Hypothesis

An estimation of minimal genome size required for life

Mitsuhiro Itaya*

Mitsubishi Kasei Institute of Life Sciences, 11-Minamiooya, Machida-shi, Tokyo 194, Japan

Received 21 February 1995

Abstract The number of indispensable chromosomal loci for a bacterium, *Bacillus subtilis* was estimated. Seventy-nine randomly selected chromosomal loci were investigated by mutagenesis. Mutation at only six loci rendered *B. subtilis* unable to form colonies. In contrast, mutants for the rest of the 73 loci retained the ability to form colonies. Mutant *B. subtilis* with multiple-fold mutations of those dispensable loci (7-, 12- or 33-fold) were not impaired in their ability to form colonies on nutritionally adequate medium, indicating that up to 33 dispensable loci were simultaneously abolished. Given the statistical analyses for the frequency of indispensable loci (6 out of 79), total indispensable genetic material would be included within about 562 kbp. The hypothetical minimum genome size lies in the range of those currently determined smallest genomes for bacteria.

Key words: Minimal genome; Indispensable gene; Mutagenesis; Dispensable gene; Colony formation; Nutrition

1. Introduction

Is the current structure of the bacterial genome a historical accident or are there selecting forces for a particular order and arrangement of the genes on the chromosome? Recent technical improvements for bacterial genome analyses revealed that their sizes ranged from 585 kb up to several thousands kbp (kilobase pairs) [1]. Supposing that there is a limited number of 'indispensable' genetic loci for bacteria, can the least genetic material or the size of a hypothetical minimal genome be deduced?

An experimental approach was carried out using *Bacillus subtilis* 168 strain, a rod-shaped Gram-positive, endospore-forming soil bacterium, as mutagenesis for *B. subtilis* is readily done by transformation [2]. Seventy-nine randomly selected chromosomal loci of the *B. subtilis* 168 chromosome were subjected to mutagenesis. Those loci included 40 *Not*I sites, 13 *Sfi*I sites, 7 *Sse*8387I sites, and 19 randomly collected DNA segments by the author (indicated in the legend to Fig. 1). The number of indispensable loci were found to be surprisingly low. The results allowed the estimation of minimal number of loci as well as minimal DNA size required for colony formation of *B. subtilis* whose determined genome size (4188 kbp) is ranked as intermediate [3]. A hypothetical minimum DNA size may account for the smallest genome size of terrestrial life.

*Corresponding author. Fax: (81) (427) 24 6316.

2. Materials and methods

B. subtilis 168trpC2, a standard *B. subtilis* 168 strain, was from the *Bacillus* Genetic Stock Center (OH, USA). A *Sfi*I–*Not*I physical map was reported [3], whose brief description is in the legend to Fig. 1.

Site-directed mutagenesis of the *B. subtilis* chromosome was carried out by transformation. Briefly, mutations (insertion of antibiotic resistance gene cassette at a particular restriction enzyme site) were constructed first in the cloned *B. subtilis* DNA segment using an *E. coli* gene engineering system [2,3]. The manipulated DNA is taken up by spontaneous competent *B. subtilis* cells and proceeds with homologous recombination with the chromosomal DNA part inside the competent cell. The mutant cells, resulting in integration of the antibiotic resistance gene into the chromosomal target site, were selected as the antibiotic resistant transformants followed by verification with Southern hybridization analyses. Competent *B. subtilis* cells were prepared as previously described [2,3].

Preparation of *B. subtilis* chromosomal DNA for pulsed field gel electrophoresis, *B. subtilis* DNA library on *E. coli* JA221, and procedures for the Southern analysis were done as previously described [3].

Type II restriction enzymes and T4 DNA ligase were obtained from Toyobo (Tokyo, Japan), except for an eight-base sequence CCTGCAGG recognizer *Sse*8387I (Takara Shuzo, Kyoto, Japan). Manipulation of cloned DNAs were done according to Maniatis et al. [4] or the supplier's manual. LB media [5] were used for selection of mutant bacteria.

