

RESEARCH ARTICLE

Open Access

# Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome

Bei-Wen Ying<sup>1†</sup>, Shigeto Seno<sup>1†</sup>, Fuyuro Kaneko<sup>1</sup>, Hideo Matsuda<sup>1</sup> and Tetsuya Yomo<sup>1,2,3\*</sup>

## Abstract

**Background:** Both large deletions in genome and heat shock stress would lead to alterations in the gene expression profile; however, whether there is any potential linkage between these disturbances to the transcriptome have not been discovered. Here, the relationship between the genomic and environmental contributions to the transcriptome was analyzed by comparing the transcriptomes of the bacterium *Escherichia coli* (strain MG1655 and its extensive genomic deletion derivative, MDS42) grown in regular and transient heat shock conditions.

**Results:** The transcriptome analysis showed the following: (i) there was a reorganization of the transcriptome in accordance with preferred chromosomal periodicity upon genomic or heat shock perturbation; (ii) there was a considerable overlap between the perturbed regulatory networks and the categories enriched for differentially expressed genes (DEGs) following genome reduction and heat shock; (iii) the genes sensitive to genome reduction tended to be located close to genomic scars, and some were also highly responsive to heat shock; and (iv) the genomic and environmental contributions to the transcriptome displayed not only a positive correlation but also a negatively compensated relationship (*i.e.*, antagonistic epistasis).

**Conclusion:** The contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome were evaluated at multiple levels. The observations of overlapping perturbed networks, directional similarity in transcriptional changes, positive correlation and epistatic nature linked the two contributions and suggest somehow a crosstalk guiding transcriptional reorganization in response to both genetic and environmental disturbances in bacterium *E. coli*.

**Keywords:** Transcriptome, Negative epistasis, Genome reduction, Chromosomal periodicity, Regulatory network, Transcriptional change, Genomic interruption, Environmental perturbation, Heat shock, Directionality

## Background

The bacterial transcriptome, as one of the important genome-level information, reflects the global cellular activity (*i.e.*, fitness). Since the fitness and/or activity of the cells is directly influenced by the environmental perturbation (*e.g.*, heat or cold stress) and genetic disturbance (*e.g.*, mutations or deletions in genome), the evaluation

of the environmental and genetic contributions to transcriptome turns to be highly essential. In recent, the transcriptome analysis (*i.e.*, gene expression profiling) of bacteria (*e.g.*, *Escherichia coli*) has been performed extensively using a range of experimental and analytical tools [1-7], to illustrate an overall picture of the biochemical events occurring inside the cells. In particular, to examine the environmental contributions to bacterial transcriptome, gene expression profiling has been tested across multiple environmental conditions in a high-throughput manner [8-10]. As a consequence, the intensive studies on global transcriptional activity contributed by environmental perturbations successfully provided conclusions regarding the molecular mechanisms involved

\* Correspondence: yomo@ist.osaka-u.ac.jp

†Equal contributors

<sup>1</sup>Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

<sup>2</sup>Graduate School of Frontier Biosciences, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan

Full list of author information is available at the end of the article

in specific pathways and/or regulatory networks [9,11-14], as well as the theoretical and/or overall outlook in regard to global dynamic changes in transcriptome [15].

However, the genetic contribution to bacterial transcriptome has been rarely reported. Although the previous studies revealed the potential links between transcriptional regulation and chromosome organization [16] and the relation between changes in DNA methylation and chromosomal structure and changes in global gene expression [3,5], whether and how the genomic scars (*e.g.*, deletions) influence the global transcriptional activity is unclear.

Both genomic and environmental perturbations could initiate a global change in the transcriptome, because both are stressful for living cells. Whether there is any potential linkage between the two contributions is an intriguing question. To address the question, a comprehensive analysis comparing the effects of genomic and environmental interruptions on the transcriptome is highly required. Such analysis would not only enable a comparison of the internal and external influences on the plasticity of genome-wide transcriptional activity but would also provide a valuable example of a global and conceptual assessment of high-throughput biological data. Therefore, a multi-level survey of genomic expression is highly in demand, to show how genomic disturbance causes the global transcriptional reorganization and whether there is any relationship between the contributions of genomic and environmental interruptions to transcriptome.

As a pilot study, we evaluated these two distinct entities, the genome and the environment, at the level of the transcriptome. We addressed how genomic and environmental perturbations contribute to the bacterial transcriptome and whether there is any interaction between the two influences. A multiple deletion *E. coli* strain, MDS42 [17], was compared with its wild-type parent strain MG1655 [18,19] in this study. Because of its lack of insertion sequences (ISs) [17,20], MDS42 has been widely used in a number of applications [21-25]; however, its detailed, genome-wide analysis has been rarely reported [21-25]. The genes used for the comparative transcriptome analysis were selected on the basis of the genome sequences of MDS42 and MG1655. The comparison of MG1655 to MDS42 provided insights into genomic disturbance-induced transcriptional reorganization. The evaluation of the heat shock response (a transient response to an elevated temperature) provided insights into environmental perturbation-induced transcriptional changes. We examined whether the losses of insertion elements in genome interrupt the genome-wide transcriptional activity (*i.e.*, the transcriptome) and whether these multiple deletions in genome and the heat shock stress exert common or different effects on transcriptional

changes. Furthermore, we performed multilevel comparative analyses of chromosomal periodicity, regulatory networks, the localization of differential gene expression, transcriptional sensitivity in response to genomic and environmental perturbations and the epistatic effect, that is, a negatively compensated relationship, on the transcriptome of the genome and the environment.

