5

RECOMBINANT GENOMES: NOVEL RESOURCES FOR SYSTEMS BIOLOGY AND SYNTHETIC BIOLOGY

Mitsuhiro Itaya

Laboratory of Genome Designing Biology, Institute for Advanced Biosciences, Keio University, 403-1 Nipponkoku, Daihoji, Tsuruoka, Yamagata 997-0017, Japan

5.1 INTRODUCTION

Cellular life is elaborately managed by linear as well as branched biochemical pathways and all underlying information required for these systems are contained in the nucleotide sequence of the genome. Modern genomes possess tremendous amounts of information selected and accumulated during responses to altering natural environmental conditions. "Genome" nomenclature for proliferating species on earth is normally given to cells higher than bacteria as illustrated in Figure 5-1. The term "Genome" is also used to describe DNA possessed by bacteriophages, viruses, plasmids, mitochondria [1], and chloroplasts [2]. The last two are believed to have an ancient bacterial origin; with these two systems, additional essential informative molecules must be supplemented by the host. Given a genomic scope limited to unicellular bacteria, which are generally regarded as simple, diversity is observed in species variations manifested not only in taxonomic classifications but also in physical structure as illustrated in Figure 5-1. Due to the vast exploring technologies of cloning that emerged in the last quarter-century, our knowledge of genes and their products including RNA molecules and

Systems Biology and Synthetic Biology Edited by Pengcheng Fu and Sven Panke

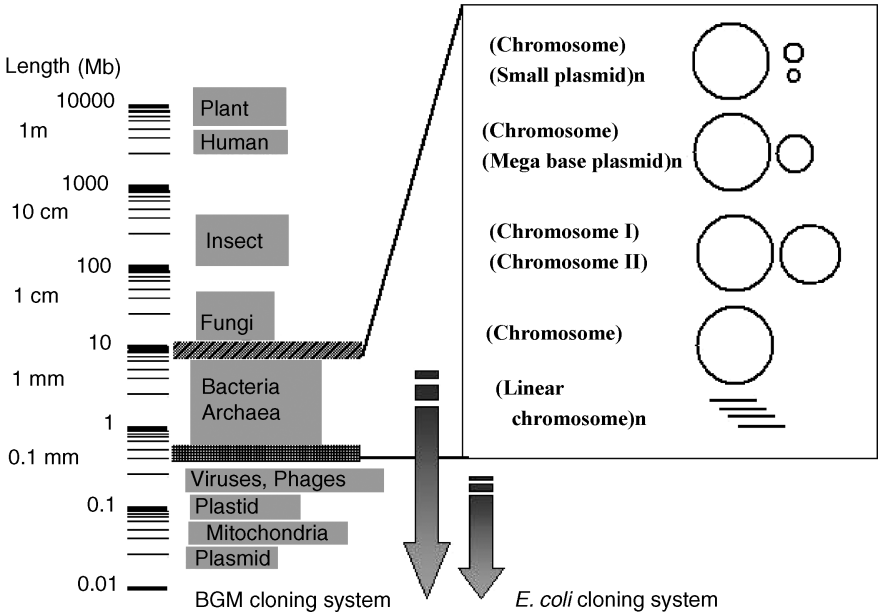


Figure 5-1 Genome variation—sizes and structure. Approximate genome size ranges are indicated for eukaryotes, bacteria, and archaea. Categorized structural types of bacterial genomes are illustrated in the insert. Sizes are in mega base pairs and the linearized lengths are scaled in the left. DNA size covered by *E. coli* plasmid vector (Section 5.8.2), and the BGM vectors (Section 5.4) are indicated.

associated metabolites has tremendously increased. Gene expressions are controlled by the complex and dynamic actions of thousands of factors as described in other chapters of this book. Perturbing small portions of gene-circuit frameworks are studied using genetically well-manipulatable organisms/cells [3].

Genome design, one of the topics highlighted in this chapter, requires at least two foundational roles: (1) Writers of novel nucleotide sequence at any length to be used as blueprint and (2) builders of high-molecular DNA who refer to the blueprint. Recombinant genome technology, not well documented elsewhere to date, has covered mostly the latter and will develop as the necessary tools for building any size DNA molecule. The former must be and is being conducted through current study of cellular life. Global attempts are described in some of other chapters where collaborative approaches among various disciplines are seen.

The DNA described in this chapter is larger in size and includes a greater number of genes than any other genetic entities handled in conventional recombinant gene technologies. Topics of this chapter will focus on emerging methodology in which complexity associated with an elevated number of genes as well as molecular constraints imposed by the increased DNA size must be treated at a time. The method utilizing *Bacillus subtilis* as a cloning host has been exploited independently from the conventional and familiar gene cloning technologies using *Escherichia coli* as a major cloning host (Fig. 5-2). Readers will be offered an overview of technical

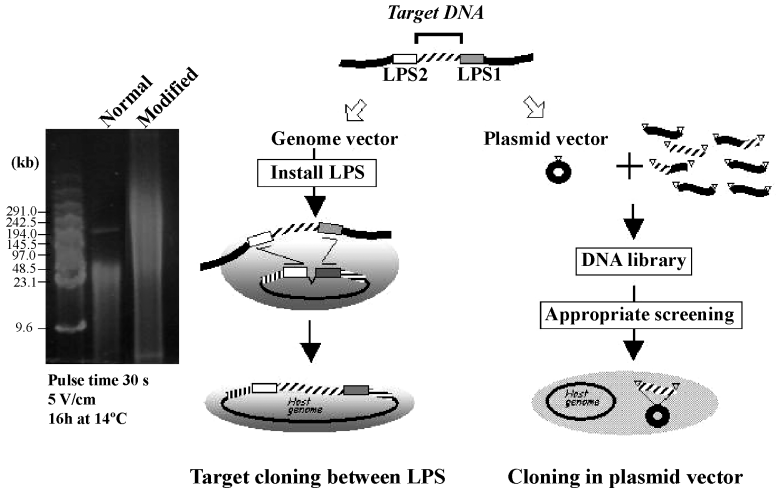


Figure 5-2 Cloning principle differs in *B. subtilis* genome vector: Essences for the *B. subtilis* genome vector (left) and the *E. coli* cloning vector (right) are in comparison. In the left path, designated/target DNA is finally guided in the host *B. subtilis* genome. Higher molecular weight DNA obtained by modified method are used for BGM cloning. *Source*: Referred to Ref. [17] and Section 5.4.2.

breakthroughs used to create recombinant genomes that would otherwise be considered impossible to obtain. The *B. subtilis* cloning system permits not only an increase in the clonable number of genes but also flexible postcloning modifications in DNA sequences, such as content, order, and orientations of genes. It will be assumed that *B. subtilis* host and its intrinsic genetic features are not familiar to most of the readers. Therefore, sections of this chapter will be attributed to basic explanation underlying this novel genome vector system. I must mention that the technological breakthroughs of several of the achievements described here are currently being employed for practical research and industrial use. Still others remain largely in nascent forms such as "Recombinant genome able to sustain life in diverse growth conditions." The intent is that the content in this chapter will not only address the genome vector protocols but also supply enough information to establish conceptual significance.

5.2 DNA (GENE) CLONING

All the primary information to sustain life is printed in the present genome DNA sequences of every organism, from bacterial to human cells, without exception [4]. According to DNA sequence determination technology with significant higher throughput [5], whole genome sequencing has extended to genomes even from nonclonal bacteria [6,7].

In parallel with the ability to determine DNA sequences with reduced cost and time, DNA cloning has been one of the most basic tools in biology to comprehend genes and gene functions. The most conventional cloning method has been developed using

E. coli as a cloning host and is summarized in Chapter 2. Tremendous accumulation of genetic and biochemical information on *E. coli* has supported versatile applications. This allows recombinant gene technologies to flourish and has made *E. coli* compatible with most research initiatives, including omics technologies introduced throughout this book. Manipulated recombinant genes that confer genetic variation, modulation, and perturbation of gene-circuit networks *in vivo* are also necessary tools for further development of systems biology. In addition, emerging *de novo* chemical synthesis technology may offer more opportunities in preparation of DNA pieces from scratch [8], yet the gene (DNA) cloning remains an inevitable step in most life science fields.

E. coli has the potential to harbor huge DNA molecules up to 350 kb in size assuming an appropriate vector choice, yet this size range remains below that of the smallest bacterial genome, 585 kb, of *Mycoplasma genitarium* [9,10]. When dealing with DNA fragments of this size, two technical skills become critically important, target cloning of the long DNA and flexibility in target sequence manipulation.

In cloning of DNAs particularly above dozens of kilo base pair, the size limit of PCR-mediated amplification method, preparation of nonsheared source DNAs is crucial. Our answer will be given in Sections 5.3–5.5 where repeated assembly of overlapping small segments leads to, via gradual elongation, final reconstruction of the target full-length DNA.

5.3 A GENOME VECTOR SUITED FOR RECOMBINANT GENOMES

Use of the 4215 kb *B. subtilis* 168 genome [11,12] as a stable cloning vector was first proposed in 1995 [13], and was supported by preceding works initiated early in 1990s [14–16]. The BGM, standing for *Bacillus G*eno*M*e vector and first coined in our related article [17], inherits a number of features absent in *E. coli* plasmid vector systems. After completion of this chapter one should be able to recognize many advantageous features of the BGM vector such as simplicity in daily handling, technical linkage to the conventional methods. Particularly, coverage for giant DNA segments and innovative potential for novel research in both fundamental and applied fields should be acknowledged.

The BGM cloning steps do not employ conventional enzymes such as restriction endonucleases and ligases that are vital in current DNA cloning methodology. On the contrary, homologous recombination plays a central role in this cloning system as well as in subsequent manipulations. As indicated in Figure 5-2, the conventional ligation step *in vitro* to directly connect DNA fragment to plasmid vector is replaced by inherent nature of the *B. subtilis* natural competence development/induced homologous recombination *in vivo* [18,19].

Readers may need some explanation as to why DNA goes into the genome of *B. subtilis* and not into the *E. coli* genome. The fundamental nature of this process is outlined in Figure 5-3. The key difference between the DNA uptake steps of both strains is clear: Only *B. subtilis* is able, under specific culture conditions, to develop a competent state where DNA outside the cell is actively incorporated in

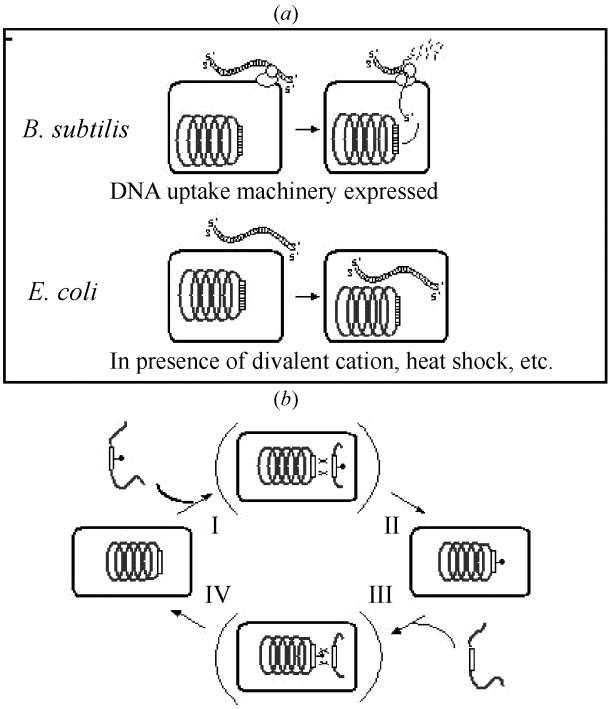


Figure 5-3 How *B. subtilis* incorporates DNA. (a) *B. subtilis* develops protein complex to actively incorporate DNA. Single-stranded DNA is the final substrate through uptake. These active mechanisms are not present in *E. coli*. (b) Insertion [I, II] and removal [III, IV] of DNA, shown by a needle with closed circle, via homologous recombination at the flunking regions of the *B. subtilis* genome.

cytoplasm [19,20]. *E. coli*, in sharp contrast, never actively transfers DNA through membranes and therefore import of DNA into cytoplasm must be induced via physicochemical treatment [21]. Another fundamental feature associated with competent *B. subtilis* is that double-stranded DNA (dsDNA) taken up by the protein complex on the cell's membrane surface undergoes processing so as to deliver a single-stranded DNA (ssDNA) molecule into the cytoplasm. The cleavage site by the nuclease and the strand selection by the transformation complex are basically random [19]. The resultant highly recombinogenic single-DNA strand promptly recombines with the counter homologous sequences if present in the genome (Fig. 5-3). The mechanism for the BGM vector to integrate/clone the target DNA is simple. As illustrated in Figures 5-2 and 5-3, two DNA sequences that sandwich the target DNA, generally termed as LPS, standing for landing pad sequence, must be present/installed in the BGM genome. Homologous recombinations between the LPS sequences of incoming DNA and of the genome result in concomitant integration of the internal target DNA region. Methods for preparing the embedded LPS and how a large target DNA integrates effectively are keys to understanding the precise cloning path of the BGM vector.

5.3.1 Domino Method: A Prototype

The first successful full-length target cloning reported was the complete *E. coli* prophage lambda DNA, one of the earliest genomes fully sequenced [22]. During the cloning of the 48.5 kb lambda *c*I857sam7 genome [13], most ideas and relevant experimental technologies that lead to what we would later call the domino method were established. The lambda DNA fragment alone never integrates in the *B. subtilis* genome because no homologous sequence is present. Also, lambda DNA possesses no inherent selection marker effective in *B. subtilis*. Therefore, appropriate LPS as well as general selection markers to navigate the lambda DNA segment had to be prepared.

