Diversity for commonality in the evolutionary changes of the reduced genome to regain the growth fitness

Kenya Hitomi[1], Yoichiro Ishii[1], Bei-Wen Ying[1,*]

[1]School of Life and Environmental Sciences, University of Tsukuba, Tennodai 1-1-1, Tsukuba, Ibaraki 305-8572, Japan

[*]Correspondence: ying.beiwen.gf@u.tsukuba.ac.jp

Abstract

As the genome encodes the information crucial for cell growth, a sizeable genomic deficiency often causes a significant decrease in growth fitness. Whether and how the decreased growth fitness caused by genome reduction could be compensated by evolution was investigated here. Experimental evolution with an *Escherichia coli* strain carrying a reduced genome was conducted in multiple lineages for approximately 1,000 generations. The growth rate, which was largely declined due to genome reduction, was considerably recovered, associated with the improved carrying capacity. Genome mutations accumulated during evolution were significantly varied across the evolutionary lineages and were randomly localized on the reduced genome. Transcriptome reorganization showed a common evolutionary direction and conserved the chromosomal periodicity, regardless of highly diversified gene categories, regulons, and pathways enriched in the differentially expressed genes. Genome mutations and transcriptome reorganization caused by evolution, which were found to be dissimilar to those caused by genome reduction, must have followed divergent mechanisms in individual evolutionary lineages. Gene network reconstruction successfully identified three gene modules functionally differentiated, which were responsible for the evolutionary changes of the reduced genome in growth fitness, genome mutation, and gene expression, respectively. The diversity in evolutionary approaches improved the growth fitness associated with the homeostatic transcriptome architecture as if the evolutionary compensation for genome reduction was like all roads leading to Rome.

Keywords: genome reduction, experimental evolution, growth fitness, transcriptome, genome mutation, chromosomal periodicity, gene expression, network analysis

Introduction

The genome encodes the information for cell growth, and its size is likely the evolutionary consequence [1, 2, 3]. To determine the essential genomic sequence of modern cells, removing redundant DNA sequences from the wild-type genome in bacteria, so-called genome reduction, has been challenged to a large extent [4, 5, 6, 7]. These efforts have been made to discover the minimal genetic requirement for a free-living organism growing under the defined conditions [8, 9, 10]. It resulted in the finding of the coordination of genome with cell growth, i.e., genome reduction significantly decreased the growth rate of *Escherichia coli (E. coli)* cells independent of culture media or growth forms [11, 12, 13]. Slow growth or fitness decline was commonly observed in the genetically reduced [13, 14] and chemically synthesized genomes [10, 15].

The growth decrease could be recovered by experimental evolution. Growth fitness generally represents the adaptiveness of the living organism to the defined environment [16]. Although the reduced genomes somehow showed differentiated evolvability compared to the wild-type genomes [17, 18, 19], their evolutionary adaptation to the environmental changes has been successfully achieved [20, 21, 22]. Experimental evolution under the defined culture condition successfully increased the decreased growth rates of the reduced genomes [17, 20, 21] and fastened the slow-growing synthetic genome [23]. The evolutionary rescue of the growth rate must be associated with the changes in genomic sequence and gene expression benefited for the growth fitness, as what happened in nature adaptive evolution [24, 25, 26]. Our previous study showed that the decreased growth rate caused by the absence of a sizeable genomic sequence could be complemented by introducing the mutations elsewhere in the genome [20, 21]. Additionally, the changes in growth rate caused by either genome reduction or experimental evolution were dependent on the genome size but not the specific gene function [20].

A genome-wide understanding of the evolutionary compensated fitness increase of the reduced genome is required. The experimental evolution compensated for the genome reduction-mediated growth changes were considered stringently related to transcriptome reorganization. Our previous studies found the conserved features in transcriptome reorganization, although the gene expression patterns were significantly disturbed by either genome reduction or experimental evolution [27, 28, 29]. It indicated that

experimental evolution rescued the growth rate disturbed by genome reduction *via* the underlying mechanisms to maintain homeostasis as a growing cell. How the growth recovery of reduced genomes was achieved and whether it was in general are unclear. To address the questions, we conducted the experimental evolution with a reduced genome in multiple lineages and analyzed the evolutionary changes of the genome and transcriptome in the present study.

Results

Fitness recovery of the reduced genome by experimental evolution

Experimental evolution of the reduced genome was conducted to regain the growth fitness, which was decreased due to genome reduction. Serial transfer was performed with multiple dilution rates to keep the bacterial growth within the exponential phase (Fig. S1), as described [17, 20]. Nine evolutionary lineages were conducted independently (Fig. S2, Table S1). A gradual increase in growth rate was observed along with the generation passage, which was combined with a rapid increase in the early evolutionary phase and a slow increase in the later phase (Fig. 1A). Most evolved populations (Evos) showed improved growth fitness. Nevertheless, the final growth rates were considerably varied (Fig. 1B, upper), and the evolutionary dynamics of the nine lineages were somehow divergent (Fig. S2). It indicated the diversity in the evolutionary approaches for improved fitness. Eight of nine Evos achieved faster growth than the genome-reduced ancestor (Anc), whereas all Evos decreased in their saturated population densities (Fig. 1B, bottom). In comparison to the wild-type strain carrying the full-length genome (WT), the primarily decreased growth rate caused by genome reduction was significantly improved by experimental evolution (Fig. 1C, upper), associated with the considerable reduction in saturated density (Fig. 1C, bottom). It demonstrated that the experimental evolution could compensate for the genome reduction with a trade-off in population size (carrying capacity), consistent with the previous findings [20, 21].

Intriguingly, a positive correlation was observed between the growth fitness and the carrying capacity of the Evos (Fig. 1D). It seemed that the present experimental evolution did not obey the *r/K* selection theory[30, 31], which was known as the trade-off

relationship (negative correlation) between the growth rate and the carrying capacity [32, 33]. Taking account of our previous finding that the colony growth rates of genome-reduced strains were proportional to the colony sizes [11], the collapse of the trade-off law likely resulted from genome reduction. As the trade-off between growth fitness and carrying capacity was proposed to balance the cellular metabolism that resulted from the cost of enzymes involved [33], the deleted genomic sequences might play a role in maintaining the metabolism balance.

Significant variation and random localization of genome mutations

Genome resequencing (Table S2) identified a total of 65 mutations fixed in the nine Evos (Table 1). The number of mutations largely varied among the nine Evos, from two to 13, and no common mutation was detected (Table S3). 51 out of 65 mutations occurred in the genes, and 45 out of 51 were SNPs. As 36 out of 45 SNPs were nonsynonymous, the mutated genes might benefit the fitness increase. In addition, the abundance of mutations was unlikely to be related to the magnitude of fitness increase. For instance, A2 accumulating only two mutations presented a highly increased growth rate compared to F2 of eight mutations (Table 1), which poorly improved the growth (Fig. 1B). B2, D2, and E2 all succeeded in fitness increase to an equivalent degree (Fig. 1B), whereas they fixed 13, 7, and 3 mutations, respectively (Table 1). The mutated genes were varied in 14 gene categories, somehow more frequently in the gene categories of Transporter, Enzyme, and Unknown function (Fig. S3). They seemed highly related to essentiality, as 11 out of 49 mutated genes were essential (Table S3). As the essential genes were known to be more conserved than nonessential ones [34, 35], the high frequency of the mutations fixed in the essential genes suggested the mutation in essentiality for fitness increase was the evolutionary strategy for reduced genome. The large variety in genome mutations and no correlation of mutation abundance to fitness improvement strongly suggested that no mutations were specifically responsible or crucially essential for recovering the growth rate of the reduced genome.

Additionally, there were no overlapped genomic positions for the 65 mutations in the nine Evos (Fig. 2A). Random simulation was performed to verify whether there was any bias or hotspot in the genomic location for mutation accumulation (Fig. 2B). The

simulation of 65 mutations randomly occurred on the reduced genome was performed 1,000 times. The distance of the mutation (mutated genomic location) to the nearest genomic scar caused by genome reduction was calculated. Welch's t-test was performed to evaluate the significance of the locational bias between the random mutations and the mutations accumulated in the Evos (Fig. 2C). As the mean of *p* values from simulations was insignificant ($\mu_p > 0.05$), there was no locational bias for mutation accumulation caused by genome reduction.

Common evolutionary direction and homeostasis in transcriptome reorganization

Since no specificity was detected in the genome mutations, whether these mutations disturbed the genome-wide gene expression pattern was investigated. Hierarchical clustering and principal component analysis (PCA) showed that the evolved transcriptomes were directed to similar patterns but divergent to the WT transcriptome (Fig. S4). The evolutionary direction of the transcriptomes of the reduced genome was not approaching the wild-type transcriptome. As a global feature of gene expression, the chromosomal periodicity of the transcriptome was analyzed using the Fourier transform as previously described [28, 36]. As a result, the transcriptomes of all Evos presented a common six-period with statistical significance, equivalent to those of the wild-type and ancestral reduced genomes (Fig. 3A, Table S4). It demonstrated that the chromosomal architecture of gene expression patterns remained highly conserved, regardless of the considerably varied mutations. The homeostatic periodicity was consistent with our previous findings that the chromosomal periodicity of the transcriptome was independent of genomic or environmental variation [27, 28].

