Synthetic genomics: from synthesis of prokaryotic genomes to synthesis of a fully functional eukaryotic chromosome

P. K. Gupta and V. Jaiswal*

Synthetic genomics is a new area of research and had its origin with the report of synthesis of the gene for yeast alanyl tRNA by Khorana and his co-workers in 1970. This was followed by the synthesis of a number of protein coding genes during 1980s and thereafter. In 1996, the concept of a minimal genome was put forward, suggesting that the synthesis of a genome by eliminating all non-essential genes should be possible and needs to be tried. Synthesis of genomes of some viruses like poliovirus and PhiX174, and those of bacteria including Mycoplasma genitalium and M. mycoides was reported during the first decade of the present century. In parallel with these, efforts underway at JC Venter Institute (JCVI), a project called Sc2.0 for the synthesis of all the 16 chromosomes of yeast, took shape at John Hopkins University in USA under the leadership of Jef Boeke and S. Chandrasegaran. Under this project, the synthesis of two chromosome arms (VIL and IXR) was reported in 2011 and that of the first functional chromosome (SynIII) was reported in April 2014, which marked the beginning of a new era in the field of synthetic genomics. Many more eukaryotic chromosomes will be synthesized in future and will be utilized for a variety of purposes.

Keywords: Eukaryotic chromosome, prokaryotic genomes, synthetic biology, synthetic genomics.

Synthetic genomics is a branch of synthetic biology (with a much wider scope) and has been variously defined. In its simplest definition, it includes manipulation and engineering of an organism at the whole genome level involving synthesis of one or more genes or the whole genome with an aim to develop products and processes of value. The history of synthetic genomics, therefore, can be traced back to 1965, when Robert Holley and his team reported the 77 nucleotide long sequence of yeast alanyl tRNA that was utilized by Har Gobind Khorana and his team for synthesis of the gene for this tRNA. It took them five years to synthesize this gene[1], but even before they reported their results, they knew that the gene they had synthesized did not represent the complete native yeast alanyl tRNA gene; the native gene was later found to be much longer in size. In 1970, it was also shown that generally a tRNA is derived from a precursor molecule, which is much longer in size than the tRNA itself, as shown in the case of *Escherichia coli* tyrosine suppressor tRNA[2]. With the availability of this information, Khorana and his team decided to synthesize yet another full-length gene responsible for the synthesis of *E. coli* tyrosine suppressor tRNA. During mid-1970s, the synthesis of 207 bp long sequence, containing the full-length 126 bp long gene for *E. coli* suppressor tRNA along with other regulatory sequences needed for its expression was completed[3,4]. Subsequently, efforts were made to synthesize protein-coding genes, and the first such gene was the 514 bp long human leukocyte interferon gene; its synthesis was completed and results reported in 1981 (ref. 5). These initial efforts for chemical synthesis of genes utilized the technique that was developed and standardized by Khorana, and included the following steps: (i) synthesis of oligonucleotides (by condensation), 5 to ~20 nucleotides long, representing both the strands of the gene of interest (GOI); (ii) arranging the oligonucleotides in a reasonable number of duplex molecules with single-stranded overhangs, each having a small complementary nucleotide sequence with similar overhangs of other duplexes, which facilitated base-pairing and thus helped in the joining of duplexes to generate much longer duplexes or the complete gene.

PCR-based methods for synthesis of genes were developed later after the discovery of PCR in 1985. Solid phase synthesis of oligonucleotides also became possible and other improved technologies, including gene synthesis machines were developed and made commercially available. Utilizing these improved technologies, thousands of protein-coding genes were later synthesized through commercial service providers. Thus, starting in

P. K. Gupta and V. Jaiswal are in the Department of Genetics and Plant Breeding, CCS University, Meerut 250 004, India
*For correspondence. (e-mail: pkgupta36@gmail.com)

Table 1. Synthetic and recoded genomes

Organism	Genome size; kb (approx.)	Reference
Poliovirus type I	7.558	7
Bacteriophage $\varphi \times 174$	5.836	8
Human endogenous retrovirus (HERV-K)	9.47	9
HIV_{CPZ}	9.74	10
Spanish influenza virus of 1918	10	11
SARS virus	29.75	12
Mycoplsma genitalium JCVI-1.0 (synthetic genome)	580	13
Mycoplsma mycoides JCVI syn-1.0 (synthetic cell)	1080	14
Phaeodactylum tricomutum (chromosomes 25 and 26)	497; 441	15
Saccharomyces cerevisiae (chromosome arms VIL and IXR)		17
S. cerevisiae (chromosome III)	316,667	19
Escherichia coli (replacement of TAG stop codons with TAA)	4600	20
E. coli (genomically recoded; stop codons replaced)	4699	21

the late 1980s, synthesis of a gene became routine, and one could outsource synthesis of a GOI at a cost, which was affordable by individual research groups.