The lambda DNAs were segmented into several pieces of DNA fragments ranging from 2.4 to 16.8 kb by either of the two restriction enzymes *Bam*HI or *Eco*RI. They are cloned in the *E. coli* pBR322 plasmid by conventional DNA cloning method of this host (Fig. 5-4). These pBR322-based clones are collectively called domino clones. All share two common structural features possession of sequence overlap with the adjacent domino clone and two identical regions of the pBR322 sequence. The two pBR halves, *amp* and *tet* half illustrated in Figure 5-4, play essential roles and are vital in BGM cloning at any stage. The same two pBR halves were integrated earlier in particular loci of the *B. subtilis* genome [15]. This genome integrated pBR form, termed as GpBR, was proven to accommodate DNA sandwiched by the two pBR halves used as LPS. The combination of the pBR part of domino clones and the GpBR assures that DNA integrated into the *B. subtilis* genome always remains flanked by *amp* and *tet* halves of GpBR. The first domino integration is illustrated in Figure 5-5.

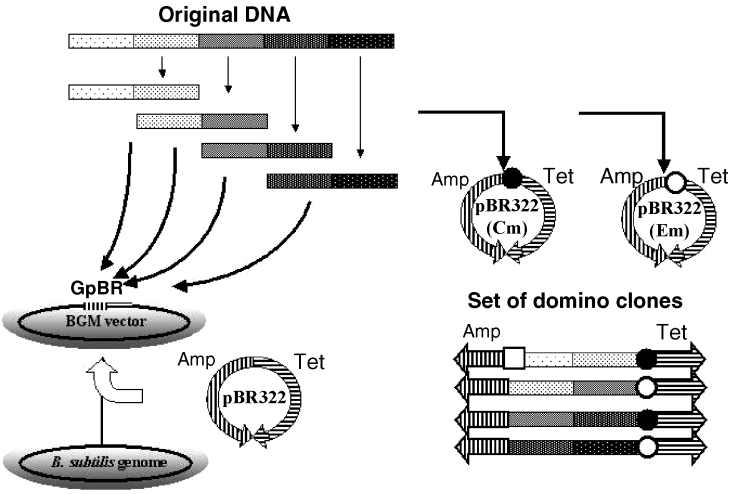


Figure 5-4 Domino method. Materials needed: Preparation of domino clones (right) and cloning locus in the BGM vector (left): Domino clones are shown in linear form. The selection markers, Cm (●) and Em (○), are set up alternately. pBR322-based domino clones all share common structural features. They are two halves of the pBR322 sequence, described as *amp* and *tet* and sequence overlap with the previous and next dominos. The two pBR halves play essential roles as illustrated in Figure 5-5.

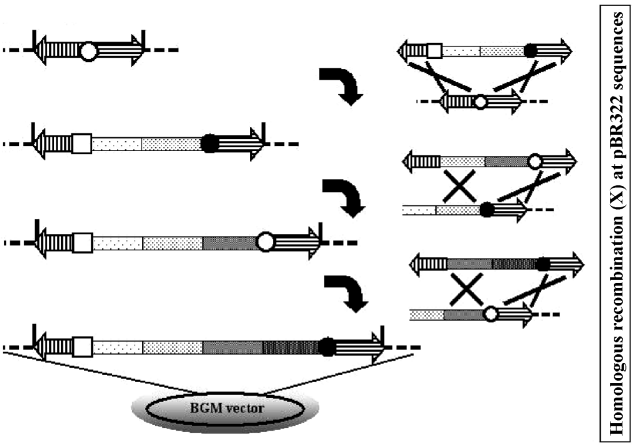


Figure 5-5 Domino method. Bottoming-up DNA cloning in the BGM vector: The first domino integrates via double homologous recombinations, indicated by X, at two pBR sequences of the domino and GpBR (shown top in the right). Integration of the second domino uses internal overlap region as one of the homologous recombination. This elongates the internal DNA and exchanges the marker for selection from Cm (●) to Em (○) (middle in the right). As elongation continues, two pBR halves always remain flunking the insert and exert reusable LPS (bottom in the right). The first domino only possesses another marker (□) to label the other end. I-*Ppo*I site indicated by short vertical bar is preinstalled and used to extract cloned DNA as described in Section 5.6.1.

If an adjacent domino clone comes in, either of the pBR halves then can serve as a reusable LPS. This automatically positions the other distal end sequence, not the other pBR half, as another LPS to result in elongation of the target region as indicated in Figure 5-5.

The full-length lambda genome DNA reconstructed using four domino clones drew little interest, probably because of unfamiliarity of the host *B. subtilis* to most *E. coli* users. Given that this success occurred immediately before several whole bacterial genome sequences become available, it may have been too early to arouse interest in the manipulation of large DNA molecules. Through the lambda-cloning experiment and the subsequent trials and errors in our laboratory, related technical problems have been refined making the domino method applicable for larger sequenced genomes.

5.3.2 Domino Method: Applications to Organelle Genomes

The domino method simply requires a full set of domino clones that cover the entire target genome. Any gaps due to lack of an available domino clone should be avoided. This problem was resolved when a PCR methodology is combined with the preparation of dominos from sequenced genomes.

Innovative applications of the domino method have been clearly proven by the challenges to obtain chloroplast and mitochondria genomes whose complete sequences are known [4]. Chloroplast genomes (cpGenome) were chosen as

candidates by (1) their sizes ranging between 100 and 200 kb [2], (2) circular form, and (3) no complete cloning report had been made. Due to their larger size, it is difficult to prepare intact unsheared circular chloroplast DNA suitable for one-step cloning. Instead, sheared shortened chloroplast DNA during biochemical isolation serves as template DNA sufficient for PCR reaction.

The set of domino clones that cover the entire sequence of cpGenome from rice (134.5 kb) were designed. A total of 31 domino clones for the rice cpGenome, 6 kb on average and 1 kb of which serve as LPS with the adjacent domino were prepared in *E. coli* via PCR amplification. Effectiveness of the method in the BGM vector was fully demonstrated for complete cloning of the rice cpGenome [23]. The recombinant rice cpGenome was stably maintained regardless of the two identical 21 kb-long inverted repeat (IR) sequences that are characteristic of higher plant cpGenome. The rice cpGenome example strongly indicates that the domino method permits any known DNA sequence up to and probably above 150 kb, to be reconstructed in the BGM vector.

Mitochondria genome (mtGenome), from another organelle present in nearly all eukaryotic cells, was similarly approached. The size range of mtGenomes exhibit great diversity in a species-dependent manner [1], compared with the limited size range of the cpGenome. Thus, our attempts to date have been limited to the well-studied mouse mtGenome. Mouse mtGenome (16.3 kb) is far smaller than other mitochondria species [1] and was sequenced at approximately the same time as the lambda genome. Interestingly, no report of its full-length cloning was found until recently, in spite of the small size, as small as one domino clone (16.8 kb) prepared for the lambda genome cloning [13]. The only previous success of mouse mtGenome cloning in pBR322 based plasmid appeared to be fully dependent on careful preparation of intact mtGenome from the mouse cell and a fortuitous insertion site selected by the transposon vectors used [24]. However, this pioneering work demonstrated that cloning of mouse mtGenome is not problematic. Cloning of the mouse mtGenome by our domino method, separated into four domino clones and incrementally reconstructed in the BGM vector, also had few problems [23].

These complete cpGenome and mtGenome stably cloned in the BGM vector were converted to a circular DNA form via a unique positional cloning method described and briefly mentioned in Section 5.6.3. The circular form of mtGenome was optionally propagated in an *E. coli* host using the shuttling nature of the plasmid, showing consistent results with those who had performed similar work [24].

5.3.3 Domino Method: General Application to Gene Assembly

The primary requirement for a domino clone is possession of overlapping sequence with the adjacent DNA. This requirement does not exclude the domino elongation step with an unlinked DNA segment; so long as an adjacent domino clone possesses the LPS portion as shown in Figure 5-4. In other words, target reconstructed DNA is not limited to continuous DNA segments of the present genome. In extreme cases, all the domino clones composed of two DNA blocks might be made from designed sequence.

The domino elongation as illustrated in Figure 5-5 does not exclude such designed DNA blocks, and consequently the final DNA is reconstructed according to the designed blueprint.

The designed assembly of DNA blocks was demonstrated for genes involved in *de novo* pigment synthesis. Eight cDNAs were prepared from *Arabidopsis thaliana* coding enzymes that catalyze a series of biochemical reactions from tryptophan to anthocyanidine, a violet-colored pigment made and stored by certain plants [25]. In our first assembly design, the order of these eight genes, dispersed in five different chromosomal loci, was as the same order of biochemical reactions as illustrated in Figure 5-6a. This primitive operon-like construct was built by progressive integrational elongation using eight gene blocks as domino clones. Similarly, the general domino method was applied to biosynthetic genes for another pigment, carotenoids, that is synthesized and stored in orange colored plants such as carrots as well as certain bacteria [26]. In a reassemble step of the genes included in the natural construct plasmid pACCAR25(Δ*crtX*) from *Erwinia uredovora* [27] presented in Figure 5-6b, the expected intermediate substance lycopene was produced in *B. subtilis* by the assembly/insertion of the first three genes in the biochemical reaction [28]. Thus, the domino method has several examples of general applications and offers rational design in DNA assembly protocol for systems biology as well as synthetic biology. The domino method will be evaluated by comparing it with other assembly methods also exploited in our groups described in Section 5.5.4.

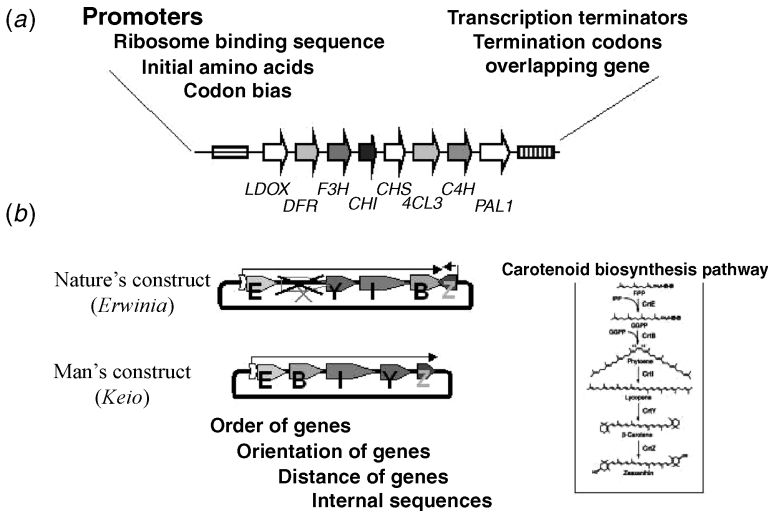


Figure 5-6 Options in gene assembly. (a) Listed are factors considered in formation of anthocyanine biosynthesis cassette by the domino method in Section 5.3.3. Location of eight genes in *Arabidopsis* chromosome is indicated in Figure 5-18. (b) Carotenoid cassette (Keio form) was constructed by the domino method [61] in Section 5.5.4. Carotenoid biosynthesis by the OGAB method in Section 5.5.4.

5.4 DIRECT TARGET CLONING: PROTOTYPE

The domino method clarified two points, (1) any DNA segments can go into the *B. subtilis* genome via the intrinsic homologous recombination system utilizing preinstalled LPS sequences, and (2) the integrated/cloned DNA segment exhibits high structural stability. The latter is accounted for by precise replication as part of the *B. subtilis* genome and subsequent accurate segregation during *B. subtilis* cell division. Sequence indiscrimination during all the integration processes of *B. subtilis* guarantees the former; which is in contrast to sequence-dependent incorporation as seen for other Gram-positive strains [20,29]. Given a quantitative evaluation for nucleotide sequence fidelity of the reconstructed mtGenome and cpGenome [23], we had come to a putative conclusion that the *B. subtilis* genome has potential to harbor significantly large DNA repertoire with great fidelity in nucleotide sequence. The primary concern on LPS-mediated integration/cloning protocol was how to select foreign DNA possessing no selection markers for *B. subtilis*. Besides the progressive mode in the domino method and an exceptional gap-sealing protocol by unmarked DNA segment practiced in the lambda DNA cloning [13], the generalized selection marker scheme was a prerequisite.

5.4.1 Counter Selection Markers for Cloning Unmarked DNA Segments

Counter selection marker makes direct and rapid isolation of the correct integrant more promising. A neomycin resistance gene has been developed as a counter selection method [30]. A protein coding sequence of the neomycin resistance gene regulated under the Pr promoter (Pr-*neo*) was constructed in *E. coli* and integrated in the BGM vector at unlinked locus from the GpBR as illustrated in Figure 5-7. The Pr-*neo* confers *B. subtilis* neomycin resistance due to full expression of the *neo* gene product. Meanwhile, a CI gene product encoded by the *c*I gene of *E. coli* bacteriophage lambda binds to the Pr promoter sequence and shuts off the promoter activity. The *c*I gene, if present and constitutively expressed in the BGM vector, renders the *B. subtilis* sensitive to neomycin. Absence of the *c*I gene restores the Pr promoter activity and makes the strain resistant to neomycin and vise versa. This small transcriptional gene-circuit worked nicely as a counter selection system known as *c*I-Pr [30, 31] and has proven useful for cloning any DNA lacking an appropriate selection marker in the BGM vector [32–35]. More importantly, the reusable *c*I-Pr system allows repeated integration of DNA segments in the same BGM vector.

