Dna Research, 2016, 23(6), 517–525
doi: 10.1093/dnares/dsw035
Advance Access Publication Date: 3 July 2016
Full Paper

Full Paper

Correlation between genome reduction and bacterial growth

Masaomi Kurokawa[1,†], Shigeto Seno[2,†], Hideo Matsuda[2], and Bei-Wen Ying[1,*]

[1]Graduate School of Life and Environmental Sciences, University of Tsukuba, Ibaraki 305-8572, Japan, and
[2]Graduate School of Information Science and Technology, Osaka University, Osaka 565-0871, Japan

*To whom correspondence should be addressed. Tel/Fax. +81 29 853 6633. Email: ying.beiwen.gf@u.tsukuba.ac.jp

[†]These authors contributed equally to this study.

Edited by Dr Katsumi Isono

Received 19 April 2016; Accepted 8 June 2016

Abstract

Genome reduction by removing dispensable genomic sequences in bacteria is commonly used in both fundamental and applied studies to determine the minimal genetic requirements for a living system or to develop highly efficient bioreactors. Nevertheless, whether and how the accumulative loss of dispensable genomic sequences disturbs bacterial growth remains unclear. To investigate the relationship between genome reduction and growth, a series of *Escherichia coli* strains carrying genomes reduced in a stepwise manner were used. Intensive growth analyses revealed that the accumulation of multiple genomic deletions caused decreases in the exponential growth rate and the saturated cell density in a deletion-length-dependent manner as well as gradual changes in the patterns of growth dynamics, regardless of the growth media. Accordingly, a perspective growth model linking genome evolution to genome engineering was proposed. This study provides the first demonstration of a quantitative connection between genomic sequence and bacterial growth, indicating that growth rate is potentially associated with dispensable genomic sequences.

Key words: growth rate, dispensable genomic sequence, genome reduction, accessory gene

1. Introduction

As a collection of genetic information, genomes are thought to be sequence redundant.[1] Since the first genome sequence was reported in bacteria,[2] experimental and theoretical studies[3–8] have attempted to determine the minimal sets of genes essential for cellular life and to define what constitutes a minimal genome.[9] A number of sound experiments have succeeded in constructing *Escherichia coli* libraries either of knockout strains, by deleting particular functional genes,[10,11] or of a variety of genome sizes, by removing the non-essential genomic sequences.[12–14] Recently, thought-provoking experiments involving synthetic genomes harbouring the smallest gene sets and exhibiting high replication efficiency[8,15] have further demonstrated that native genomic sequences are redundant in terms of the capacity for life. As a consequence, genetically engineered genomes lacking dispensable sequences are commonly used in both basic science and applied research.[16,17]

Reduced genomes have both advantages and disadvantages regarding cellular functions,[18] although few studies have commented on whether the deleted dispensable genomic sequences contribute to biological fitness. Intensive studies have shown that reduced genomes increase productivity in generating gene products,[19–21] which is a major advantage in various applications, and reduce the capacity to take up foreign DNA[22] as well as mutability,[23,24] which are likely to be disadvantageous to evolving living organisms. These results imply that the deletion of the “junk” regions of the genome increases the efficiency of biological processes but decreases cellular plasticity in living organisms.[22] Bioinformatic analysis has successfully linked the genome size to growth conditions.[25] However, insight into the relation between genome size and growth fitness is lacking. Although genome reduction has been found not to affect the capacity for growth,[12–14,26] experimental and theoretical studies of the global effects of genome reduction on cellular life are required.

Considering the reduced genomes as growing systems, the accumulated deletion of dispensable genomic sequences is assumed to be associated with growth rate, a common global parameter that represents the overall activity (output of the genomic and metabolic reactions) of a living cell. Such a global effect of bacterial growth has been demonstrated by the correlation between transcriptome and growth fitness[27,28] in both adaptation[29,30] and evolution.[31] However, studies constructing reduced genomes have reported diverse evidence regarding growth. Rough measurements of a series of reduced genomes (constructed from W3110) have not indicated any significant differences in the growth rate on the basis of optical density (OD),[14] whereas a closely related study has reported increased cell division time in another reduced genome library (constructed from MG1655)[32]. In addition, the well-known "clean" genome (MDS42)[12] has a growth rate comparable to that of its parent wild type genome (MG1655) in minimal medium,[22,33] and no general trend has been observed between genome size and growth in rich medium.[26] Whether and how the genome reduction links to the bacterial growth is still in argument, due to the different thinking on growth cost and/or genome evolution. Thus, the biological impact of genome reduction remains to be determined under different conditions in varied point views. Considering these various findings and assumptions, we sought to quantitatively evaluate the relation between cumulatively reduced genomic sequences and the changes in growth under nutritional variation.

In the present study, we addressed the question of whether and how the cumulative deletion of dispensable genomic sequences contributes to the population growth of *E. coli* in varied conditions. We used a previously constructed *E. coli* library comprising a series of genomes that were reduced in a stepwise manner,[14] and easy access to their genetic backgrounds was crucial for the analysis. Evaluation of the contribution of genomic sequences dispensable for bacterial growth was achieved through comprehensive growth assays of these *E. coli* strains. The results offer a perspective opinion of the relationship between genome size and growth fitness as a growth mode that is potentially involved in genome evolution and is applicable to genome engineering.

2. Materials and methods

2.1 Strains and media

A total of 28 *E. coli* strains with reduced genomes (KHK collection) and the wild type genome W3110 were obtained from the National BioResource Project, National Institute of Genetics, Shizuoka, Japan. The strains, which were initially cultured and stored in complete LB medium (Miller), were inoculated into M63 minimal medium and recultured until the optical density (OD_{600}) reached 0.01~0.1, representing the exponential growth phase. The cell culture was subsequently stored in 20 tubes in small aliquots (100 μl per tube) for future use. *Escherichia coli* culture stocks (newly prepared single tubes) were used once, and the remainder was discarded. The compositions of the three different media, the complete medium (LB), the minimal medium (M63), and the minimal medium supplemented with 20 amino acids (MAA), used in the present study were described previously.[34]

2.2 Acquisition of growth curves

The *E. coli* culture stocks were completely dissolved and diluted 10- to 100-fold with fresh media in a test tube. Three types of media, LB, MAA, and M63, were used, which represent the rich, supplementary, and poor growth conditions, respectively. The diluted cell mixtures of multiple dilution rates were subsequently loaded into a 96-well plate (Costar) in three to six wells per dilution rate per strain, with 200 μl of diluted cell mixture applied in each well. The 96-well plate was incubated in a plate reader (Epoch2, BioTek) with a rotation rate of 600 rpm at 37 °C. Cell growth was detected at an absorbance of 600 nm, with readings obtained at 30-min or 1-h intervals for 24 to 48 h. The growth curves were obtained for each well (Supplementary Fig. S2A). Repeated tests were performed, which resulted in 11 to 30 growth curves used for further calculations of growth rate and population density for each strain at each growth condition (medium). Growth curves were acquired in three different media: LB, M63 and MAA.

