



## OPEN ACCESS

EDITED BY  
Feng Gao,  
Tianjin University, China

REVIEWED BY  
Weiyi Li,  
Stanford University, United States  
Michihisa Maeda,  
Meiji University, Japan

\*CORRESPONDENCE  
Bei-Wen Ying  
ying.beiwen.gf@u.tsukuba.ac.jp

SPECIALTY SECTION  
This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 11 July 2022  
ACCEPTED 30 August 2022  
PUBLISHED 20 September 2022

CITATION  
Lao Z, Matsui Y, Ijichi S and Ying B-W  
(2022) Global coordination of the  
mutation and growth rates across  
the genetic and nutritional variety  
in *Escherichia coli*.  
*Front. Microbiol.* 13:990969.  
doi: 10.3389/fmicb.2022.990969

COPYRIGHT  
© 2022 Lao, Matsui, Ijichi and Ying.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Global coordination of the mutation and growth rates across the genetic and nutritional variety in *Escherichia coli*

Zehui Lao, Yuichiro Matsui, Shinya Ijichi and Bei-Wen Ying\*

School of Life and Environmental Sciences, University of Tsukuba, Ibaraki, Japan

Fitness and mutability are the primary traits of living organisms for adaptation and evolution. However, their quantitative linkage remained largely deficient. Whether there is any general relationship between the two features and how genetic and environmental variables influence them remained unclear and were addressed here. The mutation and growth rates of an assortment of *Escherichia coli* strain collections, including the wild-type strains and the genetically disturbed strains of either reduced genomes or deletion of the genes involved in the DNA replication fidelity, were evaluated in various media. The contribution of media to the mutation and growth rates was differentiated depending on the types of genetic disturbance. Nevertheless, the negative correlation between the mutation and growth rates was observed across the genotypes and was common in all media. It indicated the comprehensive association of the correlated mutation and growth rates with the genetic and medium variation. Multiple linear regression and support vector machine successfully predicted the mutation and growth rates and the categories of genotypes and media, respectively. Taken together, the study provided a quantitative dataset linking the mutation and growth rates, genotype, and medium and presented a simple and successful example of predicting bacterial growth and mutability by data-driven approaches.

## KEYWORDS

genome reduction, mutator, machine learning, support vector machine, multiple linear regression, mutation rate, growth rate

## Introduction

Growth fitness and genetic mutability are the primary traits as the driving force of adaptation and evolution (Baer et al., 2007; Basan et al., 2020). The growth rate and the mutation rate are the quantitative parameters representing the adaptiveness and evolvability of the living organisms, respectively (Elena et al., 2007;

Korem Kohanim et al., 2018; Zheng et al., 2020). Theoretical studies usually reported the fitness-dependent mutation rates upon mathematical simulation (Agrawal, 2002; Shaw and Baer, 2011). Experimental evolution often found elevated mutation rates along with fitness increase (Barrick et al., 2009; Kishimoto et al., 2010; Lenski, 2017; Shibai et al., 2017), which was practical to rescue the growth from genetic engineering in laboratory (Kurokawa and Ying, 2019). The highly increased mutation rate, i.e., mutator, usually resulted in a decrease in growth rate (Funchain et al., 2000; Ishizawa et al., 2015), and the fitness increase associated with the mutation rates (Couce et al., 2017; Nishimura et al., 2017). The experimental observations and theoretical assumptions suggested that the mutation and growth rates were associated. However, the quantitative linkages of the mutation and growth rates remained largely deficient. Whether and how the genetic and environmental changes disturb the relationship between mutation and growth rates are unclear, although both growth fitness and mutability are supposed to be constrained by the genetic and environmental conditions.

The previous studies strongly suggested that genetic disturbance influenced the mutation and growth rates. The reduced genome collections were constructed with bacterial strains (Hashimoto et al., 2005; Kato and Hashimoto, 2007; Mizoguchi et al., 2008; Karcagi et al., 2016) to discover the minimal genetic information essential for living systems (Mizoguchi et al., 2007; Reuß et al., 2017; Breuer et al., 2019; Rees-Garbutt et al., 2020) and to benefit the genetic engineering for substrate production (Sharma et al., 2007; Morimoto et al., 2008) and metabolic rewiring (Lee et al., 2009). The systematic assays showed that the genome reduction caused the decreased growth rates (Karcagi et al., 2016; Kurokawa et al., 2016) accompanied by increased mutation rates despite the regular mismatch repair (MMR) system (Nishimura et al., 2017). In comparison, the genetically disturbed MMR systems significantly induced the mutation rates, often associated with reduced growth rates (Ishizawa et al., 2015). These studies investigated the relationships between the mutation and growth rates in the genome-reduced and MMR deficient strains, demonstrating that both genome reduction and MMR deficiency participated in the changes in the mutation and growth rates.

Nevertheless, whether and how the coordination of the mutation and growth rates in the genome-reduced and MMR deficient strains responded to the environmental diversity remained unclear. Variation in growth media was representative of environmental diversity. Media variation could intuitively disturb the growth rates, as experimentally demonstrated with the wild-type (Aida et al., 2022), single-gene knockout (Liu et al., 2020), genome-reduced (Kurokawa et al., 2016), and laboratory-evolved strains (Kurokawa et al., 2022). The changes in growth media also adjusted the mutation rates of the wild-type and genome-reduced strains (Nishimura et al., 2017) and those of MMR deficient strains (Ishizawa et al., 2015). These

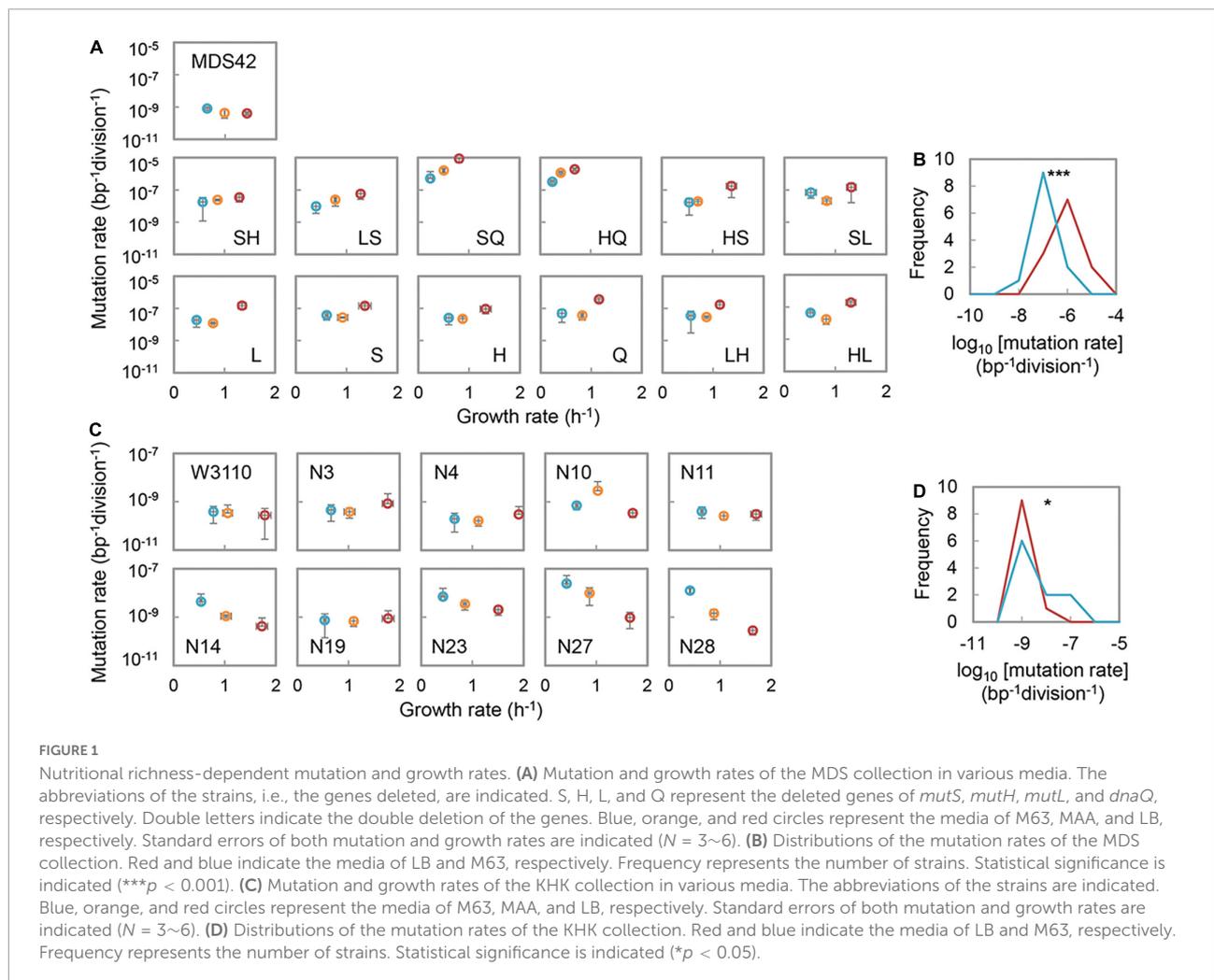
experimental findings revealed that nutritional richness affected the mutation and growth rates; resultantly, it might influence the relationship between the mutation and growth rates.