5.4.2 Quality Required as Donor DNA

Target DNA for positional cloning has to be as intact as possible [33, 35]. DNA in solution is normally fragmented into small pieces, typically dozens of kilo base pairs on average, caused by physical shearing during isolation step from cells and organelles. The relatively large DNA prepared in agarose gel matrix plug provides a more intact form with minimal breakage [12]. However, DNA inside the gel matrix is

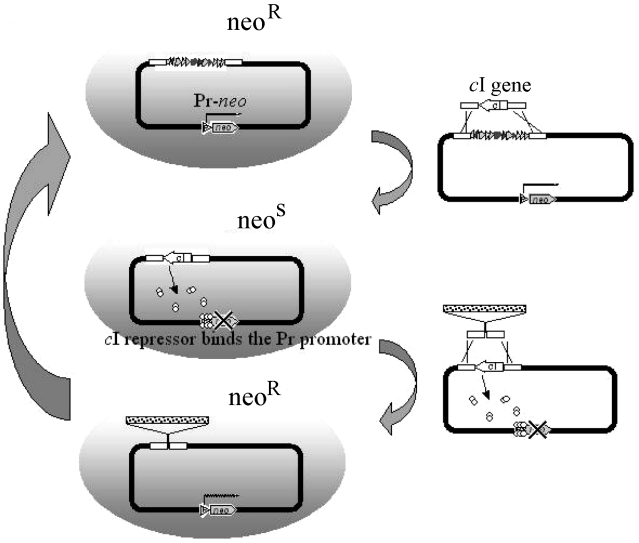


Figure 5-7 Reusable counter selection marker by a *c*I-Pr system. Presence (middle left) or absence (top left) of a repressor gene (*c*I) suppresses or induces antibiotic resistance gene (*neo*) under the Pr promoter. Replacement of the *c*I gene by foreign DNA confers resistance to neomycin (bottom left) and permits another *c*I-Pr selection. Details are described in Section 5.4.1 and [30, 31, 36].

not a good substrate for *B. subtilis* transformation. Therefore, as special protocol and gentle handling are required, we modified some DNA isolation protocols for BGM cloning. For example, a modified genome isolation method improved to yield high-quality DNA applied to the cyanobacterium *Synechocystis* PCC6803, produced DNA fragments on average of a couple of 100 kb as indicated in Figure 5-2. This quality DNA supplied sufficient length for target cloning of 30–70 kb segments [33]. Thus, the target size has to be determined by quality of prepared DNA in BGM cloning.

Ironically in contrast to the above size stipulations, the presence of huge untargeted DNA may result in adverse effects. For example, targeting 50 kb DNA of the 3500 kb *Synechocystis* genome sacrifices efficiency of cloning due to the inhibitory action by the remaining untargeted portion of the cyanobacterial genome present in the DNA preparation. This shortfall is accounted for the limited number of competent complexes formed on the competent cell surface; approximately 50 that take up and guide the DNA inside the cell [19]. Competitive inhibition by irrelevant DNA during *B. subtilis* transformation has been frequently observed, always resulting in marked reduction of the number of correct BGM recombinants [34], albeit is not entirely detrimental to the present BGM protocols.

5.4.3 Repeated Target Positional Cloning

As one can expect, no sooner was the success of target positional cloning, than target cloning of another adjacent segment results in elongation of the target region in the

BGM vector. One should recall that the *c*I-Pr dependent positional target cloning system is reusable and cloned DNA in the BGM always accompanies two halves of the pBR sequences at both ends. Furthermore, the second target cloning is performed so as to adjoin to the first BGM, as logically extended, and should permit additional positional target cloning by sliding to an adjacent region of the given BGM [36].

The LPS in target positional cloning may function as primer sequences in the PCR amplification method. If one views the two LPS sequences and intervening target DNA, analogous to the head, tail, and body of an inchworm, donor DNA acts as an inchworm walking via integration as shown in Figure 5-8. The inchworm lands first in the tail and head because only head and tail are present between the GpBR. The body part finally lands in the BGM at the completion of the cloning. Addition of a new set of LPS, new "head" aligned with the new "tail" being converted from the "head" in the previous inchworm. This sliding alteration of the two LPS guides the adjacent secondary target DNA so as to be positioned without any gaps. The DNA segment in the BGM vector is then elongated leaving the third inchworm available. Repeated application by renewing "head" and "tail" sets alternately results in progressive elongation of the target DNA until encountering certain constraints inherent in the BGM vector (see Section 5.7.2).

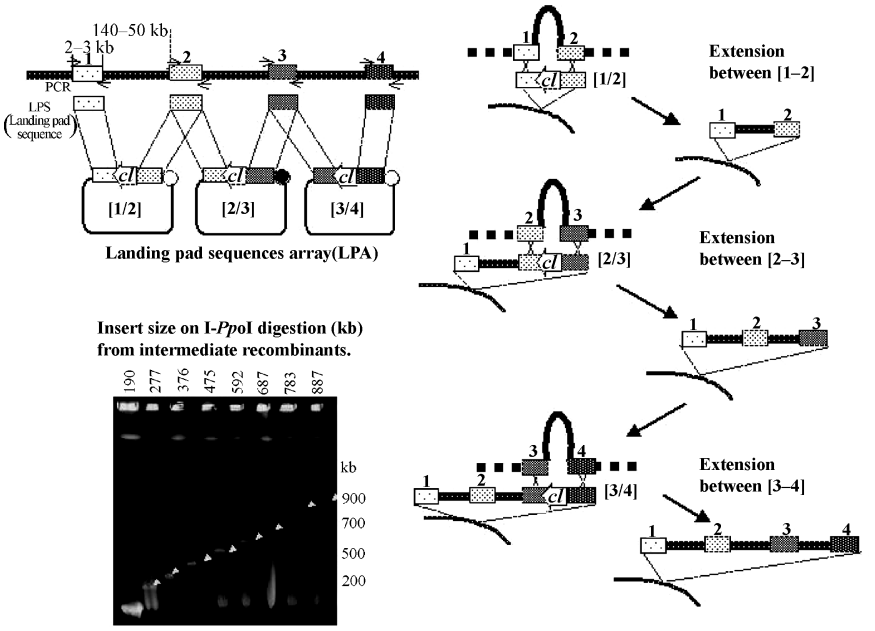


Figure 5-8 Inchworm like elongation in megacloning of the *Synechocystis* genome. Sets of two LPS (LPS array) are prepared separately. Internal sequence between LPS is integrated by the mechanism shown in the right. One cycle of inchworm starts by installation of LPS array and ends up with the incorporation of *Synechocystis* DNA. Details are provided in Sections 5.4.3 and 5.4.4. Discontinuous elongation of the *Synechocystis* genome region 2 (Fig. 5-9 is viewed by I-*Ppo*I fragment size increase. I-*Ppo*I site resides at both the ends of the GpBR in all BGM recombinants (Fig. 5-5).

5.4.4 Megacloning of the *Synechocystis* PCC6803 Genome

The method to elongate a continuous DNA by many inchworms in the BGM vector is termed as megacloning [36], invoking that the cloned DNA segment is in the mega base pair (Mb) size range. The main goal was to use the *B. subtilis* genome as a platform to clone and manipulate DNAs above 500 kb, seemingly the upper clonable limit with *E. coli* plasmids. The whole *Synechocystis* genome cloning stages are summarized in Figure 5-9. The work started in 1997, just 1 year after the whole genome sequence data were published [37]. The project evolved over the past years during which most works have been spent on refining all protocols. First, it should be mentioned as to why *Synechocystis* PCC6803 [37] was chosen as the target for megacloning. There are several reasons noted at that time: The preparation of LPS by PCR-mediated amplification requires sequence information on the whole genome. The high molecular weight genomic DNA prepared as shown in Figure 5-2 [33] was necessary for technical reasons described above. More importantly, the possible expression of cloned genes in *B. subtilis* hazardous for BGM user had to be avoided [38]. Therefore, the sequenced 3573 kb genome of the unicellular photosynthetic bacterium *Synechocystis*, thought to be nonpathogenic, was indeed the only available choice when this work started in 1997. The alleged goal ended by megacloning the whole genome of a life form, the 3.5 Mb of *Synechocystis*. Details are referred to in the recent publication [36], and results are illustrated in Figures 5-9 and 5-10.

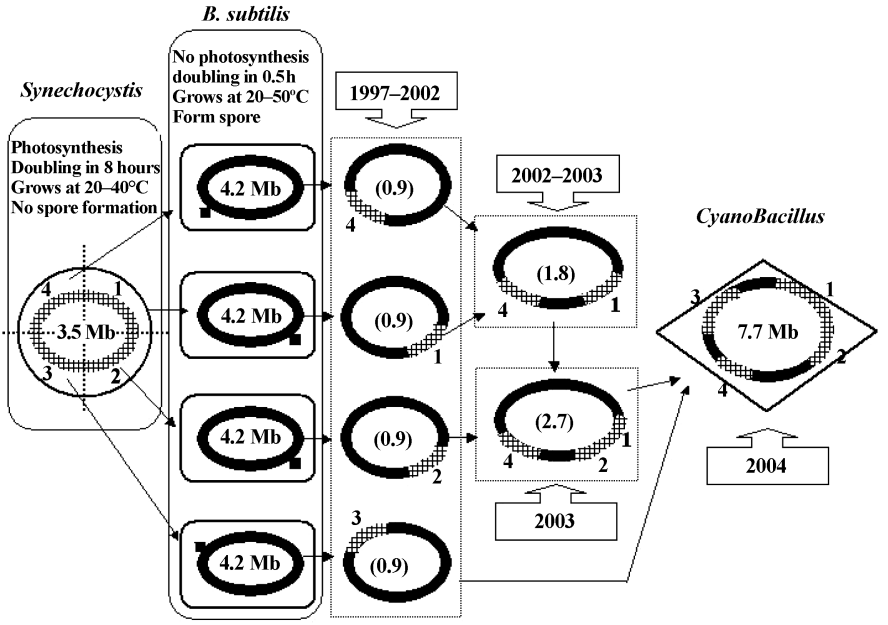


Figure 5-9 Overview of megacloning of the whole *Synechocystis* genome. *Synechocystis* genome was putatively divided in four sectors. Separately megacloned sectors in four BGM vectors were sequentially assembled in the subsequent process. This less straightforward way of the whole cloning was obligatory due to inherent structural constraints of the BGM vector (or the *B. subtilis* genome) as indicated in Section 5.7.2 and Figure 5-23.

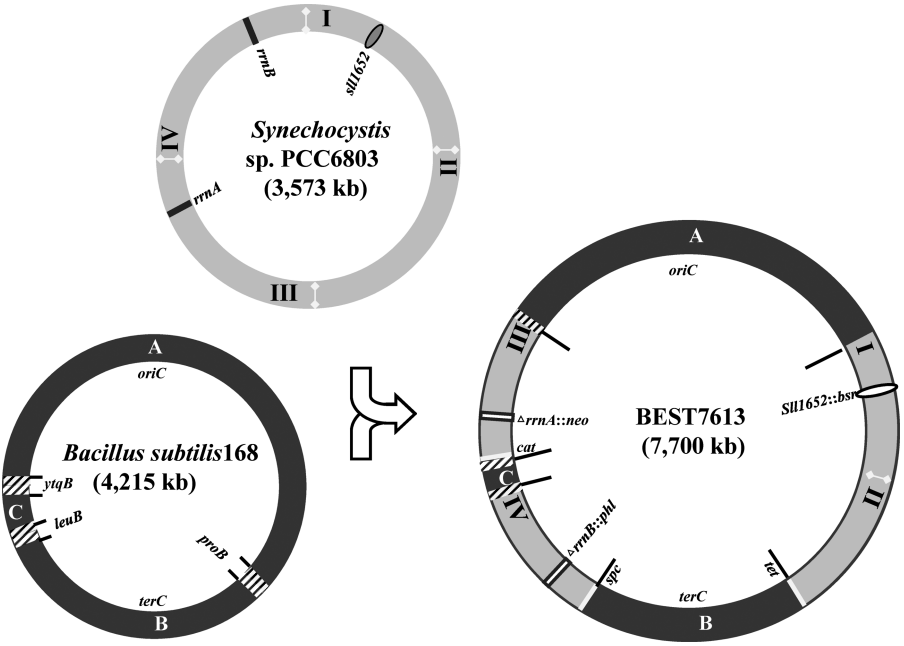


Figure 5-10 Genome anatomy of the *CyanoBacillus*. A chimera (right) of the two genomes from *Synechocystis* (top left) and *B. subtilis* (bottom left) is shown. Certain questions raised by this unprecedented organism are introduced in Section 5.4.6 and Figure 5-12.

5.4.5 Novel Method for Gene Function Analyses

During inchworm walking, various intermediate recombinants carrying different number of *Synechocystis* genes were stably obtained. This characteristic method portrays a novel technique for the investigation of gene function as well as genome function apart from other conventional DNA cloning methods. Two examples noting the discovery of adversely influential genes are worth mentioning here: *sll*1652 and *rrnA* (also equivalent *rrnB*). The gene *sll*1652, whose protein function remains unknown, apparently interfered with the sporulation process of *B. subtilis*. The sporeless phenotype was first found after a particular inchworm walk, as shown in Figure 5-11. The single gene was logically identified by subsequent deletion analysis in the BGM vector. Although the *sll*1652 gene could be found in the conventional *Synechocystis* DNA library made in *B. subtilis*, obvious phenotypic change before and after the presence of a defined region clearly specified the culprit gene. One more functional gene has been similarly suspected in a different region (Itaya M. and Fujita K. unpublished observations). In contrast to the first example, the unsuspected role for *rrnA*/*rrnB* was discovered in reverse manner. Inclusion of *Synechocystis* ribosomal RNA (5S-23S-16S) encoded in the *rrnA* or *rrnB* operon resulted in large deletions from other previously megacloned regions present.