2.3 Calculation of exponential growth rate and saturated population density

The growth rate (μ) during exponential phase and population density during stationary phase was evaluated according to the growth curves. The growth rates were calculated from every two continuous reading points in a growth curve, according to the following equation (Equation 1).

$$\mu = LN(\frac{C_j}{C_i})/(t_j - t_i) \quad (1)$$

Here, C_i and C_j represent the two reads of OD_{600} values at two continuous time points of t_j and t_i, which were at intervals of either 0.5 or 1 (h) in the present study. Every four to five continuous growth rates (μ) that exhibited the largest mean and the smallest standard deviation was averaged and taken to be the exponential growth rate of this growth curve (Supplementary Fig. S2). The mean of the exponential growth rates resulting from 11 to 30 replicated growth curves was determined as the exponential growth rate (r) of the corresponding strain and used for the correlation analyses (Supplementary Fig. S3A). The population density was calculated directly from the growth curve by averaging five continuous reads at OD_{600} that exhibited the largest means (Supplementary Fig. S2). Similarly, the mean of the population densities that result from 11 to 30 replicated growth curves was determined as the saturated population density (K) of the corresponding strain and used for the correlation analyses (Supplementary Fig. S3B).

2.4 Measurements of cell concentration and size

The *E. coli* culture stocks were inoculated in 3 ml of M63 minimal medium at a 100-fold dilution and were cultured in a bioshaker (Taitec) with a rotation rate of 200 rpm at 37 °C. Each cell culture was temporally sampled for size measurement from the late exponential to stationary phases. The cell size (diameter) and the corresponding cell concentration were measured using a cell counter (Multisizer^{TM} 4 coulter counter, Beckman) equipped with a 20-μm aperture. Every 50 μl of cell culture, a 5- to 10-fold dilution, if required, was mixed with 10 ml of COULTER ISOTON II DILUENT (Beckman) in a 25-ml disposable vial (Accuvette ST, Beckman) for measurement. The distribution of cell sizes, including the mean, medium and standard deviation of the distribution, were acquired automatically at each measurement (30,000~100,000 cells). Repeated experiments (from cell culture to size measurement) were performed for each strain. The average of multiple values of the measured medium was taken to be the representative cell size of the cell population. The stationary cell sizes were determined as the cell size acquired at concentrations $>1 \times 10^9$ cells/ml, individually. Six to eight measurements during stationary phase were obtained for each strain, and the average of these measurements was applied to the analysis.

2.5 Evaluation of growth dynamics

A total of 1,060 growth curves were rescaled to be comparable (Supplementary Fig. S5). For all growth curves, the intervals between the start of growth over the measurement under-limit and the time required to achieve the maximum OD_{600} value were standardized to a value ranging from 0 to 100. The start of the growth phase was defined as the time when the increase of OD_{600} was observed in five consecutive reading points. In addition, OD_{600} values were also standardized to a value ranging from 0 to 1, by dividing each value on a curve by individual maximum OD_{600} value. The rescaled growth rates were calculated by taking differences between consecutive two reading points on rescaled growth curves, and the timing of the maximal growth rate was regarded as the standardized timing of the maximal growth rate. This value indicates when the cells most grew most efficiently, and comparing distributions of the values between strains provides insights into the differences in growth characteristics.

2.6 Multiple linear regression analysis

The data set comprising the growth rate, the length of accumulated deletions, the population size, the angle from *ori* to *dif*, and the growth medium was subjected to multiple linear regression analysis. In this data set, only the growth media (M) was a categorical variable (i.e. LB, MAA and M63) and was included in the multiple regression model by dummy coding. Thus, an artificial variable, as a dummy variable (or indicator variable), was created to represent the variation of M (= LB, MAA or M63) and was introduced to the multiple regression analysis. To increase the regression efficiency, regression variable selection was performed on the basis of AIC (Akaike Information Criteria),[35] which evaluated the relative quality of regression models for the data set.[36] Computational analyses were performed with R,[37] and a 3D-plot was created by using the car package.[38]

2.7 Organization and graphics of reduced genomes

Information on the deleted genomic sequences (genomic position and deletion ID) was obtained from the website http://shigen.nig.ac.jp/ecoli/strain/, which distributed the *E. coli* strains (KHK collection). The accumulated length of deleted genomic sequences was calculated. The numbers of deleted genes and the corresponding gene categories,[39] which is the only gene function classification that covering the complete gene set across the whole *E. coli* genome (Supplementary Table S2, 100% assigned function), were determined (Supplementary Table S1) and subjected to the correlation analysis, as previously described.[30] Graphical representations of the reduced genomes were created using the BRIG (BLAST Ring Image Generator) tool.[40]

3. Results and discussion

3.1 Reduced genomes: a stepwise loss of dispensable genomic sequences

An assortment of *E. coli* strains with reduced genomes was used to evaluate the contribution of dispensable genomic sequences to cell growth. A total of 29 *E. coli* strains, assigned as No. 0 for the wild type genome W3110 and Nos. 1 to 28 for the reduced genomes constructed in a stepwise manner from W3110,[14] were investigated. Note that the reduction of genome size was in a complete cumulative mode, i.e. the strain No. 2 maintained the deletion in No. 1, and No. 3 remained the deletions in No. 2, and so on, as illustrated (Fig. 1A). According to the patterns of multiple deletions that occurred in the reduced genomes, the length of multiple deletions accumulated in each genome was calculated. The results revealed a gradual increase in the length of deleted genomic sequences from ∼48 to 982 kb (Fig. 1B). In addition, the genes located in the deleted regions were clustered according to the gene categories (summarized in Supplementary Table S1) and subjected to enrichment analysis to identify the significant loss of gene function in each strain (reduced genomes), as previously reported.[29,30,31] Bimodal tests revealed that the deleted genomic sequences were most significantly enriched in the gene category of phage/IS-related function (h) (Supplementary Fig. S1), thus verifying that the deleted sequences were mostly non-essential. Note that all deleted genes are designated as the accessory genes in the present study, to avoid the debated definitions of essential and non-essential. Whether and how the dispensable genomic sequences contribute to cell growth in both poor and rich conditions were subsequently investigated.