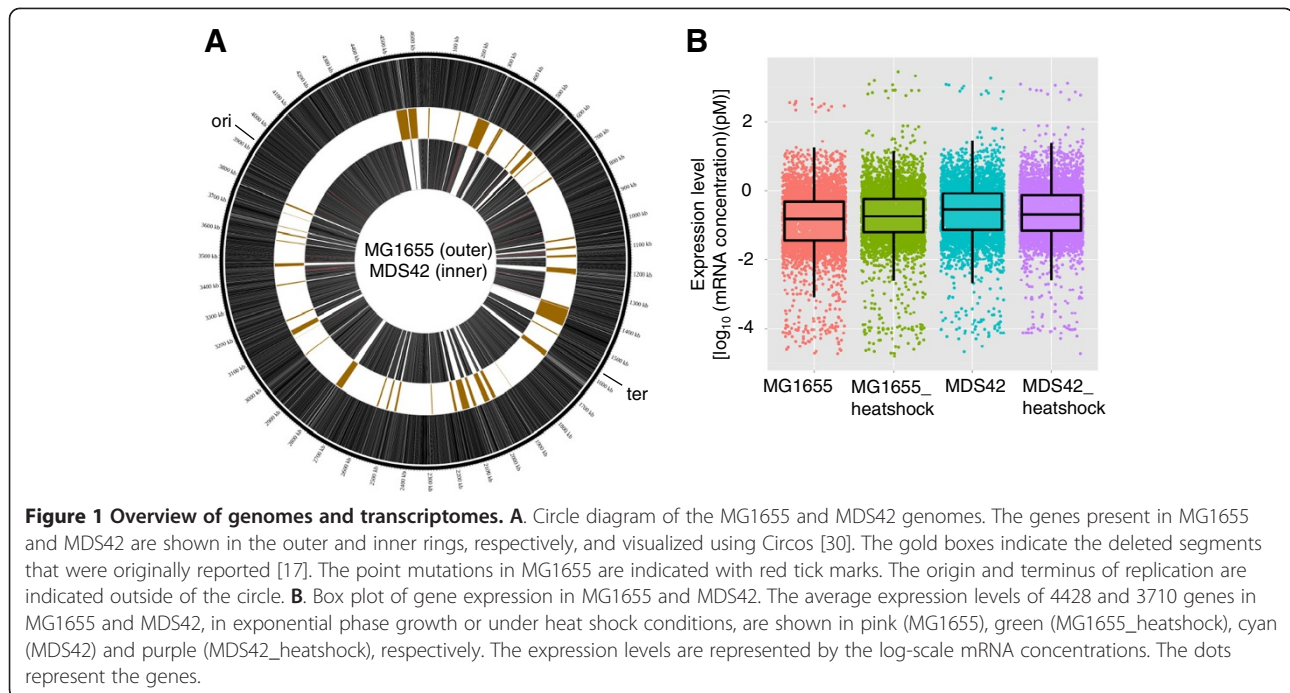
## Results

### Strains and genes subjected to transcriptome analysis

Gene expression was analyzed using a custom, high-density DNA microarray (a single nucleotide tiling array) with a precise and wide quantitative range of mRNA concentrations [26,27], as previously described for genome resequencing [28] and transcriptome analysis [29]. First, the wild-type strain MG1655 and the multiple deletion strain MDS42 [17] were used to examine whether and how removing the nonessential genomic regions in *E. coli* contributed to its transcriptome. Repeated experiments were performed with cells grown precisely to mid-exponential phase in minimal medium. The genes shared between the two strains were determined by comparing their genome sequences (MDS42: DDBJ No. AP012306; MG1655: GenBank No. U00096). In total, locus tags (b numbers) were assigned to 4485 and 3778 open reading frames (ORFs) in MG1655 and MDS42, respectively (Figure 1A), which was consistent to the original report [17]. Genes with repeated locus tags were removed from the analysis, for a final total of 4428 genes with locus tags in MG1655. Excluding the 696 deleted genes [17] and 22 mutated genes in MG1655 (determined by comparing the genome sequences, see the Materials and Methods), a total of 3710 genes were determined to be common between the two strains. Subsequently, as the heat shock stress is known as one of many external perturbations and showed distinguishable expression profiles [9,11], heat shock experiments were performed to evaluate how the external disturbance contributed to the transcriptome for comparative analyses. The average gene expression levels from repeated experiments were plotted in Figure 1B and used for further analysis. We showed that neither genomic reduction nor heat shock altered the shape of the distribution of gene expression (Figure 1B).

### Priority in chromosomal periodicity

The periodicity of genome-wide transcriptional activity was studied to provide a global view of gene expression profiling [4,6,31,32]. The average transcription levels (*i.e.*, mRNA concentrations) over the entire genome were plotted for both genomes using a 100-bp sliding window (Figure 2). The periodic patterns showed a major peak of approximately 663 kb (Figure 2A) in MG1655, consistent with previous reports [4,32]. However, these periodic



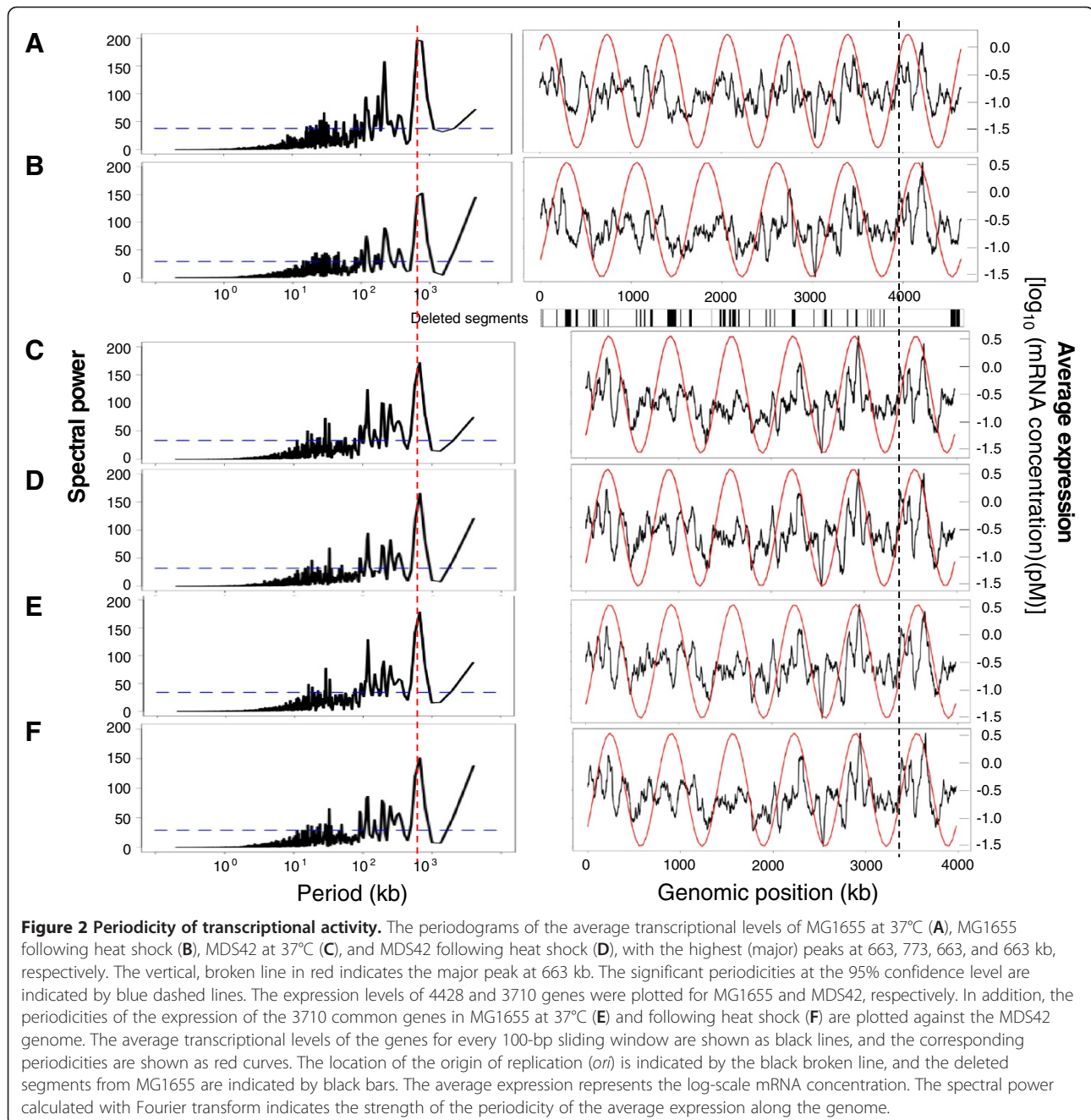
patterns were disturbed by heat shock. The most significant wavelength, as indicated by the red, vertical, broken line in Figure 2 (left panels), shifted from 663 to 773 kb (to the right of the red line, a close-up view can be found in Additional file 1: Figure S4). This slight but statistically significant change in the major peak caused a reduction in the number of periods from seven to six (Figure 2A–B, right panels). This finding indicated that the heat shock treatment initiated a reorganization of the transcriptome that gave a high priority to a periodicity of six periods.

Interestingly, the reduced genome also showed a fixed periodicity of six periods. MDS42 maintained the parental periodogram, with a major peak at 663 kb that was accompanied by a decrease in the chromosomal period number (Figure 2C). This reduced period was unchanged in MDS42 in response to heat shock perturbation (Figure 2D). In addition, when the expression of the 3710 common genes in MG1655 was plotted against the MDS42 genome, an identical chromosomal periodicity of six periods was observed independently of external conditions (Figure 2E and F). Statistical analyses confirmed that the altered periodicity was highly specific for the MDS42 genome and was not due to random deletions in the genome ( $P < 0.001$ ). The stable six periods may have been caused by the presence of six macrodomains in the chromosomal structure, as determined by recombination activity [33]. The results suggested that the absent genetic regions in MDS42 (Figure 1A, gold boxes) may contain a considerable number of genes that are not or rarely transcribed under heat shock conditions, potentially contributing to an additional period. Thus, the multiple

deletions not only reduced the genome size but also provided, probably by chance, a steady periodicity to the global transcriptional activity of the *E. coli* chromosome.

