Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast

Dejana Jovicevic[1)2)], Benjamin A. Blount[1)2)] and Tom Ellis[1)2)]*

A team of US researchers recently reported the design, assembly and in vivo functionality of a synthetic chromosome III (SynIII) for the yeast *Saccharomyces cerevisiae*. The synthetic chromosome was assembled bottom-up from DNA oligomers by teams of students working over several years with researchers as the first part of an international synthetic yeast genome project. Embedded into the sequence of the synthetic chromosome are multiple design changes that include a novel in-built recombination scheme that can be induced to catalyse intra-chromosomal rearrangements in a variety of different conditions. This system, along with the other synthetic sequence changes, is intended to aid researchers develop a deeper understanding of how genomes function and find new ways to exploit yeast in future biotechnologies. The landmark of the first synthesised designer eukaryote chromosome, and the power of its massively parallel recombination system, provide new perspectives on the future of synthetic biology and genome research.

Keywords:
DNA assembly; genome engineering; *Saccharomyces cerevisiae*; synthetic biology; synthetic genomics; yeast

DOI 10.1002/bies.201400086

[1)] Centre for Synthetic Biology and Innovation, Imperial College London, London, UK
[2)] Department of Bioengineering, Imperial College London, London, UK

***Corresponding author:**
Tom Ellis
E-mail: t.ellis@imperial.ac.uk

Abbreviation:
SCRaMbLE, synthetic chromosome rearrangement and modification by loxP-mediated evolution.

Introduction

Scientists have been manipulating DNA for over half a century in a bid to explore and exploit the nature, structure, function and interactions of all domains of life. In the past decade, a step-change in the scale of gene manipulation has occurred as part of the new discipline of synthetic biology, which applies engineering principles to the way bioscience research and technologies are developed [1, 2]. The most ambitious form of synthetic biology deals not just with a handful of genes brought together in novel arrangements, but instead deals with the hundreds and thousands of genes that make up a chromosome or the whole genome of a cell. This field, more commonly referred to as synthetic genomics, applies gene synthesis, DNA assembly and genome engineering tools to put together synthesised chromosomes or drastically modify the DNA content throughout the existing chromosomes to give new functions [3].

The team from the J. Craig Venter Institute was the first to report the synthesis and assembly of a complete bacterial genome, constructing the 582 kb *Mycoplasma genitalium* chromosome, in 2008 [4]. They followed this up in 2010 with a 1 Mb synthetic chromosome encoding the *Mycoplasma mycoides* genome, which was shown to be functional when transplanted into a host cell whose own genomic DNA had been ablated. This yielded the world's first synthetic cell, a *Mycoplasma* bacterium programmed by a synthesised genome [5]. This landmark research demonstrated that synthetic DNA could be assembled and be functional up to the genome scale; but the DNA and genetic content encoded by the synthetic genome was little changed. The genome sequence was intentionally kept almost identical to the natural genome sequence, with only a few silent watermark motifs incorporated to distinguish the synthetic genome from the natural DNA. The next major challenge for synthetic genomes, therefore, would be to build upon this achievement but to do so with *redesign*.

In 2013, a successfully redesigned genome arrived, when Farren Isaacs, George Church, and colleagues reported completion of a long-term project to systematically edit the

4.6 Mb chromosome of *Escherichia coli* to remove all instances of the TAG stop codon and replace these with the TAA stop codon, freeing-up the TAG codon to be repurposed for intentional incorporation of non-natural amino acids in desired positions of any *E. coli* expressed proteins [6]. Rather than synthesising and assembling a synthetic *E. coli* chromosome containing the DNA sequence changes required, the approach taken here was to use genome engineering tools [7, 8] to edit the natural chromosome at hundreds of positions in parallel. In contrast to the previous projects of the J. Craig Venter Institute, the new *E. coli* chromosome was not synthetically assembled, but *did* endow the bacterial cell with an impressive new ability of great value to biotechnology and bioscience research: the ability to encode amino acids beyond the canonical 20 at desired positions within proteins.