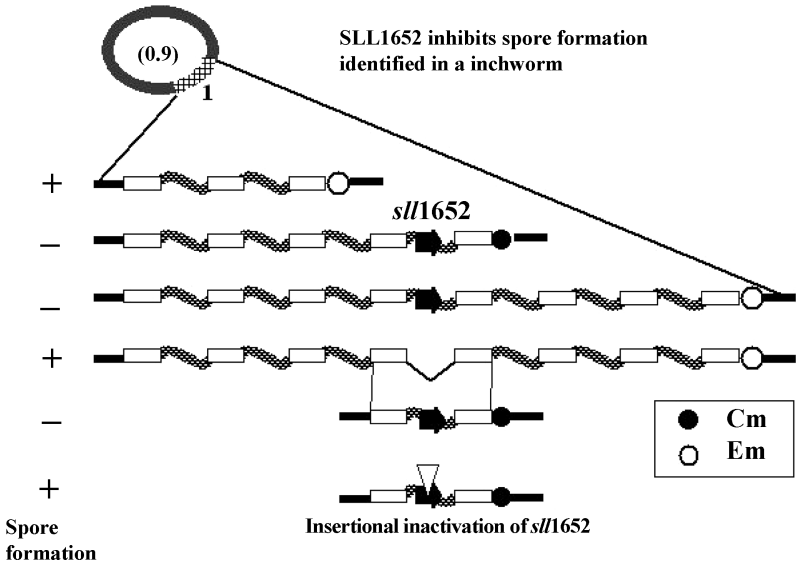


Figure 5-11 Comprehensible discovery of a gene from *Synechocystis*. An apparent sporeless phenotype rooted by a particular inchworm was scrutinized. Genomic manipulation of megacloned recombinants attributed the culprit gene to the *sll*1652 within 40 genes in the inchworm in Section 5.4.5. The sporeless feature was finally confirmed being caused by the gene only.

5.4.6 Questions Raised by Two Genomes in One Cell

Among the many questions arising from the *Bacillus–Synechocystis* chimera, putatively named as *CyanoBacillus*, the potential involvement of ribosomes is currently being intensively investigated. Lack of *Synechocystis* ribosomal RNA is consistent with our recent molecular data that a number of *Synechocystis*-originated transcripts but little translated products in the *CyanoBacillus*. These observations imply that ribosome and associated ribosomal RNA are key switching factors that determine dominant cellular gene networks as illustrated in Figure 5-12. If tremendous amounts of genes, including ribosomal RNA and protein genes are delivered by horizontal gene transfer (HGT), most of them are dormant in one environment but may be retained and utilized as functional and necessary genes when placed under different conditions. The *a priori* consensus for current molecular phylogenic analyses [39] is consistent with the difficulty of finding two natural *rrn* carriers to date [39–42]. The unusual nature of two ribosomes in one cell remains speculative and controversial. Further investigation of *CyanoBacillus* under omics analyses or creation of a second megacloning example should be conducted to access these issues.

One intriguing aspect on chimera structure of *CyanoBacillus* may be related to a proposal that genome fusions and horizontal gene transfer could be deduced from reconstruction of the phylogenetic tree of life. The hypothetical origin of the eukaryotic cell, albeit enigmatic and complex, is that it is the result of a fusion between two diverse prokaryotic genomes [43]. One fusion partner branches from

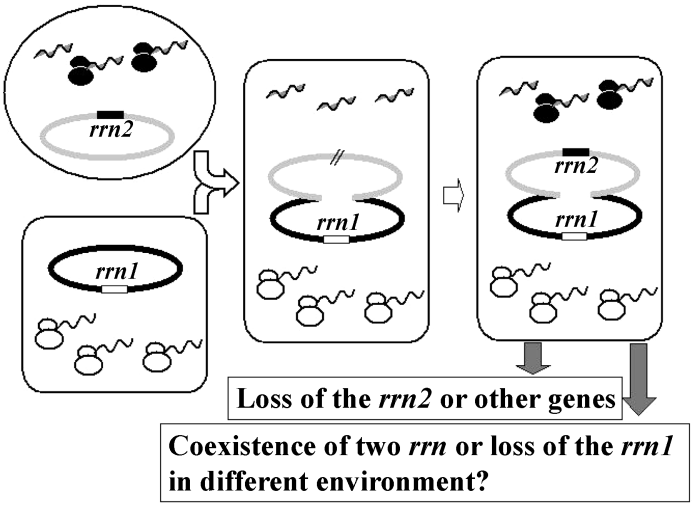


Figure 5-12 Ribosome switches genetic circuit by translational regulation? *Synechocystis* genome possessing *rrn2* (top left) and *B. subtilis* genome *rrn1* (bottom left) do not coexist in *CyanoBacillus* (center) as described in Section 5.4.6 [36]. Many numbers of transcripts from *Synechocystis* genes are observed in *CyanoBacillus* (our unpublished observation). Capturing *rrn2* might be suitable for survival in a different growth environment. The scenario employed has been limited only to complete genome fusion.

deep within an ancient photosynthetic clade, and the other is related to the archaeal prokaryotes. The eubacterial organism is either a proteobacterium, or a member of a large photosynthetic clade that includes the cyanobacteria and the Proteobacteria. This scenario may be investigated by megacloning of two candidate genomes in the BGM vector (see also Section 5.7.3).

5.5 ASSEMBLY OF GENES IN ONE DNA SEGMENT

Location of individual genes in the genome appears not to be determined by a defined set of rules [44–47]. With respect to a set of corelevant genes, typically regarded as operon, the conserved operons in the presently known bacterial genomes are very limited to certain ribosomal proteins and some metabolic pathways. Under the "selfish operon" hypothesis [48], operons are viewed as mobile genetic entities that are constantly disseminated via horizontal gene transfer.

The operon formation rule and the degree of HGT contribution remain controversial [49,50]. Apart from the enigmatic evolutional view for the present-day operons, we can technically modify the present-day gene order by the domino method as briefly mentioned above. Assembly of a number of functionally related genes may be the beginning of a drive to contemplate the ultimate man-made genome with all necessary genes assembled that function as the blueprint for engineered cellular life.

5.5.1 Why Should Genes be Assembled?

Biological processes are series of enzyme-catalyzed biochemical reactions. They include uptake or secretion of materials through the cell surface, production, or degradation of energy-coupled catabolites and metabolites in the cell, construction of cellular membranes/cell walls that are all responsible to sustain life. Biological reactions possess high potential as alternatives to traditional chemical processes for producing valuable molecules. Pioneering attempts have been made to produce materials by introducing a series of relevant genes to carry out their biological process in various hosts. To date, reports are concentrated in the two metabolically and genetically well-understood hosts: *E. coli* and *Saccharomyces cerevisiae* [51–57]. These hosts are eligible for repetitive transformation, circumventing problems associated with increased number of genes required for the biological reaction of interest. Indeed, if all the genes separately prepared from original genomes are included in a single-DNA segment in a row, delivery of such a biological reaction unit to another host becomes more popular and efficient [58]. The DNA manipulation system to construct such recombinant DNA, known as recombinogenic engineering, is limited and significantly dependent on manipulation using *E. coli* [59]. Ordering relevant genes into an all-in-one segment still requires laborious, time-consuming work. This appears to be a bottleneck in outlining a comprehensive blueprint for operon design.

5.5.2 Efficient Assembly of Genes in one DNA Segment using *B. subtilis*

We exploited a method to assemble a number of genes in one DNA segment with very few experimental steps, referred to as an ordered gene assembly in *B. subtilis* (OGAB) [60]. The method, as illustrated in Figure 5-13, stems from the unique *B. subtilis* DNA uptake characteristics. In short, multimeric forms of DNA are favorable substrates to be obtained as plasmid through a unique *B. subtilis* transformation. This consequence is due to the mode of DNA incorporation by competent *B. subtilis*. One may recall that ssDNA is taken up through the transformation apparatus [19] (Fig. 5-3). Only longer than one plasmid unit length results in replication of the full-length complementary strand, and therefore ssDNA of monomer unit length can never be converted to circular dsDNA plasmid. The DNA possessing tandemly repeated unit length is a good substrate for plasmid establishment via *B. subtilis* transformation.

5.5.3 Assembly of Various Numbers of DNA Segments in a Plasmid

How can tandemly repeated unit-length DNA (truDNA) be prepared, and how can multiple DNA fragments be assembled by BGM vector? Our solution was derived by making use of staggered dsDNA ends and modified conditions for T4-DNA ligase. In brief, all the component DNA segments have to possess protruding sequences that specifically connect only once to a singular complimentary end of another segment. As shown in Figure 5-14, for example, the variable three-base sequence within an

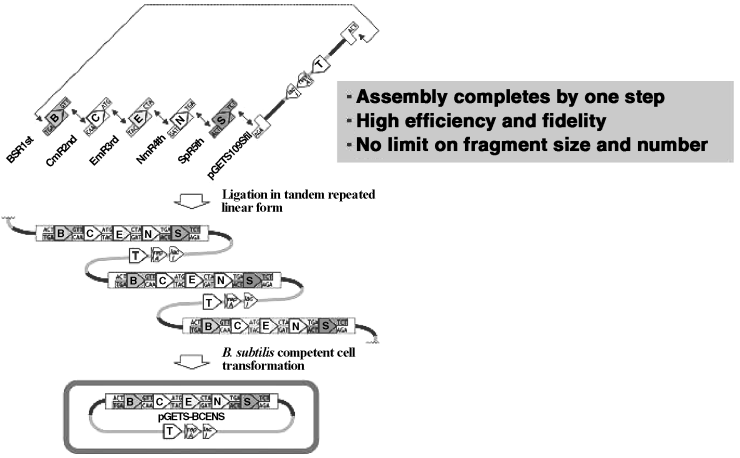


Figure 5-13 The DNA block assembly. Principle of one-step assembly of multimeric DNA fragments in *B. subtilis* plasmid is shown. The six-gene alignment comes from the result indicated in Figure 5-16. Details are stated in Section 5.5.2.

endonuclease *Sfi*I recognition sequence GGCCNNNN/NGGCC, indicated also in Figure 5-14, is suited for this aim. Linear multimer ligation products are formed preferentially in the presence of polyethylene glycol (PEG6000) and high salt concentration (sodium chloride, 150 mM). As shown in Figure 5-15, this condition dramatically suppresses formation of circular products and directs remarkable truDNA

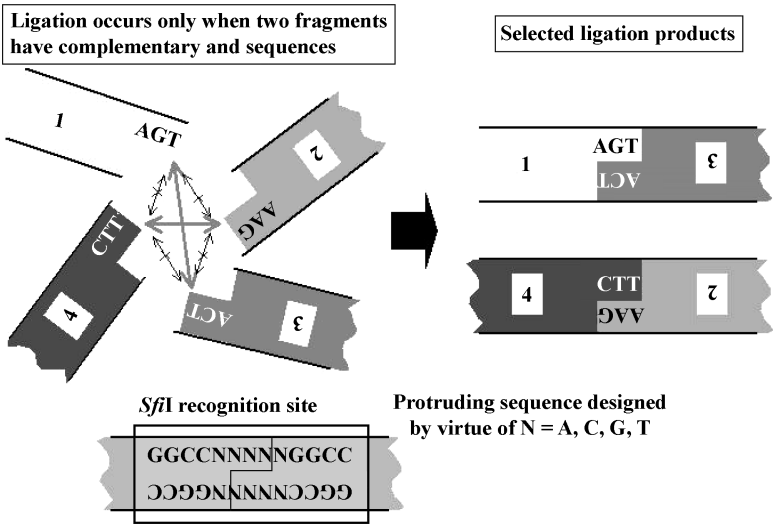


Figure 5-14 Selective ligation determines alignment of DNA blocks. Multiple protruding ends are created by restriction endonuclease such as *Sfi*I including variable nucleotide sequence within its recognition sequence. Flexible design of order and orientation of six gene blocks is possible.

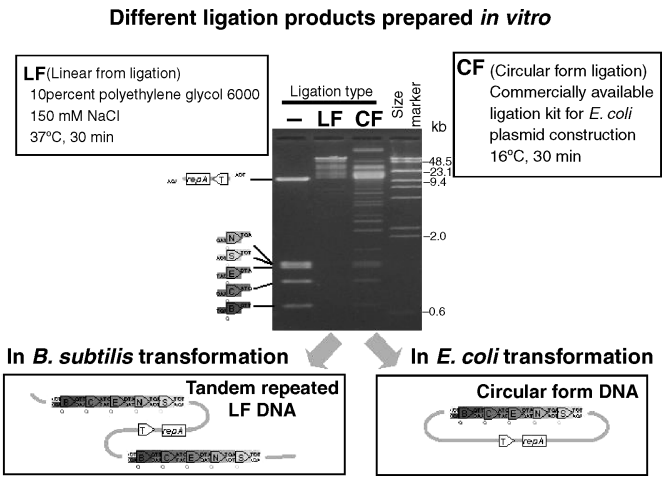


Figure 5-15 Conditions suitable for OGAB substrate on ligation. Multiple DNA blocks in linear form (LF) are preferentially produced in the presence of high molecular weight polymer and at a higher salt concentration [60]. Equivalent mole ratio for all the gene blocks is critical to form LF [60]. circular form (CF) DNA mostly formed in the regular ligation conditions appear as mixtures not suitable even for *E. coli* as indicated in Figure 5-16.

production. It should be emphasized that adjustment to equal moles of all the DNA segments is vital for ligation efficiency of truDNA. Assembly of up to six antibiotic resistance genes, comprising the multiple antibiotic resistance gene cassettes in an 18 kb plasmid, exhibited surprisingly high efficiency and fidelity compared with no equivalent plasmid obtained from *E. coli* transformation as indicated in Figure 5-16.