3.2 Accumulated deletions caused correlated changes in population growth

A negative correlation between population growth and the accumulated deletions was the most salient finding in the present study. Repeated measurements of the *E. coli* strains that were grown under three different nutritional conditions, resulted in 1,060 independent growth curves. The growth rates during exponential phase and the saturated cell densities during stationary phase were manually calculated for each growth curve (Supplementary Fig. S2). A data set of the exponential growth rates and saturated cell densities associated with reduced genomes was acquired (Supplementary Fig. S3). Statistical tests indicated a highly significant correlation ($P < 0.001$) between the exponential growth rate and the length of cumulatively deleted genomic sequences in all three nutritional conditions (Fig. 2A). Genome reduction-induced growth decreases were clearly detected. In addition, the mode of growth decrease was nutritional dependent. In comparison to a continuous gradual decrease in the poor condition (M63), a stepwise decrease in the rich condition (LB) was detected, close to the finding reported recently.[26] The variation in mode of growth decrease reflected the degree of requirement of dispensable genomic sequences for efficient growth.

A correlation was also observed between the saturated cell density and accumulated deletions ($P < 0.05$); namely, an increase in deleted length was correlated with a decrease in cell density (Fig. 2B). The reduced cell density appeared to be attributed to the reduced number of cells; however, the loss of dispensable genomic sequences influenced the cell size at the saturated concentration (Fig. 3). Because the cell density was evaluated through OD measurements, the number or the size of the cells could contribute to the changes in turbidity. Measurement of the saturated culture in M63, in which the most significant decrease in turbidity was observed, revealed that the cumulative deletions triggered the decrease in cell concentration (Fig. 3B) and the increase in cell size (Fig. 3C) at the saturated phase. Nevertheless, a positive correlation between the cell density evaluated by optical measurement and the cell concentration according to the numbers (particles) of cells (Fig. 3A)

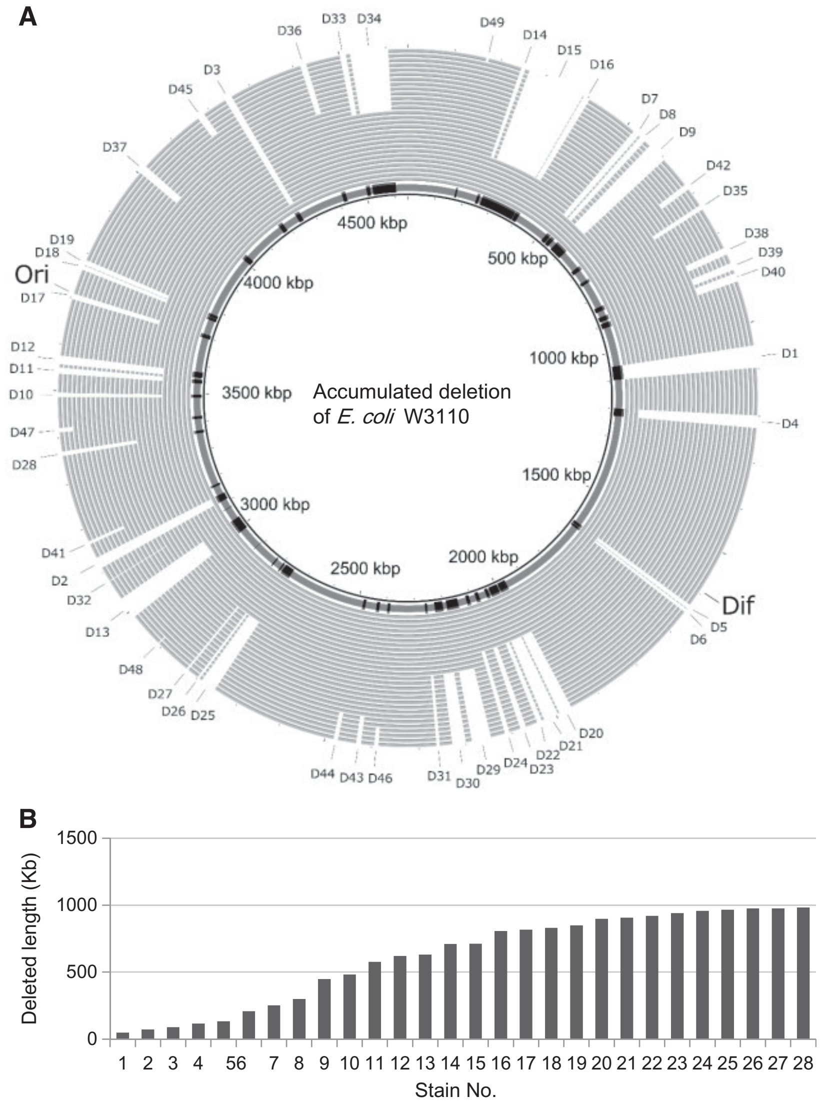

Figure 1. Reduced genomes. (A) Genomes of a total 29 *E. coli* strains. The inner bold lined ring indicates the wild type genome W3110, and black bands on the ring indicate the areas to be deleted. These deletions are also indicated by the deletion ID numbers (D1-D49) at the outside of rings. The truncated grey rings represent the reduced genomes of Nos. 1 to 28 from inside to outside, respectively. The replication initiation and termination sites are indicated by "Ori" and "Dif". (B) Accumulated length of deleted sequences. The sum of the deleted genomic length in the strains of reduced genomes (Nos. 1-28) was calculated.

confirmed the association "between accumulative genomic deletion and population decrease.