Now in 2014, a third example of a synthetic chromosome has been completed and shown to be functional in a cell. A synthetic version of the third chromosome of the yeast *Saccharomyces cerevisiae* (SynIII) was recently described [9] that incorporates both the synthesis and assembly approaches of the fore-runner projects from the J. Craig Venter Institute with the redesign and novel functionalisation concepts of the recoded *E. coli* project. The resultant synthetic chromosome is the impressive first step in the largest synthetic biology project currently underway globally: an international consortium working to construct an entirely synthetic eukaryotic genome.

Synthetic yeast

The eukaryotic organism *S. cerevisiae*, better known as Baker's or Brewer's yeast, has long been exploited by humankind, and is arguably our oldest biotechnology. It is known for being straightforward to culture, safe to work with and in the past three decades research has revealed that it has a relatively compact and stable genome, lacking much of the complexity of genomes seen in multicellular organisms. This yeast is a unicellular organism and one of the simplest examples of a eukaryote, the kingdom that incorporates animals, fungi and plants. It is widely used, both commercially and in research, where it is especially relevant as a model organism for how eukaryotic genomes operate and evolve. The 16 chromosomes that make up its genome were some of the first large scale DNA sequences to be revealed [10], and the *S. cerevisiae* genome was the first completed eukaryotic genome sequencing project, finished in 1996 [11]. Given our in-depth understanding of this organism and its genome, it is an ideal candidate to extend synthetic genomics beyond bacteria.

The international synthetic yeast genome project aims to synthesise and assemble the DNA of all 16 yeast chromosomes, and to incorporate into this sequence a significant number of design changes intended to test biological hypotheses and provide new ways to exploit yeast as a valuable technology. The project was initiated as an idea cooked up by Professors Jef Boeke and Srinivasan Chandrasegaran over a coffee break at John Hopkins University; and as it began to take shape, they sought help from computer science expert and colleague Dr Joel Bader to guide the design tools and data management such a large project would require. By 2007, the idea had metamorphosed into an active project to make a *S. cerevisiae* version 2.0; giving the project its now widely used nickname of Sc2.0.

Before any synthesis and assembly could be done, the design changes for a synthetic genome needed to be agreed upon. Based on existing knowledge of yeast and eukaryotic genetics and how DNA sequence encodes function, decisions were made as to what could be removed, relocated and added to the synthetic genome. The genome redesign has to strike the balance between making enough changes for the new yeast to be easily exploited and yet being conservative enough that the resultant yeast still grows and behaves in the lab like the strains of *S. cerevisiae* from which it is derived. The redesigned chromosomes of Sc2.0 may differ from their natural counterparts by only a few percent of the DNA bases, but across the scale of the genome this represents a large number of modifications and a significant increase compared to the minor changes made in the previously described bacterial synthetic genomics projects (see Box 1).

Building and verifying a synthetic chromosome

The full genome of *S. cerevisiae* contains 11 million base pairs of DNA, making up 16 chromosomes that vary from a few hundred thousand bases in length to over a million base pairs. As with the historic yeast genome sequencing project that ran two decades before, the first chromosome to be synthesised was one of the shortest, chromosome III. It was originally

Box 1

Synthetic yeast design changes

The design modifications the Sc2.0 project intends to write into the *S. cerevisiae* yeast genome are (i) the deletion of all known genome de-stabilising elements such as transposons and sub-telomeric repeat regions; (ii) the insertion of symmetrical loxP (loxPsym) recombination sites immediately downstream of all non-essential genes to allow the genome-content to be 'SCRaMbLEd' (synthetic chromosome rearrangement and modification by loxP-mediated evolution) by inducing expression of Cre recombinase (see further details below); (iii) the conversion of all TAG stop codons to TAA (as described for *E. coli* above); (iv) the watermarking of all protein coding sequences by synonymous base changes, so that they can be identified as synthetic genes by polymerase chain reaction (PCR) amplification; (v) the removal of all tRNA genes from the 16 chromosomes, to be translocated to a dedicated 'neochromosome'; and (vi) removal of the majority of the ~250 introns that exist in yeast genes. Discussion of all these design changes, how they are done and why they are being attempted, can be found in the earliest reported success of the project, the partial synthesis and assembly of two synthetic yeast chromosome arms, published in 2011 [12].

