Mutational effects of ultraviolet light on the genomic DNA of *Escherichia coli*

Submitted to

Graduate School of Information Science and Technology

Osaka University

January 2018

Atsushi SHIBAI

List of Publications

Published Papers

Original paper

[1] Shibai, A., Takahashi, Y., Ishizawa, Y., Motooka, Nakamura, S., Ying, B.W., Tsuru, S., "Mutation accumulation under UV radiation in *Escherichia coli*", *Scientific Reports*, 7; 14531, 2017. (Chapter 2)

Proceedings paper

[2] Shibai, A., Tsuru, S., Ying, B.W., Motooka, D., Gotoh, K., Nakamura, S., Yomo, T., "Mutation accumulation in bacteria exposed to UV radiation", *Artificial Life 14*, 757-758, 2014. (Chapter 2)

[3] Shibai, A., Motooka, D., Nakamura, S., Tsuru, S., "Reductive evolution towards primitive life: What will we see?", *Proceedings of the Artificial Life Conference 2016*, 188-189, 2016. (Chapter 4)

Presentations at International Conferences

Peer reviewed

[1] Shibai, A., Motooka, D., Nakamura, S., Tsuru, S., "Reductive evolution towards primitive life: What will we see?", ALIFE 15, Cancún, Mexico, Jul. 2016 (Oral presentation) (Chapter 4)

[2] Shibai, A., Tsuru, S., Ying, B.W., Motooka, D., Gotoh, K., Nakamura, S., Yomo, T., "Mutation accumulation in bacteria exposed to UV radiation", ALIFE 14, submission 86, New York, USA, Jul. 2014 (Poster presentation) (Chapter 3)

[3] Shibai, A., Seno, S., Motooka, D., Nakamura, S., Sakatani, Y., Otokura, M., Saito, H., Furubayashi, T., Furusawa, C., Ichihashi, N., Tsuru, S., "Prediction and experimental verification of experimental evolution of *Escherichia coli"*, ISME2016, 364A, Montreal, Canada, Aug. 2016 (Poster presentation) (Chapter 4)

Non-peer reviewed

[4] Shibai, A., Tsuru, S., Yomo, T., "Bacterial genome reduction through experimental evolution", QBiC Symposium 2015, P46, Osaka, Japan, Aug. 2015 (Poster presentation) (Chapter 2)

[5] Shibai, A., Tsuru, S., "Adaptive evolution of *E. coli* under high-mutagenicity environment", Humanware International Symposium 2017, Osaka, Japan, Jan. 2017 (Poster presentation) (Chapter 4)

Abstract

Author: Atsushi Shibai

Title: Mutational effects of ultraviolet light on the genomic DNA of *Escherichia coli*

Ultraviolet (UV) irradiation induces mutations in the DNA of living cells. UV radiation is included in sunlight and is a ubiquitous mutagen in nature. UV radiation is thought to have influenced biological evolution throughout history. Additionally, it is widely used as an artificial mutagen to accelerate experiments on the evolutionary breeding of useful microorganisms in the laboratory. It is expected that understanding the mutational effects of UV radiation will contribute both to understanding and prediction of genetic evolution.

Many studies on the mutational effects of UV radiation have been conducted using bacterial species, including *Escherichia coli*. However, most of them inferred changes in genomic DNA from phenotypic changes after UV irradiation and from DNA sequence information of corresponding reporter genes. These studies merely examined local site mutations on genomic DNA indirectly. In addition, they observed the occurrence of mutations resulting from single radiation treatments. Adaptive processes, such as fixation of mutations to cell populations and acquisition of UV tolerance, have not been studied.

I expect that applying next-generation sequencing technology, which has become widespread in recent years, will enable direct observation of DNA changes occurring at the whole-genome level. Therefore, in this study, I attempted to clarify the nature of UV mutagenesis using *E. coli* as a model organism by culturing it under UV irradiation and analyzing its genome with a next-generation sequencer.

This thesis consists of five chapters as described below.

In Chapter 1, I described the background and purpose of the research. The thesis focused on the base-pair substitution rate per generation and the spectrum of base-pair substitution, indexes affecting the speed and direction of genetic evolution, respectively. Additionally, I discussed two processes of UV mutagenesis. One is “emergence” of UV-induced mutations, and the other is “accumulation” of UV-induced mutations. These are described in Chapters 2 and 4, respectively. Accumulation of mutations required relatively long culture periods; therefore, I developed a culture system to support long and stable experiments, as described in Chapter 3.