In addition, the genomic locations of the mutations seemed irrelevant to chromosomal periodicity (Fig. 3A, red lines). No mutagenesis hotspot was observed even if these mutations were accumulated on a single genome (Fig. 2A). Alternatively, the expression levels of the mutated genes were somehow higher than those of the remaining genes (Fig. 3B), which sounded reasonable. The ratio of the essential genes in the mutated genes was ~ 22% (Table S3), much higher than the ratio (~9%) of essential to all genes (302 out of 3,290) in the reduced genome. As the essential genes showed higher expression levels than the nonessential ones [37], the high essentiality of

the mutated genes might result in a higher mean expression level. On the other hand, the high frequency of the mutations fixed in the essential genes was unexpected because the essential genes were generally more conserved than nonessential ones [34]. It strongly suggested the essentiality mutation for homeostatic transcriptome architecture happened in the reduced genome.

Diversified functions and pathways of the differentially expressed genes

As the evolved transcriptomes were differentiated from those of the WT and reduced genomes (Fig. S4), the differentially expressed genes (DEGs) were further identified to discover the gene functions or biological processes contributing to the fitness changes. The abundance of DEGs among the Evos varied from 333 to 1,130 genes and of few overlaps (Fig. 4A). Most DEGs were unique to each evolutionary lineage, and the common DEGs across all Evos were only 108 genes (Table S5). The number of DEGs partially overlapped among the Evos declined significantly along with the increased lineages of Evos (Fig. 4B). Enrichment analysis showed that only the histidine-related pathways were significantly enriched in the common DEGs (Fig. 4C). Functional enrichment of the DEGs in individual Evos showed that the amino acid metabolism considerably participated and that the enriched pathways were poorly overlapped (Fig. S5). These findings strongly suggested no universal rule for evolutionary changes of the reduced genome.

In comparison, 1,226 DEGs were induced by genome reduction. The common DEGs of genome reduction and evolution varied from 168 to 540, fewer than half of the DEGs responsible for genome reduction in all Evos (Fig. 5A). The conclusion remained consistent even if the DEGs were determined with an alternative method, RankProd (Fig. S6). Functional enrichment of the DEGs observed the specific transcriptional regulation and metabolic pathways participating in the transcriptome reorganization in response to genome reduction and evolution. Only σ38 was enriched in the genome reduction-mediated DEGs, which was the only regulon that partially overlapped with the Evos (Fig. 5B). No regulon in common was enriched, besides a few partially overlapped regulons, i.e., GadW, GadX, and RcsB (Fig. 5B). In addition, both the number of enriched pathways and their overlaps were significantly differentiated among

the nine Evos or between Evos and reduced genome (Fig. 5C). No common pathways were commonly enriched between the genome reduction-mediated DEGs and Evos, no matter annotated the metabolic pathways with GO or KEGG (Fig. S7). The flagellar assembly, the only enriched pathway in genome reduction-mediated DEGs, was absent in all Evos (Fig. 5D). Alternatively, the amino acids-related metabolisms were frequently detected in the Evos, e.g., histidine metabolism, biosynthesis of amino acids, etc. (Fig. 5D, Fig. S7). The variable pathways in the Evos indicated that evolution compensated for the genome reduction in various ways, which differed from how the genome reduction was caused.

Gene modules responsible for the evolutionary changes of the reduced genome

Genome mutation analysis and transcriptome analysis failed to identify the common gene categories or pathways that correlated to the evolution of the reduced genome; thus, the gene modules correlated to evolution were newly evaluated. The weighted gene co-expression network analysis (WGCNA) [38] was performed toward the evolved transcriptomes as tested previously [27]. Reconstruction of the 3,290 genes in the reduced genome led to 21 gene modules, comprising 8 to 320 genes per module (Fig. S8). Hierarchical clustering of these modules showed that roughly three major classes could be primarily divided (Fig. 6A). Functional correlation analysis showed that three of 21 modules (M2, M10, and M16) were highly significantly correlated to the growth fitness, the number of DEGs, and mutation frequency, respectively (Fig. 6A). It indicated that the three modules were highly essential and functionally differentiated for growth control, transcriptional change, and mutagenesis.

Enrichment analyses further identified the gene categories, regulons, and metabolisms that significantly appeared in the three modules (Fig. 6B). Two gene categories of nonessential function were enriched in M10, and the module correlated to the number of DEGs; in contrast, no gene category was enriched in M2 and M16 (Fig. 6B, left). All three modules successfully enriched the regulons without overlaps (Fig. 6B, middle). It indicates that the main regulatory mechanisms participating in the three modules for growth control, transcriptional change, and mutagenesis were divergent. GO enrichment resulted in various biological processes in the three modules, roughly

related to transport, transposition, and translation in M2, M10, and M16, respectively (Fig. 6B, right). Compared to the enriched functions of the genes that disappeared due to genome reduction (Fig. 6B, bottom), the gene categories of phage and unknown function and the biological processes related to DNA transposition and integration were commonly identified in M10. It strongly indicated that M10 was responsible for genome reduction. The newly constructed gene networks successfully identified three modules correlated to mutation, DEGs, and growth, revealing the functional differentiation responsible for evolution to maintain the homeostatic transcriptome architecture for a growing cell.

Discussion

The evolutionary compensation for genome reduction was directed toward an identical goal of increased fitness but differentiated in the manner of genomic changes. Firstly, various genetic functions seemed to trigger the increased growth rates of the Evos. A few mutations could compensate for the sizeable genomic deficiency. The considerable variation in the fixed mutations without overlaps among the nine Evos (Table 1) implied no common mutagenetic strategy for the evolutionary improvement of growth fitness. It was supported by the fact that no genomic locational bias for mutations fixed in the evolution (Fig. 2). Secondly, the transcriptomes presented conserved chromosomal architectures (Fig. 3) and universal directional changes (Fig. S4), regardless of the significant and differentiated changes in gene expression in response to evolution (Fig. 4, Fig. S5). Although the periodicity of chromosomal architecture was well known [39, 40] and coordinated to the bacterial growth rate[28, 41], employing its homeostasis as the evolutionary consequence provided a conceptual and unique understanding of growing cells.

It's unclear whether the differentiation in evolutionary paths was particularly significant for the reduced genome used in the present study. Evolution studies often focus on finding the common mutations accumulated in multiple evolutionary lineages to obtain the reasonable mechanism responsible for the adaptation to the defined condition (Fig. 7, i). Common mutations[22, 42] or identical genetic functions[43] were reported in the experimental evolution with different reduced genomes. Nevertheless,

divergent evolutionary mechanisms were proposed as not all mutations contribute to the evolved fitness [22, 43]. The present study accentuated the variety of mutations fixed during evolution. Considering the high essentiality of the mutated genes (Table S3), most or all mutations were assumed to benefit the fitness increase, partially demonstrated previously [20]. The differentiated mutations and DEGs guided to a homeostatic (Fig. 7, iii) rather than a variable consequence (Fig. 7, ii). Multiple evolutionary paths for the reduced genome to improve growth fitness were likely all roads leading to Rome.

In addition, the transcriptome reorganization for fitness increase triggered by evolution differed from that for fitness decrease caused by genome reduction. General analyses failed to detect the regulatory network or genetic function mediated by genome reduction and evolution in common. Instead, the newly constructed gene modules successfully enriched the gene categories of mobile elements and unknown functions (Fig. 6B, left) as the evolutionary compensation for genome reduction. The represented mobile elements, flagella, were known to be responsive to environmental stresses such as hypoosmotic pressure or pH [44, 45]. Genome reduction and evolution seemed equivalent to the stress response in *E. coli*. These findings were reasonable as enterobactin protected *E. coli* from oxidative stress [46, 47], and enterobactin biosynthesis was upregulated for biofilm formation in genome-reduced *E. coli* [48]. The compensation of evolution to genome reduction not only verified the known function and mechanism from a global regulatory viewpoint but also revealed a novel understanding of the molecular mechanisms and gene functions.