317 kb in size, but the new computer-aided SynIII design resulted in a more compact 273 kb chromosome: smaller than the synthetic chromosomes previously described by Gibson et al., but still a major undertaking. So how to build this? With the decreasing costs of custom DNA synthesis [13] one way would be to outsource synthesis of 1–10 kb DNA Chunks to a company, as previously done by the J. Craig Venter Institute [4]. These could then be stitched together using a variety of different DNA assembly protocols [14]. However, for SynIII the team, then based at Johns Hopkins University, took a more hands-on, bottom-up approach. They founded an undergraduate class called Build-a-Genome (BAG), where students were supervised whilst working to gradually assemble all the 60 mer DNA oligonucleotides that encode an entire chromosome. Students initially used polymerase cycling assembly [15] to stitch together 10–20 of their oligos into ~750 bp building blocks and then worked in teams to ligate these together into 'Minichunks' of DNA, approximately 2–3 kb in size, using DNA assembly methods such as Gibson assembly [16] and USER assembly [17]. The student-made Minichunks are sequence-verified and then either ligated with each other to form larger ~8 to 10 kb 'Chunks' or transformed directly into yeast, as they are designed to have significant areas of homology with one another and with the existing genome, hence allowing recombination when transformed into the cell (Fig. 1). Unlike the previous all-in-one synthetic *Mycoplasma* genome projects, in the case of Sc2.0 the synthetic genome is added to yeast chunk-by-chunk, creating a series of hybrid part-natural-part-synthetic chromosomes along the way to completion, each of which is then checked for sequence fidelity and growth in multiple conditions. The Minichunks and Chunks made by the BAG class were integrated into chromosome III using yeast's native homologous recombination machinery, which efficiently assembles together DNA ends that share at least 40 bp of equivalent sequence. Successful integration of synthetic DNA is then verified by using PCR amplification of yeast genomic DNA with oligos designed to bind only to the synthetic watermarks encoded into each protein coding sequence. If amplification of genomic DNA occurs with the oligos matching synthetic DNA, but not with corresponding oligos that match the natural DNA, then synthetic DNA incorporation is successful. These verification steps following integration are the most time-consuming steps of the process and have to be done every round before the next integration can begin.

Having already built parts of chromosomes VI (synVIL) and IX (synIX) using different DNA assembly approaches [12], the team were able to apply the techniques used to verify their previous semi-synthetic chromosomes to the completed synthetic chromosome III when the years of student-powered DNA assembly and integration were finished. The meticulous verification of the chromosome, detailed in their recent publication [9] proves that yeast with SynIII no longer contains the natural chromosome III or any of its natural genes, only the watermarked synthetic versions. Whole-genome sequencing also shows that the chromosome DNA sequence is almost exactly as designed, and a variety of assays in the lab demonstrates that this new yeast with one chromosome entirely replaced by its synthetic counterpart can grow in a variety of different conditions with equal efficiency as the standard lab yeast from which it was redesigned. The fact that so many design changes can be applied to the DNA sequence of the chromosome with such minimal impact on cell fitness makes this breakthrough truly impressive: the radical recoding inherent to the design of the synthetic yeast genome project works, even at the full-chromosome scale. This paves the way for future synthetic genomics projects that seek to rearrange and rewrite other chromosomes and genomes.

The Sc2.0 SynIII chromosome took several years and nearly a hundred students and researchers to complete. For the same team to realise their dream of a completely synthetic genome, they realised they would need more help. Following the historic model of the international collaboration to sequence the yeast genome in the 1990s, Sc2.0 went global in 2011, with groups from around the world joining in the effort to synthesise, assemble and verify all the remaining chromosomes for a synthetic yeast. The International Synthetic Yeast Genome Project now includes groups from countries as diverse as China, Singapore, Australia, the United Kingdom and across the USA, and each team is being allocated work on individual chromosomes with the aim of bringing these together and finishing the genome project before the end of the decade. By then, it is expected that the cost of synthetic DNA will have decreased by orders of magnitude, especially if innovations in constructing synthetic DNA directly from printed oligo arrays continue. Considerably, lower costs will make it likely that chromosome synthesis and genome re-design will be within the reach of the average research group by 2020. The Sc2.0 project therefore provides a framework to develop the methods and technologies for work that will become commonplace in the future. Innovations on parallelising the step-by-step integration (e.g. by exploiting meiotic recombination between multiple versions of semi-synthetic chromosomes) and using new technologies to accelerating verification steps (e.g. next generation sequencing) will certainly aid in making future projects not only cheaper, but also faster.