3. Results

At least 40 *Sse*8387I sites were found in the *B. subtilis* chromosome (unpublished observations). Seven *Sse*8387I-linking clones (linking clones are defined as cloned *B. subtilis* DNA fragment carrying an *Sse*8387I restriction enzyme site) were obtained from the *B. subtilis* DNA library by the same method to clone *Not*I- and *Sfi*I-linking clones as previously described [3]. Their chromosomal locations were determined by hybridization as well as physical mapping technique (data not shown, but the result is presented in Fig. 1). Mutagenesis at these *Sse*8387I sites was done by an insertion of the neomycin resistance gene cassette prepared from pBEST509 with *Pst*I digestion [3]. Neomycin resistant *B. subtilis* transformants (insertional mutation at each of the cloned *Sse*8387I sites) were selected at the neomycin concentration of 15 μg/ml without obvious phenotypes. The observation demonstrated that the seven *Sse*8387I sites were all dispensable for colony formation. Subsequently, a 7-fold mutant strain (BEST3094) was con-

SSDI 0014-5793(95)00233-2

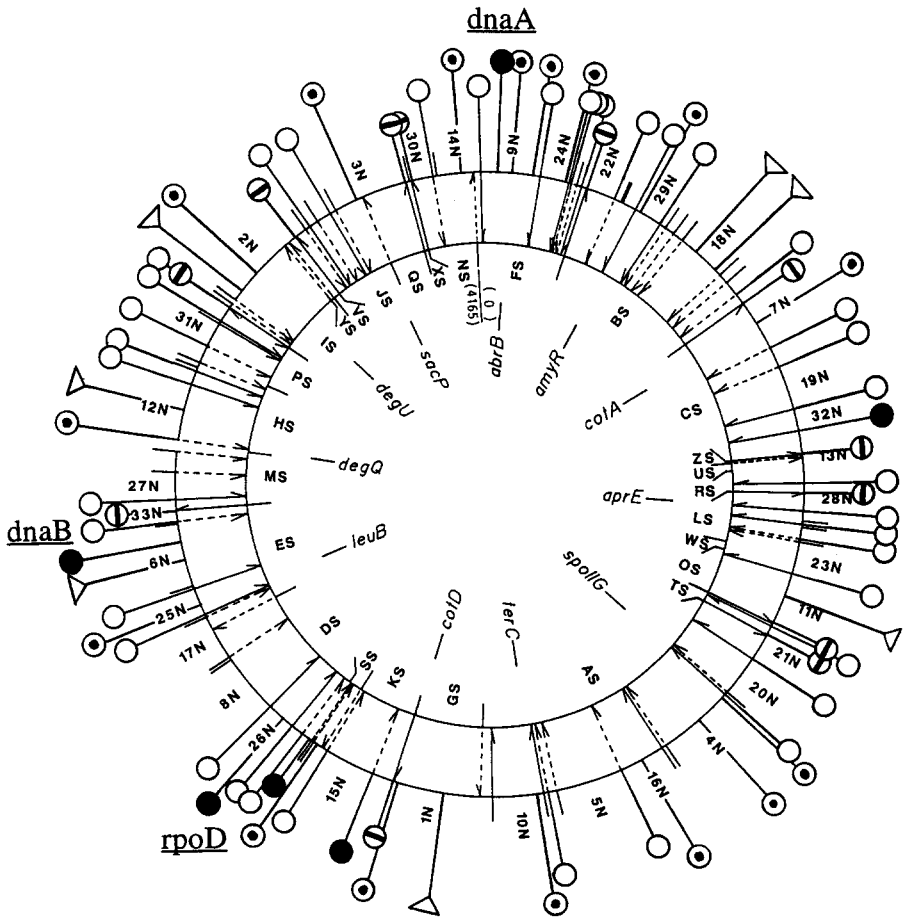

Fig. 1. Seventy-nine loci of the *B. subtilis* chromosome examined for mutagenesis. The *Not*I–*Sfi*I physical map was re-drawn with minor modifications. Briefly, 26 *Sfi*I fragments (AS–ZS) were aligned in the inner circle and 72 *Not*I fragments (1N–33N and smaller) in the outer circle. The zero point of the 4188 kb map was designated as the *Sfi*I site between fragments FS and NS. Loci subjected for mutagenesis analysis are indicated by small open circles or open triangles. Small circles (◐—), (○—) or (⊙—), represent the loci for 13 *Sfi*I, 40 *Not*I, and 19 gene loci, respectively. Details are described in the text and in [3,6,12]. Seven open triangles indicate the *Sse*8387I sites that were mapped on the physical map in this study. Six closed circles indicate indispensable loci, three of which were assigned as genes for *dnaA*, *dnaB*, and *rpoD*.

structed by a method designed to introduce multiple mutations in *B. subtilis* [2,3]. Ability of BEST3094 to form colonies indicate that all these seven *Sse*8387I loci were simultaneously abolished (Fig. 2).