As the genetic and environmental variables contributed to the mutation and growth rates and probably interrupted their coordination, whether and how these variables and parameters related to each other are intriguing questions. In the present study, we investigated whether there's any global pattern among the mutation and growth rates, genetic and environmental variables, and, if applicable, how they coordinated with each other. Instead of the mechanistic interpretation, we attempted to construct simple models for quantitative understanding of the genetic and environmental contributions to the coordination of the mutation and growth rates. As a pilot survey, the genome reduced strains, and the newly constructed mutator strains were assayed to examine the genetic contribution. Three media representing varied nutritional richness were tested to investigate the environmental contribution. Both theoretical regression and machine learning were applied to discover a quantitative and global linkage taking the genotype, the medium, and the mutation and growth rates into account.

## Results and discussion

### Nutritional richness mediated changes in the mutation and growth rates

To investigate the nutritional-dependent changes in mutation rate, two *E. coli* strain collections were analyzed. The MDS collection was newly constructed from the cleaned genome strain of MDS42 by deleting the genes participating in the mismatch repairing and DNA replication fidelity systems, i.e., *mutS*, *mutH*, *mutL*, and *dnaQ* (Echols et al., 1983; Rewinski and Marinus, 1987; Yang, 2000). A total of 13 strains were used in the present study. The KHK collection was a reduced genome library constructed from the wild-type strain of W3110, previously (Mizoguchi et al., 2008). Ten strains of varied genome sizes in the KHK collection were chosen for the test, as described previously (Nishimura et al., 2017). Repeated tests showed that an increase in mutation rates in response to the nutritional enrichment was observed in the MDS collection. However, the mutation rate of the reduced genome, i.e., their parent strain MDS42, was likely to be decreased in rich media (Figure 1A). Although the degrees of change in mutation rates were varied, the directional shift of the distributions of mutation rates from poor (M63) to rich (LB) media was highly significant ( $p = 8e-7$ ) (Figure 1B). The tendency was consistent with the directional changes in mutation rates of the MG collection in response to the changes in media (Supplementary Figure 1), although the dataset was obtained with different assay methods previously (Ishizawa et al., 2015). Nutritional enrichment



increased the mutation rates of the mutators, independent of the genomes.

On the other hand, a decrease in mutation rates of the KHK collection in response to the nutritional enrichment was identified (Figure 1C). The distribution of mutation rates was slightly but significantly ( $p = 0.03$ ) shifted from low to a high level in response to the medium alteration from LB to M63 (Figure 1D). The nutritional richness mediated changes in mutation rates were somehow reasonable. Increased mutation rates of the wild-type genomes and the mutators more often caused the mutations that triggered the damage to the cells; however, the sufficient nutrients in the rich media could compensate for the damages caused by mutation. The specific nutrients in the medium might have buffered the fitness decrease caused by the mutations (Casaneuva et al., 2012; Kinsler et al., 2020; Eisner et al., 2022), which allowed high mutability. Increased mutation rates of the reduced genomes in the poor media could be elucidated by the severe environmental stress-induced mutagenesis (Maharjan and Ferenci, 2017; Frenoy and Bonhoeffer, 2018). The differentiated directions of the

changes in mutation rates responding to the nutritional changes suggested that the large deletion of genomic fragments and the interruption specifically in replication fidelity contributed to the mutability differentiation. Note that the growth rates were universally increased in the order of M63, MAA, and LB, regardless of the genotypes (Figures 1A,C), which well-reflected the expected contribution of the nutritional richness to the growth fitness.

## Genetic disturbance mediated changes in the mutation and growth rates

A common negative correlation of growth rate to mutation rate was highly significant in all media, despite the variation in genotypes, including the reduced genomes and mutators (Figure 2A and Supplementary Figure 2). The directional changes in mutation rate were associated with the changes in growth rate. Combining the reduced genomes and the mutators resulted in the global correlation between mutation and growth

rates in the three media (Figure 2A). Even if the three datasets were combined, both the trend of negative correlation between the mutation and growth rates and the differentiation in the slopes remained clearly (Supplementary Figure 2). The decreased growth rate accompanied by increased mutation rate was independent of the nutritional richness. The global parameters of the mutation and growth rates were stringently connected, as a common phenomenon evidently in reduced genomes, i.e., the MDS and KHK collections. Although the fitness effect of mutations depended on the genomic background (Wang et al., 2016), the trajectory across the various genotypes indicated the coordination of the mutability to the fitness.

A simple regression (Eq. 1) showed that all strains of various genotypes followed a common trajectory formed by the mutation and growth rates, and the magnitudes of the correlated changes were medium-dependent.

$$\mu_i = LN(M_i/M_\infty)/\alpha \quad (1)$$

Here,  $M_i$  and  $\mu_i$  represent the mutation rate and the corresponding growth rate in a defined condition.  $M_\infty$  and  $\alpha$  indicate the maximal mutation rate when the growth rate dropped to zero and the magnitude of the growth decrease caused by the increased mutation rate (slope), respectively. Both  $M_\infty$  and  $\alpha$  are nutritional dependent but genotype independent and could be calculated according to the experimental data sets. The slope of  $\alpha$  was  $-9.5$ ,  $-12.6$ , and  $-12.8$  in LB, MAA, and M63, respectively (Supplementary Figure 2). Increased mutation rate led to a similar degree of growth decrease in MAA and M63 but a more significant reduction in LB. In addition,  $M_\infty$  were  $1e-3$ ,  $4e-4$ , and  $7e-6$   $\text{bp}^{-1} \text{division}^{-1}$  in LB, MAA, and M63, respectively, suggesting that the nutritional richness decided the maximal mutation rate. Since the genome size was approximately  $4 \times 10^6$  bps, the poor nutritional condition allowed only a few mutation(s) per genome compared to the rich nutritional condition that allowed more than 1,000 mutations per genome.