Synthetic chromosomes that can be SCRaMbLEd to order

The most intriguing and potentially valuable design change added to the synthetic genome is the SCRaMbLE system, which uses an inducible Cre recombinase to shuffle-up regions of the genome where loxPsym sites have been introduced. The Cre-Lox recombination system naturally originates from the prokaryotic bacteriophage P1 as a mechanism to infect its bacterial host with its own DNA [18]. Shortly after its discovery it was swiftly repurposed as a new site-specific recombination tool, allowing genes to be integrated into the chromosomes of many organisms including eukaryotes like yeast [19]. Even now, Cre-Lox recombinases are still the primary tool for inserting gene expression cassettes into the genome of transgenic mice [20].

The chromosomes of the Sc2.0 genome, including SynIII, incorporate symmetrical 34 bp loxP sites (loxPsym) downstream of all non-essential genes. This size is below the threshold for yeast to perform its own homologous

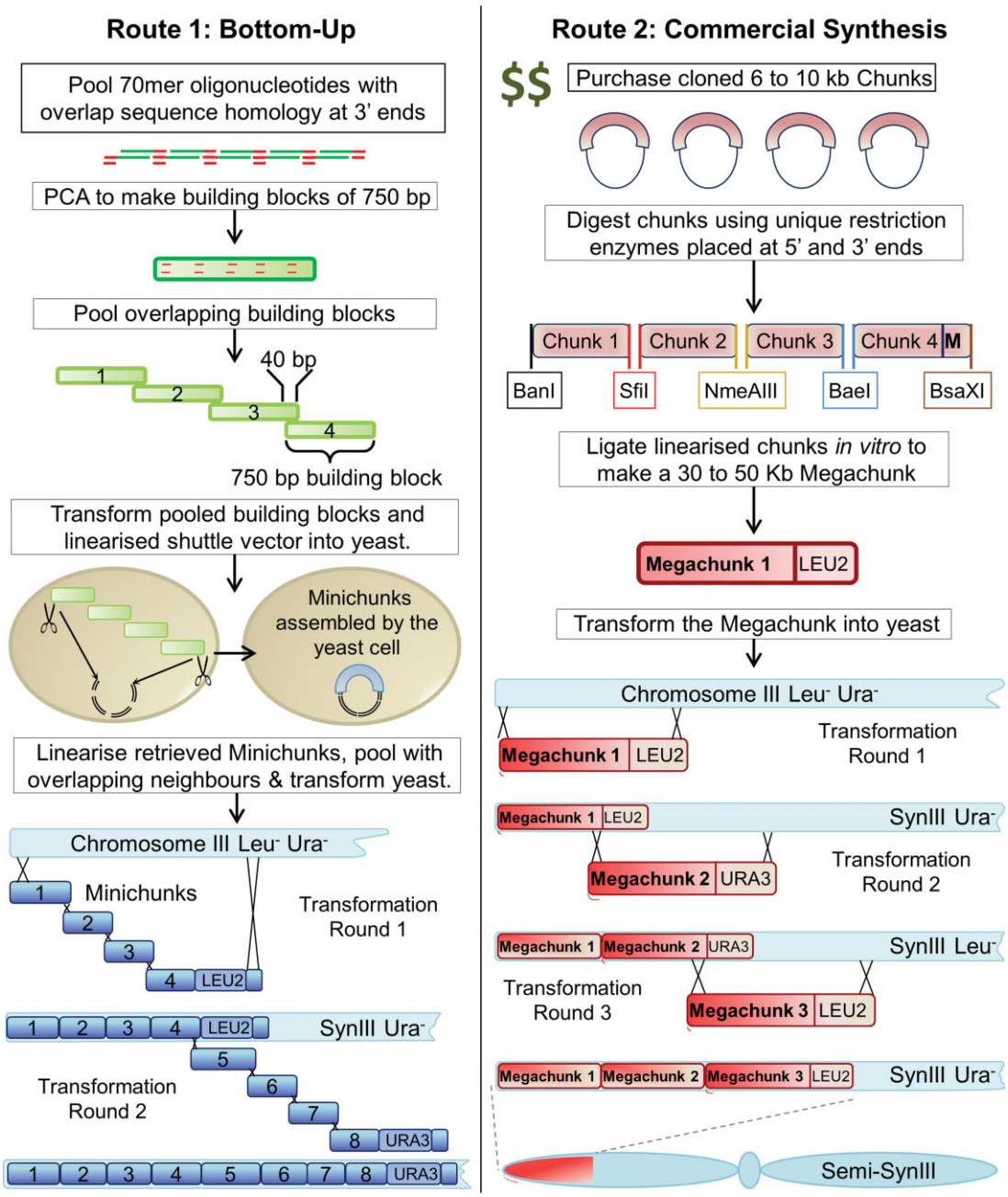

Figure 1. Assembly of a synthetic chromosome. Two routes towards the assembly of semi-synthetic yeast chromosome III are shown. **Route 1** shows the steps taken by Annaluru et al. [9] to assemble synthetic regions from pools of synthesised oligonucleotides using polymerase cycling assembly (PCA) to create overlapping building blocks that can be assembled into Minichunks using native yeast homologous recombination. Several overlapping Minichunks can then be introduced into yeast with a selectable auxotrophic marker gene (LEU2 or URA3) and recombine with the native chromosome to replace natural sequence with synthetic sequence. Repeating this process with interchangeable markers allows gradual construction of a synthetic chromosome, replacing ~10 kb of sequence each time. **Route 2** is a modified and more rapid approach that is currently used for other synthetic yeast chromosomes, where commercial DNA synthesis is used to create 10 kb Chunks that can be simply ligated together to produce Megachunks that replace ~50 kb of sequence each time.

recombination between multiple copies of this sequence, so instead their recombination is dependent on the Cre recombinase being expressed. The symmetrical nature of these loxPsym sequences means that when two loxPsym sites are brought together by a Cre recombinase, the DNA between can undergo either an insertion, a deletion, a translocation or an inversion (Fig. 2). If, for example a non-essential gene were located between two loxPsym sites on the chromosome, then expression of the Cre recombinase (induced by addition of estradiol to the yeast growth media) could lead to this gene