Similar observations had been previously reported without the particular emphasis of the subject in this article [3]. Briefly, similar insertional mutagenesis was carried out at 40 *Not*I sites, 13 *Sfi*I sites, and 19 additional gene loci collected for the mapping purpose. Only two out of the 40 *Not*I, one out of the 13 *Sfi*I, and three out of the 19 gene loci were found to be indispensable [3,6]. Those previously examined loci are included in Fig. 1.

The unexpectedly few number of indispensable loci urged me to estimate the total number of indispensable loci. 'Indispensability' was defined in this report as *B. subtilis* chromosomal loci required for colony formation on LB media plate, at 37°C. Three of the six indispensable loci were assigned to those directly involved in macromolecule synthesis (*dnaA*, *dnaB*, and *rpoD*: Fig. 1), yet the other three (two *Not*I sites and an *Sfi*I site) remained to be characterized [3]. Auxotrophic genes were expected to be included in dispensable loci, because cells do not need gene products for de novo synthetic pathways, due to adequate nutritional ingredients. It was verified that three amino acid auxotrophic mutants were found among dispensable *Not*I mutants [3].

The *B. subtilis* strain with 33 dispensable *Not*I mutations (BEST4133), or 12 dispensable *Sfi*I mutations (BEST3055) retained the ability to form colonies (Fig. 2). Although their colony morphology were heavily altered, the viability of multiple-fold mutants indicated that most of the dispensable loci could be concurrently discarded for bacterial cell growth on rich media and substantiate the view that dispensable loci exhibiting synthetic lethal phenotypes are few.

A statistical analysis of the present mutagenesis data (frequency of indispensable loci = 6/79) showed that the mean value was 7.59% and the maximum frequency was 13.4% (normal distribution, $P = 0.95$). Indispensable DNA size was thus calculated at 318 kbp and a maximum of 562 kbp based on the determined 4188 kb chromosome size [3].

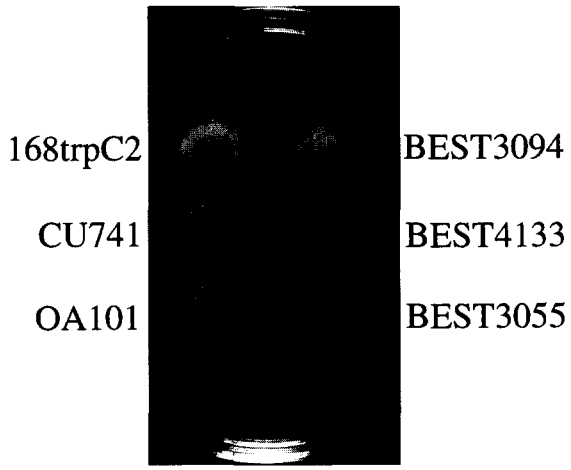

Fig. 2. *B. subtilis* mutants carrying multiple mutations. 168trpC2, CU741, and OA101 are standard *B. subtilis* strains [13]. Colonies of BEST3094 (seven *Sse*8387I sites of 168trpC2 are mutated, in Fig. 1), BEST4133 (thirty-three *Not*I sites of CU741 mutated, in [3]), and BEST3055 (twelve *Sfi*I sites of OA101 mutated, in [3]) were grown on LB medium kept at 30°C for 3 days.

4. Discussion

Two hypotheses are proposed.

(I) The number of genetic loci to support bacterial colony formation on nutritionally rich media is strikingly small.

(II) The bacterial chromosome could be reduced in size to a minimum under defined growth conditions.

The integral characteristic for experiments from which these two hypotheses stem was that all the mutagenesis processes and mutant selection were carried out under well defined conditions, temperature, nutritional media, concentration of antibiotics, etc. The calculation for frequency of indispensable loci was based on apparent random distribution of the 79 loci investigated (Fig. 1) by disregarding (1) the actual characteristic of individual genes or non-functional regions, and (2) possible bias for high GC content of the *Not*I site (GCGGCCGC). The significance for the hypothesis (I) remains to be demonstrated by investigating the first drawback. The second possibility seems less influential as there was no apparent bias for the frequency of indispensability between *Not*I sites and the other two GC-less *Sfi*I (GGCCNNNNNGGCC) or *Sse*8387I (CCTGCAGG) sites.