The discriminated functions of gene modules might play a crucial role in response to genomic and evolutionary changes. WGCNA was conducted to discover the potential correlation of gene expression to growth fitness. It succeeded in finding the genes participating in the evolutionary changes of the reduced genome to regain growth fitness. Three enriched gene modules were assumed separately responsible for replication, transcription, and population dynamics (Fig. 6B). The growth-correlated gene module significantly enriched the iron-related biological functions (M2). Although the translation was commonly reported to be correlated to the growth rate [49, 50], it was enriched in the gene module coordinated to transcriptional changes (M10). Such

functional differentiation of the gene modules might connect with the differentiated medium components responsible for varied bacterial growth phases, which was observed using the high-throughput growth assay in hundreds of medium combinations combined with machine learning [51, 52]. We assumed that the various chemicals disturbed different metabolic fluxes in which different gene modules might have participated. The biological meaningfulness of the gene modules suggested an alternative genetic classification besides the commonly used clustering criteria, such as Gene Orthology [53] and Regulon [54]. In summary, the present study provided a representative example showing multiple evolutionary paths (i.e., gene mutation and expression) directed the reduced genome to the improved fitness with the homeostatic transcriptome.

Materials and Methods

E. coli strains and culture medium

The *E. coli* K-12 W3110 wild-type and its genome-reduced strains were initially distributed by the National BioResource Project of the National Institute of Genetics. The reduced genome was approximately 21% smaller than the wild-type genome. The minimal medium M63 was used as described in detail elsewhere [13, 55].

Experimental evolution

The genome-reduced *E. coli* strain was evolved in 2 mL of the M63 medium by serial transfer, as previously described [17, 20]. Nine evolutionary lineages were all initiated from the identical culture stock prepared in advance. The 24-well microplates specific for microbe culture (IWAKI) were used, and every four wells of four tenfold serial dilutions, e.g., 10^3~10^6, were used for each lineage. The microplates were incubated in a microplate bioshaker (Deep Well Maximizer, Taitec) at 37°C, with rotation at 500 rpm. The serial transfer was performed at ~24-h intervals. Only one of the four wells (dilutions) showing growth in the early exponential phase (OD_{600} = 0.01-0.1) was selected and diluted into four wells of a new microplate using four dilution ratios. Serial transfer was repeated until all evolutionary lineages reached approximately 1,000 generations, which required approximately two months per lineage. A total of nine lineages were conducted, and the daily records of the nine lineages were summarized in

Table S1. The evolutionary generations (G) and the growth rates (μ) were calculated according to the following equations (Eq. 1 and Eq. 2).

$$G = \log_2\left(\frac{N_t}{N_0}\right) \qquad \text{(Eq. 1)}$$

$$\mu = \frac{\ln\left(\frac{N_t}{N_0}\right)}{\Delta t} \qquad \text{(Eq. 2)}$$

Fitness assay

Both the ancestor and the evolved *E. coli* populations of the reduced genome were subjected to the fitness assay, as previously described [13, 55]. In brief, every 200 µl of culture was dispensed to each well in the 96-well microplate (Coaster). The microplate was incubated at 37°C in a plate reader (EPOCH2, BioTek), shaking at 567 cpm (cycles per minute) for 48 h. The temporal changes in OD_{600} were measured at 30-min intervals. The growth fitness (r) was calculated using the following equation between any two consecutive points (Eq. 3).

$$r_i = \frac{ln\left(\frac{C_{i+1}}{C_i}\right)}{t_{i+1} - t_i} \qquad \text{(Eq. 3)}$$

Here, t_i and t_{i+1} are the culture times at the two consecutive measurement points, and C_i and C_{i+1} are the OD_{600} at time points t_i and t_{i+1}. The growth rate was determined as the average of three consecutive ri, showing the largest mean and minor variance. The mean of the biological triplicates was defined as the growth fitness and used in the analyses.

Genome resequencing and mutation analysis

The *E. coli* cells were collected at the stationary growth phase (i.e., $OD_{600} > 1.0$) and subjected to genome resequencing, as described previously [20]. In brief, the bacterial culture was stopped by adding rifampicin at 300 µg/mL. The cell pellet was collected for genomic DNA purification using a Wizard Genomic DNA Purification Kit (Promega) under the manufacturer's instructions. The sequencing library was prepared using the Nextera XT DNA Sample Prep Kit (Illumina), and the paired-end sequencing (300 bp × 2) was performed using the Illumina HiSeq platform. The raw datasets were deposited in the DDBJ Sequence Read Archive under the accession number DRA013662. The sequencing reads were aligned to the reference genome *E. coli*

W3110 (AP009048.1, GenBank), and the mutation analysis was performed with the Breseq pipeline [56, 57]. The DNA sequencing and mapping statistics are summarized in Table S2. The mutations fixed in the coding regions due to evolution are shown in Table S3.

Calculation and simulation of the genomic positions of mutations

The distances from the genomic positions of the genome mutations fixed in the nine Evos to the nearest genomic scars caused by the genome reduction were calculated as described previously [13, 36]. Random mutations that occurred on the reduced genome were simulated 1,000 times, and the distances of these mutations to the nearest genomic scars were calculated. Note that the number of mutations in the simulation remained equivalent to that detected in the Evos. Welch t-tests were performed to evaluate the statistical significance (*p* values) of the bias in the distances observed in the Evos and the simulations. The distribution and the mean of 1,000 *p* values acquired from the simulation were calculated to evaluate the locational bias of the mutations fixed in Evos.

RNA sequencing

The *E. coli* cells were collected at the exponential growth phase (i.e., $5\times10^7 \sim 2\times10^8$ cells/mL) and subjected to RNA sequencing, as described previously [27, 41]. In brief, the bacterial growth was stopped by mixing with the iced 10% phenol ethanol solution. The cell pellet was collected to purify the total RNAs using the RNeasy Mini Kit (QIAGEN) and RNase-Free DNase Set (QIAGEN) according to product instructions. The paired-end sequencing (150 bp × 2) was performed using the Novaseq6000 next-generation sequencer (Illumina). The rRNAs were removed from the total RNAs using the Ribo-Zero Plus rRNA Depletion Kit (Illumina), and the mRNA libraries were prepared using the Ultra Directional RNA Library Prep Kit for Illumina (NEBNext). Biological replicates were performed for all conditions (N=2~4). The raw datasets were deposited in the DDBJ Sequence Read Archive under the accession number DRA013662.

Data processing and normalization

The FASTQ files were mapped to the reference genome W3110 (accession number AP009048.1, GenBank) using the mapping software Bowtie2 [58], as described previously [27, 41]. The obtained read counts were converted to FPKM values according to the gene length and total read count values. Global normalization of the FPKM values was performed to reach an identical mean value in the logarithmic scale in all datasets. The gene expression level was determined as the logarithmic value of FPKM, and the mean values of biological replicates were used for the following analyses (Table S6).

Computational analysis

The normalized datasets were subjected to the computational analyses performed with the R statistical analysis software. A total of 3290 genes were used for hierarchical clustering and principal component analysis (PCA), which were performed with the R functions of "hclust" and "prcomp", respectively, as described previously[27]. The corresponding parameters of "method" and "scale" were set as "average" and "F", respectively. The R package of DESeq2[59] was used to determine the differentially expressed genes (DEGs), based on the false discovery rate (FDR < 0.05) [60]. The read counts were used as the input data for DESeq2, in which the data normalization was performed at each run for pair comparison.

Enrichment analysis

Functional enrichments were performed according to the features of gene category[61], transcriptional regulation[54], gene ontology[53], and metabolic pathways[62, 63]. Twenty-one gene categories and 46 regulons, which comprised more than 15 genes and 15 regulatees, were subjected to the enrichment, respectively. The statistical significance was evaluated by the binomial test with Bonferroni correction. The enrichment analysis of gene ontology (GO terms) [53, 64] and metabolic pathway (Kyoto encyclopedia of genes and genomes, KEGG) [62, 63] was performed using DAVID, a web-based tool for visualizing the characteristics of gene clusters with expression variation [65 66]. The statistical significance was according to FDR.

Chromosomal periodicity analysis

Fourier transform was used to evaluate the chromosomal periodicity of the transcriptome, as previously described [28, 41]. The genome was divided into compartments of 1 kb each, and the mean expression level of the genes within the corresponding sections was calculated. Gene expression levels were smoothed with a moving average of 100 kb and subjected to the periodicity analysis using the function "periodogram" in R. The max peak (periodic wavelength) of the periodogram was fitted to the gene expression data using the function "nls" in R by the least squares method according to the following equation (Eq. 4).

$$\mathrm{Exp}\ (x) = a * \sin\left(x + \frac{b}{T}\right) + c \quad \text{(Eq. 4)}$$

Here, a, b, T, and c represent the periodic amplitude, the periodic phase, the periodic wavelength indicated by the max peak, and the mean expression level of the whole transcriptome as a constant, respectively. The statistical significance of the periodicity was assessed with Fisher's g test (Table S4), as described previously [28, 41]. The genomic position of *ori* was according to the previous reports [41, 67]. In addition, the function "abline" in R was used to point out the genomic positions of the mutations.