Together, our findings indicated that longer accumulative lengths of deleted dispensable genomic sequences not only led to the decrease in growth fitness during exponential phase but also decreased the final population size during stationary phase. Although the dispensable genomic sequences were non-essential for life, the length-dependent changes illustrated the slight but significant contribution of the accumulated loss of these sequences to bacterial growth.

3.3 The degree of change in population growth was nutritional dependent

A quantitative association between the accumulated length of multiple deletions and bacterial growth was constructed. Normalization of both the exponential growth rate and the saturated cell density indicated gradual changes in the magnitude of the decrease (Fig. 2C). Here, the exponential growth rate and the saturated cell density of the wild type genome (Strain No. 0) were considered as one unit. Linear regression was simply applied to fit the following equations (Equations 2 and 3).

$$r = 1 - \alpha L_d \qquad (2)$$

$$K = 1 - \beta L_d \qquad (3)$$

Here, L_d, r and K represent the length of the accumulated deletions (kb), the normalized growth rate and the cell density, respectively. The constants α and β indicate the magnitudes of the decrease in population growth in relation to the length of multiple deletions. In three different media, the slopes of α_{M63}, α_{MAA} and α_{LB} were 5e-4, 1e-4 and 9e-5 per kb, respectively (Fig. 2C, left). The slopes of β_{M63}, β_{MAA} and β_{LB} were 2e-4, 1e-4 and 7e-5 per kb, respectively (Fig. 2C, right). Thus, every 1-kb deletion would cause a 0.05%, 0.01%, and 0.009% fitness decrease and a 0.02%, 0.01%, and 0.007%

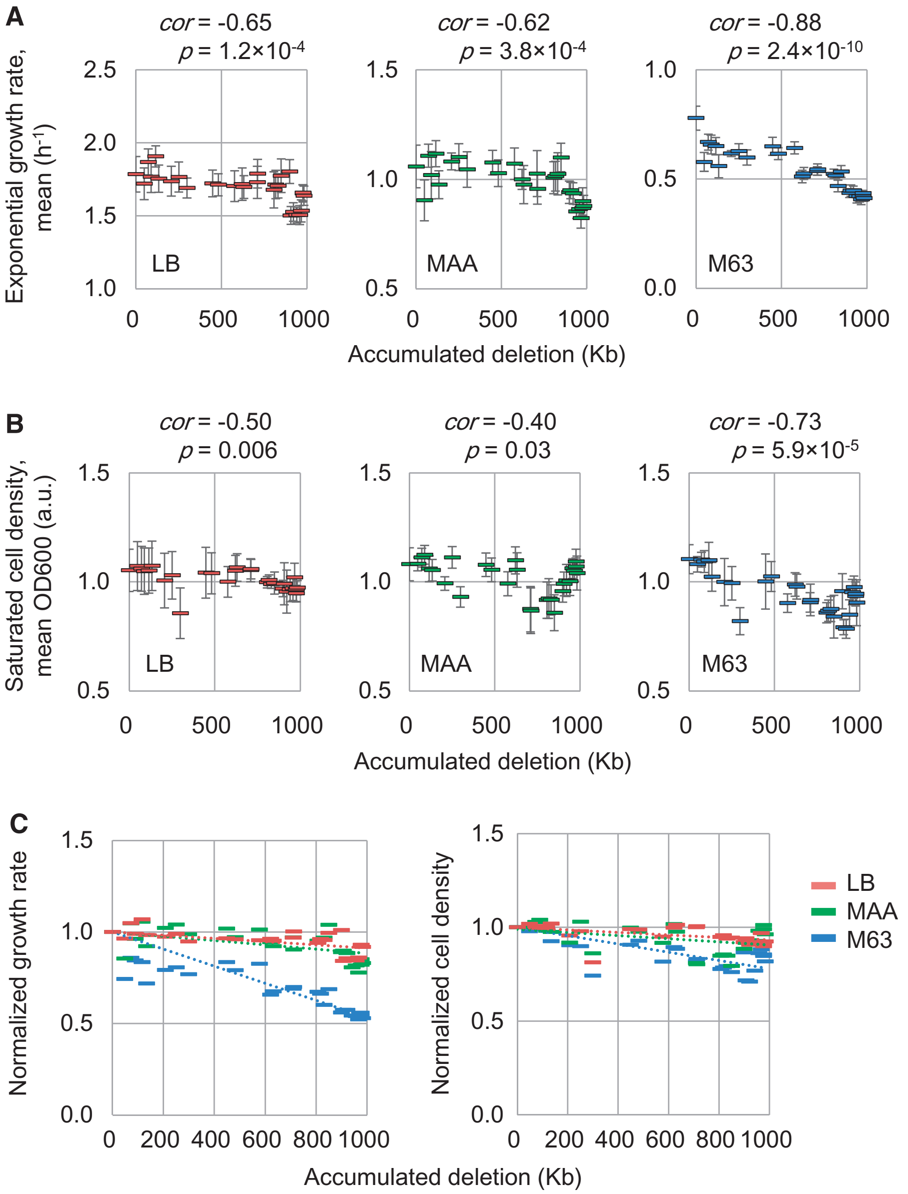

Figure 2. Accumulated deletion-dependent changes in population growth. (A) Correlated changes in growth rate. The calculated exponential growth rates are plotted against the length of accumulated deletions. (B) Correlated changes in saturated cell density. The calculated maximal OD_{600} values are plotted against the length of accumulated deletions. Standard errors ($N = 11{\sim}30$), growth media, correlation coefficients and *P* values are indicated. M63, MAA and LB indicate the growth media. (C) Nutritional dependent changes in population growth. The left and right panels represent the normalized decease in growth rate and the normalized decease in saturated cell density, respectively. Red, green and blue represent the growth in LB, MAA and M63, respectively.

population reduction in the conditions of M63, MAA, and LB, respectively. Both α and β exhibited gradual changes in the order of the richness of growth media, thus indicating that the extent of the length-dependent changes in population growth was nutritional dependent. The deleted genomic sequences affected both the growth fitness (r) and the environmental capacity (K), and this effect was more substantial in poor environments.