The 318–562 kbp accounts for most of what we know about the genome size. First, the currently determined small genomes are found in mycoplasma species (585–1330 kbp [1]), or spirochaete species (900–1000 kbp [7]), followed by an archaebacteria (*Methanobacterium thermoautotrophicum* 1623 kbp [8]). Direct comparison with mycoplasma spp. and spirochaete spp. may, however, be inappropriate because they exhibit moderate symbiotic characteristics as well as pathogenic phenotypes. The size-related view remains to be modified or strengthened by discoveries of more microbial species that have various genome sizes, since a tremendous number of bacterial species remain uncultivated [9]. Second, it is intriguing to picture that 318–562 kb DNA may provide a border to discriminate the living and non-living organism. The largest virus genome reported to date was about 370 kbp [10]. It is necessary to define the genome as the sum of all the genes of an organism but viruses cannot by themselves produce their offspring in any media.

An argument was raised that living organisms share a minimum amount of genetic material or more than necessary. Given that the least amount of genetic material to be sufficient to support 'Life' in suitable circumstances, by which they manage their for the obligatory production of offspring, present bacteria must have the luxury of having developed sophisticated mechanisms in addition to the fundamental machinery such as macromolecule synthesis (DNA, RNA, proteins), energy metabolism, sustaining the compartment. Thus, the genome size could be adaptively increased by duplications, horizontal transfer, etc., in the course of evolution, resulting in suitable regulation of gene expression or number of the genome adaptive to their environmental disturbance. These genome parts, probably, determine the versatility of a bacterial life style. For example, there has been an estimate that *E. coli* has about 1800 functional gene products out of approximately 4000 genes [11]. The hypothesis (I) predicts that a number of genes could be discarded in rich media as precursor metabolites permeate the cell membrane.

Hypothesis (II) comes from the feature of the viability of multiple mutant strains (Fig. 2). Morphological changes for multiple-fold mutants (caused by up to 33-fold mutations) was an obligatory part associated with the lower number of functional genetic loci. This hypothesis raised a practical question if a minimal genome size could be achieved from *B. subtilis* 168 by progressive DNA deletions. For example the number of *rrn* operons, presently ten for wild type *B. subtilis*, could be reduced at the expense of rapid growth [12]. Several spontaneous deletion formations were found at various regions of the *B. subtilis* chromosome [3,13]. The strain that carries the smallest *B. subtilis* genome was already the case to date (4030 kb; decreased by 158 kb [13]). It remains to be tested if other reported deletions could be simultaneously formed in *B. subtilis*.

As the entire nucleotide sequence of the *B. subtilis* chromosome is being determined [14,15], we will learn more from further analysis of the *B. subtilis* genome.

Acknowledgements: I thank Drs. T. Tanaka, M. Ogura, N. Ogasawara, T. Toda, F. Kawamura, L. Hirshbein, M. Riley, L. Sympson, A. Weiner, R. Losick, Y. Kikuchi, T. Yamada, and H. Yanagawa for discussions. I also thank Dr. S. Inoue for discussion of statistical analyses.

References

[1] Krawiec, S. and Riley, M. (1990) Microbiol. Rev. 54, 502–539.
[2] Itaya, M. and Tanaka, T. (1990) Mol. Gen. Genet. 223, 268–272.
[3] Itaya, M. and Tanaka, T. (1991) J. Mol. Biol. 220, 631–648.
[4] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Press, Cold Spring Harbor, NY.
[5] Miller, J.H. (1972) in: Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[6] Itaya, M., Laffan, J.J. and Sueoka, N. (1992) J. Bacteriol. 174, 5466–5470.
[7] Saint Girons, I., Old, I.G. and Davidson, B.E. (1994) Microbiology 140, 1803–1816.
[8] Stettler, R. and Leisinger, T. (1992) J. Bacteriol. 174, 7227–7234.

[9] Ward, D.M., Bateson, M.M., Weller, R. and Ruff-Roberts, A.L. (1992) in: Advances in Microbial Ecology (Marshall, K.C., Ed.) vol. 12, pp. 219–286, Plenum, New York.
[10] Yamada, T., Higashiyama, T. and Fukuda, T. (1991) Appl. Environ. Microbiol. 57, 3433–3439.
[11] Niedhardt, F.C., Ingrahm, J.L. and Schaechter, M. (1990) in: Physiology of the Bacterial Cell: A Molecular Approach, pp. 463–481, Sinaur Associates, Inc., Sunderland, MA.
[12] Itaya, M. (1993) Biosci. Biotech. Biochem. 57, 1611–1614.
[13] Itaya, M. (1993) J. Bacteriol. 175, 741–749.
[14] Kunst, F., Vassarotti, A. and Danchin, A. (1995) Microbiology 141, 249–255.
[15] Ogasawara, N., Fujita, Y., Kobayashi, Y., Sadaie, Y., Tanaka, T., Takahashi, H., Yamane, K. and Yoshikawa, H. (1995) Microbiology 141, 257–259.