The research activity involving synthesis of genes was followed by synthesis of whole genomes, leading to synthesis/recoding of several genomes (Table 1). The latest in this research area is the synthesis of a fully functional eukaryotic chromosome (SynIII), which is the first of 16 yeast chromosomes, which will be synthesized under the project Sc2.0. A timeline for important events in the history of synthetic genomics during 2001–2014 is presented in Figure 1. This article gives a brief account of the research conducted in the field of synthetic genomics with major emphasis on the synthesis of a fully functional yeast chromosome 3, led by a scientist of Indian origin, S. Chandrasegaran in collaboration with Jef Boeke, both of whom were then at John Hopkins University (JHU), USA. The account presented in this article is restricted to synthetic genomics mainly involving the synthesis of genomes and a fully functional eukaryotic chromosome; no effort has been made to cover the whole area of synthetic biology (for synthetic biology, consult a collection of the literature compiled in a special issue of *Nature*, 17 June 2014; also see text below.

Synthetic biology versus synthetic genomics

Synthetic biology includes development of novel organisms through both, the genetic or metabolic engineering of existing organisms (top-down approach), and assembly of organisms with novel synthetic genomes (bottom-up approach). Thus it is apparent that synthetic biology may not always involve the development and use of synthetic genes and genomes. The top-down approach of synthetic biology is exemplified by the development of either the insect-resistant transgenic crops, where cloned *Cry* genes have been utilized, or the antimalarian drug artemisinin, where the approach of metabolic engineering was utilized for transforming yeast cells through as many as ~50 modifications in the existing genome. However, in this article, we will restrict ourselves with the bottom-up approach of synthetic biology, which includes synthetic genomics.

Synthesis of whole genomes

Starting in 1996, synthesis of a minimal genome, by dropping out the non-essential genes, was conceived in several organisms[6]. Later, whole genomes of several viruses and bacteria were synthesized, which will be briefly discussed.

Synthesis of viral genomes

Starting in 2002, success was achieved in the synthesis of complete genomes of several viruses[7–12], including the following two important viruses: (i) 7558 bp long genome of poliovirus[7] and (ii) 5836 bp long genome[8] of single-stranded phage φX174 (also see Table 1).

Synthesis of bacterial genomes

In 2005, Craig Venter started synthesis of bacterial genomes and in 2008 reported results of the synthesis of the complete genome of *Mycoplasma genitalium*[13]. Later in 2010, the group reported the results of design, synthesis and assembly of the 1.08 megabase (Mb) pairs long genome of *Mycoplasma mycoides* JCVI-syn1.0 and its transplantation in a recipient cell belonging to *M. capricolum*. The new cells with synthetic genome had expected phenotypic characteristics and were capable of self-replication, suggesting that the synthetic genome functioned like a native genome within a cell[14].

An interesting feature of the synthesis of a bacterial genome at the J. Craig Venter Institute (JCVI) was also the use of watermarks, to provide evidence that the genome is synthetic and not a native genome. This also provides protection against infringement of an intellectual property in the form of synthetic genome. These watermarks were specially designed segments of DNA sequences, which used the alphabets of protein that enable to spell out words or phrases (words could be spelt using the letters for the 20 essential amino acids that would be produced by DNA triplets used in the watermarks following the genetic code). In 2008, to design watermarks, JCVI used codons, which coded for 20 letters (which are routinely used for 20 amino acids); for the remaining six