Combining the two equations (Equations 2 and 3) resulted in the following equation (Equation 4), and further transformation made the coordinated relationship between r and K clearer (Equation 5).

$$K = 1 - \frac{\beta}{\alpha}(1 - r) \tag{4}$$

$$K = C_M + k_M r \tag{5}$$

Here, C_M and k_M represent the medium-dependent population capacity and correlation, respectively. Although a positive correlation between population size and growth fitness could be estimated ($k_M > 0$), significant correlations were identified in the conditions of M63 and LB ($P < 0.01$) but not MAA (Supplementary Fig. S4). It might because of the relatively low significance in the correlation between the saturated cell density and genome reduction in MAA (Fig. 2B). The results suggest that the nutritionally differentiated bias in the relationships among growth parameters is likely to be related to genomic deletions.

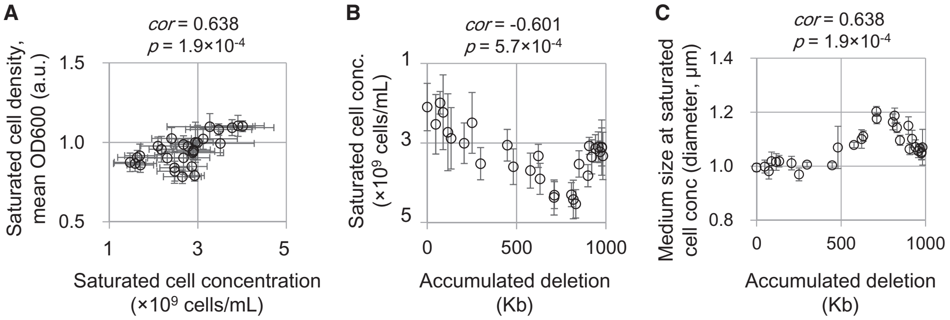

Figure 3. Changes in cell concentration and cell size. (A) Correlation between cell concentration and cell density. The saturated cell density calculated from optical measurement is plotted against the saturated cell concentration as measured by cell counter. (B) Correlations with changes in cell concentration. The saturated cell concentration is plotted against the length of accumulated deletions. (C) Correlations with changes in cell size. The mean cell size of the saturated population is plotted against the length of accumulated deletions. The cell size is represented by the diameter of the cell particle. Standard errors, correlation coefficients and statistical significance are indicated.

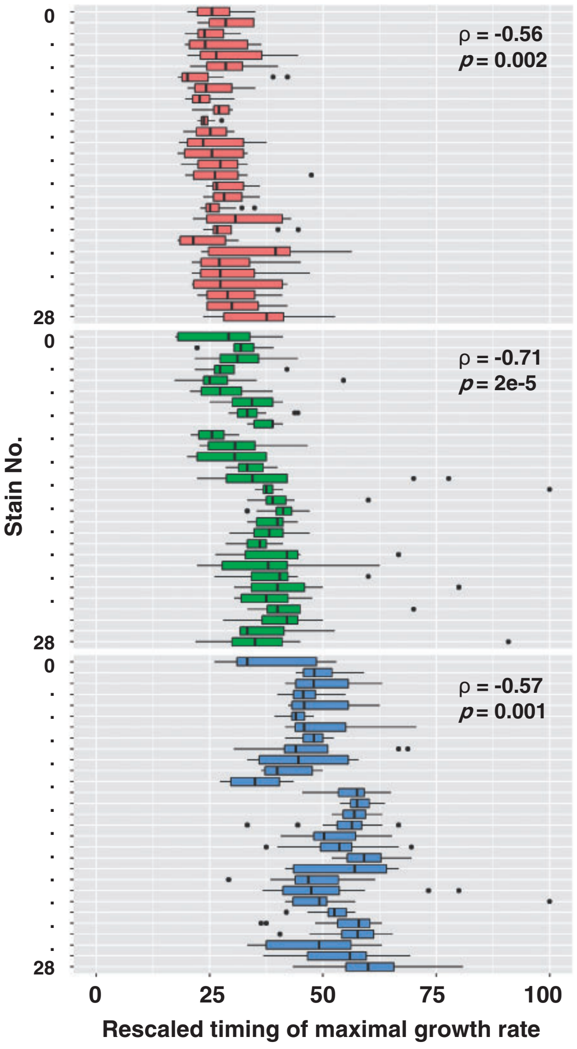

Figure 4. Rescaled timing of the maximal growth rate. The time required for the cell populations to achieve the best growth performance was calculated following the rescaling. Box plots represent every 11 to 30 repeated measurements for each condition. Panels from top to bottom represent growth in LB, MAA and M63, respectively. The genomes (Strain No.) from Nos. 0 to 28 are indicated in order from the top to bottom in each panel. Spearman's rank correlation coefficients between the mean timing and the length of accumulated deletions are indicated along with the statistical significance.

3.4 Accumulated deletions also caused correlated changes in growth dynamics

In addition to the speed (r) and the maximum (K) of population growth, the patterns of growth dynamics were also observed in correlated changes (Fig. 4) evaluated by rescaling the growth curves (Supplementary Fig. S5). The saturated cell density was rescaled in a common unit such that it had an increased effect on the low K and r in the reduced genomes. If the growth dynamics was similar, the rescaling would lead to the overlapped growth curves of all strains despite the variation in the raw growth curves (Supplementary Fig. S5). That is, the time required to achieve the maximal growth rate (τ), which was considered a representative parameter of growth dynamics, should be equivalent. However, the analysis resulted in high variation in the timing and the negative correlations between the averaged timing (τ) and the deleted length (L_d) in all three media ($P < 0.005$) (Fig. 4). The accumulative deletion caused the changes in growth patterns; i.e. the time from the growth initiation to the highest growth efficiency was increased, and the time from the maximal growth rate to the population saturation was decreased. Because growth is thought to be controlled by metabolic reactions, the global biochemical reactions achieved the highest efficacy more slowly and shut down more rapidly. These findings implied that the deleted sequences participated in metabolic efficiency in a dose-dependent manner. Hence, the qualities of life processes (e.g. growth fitness, metabolic activity) are associated with the quantity of genetic material (i.e. genomic length).