In addition, the positive correlations of growth and mutation rates between any pair of the media were observed (Figure 2B). It demonstrated that the changes in nutritional richness maintained the order of the growth fitness and mutability. However, the direction of the changes in mutation rates due to the genetic disturbance was somehow differentiated (Figure 3A). The changes in mutation rates caused by MMR deficiency were more significant in rich media (Figure 3A, Upper) compared to the changes mediated by genome reduction, which were more prominent in poor media (Figure 3A, Bottom). Nevertheless, the changes in growth rates in response to the nutritional alteration were roughly identical in both collections (Figure 3B). Taken together, it was highly intriguing that the negative correlations between the growth rate and the mutation rate were in common once the medium was determined; however, the directions of the nutritional

responsivity of the mutation rate were reversed decided by the types of genetic disturbance.

## Comprehensive association of the mutation and growth rates with genotypes and media

The exhaustive tests identified a common trajectory directed the correlated changes in the mutation and growth rates across varied genotypes at all nutritional levels. The results were highly consistent with the previous studies on the correlation of the mutation and growth rates in MG1655 (Ishizawa et al., 2015), as well as the coordinated changes in the mutation and growth rates mediated by genome reduction (Nishimura et al., 2017). It suggested a trade-off relationship between the growth rate and the mutation rate in common, which was feasible in evolution (Nishimura et al., 2017; Kang et al., 2019). Experimental evolution of the KHK collection demonstrated that equivalent generation led to equal numbers of genome mutations fixed on the reduced genomes but none in the wild-type genome; nevertheless, their growth rates were improved comparably (Kurokawa et al., 2022). To further confirm the generality of the relationship between the mutation and growth rates, the evolved KHK collection strains were additionally tested. The increased growth rates resulting from the experimental evolution were associated with reduced mutation rates (Supplementary Figure 3). However, the mutations fixed in experimental evolution were unrelated to the DNA replication fidelity and MMR system (Kurokawa et al., 2022). The correlated changes in the mutation and growth rates were in common. The trade-off mechanisms were crucial in bacterial competition and coexistence (Ferenci, 2016) and shaped the diversity of species in eco-evolution (Farahpour et al., 2018). The negative correlation of the mutation and growth rates across the genetic variation well-agreed with those reported trade-offs, which might be relied on the cost-benefit working principle in living systems (Eames and Kortemme, 2012; Weisse et al., 2015; Erickson et al., 2017).

Intriguingly, genome reduction and MMR deficiency showed differentiated directions of the correlated changes in the mutation and growth rates in response to nutritional variety (Figure 4A, Upper). The positive correlation of the mutation rate to the growth rate triggered by MMR deficiency supported the theoretical framework, which proposed that the mutation accumulation rate increased with the cell division rate across species (Gao et al., 2016). In contrast, both genetic disturbances presented the universal direction in defined media (Figure 4A, Bottom). The present study first observed the reversibility in the direction of the medium-dependent changes in mutation rate. The compensability between genome reduction and MMR deficiency might be tuning the DNA replication errors regulated by the gene expression to maintain

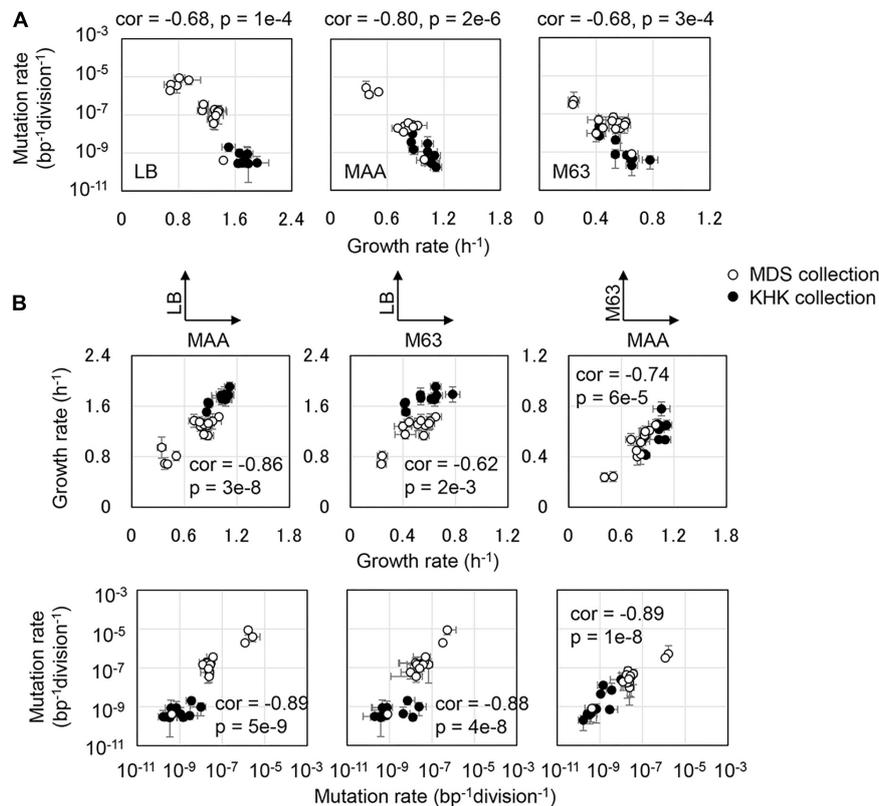


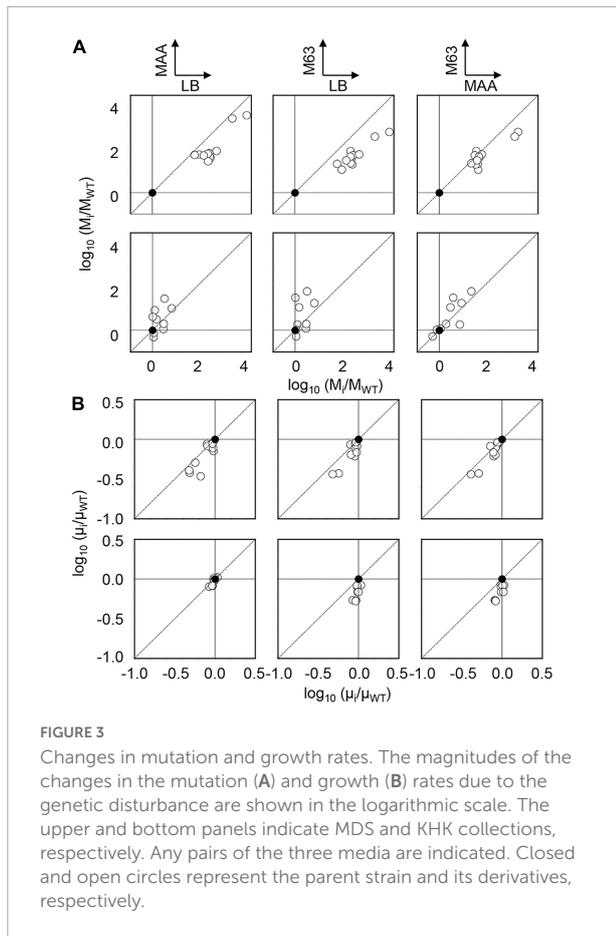
FIGURE 2

Coordinated changes in mutation and growth rates. (A) Correlation between growth rate and mutation rate in the defined medium. Pearson correlation coefficients, the  $p$  values, and the media are indicated. (B) Correlated changes of the mutation and growth rates in various media. The upper and bottom panels show the correlated changes in growth and mutation rates between any two different media, respectively. Open and closed circles stand for the collection of MDS and KHK collections, respectively.