Gene network analysis

The weighted gene co-expression network analysis (WGCNA) of the nine Evos was performed with the R package of WGCNA [38], as described previously [27]. A Step-by-Step method was used to determine the parameters for constructing the gene networks. The soft threshold was set at 12, where the R^2 of Scale Free Topology Model Fit was approximately 0.9 recommended by the developer's instruction. The resultant gene networks were clustered with the "hclust" function (method=average) and reconstructed by merging similar modules using the "mergeCloseModules" function with a height cut of 0.25 in the "dynamic tree cut" method. Finally, a total of 21 modules were determined for 3,290 genes. The correlation coefficients and p-values between the expression of the gene modules and the other global features (growth rates, DEGs, and mutations) were evaluated using the functions "cor" and "corPvalueStudent" in R, respectively. FDR correction was applied to the p-values to account for multiplicity. Functional enrichment of the gene modules was performed, and the

statistical significance was evaluated by the binomial test with Bonferroni correction as described above.

Acknowledgments

We thank NBRP for providing the *E. coli* strains carrying the wild-type and reduced genomes (W3110 and KHK collection). This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) (grant number 19H03215) and partially by Grant-in-Aid for Challenging Exploratory Research (grant number 21K19815).

Competing interests

The authors declare that there are no competing interests.

References

1. Koonin EV. Evolution of genome architecture. *Int J Biochem Cell Biol* **41**, 298-306 (2009).
2. Lynch M. Streamlining and simplification of microbial genome architecture. *Annu Rev Microbiol* **60**, 327-349 (2006).
3. Lynch M, Conery JS. The origins of genome complexity. *Science* **302**, 1401-1404 (2003).
4. Kotaka Y, Hashimoto M, Lee KI, Kato JI. Mutations identified in engineered Escherichia coli with a reduced genome. *Front Microbiol* **14**, 1189877 (2023).
5. Kato J, Hashimoto M. Construction of consecutive deletions of the Escherichia coli chromosome. *Mol Syst Biol* **3**, 132 (2007).
6. Pósfai G, *et al.* Emergent properties of reduced-genome Escherichia coli. *Science* **312**, 1044-1046 (2006).
7. Hashimoto M, *et al.* Cell size and nucleoid organization of engineered Escherichia coli cells with a reduced genome. *Mol Microbiol* **55**, 137-149 (2005).
8. Aida H, Ying B-W. Efforts to Minimise the Bacterial Genome as a Free-Living Growing System. *Biology* **12**, 1170 (2023).
9. Breuer M, *et al.* Essential metabolism for a minimal cell. *Elife* **8**, (2019).
10. Hutchison CA, 3rd, *et al.* Design and synthesis of a minimal bacterial genome. *Science* **351**, aad6253 (2016).

11. Hitomi K, Weng J, Ying BW. Contribution of the genomic and nutritional differentiation to the spatial distribution of bacterial colonies. *Front Microbiol* **13**, 948657 (2022).
12. Xue H, Kurokawa M, Ying BW. Correlation between the spatial distribution and colony size was common for monogenetic bacteria in laboratory conditions. *BMC Microbiol* **21**, 114 (2021).
13. Kurokawa M, Seno S, Matsuda H, Ying BW. Correlation between genome reduction and bacterial growth. *DNA Res* **23**, 517-525 (2016).
14. Karcagi I, *et al.* Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. *Mol Biol Evol* **33**, 1257-1269 (2016).
15. Gibson DG, *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* **329**, 52-56 (2010).
16. Orr HA. Fitness and its role in evolutionary genetics. *Nat Rev Genet* **10**, 531-539 (2009).
17. Nishimura I, Kurokawa M, Liu L, Ying BW. Coordinated Changes in Mutation and Growth Rates Induced by Genome Reduction. *mBio* **8**, (2017).
18. Csorgo B, Feher T, Timar E, Blattner FR, Posfai G. Low-mutation-rate, reduced-genome Escherichia coli: an improved host for faithful maintenance of engineered genetic constructs. *Microbial cell factories* **11**, 11 (2012).
19. Umenhoffer K, *et al.* Reduced evolvability of Escherichia coli MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microbial cell factories* **9**, 38 (2010).
20. Kurokawa M, Nishimura I, Ying BW. Experimental Evolution Expands the Breadth of Adaptation to an Environmental Gradient Correlated With Genome Reduction. *Front Microbiol* **13**, 826894 (2022).
21. Choe D, *et al.* Adaptive laboratory evolution of a genome-reduced Escherichia coli. *Nat Commun* **10**, 935 (2019).
22. Suzuki S, Horinouchi T, Furusawa C. Prediction of antibiotic resistance by gene expression profiles. *Nat Commun* **5**, 5792 (2014).
23. Moger-Reischer RZ, *et al.* Evolution of a minimal cell. *Nature* **620**, 122-127 (2023).
24. Ishikawa A, *et al.* Different contributions of local- and distant-regulatory changes to transcriptome divergence between stickleback ecotypes. *Evolution* **71**, 565-581 (2017).
25. Rozen DE, de Visser JA, Gerrish PJ. Fitness effects of fixed beneficial mutations in microbial populations. *Curr Biol* **12**, 1040-1045 (2002).

26. Lynch M. Evolution of the mutation rate. *Trends Genet* **26**, 345-352 (2010).
27. Matsui Y, Nagai M, Ying BW. Growth rate-associated transcriptome reorganization in response to genomic, environmental, and evolutionary interruptions. *Front Microbiol* **14**, 1145673 (2023).
28. Nagai M, Kurokawa M, Ying BW. The highly conserved chromosomal periodicity of transcriptomes and the correlation of its amplitude with the growth rate in Escherichia coli. *DNA Res* **27**, (2020).
29. Ying BW, Yama K. Gene Expression Order Attributed to Genome Reduction and the Steady Cellular State in Escherichia coli. *Front Microbiol* **9**, 2255 (2018).
30. Engen S, Saether BE. r- and K-selection in fluctuating populations is determined by the evolutionary trade-off between two fitness measures: Growth rate and lifetime reproductive success. *Evolution* **71**, 167-173 (2017).
31. Luckinbill LS. r and K Selection in Experimental Populations of Escherichia coli. *Science* **202**, 1201-1203 (1978).
32. Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* **5**, 323 (2009).
33. Wortel MT, Noor E, Ferris M, Bruggeman FJ, Liebermeister W. Metabolic enzyme cost explains variable trade-offs between microbial growth rate and yield. *PLoS Comput Biol* **14**, e1006010 (2018).
34. Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. *Genome research* **12**, 962-968 (2002).
35. Zhang J. Important genomic regions mutate less often than do other regions. *Nature* **602**, 38-39 (2022).
36. Ying BW, Seno S, Kaneko F, Matsuda H, Yomo T. Multilevel comparative analysis of the contributions of genome reduction and heat shock to the Escherichia coli transcriptome. *BMC Genomics* **14**, 25 (2013).
37. Ying BW, Seno S, Matsuda H, Yomo T. A simple comparison of the extrinsic noise in gene expression between native and foreign regulations in Escherichia coli. *Biochem Biophys Res Commun* **486**, 852-857 (2017).
38. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
39. Mathelier A, Carbone A. Chromosomal periodicity and positional networks of genes in Escherichia coli. *Mol Syst Biol* **6**, 366 (2010).
40. Krogh TJ, Moller-Jensen J, Kaleta C. Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes. *Front Microbiol* **9**, 2019

(2018).
41. Liu L, Kurokawa M, Nagai M, Seno S, Ying BW. Correlated chromosomal periodicities according to the growth rate and gene expression. *Sci Rep* **10**, 15531 (2020).
42. Choe D, *et al.* Adaptive laboratory evolution of a genome-reduced Escherichia coli. *Nature Communications* **10**, (2019).
43. Lu H, *et al.* Primordial mimicry induces morphological change in Escherichia coli. *Commun Biol* **5**, 24 (2022).
44. Ikeda T, *et al.* Hypoosmotic stress induces flagellar biosynthesis and swimming motility in Escherichia albertii. *Commun Biol* **3**, 87 (2020).
45. Maurer LM, Yohannes E, Bondurant SS, Radmacher M, Slonczewski JL. pH regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia coli K-12. *J Bacteriol* **187**, 304-319 (2005).
46. Adler C, Corbalan NS, Peralta DR, Pomares MF, de Cristóbal RE, Vincent PA. The alternative role of enterobactin as an oxidative stress protector allows Escherichia coli colony development. *PLoS One* **9**, e84734 (2014).
47. Peralta DR, Adler C, Corbalán NS, Paz García EC, Pomares MF, Vincent PA. Enterobactin as Part of the Oxidative Stress Response Repertoire. *PLoS One* **11**, e0157799 (2016).
48. May T, Okabe S. Enterobactin is required for biofilm development in reduced-genome Escherichia coli. *Environ Microbiol* **13**, 3149-3162 (2011).
49. Scott M, Hwa T. Shaping bacterial gene expression by physiological and proteome allocation constraints. *Nat Rev Microbiol*, (2022).
50. Dai X, *et al.* Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat Microbiol* **2**, 16231 (2016).
51. Aida H, Hashizume T, Ashino K, Ying BW. Machine learning-assisted discovery of growth decision elements by relating bacterial population dynamics to environmental diversity. *Elife* **11**, (2022).
52. Ashino K, Sugano K, Amagasa T, Ying BW. Predicting the decision making chemicals used for bacterial growth. *Sci Rep* **9**, 7251 (2019).
53. Ashburner M, *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25-29 (2000).
54. Salgado H, *et al.* RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res* **41**, D203-213 (2013).
55. Kurokawa M, Ying BW. Precise, High-throughput Analysis of Bacterial Growth.