5.5.4 Application to Assemble Functionally Related Genes

In principle, the OGAB-mediated gene assembly has no limit on the number of included genes and final DNA size. In addition, order and orientation of these DNA fragments can be easily altered by sequence design of the protruding ends. Two recent achievements leading to what we believe is an initial framework for "operon designing" are described here. Two selected metabolic pathways, a pigment biosynthesis and a fungicidal substance production, require five and seven enzyme-coding genes, respectively. The pigment, a carotenoid, and the genes required for its production are basically the same as those used in the domino method (Fig. 5-6b). pACCAR25(Δ*crtX*), a natural version, containing six genes under a unique promoter [27], produces the pigment in *E. coli*, albeit the order of genes is not aligned with that of biochemical reactions *in vivo*. Five out of the six genes that catalyze the five biochemical reactions from FPP to give zeaxanthin were prepared as DNA pieces and aligned in parallel with that of the biochemical reactions in a single-DNA segment by OGAB method. Due to the OGAB plasmid that carries dual replication origins for *B. subtilis* and *E. coli*, the recombinant construct when expressed in *E. coli* resulted in

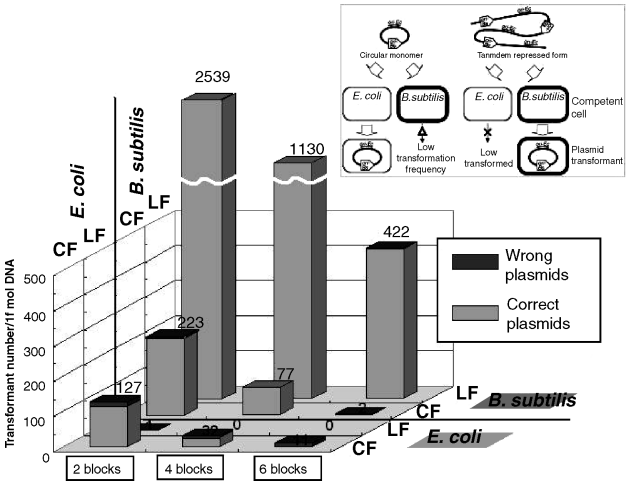


Figure 5-16 Powerful gene assembly by OGAB method. LF preferentially yields correct OGAB plasmid in *B. subtilis*, whereas CF dose not produce the expected ones in *E. coli* as the number of gene blocks increases. Simplified schematic view is shown in the insert.

surprisingly higher zeaxanthin than the original natural construct [61]. This simple and clear result showed that the expression level of a gene-circuit readily varies with the gene order and orientation and raises questions on how to approach the design of such operons, unless the natural operon is available.

We are learning lessons in approaching another biological process where extremely large proteins play roles to synthesize plipastatin, a lipopeptide fungicidal substance whose peptide portion is a nonribosomal peptide (NRP) produced by *B. subtilis* shown in Figure 5-17 [62]. The peptide portion of plipastatin is synthesized by five NRP synthases. These five large enzymes, 289, 290, 287, 407, and 145 kDa, are encoded in 38 kb-long genomic *ppsABCDE* operon in Figure 5-17. Difficulties in manipulating the huge operon and, more importantly, the resultant large transcript are expected to allow for novel insight into gene alignment and associated mRNA design.

For plipastatin production in *B. subtilis,* at least two more genes, *sfp* and *degQ,* are required in addition to the five genes in the *ppsABCDE* operon [62]. The *sfp* (0.9 kb) gene encodes 4′-phosphopantetheinyl transferase that catalyses transfer of 4′-phosphopantetheinyl to apo-peptide synthetases to convert the peptide to the holoenzyme form. The *degQ* (0.6 kb) gene that encodes a polypeptide composed of 42 amino acids shown to be a possible regulator for the *ppsABCDE* operon expression. We used the term "gene block" to define that the DNA fragments can be significantly variable both in number of genes present and in size. It was demonstrated that the OGAB method permits efficient assembly of three gene blocks ranging from 0.6 to 38 kb as predicted [62]. The two examples of zeaxanthin [61] and plipastatin [62] encourage us to enrich our disciplines toward rational operon design that was previously considered quite difficult.

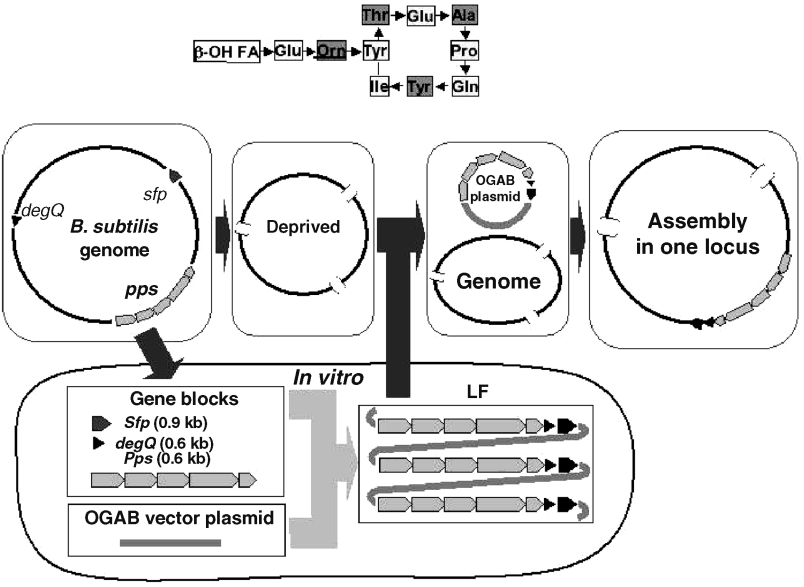


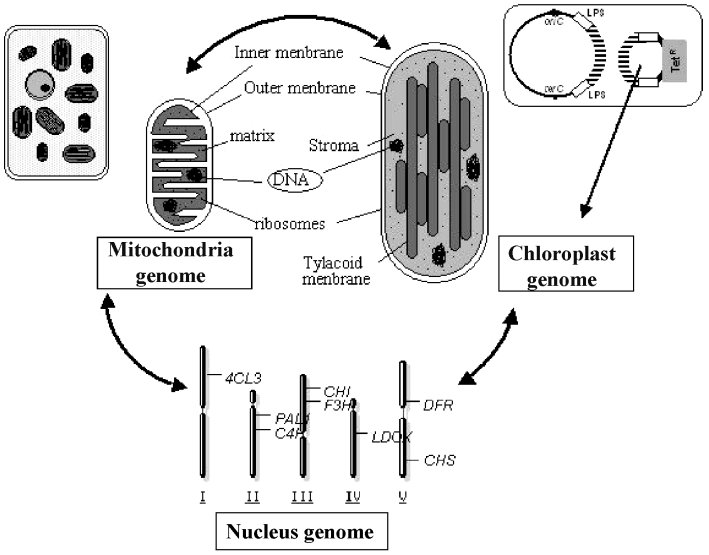
Figure 5-17 A design of operon for plipastatin bioprocess. Molecular structure of plipastatin, an antifungal agent, is shown on top [61]. Combined structure of lipid and peptides containing amino acid with D-configuration and ornithin is synthesized from components by three relevant gene blocks [62]. *B. subtilis* plipastatine gene originally dispersed in three locations are assembled by OGAB and function as stable operon in the genome.

5.5.5 Integration of OGAB Construct in the BGM Vector

The OGAB method yields assembled DNA segment in plasmid form due to its formation mechanism in *B. subtilis* (see the insert in Figure 5-16). The plasmid normally possesses a specific sequence used for initiation of DNA replication (*ori*) independent from the chromosomal counterpart (*oriC*). This physically independent character in replication as well as segregation during cell division ensures properties advantageous for gene cloning; for example, rapid extraction of cloned plasmid. However, in retrospect, the plasmid format sacrifices genetic stability compared with genes integrated in the genome. Thus, function of the newly made operons should be characterized as integrated form in the BGM vector. Attempts to integrate them in the BGM vector are currently pending.

5.6 HOW TO PURIFY RECOMBINANT GENOMES

Much focus has been placed on the DNA-cloning process in the BGM vector. These assembled DNA molecules then have to be delivered to another host system to diversify applications. In particular, biomaterial production, which is regulated by a number of metabolic genes, is advantageous if efficiently expressed in lower costing hosts. Engineering of organelle genomes, mitochondria and chloroplast, is much



BGM approach toward organella and nucleus genomes

Figure 5-18 Organelle genomes and nucleus genome. Nucleus and two organelle genomes, mitochondria and chloroplast, are schematically viewed. The eight genes relevant for antocyanine bioprocess described in Figure 5-6 are shown. Chloroplast genomes from rice and tobacco are in circular form retrieved out of the BGM vector described in Section 5.6.3.

likely to establish an exciting field because organelle genomes are shown to interfere and exchange genes among genomes *in vivo*, such as transfer with plant nuclear genes [63] shown in Figure 5-18. Three methods for retrieval of the cloned segment from the BGM vector are summarized in Figure 5-19 and stated below.

5.6.1 General Method by Sequence Specific Endonucleases

Digestion by endonuclease of the cloned BGM recombinant and subsequent isolation/purification of the cloned segment appear to be the most simple and straightforward method. Extreme infrequent recognition sequences, 23-base [ATGACTCTCTTAA/GGTAGCCAAA] for I-*Ppo*I [64], and 18-base [TAGGGATAA/CAGGGTAAT] for I-*Sce*I [65] have been shown valuable in the BGM vector [16,36]. Owing to I-*Ppo*I site preinstalled at both ends of the GpBR of all the BGM vectors (Fig. 5-5), linearized DNAs produced on I-*Ppo*I digestion are readily isolated from agarose gel resolved by pulsed-field gel electrophoresis and can then be concentrated in liquid form. Because of such simplicity as depicted in Figure 5-19, the method has been of great use in primary analyses as well as pilot preparation of the cloned DNA [23,36].

5.6.2 Dissection of the *B. subtilis* Genome

The second method, genome dissection, largely depends on the *B. subtilis* genetic systems. This method originated from the study on diversity of multiple chromosomes

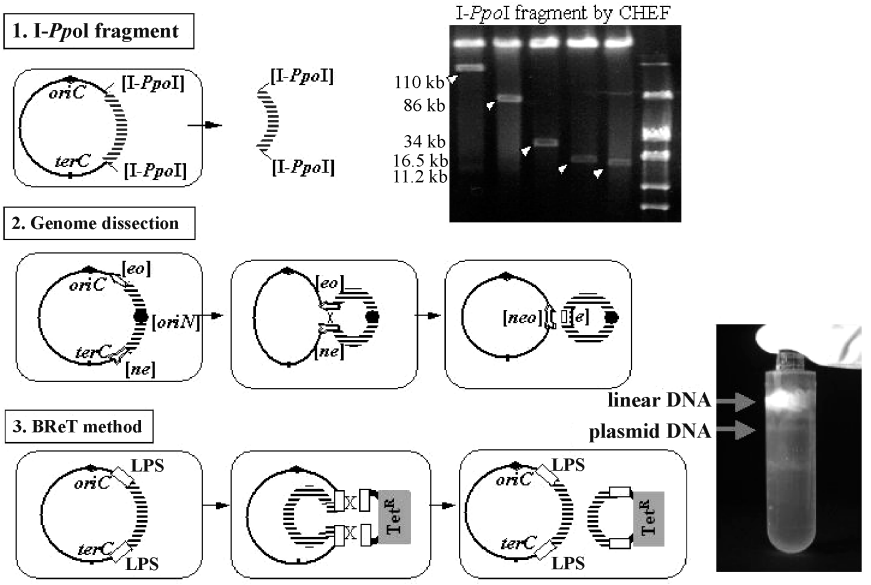


Figure 5-19 Three methods to extract the cloned DNA out of the BGM vector. Three methods are currently BGM vector specific. Examples for DNA resolved by the first method (Section 5.6.1) are shown in top right modified from the data in [90]. The extracted DNA by the second (Section 5.6.2) and third method (Section 5.6.3) is purified on density gradient of cesium chloride formed on ultracentrifugation [21]. Retrieved plasmid DNA is clearly separated from linear DNA mostly comes from sheared genome DNA. The molecular apparatus in the second method is also referred to in Section 5.7.2 and Figure 5-22.

in bacteria illustrated in Figure 5-1. The molecular apparatus briefly described in Figure 5-19 made it possible to physically disconnect long genomic DNA segments. Intrachromosomal homologous recombination between the two DNA repeats separated by 300 kb simply produced the intervening 300 kb as a second chromosome [66,67]. The extreme stability of the reported 300 kb second chromosome, termed the subgenome in the original report [66], exhibited extreme genetic stability via unknown mechanisms. The observation was inconsistent with the somewhat lower stability of the original plasmid, pLS32, whose replicational origin sequence (*oriN*) was required for subgenome replication. Later, presence of one of the essential genes reported by Kobayashi et al. [68] in the subgenome was shown to account for the genetic stability. In spite of the potential ability to make available circular subgenomes larger than 300 kb DNA by the method, broad application remains provisional due to its somewhat complicated procedure (Itaya M. and Fujita K., unpublished observations). Use of the conjugational transfer plasmid may make the delivery process more convenient [69].

5.6.3 Retrieval by Copying the Segment of the *B. subtilis* Genome

Compared with the somewhat elaborate genome dissection method, the third method proceeds through a yet more complicated genetic process referred to as *Bacillus*

recombinational transfer (BReT) [70]. Indeed, with the BReT system, copying a DNA segment from the genome and pasting it into the incomplete plasmid, causes the DNA transfer from the genome to plasmid in an apparent reverse direction as that of the domino or inchworm cloning method. The recipient plasmid should possess two landing pad sequences (LPS). If they are the two half pBR sequences, the intervening/cloned DNA segment between the two GpBR halves bridges the gap by landing to the plasmid. As illustrated in simplified manner in Figure 5-19, the complete circular form given only by the BReT pathway should be selected by plasmid-linked markers. Standard extraction protocol for plasmid DNA has resulted in purification of complete recombinant genomes of lambda [70] and organelle genomes from mitochondria and chloroplast [23] as indicated in Figure 5-18. These are the first recombinant genomes shown to be convertible to another host [23]. The BReT system has been widely used to retrieve certain *B. subtilis* genome regions due to technical simplicity [71,72].