#### Overlaps in perturbed regulatory networks

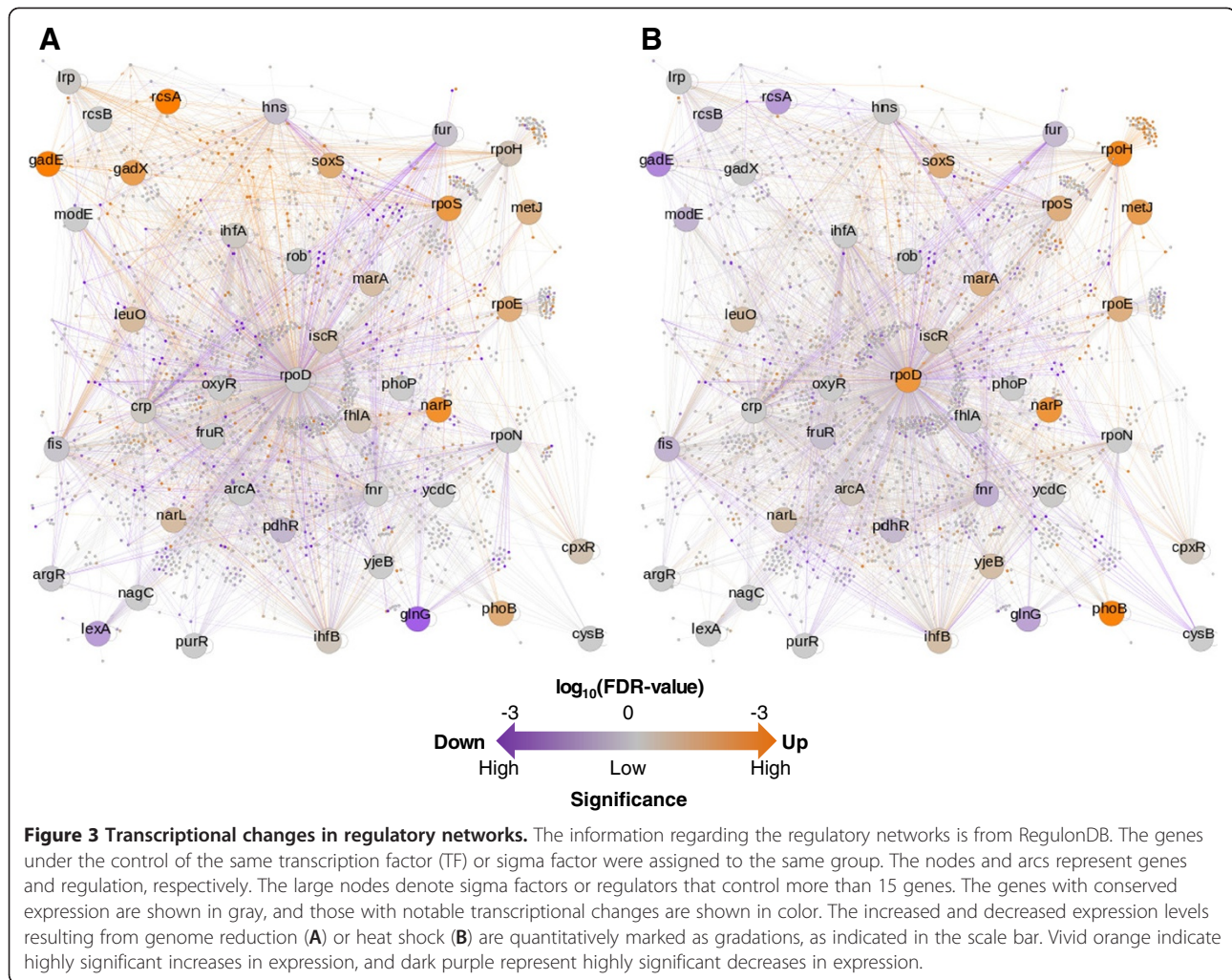
Regulatory network maps comprising the transcription factors (TFs, or regulators) that control more than 15 genes (42 of 175 regulatory networks) were constructed as shown in Figure 3. Significant changes in gene expression are represented as gradations according to their FDR values, which were determined using the rank product method [34,35]. Overall, both genome reduction and heat shock exerted broad effects on transcriptional reorganization: the expression levels of a number of genes were changed (Figure 3). A small number of regulators showed remarkable, independent transcriptional changes; for example, *gadX* was induced only by genome reduction (A) and *rpoD* only by heat shock (B). It indicated that genome structural changes might activate the primary activator of the *gad* system but fairly influenced the global transcriptional factor (sigma factor). In addition, a comparable number of regulators were disturbed by both perturbations in a correlated manner: those induced by genome reduction were also induced by heat shock. The changes in these overlapping perturbed regulators occurred in either the same or the reverse direction; e.g., the expression level of *phoB* was upregulated by both genome reduction and heat shock (orange in both A and B), whereas, the expression level of *gadE* was upregulated by genome reduction but repressed by heat shock (orange in A, purple in B).



To identify these regulators (regulatory networks) upstream of genes with marked changes in transcription induced by either genome reduction or heat shock, a gene set enrichment analysis (GSEA) was applied [36]. Among the 175 regulatory networks controlled by 5 sigma factors and 170 TFs, 24 and 12 networks showed clear transcriptional changes (FDR < 0.25) associated with genome reduction and heat shock, respectively (Table 1; details in Additional file 2: Table S1). Heat shock interrupted approximately equal numbers of the induced and repressed regulatory networks (7 vs. 5),

whereas genome reduction triggered many more upregulated than downregulated networks (22 vs. 2). Moreover, most of the regulatory networks that responded to heat shock were highly sensitive to genome reduction as well (9 out of 12). These overlapping regulatory networks (e.g., *rpoH* and *purR*) showed a marked tendency to undergo transcriptional changes in the same direction (i.e., increased or decreased) in response to genomic and heat shock perturbations. The exception (with changes in opposite directions) was found in the network controlled by *gadX*, where the downstream genes were





induced by genome reduction but suppressed by heat shock (FDR < 0.02).

Taken together, the transcriptional reorganization observed in the regulatory networks was not only secondary to the expression changes in the regulators (e.g., *rpoH*) but also occurred in networks controlled by the regulators of steady expression (e.g., *purR*). The latter case suggested either weakened transcriptional control by the regulators or a high level of sensitivity in the downstream genes (i.e., slight transcriptional changes in the regulators resulted in profound transcriptional changes in the downstream genes). The discovery of regulatory networks (and/or regulators) that were universally involved in both perturbations indicated that transcriptional reorganization responds similarly to both endogenous and exogenous disruptions.

#### Genome location dependency and gene category enrichment of highly responsive genes

Genes with significant transcriptional changes (differentially expressed genes, DEGs) among the 3710 common

genes were also determined with the rank product method. In all, 159 and 95 genes were identified that showed either up- or downregulated expression in response to genome reduction (DEGs<sub>gr</sub>) or heat shock (DEGs<sub>hs</sub>), respectively (FDR < 0.001). The gene names, gene categories, locus tags, and directions of the transcriptional changes are summarized in Additional file 3: Table S2, and the genome locations are marked in Additional file 1: Figure S1. An analysis of the distances between the DEGs and the genome scars caused by genome reduction showed that the DEGs<sub>gr</sub>, but not the DEGs<sub>hs</sub>, were located much closer to the nearest genome scar than were the genes with conserved expression (nonDEGs) ( $P < 0.001$ ) (Figure 4A and B). This tendency indicated that the genomic disturbance caused by genome reduction brought about considerable transcriptional changes in neighboring genes. In particular, the transcriptional changes were highly significant in the direction of upregulation (Figure 4A, orange), which suggests that the removal of the dispensable genetic segment might increase the expression efficiency of the neighboring genes.