J Vis Exp, (2017).

56. Deatherage DE, Traverse CC, Wolf LN, Barrick JE. Detecting rare structural variation in evolving microbial populations from new sequence junctions using breseq. *Front Genet* **5**, 468 (2014).
57. Barrick JE, *et al.* Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* **15**, 1039 (2014).
58. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
59. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
60. Storey JD. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 479-498 (2002).
61. Riley M, *et al.* Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Res* **34**, 1-9 (2006).
62. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462 (2016).
63. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
64. Carbon S, *et al.* AmiGO: online access to ontology and annotation data. *Bioinformatics* **25**, 288-289 (2009).
65. Huang DW, *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* **35**, W169-175 (2007).
66. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
67. Bryant JA, Sellars LE, Busby SJ, Lee DJ. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res* **42**, 11383-11392 (2014).

Table 1 Overview of fixed genome mutations. The number of mutations in the nine Evos are shown separately and summed**.** SNP, N and S indicate single nucleotide substitution, nonsynonymous and synonymous SNP, respectively.

Evos	All	Intergenic	Genic		Genic SNP	
			Indel	SNP	N	S
A1	3	0	0	2	1	1
A2	2	1	0	1	1	0
B2	13	0	1	11	9	2
C2	11	1	0	9	8	1
D2	7	0	1	5	4	1
E2	3	1	0	1	1	0
F2	8	1	0	6	5	1
G2	12	2	2	6	4	2
H2	6	0	2	4	3	1
Sum	65	6	6	45	36	9

Figure legends

Figure 1 Fitness increase of the reduced genome mediated by experimental evolution. A. Temporal changes in growth rate. Color variation indicates the nine evolutionary lineages. **B.** Growth rate and maximal population size of the reduced genome. Blue and pink indicate the common ancestor and the nine evolved populations, respectively. Standard errors are shown according to the biological replications (N=4~6). **C.** Boxplots of growth rate and maximum. Cross and open circles indicate the mean and individual values, respectively. Statistic significance evaluated by Mann-Whitney U tests is indicated. **D.** Correlation between growth rate and maximum. Spearman's rank correlation coefficient and p-value are indicated.

Figure 2 Genomic localization of mutations. A. Normalized genomic positions of all mutations. The vertical lines highlight the total 65 mutations fixed in the nine Evos. Color variation indicates the nine Evos. WT and Reduced represent the wild-type and reduced genomes used in the present study. **B.** Normalized genomic positions of random mutations. The simulation of 65 mutations randomly fixed in the reduced genome was performed 1,000 times. As an example of the simulation, the genomic positions of 65 random mutations are shown. The vertical lines in purple indicate the mutations. **C.** Statistic significance of the genome locational bias of mutations. The distance from the mutated location to the nearest genomic scar caused by genome reduction was calculated. The mutations accumulated in the nine Evos and the 1,000-time random simulation were all subjected to the calculation. The significance of genome locational bias of the mutations in Evos was evaluated by Welch's t-test. The histogram of 1,000 tests for 1,000 simulated results is shown. The mean of *p*-values (μ_p) is indicated, which is within the 95% confidence interval ($0.07 < \mu_p < 0.09$).

Figure 3 Chromosomal periodicity of transcriptome and mutated gene expression. A. Chromosomal periodicity of transcriptomes. The transcriptomes of the nine Evos are shown. Black lines, red curves, and red vertical lines indicate the gene expression levels, fitted periods, and locations of mutations, respectively. *Ori* and *dif* are indicated with the vertical broken lines. **B.** Boxplot of gene expression levels. Gene expression levels

of the 49 mutated genes in the nine Evos and the remaining 3,225 genes are shown. Statistic significance evaluated by Welch's t-test is indicated.

Figure 4 Differentially expressed genes (DEGs) and their enriched functions. A. Commonality of DEGs in the nine Evos. Closed circles represent the combinations of the Evos. Vertical and horizontal bars indicate the number of the overlapped DEGs in the combinations and the number of all DEGs in each Evos, respectively. The combinations with more than 20 DEGs in common are shown. **B.** The number of DEGs overlapped among the Evos. The numbers of DEGs overlapped across 2 to 9 Evos are shown. The number of Evos detected in the single Evos is indicated as 1. **C.** Enriched function in common. The KEGG and GO terms enriched in the common DEGs across the nine Evos are shown. The statistical significance (FDR) of the enriched pathways and biological processes is shown on a logarithmic scale represented by color gradation.

Figure 5 Transcriptome comparison between genome reduction and evolution. A. Venn diagrams of DEGs induced by genome reduction and evolution. The numbers of individual and overlapped DEGs are indicated. **B.** Heatmap of enriched regulons. Statistically significant regulons are shown with the FDR values on a logarithmic scale. **C.** Number of enriched functions in common. Left and right panels indicate the numbers of enriched GO terms and KEGG pathways caused by genome reduction and evolution, respectively. **D.** Enriched functions in common. The overlapped GO terms enriched in the nine Evos and genome reduction are shown. Blue and pink represent genome reduction and evolution, respectively.

Figure 6 Reconstructed gene modules. A. Cluster dendrogram of the gene modules reconstructed by WGCNA. A total of 21 gene modules (M1~M21) were reconstructed. The significance of the correlation coefficients of the gene modules to growth, mutation, and expression is represented in purple gradation. From light to dark indicates the logarithmic p-values from high to low. **B.** Enriched functions of gene modules and deletion. Enriched gene categories, regulons, and GO terms are shown from left to right. The numbers of the genes assigned in the three gene modules and the genomic deletion

for genome reduction are indicated in the brackets. Color gradation indicates the normalized *p* values on a logarithmic scale.

Figure 7 Schematic drawing of evolutionary approaches for the reduced genome.

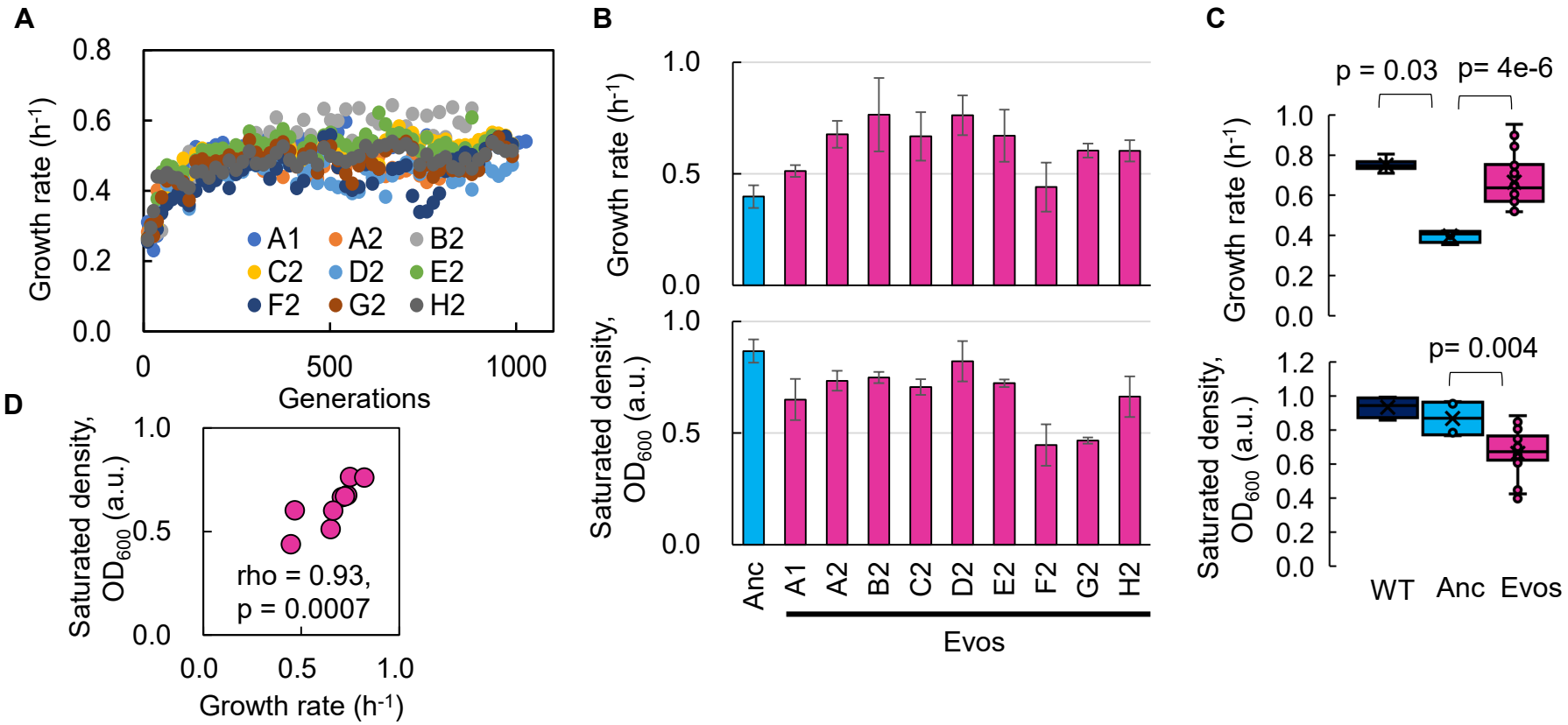
A
Growth rate (h-1)
0.8
0.6
0.4
0.2
0.0
0
500
1000
Generations
A1
A2
B2
C2
D2
E2
F2
G2
H2
B
Growth rate (h-1)
1.0
0.5
0.0
Saturated density, OD600 (a.u.)
Anc
A1
A2
B2
C2
D2
E2
F2
G2
H2
Evos
C
p = 0.03
p= 4e-6
p= 0.004
WT
Anc
Evos
D
rho = 0.93,
p = 0.0007
Growth rate (h-1)