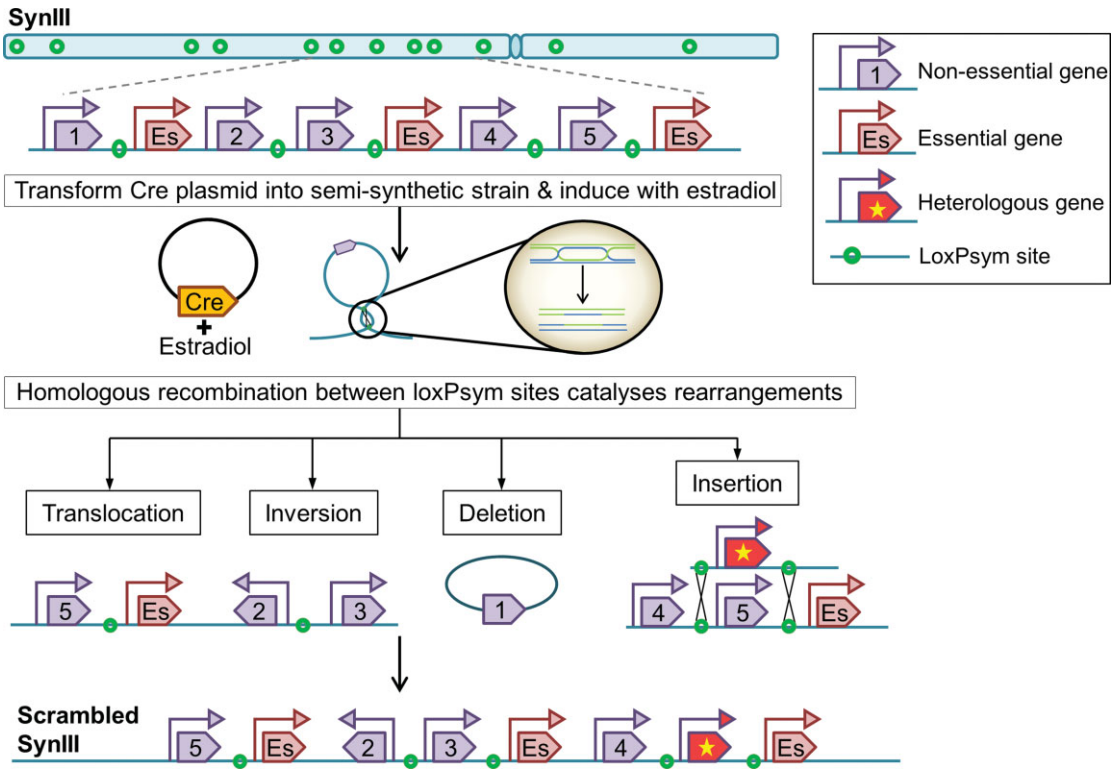

Figure 2. The SCRaMbLE system of synthetic yeast. Incorporating loxPsym sequences at sites downstream of non-essential genes in synthetic yeast chromosome allows for Cre-mediated recombination of the genomic DNA. Cre recombinase is provided on a plasmid and its expression induced by estradiol. It catalyses rearrangements between genomic loxPsym sites that can lead to translocations, inversions and deletions of genes. If a heterologous gene flanked by loxPsym sites is also provided at the time of recombination then this may become inserted at the expense of non-essential genes if it provides a growth advantage.

being deleted, or having its orientation inverted, or being translocated to another chromosomal region with loxPsym sites. This is precisely what has been seen in SCRaMbLE verification experiments performed previously with the semi-synVI and semi-synIX chromosomes described in 2011. Genes flanked by loxPsym sites that encode enzymes in the amino acid biosynthesis pathways, for example are lost when the Cre-recombinase is expressed in growth media where the amino acids are provided; cells that delete the biosynthesis genes can survive just fine as long as the amino acid is given [12]. Essentially SCRaMbLE is a tool for simultaneously studying the necessity for individual genes to be on the genome, and the effects caused by the loss or rearrangement of not just one gene, but many combinations of genes together.

Previous efforts in the yeast community have investigated the role each gene plays within the genome by first deleting individual gene one-by-one [21], and then by deleting every pair of genes and measuring the effect on the yeast's fitness [22]. The SCRaMbLE system built into the synthetic genome allows billions of different combinations of gene deletions, inversions and translocations to be sampled in a test-tube and only the genome rearrangements that allow the cells to keep growing competitively in the conditions of the experiment will be seen after recovery of the yeast cells. The number of possible rearrangements that can result from even just one chromosome being SCRaMbLEd is mind-boggling. The diversity that can be sampled when the entire Sc2.0 project is finished with thousands of loxPsym sites placed throughout the synthetic genome will be ground-breaking for research. In theory, this strain could be used to determine the minimal gene set required to have a functioning eukaryotic cell growing in the lab. The strain would simply need to be subjected to continual rounds of Cre recombinase rearrangements in rich growth media, and over time all non-essential genes would be deleted.