Chapter 2 details a 28-day culture experiment with UV irradiation that I conducted, focusing on the emergence process of UV mutagenesis described above. I then analyzed the *E. coli* genome after the culture experiment. The results showed that the rate of emergence of base-pair substitution was approximately 100 times higher in UV-irradiating conditions compared to that in a typical environment without UV radiation. Additionally, I showed the spectrum of UV-induced base-pair substitutions. The substitution of G:C to A:T base pairs was most frequent, occurring more than half the time, though 4 of the 6 types of base-pair substitutions occurred at different rates.

Chapter 3 describes the development of a culture system capable of supporting longer experiments with UV irradiation that I used to observe the accumulation of mutations in Chapter 4. In the experiments in Chapter 2, the UV radiation dose per unit time was manually adjusted. However, in order to prevent extinction of the *E. coli* by underestimation of the UV radiation dosage, I conducted experimental manipulation for several hours at a fixed time every day. I reasoned that a system capable of measuring cell growth and applying UV radiation automatically would be required to conduct longer culture experiments stably. By developing such a system, it was possible to apply the maximum radiation dosage while preventing extinction of the cell population from excessive UV irradiation.

Chapter 4 describes a UV-irradiated culture experiment of 700 days carried out using the system developed in Chapter 3, focusing on the process of accumulated mutations described above. The *E. coli* genome was analyzed using a next-generation sequencer at several points during the experiment. The results showed that the base-pair substitution rate was maintained 100 times higher than that in a typical environment without UV radiation for one year, after which time the UV radiation tolerance of *E. coli* improved. Additionally, I analyzed the genes on which mutations accumulated. The results showed that function-inactivating mutations accumulated on a gene predicted to be nonessential in a previous study.

Chapter 5 summarizes the results obtained from this work and also discusses its future prospects and applications. In this study, I demonstrated that the base-pair substitution rates of *E. coli* were raised about 100 times by UV radiation exposure using culture experiments. Additionally, the spectrum of base-pair substitutions induced by UV radiation is described in my results. Further, I showed that mutations with inactivating effects on gene function accumulated intensively on purportedly nonessential genes. These findings will contribute to the planning of laboratory evolution experiments that require large numbers of mutations.

Contents

Chapter 1: Introduction 1
1.1 Background 1
1.2 Previous studies on UV mutagenesis and their limitations 2
1.3 Purpose 4
1.4 Layout of this thesis 5
Chapter 1 5
Chapter 2 5
Chapter 3 5
Chapter 4 5
Chapter 5 5
Chapter 2: Mutational profiles of UV exposure on *E. coli* genome 7
2.1 Introduction 7
2.2 Methods 8
2.2.1 Bacterial strains 8
2.2.2 Culture conditions 8
2.2.3 Measurement of maximal growth rate 8
2.2.4 Mutation accumulation experiments 9
2.2.5 Whole-genome sequencing 11
2.2.6 Measurement of base-pair substitution rate during the evolution experiment 11
2.2.7 Calculating dN/dS values 12
2.3 Results 13
2.3.1 Mutation accumulation experiment 13
2.3.2 Changes in maximal growth rate 16
2.3.3 Rate and spectrum of accumulated base-pair substitutions 18
2.3.4 dN/dS values 21
2.4 Discussion 22
2.4.1 Did the genetic background influence UV mutagenesis? 22
2.4.2 Comparing UV-induced mutations and internally emerging mutations 22
2.4.3 Determinants of mutation accumulation rate 23
2.5 Summary 25
Chapter 3: Development of an automated UV-irradiated culture system 26
3.1 Introduction 26
3.2 Methods 28