Figure 1

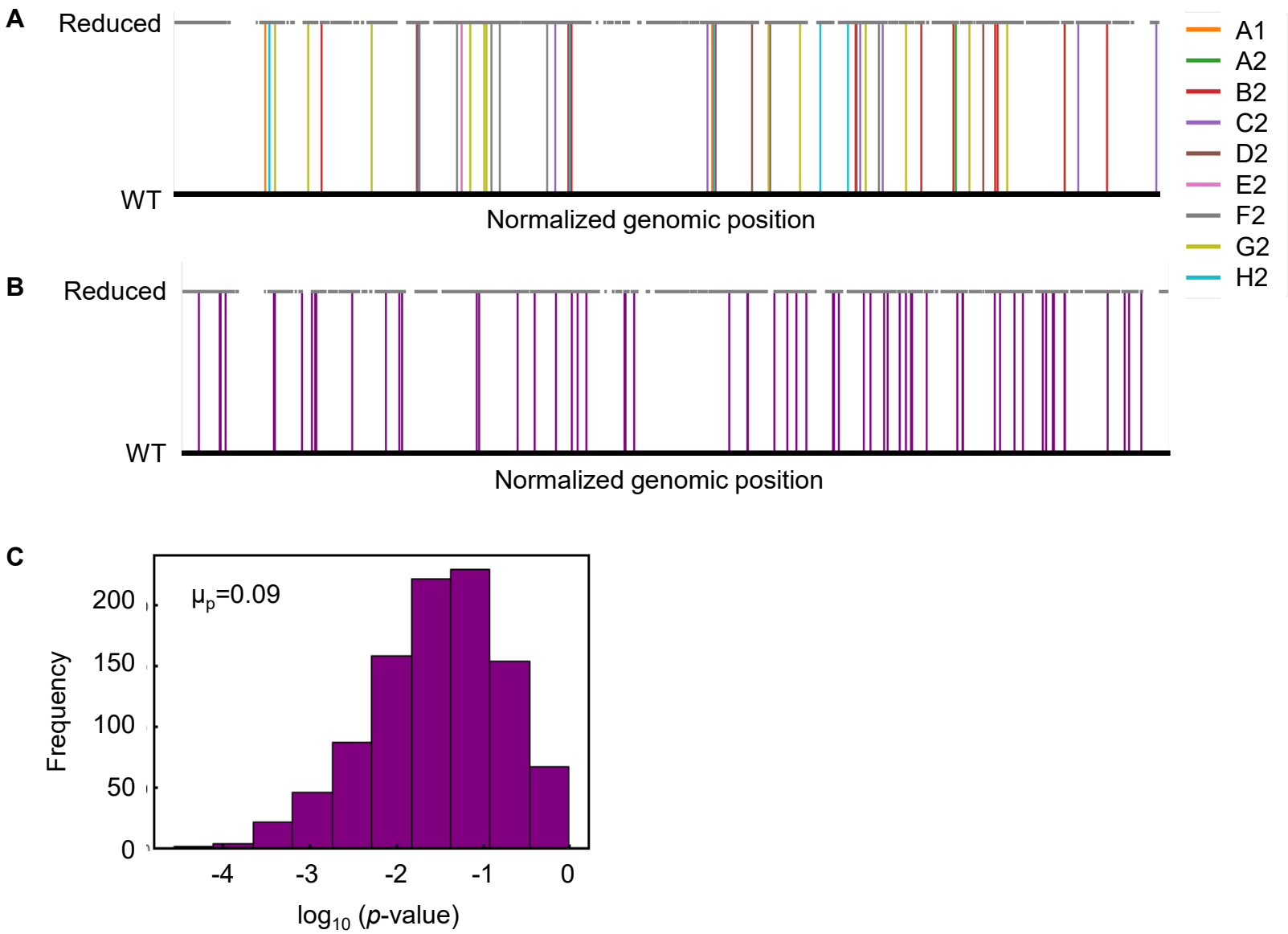
A
Reduced
WT
Normalized genomic position
A1
A2
B2
C2
D2
E2
F2
G2
H2
B
Reduced
WT
Normalized genomic position
C
μp=0.09
Frequency
200
150
100
50
0
-4
-3
-2
-1
0
log10 (p-value)

Figure 2

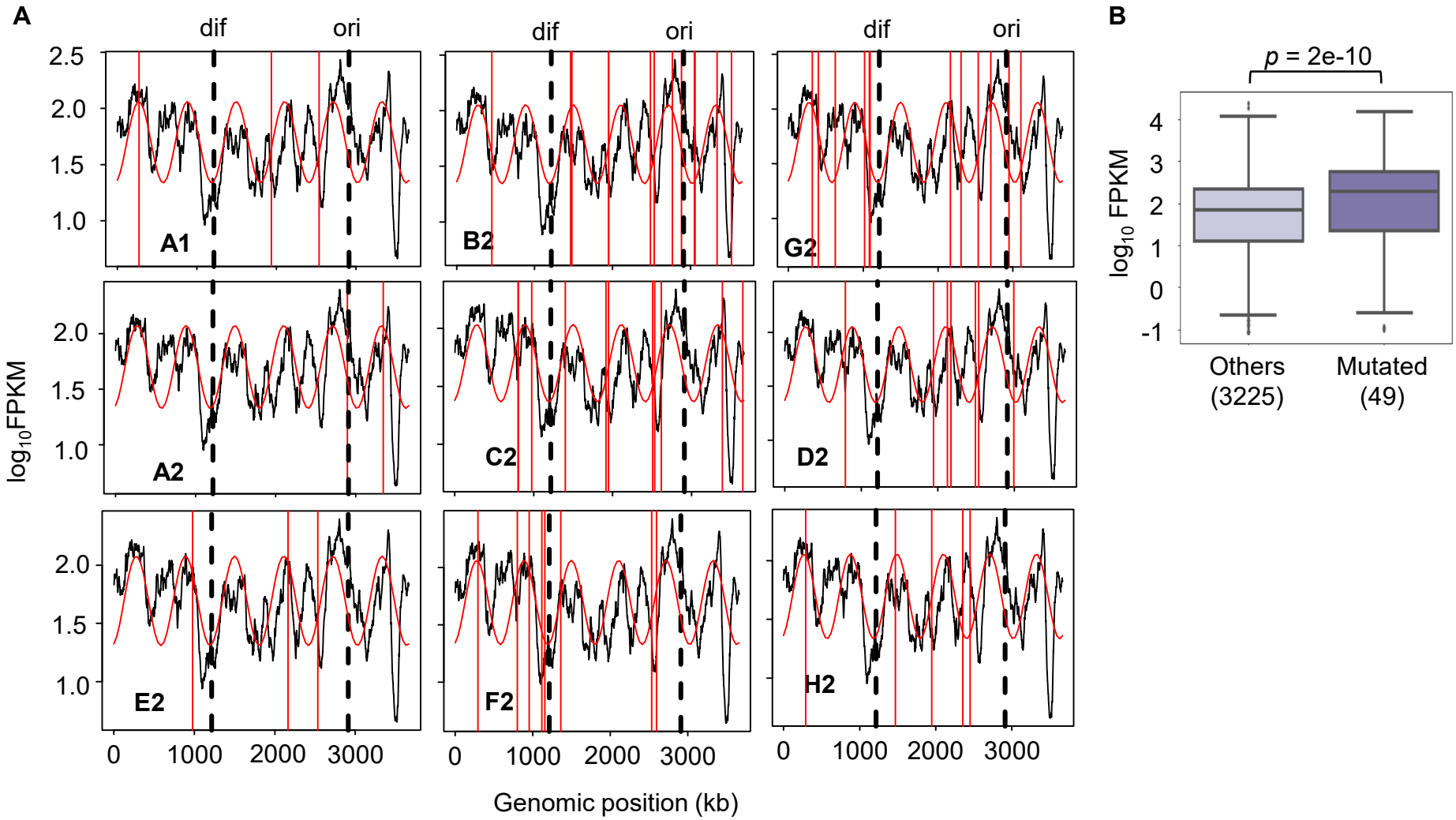
A
dif
ori
A1
B2
G2
A2
C2
D2
E2
F2
H2
log10FPKM
Genomic position (kb)
B
p = 2e-10
log10 FPKM
Others
(3225)
Mutated
(49)