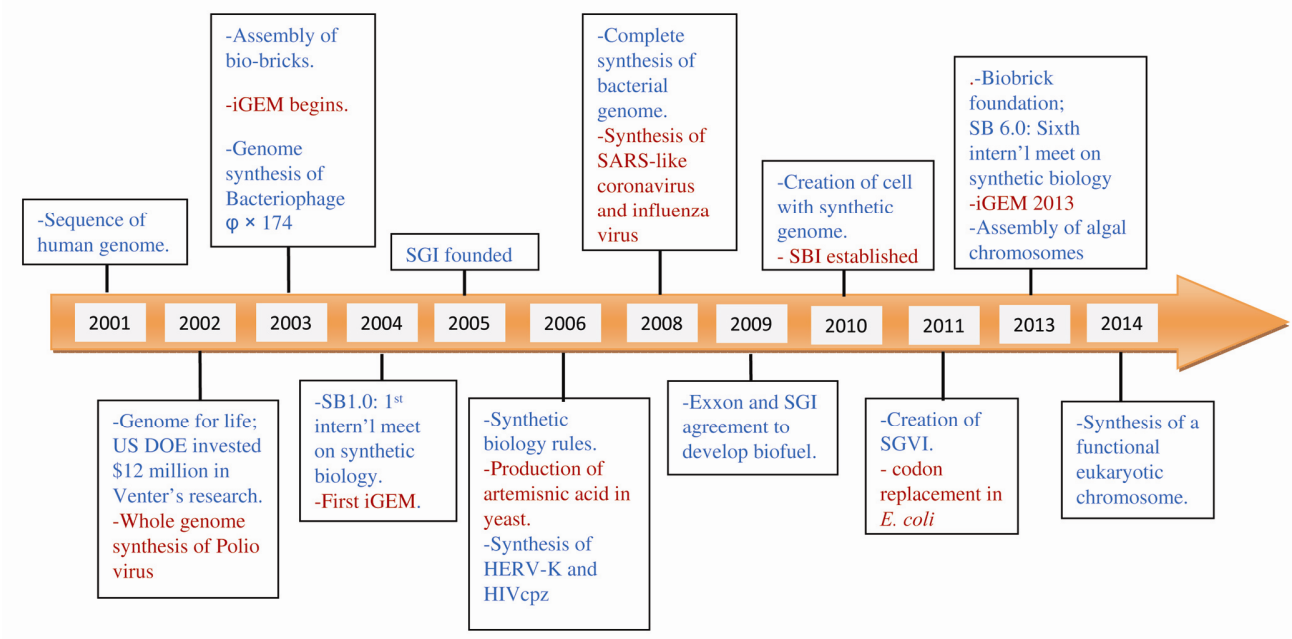

Figure 1. A timeline for events in the history of synthetic genomics.

alphabets, alternatives were used. For instance, one watermark was written as 'VenterInstitvte' (because there is no amino acid represented by the letter 'u', v was used for u in the word 'institute').

However, in 2010, for the creation of *M. mycoides* JCVI-syn1.0, the JCVI produced much larger and more elaborate watermarks. Each of these watermarks was more than 1000 base pairs long (1246, 1081, 1109 and 1222) and encoded all the 26 letters and a variety of punctuations for the English language. The watermarks used carried the following messages: (i) an explanation of the coding system used, (ii) URL address for sending a message, (iii) a list of 46 authors and contributors, and (iv) a series of famous quotes. In order to minimize the effect of watermarks on biological phenomena, they were inserted in intergenic regions that are also tolerant to transposon insertion. However, they did not reveal how the watermarks were encoded. Watermarks were included as DNA sequences, but how to get the meaning out of this was left as a puzzle to be decoded by the readers. By 2012, more than 100 people had already decoded these watermarks, and it still remains a puzzle for those who want to try and decode it.

Synthesis of eukaryotic chromosomes

When the synthesis of a bacterial genome was being undertaken at JCVI, a project for the synthesis of 12 Mb yeast genome involving synthesis of all the 16 yeast chromosomes was also conceived elsewhere (JHU, USA); the project was named Sc2.0 (Figure 2). For the success of this project, techniques were also developed for the synthesis of long designer DNA segments up to 1.8 Mb (by adding yeast replication origin every 100 kb) followed by their cloning in yeast cells. Also, when the synthesis of two individual chromosome arms was already completed (see later), and that of an entire eukaryotic chromosome was in progress at JHU, synthesis and assembly of two algal chromosomes belonging to a diatom was also completed at JCVI. The synthesis of yeast chromosome III was completed later at JHU. Both these works will be briefly described.

Synthesis of two algal chromosomes

The model diatom, *Phaeodactylum tricomutum* has a 27.4 Mb genome, consisting of 33 chromosomes. Synthesis of two chromosomes of this model alga, namely chromosomes 25 (497 kb) and 26 (441 kb), was reported in 2013 from JCVI. For the synthesis of these two chromosomes, assembly-based approach was used, which involved cloning of small fragments in yeast cells followed by assembly of full-length chromosomes. Fragments longer than ~150 kb were maintained by addition of yeast replication origins. Each chromosome was first cloned as five fragments, each ~150 kb long, with overlapping fragments of 10 to 40 kb. This demonstrated for the first time that assembly and maintenance of eukaryotic chromosomes is possible in yeast cells[15].