the balance between fitness and evolvability in response to environmental changes. Suppose genome reduction and MMR deficiency had caused the correlated changes of the mutation and growth rates in an identical but not reverse direction. In that case, it must have been disadvantageous for survival, as the high mutation rates might restrict the adaptive evolution for growth improvement (Sprouffske et al., 2018). Theoretically, the maximal mutation rate was assumed to be  $10^{-5}\sim 10^{-3}/\text{bp}/\text{division}$  in different growth media (Eq. 1, Figure 2), which were higher than those of RNA viruses (Krašovec et al., 2017; Duffy, 2018). However, the experimentally acquired mutation rates were lower than  $\sim 10^{-6}/\text{bp}/\text{division}$ , independent of the genetic backgrounds of *E. coli* (Ishizawa et al., 2015; Krašovec et al., 2017; Shibai et al., 2017; Ramiro et al., 2020). As the genome size of *E. coli* was  $\sim 4$  Mb, the mutation rate detrimental to survival was assumed to be  $\sim 10^{-6}/\text{bp}/\text{division}$ , which led to roughly one division for one mutation.

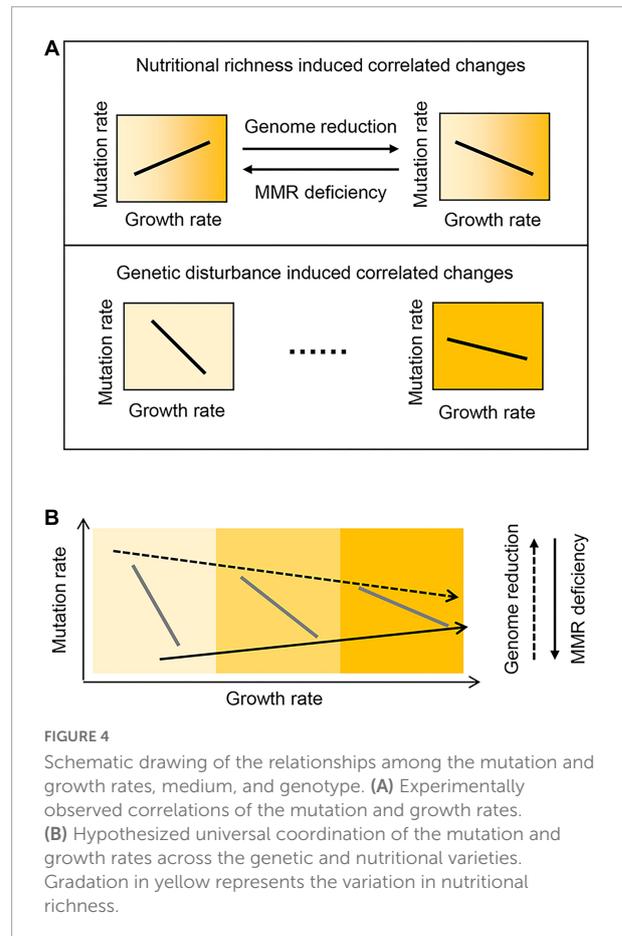
The comprehensive association of the mutation and growth rates with the genotype (genetic interruption) and medium (nutritional richness) was supposed to be universal (Figure 4B).

The biological mechanism in charge of their connections, if it existed, was largely unknown. It was somehow hard to be addressed so far. A preliminary analysis showed that the expression levels of the genes participating in the DNA replication and mismatch repair were either increased or decreased in the KHK collection strains (Supplementary Figure 4). However, these reduced genomes maintained the regular DNA replication and mismatch repair systems (Mizoguchi et al., 2008; Kurokawa et al., 2016). It indicated that the biological mechanisms could not directly explain the correlated changes in mutation rates. High-throughput transcriptome analysis was required to acquire the big data linking the global changes in gene expression to the mutation and growth rates. Instead, the theoretical understanding of the relationships might be practical. Suppose the relationships between the mutation and growth rates, the medium, and the genotype were universal. Any of them might be theoretically estimated according to the other three parameters. The prediction with the regression and machine learning approaches was challenged to achieve a quantitative understanding of the cooperative relationships among the four parameters.



## Estimation of the mutation and growth rates by multivariable regression

Whether the mutation and growth rates could be estimated according to the other parameters was tested. Multiple linear regression (MLR) was applied, where the logarithmic values of mutation rates were used, and both the media and the genotypes were set as the numerals ([Supplementary Table 2](#)). Note that the genotype, which represented the combination of the reduced genome and mutator genotype, was evaluated in two different modes in MLR. Whether the genome reduction and the MMR deficiency interacted or not were defined as interactive and additive modes, respectively. MLR showed that the mutation rate could be well-estimated according to the growth rate, the medium, and the genotype ([Figure 5A](#)), independent of the mode applied for the genotype ([Supplementary Table 3](#)). The growth rate was also successfully predicted by MLR according to the mutation rate, the medium, and the genotype in the mode of interactive ([Figure 5B](#)). Although the prediction accuracy of the additive mode remained equivalent to that of the interactive mode, the contribution of the genotype to the regression (prediction) was insignificant ([Supplementary Table 3](#)). The results suggested genome reduction and deficient MMR.



Moreover, another dataset was employed to address whether the predictivity of the mutation and growth rates was general, as varied strains and methods might cause different consequences. The *E. coli* strain collection derived from MG1655 was adopted from the previous study ([Ishizawa et al., 2015](#)). MLR of the three different datasets comprising the MDS, KHK, and MG collections, that is, the W3110, MDS42, and MG1655 derivatives, showed that the prediction of the mutation rate was statistically significant. However, the accuracy was much lower ([Supplementary Table 4](#)). On the other hand, the growth rate prediction was somehow insignificant ([Supplementary Table 4](#)), probably because of the differentiation in the methods of growth assay. In conclusion, the mutation and growth rates were roughly predictable to each other when the nutritional richness and genotype were decided.

## Classification of the medium and genotype by machine learning

Alternatively, whether the medium and the genotype could be clustered according to the mutation and growth rates were investigated by machine learning ([Jordan and Mitchell, 2015](#)),