Figure 3

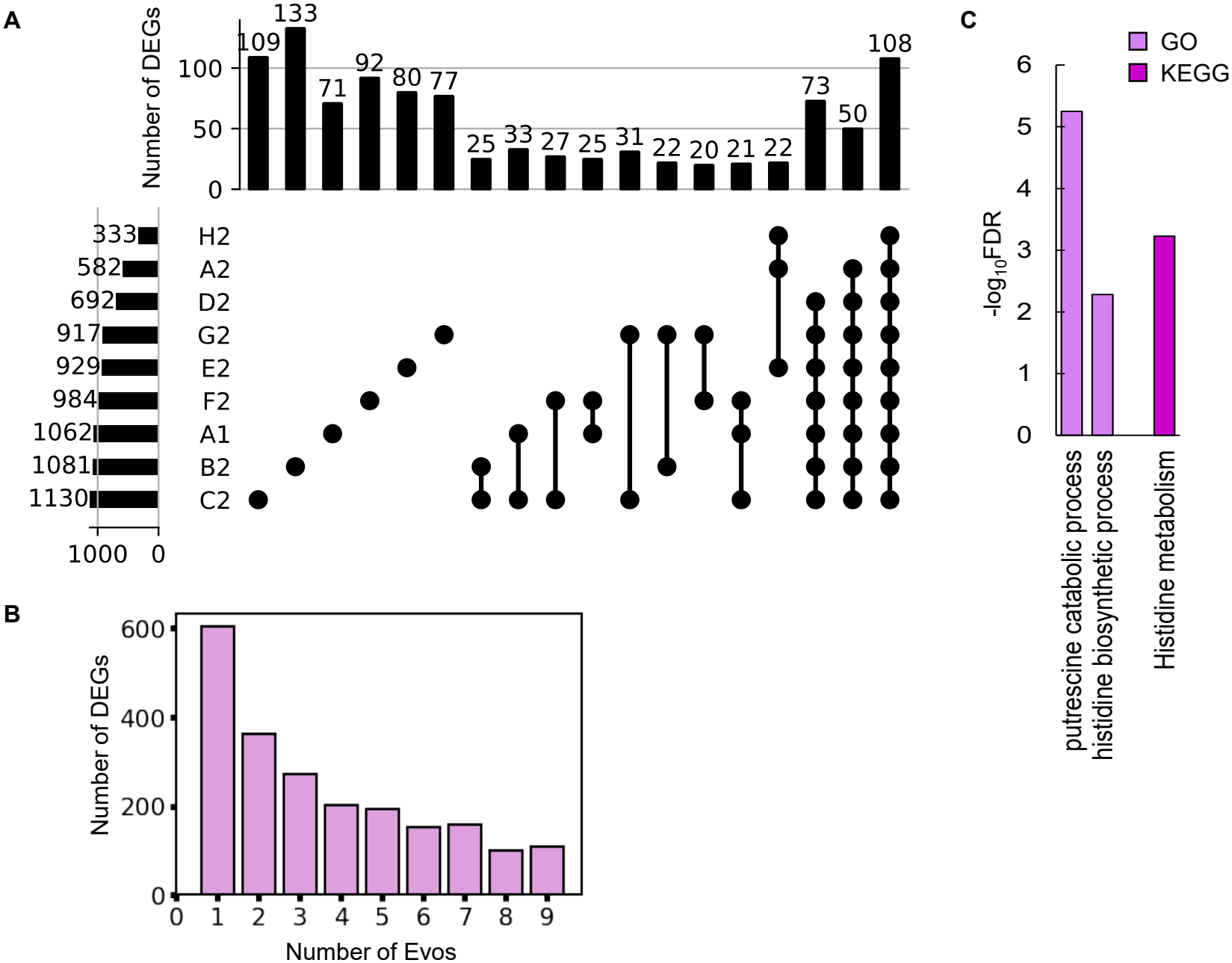
A
Number of DEGs
109
133
71
92
80
77
25
33
27
25
31
22
20
21
22
73
50
108
333
582
692
917
929
984
1062
1081
1130
H2
A2
D2
G2
E2
F2
A1
B2
C2
1000
0
B
Number of DEGs
Number of Evos
C
GO
KEGG
-log10FDR
putrescine catabolic process
histidine biosynthetic process
Histidine metabolism

Figure 4

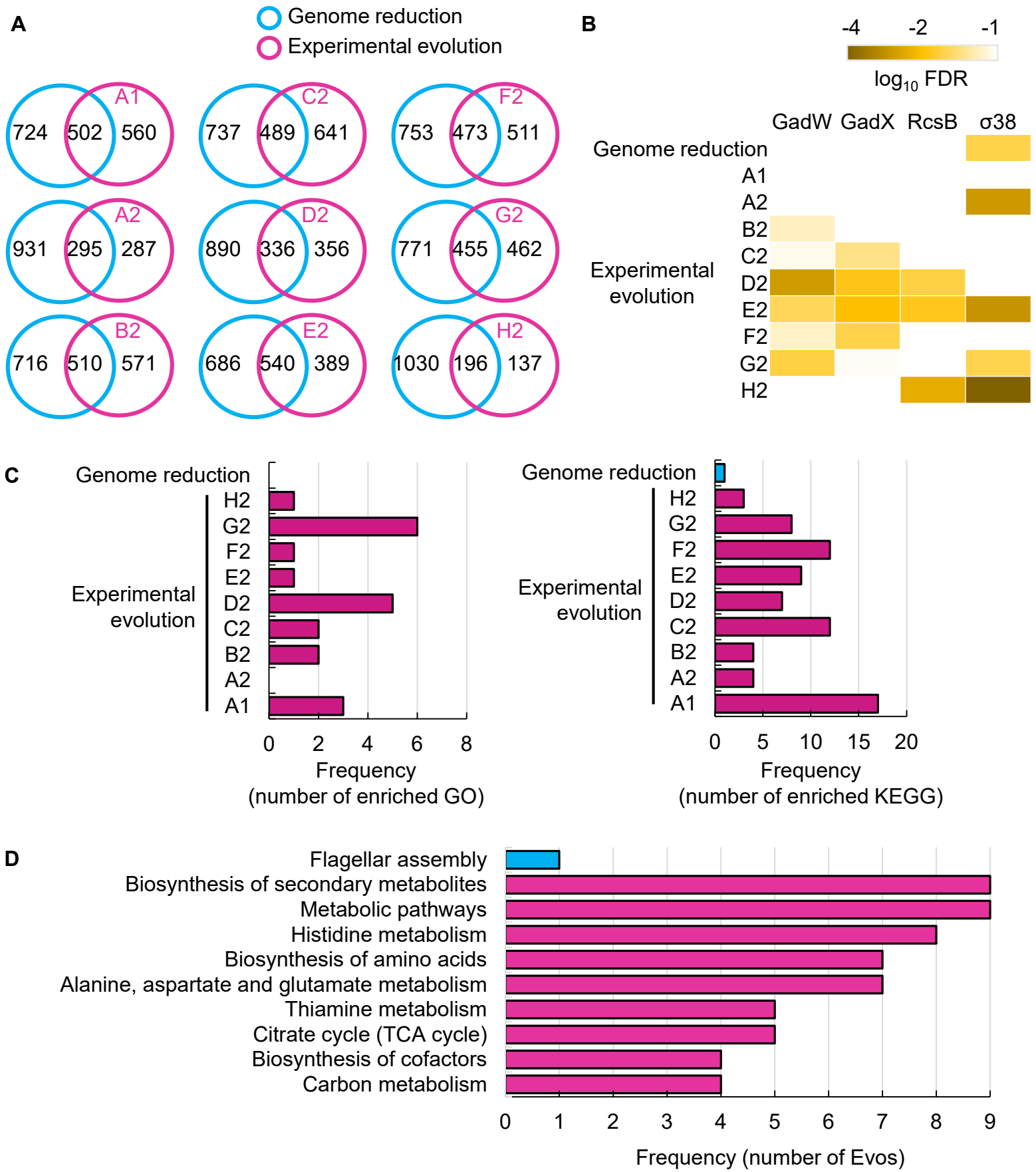
A
Genome reduction
Experimental evolution
A1
724
502
560
C2
737
489
641
F2
753
473
511
A2
931
295
287
D2
890
336
356
G2
771
455
462
B2
716
510
571
E2
686
540
389
H2
1030
196
137
B
-4
-2
-1
log10 FDR
GadW
GadX
RcsB
σ38
Genome reduction
Experimental evolution
A1
A2
B2
C2
D2
E2
F2
G2
H2
C
Genome reduction
Experimental evolution
H2
G2
F2
E2
D2
C2
B2
A2
A1
0
2
4
6
8
Frequency
(number of enriched GO)
Genome reduction
Experimental evolution
H2
G2
F2
E2
D2
C2
B2
A2
A1
0
5
10
15
20
Frequency
(number of enriched KEGG)
D
Flagellar assembly
Biosynthesis of secondary metabolites
Metabolic pathways
Histidine metabolism
Biosynthesis of amino acids
Alanine, aspartate and glutamate metabolism
Thiamine metabolism
Citrate cycle (TCA cycle)
Biosynthesis of cofactors
Carbon metabolism
0
1
2
3
4
5
6
7
8
9
Frequency (number of Evos)