3.2.1 Strain and media 28
3.2.2 Automated device....28
3.2.3 Operation procedure....30
3.2.4 Culturing various UV doses per exposure....32
3.2.5 Mutation accumulation experiment....32
3.2.6 Whole-genome sequencing....32
3.2.7 Measurement of base-pair substitution rate during the evolution experiments....32
3.2.8 Calculating dN/dS values....33
3.3 Results....34
3.3.1 Calibration curve for optical density measurement....34
3.3.2 UV killing curve....34
3.3.3 Culturing various UV doses per exposure....35
3.3.4 Mutation accumulation experiment....36
3.3.5 Detection of accumulated mutations....38
3.4 Discussion....40
3.4.1 Does feedback control work?....40
3.4.2 Changes in the tolerable UV dose via mutation accumulation experiments....40
3.4.3 Mutational profile of our automated UV-irradiating culture device....41
3.5 Summary....42
Chapter 4: Long-term evolution experiment with UV exposure....43
4.1 Introduction....43
4.2 Methods....44
4.2.1 Strain and media....44
4.2.2 Evolution experiment....44
4.2.3 Measurement of maximal growth rate....44
4.2.4 Whole-genome sequencing....44
4.2.5 Measurement of base-pair substitution rate during the evolution experiment....45
4.2.6 Calculating dN/dS values....45
4.2.7 Comparing mutational pattern and essentiality of each gene in *E. coli*....45
4.2.8 Gene ontology enrichment analysis for mutated genes....46
4.3 Results....47
4.3.1 Evolution experiment....47
4.3.2 Change in maximal growth rates....48
4.3.3 Rate and spectrum of accumulated base-pair substitutions....48
4.3.4 dN/dS values in the evolution experiment....49

4.3.5 Essentiality of *E. coli* genes 50
4.4 Discussion 52
4.4.1 How did *E. coli* adapt to its UV-irradiated environment? 53
4.4.2 *E. coli* genes are redundant in the laboratory environment 56
4.5 Summary 57
Chapter 5: Conclusion 58
5.1 Summary of this thesis 58
5.1.1 Mutational profiles of UV exposure on the *E. coli* genome 58
5.1.2 Development of an automated UV-irradiated culture system 58
5.1.3 Long-term evolution experiment with UV exposure 59
5.2 Discussion 60
5.2.1 Application for high-mutagenicity evolution experiments 60
5.2.2 Genome reduction driven by random mutagenesis 60
References 61
Acknowledgements 66

Chapter 1: Introduction

1.1 Background

All living organisms have highly sophisticated, complex gene networks. Humans have approximately 22,000 genes, and even prokaryotic *Escherichia coli* have about 4,000 genes[1]. Genes interact with one another in complex ways to realize organismal activities[2-4]. In addition, each gene is encoded by a highly sophisticated nucleotide sequence of roughly 1000 bases. Each gene is thought to be very vulnerable to mutation, because most mutations occurring on the genomes of microorganisms have harmful effects[5,6]. Therefore, it is thought that living organisms maintain the function of these complex networks of fragile genes by suppressing mutations through a gene repair mechanism.

Mutations are induced not only by internal factors, such as errors during DNA replication, but also by external factors, such as ultraviolet (UV) irradiation. Therefore, knowing the nature of mutations induced by each factor is important to understanding the evolutionary processes of living organisms in nature and in the laboratory. However, while the influence of internal factors has been investigated extensively[7-10], the mutations induced by external factors have not been examined on the genome scale. In addition, most previous studies have focused on the short-term effects of a single round of mutagenesis by external factors, for example, by observing the fixation of mutations in a population and acquisition of UV radiation tolerance[11,12]. To date, the long-term effects of continuous mutagenesis by external factors have not been investigated.

E. coli is a model organism widely used in the field of molecular biology. Examining the influence of external mutagens on *E. coli* will aid understanding of the long-term effects of mutagenesis by external factors. Various methods can be used to induce mutations in *E. coli*, such as administration of a mutagen or exposure to radiation. UV irradiation is considered to be safe and quantitative. Many studies have investigated the effects of UV irradiation on *E. coli* and other bacteria[11-15]. However, as mentioned above, few studies have conducted long-term subcultures, and those that did simply observed adaptations to UV irradiation and did not focus on the occurrence of mutations. Therefore, clarifying mutations by UV irradiation and their long-term effects will contribute to the understanding of the evolution of organisms in nature and to the development of more efficient evolutionary engineering.

1.2 Previous studies on UV mutagenesis and their limitations

UV light is known to kill *E. coli* by causing harmful mutations, and Vermeulen et al.[14] showed that the effect was highest at a wavelength of 265 nm. Additionally, Mundhada et al.[38] recently showed that UV mutagenesis can be used to generate an L-serine-producing *E. coli* strain for industrial applications. The accurate identification of the mutational profiles of mutagens is required to control mutation speed more precisely and predictably. As noted above, there have been many studies related to the mechanisms of UV mutagenesis and *E. coli*. Using the *lacI* gene on the *E. coli* genome as a reporter gene, Miller[39] showed that the most common base-pair substitution occurring on that sequence is the G:C to A:T substitution. Additionally, Ikehata and Ono[37] argued that substitution of C to T occurs specifically with UV irradiation at a position of two or more adjacent pyrimidine bases on the sequence. Lin and Wang[40] also showed that the *lacZ*, *rpoB*, *ompF*, and *ompA* genes can be used as UV mutagenesis reporter genes.