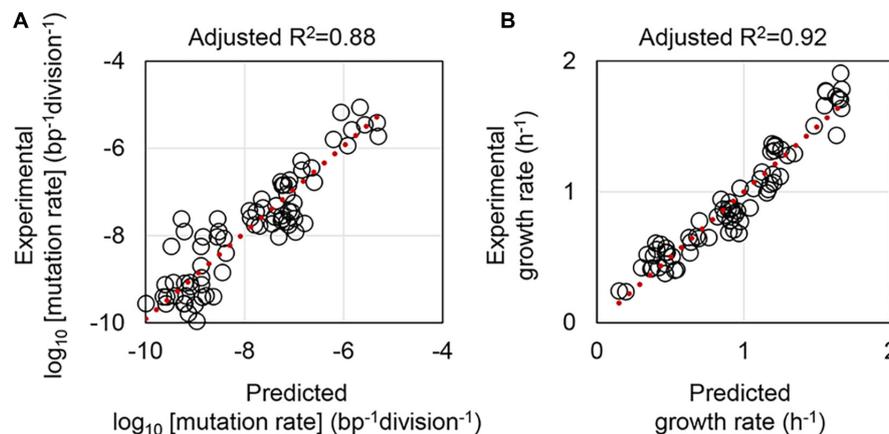


FIGURE 5

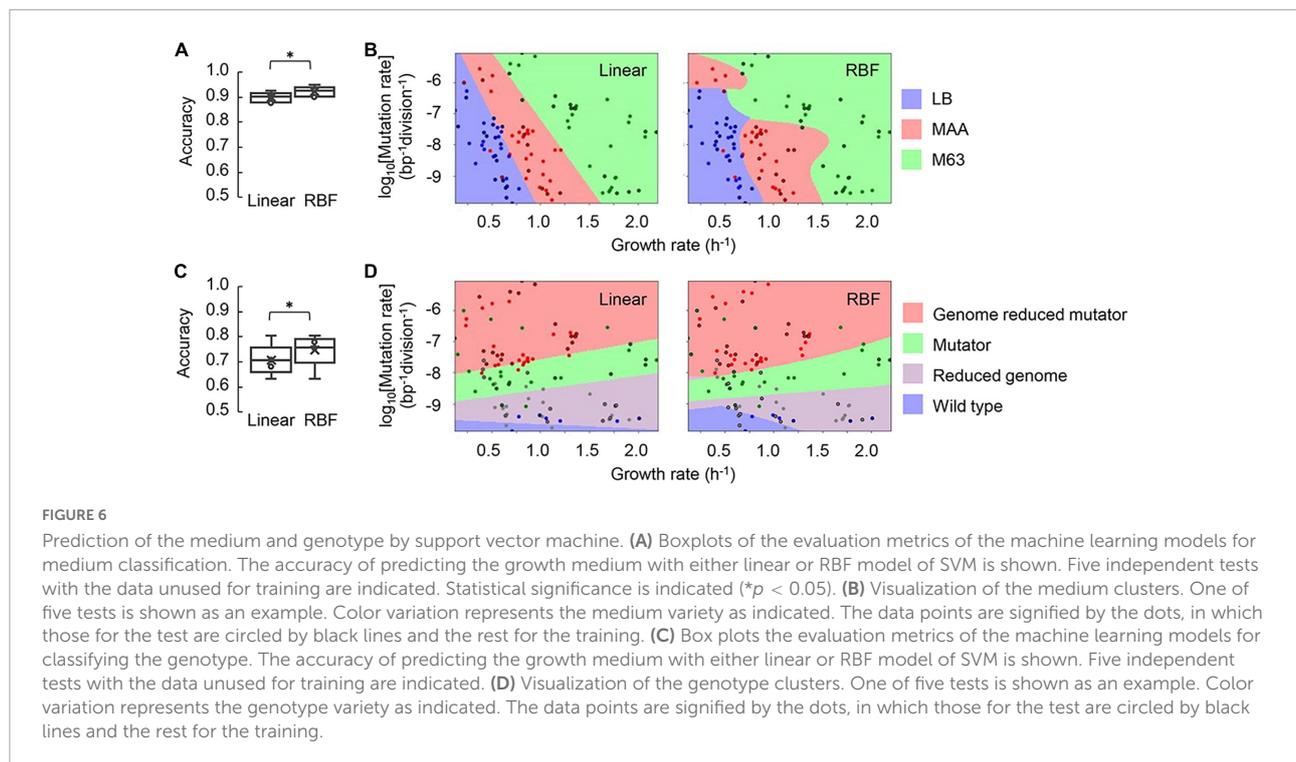
Prediction of the mutation and growth rates by multiple linear regression. Adjusted R<sup>2</sup> denotes the accuracy of multiple linear regression. Red dotted lines indicate the slope of 1. (A,B) Represent the mutation and growth rates, respectively. The corresponding statistics are summarized in [Supplementary Table 3](#).

which was beneficial for discovering the dataset with unclear mechanisms relating to metabolism (Cuperlovic-Culf, 2018; Kim et al., 2020), genetics (Libbrecht and Noble, 2015; Schrider and Kern, 2018), evolution (Wang et al., 2018) and population dynamics (Ashino et al., 2019; Cao et al., 2020; Gilpin et al., 2020). Here, the support vector machine (SVM) was applied to classify the three media (i.e., LB, MAA, and M63) and the four genotypes (i.e., reduced genome, mutator, genome reduced mutator, and wild type). All three datasets (i.e., the MDS, KHK, and MG collections) were subjected to SVM machine learning. The linear regression and the radial basis function (RBF) models were tested, as both the quantitative parameters of the mutation and growth rates and the qualitative parameters of the genotype and medium were comprised in the datasets. The growth medium was well-predicted from the mutation and growth rates by both models (Figure 6A). The RBF model presented a higher accuracy of medium classification, either in testing (Figure 6A) or in training (Supplementary Figure 5). Visualization of the medium variety clearly showed the discontinuous edges/areas of MAA in the landscape formed by the mutation and growth rates in the RBF model (Figure 6B). The unique landscape in the MAA medium might be attributed to either nutritional or generic biases. The MAA medium was rich in amino acids and remained poor in other nutritional elements, such as glucose. As fast growth preferred catabolism of amino acids and reduced glucose uptake in *E. coli* (Zampieri et al., 2019), the enriched amino acids in MAA possibly benefited the growth even with a high mutation rate. In addition, the reduced genomes had low redundancy in non-essential sequences. As the genomic sequences were related to the turnovers of carbon, nitrogen, and sulfur (Baker et al., 2015), genome reduction might disturb the cycling of these essential elements, causing slow growth even with a low mutation rate.

Taken together, the coordination of amino acids and glucose catabolism and the reduced redundancy of the genome might trigger the discontinued landscape of the mutation and growth rates.

In addition, the two SVM models predicted the genotype with comparable accuracy in the training (Supplementary Figure 5). Nevertheless, the prediction accuracy of the RBF model was significantly higher than that of the linear model in the test (Figure 6C). Both models could roughly categorize the four genotypes (Figure 6D). It demonstrated that a simple SVM machine learning model could provide an acceptable classification of the genotype, despite the prediction accuracy being lower than that of classifying the medium. An increased number of data points were required to achieve better precision.

In summary, the present study found an intriguing global correlation of the growth rate to the mutation rate across a wide genetic variety and nutritional variation. As the mutation rate was proposed as a plastic trait associated with population density across domains of life (Krašovec et al., 2017), the cooperative relationship between the mutation and growth rates was supposed to be universal in living systems (Shaw and Baer, 2011). The results expanded the previous findings of the correlated changes in the mutation and growth rates with the defined genotypes (Wang et al., 2018) or the defined conditions (Ishizawa et al., 2015; Nishimura et al., 2017). It indicated the fundamental working principle in maintaining cellular homeostasis. Genetic disturbance in the genome length or the MMR system led to the coordinated changes of mutation rate and growth fitness in common but the reverse directional changes in response to nutritional richness. It strongly suggested that the genetic deficiency could partially compensate for each other, that is, the genome reduction might recover the damage caused by the disturbed MMR system



to some extent. The simple regression and machine learning models connected the growth rate, mutation rate, genotype, and medium well. The successful connection among these parameters indicated that the fitness and evolvability were intrinsically associated in living cells, regardless of the genetic, and environmental interruptions. To develop an advanced ML model taking the underground biological mechanisms into account, the transcriptome analysis is required to connect the mutation and growth rates to the gene expression. The ML predicted mechanisms remain to be challenged in the future.