**Table 1 Regulatory networks with significant perturbations**

Perturbation	Direction	Locus tag	Name of TF	Size	FDR q-val		
Genome reduction	Increased	b1951	rcaA	26	0.0000		
		b2217	rcaB	29	0.0000		
		b3938	metJ	15	0.0000		
		b3516	gadX	25	0.0000		
		b3512	gadE	49	0.0000		
		b0889	lrp	83	0.0038		
		b3461	rpoH	150	0.0065		
		b0761	modE	45	0.0061		
		b1237	hns	136	0.0145		
		b3261	fis	216	0.0513		
		b2193	narP	49	0.0870		
		b1221	narL	113	0.1043		
		b3912	cpxR	51	0.0972		
		b2531	iscR	26	0.1580		
		b0399	phoB	36	0.1572		
		b1712	ihfA	191	0.1477		
		b2741	rpoS	215	0.1448		
		b0912	ihfB	191	0.1398		
		b0076	leuO	20	0.1703		
		b1130	phoP	48	0.1627		
		b1334	fnr	272	0.1687		
		b1531	marA	36	0.2127		
		Genome reduction	Decreased	b1658	purR	31	0.0071
b0683	fur			81	0.0201		
Heat shock	Increased	b3461	rpoH	150	0.0000		
		b3912	cpxR	51	0.0004		
		b3938	metJ	15	0.0026		
		b2193	narP	49	0.0837		
		b0399	phoB	36	0.1384		
		b0761	modE	45	0.2134		
		b0076	leuO	20	0.2345		
		Heat shock	Decreased	b1658	purR	31	0.0000
				b1275	cysB	19	0.0000
				b0683	fur	81	0.0000
				b3516	gadX	25	0.0123
		b3868	glgG	44	0.0118		

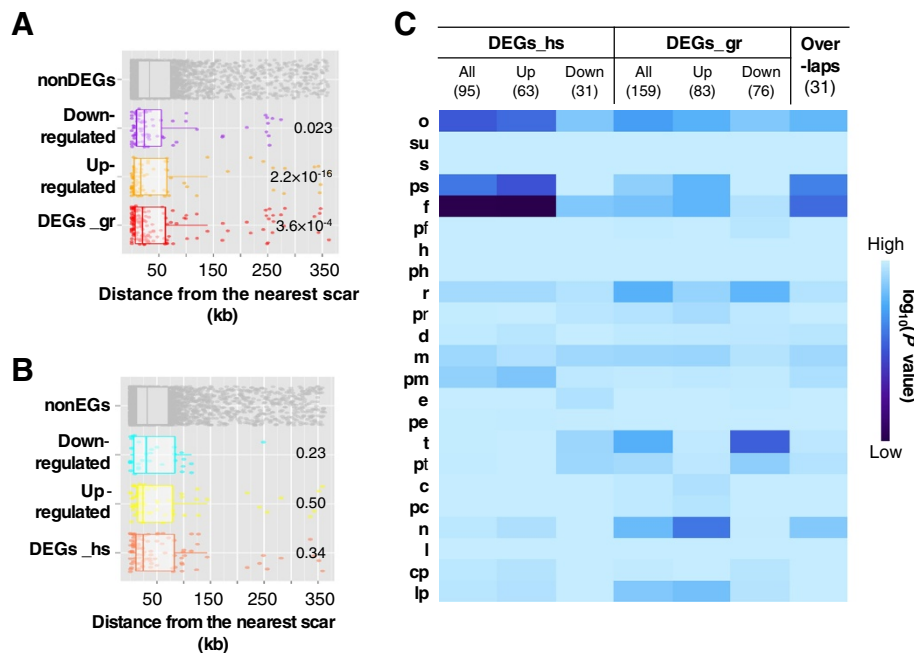
The information pertaining to transcription factors (TFs), sigma factors and the downstream genes under their control is from RegulonDB. The transcriptional changes in the downstream genes were evaluated with gene set enrichment analysis (GSEA) [36], and the disturbed regulatory networks were determined accordingly (FDR < 0.25). Perturbation and Direction refer to the type of disturbance to the transcriptome and the directionality of the changes in gene expression, respectively. Locus tag, Name of TF, Size and FDR q-val represent the gene ID, the corresponding regulator gene, the number of regulated downstream genes and the statistical significance, respectively.

We assumed that the deleted genes most likely perturbed the local superhelical density and thereby affected the transcriptional activity of the neighboring genes, although the functions of the deleted genes near the deletion junctions or mutations remain to be further examined.

The gene categories that were enriched in the DEGs were analyzed to address the question of whether the genomic and heat shock perturbations had affected any specific gene categories (details in Additional file 3: Table S2 and Additional file 4: Table S3). The heat map (Figure 4C) illustrates that the heat shock affected only a limited number of gene categories (*i.e.*, Unknown function and Factor) with high significance ( $P < 0.001$ ). However, the genome reduction influenced a relatively large number of gene categories with diverse directions in transcriptional changes; for example, the transporter genes (t) and the RNA-coding genes (n) were highly enriched in the DEGs<sub>gr</sub> in the directions of down- ( $P < 0.001$ ) and upregulation ( $P < 0.002$ ), respectively. Only the “Unknown function” (o) category was significantly enriched in both the DEGs<sub>gr</sub> and the DEGs<sub>hs</sub> (All,  $P < 0.05$ ). This result suggested that the genes with unknown functions were highly sensitive to both genomic and environmental perturbations, as might result from a loose control of gene expression. The genes in the Factor category (f), of which most are heat shock proteins (chaperones), were not only enriched in the upregulated DEGs<sub>hs</sub> (Up,  $P < 0.001$ ), as extensively reported, but were also detected in the upregulated DEGs<sub>gr</sub> (Up,  $P < 0.05$ ). As a result, Factor (f) was the most highly enriched category among the 31 overlapping genes ( $P < 0.001$ ). Intriguingly, some well-known molecular chaperones, such as *clpB* and *ibpA*, were highly responsive to both perturbations (Additional file 3: Table S2), indicating a universal response mechanism to internal and external stressors in these genes.

#### Positive correlations in changes of gene expression

The directions of the transcriptional changes demonstrated by the DEGs were further analyzed, in particular, on the gene expression patterns by means of correlations. In a global view, the DEGs<sub>gr</sub> (Figure 5A, upper) also showed remarkable transcriptional changes in response to heat shock (Figure 5B, upper), as evidenced by the significantly lower correlation of the DEGs<sub>gr</sub> (0.672) compared with that (0.876) of the 3710 common genes ( $P < 0.001$ , Additional file 1: Figure S2). Similarly, the DEGs<sub>hs</sub> showed large transcriptional changes in response to genomic perturbation (Figure 5B, bottom); their correlation (0.893) was significantly lower than that (0.968) of the 3710 common genes ( $P < 0.001$ , Additional file 1: Figure S2). Furthermore, the upregulated DEGs<sub>hs</sub> (Figure 5A, yellow) were also induced by genome reduction (Figure 5B, yellow), and the upregulated DEGs<sub>gr</sub> (Figure 5A, orange) were also induced by heat shock



**Figure 4 Differentially expressed genes (DEGs).** A-B. Distance between genes and genomic scars. The distance from each gene to the nearest genomic scar resulting from genome reduction is plotted. The genes with conserved (nonDEGs), increased, or reduced expression caused by genome reduction are shown in gray, orange, and purple, respectively (A). The genes with conserved (nonDEGs), upregulated, or downregulated expression due to heat shock are shown in gray, yellow and cyan, respectively (B). All the genes of the DEGs\_gr and DEGs\_hs are shown in red and brown, respectively. Significance was determined with the Wilcoxon-Mann-Whitney test, and the P-values are indicated. C. Enriched gene categories. The DEGs were divided into 23 gene categories, according to a previous report [37]. One-tailed binominal tests were performed to determine the significance of the enrichment of each category. The log-scale P-values are displayed as a heat map with gradations of blue; darker blue indicates higher significance.

(Figure 5B, orange). These correlated patterns of changes in gene expression deduced from the overall evaluation of the DEGs (Figure 5A and B) implied that the overlapping genes (Figure 4C) might play a crucial role in producing such transcriptional fluctuation because not all the DEGs changed significantly when moving to the other condition. Among the 31 overlapping genes (Figure 4C, Additional file 4: Table S3), 20 and 10 genes with up- or downregulated expression in response to genome reduction were up- or downregulated by heat shock in the same direction, respectively. Only a single gene (*yafF*, a conserved protein) reversed the direction of its transcription; *i.e.*, it was induced by heat shock but repressed by genome reduction. The overall directional similarities implied that the genes that were sensitive to genomic disturbance tended to be highly responsive to external perturbation and *vice versa*.