5.7 WHAT IS BACTERIAL GENOME?

The ongoing in-depth genome sequencing analyses have unveiled a number of factors in gene composition, location, orientation, accessory sequences, promoters, and so on. Numerous examples deduced from whole genome sequencing results unveiled that HGT plays a significant role in generating subpopulations even today [9,29,39,43–47,50]. Is the concept underlying the present bacterial genomes strengthened or amended if more number of whole genome sequence data are added? Evolutional processes responsible for the mitochondrial genome development from the ancestor alpha purple bacteria [1] and chloroplast development from photosynthetic cyanobacteria [2] might represent the largest HGT in history. Gene capturing suggested in some cases [73] has prompted us to mimic the process in using laboratory expertise and aiming at plausible phenotypic conversion upon HGT. A more extensive review on detailed diversity among the sequence-known genomes for subspecies and/or variant genomes will be avoided here, instead focus will be on our experimental approaches, such as genome laundering to create stable mosaic subspecies [74,75] (Figs 5-20 and 5-21) and inversion mutations to expand genome structural variants (Fig. 5-22). It should be mentioned that these two works started before the *B. subtilis* whole genome sequence were determined in 1997 [11]. Although little conclusive gene function based interpretation was made, I believe our primary attempt to convert nascent concepts of plastic genomes to the real examples of gene assembly is demonstrated.

5.7.1 Genome Conversion: Stable Mosaic Genomes by HGT

Obvious traits among closely related strains, where one presents whilst another lacks, may be explained by gain or loss of relevant genes. One of our earlier studies on HGT [74], monitoring terminal phenotypes for *Natto*-production bioprocess, unveiled both predicted and novel constraints underlying the gene content per genome. We used *Natto*, a traditional Japanese food made from boiled soybean. On the soybean surface, a growing strain of *Bacillus natto* produces viscous biofilm-like materials [74].

B. natto whose genome has not been sequenced exhibits high similarity to that of *B. subtilis* with respect to a physical map-based genome comparison [76]. Ability to ferment soybean is only possessed by *B. natto,* and several relevant *Natto* genes have been identified from *B. natto*. Because of the similar context of promoters and identical ribosomal RNA sequences between *B. subtilis* and *B. natto*, these *Natto*-relevant genes must be lost from or deficient in *B. subtilis*. Therefore, consecutive replacement or displacement of *B. subtilis* genome regions by *B. natto* genome DNA incorporated through homologous recombination was performed. The genomic DNA replacement, termed as genome laundering [74], may permit assembly of the lost or deficient *Natto*-genes in the genome of nonproducer *B. subtilis* as illustrated in Figure 5-20. *B. natto* DNA randomly launders the multiple *B. subtilis* genome loci via fragments of approximately 50 kb on average—equivalent to about 50 genes. When the genome laundering process was repeated, the *B. subtilis* genome gradually and discontinuously becomes mosaic as a greater number of genes are introduced. Results summarized in Table 5.1 clearly show that as degree of mosaic increases, phenotypes specific to *B. subtilis* are subdued in parallel with the appearance of *Natto* characteristics. The highest mosaic strain nicknamed as *Natsuko7* exhibiting the most *Natto* traits carries approximately 350 kb of *B. natto*-originated DNA segment, 8 percent of the total 4215 kb *B. subtilis* 168 genome [74]. The estimated degree of DNA heterogeneity was later found not consistent with the expressed proteins profile. As shown in Figure 5-21, the apparent similarity of proteins expressed in *Natusko*7 to those of *B. natto* is biased far greater than the 8 percent estimated from the DNA. *Natusko*7 thus might be classified as *B. natto*. This observation postulates that the global gene networks are

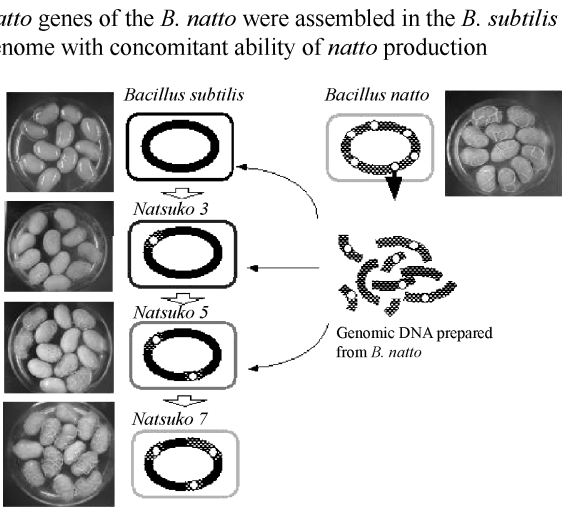


Figure 5-20 Mosaic genomes yielded by HGT. Genome parts from the *B. natto* transferred to and incorporated by replacement in the *B. subtilis* genome. Gradual and stepwise accumulation of *Natto* relevant genes converts non-*Natto* producer to the producer in proportion to the degree of mosaic indicated in Table 5.1. See the details in Section 5.7.1.

Table 5-1 ***Natsuko*: Stable intermediate strain retaining both parental traits. Traits (1,2) or (3–6) are specific to *B. subtilis* or *B. natto*.**

Strain	(1)	(2)	(3)	(4)	(5)	(6)
B. subtilis	+ + +	+ + +	—	—	—	—
*Natsuko*1	+ + +	+ + +	+	—	—	—
*Natsuko*2	+ + +	+ + +	+ +	—	—	—
*Natsuko*3	+ + +	+ + +	+ +	—	—	—
*Natsuko*4	+ + +	+ + +	+ +	+	—	—
*Natsuko*5	+ + +	+ +	+ +	+ + +	+ + +	+ +
*Natsuko*6	+ +	+ +	+ + +	+ + +	+ + +	+ + +
*Natsuko*7	+	+	+ + +	+ + + +	+ + + +	+ + +
B. natto	—	—	+ + + +	+ + + +	+ + + +	+ + + +

(1) Ability to develop competent.
(2) Growth on Spizizen plate.
(3) Viscosity of colonies on GSP plate at 42°C for 24 h.
(4) Natto fermentation.
(5) Protease secretion.
(6) Natto fragrance 24 h at 42°C.

physiologically and metabolically dominated by the activity of translation even between closely related species. In this sense, *Natusko*7 may represent the earliest attempt to combine traits form the two separate genomes in one cell. But it does not constitute a recombinant genome because the genome size as well as junction region remain largely indiscriminatory. Rather, the *Natsuko* series strains listed in Table 5.1 invoked an idea for relevant DNA/gene assembly and may be relevant in "the first truly engineered bacterial genome."

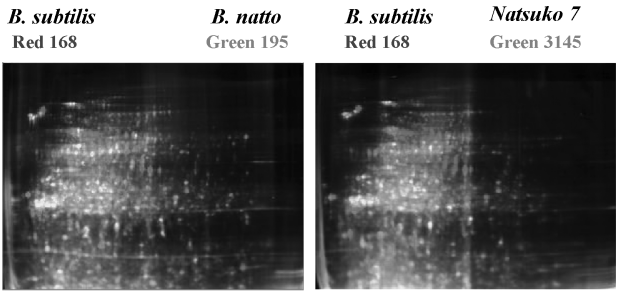


Figure 5-21 Similarity in expressed protein population. Separately labeled proteins prepared from two strains are mixed and run. In proteins resolved by two-dimensional electrophoresis, green spots display major expression in *B. natto* (left) and *Natuko*7 (right). Similar protein profile strongly indicates the highest *Natto*-producer, *Natsuko*7, virtually *B. natto* in spite of the expected degree of DNA converted. See the details in Section 5.7.1.

5.7.2 Rearrangements of Genome Structure by Inversion

Structure of the bacterial genome, a backbone for a set of genes, appears to be stably maintained after a number of replication cycles and cell divisions. The primary structural constraint is clearly exhibited by the diverse modes of replicons as shown in Figure 5-1. A secondary constraint is shown by the symmetry of the genome structure. Two replication arms divided by the *oriC-terC* axis, opposite the location of an initiation and termination locus, are well defined for certain bacterial genomes [44–46].

This structural symmetry together with gene alignment, orientation, and location seems conserved [44].

Bacterial genome plasticity, proposed and argued by pioneers many years ago, does not necessarily dictate apparent phenotypic changes. Thus, in-depth analysis must wait for development of reliable and easily accessible analytical methods such as physical map construction, and comparable sequence determination for closely related genomes. In line with these acknowledgements, the *B. subtilis* genome, an essential component of the BGM vector, is also considered to be plastic. Given analogies to plasmid vectors, genome vectors should surely be tolerant to structural disorder associated with DNA cloning and subsequent maintenance.

Among the naturally occurring events that induce considerable structural disorder listed in Figure 5-22, only inversion does not produce obvious genome size change; accordingly, resulting in little fluctuation of the gene set per genome. Yet, inversion may alter gene order and/or orientations, as well as relative gene locations in the genome. Experimental generation of systematic inversion mutants in the past decades has supported the concept of bacterial genome plasticity [77–79]. Focusing on the *B. subtilis* genome, there is evidence that supports both the stability [15,16,76] and the plasticity of the 4.2 Mb primary sequence [77–79].

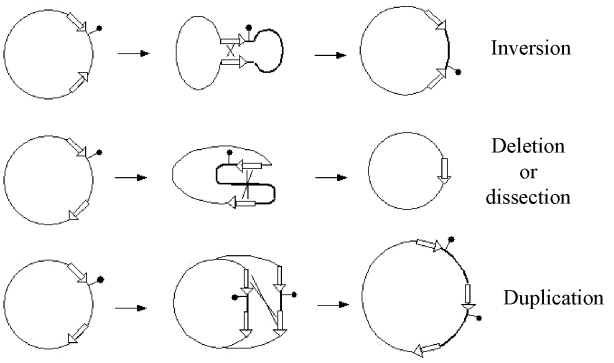


Figure 5-22 To induce large disorder of the genome structure. Two identical sequences separately embedded in the different loci of the genome induce inversion, deletion/dissection, or duplication by intrachromosomal homologous recombination. Resulted rearrangement is sequence orientation dependent. The second method to recover the cloned DNA (Fig. 5-19) is aided by the middle protocol here. Inversion mutagenesis is described in Section 5.7.2.

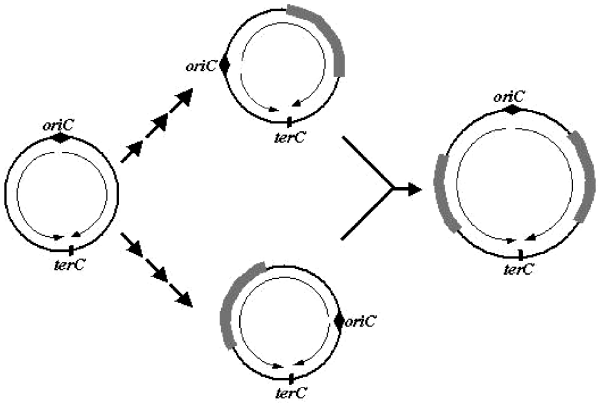


Figure 5-23 Requirement of asymmetry of the bacterial genome for rapid growth is true. Asymmetry around *oriC-terC* axis observed in certain bacterial genome is obstinate to ensure growth. Growth reduction by large DNA insertion is compensated by DNA insertion into opposite half to restore symmetry. This was proven in *Synechocystis* genome megacloning [36] also mentioned in Figure 5-9.

From our attempts to invert 28 various regions of the *B. subtilis* genome, some general conclusions can be drawn. Regions from the smallest 300 kb to the largest 1900 kb, covering any region of the 4215 kb *B. subtilis*168 genome, have been shown to be invertible without losing mutant viability (Toda T. and Itaya M., unpublished observations). All the inversion mutants induced by the molecular apparatus commonly used to create subgenomes in Figure 5-19 stably grew. The ability to form spores, one of the phenotypes strongly associated with *B. subtilis,* was maintained in all cases. In contrast, competency, another intrinsic feature important to megacloning, shows reduction specific to the inversion of particular regions. Significant growth reduction was also observed in mutants possessing dramatic asymmetry around the *oriC-terC* axis, and therefore restoration of growth rate seems solely dependant on symmetry. This working hypothesis has been clearly evidenced by growth recovery in inversion mutants that restore symmetry (Kuroki A. and Itaya M., unpublished observations) and relocation of the origin of replication so as to make *oric-terC* axis normal (Tomita, S. and Itaya, M., unpublished observation). This was highlighted in a very practical case where elongation of *Synechocystis* DNA above 1000 kb was stalled during megacloning as described in Section 5.4.3 [36]. Additional megacloning in the opposite arm alleviated the asymmetry and allowed for the complete cloning as illustrated in Figure 5-23. The symmetry rule seemed decisively dominant in spite of genome size [36]. On the contrary, the yet undetermined *oriC-terC* loci of *Synechocystis* PCC6803 create a particular concern, if this strain is eligible to harbor the 4.2 Mb *B. subtilis* genome in a similar fashion as shown in Figures 5-9 and 5-10.

5.7.3 The Largest Bacterial Genome

What is the upper size limit of a bacterial genome? The largest evidenced bacterial genomes are from the *Streptomycete* genus with 9.7 Mb [4]. Considering *B. subtilis*,

where the cognate 4.2 Mb became a 7.7 Mb hybrid-genome, the following question arises: Can this *CyanoBacillus* accept another 4.8 Mb of DNA to yield a 12.5 Mb hybrid-genome? This size is comparable to the smallest natural eukaryote, *S. cerevisae*, with a 12.5 Mb genome [83]. Why would someone want to do this? Is it just for pure scientific curiosity or some industrial application of these organisms with double or triple the amount of normal DNA? Two points should be addressed here. The long-standing question of the plausible size boundary to discriminate eukaryotes from prokaryotes will be examined experimentally [83]. The large extra DNA serves as a broad palette for various applications, for example, the recurrent use of existing genes through gene duplication or lateral transfer is the most common evolutionary mechanism to generate new protein-coding genes in bacteria [84].