## Materials and methods

### Genetic construction of the MDS mutators

Genome-reduced mutators were constructed by deleting the genes that participated in mismatch repair or proofreading, i.e., *mutH*, *mutS*, *mutL*, and *dnaD*, from the reduced genome MDS42 (Pósfai et al., 2006). Genetic deletion of the single genes and any pairs of these genes was performed as described previously (Ying et al., 2014; Ishizawa et al., 2015). Briefly, the deletions were induced *via* standardized  $\lambda$ -red homologous recombination (Kirill and Barry, 2000; Ying et al., 2010). The chloramphenicol resistance gene was used repeatedly as a selection marker in each deletion. Genetic and phenotypic

verification of the transformants was carried out as previously described (Ying et al., 2014; Ishizawa et al., 2015). The primers for genetic deletion and PCR confirmation were reported previously (Ishizawa et al., 2015). A total of 13 mutator strains were successfully constructed. One out of 13 mutators failed to grow in the minimal medium, as the genetic construction was performed with the rich medium. Note that the gene circuit sequence was previously used in various studies (Ying et al., 2014, 2015, 2017; Ishizawa et al., 2015; Kishimoto et al., 2015; Shibai et al., 2017; Lu et al., 2022), and no experimental bias was observed.

### Media

Three different media of LB, MAA, and M63 were used for cell culture, representing rich, supplementary, and poor nutritional conditions. The LB medium was commercially available (Luria-Bertani, Sigma). The M63 and MAA media were the minimal medium and the minimal medium supplied with 20 amino acids, respectively, prepared as described previously (Kurokawa et al., 2016; Nishimura et al., 2017).

### Growth assay

The growth dynamics in the three media were assayed as described previously (Kurokawa et al., 2016; Kurokawa and Ying, 2017). The cell culture was diluted and

loaded to the 96-well plate (Costar), which was incubated in a plate reader (Epoch2, BioTek) with a rotation rate of 600 rpm at 37°C. The cell growth was detected at an absorbance of 595 nm, reading at an interval of 30 min or 1 h for 24 to 48 h. The growth curves were obtained for each well. Repeated tests ( $N = 6\sim 12$ ) were performed to acquire the growth curves in each condition. The growth rates were calculated as described previously (Kurokawa et al., 2016; Nishimura et al., 2017).

## Fluctuation test

The mutation rate was estimated by the fluctuation test according to resistance to the antibiotic nalidixic acid, as described previously (Ishizawa et al., 2015; Nishimura et al., 2017). In brief, the number of cells was counted using a CFU assay. The *E. coli* cell cultures in the exponential phase were diluted and plated onto  $\sim 10^8$  LB plates for CFU assay. Only the number of colonies per plate ranging from 10 to 500 was considered reliable for calculating the CFU. Approximately 30 tubes of identical cell culture were used to evaluate the frequency of mutagenesis for each test. At least three repeated tests were performed for each strain at each medium, and more than 8,000 agar plates were used. The mutation rates were calculated as described previously (Kishimoto et al., 2010; Ishizawa et al., 2015). Note that the mutation rate was evaluated based on the emerging frequency of nalidixic acid resistance; nevertheless, we previously verified that the relative mutation rates did not change in response to different antibiotics (Ishizawa et al., 2015).

## Data acquisition

The mutation and growth rates of the MDS collection (16 strains) were obtained in the present study. The reduced genomes of *E. coli* strains, i.e., the derivatives of W3110, were randomly selected from the KHK (Kyowa Hakko Kirin) collection (Mizoguchi et al., 2008), as described previously (Nishimura et al., 2017). Note that the KHK collection strains were constructed in an accumulative deleted manner. The mutation and growth rates of the ten KHK strains were partially adopted from our previous report (Nishimura et al., 2017). Additionally, the mutation and growth rates of the MG collection, which included the wild-type genome MG1655 and the nine derivative mutators, were acquired from the previous study (Ishizawa et al., 2015). The data details of the three collections are summarized in **Supplementary Table 2**. The RNA sequencing dataset of the KHK collection grown in the M63 medium was acquired from the DNA Data Bank of Japan (DDBJ) under the accession number DRA13430. Global normalization of the read counts (raw data) was performed as described previously (Ying and

Yama, 2018). The relative expression levels of the genes participating in the DNA replication fidelity was analyzed (**Supplementary Table 5**).

## Multiple linear regression

Multiple linear regression (MLR) was performed with Python, as described previously (Aida et al., 2022). The media of LB, MAA, and M63 were represented as 1, 0, and  $-1$ , respectively. The *E. coli* strains of the reduced genome or MMR deficiency (mutator) were commonly indicated as one unit, and those of wild-type genome or non-mutator were set to zero. Genome reduction and MMR deficiency were categorized in the genotype, which was calculated by multiplying or adding the values of 1 or 0. Whether their relationship was considered interactive or additive, the two values representing genome reduction and MMR deficiency were multiplied or added, respectively. The logarithmic values of mutation rates were used in the MLR analysis. A total of four global parameters, i.e., growth rate, mutation rate, genotype, and medium, were finally subjected to the analysis. Ordinary least squares (OLS) regressions of the mutation and growth rates were performed with the three parameters of medium, genotype, and mutation or growth rate, respectively. The package of “stats” in Python was used, and the parameter estimation method of “ols” was applied.

## Machine learning with support vector machine

Support vector machine (SVM) was performed with Python and using the “svm” module in the package of scikit-learn, as previously reported (Aida et al., 2022). Briefly, the data points were randomly divided into two sets for the training and testing in ML as commonly performed (Zhou, 2021). The “random” function in the “svm” module was used to divide the whole dataset into the training and test datasets in the 60 to 40%. Five repeated training and test were conducted to evaluate the reliability of the SVM models. The linear and the radial basis function (RBF) methods were tested. Five-fold nested cross-validation searching the hyperparameters of C and gamma from 0.001 to 100 in increments of 10-fold was performed. The other hyperparameters were all used as default. Finally, in the “linear” method, C was set to 10 and 100 for classifying/predicting the medium and the genotype, respectively. In the “RBF” method, C and gamma were set to 10 and 1 and 100 and 0.01 for categorizing/predicting the medium and the genotype, respectively. The score of fivefold nested cross-validation was calculated by macro-averaging. The accuracy of the models was estimated according to the confusion matrix.

## Data availability statement

The original contributions presented in this study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

## Author contributions

B-WY conceived the research. ZL, YM, and B-WY analyzed the data. SI, YM, and B-WY performed the experiments. B-WY and ZL wrote the manuscript. All authors have read and approved the final manuscript.

## Funding

This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) (grant number: 19H03215). The funder has no role in the research.

## Acknowledgments

We thank NBRP for providing the KHK collection and the lab alumnae for experimental support.