3.5 Growth prediction verified that genome reduction is the main factor underlying decreased growth

To determine the main factors involved in the growth decrease, growth prediction by multiple regression analysis was performed, in which five parameters were used:, the length of accumulated deletions (L_d), the population size (K), the culture media (M), the timing of growth maximal (τ) and the *ori-dif* angle (θ), a newly introduced parameter representing the symmetry in genome replication from *ori* to *dif* (Supplementary Fig. S6). Multiple deletions might disturb the symmetry of genome replication and cause growth changes. To avoid over regression, model selection among the five predictors (L_d, K, M, τ, θ) according to Akaike Information Criteria[35] was applied. The results showed that the most efficient regression (adjusted R^2, 0.9818) was achieved by choosing L_d, M and θ (Supplementary Table S3). Rejection of K and τ from the growth

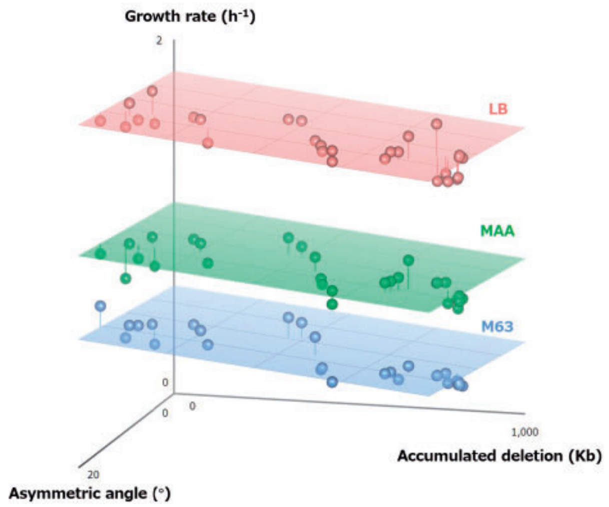

Figure 5. Multiple linear regression analysis of growth rate. Best prediction model of the growth rate (h^{-1}, *y*-axis) with the length of accumulated deletions (Kb, *x*-axis), the asymmetric angle (°, *z*-axis) and the growth media (LB, MAA and M63) is presented. Colour planes and spheres represent the predicted and the experimental values, respectively. Residuals between the predicted and the experimental values are represented by the lines from spheres to the corresponding plane.

prediction suggested the multicollinearity among K, τ and L_d, consistently with the previous discussion (Equations 2∼5 and Fig. 5).

Thus, the best regression for global effect on cell growth was defined as follows (Equation 6).

$$r = (-2.00e-4)L_d + 0.005\theta + \begin{cases} 1.741, & M = LB \\ 1.741 - 0.809, & M = MAA \\ 1.741 - 1.158, & M = M63 \end{cases} \tag{6}$$

The predicted growth rates were similar to those acquired experimentally (Fig. 4). Notably, the additional introduction of the angle to the regression model improved the prediction accuracy (Supplementary Table S1). However, the direct comparison between the angle and growth changes did not detect any significant correlation ($P > 0.1$, Supplementary Fig. S6). The slight but positive contribution of θ to r appeared to increase the prediction accuracy. Thus, the deterministic factors for growth fitness included genome reduction and growth conditions. Although the mechanism of the predictable growth changes was unclear and the theoretical relation might be strain- or reduction-mode-dependent, the present findings reveal a novel quantitative principle involved in the bacterial genome as a propagating machine.

3.6 *A hypothesis connecting genome reduction with growth changes*

Given that there are dispensable regions in bacterial genomes, harbouring the accessory genes involved in genome evolution,[41,42] we propose a hypothesis to explain the correlation between accumulative genomic reduction and gradual growth changes, particularly in severe environments (Fig. 6A). First, the length-dependent changes may be explained by considering genome evolution as a molecular clock.[43] Genome evolution should promote an enlarged genome size by incorporating non-essential accessory genes (e.g. horizontal gene transfer, HGT),[44,45] which might be disadvantageous for the growth fitness, given the additional cost for the replication and expression of the newly acquired sequences. Nevertheless, the evolutionary selection on genomic sequences was based on growth fitness. The enlarged genome size was subsequently optimized (e.g. mutagenesis) to generate novel genetic interactions for beneficial cooperation from the new sequences (i.e. accessory genes) to achieve the growth recovery (fitness increase). Thus, the wild type genome was the evolutionary consequence representing a balance between the additional cost and the novel cooperation, as well as a balanced distribution of accessory genes in genome. Given that genomic sequence elongation correlates with the evolutionary period (clock), the deletion-length dependent fitness decrease is reasonable. Second, multiple deletions of the evolved genomic sequence in the laboratory clearly disturbed the cooperative accessory networks formed during evolution, thus leading to decreased fitness regardless of the remaining growing capacity. Additional steps for sequence optimization or complete systematic reduction of accessory networks was required to retain the growth fitness of the small genome, thus potentially explaining why the reduced genome MDS42 grew as fast as the wild type genome did in minimal medium.[12,22,33] Because MDS42 exhibited a fixed periodicity for efficient growth in the transcriptome[33] and the sequence differences from the MG1655 counterpart ranging from single nucleotide substitutions to complete deletions,[46] we assumed that either a successful systematic loss of accessory networks or the sequence optimization owing to the serial transfer was achieved.

A supportive line of evidence was that the gene function correlation analysis did not detect any single gene category that was specifically responsible for the changes in population growth (Fig. 6B). We analysed whether any gene categories specifically contributed to the decrease in growth rate. The gene categories comprising >10 deleted genes in strain No. 28 (independent of the significance evaluated in Supplementary Fig. S1) were subjected to the evaluation. The correlations between the decreasing growth rates from strain No. 1 to 28 (Fig. 2A) and the increasing numbers of deleted genes in each gene categories (Supplementary Table S1, 14 out of 23 gene categories) were evaluated. High significance ($P < 0.001$) was detected in all 14 gene categories under the conditions of M63 and LB, thus indicating that all gene categories contributed to the growth decrease in both rich and poor nutritional conditions. The addition of 20 amino acids (MAA) decreased the significance of such correlation, thus suggesting that these amino acids compensated for the loss of the genes broadly classified in not only the central gene categories of Enzyme (e), Structural component (s), and Regulator (r), but also the accessory gene categories of Unknown function (o), Phage/IS (h), Pseudogenes (su) and predicted functions (pf, pr). Although a single accessory gene might have an undetectable contribution to the growth (∼0.03–0.001%), a cluster of these genes might reach a quantitatively predictive level regardless of the variation in gene categories and/or functional mechanisms. Note that other gene function analyses failed to acquire any significant correlated contribution, which was largely due to the missing functional assignment of the deleted accessory genes (Supplementary Table S2). The broad but not specific correlations between the gene category and the growth rate not only explained the finding of that the growth rate could be predicted on the basis of the accumulatively deleted length alone but also supported the assumption in which evolutionary sequence optimization can build the functional interactions involving accessory genes and thus result in increased fitness.