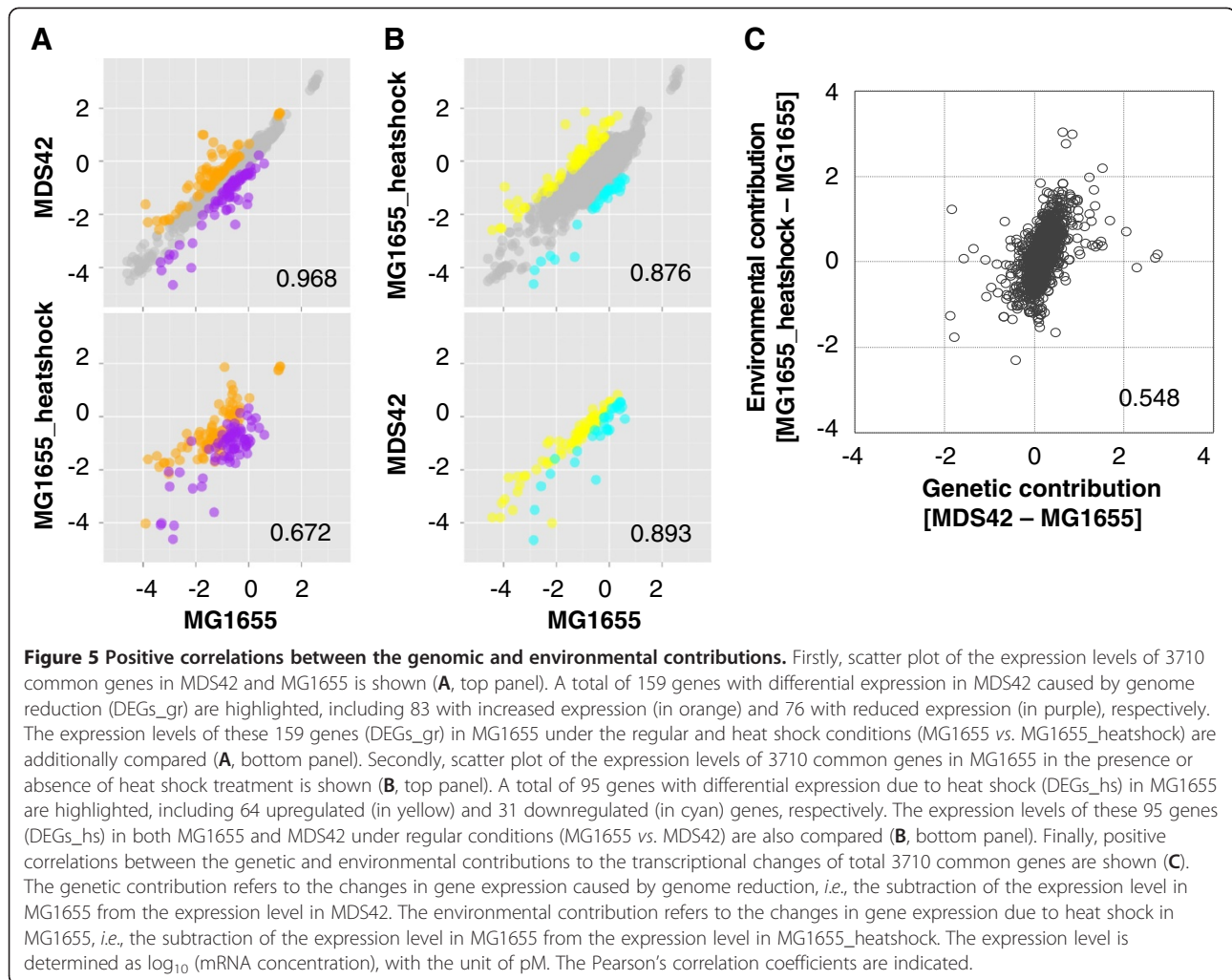
Furthermore, the transcriptional changes of the 3710 genes that were induced by genome reduction were plotted against those that were induced by heat shock (Figure 5C). A clear, positive correlation between the transcriptional changes produced by the genetic and environmental contributions was observed ( $P < 0.001$ , Additional file 1: Figure S2). In other words, larger degrees of

transcriptional fluctuation triggered by genetic interruption were correlated with greater changes in gene expression induced by environmental perturbation. The positive correlation between these two contributions to the transcriptome implies somehow a linkage in transcriptional reorganization in response to both genetic and environmental stresses.

#### Antagonistic epistasis in the transcriptome

Theoretically, the positive correlation in the transcriptional changes triggered by the genetic and environmental interruptions (Figure 5C) potentially had a tendency to amplify the magnitude of the changes in gene expression that occurred in response to both genetic and environmental perturbations. For instance, the gene expression levels of *dnaJ* were approximately 2.6 and 30 folds up-regulated due to genome reduction and heat shock, respectively. Owing to the positive correlation, the change in the expression of *dnaJ* was assumed to be multiplied, up to 78 folds, by the two simultaneously occurred perturbations. Actually, it was a 21-fold increase, a highly suppressed value, in the transcriptional change. To understand such repressed change in gene expression, the genetic concepts of epistasis and additivity were



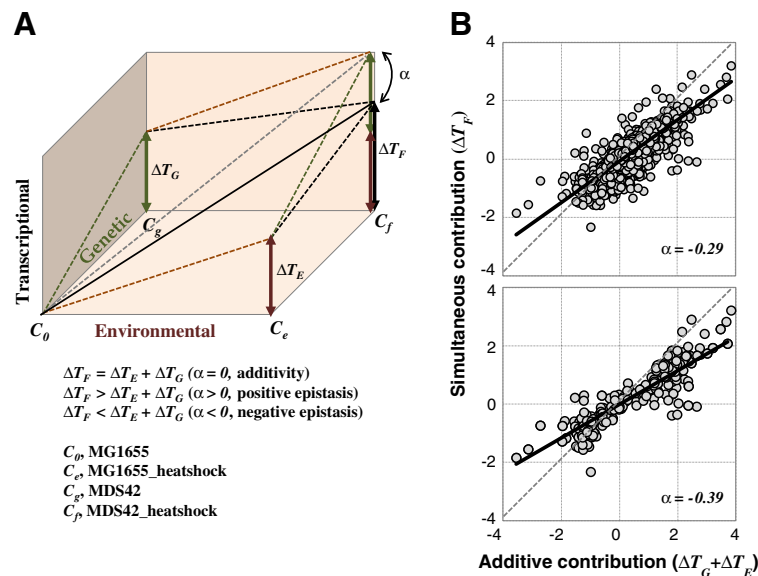


employed, as represented schematically in Figure 6A. These concepts are widely used to determine whether the final fitness or output is simply the sum of the effects of two genes (additivity) or whether one gene partially inhibits (negative epistasis) or enhances (positive epistasis) the effects of the other gene. The extended definition of epistasis could be applied to the interactions among the chemicals and molecules in highly complex biochemical reactions [38]. Here, the two analyzed effects were the transcriptional changes contributed by genome reduction ( $\Delta T_G$ ) and heat shock ( $\Delta T_E$ ) individually, and the final output was defined as the transcriptional changes that were produced when the genomic and heat shock perturbations occurred simultaneously ( $\Delta T_F$ ), *i.e.*, the reduced genome undergoing heat shock (MDS42<sub>heatshock</sub>). The slopes represent the transcriptional changes ( $\Delta T_F$ ) perturbed by the genome and environment in an additive or epistatic manner (Figure 6A, indicated by the solid black line and the broken gray line). The difference in the slopes, defined as  $\alpha$ , represented the relative magnitude of the difference

(repressed or increased) between the additive and simultaneous contributions.