SCRaMbLE-ing has been designed to lead to loss of functions, desired gene deletions and/or gene rearrangements, in order to study how the gene layout and content of the genome defines how it works. However, the SCRaMbLE system could also be put to task for more biotechnology-focused aims in the future, not just by catalysing rearrangements or deletions but as a vehicle for incorporating new genes into the synthetic genome. Consider the situation where synthetic genes encoding a desired property – the ability to grow on an alternative carbon source, for example – are flanked by loxPsym sites and made available to the yeast cells when the SCRaMbLE system is activated (e.g. by co-transforming their DNA into yeast). If only the alternative carbon source is present in the growth media at this point then genomes that rearrange to include the genes to utilise this source will have a significant growth advantage, and will thrive following SCRaMbLE-ing at the expense of those that

do not. Thus, by applying certain environmental pressures on the yeast, it becomes favourable for synthetic yeast to incorporate any provided genes that can give the cells an advantage during the SCRaMbLE period (Fig. 2). One can imagine that loxPsym-formatted synthetic genes adapted from a diverse range of sources such as plants, bacteria or fungi could be deliberately mixed into synthetic yeast strains undergoing SCRaMbLE-ing in conditions where it would be desirable to evolve a synthetic yeast cell to survive somehow, such as in high concentrations of ethanol or on non-classic carbon sources such as those from lignocellulosic biomass. Synthetic yeast and its SCRaMbLE system offer a new kind of evolution to be exploited: a 'hyper-evolution' where gene content and genetic arrangement is radically altered, rather than individual DNA bases. This fits very much with the ethos of synthetic biology, where organism design is modular, and new functions can be derived by rewiring genetic modules into new combinations and arrangements [23, 24].