## References

- Agrawal, A. F. (2002). Genetic loads under fitness-dependent mutation rates. *J. Evol. Biol.* 15, 1004–1010.
- Aida, H., Hashizume, T., Ashino, K., and Ying, B. W. (2022). Machine learning-assisted discovery of growth decision elements by relating bacterial population dynamics to environmental diversity. *Elife* 11:e76846. doi: 10.7554/eLife.76846
- Ashino, K., Sugano, K., Amagasa, T., and Ying, B. W. (2019). Predicting the decision making chemicals used for bacterial growth. *Sci. Rep.* 9:7251.
- Baer, C. F., Miyamoto, M. M., and Denver, D. R. (2007). Mutation rate variation in multicellular eukaryotes: Causes and consequences. *Nat. Rev. Genet.* 8, 619–631.
- Baker, B. J., Lazar, C. S., Teske, A. P., and Dick, G. J. (2015). Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. *Microbiome* 3:14. doi: 10.1186/s40168-015-0077-6
- Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., et al. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461, 1243–1247. doi: 10.1038/nature08480
- Basan, M., Honda, T., Christodoulou, D., Hörl, M., Chang, Y. F., Leoncini, E., et al. (2020). A universal trade-off between growth and lag in fluctuating environments. *Nature* 584, 470–474. doi: 10.1038/s41586-020-2505-4
- Breuer, M., Earnest, T. M., Merryman, C., Wise, K. S., Sun, L., Lynott, M. R., et al. (2019). Essential metabolism for a minimal cell. *Elife* 8:e36842.
- Cao, Y. Y., Yomo, T., and Ying, B. W. (2020). Clustering of bacterial growth dynamics in response to growth media by dynamic time warping. *Microorganisms* 8:331. doi: 10.3390/microorganisms8030331
- Casanueva, M. O., Burga, A., and Lehner, B. (2012). Fitness trade-offs and environmentally induced mutation buffering in isogenic *C. elegans*. *Science* 335, 82–85. doi: 10.1126/science.1213491
- Couce, A., Caudwell, L. V., Feinauer, C., Hindré, T., Feugeas, J. P., Weigt, M., et al. (2017). Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 114, E9026–E9035. doi: 10.1073/pnas.1705887114
- Cuperlovic-Culf, M. (2018). machine learning methods for analysis of metabolic data and metabolic pathway modeling. *Metabolites* 8:4.
- Duffy, S. (2018). Why are RNA virus mutation rates so damn high? *PLoS Biol.* 16:e3000003. doi: 10.1371/journal.pbio.3000003
- Eames, M., and Kortemme, T. (2012). Cost-benefit tradeoffs in engineered lac operons. *Science* 336, 911–915. doi: 10.1126/science.1219083
- Echols, H., Lu, C., and Burgers, P. M. (1983). Mutator strains of *Escherichia coli*, mutD and dnaQ, with defective exonucleolytic editing by DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. U.S.A.* 80, 2189–2192. doi: 10.1073/pnas.80.8.2189
- Eisner, S. A., Velicer, G. J., and Yu, Y.-T. N. (2022). Mutation of rpoB shifts the nutrient threshold triggering myxococcus multicellular development. *Front. Microbiol.* 13:817080. doi: 10.3389/fmicb.2022.817080
- Elena, S. F., Wilke, C. O., Ofria, C., and Lenski, R. E. (2007). Effects of population size and mutation rate on the evolution of mutational robustness. *Evolution* 61, 666–674.
- Erickson, D. W., Schink, S. J., Patsalo, V., Williamson, J. R., Gerland, U., and Hwa, T. (2017). A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*. *Nature* 551, 119–123. doi: 10.1038/nature24299
- Farahpour, F., Saeedghalati, M., Brauer, V. S., and Hoffmann, D. (2018). Trade-off shapes diversity in eco-evolutionary dynamics. *Elife* 7:e36273.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.990969/full#supplementary-material>