However, most studies that have assessed the mutational profiles of mutagens were based on fluctuation tests with a few reporter loci on the genome. This approach is conventional but inaccurate for assessing the representative profiles of the whole genome. Moreover, the mutational profiles, mutation rate, and mutational spectrum of the restricted reporter loci and the whole genome can be very different. For example, Tsuru et al. reported that the trends in base-pair substitutions seen using the *lacI* reporter gene was not seen on the whole-genome level[9,41]. Additionally, when inducing mutations with UV light, the amounts of mutants detected by the *rpoB* and *ompF* reporter genes were largely different[40]. These findings indicate the importance of capturing the mutation profiles of mutagens over the whole genome.

In this study, by using next-generation sequencing technology, which has become popular in recent years, it was possible to observe DNA changes occurring on the whole-genome scale[9]. Therefore, I attempted to clarify the nature of UV mutagenesis using *E. coli* as a model organism cultured under UV irradiation by analyzing its genome with a next-generation sequencer.

In addition, previous researches proposed some mechanisms of UV mutagenesis[14,37,39]. The process of UV mutagenesis occurs following two steps; (1) occurrence of damages in DNA by UV light, (2) nucleotide substitutions during replication caused by damaged DNA. Here I explained each step according to a recent review (Ikehata and Ono, 2011[37]).

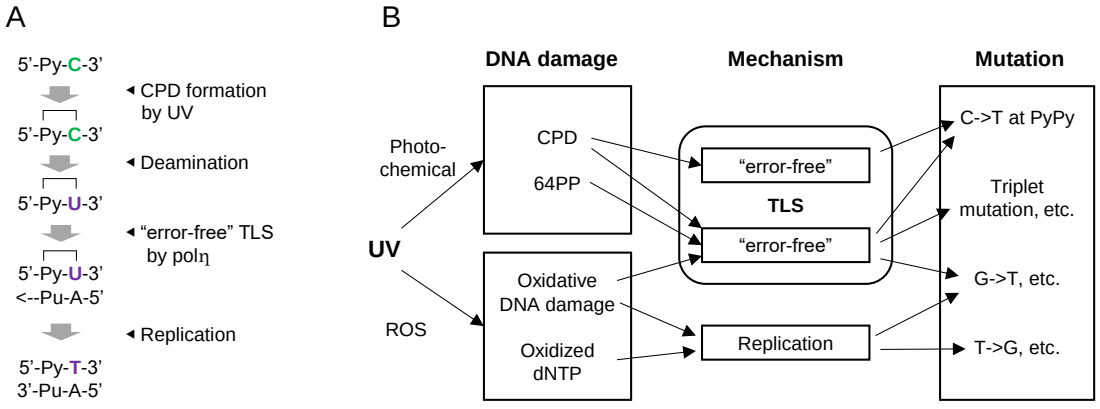
(1) Occurrence of damage in DNA by UV light

UV light directly affects chemical bonds in DNA depending on the sequence. Specifically, the bases are altered to form a dimer in consecutive pyrimidine bases (C or T), distorting the helix of that position. There are two kinds of the dimers, cyclobutane-pyrimidine dimers (CPDs, Fig. 1.1A) and pyrimidine (6-4) pyrimidine photoproducts (64 PPs) depending on their modes of bonding. Also, UV light indirectly causes damages in nucleic acids (DNA and dNTPs) via generating reactive oxygen.

(2) Nucleotide substitutions during replication caused by damaged DNA

There have been proposed two mutagenesis models for damaged nucleotides during DNA replication. In the first "error-free" bypass model, during translesion DNA synthesis (TLS, Fig. 1.1B), the DNA polymerase η binds to the damaged parts of DNA and causes C to T substitution at the position of CPD, and also causes base-pair substitutions such as G to T, T to G, etc., at the nucleotides damaged by reactive oxygen. In the second two-step "error-prone" bypass model, the combination of TLS with replicative DNA polymerases caused base-pair substitutions at the position of CPD and 64PP.