Synthesis of two chromosome arms and a fully functional designer yeast chromosome

It is known that baker's yeast (*Saccharomyces cerevisiae*) has a haploid complement of 16 chromosomes, with a

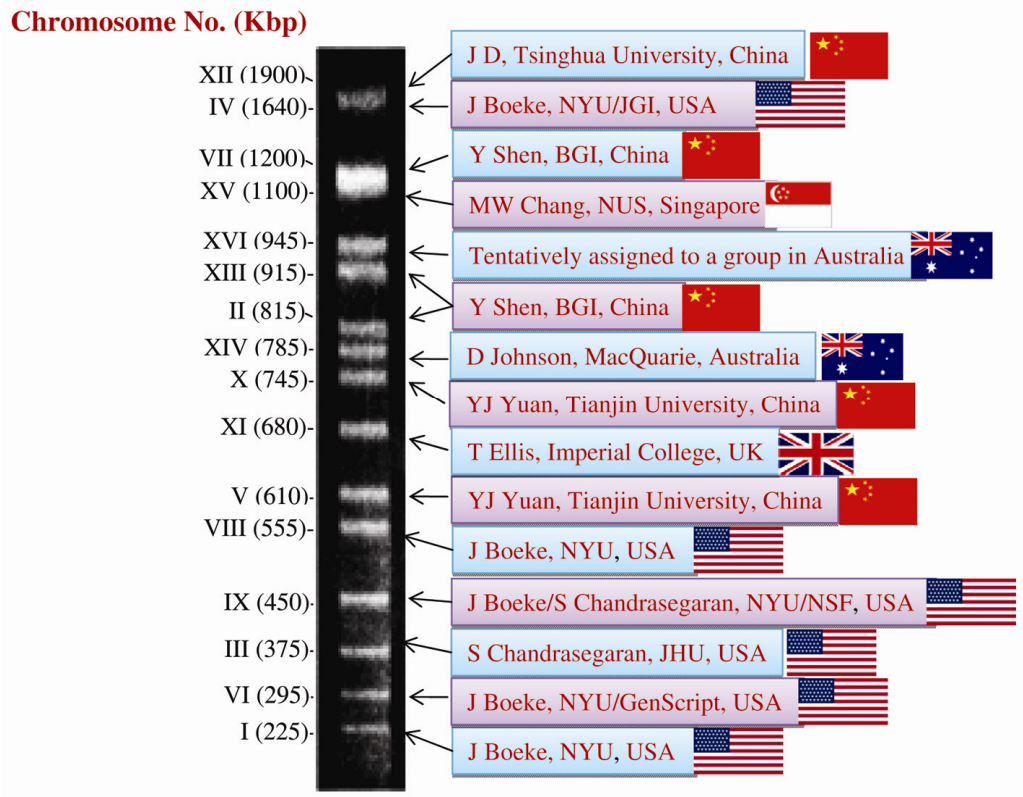

Figure 2. Assignment of 16 different yeast chromosomes (I to XVI) to different laboratories (with names of group leaders). From extreme left to right are chromosome numbers (with size of chromosomes), pulse field gel electrophoresis (PFGE) image showing 14 bands representing 16 chromosomes, and the group leaders with associated laboratories engaged in the synthesis (based on http://syntheticyeast.org/collaborators/ and S. Chandrasegaran, pers. commun.).

genome size of 11–12 Mb carrying 6000 genes; of these as many as 5000 genes are known to be non-essential[16]. With this background information, in parallel with the efforts at JCVI to assemble algal chromosomes, efforts were initiated at JHU and other centres (spread world over) to synthesize yeast chromosomes under the project Sc2.0 (Figure 2). In order to reduce the cost, a large number of undergraduate and many graduate students were also involved in the project. Boeke also taught an undergraduate course 'Build-A-Genome' (BAG), where students synthesized oligonucleotides that were 750 bp long and were needed as building blocks (BBs) for the Sc2.0 project.

In 2011, as a part of Sc2.0, a team of workers led by Chandrasegaran and Boeke at JHU (Boeke is now at New York University Langone Medical Centre) reported the synthesis of two chromosome arms of different yeast chromosomes (synVIL and synIXR)[17]. In this study, the utility of an entirely new inducible system called SCRaMbLE (Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution) was also demonstrated[18]; the system was later utilized for creating diverse forms during the synthesis of the entire chromosome III (see below for details).

In 2014, the same group reported the synthesis of the first fully functional eukaryotic chromosome, the yeast chromosome III, which is one of the smallest yeast chromosomes[19]. The synthetic chromosome (designated as synIII) first designed *in silico*, and then assembled in pieces and used to replace the native chromosome in an yeast cell using the following steps: (i) 60- to 79-mer oligonucleotides, in large numbers, covering the entire length of synIII were synthesized as the starting point; (ii) 367 BBs (each ~750 bp long) were produced from overlapping 60- to 79-mer oligonucleotides using PCR; this was done by undergraduate students in the BAG course taught by Boeke at JHU; 133 BBs belonged to the left arm of chromosome III (synIIIL), and 234 BBs belonged to the right arm (synIIIR); (iii) 127 BBs were assembled into overlapping minichunks (2–4 kb in length; 44 for synIIIL and 83 for synIIIR); (iv) the minichunks were assembled *in vivo* to give synIII by homologous recombination. Using an average of 12 minichunks and selectable markers in each experiment, native chromosome III was replaced by synIII through 11 successive rounds of transformation. Yeast cells with synIII were found to function normally, as shown by comparing them with native cells for a number of characteristics,

