Journal of Biotechnology 160 (2012) 72–79

Contents lists available at SciVerse ScienceDirect

Journal of Biotechnology

journal homepage: www.elsevier.com/locate/jbiotec

In the fast lane: Large-scale bacterial genome engineering

Tamás Fehér[a], Valerie Burland[b,c], György Pósfai[a,*]

[a] Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, 62 Temesvari krt., H-6726 Szeged, Hungary
[b] Scarab Genomics LLC, 1202 Ann Str., Madison, WI 53713, USA
[c] Department of Genetics, University of Wisconsin, 425-G Henry Mall, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 4 November 2011
Received in revised form 16 February 2012
Accepted 21 February 2012
Available online 1 March 2012

Keywords:
Genome engineering
High throughput
Bacteria
Synthetic biology

ABSTRACT

The last few years have witnessed rapid progress in bacterial genome engineering. The long-established, standard ways of DNA synthesis, modification, transfer into living cells, and incorporation into genomes have given way to more effective, large-scale, robust genome modification protocols. Expansion of these engineering capabilities is due to several factors. Key advances include: (i) progress in oligonucleotide synthesis and in vitro and in vivo assembly methods, (ii) optimization of recombineering techniques, (iii) introduction of parallel, large-scale, combinatorial, and automated genome modification procedures, and (iv) rapid identification of the modifications by barcode-based analysis and sequencing. Combination of the brute force of these techniques with sophisticated bioinformatic design and modeling opens up new avenues for the analysis of gene functions and cellular network interactions, but also in engineering more effective producer strains. This review presents a summary of recent technological advances in bacterial genome engineering.

1. Introduction

In the past two decades, our encyclopedic knowledge of the gene-pool of the microbial world has been vastly expanded. Thanks to dramatic improvements in DNA sequencing techniques (Niedringhaus et al., 2011) and bioinformatic services, the genetic composition of a microbe can now be deciphered within days at an ever decreasing cost. The next great challenge is, however, putting this knowledge to use by building new, synthetic biological constructs, ranging from mutant genes to complex genetic modules, or even to whole new organisms. Building functional biological circuits is, however, a much more complex task than performing a simple read-out of the genetic composition. Beyond the gaps in our detailed knowledge of host biology, creation of novel biological systems may face intrinsic challenges stemming from the application of untested genetic components that have never evolved in nature.

With remarkable foresight, geneticist Waclaw Szybalski wrote in 1974: "Let me now comment on the question "what next". Up to now we are working on the descriptive phase of molecular biology. . . . But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements . . . or build up wholly new genomes. This would be a field with the unlimited expansion potential and hardly any limitations to building "new better control circuits" and . . . finally . . . "synthetic" organisms, like a "new better mouse"" (Szybalski, 1974).

Although a truly synthetic, "new better mouse" is still out of reach, building new, better microbes, much less complex than a mouse, became a real-time endeavor. The last few years have brought about a quantum leap in bacterial genome engineering. The long-established manual tools and laborious methods of genome engineering (e.g., transposon mutagenesis, conjugation, P1 transduction, gene targeting by suicide plasmids), together with the more recent lambda Red system-mediated homologous recombination methods, while still remaining useful techniques, can now be optimized and scaled up to fast, automated, and highly parallel, robust genome modification protocols. Fast and efficient genome engineering, on one hand, would advance our knowledge of microbial systems by rapidly mapping phenotypes to genetic modifications on the genome scale. On the other hand, robust, parallel rewiring of large gene-sets, coupled with efficient selection and analysis of useful variants may yield new, improved producer cells of valuable biomolecules.

Here we discuss recent technological advances and review the new tools now available for bacterial genome engineering. First, important developments in the synthesis of oligonucleotides, joining of oligos and PCR fragments into larger assemblies, and optimization of recombineering technology are described. This is followed by discussion of complex procedures where these tools

* Corresponding author. Tel.: +36 62 599 778; fax: +36 62 433 506.
E-mail address: posfai@brc.hu (G. Pósfai).

0168-1656/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2012.02.012

Table 1
Companies providing custom microarray-based oligonucleotide library synthesis.

Vendor	Features	Length (nt)
Agilent	55,000	200
Atactic	3900	90
LC sciences	<10,000	100
MYcroarray	20,000	150
Nimblegen	4,200,000	50–75

were integrated for high-throughput, to generate highly engineered, re-programmed bacterial strains.

2. Oligonucleotides, the elementary tools of genome engineering

Traditional applications of short, synthetic oligos in recombinant DNA technology include the introduction of restriction enzyme recognition sites for cloning, mutagenesis of a DNA segment, priming of PCR amplifications, and pasting together recombinant molecules. Today, the large scale and relatively low cost of short oligo synthesis with new oligo assembly techniques has resulted in replacing classic construct engineering by DNA synthesis, now frequently outsourced to commercial providers.

Most DNA synthesizers available in the market use phosphoramidite chemistry to elongate polystyrene- or controlled pore glass-bound oligos in the $3' \rightarrow 5'$ direction (Caruthers, 1985; Caruthers et al., 1983, 1987). Certain characteristics of this scheme hinder its use in genome-scale applications. The pace of cost-reduction of column-based oligo synthesis seems to be slowing down, and the major part of oligo stocks is mostly wasted (more than is required) even when purchased in the smallest possible scale (Gibson et al., 2010b). In addition, accidental deletions, insertions, and depurinations may strongly reduce the fidelity of oligos, and consequently limit the length of robust synthesis (Hughes et al., 2011). Developments are, however, already on the way to control depurination (a common side reaction), allowing the reliable synthesis of oligos of 150 nt or longer (LeProust et al., 2010; Sierzchala et al., 2003), but the most promising ways of decreasing costs and increasing throughput are microarray-based synthesis and microfluidics.