Among these mutagenesis models, those related to CPD and 64PP is considered as DNA-sequence dependent. Therefore, knowledge on the occurrence of genome-scale mutation by UV light which I tried to clarify in this thesis has the potential to contribute to the understanding of UV mutagenesis mechanism.



(arranged from Ikehata and Ono, 2011[37])

Fig. 1.1 Representative mechanisms of UV mutagenesis

(A) An example of C->T base-pair substitution process by "error-free" model. (B) The previously proposed models of UV mutagenesis. These figures were edited and arranged from the previously reported review (Ikehata and Ono, 2011[37]).

1.3 Purpose

In this study, I aimed to clarify what types of mutations were generated on the genomic DNA of *E. coli* by UV irradiation. In addition, I investigated how the mutation profiles of *E. coli* shifted in highly mutagenic environments over a relatively long term. To accomplish these aims, subculture experiments of *E. coli* populations with UV irradiation were performed. A conceptual diagram of the subculture experiments is shown in Fig. 1.2. I analyzed the phenotypes and genotypes of the ancestral strain and the evolution strains to detect mutations accumulating on the *E. coli* genome. First, experiments were conducted over a short time period to obtain the mutation profile of UV light. Subsequently, similar experiments were performed over a longer period, and changes to the mutation profile were examined. The aims of these experiments were to clarify short-term trends in UV mutagenesis on the *E. coli* genome and to examine how these short-term trends change as mutations accumulate on the genome.

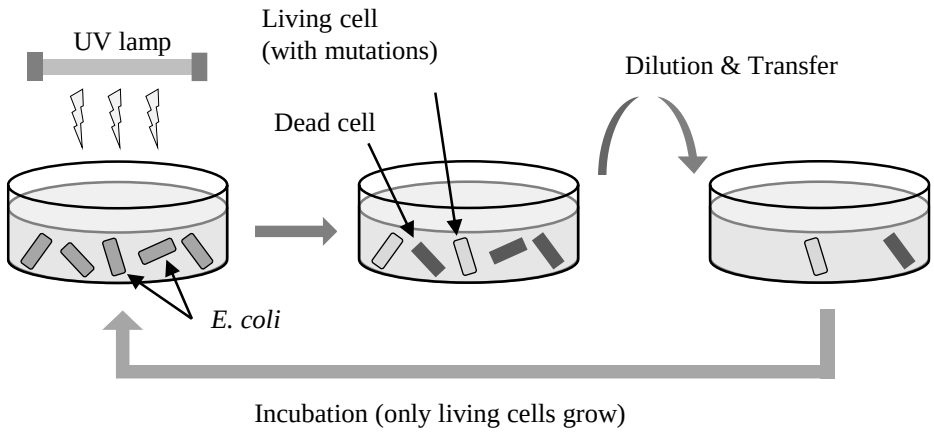


Fig. 1.2 Conceptual scheme of the UV-irradiating serial culture experiment
An *E. coli* population is irradiated with UV light (left). Many mutations are generated on the *E. coli* genomes. *E. coli* cells that obtained lethal mutations die, while those that obtained nonlethal mutations survive (middle). When diluting the cell culture with fresh medium (right), the surviving cells grow (left again). This grown population is expected to inherit the mutations obtained by the survivors. By repeating this cycle, the number of mutations on the *E. coli* genomes is increased.

1.4 Layout of this thesis

Chapter 1

Chapter 1 explains the background and purpose of this research.

Chapter 2

To clarify the profiles of mutations generated on the *E. coli* genome by UV irradiation, short-term mutation accumulation experiments were conducted by repeatedly exposing *E. coli* populations to UV light. Genomic DNA sequences were analyzed before and after the experiments using a next-generation sequencer. The results showed that the mutation accumulation rate with UV irradiation was approximately 100 times higher than that in non-irradiation conditions and several times higher than that of typical mutator strains, which have internal mutagenesis factors. Additionally, I showed that the spectrum of base-pair substitutions induced by UV irradiation varies more than that of the mutator strains.

Chapter 3

In order to enable long-term evolution experiments with recursive UV irradiation, I developed a culture system that automatically performs cell growth and adjustable ultraviolet irradiation. This system makes it possible to apply the maximum amount of UV radiation while preventing extinction of the cell population due to excessive irradiation.