Figure 5

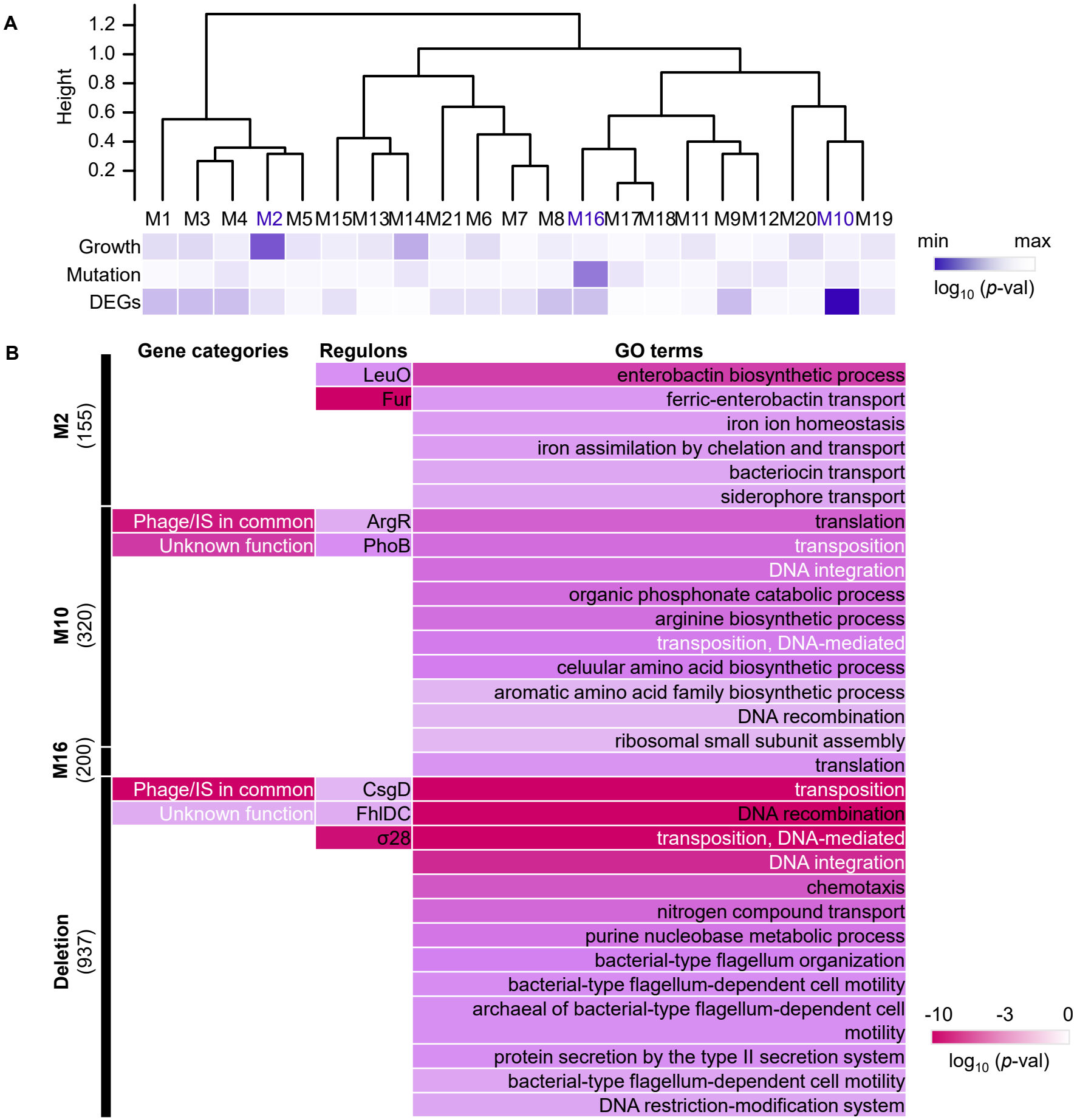
A
Height
1.2
1.0
0.8
0.6
0.4
0.2
M1 M3 M4 M2 M5 M15 M13 M14 M21 M6 M7 M8 M16 M17 M18 M11 M9 M12 M20 M10 M19
Growth
Mutation
DEGs
min
max
log10 (p-val)
B
Gene categories
Regulons
GO terms
M2 (155)
M10 (320)
M16 (200)
Deletion (937)
LeuO
Fur
Phage/IS in common
Unknown function
ArgR
PhoB
Phage/IS in common
Unknown function
CsgD
FhlDC
σ28
enterobactin biosynthetic process
ferric-enterobactin transport
iron ion homeostasis
iron assimilation by chelation and transport
bacteriocin transport
siderophore transport
translation
transposition
DNA integration
organic phosphonate catabolic process
arginine biosynthetic process
transposition, DNA-mediated
celuular amino acid biosynthetic process
aromatic amino acid family biosynthetic process
DNA recombination
ribosomal small subunit assembly
translation
transposition
DNA recombination
transposition, DNA-mediated
DNA integration
chemotaxis
nitrogen compound transport
purine nucleobase metabolic process
bacterial-type flagellum organization
bacterial-type flagellum-dependent cell motility
archaeal of bacterial-type flagellum-dependent cell motility
protein secretion by the type II secretion system
bacterial-type flagellum-dependent cell motility
DNA restriction-modification system
-10
-3
0
log10 (p-val)

Figure 6

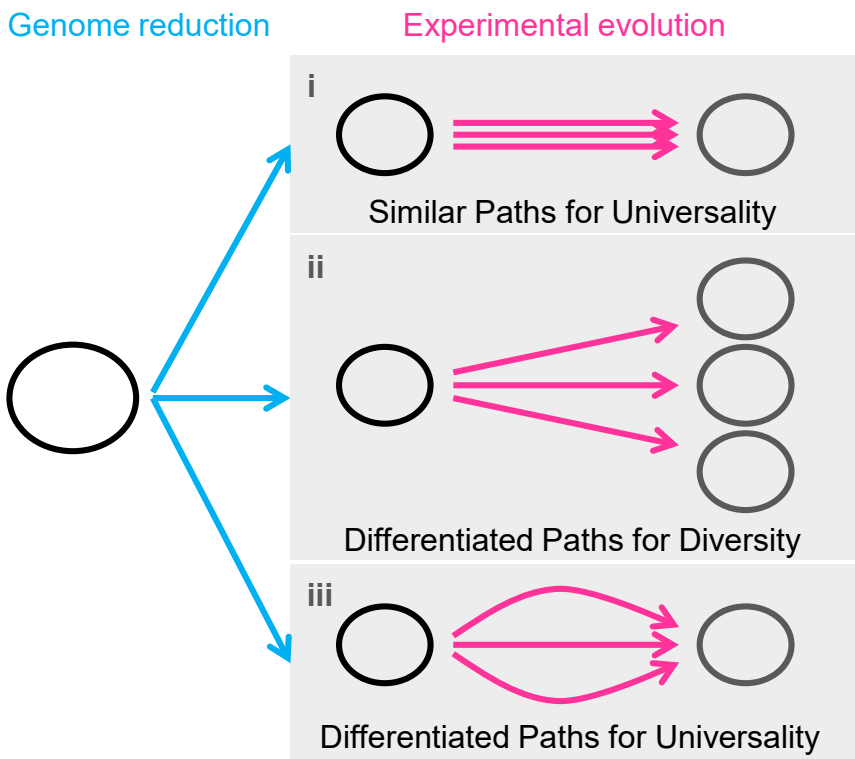
Genome reduction
Experimental evolution
i
Similar Paths for Universality
ii
Differentiated Paths for Diversity
iii
Differentiated Paths for Universality

Figure 7