Accordingly, the relationship between the contributions of genome reduction and heat shock to transcriptome was subsequently analyzed in a genomic view. Considering the difference in the number of genes in MG1655 and MDS42, the common genes (3710 genes) shared with both strains were applied in the following analyses. The transcriptional changes due to these two independent contributors (genome and environment, as shown in Figure 5C) were added, and the sum was considered to be the additive contribution ( $\Delta T_G + \Delta T_E$ ). The changes in gene expression between MDS42<sub>heatshock</sub> and MG1655 represented simultaneous genetic and environmental interruptions (simultaneous contribution,  $\Delta T_F$ ). The additive contribution was then plotted against the simultaneous contribution, as shown in Figure 6B. The simultaneous contribution showed a smaller magnitude of transcriptional changes than did the additive contribution ( $\alpha < 0$ ) (Figure 6B, upper panel), approximately 30% suppressive effect ( $\alpha = -0.29$ ,  $p < 0.001$ ). It





**Figure 6 Epistasis between genomic and environmental perturbations.** **A.** Schematic illustration of epistasis and additivity. The environmentally and genetically induced transcriptional changes are defined as  $\Delta T_E$  and  $\Delta T_G$ , respectively. When genetic and environmental perturbations occurred together, the simultaneous change in expression is defined as  $\Delta T_F$ . The relationships among these three parameters were described with the indicated equations, where  $\alpha$  represents the magnitude of the epistatic effect. **B.** Antagonistic epistasis in the transcriptome. The sum of the transcriptional changes caused separately by genetic and environmental perturbations (the additive contribution,  $\Delta T_E + \Delta T_G$ ) is plotted against the transcriptional changes caused simultaneously by both perturbations (the simultaneous contribution,  $\Delta T_F$ ), i.e., the changes in gene expression between MDS42\_heatshock and MG1655. The 3710 common genes and the genes showing significant fluctuations induced by either genetic or environmental perturbation (both DEGs\_gr and DEGs\_hs) are plotted in the upper and bottom panels, respectively. The values of  $\alpha$  were calculated as follows:  $\alpha = 1 - k$ , where  $k$  is the slope of the dot plot fit by linear regression, as indicated by the solid black line.

strongly indicated a negatively compensated relationship between the genomic and environmental contributions to the changes in the transcriptome, as so-called “antagonistic epistasis”. Moreover, the magnitude of the compensatory effect became larger ( $\alpha = -0.39$ ) in the gene groups of the DEGs\_gr and the DEGs\_hs (Figure 6B, bottom panel), suggesting that the negative epistasis became more significant, approximately 40%. Both the genetic and environmental perturbations alone triggered fluctuations in the transcriptome, whereas either perturbation repressed or limited the fluctuation caused by the other when both occurred simultaneously. These results indicated that negative epistasis may be universally involved in biological processes, although the underlying mechanisms and theoretical principles remain to be determined.

## Discussion

This study demonstrated a considerable interaction between the genetic and environmental contributions to the changes in the transcriptome, as evidenced by the identical preferred chromosomal periodicity, the directional similarities in the transcriptional changes, the overlaps between the perturbed regulatory networks, the positive correlation in transcriptional changes induced by the genetic and environmental contributions and the

negative epistasis. Thus, our multilevel comparative analysis provided a global comparison of the effects of genomic and heat shock perturbations to the *E. coli* transcriptome. A more comprehensive evaluation based on the additional experiments under cross multiple conditions and a cross validation using the fruitful reported data sets are essential to figure out the common features of the genomic and environmental contributions to the reorganization in transcriptome.

The transcriptome analysis caught the genome reduction effect. The deleted regions were assumed to have crucial functions, on the basis of our novel observations of the reduced genome, along with the previous reports of its advantages in various applications [17,21,23] but its limited evolvability [20]. As these regions were largely occupied by genes that were derived from the IS/phage, that encoded structural components, or that had unknown or predicted functions, it highlighted the potential importance of these genes. We observed that the deletions altered the transcriptional efficiency, the stability of transcriptional patterns and the specificity of transcriptional changes. The high average levels of gene expression (Figure 1B) and the strong significance of the DEGs\_gr upregulation (Figure 4A) indicated that the elimination of the redundant genome sequence may save materials and energy, leading to increases in the transcriptional

efficiency of the remaining genes. The observation that the fixed chromosomal periodicity (Figure 2C and D) was identical to the number of chromosomal macrodomains [33] indicated that the global transcriptional pattern may be stabilized by the complete removal of the flexible IS/phage regions. The slight but significant increase in the cell growth rate most likely resulted from this steady expression pattern, which potentially accelerated cell division [39]. The significant enrichment of genes of unknown function in the DEGs<sub>gr</sub> and DEGs<sub>hs</sub> (Figure 4C, Additional file 4: Table S3) suggested that these genes are easily disturbed. Additional analyses revealed an apparent bias in the genomic positions of the unknown function genes; these genes were preferentially located in areas near the deleted regions (Additional file 1: Figure S3), which explained the finding of genomic scar dependency in the DEGs<sub>gr</sub> (Figure 4A). The genomic distribution of the functionally undefined genes reflected shared gene organization strategies [40] that were most likely shaped by evolution. Taken together, the data suggest the possibility of a relationship between the identities of the deleted genes (in particular, the genes of unknown function), the chromosomal distribution of the deleted genes, the global expression patterns (periodicity), and the cell growth (fitness).

The genomic and environmental contributions to transcriptome reorganization indicated a negative epistatic relationship (Figure 6B), that is, a negatively compensative relation of the transcriptional changes caused by the two contributors. Such relationship would inhibit the transcriptional fluctuation that was amplified by the positive correlation between the changes in gene expression triggered by the genetic and environmental interruptions (Figure 5C). Epistasis has largely been studied within the framework of genetics, *i.e.*, how a mutation in one gene masks the phenotypic effect of a mutation in another gene [41,42]. The present study was the first to compare the contributions of the genome and the environment to the transcriptome and to include an analysis of the antagonistic (negative) epistasis involved in the reorganization of transcriptome. Because MDS42 was synthetically constructed and is therefore not a product of evolution in the wild nature, the antagonistic epistatic relationship between the genomic and environmental contributions suggested a universal innate principle of cell biology. This principle could be described as the need to restrict the consequences of two positively correlated effects to physiologically tolerable limits. The magnitude of the compensatory effect detected in the DEGs was larger than that in the 3710 common genes (Figure 6B) strongly supported this theory. The adage that the sum is less than its parts (*e.g.*,  $\Delta T_G + \Delta T_E < \Delta T_F$ ) appears to be applicable to the ability of living organisms to survive in severe environments for millions of years

and to an earlier report of the robustness and/or plasticity of living systems [43], although more data are required to support this hypothesis.

## Conclusion

This study investigated the relationship between the genetic and environmental contributions to the bacterial transcriptome. Intriguingly, in addition to similarities and overlaps in periodicity, regulatory networks, DEGs and directionality, both positive correlations and negative epistasis between the genomic and environmental contributions to the global transcriptional activity were observed. These results linked the genome and the environment at the level of the transcriptome and revealed the epistatic nature of the genome-wide transcriptional reorganization that occurs universally in living cells in response to both endogenous and exogenous disturbances.

## Materials and methods

### Strains and culture conditions

The wild-type *E. coli* strain MG1655 was provided by the National BioResource Project, National Institute of Genetics, Shizuoka, Japan. The reduced-genome *E. coli* strain MDS42 was purchased from Scarab Genomics. The genome sequences of MG1655 and MDS42 were retrieved from the information deposited in the GenBank and DDBJ databanks and have accession IDs of U00096 and AP012306, respectively. When comparing the two deposited genome sequences, we found that 696 genes presented in MG1655 were absent in MDS42 (*i.e.*, 696 deleted genes), consistent with the original report [17]. In addition, 22 genes showed mismatched sequences between two genomes (*i.e.*, 22 mutated genes). Thus, a total of 3710 genes (if including *oriC*, then 3711 genes) of perfect matched sequences were determined to be common between the two strains. The bacterial cells were cultured in 5 mL of M63 minimal medium (62 mM  $K_2HPO_4$ , 39 mM  $KH_2PO_4$ , 15 mM  $(NH_4)_2SO_4$ , 2  $\mu$ M  $FeSO_4 \cdot 7H_2O$ , 15  $\mu$ M thiamine hydrochloride, 203  $\mu$ M  $MgSO_4 \cdot 7H_2O$  and 22 mM glucose) at 37°C with shaking at 200 rpm in a BR-21FP air incubator (Taitec).