Chapter 4

Long-term evolution experiments were carried out using the developed system. As the result of two years of evolution experiments, I found that the mutation rate on the *E. coli* genome was maintained at an approximately 100 times higher level than that under non-irradiation conditions. This result showed that the highly mutagenic environment produced by UV irradiation can be maintained for long periods.

Chapter 5

Chapter 5 summarizes the results in previous chapters and discusses the achievements and progress attained by this research.

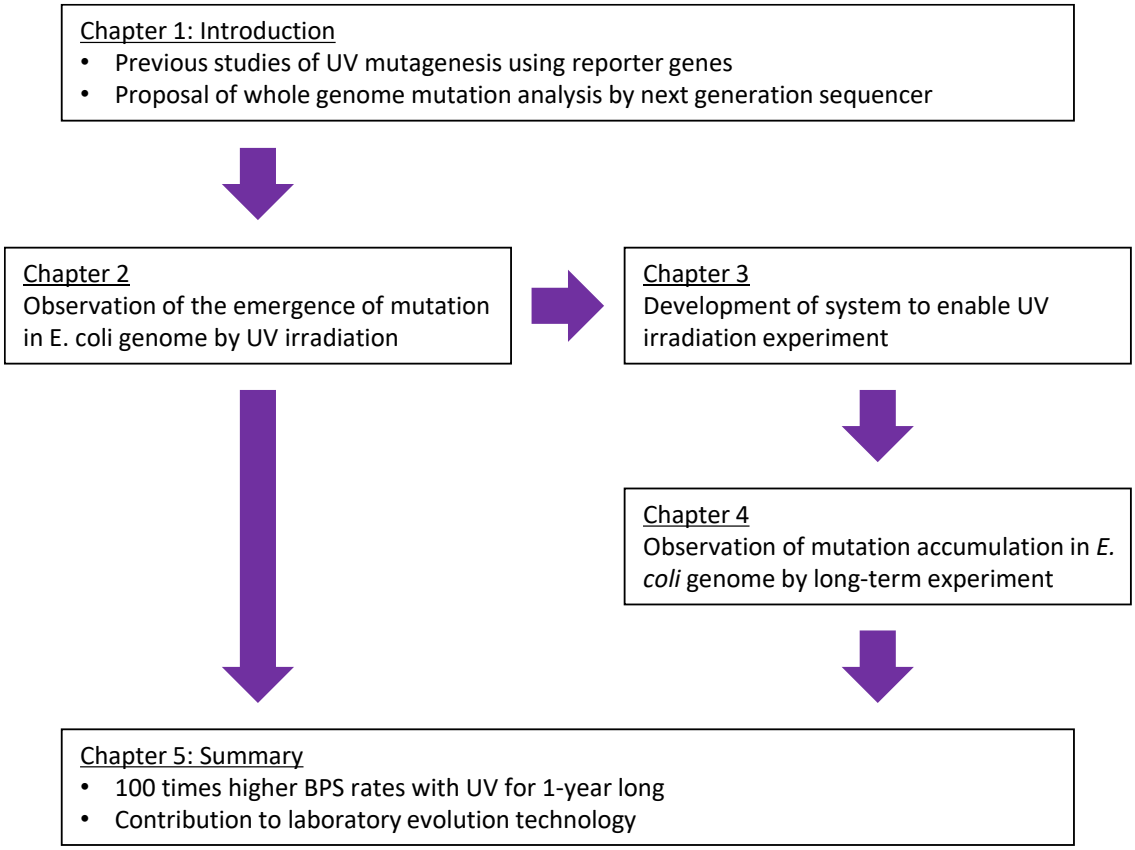


Fig. 1.3 Layout of this thesis

Chapter 2: Mutational profiles of UV exposure on *E. coli* genome

(1) The rate of emergence of base-pair substitution was approximately 100 times higher in UV irradiating conditions than in a typical environment without UV light.

(2) The substitution of the G:C to A:T base pair was the most frequent on the genomic level, consistently with previous research that only used *lacI*[39], and represented more than half of the observed substitutions, although 4 of all 6 types of base-pair substitutions occurred at a different rates.

(3) These result enabled estimation of the genome-wide emergence of mutations induced by UV irradiation.