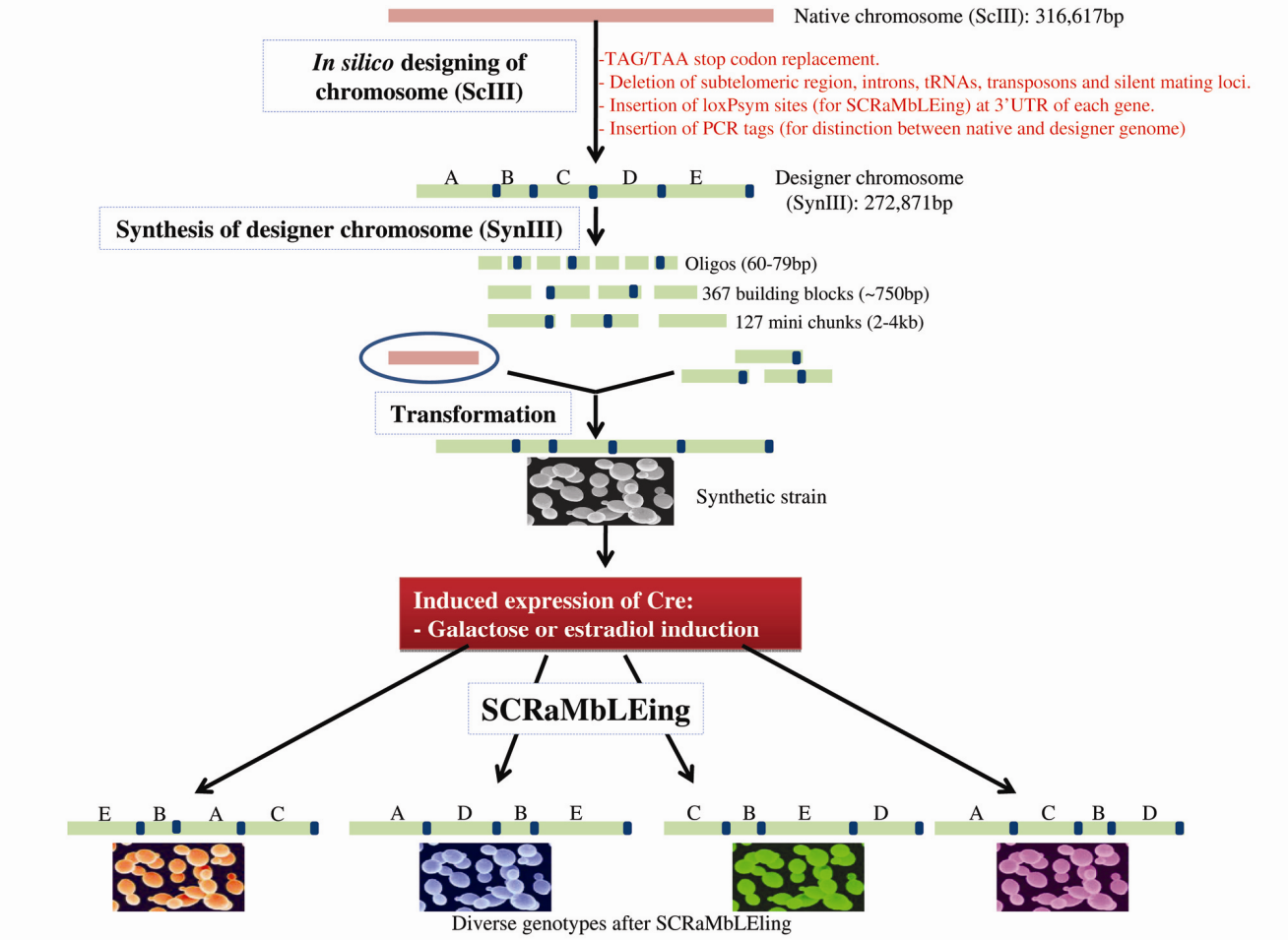

Figure 3. Schematic representation of different steps involved in SCRaMbLEing. At the bottom, four diverse genomes produced due to SCRaMbLEing are shown by different colours; the corresponding chromosomes carry loxPsym sites (shown by blue rectangles).

which included colony size, growth curves, and morphology under different conditions. The results were published in an on-line issue of *Science Express* on 27 March 2014, and later published in *Science*[19] on 4 April 2014.