### Cell culture conditions for the precise measurement of mRNA levels and growth rates

The cell cultures were established from a glycerol stock with two preculture steps, and serial transfers were performed during the exponential phase. The dilution rate needed to keep the final cell concentration at approximately  $1-2 \times 10^8$  cells/mL was calculated according to the growth rate observed during the previous day. The initial cell concentration (*C*<sub>0</sub>) was determined on the basis of the concentration of the preculture used for transfer and the dilution rate. The final cell concentration (*C*<sub>t</sub>) was measured using a flow cytometric analysis (Canto II; Becton,

Dickinson and Company) of the remainder of the cell cultures that had been sampled for microarray analysis. The growth rate was calculated according to the following formula:  $\ln(Ct/CO)/t$ , where  $Ct$ ,  $CO$ , and  $t$  represent the final and initial cell concentrations (cells/mL) and the culture time (h), respectively. This formula has been previously applied in the quantitative evaluation on bacterial growth in exponential phase [28,29,44].

### Heat shock experiment

As previous studies showed that heat shock response in *E. coli* was evidently induced within a few minutes (*i.e.*, 5–10 minutes) after the temperature upshift from 30–37°C to 42–50°C [9,45–47], the condition for heat shock experiment was determined as a 5-min incubation following the temperature upshift from 37°C to 45°C [28]. Exponentially growing cells at approximately  $10^8$  cells/mL were rapidly transferred to an adjacent water bath incubator (Personal-11EX; Taitec) set at 45°C. Following a 5-min incubation at 45°C, the cell culture was immediately poured into a cold phenol-ethanol solution to prepare the samples for microarray analysis. Each heat shock experiment was performed separately to enable the precise control of the timing of the heat shock to ensure that the mRNA levels accurately reflected the stress response.

### Microarrays and data extraction

A high-density DNA microarray covering the entire genome of the *E. coli* W3110 strain was utilized [26,27] as described elsewhere [28,29]. The sample preparation, the microarray analysis with the Affymetrix GeneChip system, and the data extraction based on the finite hybridization (FH) model [26] were performed as previously described [29]. To avoid any significant errors resulting from the diverse signal intensities of the GeneChips (*i.e.*, to minimize experimental errors), only array results with highly similar distributions of probe fluorescence intensities were included in the subsequent expression analysis. The results of 20 arrays for the two strains under the two conditions were used (7 and 3 replicas of exponential growth under 37°C and 45°C heat shock conditions, respectively). The raw data from these 20 microarrays were deposited in the NCBI Gene Expression Omnibus database under the GEO Series accession number GSE33212 (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33212>). The gene names are based on the genome information for W3110 and represent the genes in common between strains W3110 and MG1655 [37]. The entire dataset of gene names and categories is from GenoBase, Japan (<http://ecoli.aist-nara.ac.jp/gb6/Download.html>).

### Computational analyses and graphics

The transcription levels were determined as the log-scale mRNA concentrations (pM). The transcriptional changes (fluctuations) were calculated as the difference between the two transcription levels. Although several methods for the determination of chromosomal periodicity have been reported [4,32], a simple approach was used here. Each expression value that had been determined with the method described in the previous section was projected onto the genome site that corresponded to the gene position, using 100-base bins. Next, the series of expression levels along the genome was smoothed with a moving average of 500 bins (50 kb). The periodicity was calculated using a standard Fourier transform, and the significance of the periodicity was assessed with the chi-squared test. The approximate line of the periodicity was calculated using the highest peak (statistic significance) of the periodogram and was fitted by minimizing the square error between the approximate line and the series of expression values. Gene set enrichment analysis (GSEA) was performed according to the original report [36] using the available online tools (<http://www.broadinstitute.org/gsea/index.jsp>). The TF and sigma factor gene regulation datasets were from RegulonDB. To compare the responses of MG1655 and MDS42, we limited the gene sets to those genes that were included among the 3710 common genes. The pre-ranked gene lists used as the input for GSEA comprised genes filtered by the absolute value of their expression difference between MG1655\_hs and MG1655 (or MDS42 and MG1655). The regulatory networks were visualized using Gephi software (<http://gephi.org>). Subsequently, the Bioconductor software package RankProd [34], which is based on the rank product method, was employed to identify the differential gene expression caused by genome reduction and the heat shock response. The rank product method is a nonparametric statistical method derived from biological reasoning that detects items that are consistently ranked high or low in a number of lists. An advantage of this method is its ability to identify biologically relevant expression changes among a relatively small number of samples [35]. Finally, the statistical analyses, with the exception of GSEA, were performed using R software [48] (<http://www.r-project.org>), and the array plot (heat map) of the gene categories was constructed using Mathematica 8 (Wolfram Research).

### Additional files

**Additional file 1: Figure S1.** Chromosomal locations of the DEGs. **Figure S2.** Statistical analysis. **Figure S3.** Genes of unknown function. **Figure S4.** A close-up view of the transition in major peaks.

**Additional file 2: Table S1.** Detailed results of the gene set enrichment analysis.

**Additional file 3: Table S2.** List of the DEGs.

**Additional file 4: Table S3.** Analysis of the enriched gene categories.

#### Abbreviations

*E. coli*: *Escherichia coli*; TF: Transcription factor; bp: Base pair; kb: Kilobase; gr: Genome reduction; hs: Heat shock; DEGs: Differentially expressed genes; FDR: False discovery rate; GSEA: Gene set enrichment analysis.

#### Competing interests

The authors declare that there is no competing interest.

#### Authors' contributions

BWY and TY conceived the research; BWY designed and performed the experiments; BWY, SS and TY analyzed the data; FK, HM and TY provided the analytic tools; and BWY, SS and TY wrote the manuscript. All the authors read and approved the final manuscript for publication.

#### Acknowledgements

We thank Junko Asada, Natsuko Yamawaki, Shingo Suzuki, Naoaki Ono and Chikara Furusawa for technical assistance and analytical tools. We thank the anonymous reviewers for helping us to improve the clarity of the manuscript. BWY was supported by the project of Novel and Innovative R&D making use of brain structures, from the Ministry of Internal Affairs and Communications, Japan.

#### Author details

<sup>1</sup>Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan. <sup>2</sup>Graduate School of Frontier Biosciences, Osaka University, 1-5 Yamadaoka, Suita, Osaka 565-0871, Japan. <sup>3</sup>Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency (JST), Suita, Osaka 565-0871, Japan.