- Ferenci, T. (2016). Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol.* 24, 209–223.
- Frenoy, A., and Bonhoeffer, S. (2018). Death and population dynamics affect mutation rate estimates and evolvability under stress in bacteria. *PLoS Biol.* 16:e2005056. doi: 10.1371/journal.pbio.2005056
- Funchain, P., Yeung, A., Stewart, J. L., Lin, R., Slupska, M. M., and Miller, J. H. (2000). The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* 154, 959–970. doi: 10.1093/genetics/154.3.959
- Gao, Z., Wyman, M. J., Sella, G., and Przeworski, M. (2016). Interpreting the dependence of mutation rates on age and time. *PLoS Biol.* 14:e1002355. doi: 10.1371/journal.pbio.1002355
- Gilpin, W., Huang, Y., and Forger, D. B. (2020). Learning dynamics from large biological data sets: Machine learning meets systems biology. *Curr. Opin. Syst. Biol.* 22, 1–7.
- Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., et al. (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Mol. Microbiol.* 55, 137–149. doi: 10.1111/j.1365-2958.2004.04386.x
- Ishizawa, Y., Ying, B. W., Tsuru, S., and Yomo, T. (2015). Nutrient-dependent growth defects and mutability of mutators in *Escherichia coli*. *Genes Cells* 20, 68–76. doi: 10.1111/gtc.12199
- Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science* 349, 255–260.
- Kang, M., Kim, K., Choe, D., Cho, S., Kim, S. C., Palsson, B., et al. (2019). Inactivation of a mismatch-repair system diversifies genotypic landscape of *Escherichia coli* during adaptive laboratory evolution. *Front. Microbiol.* 10:1845. doi: 10.3389/fmicb.2019.01845
- Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., et al. (2016). Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. *Mol. Biol. Evol.* 33, 1257–1269. doi: 10.1093/molbev/msw009
- Kato, J., and Hashimoto, M. (2007). Construction of consecutive deletions of the *Escherichia coli* chromosome. *Mol. Syst. Biol.* 3:132. doi: 10.1038/msb4100174
- Kim, G. B., Kim, W. J., Kim, H. U., and Lee, S. Y. (2020). Machine learning applications in systems metabolic engineering. *Curr. Opin. Biotechnol.* 64, 1–9.
- Kinsler, G., Geiler-Samerotte, K., and Petrov, D. A. (2020). Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *Elife* 9:e61271. doi: 10.7554/eLife.61271
- Kirill, A. D., and Barry, L. W. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297
- Kishimoto, T., Iijima, L., Tatsumi, M., Ono, N., Oyake, A., Hashimoto, T., et al. (2010). Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. *PLoS Genet.* 6:e1001164. doi: 10.1371/journal.pgen.1001164
- Kishimoto, T., Ying, B. W., Tsuru, S., Iijima, L., Suzuki, S., Hashimoto, T., et al. (2015). Molecular clock of neutral mutations in a fitness-increasing evolutionary process. *PLoS Genet.* 11:e1005392. doi: 10.1371/journal.pgen.1005392
- Korem Kohanim, Y., Levi, D., Jona, G., Towbin, B. D., Bren, A., and Alon, U. A. (2018). Bacterial growth law out of steady state. *Cell Rep.* 23, 2891–2900. doi: 10.1016/j.celrep.2018.05.007
- Krašovec, R., Richards, H., Gifford, D. R., Hatcher, C., Faulkner, K. J., Belavkin, R. V., et al. (2017). Spontaneous mutation rate is a plastic trait associated with population density across domains of life. *PLoS Biol.* 15:e2002731. doi: 10.1371/journal.pbio.2002731
- Kurokawa, M., Nishimura, I., and Ying, B. W. (2022). Experimental evolution expands the breadth of adaptation to an environmental gradient correlated with genome reduction. *Front. Microbiol.* 13:826894. doi: 10.3389/fmicb.2022.826894
- Kurokawa, M., Seno, S., Matsuda, H., and Ying, B. W. (2016). Correlation between genome reduction and bacterial growth. *DNA Res.* 23, 517–525.
- Kurokawa, M., and Ying, B. W. (2017). Precise, high-throughput analysis of bacterial growth. *J. Vis. Exp.* 127:56197. doi: 10.3791/56197
- Kurokawa, M., and Ying, B. W. (2019). Experimental challenges for reduced genomes: The cell model *Escherichia coli*. *Microorganisms* 8:3. doi: 10.3390/microorganisms8010003
- Lee, J. H., Sung, B. H., Kim, M. S., Blattner, F. R., Yoon, B. H., Kim, J. H., et al. (2009). Metabolic engineering of a reduced-genome strain of *Escherichia coli* for L-threonine production. *Microbial Cell Fact.* 8:2.
- Lenski, R. E. (2017). Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. *ISME J.* 11, 2181–2194.
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332.
- Liu, L., Kurokawa, M., Nagai, M., Seno, S., and Ying, B. W. (2020). Correlated chromosomal periodicities according to the growth rate and gene expression. *Sci. Rep.* 10:15531. doi: 10.1038/s41598-020-72389-6
- Lu, H., Aida, H., Kurokawa, M., Chen, F., Xia, Y., Xu, J., et al. (2022). Primordial mimicry induces morphological change in *Escherichia coli*. *Commun. Biol.* 5:24. doi: 10.1038/s42003-021-02954-w
- Maharjan, R. P., and Ferenci, T. (2017). A shifting mutational landscape in 6 nutritional states: Stress-induced mutagenesis as a series of distinct stress input-mutation output relationships. *PLoS Biol.* 15:e2001477. doi: 10.1371/journal.pbio.2001477
- Mizoguchi, H., Mori, H., and Fujio, T. (2007). *Escherichia coli* minimum genome factory. *Biotechnol. Appl. Biochem.* 46, 157–167. doi: 10.1042/BA20060107
- Mizoguchi, H., Sawano, Y., Kato, J., and Mori, H. (2008). Superpositioning of deletions promotes growth of *Escherichia coli* with a reduced genome. *DNA Res.* 15, 277–284. doi: 10.1093/dnares/dsn019
- Morimoto, T., Kadoya, R., Endo, K., Tohata, M., Sawada, K., Liu, S., et al. (2008). Enhanced recombinant protein productivity by genome reduction in *Bacillus subtilis*. *DNA Res.* 15, 73–81. doi: 10.1093/dnares/dsn002
- Nishimura, I., Kurokawa, M., Liu, L., and Ying, B. W. (2017). Coordinated changes in mutation and growth rates induced by genome reduction. *mBio* 8:e00676–17. doi: 10.1128/mBio.00676-17
- Pósfai, G., Plunkett, G. III, Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., et al. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science* 312, 1044–1046.
- Ramiro, R. S., Durão, P., Bank, C., and Gordo, I. (2020). Low mutational load and high mutation rate variation in gut commensal bacteria. *PLoS Biol.* 18:e3000617. doi: 10.1371/journal.pbio.3000617
- Rees-Garbutt, J., Chalkley, O., Landon, S., Purcell, O., Marucci, L., and Grierson, C. (2020). Designing minimal genomes using whole-cell models. *Nat. Commun.* 11:836.
- Reuß, D. R., Altenbuchner, J., Mäder, U., Rath, H., Ischebeck, T., Sappa, P. K., et al. (2017). Large-scale reduction of the *Bacillus subtilis* genome: Consequences for the transcriptional network, resource allocation, and metabolism. *Genome Res.* 27, 289–299. doi: 10.1101/gr.215293.116
- Rewinski, C., and Marinus, M. G. (1987). Mutation spectrum in *Escherichia coli* DNA mismatch repair deficient (mutH) strain. *Nucleic Acids Res.* 15, 8205–8215. doi: 10.1093/nar/15.20.8205
- Schrider, D. R., and Kern, A. D. (2018). Supervised machine learning for population genetics: A new paradigm. *Trends Genet.* 34, 301–312. doi: 10.1016/j.tig.2017.12.005
- Sharma, S. S., Blattner, F. R., and Harcum, S. W. (2007). Recombinant protein production in an *Escherichia coli* reduced genome strain. *Metab. Eng.* 9, 133–141. doi: 10.1016/j.jymben.2006.10.002
- Shaw, F. H., and Baer, C. F. (2011). Fitness-dependent mutation rates in finite populations. *J. Evol. Biol.* 24, 1677–1684. doi: 10.1111/j.1420-9101.2011.02320.x
- Shibai, A., Takahashi, Y., Ishizawa, Y., Motooka, D., Nakamura, S., Ying, B. W., et al. (2017). Mutation accumulation under UV radiation in *Escherichia coli*. *Sci. Rep.* 7:14531. doi: 10.1038/s41598-017-15008-1
- Sprouffske, K., Aguilar-Rodríguez, J., Sniegowski, P., and Wagner, A. (2018). High mutation rates limit evolutionary adaptation in *Escherichia coli*. *PLoS Genet.* 14:e1007324. doi: 10.1371/journal.pgen.1007324
- Wang, X., Zorraquino, V., Kim, M., Tsoukalas, A., and Tagkopoulos, I. (2018). Predicting the evolution of *Escherichia coli* by a data-driven approach. *Nat. Commun.* 9:3562.
- Wang, Y., Diaz Arenas, C., Stoebl, D. M., Flynn, K., Knapp, E., Dillon, M. M., et al. (2016). Benefit of transferred mutations is better predicted by the fitness of recipients than by their ecological or genetic relatedness. *Proc. Natl. Acad. Sci. U.S.A.* 113, 5047–5052.
- Weisse, A. Y., Oyarzun, D. A., Danos, V., and Swain, P. S. (2015). Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U.S.A.* 112, E1038–E1047. doi: 10.1073/pnas.1416533112
- Yang, W. (2000). Structure and function of mismatch repair proteins. *Mutat. Res. DNA Repair* 460, 245–256.
- Ying, B. W., Honda, T., Tsuru, S., Seno, S., Matsuda, H., Kazuta, Y., et al. (2015). Evolutionary consequence of a trade-off between growth and maintenance along with ribosomal damages. *PLoS One* 10:e0135639. doi: 10.1371/journal.pone.0135639

- Ying, B. W., Ito, Y., Shimizu, Y., and Yomo, T. (2010). Refined method for the genomic integration of complex synthetic circuits. *J. Biosci. Bioeng.* 110, 529–536. doi: 10.1016/j.jbiosc.2010.05.014
- Ying, B. W., Seno, S., Matsuda, H., and Yomo, T. (2017). 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. doi: 10.1016/j.bbrc.2017.03.148
- Ying, B. W., Tsuru, S., Seno, S., Matsuda, H., and Yomo, T. (2014). Gene expression scaled by distance to the genome replication site. *Mol. bioSystems* 10, 375–379. doi: 10.1039/c3mb70254e
- Ying, B. W., and Yama, K. (2018). Gene expression order attributed to genome reduction and the steady cellular state in *Escherichia coli*. *Front. Microbiol.* 9:2255. doi: 10.3389/fmicb.2018.02255
- Zampieri, M., Hörl, M., Hotz, F., Müller, N. F., and Sauer, U. (2019). Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in *Escherichia coli*. *Nat. Commun.* 10:3354. doi: 10.1038/s41467-019-11331-5
- Zheng, H., Bai, Y., Jiang, M., Tokuyasu, T. A., Huang, X., Zhong, F., et al. (2020). General quantitative relations linking cell growth and the cell cycle in *Escherichia coli*. *Nat. Microbiol.* 5, 995–1001. doi: 10.1038/s41564-020-0717-x
- Zhou, Z. H. (2021). *Machine learning*. Singapore: Springer