The synthetic chromosome, synIII was 272,871 bp long and did not entirely resemble the native chromosome III (316,667) found in yeast cells. The assembled chromosome had more than 500 alterations (deletions and rearrangements), including the following: (i) removal of repeating segments of ~50,000 bp, which were considered unnecessary for cell function and its reproduction; (ii) removal of 'junk DNA' and 'jumping genes'; (iii) replacing the deleted sequences with entirely new sequences that were absent in the native chromosome. SCRaMbLE approach was used for the so-called chromosome SCRaMbLEing (see next paragraph), which involved rearrangement of nucleotides within the chromosome, thus making it possible to produce a large number of variants at will. The synthetic chromosome, assembled within a cell, was shown to function like a native chromosome, suggesting that novel chromosomes can be synthesized and used for the production of novel products like biofuel, drugs, vaccines, etc.

SCRaMbLE approach

SCRaMbLE is an inducible evolution system that restructures a synthetic genome at will, so that it became possible to develop and synthesize new genome variants that are not available in nature[18]. The following steps were involved in SCRaMbLEing (Figure 3): (i) the synthetic chromosome was first designed *in silico*, which involved listing of sequences of oligos to be synthesized (taking into consideration the segments to be deleted, replaced or added), (ii) insertion of loxPsym sites (for SCRaMbLEing) and (iii) insertion of PCR tags for distinction between native and designer chromosomes. The complete synthetic chromosome was actually assembled *in vivo* within a cell through transformation of yeast cells, where expression of Cre was induced through galactose or estradiol treatment. This promoted SCRaMBbLEing, leading to the production of diverse synthetic genotypes.

Synthesis of all the 16 yeast chromosomes in future

Once chromosome III of yeast was successfully synthesized and assembled *in vivo* to function normally within a cell, efforts are now being made to synthesize all the 16 chromosomes of yeast in different research laboratories located in five countries, namely USA, UK, China, Singapore and Australia. Out of 16 chromosomes, 10 (1–6, 8, 9, 11 and 12) are in various stages of synthesis and assembly; however, synthesis of the remaining 6 chromosomes is yet to start. Details of laboratories involved in the synthesis of each of the 16 chromosomes are given in Figure 2.

Genomically recoded *E. coli*

Another impressive and landmark achievement reported in 2013 involved recoding of the genome of *E. coli*, even though it did not amount to *de novo* synthesis of a genome. This involved editing of the 4.6 Mb long *E. coli* genome involving replacement of all UAG stop codons by synonymous UAA codons. This permitted the elimination of release factor 1 (RF1) and thus also allowed reassignment of the UAG translational function (by incorporating appropriate tRNAs and aminoacyl-tRNA synthetases for these novel amino acids). This allowed incorporation of some non-standard amino acids in the proteins, thus expanding their chemical diversity[20,21].

Applications of synthetic genomics

Synthetic genomics has many possible applications, although relative to the applications of synthetic biology, which is a much broader area, these applications are yet to be realized at a measurable level. One of the most widely discussed examples of synthetic biology is the semi-synthetic production of an antimalarial drug 'artemisinin' that was produced using engineered yeast[22]. A similar example is the production of live attenuated viral vaccines by means of computer-aided rational design followed by the introduction of genome-wide changes through synthesis of long stretches of DNA[23,24]. However, these two products may not be suitable examples of the use of synthetic genomics, since mainly metabolic engineering rather than synthetic genomics was exploited for these products.

Possible future applications of synthetic genomics have also been widely discussed, although there seems to be hardly any example available, where it has already been exploited at a commercial scale. The possible applications of synthetic genomics can be broadly classified into two categories. First, the generation of knowledge, which may include either knowledge about the maximum number of non-essential genes, which can be deleted in an organism, or an enhancement of basic understanding of life itself. Second and the more important application of synthetic genomics includes exploitation of the products of synthetic genomics at an industrial scale. These future possible applications include the following: (i) development of major alternative sources of renewable energy, including those for petrol-chemical industry; (ii) development of microbes or plants for bioremediation; (iii) discovery and production of novel antibiotics, vaccines, drugs and other novel chemicals and biomaterials. These possible applications will certainly be achieved within the next one or two decades, so that we are now at the threshold of a new era, where synthesis of novel genomes will be exploited in a big way for a variety of purposes.