5.7.4 Minimal Set of Genes for Life

The availability of a large number of complete genome sequences raises the question of how many genes are essential for cellular life [9,10,85]. Attempts to reconstruct the core of the protein-coding gene set for hypothetical minimal bacterial cell performed by computational as well as experimental analyses are not detailed here. The main features of such a minimal gene set that must be present in the hypothetical minimal cell would be informational genes (genes involved in transcription, translation, and other related processes) and operational genes (genes involved in cellular metabolic processes such as amino acid biosynthesis, cell envelope and lipid synthesis, and so on).

With regard to terminology "minimal set of genes (MSG)" is not equivalent to minimal genome. This is simply because MSG is a sum of the number of genes elucidated and listed, but a minimal genome requires a real body in which the complete nucleotide sequences must be contained. Apart from the determination that the MSG estimated by *in silico* analyses varies from 206 to 254 [85] gene in number, it is informative to discuss here how to experimentally construct the real minimal genome DNA body. The approach taken may be either top–down or bottom–up. The generation of mutation in all of the nonessential ORFs may be the beginning of the former case, as *B. subltilis* is shown to possess the 270 essential genes required for fixed growth conditions [68,86]. I focus on the latter case only by suggesting how to extract all the essential genes, less than 300 in number, and combine them together in a single-DNA sequence, preferably circular. One can imagine that it is sufficient that given DNA is anchored in a cellular bag, and the bag is exposed in nutritionally sufficient media. But the task is not that simple. Certainly, even limited in DNA structure, millions of questions arise in *de novo* assembly of more than 200 genes or gene blocks relating to the biological meaning when examined *in vivo*. For example, in the case where only 10 genes should be aligned in one DNA segment, the OGAB or possibly the domino method experimentally permits the designed DNA assembly. However, with lacking of information included in Figure 5-6a, no one writes the *de novo* nucleotide sequence.

Intermediate approaches that may be productive and timesaving may start from genomes with a nearly full set of essential genes present but lacking some

of those predicted from the *in silico* experimentation. The *Buchnella* species [85], symbionts in certain aphids, may provide several starting points en route to a full set of essential genes. Given the *Buchnella* genome, being several hundred kilo base pairs, it is plausible that by megacloning method or domino method or other combinatorial use, it may serve as the initial body for addition of a certain number of relevant yet lacking genes. By this scenario, modified *Buchnella* genome can be clearly provided in *B. subtilis* cell after genome dissection method or BReT method. The approach sounds at present controversial and remained to be experimentally investigated.

5.8 MISCELLANEOUS USE OF THE BGM VECTOR

5.8.1 Extension of Target DNA to Present DNA Libraries

Our primary assumption of no sequence discrimination during cloning of increased DNA size was strengthened by successful results demonstrated to date. The source of natural DNAs targeted to our BGM vector has been expanding rapidly. Also, the use of currently available cloned DNA resources is progressing. The most likely candidates are DNAs stocked in the prominent *E. coli* cloning vectors, the bacterial artificial chromosome (BAC) [87]. In spite of the nomenclature "chromosome," the structures do not actually resemble true bacterial chromosomes, but are really a large plasmid given the use of the origin of replication from the large stable conjugative plasmid F [87]. Indeed, due to the potential to clone DNA far larger than those considered at the time of introduction in 1991 and its technological simplicity, BACs have greatly attributed to DNA library construction, from mouse to the human genomes, and are widely available. Those DNAs cloned into an *E. coli* BAC vector and directly transferred to the BGM would be another DNA resource or method for DNA library generation. A BGM vector specialized for this aim has been developed [35]. Changing the familiar GpBR sequences to genomic BAC (GBAC) sequences permit direct transfer of a BAC insert ranging up to 200 kb as illustrated in Figure 5-24. It should be emphasized that the DNAs cloned into the BGM vector, in general, automatically serve as a long-term preservation system; owing to the ability for *B. subtilis* to form spores as indicated in Figure 5-24 [35].

5.8.2 Mouse Gneomic DNA in BGM

It is often difficult within BAC libraries to find a single BAC clone that covers an entire genomic gene of interest that possesses a number of introns of various lengths and controlling elements particular to the genes of higher eukaryotes. Connecting together two or more overlapping BAC inserts into a single clone can provide a full-length copy of the gene of interest. Indeed, highly efficient homologous recombination systems have been exploited in *E. coli* that allow for modifications of large BACs without the use of restriction enzymes and ligases, a process called recombineering [88,89]. We have demonstrated similar molecular processing of mouse genomic DNA in the BGM

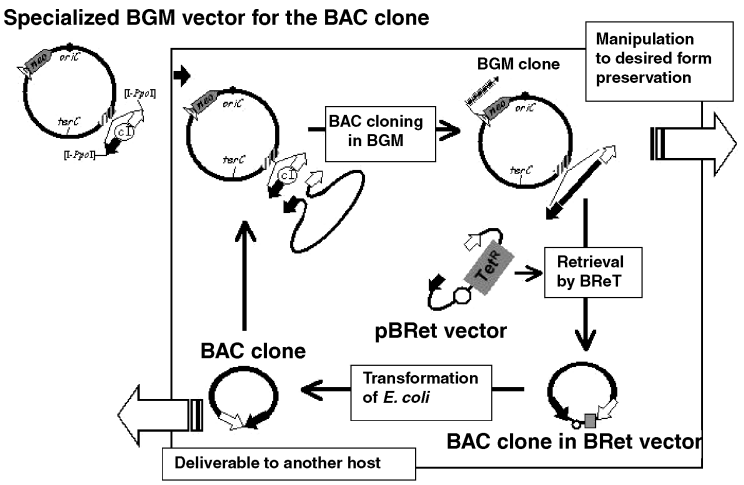


Figure 5-24 Application of the BGM vector to effective transfer from BAC library. A BGM vector providing BAC vector sequence as LPS, open and closed arrow, instead of pBR322 was made (left). Effectiveness of the present BAC library to more versatile use, including long-term preservation as spore form, appears successful [35]. These are mentioned in Section 5.8.1.

vector [34,90]. Joining of the two overlapping BACs to result in a 350 kb continuous mouse DNA segment in the BGM vector (Kaneko S. and Itaya M., unpublished data). Both BAC and BGM methods are similar given that homologous recombination is employed as the basic molecular mechanism. However, I believe our BGM system has super flexible handling to postcloning manipulation.

5.8.3 Sequence Fidelity of Recombinant Genomes

As mentioned above, genomes larger than 100 kb in purified form in solution are not only subjected to physical shearing but also sensitive to contaminating nucleases. The former case may invoke the problem associated with geographical preservation [29], and the latter are technically significant considering laboratory use [21]. From our experience in handling large DNA, most nucleases are removed via careful purification steps, for example, washing two times as a part of the standard protocol wash, wearing gloves, avoiding physical contact with unwrapped instrument, preservation in the presence of nuclease inhibitors such as chelating chemicals EDTA, etc.. Highly concentrated DNA seems resistant to residual nucleases probably due to competitive inhibition.

One fundamental concern for the rGenome is fidelity of the nucleotide sequences of recombinant progenies. Of course deterioration via mutations under given growth conditions must be selected during DNA replication, mutation inevitably accumulates even with the inherent sophisticated repair systems to repair mismatch of bases and removal and replacement of misincorporated nucleotides in the natural genome. Two *E. coli* K-12 derivatives separately cultivated over 40 year

revealed nucleotide changes of 10^{-7} per base per year (or 1 in every 10^7 bases) [91]. Hence, we reach a very serious dilemma during construction of the designed rGenomes. Unsuspected mutations may freely accumulate in proportion to both the increase of the DNA size and number of cell divisions during prolonged cultivation. As mentioned in Section 5.4, freshly prepared mtGenome and cpGenome exhibited no obvious nucleotide alteration [23]. One may evade this intrinsic problem by resequencing the DNA whose time and cost turns relevant recently [5].

5.9 SUMMARY

DNA cloning is one of the most basic and inevitable tools to investigate genes and gene functions in biological sciences. Cloning strategy has changed depending on DNA sources, cloning vehicles, and cloning hosts. Innovation of the PCR method allows amplification of unlimited regions of DNA from various species, as well as reducing the required amount and purity of the template DNA. Recently, chemically synthesizing DNA without template DNA above dozens of kilo base pair in size has been shown to be possible and may soon be widely available. The BGM vector is a powerful tool for cloning the present genomes as a whole and for conferring subsequent modifications and reconstruction according to the purpose of the users. General schemes from cloning, pinpoint manipulation, and retrieval protocols are briefly described in Figures 5-2 and 5-19, and examples of progressive elongation of fragments with or without overlapped regions are described in Section 5.3, or in Section 5.4, respectively.

These two examples show reconstruction in the BGM vector of significantly large and continuous target DNA or genomes. Another method to assemble a number of DNA fragments in unit DNA blocks is to form gene clusters, so-called man-made operon, and is described in Section 5.5. Genomes from bacteriophage lambda (48.5 kb), mouse mitochondrial genome (16.3 kb), rice chloroplast genomes (135 kb), mouse genomic genes (up to 350 kb), and the whole genome of a photosynthetic bacterium (3500 kb) were cloned in the BGM vector; this is far above results obtained by conventional gene cloning and technology. The new cloning concept, readily applicable to any sequenced DNA, confers an experimental basis for not only provisional organelle genome engineering but also emerging recombinogenic engineering. Furthermore, the method could be used to assemble sequence-designed DNAs that are made from scratch.

The ultimate goal based on these technologies should be of course to unveil the nature underlying this still somewhat fragile iceberg; as well as to show how to plan and conduct new material production systems beneficial for humans and the earth. Efforts are being undertaken to make all protocols more conventional, hoping that the BGM vector users work just as comfortably as listening to background music.

5.10 FUTURE PERSPECTIVES

Life activities are controlled by the complex and dynamic actions of thousands of genes encoded by their genome. With the sequencing of many genomes, the key

problem has shifted from identifying genes to knowing what the genes do; we need a framework for expressing that knowledge. Further comprehensive and systematic investigations regarding other factors listed in Figure 5-6 are required to draw at least some foundational rules on the design of the gene cluster. Gene clusters starting from an even small gene content exhibit various, and far different in some cases, levels of metabolites that may influence terminal phenotypes of the cell where a number of other gene networks may be perturbed. Gene cluster designing and molecular construction for certain basic metabolic pathways such as glycolysis have begun in Keio University (Tsuge et al., unpublished). It should be addressed here that the experimental basis for OGAB and/or domino is offering to provide a substantially large number of isomers with the same gene content. Although it seems a long-way off for creation of genome variations and establishment of subsequent assays for selection, this challenge would be crucial in bottoming up in facilitating the designing of life. The idea of making the appropriate *B. subtilis* genome into a linear chromosome vector similar to the eukaryotic stable vectors, such as YAC, remains to be initiated.

Production of designed genomes and consequently cellular life is one of the most challenging tasks in systems biology and synthetic biology, which still remains a nascent field. As deduced from the impact caused by gene cloning and recombinant genetics in the last century, once technologies show any potential, many start thinking how to use a technology to accomplish their goal. In this context, the present achievement shows that whole bacterial genomes have now become a target for cloning that was previously not an available option. I personally think the present two-genome cell is just a start for teaching many readers that "genome cloning or recombinant genomes" is becoming a reality. I personally believe that modification/manipulation of the recombinant Genomes of *Synechocystis*, would be much faster than genetic conversion of *Synechocystis* itself in traditional manner.

I always recall the aim of recombinant genomes research clearly addressed 35 years ago by late Dr Fujio Egami, primary director of the institute to which the author previously worked for 20 years. He said "The primary aim of life reorganization is to bring about fulfilled and comfortable lives for present and future human. Thus, there is no need to make the same cells as those currently available. As a consequence, what we need is to provide simpler and more beneficial cells for us."

The recombinant genome technology will markedly change the style of research not only limited to biomedical and biomaterial production, but also yielding a significant scientific, economic, and cultural impact. Cloning DNA to some extent mimics HGT. Finally, I would like to say that I make it a rule to keep learning from the greatest tutor, *Nature*.

ACKNOWLEDGEMENTS

I thank Drs Hoshino T., Ikeuchi M., Takeuchi T., Tanaka T., Yanagawa H., and Yoshikawa H. for their useful comments and many discussions. I also thank my

colleagues, Drs Kaneko S., Ohashi Y., Oshima H., Tsuge K., and many other collaborators for their daily works and number of discussions. Many technical staffs and students of graduate as well as undergraduate in my laboratories at Mitsubishi Kagaku Institute of Life Sciences and Keio University are specially acknowledged for their great contributions to improve the BGM vector. I particularly thank Prof. Tomita M., Director, Institute of Advanced Biosciences of Keio, for his continuous encouragement and opportunities for genome designing works. Finally, special thanks are addressed to Ms Fujita. K. whose contribution year to genome designing project is the longest.