Concluding remarks

The purpose of the design and engineering of Sc2.0 is to improve understanding of eukaryotic genome structure and function, and to endow yeast with new in-built possibilities, such as the SCRaMbLE system. The radical re-design of the synthetic genome has so far been validated in terms of cell fitness, and this has been verified first for regions of chromosomes [12] and now for the entire chromosome III of yeast [9]. This shows that synthetic genomics is no longer limited to making a copy of an existing genome with watermarks but can extend to human-design of multiple features. It is to be hoped that this landmark development will encourage synthetic biologists to advance their work on metabolic pathways and gene regulation networks up to the chromosomal scale, where hundreds of genes could be combined together to produce novel drugs, alternative renewable energy sources and new chemicals and biomaterials needed in the forthcoming century.

References

1. **O'Malley MA, Powell A, Davies JF, Calvert J**. 2008. Knowledge-making distinctions in synthetic biology. *BioEssays* **30**: 57–65.
2. **Cheng AA, Lu TK**. 2012. Synthetic biology: an emerging engineering discipline. *Annu Rev Biomed Eng* **14**: 155–78.
3. **Gibson DG**. 2014. Programming biological operating systems: genome design, assembly and activation. *Nat Methods* **11**: 521–6.
4. **Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA**, et al. 2008. Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* **319**: 1215–20.
5. **Gibson DG, Glass JI, Lartigue C, Noskov VN**, et al. 2010. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**: 52–6.
6. **Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR**, et al. 2013. Genomically recoded organisms expand biological functions. *Science* **342**: 357–60.
7. **Isaacs FJ, Carr PA, Wang HH, Lajoie MJ**, et al. 2011. Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**: 348–53.
8. **Wang HH, Isaacs FJ, Carr PA, Sun ZZ**, et al. 2009. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**: 894–8.
9. **Annaluru N, Muller H, Mitchell LA, Ramalingam S**, et al. 2014. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**: 55–8.
10. **Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M**, et al. 1992. The complete DNA sequence of yeast chromosome III. *Nature* **357**: 38–46.
11. **Goffeau A, Barrell BG, Bussey H, Davis RW**, et al. 1996. Life with 6000 genes. *Science* **274**: 63–7.
12. **Dymond JS, Richardson SM, Coombes CE, Babatz T**, et al. 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477**: 471–6.
13. **Carlson R**. 2009. The changing economics of DNA synthesis. *Nat Biotechnol* **27**: 1091–4.
14. **Ellis T, Adie T, Baldwin GS**. 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integr Biol (Camb)* **3**: 109–18.
15. **Stemmer WP, Crameri A, Ha KD, Brennan TM**, et al. 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**: 49–53.
16. **Gibson DG, Young L, Chuang RY, Venter JC**, et al. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343–5.
17. **Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA**. 2007. USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. *Nucleic Acids Res* **35**: e55.
18. **Lobocka MB, Rose DJ, Plunkett G III, Rusin M**, et al. 2004. Genome of bacteriophage P1. *J Bacteriol* **186**: 7032–68.
19. **Sauer B**. 1987. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**: 2087–96.
20. **Schwenk F, Baron U, Rajewsky K**. 1995. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res* **23**: 5080–1.
21. **Giaever G, Chu AM, Ni L, Connelly C**, et al. 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387–91.
22. **Costanzo M, Baryshnikova A, Bellay J, Kim Y**, et al. 2010. The genetic landscape of a cell. *Science* **327**: 425–31.
23. **Purnick PE, Weiss R**. 2009. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **10**: 410–22.
24. **Blount BA, Weenink T, Ellis T**. 2012. Construction of synthetic regulatory networks in yeast. *FEBS Lett* **586**: 2112–21.