In the UK, at the University of Edinburgh, a BBSRC-funded project named IESY (Inducible Evolution of Synthetic Yeast genomes) has been initiated. Under this project, yeast strains will be produced for industrially high-value phenotypes involving the following three attributes: (i) growth in glucose-limiting medium, to become relevant for use in industrial bioreactors; (ii) growth in high ethanol concentrations, that will prove useful for production of biofuels, and (iii) production of a carotenoid or other high-value metabolites. Synthetic genomics involving SCRaMbLE approach will also provide immense opportunities for producing diverse forms.

Ethical issues and risks involved in using synthetic genomics

Synthetic genomics, according to some, amounts to the creation of new artificial life forms. Therefore, by its very nature, synthetic genomics raises ethical issues like scientists 'playing God'. It also raises concerns about the risks of its use in bioterrorism. In response to these issues, a number of scientists have proposed a modified version of Turing's test for life imitation, which is prescribed for distinguishing artificial forms from the naturally occurring forms. However, it is unclear whether such moves to redefine life will ameliorate deeper fears about blurring of the boundary between artificial and natural life. Craig Venter and his team have been concerned with the societal issues surrounding the work involving synthetic genomics. In 1995, while the team was doing research on the minimal genome, the work underwent significant ethical review, and concluded that the prospect of constructing minimal and new genomes does not violate any fundamental moral precepts or boundaries, but does raise questions that need to be examined before the technology advances further[25].

When new forms of life are created through synthetic genomics, accidental release of these new life forms into the environment, and their subsequent evolution, proliferation and unexpected interactions might alter the ecosystem. This is an ethical issue and also a risk perceived by some. Synthetic genomics can also be used to produce

new microorganisms that are specifically designed to be hostile to humans, as evident from the synthesis of the poliovirus and the pandemic Spanish flu virus of 1918. Such research has been described as dual use research of concern (DURC), which has been adequately addressed[26]. However, major concern arises from state-level biological warfare programmes. A number of proposals have been made by both scientific groups and the government agencies to address the dual use (military/civilian) nature of synthetic genomics. This will include the overall control over both, the commercial DNA synthesis or public research, and the impact of synthetic genomics on international bioweapons conventions.

Another concern about the use of synthetic genomics is the possibility of commercial monopolies that may emerge due to patenting of the products and processes developed through the use of synthetic genomics. One such area is the drive to create a microorganism that can turn biomass into fuels such as ethanol or hydrogen. In response, there have been moves to develop an open-source movement (based on so-called BioBricks) involving creation of a 'commons' that will facilitate open scientific research.

Regulatory issues in using synthetic genomics

Since there are perceived risks involved in using synthetic genomics for useful purposes, the products of this technology also need to be subjected to regulation. In order to increase awareness among providers and users involved in synthetic genomics, in 2006 the National Science Advisory Board on Biosecurity (USA) published a report entitled 'Addressing biosecurity concerns related to the synthesis of select agents' (www.biosecurityboard.gov/links.asp). This document provides guidance for the regulation of select agents for those who create and use the products of synthetic genomics (select agents are microbes and toxins, which pose a significant risk). The regulations implement the provisions of different Acts, which have been promulgated for this purpose. Individuals applying for access to select agents must undergo a security risk assessment by the Federal Bureau of Investigation, Criminal Justice Information Service (FBI/CJIS). Information on the select agent regulations is available at the US national select agent website (www.selectagents.gov).

Since the US and British research centres are in the forefront of synthetic genomics research, it would be expected that they would be heavily involved in the development of new regulations, if needed; they would also examine whether current regulations are adequate to address the risks involved. The development of regulatory system for synthetic genomics in other countries may take some time.

Patenting of the products of synthetic genomics is another important issue, which has been addressed. Watermarks inserted in synthetic genomes by JCVI are an example of, how protection of intellectual property can be achieved. This may be necessary to encourage research in this area leading to development of industrial products of value. However, as mentioned earlier, this protection may lead to commercial monopolies, so that a balance may have to be struck between protection and the much needed open source for products of public good (e.g. biofuel).

