., Deerinck, T. J., Ellisman, M. H., Gill, J., Kannan, K., Karas, B. J., Ma, L., *et al.* (2016) Design and synthesis of a minimal bacterial genome. Science, 351, aad6253

22. Schatz, M. C. and Phillippy, A. M. (2012) The rise of a digital immune system. Gigascience, 1, 4

23. Kosuri, S., Eroshenko, N., LeProust, E. M., Super, M., Way, J., Li, J. B. and Church, G. M. (2010) Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. Nat.

Biotechnol., 28, 1295–1299

24. Matzas, M., Stähler, P. F., Kefer, N., Siebelt, N., Boisguérin, V., Leonard, J. T., Keller, A., Stähler, C. F., Häberle, P., Gharizadeh, B., *et al.* (2010) High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing. Nat. Biotechnol., 28, 1291–1294

25. Kim, H., Kim, J., Kim, E.-G., Heinz, A. J., Kwon, S. and Chun, H. (2010) Optofluidic *in situ* maskless lithography of charge selective nanoporous hydrogel for DNA preconcentration. Biomicrofluidics, 4, 043014

26. Lin, Q., Jia, B., Mitchell, L. A., Luo, J., Yang, K., Zeller, K. I., Zhang, W., Xu, Z., Stracquadanio, G., Bader , J., *et al.* (2014) RADOM, an efficient *in vivo* method for assembling designed DNA fragments up to 10 kb long in *Saccharomyces cerevisiae*. ACS Synth. Biol. 4, 213–220

27. Gibson, D. G., Smith, H. O., Hutchison, C. A. Ⅲ, Venter, J. C. and Merryman, C. (2010) Chemical synthesis of the mouse mitochondrial genome. Nat. Methods, 7, 901–903

28. Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. PLoS One, 3, e3647

29. Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS One, 4, e5553

30. Shao, Z., Zhao, H. and Zhao, H. (2009) DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways. Nucleic Acids Res., 37, e16

31. Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Wang, L., Shen, Y., Luo, Y., Zheng, Y., *et al.* (2015) YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. Nucleic Acids Res., 43, e88

32. Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., *et al.* (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. Nature, 477, 471–476

33. Shen, Y., Stracquadanio, G., Wang, Y., Yang, K., Mitchell, L. A., Xue, Y., Cai, Y., Chen, T., Dymond, J. S., Kang, K., *et al.* (2016) SCRaMbLE generates designed combinatorial stochastic diversity in synthetic chromosomes. Genome Res., 26, 36–49

34. Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., *et al.* (2014) Total synthesis of a functional designer eukaryotic chromosome. Science, 344, 55–58

35. Mercy, G., Mozziconacci, J., Scolari, V. F., Yang, K., Zhao, G., Thierry, A., Luo, Y., Mitchell, L. A., Shen, M., Shen, Y., *et al.* (2017) 3D organization of synthetic and scrambled chromosomes. Science, 355, eaaf4597

36. Mitchell, L. A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X.,

Yang, K., Richardson, S., Martin, J. A., Zhao, Y., Walker, R., *et al.* (2017) Synthesis, debugging and consolidation of synthetic chromosomes in yeast: synVI and beyond. Science, 355, eaaf4831

37. Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., Lee, D., Huang, C. L. V., Chandrasegaran, S., Cai, Y., *et al.* (2017) Design of a synthetic yeast genome. Science, 355, eaaf4557

38. Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., *et al.* (2017) Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. Science, 355, eaaf4791

39. Wu, Y., Li, B. Z., Zhao, M., Mitchell, L. A., Xie, Z. X., Lin, Q. H., Wang, X., Xiao, W. H., Wang, Y., Zhou, X., *et al.* (2017) Bug mapping and fitness testing of chemically synthesized chromosome X. Science, 355, eaaf4706

40. Xie, Z. X., Li, B. Z., Mitchell, L. A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B. X., Liu, H. M., *et al.* (2017) "Perfect" designer chromosome V and behavior of a ring derivative. Science, 355, eaaf 4704

41. Zhang, W., Zhao, G., Luo, Z., Lin, Y., Wang, L., Guo, Y., Wang, A., Jiang, S., Jiang, Q., Gong, J., *et al.* (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. Science, 355, eaaf 3981

42. Dymond, J. S., Scheifele, L. Z., Richardson, S., Lee, P., Chandrasegaran, S., Bader, J. S. and Boeke, J. D. (2009) Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary build-a-genome course. Genetics, 181, 13–21

43. Mitchell, L. A., Cai, Y., Taylor, M., Noronha, A. M., Chuang, J., Dai, L. and Boeke, J. D. (2013) Multichange isothermal mutagenesis: a new strategy for multiple site-directed mutations in plasmid DNA. ACS Synth. Biol., 2, 473–477

44. Mitchell, L. A. and Boeke, J. D. (2014) Circular permutation of a synthetic eukaryotic chromosome with the telomerator. Proc. Natl. Acad. Sci. USA, 111, 17003–17010

45. Cai, Y., Agmon, N., Choi, W. J., Ubide, A., Stracquadanio, G., Caravelli, K., Hao, H., Bader, J. S. and Boeke, J. D. (2015) Intrinsic biocontainment: multiplex genome safeguards combine transcriptional and recombinational control of essential yeast genes. Proc. Natl. Acad. Sci. USA, 112, 1803–1808

46. DiCarlo, J. E., Chavez, A., Dietz, S. L., Esvelt, K. M. and Church, G. M. (2015) Safeguarding CRISPR-Cas9 gene drives in yeast. Nat. Biotechnol., 33, 1250–1255

47. Gallagher, R. R., Patel, J. R., Interiano, A. L., Rovner, A. J. and Isaacs, F. J. (2015) Multilayered genetic safeguards limit growth of microorganisms to defined environments. Nucleic Acids Res., 43, 1945–1954

48. Boeke, J. D., Church, G., Hessel, A., Kelley, N. J., Arkin, A., Cai, Y., Carlson, R., Chakravarti, A., Cornish, V. W., Holt, L., *et al.* (2016) The Genome Project-Write. Science, 353, 126–127