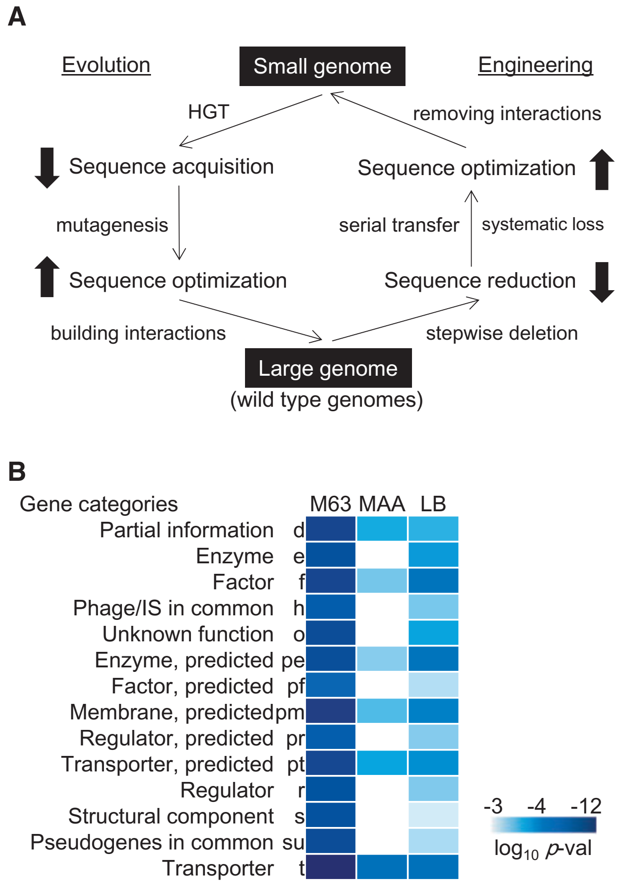

Figure 6. Hypothesis for genome reduction-correlated growth changes. (A) Perspective of genome size. Left and right flowcharts illustrate the processes of genome evolution and genome reduction, respectively. Upward and downward bold arrows indicate the increased and decreased changes in growth fitness, respectively. A detailed explanation is provided in the main text. (B) Correlation between the growth rate and the number of deleted genes. The deleted genes were clustered according to the gene categories. Statistical significance of the correlation coefficients between the growth rates and the numbers of deleted genes in each gene categories is presented as a heat map with a logarithmic scale. The progression from dark to light represents the significance from high to low, and the white colour represents no significance. M63, MAA and LB indicate the growth media. Gene categories are noted in both full names and abbreviations.

In summary, the accumulative loss of dispensable genomic sequences contributes to bacterial growth in a dose-dependent manner, significantly in poor conditions. The data sets and the theoretical underpinnings suggest a quantitative linkage between genomic sequences and growth fitness and should be valuable in understanding genome evolution and constructing the desired genome for a growing system.

Acknowledgements

We thank the National BioResource Project and the National Institute of Genetics (Shizuoka, Japan) for providing the *E. coli* strains and our lab members for fruitful discussions.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This work was partially supported by a Grant-in-Aid for Scientific Research (C) no. 26506003 (to BWY) and a Grant-in-Aid for Scientific Research on Innovative Areas no. 26119702 (to BWY) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. Funding to pay the Open Access publication charges for this article was provided by the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Koob, M.D., Shaw, A.J. and Cameron, D.C. 1994, Minimizing the genome of *Escherichia coli*. Motivation and strategy, *Ann. N. Y. Acad. Sci.*, **745**, 1–3.
2. Blattner, F.R., Plunkett, G. 3rd, Bloch, C.A., et al. 1997, The complete genome sequence of *Escherichia coli* K-12, *Science*, **277**, 1453–62.
3. Kolisnychenko, V., Plunkett, G. 3rd, Herring, C.D., et al. 2002, Engineering a reduced *Escherichia coli* genome, *Genome Res.*, **12**, 640–7.
4. Yu, B.J., Sung, B.H., Koob, M.D., et al. 2002, Minimization of the *Escherichia coli* genome using a Tn5-targeted Cre/loxP excision system, *Nat. Biotechnol.*, **20**, 1018–23.
5. Feher, T., Karcagi, I., Gyorfy, Z., Umenhoffer, K., Csorgo, B. and Posfai, G. 2008, Scarless engineering of the *Escherichia coli* genome, *Methods Mol. Biol.*, **416**, 251–9.
6. Feist, A.M., Henry, C.S., Reed, J.L., et al. 2007, A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information, *Mol. Syst. Biol.*, **3**, 121.
7. Mizoguchi, H., Mori, H. and Fujio, T. 2007, *Escherichia coli* minimum genome factory, *Biotechnol. Appl. Biochem.*, **46**, 157–67.
8. Hutchison, C.A. 3rd, Chuang, R.Y., Noskov, V.N., et al. 2016, Design and synthesis of a minimal bacterial genome, *Science*, **351**, aad6253.
9. Xavier, J.C., Patil, K.R. and Rocha, I. 2014, Systems biology perspectives on minimal and simpler cells, *Microbiol. Mol. Biol. Rev.*, **78**, 487–509.
10. Baba, T., Ara, T., Hasegawa, M., et al. 2006, Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection, *Mol. Syst. Biol.*, **2**, 2006 0008.
11. Yamamoto, N., Nakahigashi, K., Nakamichi, T., et al. 2009, Update on the Keio collection of *Escherichia coli* single-gene deletion mutants, *Mol. Syst. Biol.*, **5**, 335.
12. Posfai, G., Plunkett, G. 3rd, Feher, T., et al. 2006, Emergent properties of reduced-genome *Escherichia coli*, *Science*, **312**, 1044–6.
13. Kato, J. and Hashimoto, M. 2007, Construction of consecutive deletions of the *Escherichia coli* chromosome, *Mol. Syst. Biol.*, **3**, 132.
14. Mizoguchi, H., Sawano, Y., Kato, J. and Mori, H. 2008, Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome, *DNA Res.*, **15**, 277–84.
15. Gibson, D.G., Glass, J.I., Lartigue, C., et al. 2010, Creation of a bacterial cell controlled by a chemically synthesized genome, *Science*, **329**, 52–6.
16. Feher, T., Papp, B., Pal, C. and Posfai, G. 2007, Systematic genome reductions: theoretical and experimental approaches, *Chem. Rev.*, **107**, 3498–513.
17. Pal, C., Papp, B. and Posfai, G. 2014, The dawn of evolutionary genome engineering, *Nat. Rev. Genet.*, **15**, 504–12.
18. Yus, E., Maier, T., Michalodimitrakis, K., et al. 2009, Impact of genome reduction on bacterial metabolism and its regulation, *Science*, **326**, 1263–8.
19. Sharma, S.S., Campbell, J.W., Frisch, D., Blattner, F.R. and Harcum, S.W. 2007, Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains, *Biotechnol. Bioeng.*, **98**, 1056–70.
20. Morimoto, T., Kadoya, R., Endo, K., et al. 2008, Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*, *DNA Res.*, **15**, 73–81.