REFERENCES

1. Gray MW, Burger G, Lang BF. Mitochondrial evolution. *Science* 1999;283:1476–1481.
2. Bendich A. Circular chloroplast chromosomes: the grand illusion. *Plant Cell* 2004; 161:1661–1666.
3. Endy D. Foundations for engineering biology. *Nature* 2005;438:449–453.
4. http://www.ncbi.nlm.nih.gov
5. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, LA, Bemben. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437: 376–380.
6. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 2004;428:37–43.
7. Venter JC. et al. Environmental genome shotgun sequencing of the Sargasso sea. *Science* 2004;304:66–74.
8. Cello J, Paul AV, Wimmer E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science* 2002;297:1016–1018.
9. Kooning EV. Minimal set of genes. *Nat Rev Microbiol* 2003;2:127–136.
10. Gil R, Latorre A, Moya A. Bacterial endosymbionts of insects: insights from comparative genomics. *Environ Microbiol* 2004;6:1109–1122.
11. Kunst F. et al. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 1997;390:249–256.
12. Itaya M, Tanaka T. Complete physical map of the *Bacillus subtilis* 168 chromosome constructed by a gene-directed mutagenesis method. *J Mol Biol* 1991;220:631–648.
13. Itaya M. Toward a bacterial genome technology: integration of the *Escherichia coli* prophage lambda genome into the *Bacillus subtilis* 168 chromosome. *Mol Gen Genet* 1995;248:9–16.
14. Itaya M, Tanaka T. Gene-directed mutagenesis on the chromosome of *Bacillus subtilis*. *Mol Gen Genet* 1990;223:268–272.
15. Itaya M. Integration of repeated sequences (pBR322) in the *Bacillus subtilis* 168 chromosome without affecting the genome structure. *Mol Gen Genet* 1993;241: 287–297.
16. Itaya M. Stability and asymmetric replication of the *Bacillus subtilis* 168 chromosome. *J Bacteriol* 1993;175:741–749.

17. Itaya M, Shiroishi T, Nagata T, Fujita K, Tsuge K. Efficient cloning and engineering of giant DNAs in a novel *Bacillus subtilis* genome vector. *J Biochem* 2000;128:869–875.
18. Spizizen J. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proc Natl Acad Sci USA* 1958;44:1072–1078.
19. Dubnau DA. DNA uptake in bacteria. *Annu Rev Microbiol* 1999;53:217–244.
20. Chen I. Christie PJ, Dubnau DA. The ins and outs of DNA transfer in bacteria. *Science* 2005;310:1456–1460.
21. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
22. Sanger F, Coulson AR, Hong GF, Hill DF, Peterson GB. Nucleotide sequence of bacteriophage lambda DNA. *J Mol Biol* 1982;162:729–773.
23. Itaya M. Fujita K, Kuroki A, Tsuge K. Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nature Methods* 2008;5:41–43.
24. Yoon YG, Koob MD. Efficient cloning and engineering of entire mitochondrial genomes in *Escherichia coli* and transfer into transcriptionally active mitochondria. *Nucleic Acids Res* 2003;31:1407–1415.
25. Winkel-Shirley B. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* 2001;126:485–493.
26. Barkovich R, Liao JC. Metabolic engineering of isoprenoids. *Metab Eng* 2001;3:27–39.
27. Misawa N, Shimada H. Metabolic engineering for the production of carotenoids in non-carotenogenic bacteria and yeast. *J Biotech* 1998;59:169–181.
28. Yoshikawa H, personal communications.
29. Lorenz MG, Wackernagel W. Bacterial gene transfer by natural genetic transformation in the environment. *Microbiol Rev* 1994;58:563–602.
30. Itaya M. Effective cloning of unmarked DNA fragments in the *Bacillus subtilis* 168 genome. *Biosci Biotech Biochem* 1999;63:602–604.
31. Uotsu-Tomita R, Kaneko S, Tsuge K, Itaya M. Insertion of unmarked sequences in multiple loci of the *Bacillus subtilis* 168 genome: an efficient selection method. *Biosci Biotechnol Biochem* 2005;69:1036–1039.
32. Itaya M, Nagata T, Shiroishi T, Fujita K, Tsuge K. Efficient cloning and engineering of giant DNAs in a novel *Bacillus subtilis* genome vector. *J Biochem* 2000;128:869–875.
33. Itaya M, Fujita K, Koizumi M, Ikeuchi M, Tsuge K. Stable positional cloning of long continuous DNA in the *Bacillus subtilis* genome vector. *J Biochem* 2003;134:513–519.
34. Yonemura I, Nakada K, Sato A, Hayashi J-I, Fujita K, Kaneko S, Itaya M. Direct cloning of full-length mouse mitochondrial DNA via *Bacillus subtilis* genome vector. *Gene* 2007;391:171–177.
35. Kaneko S, Akioka M, Tsuge K, Itaya M. DNA shuttling between plasmid vectors and a genome vector: systematic conversion and preservation of DNA libraries using the *Bacillus subtilis* genome (BGM) vector. *J Mol Biol* 2005;349:1036–1044.
36. Itaya M, Tsuge K, Koizumi M, Fujita K. Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc Natl Acad Sci USA* 2005;102:15971–15976.

37. Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S. et al. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II, Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res* 1996;3:109–136.
38. Nelson KE, Paulsen IT, Heidelberg JF Fraser CM. Status of genome projects for nonpathogenic bacteria and archaea. *Nat Biotechnol* 2000;18:1049–1054.
39. Doolittle WF. Phylogenetic classification and the universal tree. *Science* 1999;284: 2124–2129.
40. Berkum PB, Terefework Z, Paulin L, Suomalainen S, Lindstro K, Eardly BD. Discordant Phylogenies within the *rrn* Loci of Rhizobia. *J Bacteriol* 2003;185:2988–2998.
41. Yap WH, Zhang Z, Wang Y. Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol* 1999;181:5201–5209.
42. Evguenieva-Hackenberg E. Bacterial ribosomal RNA in pieces. *Mol Microbiol* 2005;57: 318–325.
43. Riviera MC, Lake J. The ring of life provides evidence for a genome fusion origin of eukaryote. *Nature* 2004;431:152–155.
44. Rocha EP. Is there a role for replication fork asymmetry in the distribution of genes in bacterial genomes? *Trends Microbiol* 2002;10:393–395.
45. Rocha EP, Danchin A. Essentiality, not expressiveness, drives gene-strand bias in bacteria. *Nat Genet* 2003;34:377–378.
46. Rocha EP, Danchin A, Viari A. Universal replication biases in bacteria. *Mol Microbiol* 1999;32:11–16.
47. Rocha EP, Guerdoux-Jamet P, Moszer I, Viari A, Danchin A. Implication of gene distribution in the bacterial chromosome for the bacterial cell factory. *J Biotechnol* 2000;78:209–219.
48. Lawrence JG. Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* 2002;110:407–413.
49. Price MN. et al. Operon formation is driven by co-regulation and not by horizontal gene transfer. *Genome Res* 2003;15:809.
50. Omelchenko MV, Makarova KS, Wolf YI, Rogozin IB, Koonin EV. Evolution of mosaic operons by horizontal gene transfer and gene displacement *in situ*. *Genome Biol* 2003;4:R55.
51. Khosla C, Keasling JD. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov* 2003;2:1019–1025.
52. Wang CW, Oh MK, Liao JC. Direct evolution of metabolically engineered *Escherichia coli* for carotenoid production. *Biotechnol Prog* 2000;16:922–926.
53. Kao CM, Katz L, Khosla C. Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* 1994;265:509–512.
54. Pfeifer BA, Admiraal S, Gramajo H, Cane DE, Khosla C. Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* 2001;291:1790–1792.
55. Szczebara FM, Chandelier C, Villeret C, Masurel A, Bourot S, Duport C, Blanchard S, Groisillier A, Testet E, Costaglioli P. et al. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat Biotechnol* 2003;21:143–149.
56. Alper H, Miyaoku K, Stephanopoulos G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol* 2005;23:612–616.

57. Watanabe K, Rude MA, Walsh CT, Khosla C. Engineered biosynthesis of an ansamycin polyketide precursor in *Escherichia coli*. *Proc Natl Acad Sci USA* 2003;100: 9774–9778.
58. Wenzel SC, Gross F, Zhang Y, Fu J, Stewart AF, Müller R. Heterologous expression of a Myxobacterial natural products assembly line in Pseudomonads via Red/ET recombineering. *Chem Biol* 2005;12:349–356.
59. Kumar P, Khosla C, Tang Y. Manipulation and analysis of polyketide synthases. *Methods Enzymol* 2005;388:269–293.
60. Tsuge K, Matsui K, Itaya M. One step assembly of multiple DNA fragments with designed order and orientation in *Bacillus subtilis* plasmid. *Nucleic Acids Res* 2003;31(21): e-133.
61. Nishizaki T, Tsuge K, Itaya M, Doi N, Yanagawa H. Metabolic engineering of carotenoid biosynthesis in *Escherichia coli* by ordered gene assembly in *Bacillus subtilis* (OGAB). *Appl Environ Microbiol* 2007;73:1355–1361.
62. Tsuge K, Matsui K, Itaya M. Production of the non-ribosomal peptide plipastatin in *Bacillus subtilis* regulated by 3 relevant gene blocks assembled in a single movable DNA segment. *J Biotechnol* 2007;129:592–603.
63. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, Hirai A, Kadowaki K. The complete sequence of the rice (*Oryza sativa* L,) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Gen Genomics* 2002;268:434–445.
64. Argast GM, Stephens KM, Emond MJ, Monnat RJ. I-*Ppo*I and I-*Cre*I homing site sequence degeneracy determined by random mutagenesis and sequential *in vitro* enrichment. *J Mol Biol* 1998;280:345–353.
65. Moure CM, Gimble FS, Quiocho FA. The crystal structure of the gene targeting homing endonuclease I-*Sce*I reveals the origins of its target site specificity. *J Mol Biol* 2003;334: 685–695.
66. Itaya M, Tanaka T. Experimental surgery to create subgenomes of *Bacillus subtilis* 168. *Proc Natl Acad Sci USA* 1997;94:5378–5382.
67. Itaya M, Tanaka T. Fate of unstable *Bacillus subtilis* subgenome: reintegration and amplification in the main genome. *FEBS Lett* 1999;448:235–238.
68. Kobayashi K, Ehrich SD, Aibertini A, Amati G, Andersen KK. et al. Essential *Bacillus subtilis* gene. *Proc Natl Acad Sci USA* 2003;100:4678–4683.
69. Itaya M, Sakaya N, Matsunaga S, Fujita K, Kaneko S. Conjugational transfer kinetics for *Bacillus subtilis* in liquid culture. *Biosci Biotechnol Biochem* 2006;70:740–742.
70. Tsuge K, Itaya M. Recombinational transfer of 100-kb genomic DNA to plasmid in *Bacillus subtilis* 168. *J Bacteriol* 2001;183:5453–5456.
71. Tomita S, Tsuge K, Kikuchi Y, Itaya M. Targeted isolation of a designated region of the *Bacillus subtilis* 168 genome by Recombinational transfer. *Appl Environ Microbiol* 2004;70:2508–2513.
72. Tomita S, Tsuge K, Kikuchi Y, Itaya M. Regional dependent efficiency for recombinational transfer of the *Bacillus subtilis* 168 genome. *Biosci Biotech Biochem* 2004;68:1382–1384.
73. Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L. et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* 2000;406:477–483.
74. Itaya M, Matsui K. Conversion of *Bacillus subtilis*. 168: natto producing *Bacillus subtilis* with mosaic genomes. *Biosci Biotech Biochem* 1999;63:2034–2037.

75. Itaya M, Tanaka T, Predicted and unsuspected alterations of the genome structure of genetically defined *Bacillus subtilis* 168 strains. *Biosci Biotech Biochem* 1997;61:56–64.
76. Qiu D, Fujita K, Sakuma Y, Tanaka T, Ohashi Y, Ohshima H, Tomita M, Itaya M. Comparative analysis of the structure of complete physical maps of four *Bacillus subtilis* (natto) genome. *Appl Environ Microbiol* 2004;70:6247–6256.
77. Rebollo JE, François V, Louarn JM. Detection and possible role of two large nondivisible zones on the *Escherichia coli* chromosome. *Proc Natl Acad Sci USA* 1988;85:9391–9395.
78. Segall A, Mahan MJ, Roth JR. Rearrangement of the bacterial chromosome: forbidden inversions. *Science* 1988;241:1314–1318.
79. Campo N, Dias MJ, Daveran-Mingot ML, Ritzenthaler P. Le Bourgeois P, chromosomal constraints in Gram-positive bacteria revealed by artificial inversions. *Mol Microbiol* 2004;51:511–522.
80. Itaya M. First evidence for homologous recombination-mediated DNA inversion of the *Bacillus subtilis* 168 chromosome. *Biosci Biotech Biochem* 1994;58:1836–1841.
81. Toda T, Itaya M. I-*Ceu*I recognition sites in the *rrn* operons of the *Bacillus subtilis* 168 chromosome: inherent landmarks for genome analysis. *Microbiology* 1995;141: 1937–1945.
82. Toda T, Tanaka T, Itaya M. A method to invert DNA segments of the *Bacillus subtilis* 168 genome by recombination between two homologous sequences. *Biosci Biotech Biochem* 1996;60:773–778.
83. Bendich A, Drlica K. Prokaryotic and eukaryotic chromosomes: what's the difference? *Bioessays* 2000;22:481–486.
84. Gevers D, Vandepoele K, Simillon C. Van de Peer Y, Gene duplication and biased functional retention of paralogs in bacterial genomes. *Trends Microbiol* 2004;12:148–154.
85. Gil R, Silva FJ, Pereto J, Moya A. Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev* 2004;68:518–537.
86. Itaya M. An estimation of minimal genome size required for life. *FEBS Lett* 1995;362:257–260.
87. Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F- factor-based vector. *Proc Natl Acad Sci USA* 1992;89:8794–8797.
88. Copeland NG, Jenkins NA, Court DL. Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* 2001;2:769–779.
89. Kotzamanis G, Huxley C. Recombining overlapping BACs into a single large BAC. *BMC Biotechnol* 2004;4:I.
90. Kaneko S, Tsuge K, Takeuchi T, Itaya M. Conversion of submegasized DNA to desired structures using a novel *Bacillus subtilis* genome vector. *Nucleic Acids Res* 2003;31(18): e-112.
91. Itoh T, Okayama T, Hashimoto H, Takeda J, Davis RW, Mori H, Gojobori T. A low rate of nucleotide change in *Escherichia coli* K-12 estimated from a comparison of the genome sequences between two different substrains. *FEBS Lett* 1999;450:72–76.