Microarray-based synthesis uses basically the same chemistries, but significantly reduces reagent-requirement and waste. At present, three different technologies are used for selective elongation of oligo chains in parallel: photolitography, the light-directed control of certain synthesis steps (Fodor et al., 1991; Barone et al., 2001; Gao et al., 2001; Singh-Gasson et al., 1999), inkjet printing, the confinement of synthesis to discrete microdroplets (Blanchard et al., 1996; Hughes et al., 2001), and electrochemically directed synthesis (Chow et al., 2009; Egeland and Southern, 2005). Microchip synthesis can produce a vast number (up to 4.2 million) of different oligos at fractional costs (0.001 cent/nt (Carr and Church, 2009) vs. 10 cents/nt (Cheong et al., 2010)). However, each oligo (only 10^6–10^7 molecules) (Tian et al., 2009) must be amplified prior to use, which requires assigning a significant portion of their length (~40 nt) as primer binding sites. This sacrifice nevertheless pays off, since the inclusion of specific primer binding sites in oligo subsets permits their selective amplification for downstream applications. It must be noted, however, that longer oligos pose a higher unit cost (over 0.04 cents/nt) and so far are available only in lower feature densities (i.e., carrying fewer distinct sequences per chip) (Table 1). The key to low-background subclass amplification and efficient downstream applications is 'orthogonal' primer design. This is a computer-automated, high throughput oligo design scheme that matches melting points and avoids cross-hybridization and secondary structure in the primer sequences. George Church's group provided an illustrative example of subset amplification in a two-stage PCR starting from a set of 13,000 oligos encoding a total of 2.5 Mbp. Each amplified subclass was then used for PCR-mediated assembly resulting in a total of 47 synthetic genes (Kosuri et al., 2010).

Microfluidic devices could provide an alternative technology for small-scale parallel synthesis of oligos (Blair et al., 2006; Lin et al., 2009). The large surface area to volume ratio and stringent mass and temperature control allow much lower reagent consumption and less waste. Compared to the conventional in-column reaction, up to 100-fold decrease in reagent requirement can be achieved (Huang et al., 2007; Lee et al., 2010). Consecutive oligo-assembly and amplification is carried out on specialized chips that perform as miniature thermal cyclers (Huang et al., 2009). An integrated microfluidic reactor array may execute thousands of light-controlled oligo synthesis reactions at once (Srivannavit et al., 2009).

3. Assembly of genes and genomes

Complex genetic engineering tasks might require synthetic DNA pieces of lengths much beyond the capacity of simple oligonucleotide synthesis. Assembly of oligos into larger DNA segments has come far since the synthesis of the 77-bp gene encoding yeast alanine tRNA (Khorana et al., 1972). In vitro techniques, representing basically two approaches simplify and accelerate this process. First, ligation-based assembly uses a bacterial ligase to stitch together overlapping phosphorylated oligos that completely span both strands (Grundstrom et al., 1985). Thermostable ligases permitted automation of the annealing/ligation process in a cyclic fashion (ligase chain reaction, LCR, analogous to PCR) (Au et al., 1998). The technique is simple, has a low rate of mutagenesis, but is also relatively expensive and requires phosphorylation. The second approach, PCR-mediated assembly (Stemmer et al., 1995) reduces synthesis costs by leaving gaps between the adjacent (non-phosphorylated) oligos on each strand, which are then filled by overlap extension (Horton et al., 1989) using the complementary strand as template. The DNA polymerases available at that time introduced some errors by nucleotide misincorporation. Proofreading polymerases have reduced the frequency of this error type by over 50-fold (Meng et al., 2006). For longer or more complex genes, assembly is carried out in subsets, and the fragments are joined in subsequent reactions in a hierarchical fashion. Versions and combinations of the two approaches vary the length, ratio, orientation and overlap of the corresponding oligos, (reviewed recently by Hughes et al. (2011)). These methods have an error rate of 1–10 mutations/kbp, necessitating routine sequence verification (Hughes et al., 2011). The majority of these errors occur during oligo synthesis, resulting primarily in single nucleotide deletions and insertions (Tian et al., 2009). Additionally, the fidelity of microarray-based oligo synthesis is inferior to that of conventional techniques, further emphasizing the requirement of efficient error-correction methods (Kosuri et al., 2010).

Up to now, several strategies have been developed to reduce the error rate of gene assembly: oligo purification (Xiong et al., 2006), selection for the correct reading frame using downstream reporter genes (Cox et al., 2007; Gerth et al., 2004; Lutz et al., 2002; Seehaus et al., 1992), or applying mismatch-binding proteins (Bang and Church, 2008; Binkowski et al., 2005; Carr et al., 2004; Smith and Modrich, 1997). A cocktail of mismatch-cleaving enzymes is commercially available for this purpose (ErrASE™). Recently, high throughput pyrosequencing was used to find correct members within a library of oligos. Sequence-verified oligos were then extracted for downstream assembly, reducing error rates 500-fold (Matzas et al., 2010). In highly parallel gene synthesis, this technique can reduce costs by an order of a magnitude compared to