21. Lee, J.H., Sung, B.H., Kim, M.S., et al. 2009, Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production, *Microb. Cell Factories*, **8**, 2.
22. Akeno, Y., Ying, B.W., Tsuru, S. and Yomo, T. 2014, A reduced genome decreases the host carrying capacity for foreign DNA, *Microb. Cell Factories*, **13**, 49.
23. Csorgo, B., Feher, T., Timar, E., Blattner, F.R. and Posfai, G. 2012, Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs, *Microb. Cell Factories*, **11**, 11.
24. Umenhoffer, K., Feher, T., Baliko, G., et al. 2010, Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications, *Microb. Cell Factories*, **9**, 38.
25. Sabath, N., Ferrada, E., Barve, A. and Wagner, A. 2013, Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation, *Genome Biol. Evol.*, **5**, 966–77.
26. Karcagi, I., Draskovits, G., Umenhoffer, K., et al. 2016, Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining, *Mol. Biol. Evol*, **33**, 1257–69.
27. Nahku, R., Valgepea, K., Lahtvee, P.J., et al. 2010, Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures, *J. Biotechnol.*, **145**, 60–5.
28. Ying, B.W., Yama, K., Kitahara, K. and Yomo, T. 2016, The *Escherichia coli* transcriptome linked to growth fitness, *Genomics Data*, **7**, 1–3.
29. Murakami, Y., Matsumoto, Y., Tsuru, S., Ying, B.W. and Yomo, T. 2015, Global coordination in adaptation to gene rewiring, *Nucleic Acids Res.*, **43**, 1304–16.
30. Matsumoto, Y., Murakami, Y., Tsuru, S., Ying, B.W. and Yomo, T. 2013, Growth rate-coordinated transcriptome reorganization in bacteria, *BMC Genomics*, **14**, 808.
31. Ying, B.W., Matsumoto, Y., Kitahara, K., et al. 2015, Bacterial transcriptome reorganization in thermal adaptive evolution, *BMC Genomics*, **16**, 802.
32. Hashimoto, M., Ichimura, T., Mizoguchi, H., et al. 2005, Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome, *Mol. Microbial.*, **55**, 137–49.
33. Ying, B.W., Seno, S., Kaneko, F., Matsuda, H. and Yomo, T. 2013, Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome, *BMC Genomics*, **14**, 25.
34. Ishizawa, Y., Ying, B.W., Tsuru, S. and Yomo, T. 2015, Nutrient-dependent growth defects and mutability of mutators in *Escherichia coli*, *Genes Cells*, **20**, 68–76.
35. Akaike, H. 1974, A new look at the statistical model identification, *IEEE Trans. Automatic Control*, **19**, 8.
36. Aho, K., Derryberry, D. and Peterson, T. 2014, Model selection for ecologists: the worldviews of AIC and BIC, *Ecology*, **95**, 631–6.
37. Ihaka, R. and Gentleman, R. 1996, R: a language for data analysis and graphics, *J. Comput. Graph. Stat.*, **5**, 299–314.
38. Fox, J. and Weisberg, S. 2011, *An R Companion to Applied Regression*, Sage, Thousand Oaks CA.
39. Riley, M., Abe, T., Arnaud, M.B., et al. 2006, *Escherichia coli* K-12: a cooperatively developed annotation snapshot–2005, *Nucleic Acids Res.*, **34**, 1–9.
40. Alikhan, N.F., Petty, N.K., Ben Zakour, N.L. and Beatson, S.A. 2011, BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons, *BMC Genomics*, **12**, 402.
41. Lawrence, J.G. and Ochman, H. 1998, Molecular archaeology of the *Escherichia coli* genome, *Proc. Natl. Acad. Sci. U. S. A.*, **95**, 9413–7.
42. Puigbo, P., Lobkovsky, A.E., Kristensen, D.M., Wolf, Y.I. and Koonin, E.V. 2014, Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes, *BMC Biol.*, **12**, 66.
43. Novichkov, P.S., Omelchenko, M.V., Gelfand, M.S., Mironov, A.A., Wolf, Y.I. and Koonin, E.V. 2004, Genome-wide molecular clock and horizontal gene transfer in bacterial evolution, *J. Bacteriol.*, **186**, 6575–85.
44. Koonin, E.V. and Wolf, Y.I. 2008, Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world, *Nucleic Acids Res.*, **36**, 6688–719.
45. Kurland, C.G., Canback, B. and Berg, O.G. 2003, Horizontal gene transfer: a critical view, *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 9658–62.
46. Ying, B.W.Yomo, T. 2016, Comparative analyses of bacterial transcriptome reorganisation in response to temperature increase. In: F.D., Bruijn (ed.), *Stress and Environmental Control of Gene Expression in Bacteria*, Wiley-VCH, Weinheim.