Received: 15 July 2012 Accepted: 29 December 2012

Published: 16 January 2013

#### References

- Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR: **Genome-wide expression profiling in *Escherichia coli* K-12.** *Nucleic Acids Res* 1999, **27**:3821–3835.
- Tao H, Bausch C, Richmond C, Blattner FR, Conway T: **Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media.** *J Bacteriol* 1999, **181**:6425–6440.
- Lobner-Olesen A, Marinus MG, Hansen FG: **Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis.** *Proc Natl Acad Sci U S A* 2003, **100**:4672–4677.
- Allen TE, Hergard MJ, Liu M, Qiu Y, Glasner JD, Blattner FR, Palsson BO: **Genome-scale analysis of the uses of the *Escherichia coli* genome: model-driven analysis of heterogeneous data sets.** *J Bacteriol* 2003, **185**:6392–6399.
- Peter BJ, Arsuaga J, Breier AM, Khodursky AB, Brown PO, Cozzarelli NR: **Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*.** *Genome Biol* 2004, **5**:R87.
- Jeong KS, Ahn J, Khodursky AB: **Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*.** *Genome Biol* 2004, **5**:R86.
- Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH: **Structure and complexity of a bacterial transcriptome.** *J Bacteriol* 2009, **191**:3203–3211.
- Faith JJ, Driscoll ME, Fusaro VA, Cosgrove EJ, Hayete B, Juhn FS, Schneider SJ, Gardner TS: **Many Microbe Microarrays Database: uniformly normalized Affymetrix compendia with structured experimental metadata.** *Nucleic Acids Res* 2008, **36**:D866–D870.
- Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, Steinhauser D, Selbig J, Willmitzer L: **Metabolomic and transcriptomic stress response of *Escherichia coli*.** *Mol Syst Biol* 2010, **6**:364.
- Nicolas P, Mader U, Dervyn E, Rochat T, Leduc A, Pigeonneau N, Bidnenko E, Marchadier E, Hoebek M, Aymerich S, et al: **Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*.** *Science* 2012, **335**:1103–1106.
- Gunasekera TS, Csonka LN, Paliy O: **Genome-wide transcriptional responses of *Escherichia coli* K-12 to continuous osmotic and heat stresses.** *J Bacteriol* 2008, **190**:3712–3720.
- Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT, Conway T: **The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*.** *Mol Microbiol* 2008, **68**:1128–1148.
- Durfee T, Hansen AM, Zhi H, Blattner FR, Jin DJ: **Transcription profiling of the stringent response in *Escherichia coli*.** *J Bacteriol* 2008, **190**:1084–1096.
- Rutherford BJ, Dahl RH, Price RE, Szmidi HL, Benke PI, Mukhopadhyay A, Keasling JD: **Functional genomic study of exogenous n-butanol stress in *Escherichia coli*.** *Appl Environ Microbiol* 2010, **76**:1935–1945.
- Guell M, Yus E, Lluch-Senar M, Serrano L: **Bacterial transcriptomics: what is beyond the RNA horizo-ome?** *Nat Rev Microbiol* 2011, **9**:658–669.
- Hershberg R, Yeger-Lotem E, Margalit H: **Chromosomal organization is shaped by the transcription regulatory network.** *Trends Genet* 2005, **21**:138–142.
- Posfai G, Plunkett G 3rd, Feher T, Frisch D, Keil GM, Umehoffner K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, et al: **Emergent properties of reduced-genome *Escherichia coli*.** *Science* 2006, **312**:1044–1046.
- Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277**:1453–1462.
- Gerdes SY, Scholle MD, Campbell JW, Balazsi G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, et al: **Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655.** *J Bacteriol* 2003, **185**:5673–5684.
- Umehoffner K, Feher T, Baliko G, Ayaydin F, Posfai J, Blattner FR, Posfai G: **Reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications.** *Microb Cell Fact* 2010, **9**:38.
- Lee JH, Sung BH, Kim MS, Blattner FR, Yoon BH, Kim JH, Kim SC: **Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production.** *Microb Cell Fact* 2009, **8**:2.
- Chakiath CS, Esposito D: **Improved recombinational stability of lentiviral expression vectors using reduced-genome *Escherichia coli*.** *Biotechniques* 2007, **43**:466, 468, 470.
- Sharma SS, Blattner FR, Harcum SW: **Recombinant protein production in an *Escherichia coli* reduced genome strain.** *Metab Eng* 2007, **9**:133–141.
- Sharma SS, Campbell JW, Frisch D, Blattner FR, Harcum SW: **Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome *Escherichia coli* strains.** *Biotechnol Bioeng* 2007, **98**:1056–1070.
- Chae HS, Kim KH, Kim SC, Lee PC: **Strain-dependent carotenoid productions in metabolically engineered *Escherichia coli*.** *Appl Biochem Biotechnol* 2010, **162**:2333–2344.
- Ono N, Suzuki S, Furusawa C, Agata T, Kashiwagi A, Shimizu H, Yomo T: **An improved physico-chemical model of hybridization on high-density oligonucleotide microarrays.** *Bioinformatics* 2008, **24**:1278–1285.
- Suzuki S, Ono N, Furusawa C, Kashiwagi A, Yomo T: **Experimental optimization of probe length to increase the sequence specificity of high-density oligonucleotide microarrays.** *BMC Genomics* 2007, **8**:373.
- Kishimoto T, Iijima L, Tatsumi M, Ono N, Oyake A, Hashimoto T, Matsuo M, Okubo M, Suzuki S, Mori K, et al: **Transition from positive to neutral in mutation fixation with continuing rising fitness in thermal adaptive evolution.** *PLoS Genet* 2010, **6**:e1001164.
- Tsuru S, Yasuda N, Murakami Y, Ushioda J, Kashiwagi A, Suzuki S, Mori K, Ying BW, Yomo T: **Adaptation by stochastic switching of a monostable genetic circuit in *Escherichia coli*.** *Mol Syst Biol* 2011, **7**:493.
- Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639–1645.
- Wright MA, Kharchenko P, Church GM, Segre D: **Chromosomal periodicity of evolutionarily conserved gene pairs.** *Proc Natl Acad Sci U S A* 2007, **104**:10559–10564.
- Mathelier A, Carbone A: **Chromosomal periodicity and positional networks of genes in *Escherichia coli*.** *Mol Syst Biol* 2010, **6**:366.
- Valens M, Penaud S, Rognon M, Cornet F, Boccard F: **Macrodomain organization of the *Escherichia coli* chromosome.** *EMBO J* 2004, **23**:4330–4341.
- Hong J, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J: **RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis.** *Bioinformatics* 2006, **22**:2825–2827.
- Breitling R, Armengaud P, Amtmann A, Herzyk P: **Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments.** *FEBS Lett* 2004, **573**:83–92.



36. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: **Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.** *Proc Natl Acad Sci U S A* 2005, **102**:15545–15550.
37. Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, et al: **Escherichia coli K-12: a cooperatively developed annotation snapshot–2005.** *Nucleic Acids Res* 2006, **34**:1–9.
38. Matsuura T, Kazuta Y, Aita T, Adachi J, Yomo T: **Quantifying epistatic interactions among the components constituting the protein translation system.** *Mol Syst Biol* 2009, **5**:e297.
39. Viollier PH, Shapiro L: **Spatial complexity of mechanisms controlling a bacterial cell cycle.** *Curr Opin Microbiol* 2004, **7**:572–578.
40. Lawrence JG: **Shared strategies in gene organization among prokaryotes and eukaryotes.** *Cell* 2002, **110**:407–413.
41. Van Driessche N, Demsar J, Booth EO, Hill P, Juvan P, Zupan B, Kuspa A, Shaulsky G: **Epistasis analysis with global transcriptional phenotypes.** *Nat Genet* 2005, **37**:471–477.
42. Sanjuan R, Elena SF: **Epistasis correlates to genomic complexity.** *Proc Natl Acad Sci U S A* 2006, **103**:14402–14405.
43. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, et al: **Multiple high-throughput analyses monitor the response of E. coli to perturbations.** *Science* 2007, **316**:593–597.
44. Shimizu Y, Tsuru S, Ito Y, Ying BW, Yomo T: **Stochastic switching induced adaptation in a starved Escherichia coli population.** *PLoS One* 2011, **6**:e23953.
45. Tilly K, Erickson J, Sharma S, Georgopoulos C: **Heat shock regulatory gene rpoH mRNA level increases after heat shock in Escherichia coli.** *J Bacteriol* 1986, **168**:1155–1158.
46. Nagai H, Yuzawa H, Kanemori M, Yura T: **A distinct segment of the sigma 32 polypeptide is involved in DnaK-mediated negative control of the heat shock response in Escherichia coli.** *Proc Natl Acad Sci U S A* 1994, **91**:10280–10284.
47. Yura T, Nakahigashi K: **Regulation of the heat-shock response.** *Curr Opin Microbiol* 1999, **2**:153–158.
48. Ihaka R, Gentleman R: **R: A language for data analysis and graphics.** *J Comput Graph Stat* 1996, **5**:299–314.

doi:10.1186/1471-2164-14-25

**Cite this article as:** Ying et al.: Multilevel comparative analysis of the contributions of genome reduction and heat shock to the *Escherichia coli* transcriptome. *BMC Genomics* 2013 **14**:25.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