Summary and conclusion

Synthetic genomics is a new area of research, which has emerged recently with the synthesis of whole viral/bacterial genomes and a fully functional yeast chromosome. The research area has also witnessed the use of some novel approaches, which include insertion of watermarks during synthesis to distinguish a synthetic genome from a native genome, and the use of LoxPsym sites to facilitate the so-called SCRaMbLEing for generating a variety of genotypes. There are also ethical issues, the risks of bioterrorism involved and the regulatory measures, which need to be addressed. Despite these limitations, we anticipate that, synthetic genomics will grow rapidly and will be extensively used in future for a variety of purposes. Some of the important and widely discussed uses of synthetic genomics include the development of novel vaccines and new sources of bioenergy/biofuel. This is apparent by the establishment of the Synthetic Genomics Vaccine Inc. (SGVI) by Synthetic Genomics Inc. (SGI) and JCVI in 2010 at La Jolla, California, USA, for developing vaccines, and by the agreement between ExxonMobil and SGI for the production of biofuel from algae using synthetic genomics approach. SGVI is already collaborating with Novartis for the development of influenza vaccines using synthetic genomics approach. Thus, synthetic genomics has already become an important area of research and will also be used in future to answer basic questions of biology. Although scientists in USA and Europe (including the UK) are actively involved in synthetic genomics research, other countries including India, will certainly take part in this new and exciting research area in the near future. India has already taken part in sequencing several genomes of crop plants, but there is need to allocate funds for initiating projects in the fields of synthetic genomics, although it lost the opportunity to be a part of Sc2.0. The subject also needs to find a place in undergraduate and postgraduate syllabi of biology in the universities (by designing courses like BAG) and students have to take part in competitions like iGEM to enjoy the excitement of this new area of research.

1. Agarwal, K. L. *et al*., Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. *Nature*, 1970, **227**, 27–34.
2. Goodman, H. M., Abelson, J. N., Landy, A., Zadrazil, S. and Smith, J. D., The nucleotide sequences of tyrosine transfer RNAs of *Escherichia coli*. *Eur. J. Biochem*., 1970, **13**, 461–483.
3. Khorana, H. G. *et al.*, Total synthesis of the structural gene for the precursor of a tyrosine suppressor transfer RNA from *Escherichia coli*. 1. General introduction. *J. Biol. Chem*., 1976, **251**, 565–570.
4. Khorana, H. G., Total synthesis of a gene. *Science*, 1979, **203**, 614–625.
5. Edge, M. D. *et al*., Total synthesis of a human leukocyte interferon gene. *Nature*, 1981, **292**, 756–762.
6. Maniloff, J., The minimal cell genome: 'On being the right size'. *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 10004–10006.
7. Cello, J., Paul, A. V. and Wimmer, E., Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science*, 2002, **297**, 1016–1018.
8. Smith, H. O., Hutchison III, C. A., Pfannkoch, C. and Venter, C. J., Generating a synthetic genome by whole genome assembly: PhiX174 bacteriophage from synthetic oligonucleotides. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 15440–15445.
9. Dewannieux, M. *et al*., Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res*., 2006, **16**, 1548–1556.
10. Keele, B. F. *et al.*, Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science*, 2006, **313**, 523–526.
11. Taubenberger, J. K. and Morens, D. M., The pathology of influenza virus infections. *Annu. Rev. Pathol*., 2008, **3**, 499–522.
12. Becker, M. M. *et al*., Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 19944–19949.
13. Gibson, D. G. *et al*., Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, 2008, **319**, 1215–1220.
14. Gibson, D. G. *et al.*, Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010, **329**, 52–56.
15. Karas, B. J. *et al.*, Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng*., 2013, **7**, 30.
16. Goffeau, A. *et al*., Life with 6000 genes. *Science*, 1996, **274**, 546–567.
17. Dymond, J. S. *et al*., Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, 2011, **477**, 471–476.
18. Dymond, J. S. and Boeke, J., The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioeng. Bugs*, 2012, **3**, 168–171.
19. Annaluru, N. *et al*., Total synthesis of a functional designer eukaryotic chromosome. *Science*, 2014, **344**, 55–58.
20. Issacs, F. J. *et al.*, Precise manipulation of chromosomes *in vivo* enables genome wide codon replacement. *Science*, 2011, **333**, 348–353.
21. Lajoie, M. J. *et al*., Genomically recoded organisms expand biological functions. *Science*, 2013, **342**, 357.
22. Paddon, C. J. *et al*., High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 2013, **496**, 528–532.
23. Wimmer, E., Mueller, S., Tumpey, T. M. and Taubenberger, J. K., Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nature Biotechnol*., 2009, **27**, 1163–1172.
24. Mueller, S. *et al*., Live attenuated influenza virus vaccines by computer-aided rational design. *Nature Biotechnol*., 2010, **28**, 723–726.
25. Cho, M. K., Magnus, D., Caplan, A. L. and McGee, D., Ethical considerations in synthesizing a minimal genome. *Science*, 1999, **286**, 2087–2090.
26. Gupta, P. K., Regulating the dual-use and dual-impact life science research: influenza virus versus biotech crops. *Curr. Sci*., 2012, **103**, 995–1002.

ACKNOWLEDGEMENTS. We thank the National Academy of Sciences India for the Award of Platinum Jubilee Senior Scientist Fellowship. We also thank the Head, Department of Genetics and Plant Breeding, CCS University, Meerut for providing the necessary facilities.

Received 31 March 2014; revised accepted 2 September 2014