2.1 Introduction

In the study detailed in this chapter, I aimed to investigate the profile of mutations on the *E. coli* genome caused by UV irradiation. Among the possible traits of the profile, I focused on (1) the mutational spectrum of UV light; (2) mutation rate, which regulates the speed of evolution; and (3) dN/dS values, which indicate the type of UV-induced mutations fixed in the population. *E. coli* populations were irradiated with UV light, and their genome sequences were analyzed by next-generation sequencing technology to detect mutations. I analyzed both the types and the number of mutations. Since not all mutations occurring on the *E. coli* genomes were fixed in the population, I also described the fixing mechanism of mutations based on the fixation bias pattern.

2.2 Methods

2.2.1 Bacterial strains

I used *E. coli* MDS42 and two mutator strains, MDS42*ΔmutS::Cm* and MDS42*ΔmutH,ΔmutS,ΔuvrB::Cm*. These strains were used to examine how UV radiation affected inherent mutability, spontaneous mutation rate in the absence of UV exposure, and UV sensitivity. They were named WT, ΔS, and ΔHSB. ΔS and ΔHSB were constructed by deletions of mismatch repair genes (*mutS*, *mutH*) and a UV-resistant gene (*uvrB*). These deletions were examined by standardized λ-red homologous recombination using the pKD46 plasmid. The *mutS* and *mutH* genes are involved in the mismatch repair system, so deletion of either or both genes increases mutation rate. The *uvrB* gene is a nucleotide excision repair genes. Defects in *uvrB* reduce native UV resistance. Therefore, ΔHSB was designed to be a UV-sensitive mutator strain. The MDS42 strain is a strain in which 14.3% of the genome has been removed by the genetic engineering operations against the K12 strain; in particular, the global structure of the genome is thought to be stable, because most of its IS elements are lost[16]. Therefore, I expect this to be suitable for whole-genome sequencing analysis using next-generation sequencing.

2.2.2 Culture conditions

I used a chemically defined medium, mM63, which contained 62 mM K_2HPO_4, 39 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 2 μM $FeSO_4 \cdot 7H_2O$, 15 μM thiamine hydrochloride, 203 μM $MgSO_4 \cdot 7\ H_2O$, and 22 mM glucose[25].

2.2.3 Measurement of maximal growth rate

Frozen stock cells were inoculated into 100 μl of mM63 broth on a 96-well microplate and cultured at 37 °C with shaking at 550 rpm for 12 h. The cell culture was then diluted 100 times with fresh mM63 broth and poured into wells in a microplate so that the culture volume was 100 μl/well (10 wells for each evolved strain and 60 wells for each ancestral strain). The microplate was incubated at 37 °C with shaking in a plate reader (Infinite F200 PRO, Tecan, Mannedorf, Switzerland). The plate reader measured OD_{595} every 15 min. I calculated the maximum growth rate [h^{-1}] by fitting the slopes of the growth curves during the exponential growth phase (OD_{595} = 0.01~0.06) to the standard Malthusian growth model[51].

2.2.4 Mutation accumulation experiments

I conducted subculture experiments for 28 days, in which one round corresponded to one day. A conceptual diagram of the subculture cycle with UV irradiation is shown in Fig. 2.1. Each round operation consisted of OD measurement, dilution, and UV irradiation. The OD_{595} of the 96-well microplate that had been cultured for 1 day from the previous round was measured using a plate reader (Infinite F200 PRO). On the microplate, there were five sets of wells irradiated with different doses of UV light in the previous round. The well with the largest dose satisfying $OD_{595} > 0.1$ was selected, and μl of its culture was sampled. After a 100-fold dilution, 100 μl of the culture solution was dispensed into 5 wells on a new 96-well plate. Each of the 5 wells was irradiated with UV light of a different intensity. The microplate was covered with a lid and sealed. The plate was then incubated at 37 ºC with shaking at 550 rpm for 1 day. For UV irradiation, a sterilizing lamp in a clean bench was used. Adjustment of UV intensity was performed by covering 5 wells with different numbers of filters and simultaneously exposing them to UV light. The 5 wells were covered by 0, 2, 4, 6, and 8 filters; each filter level allowed 0.6 times as much UV light through as the previous level. The maximum UV dose in each round, which was the dose in the well without filters, was increased as the populations obtained UV tolerance. Irradiation intensity was recorded before irradiation using a dosimeter. I made a 15% glycerol stock every 7 rounds and stored it at –80 ºC. The UV irradiation intensity of the 0-filter well of the microplate was 1.0 to 2.1 [W/m^2]. For each of the three ancestral strains (WT, ΔS, ΔHSB), 6 lineage evolution experiments were performed simultaneously using one 96-well plate.

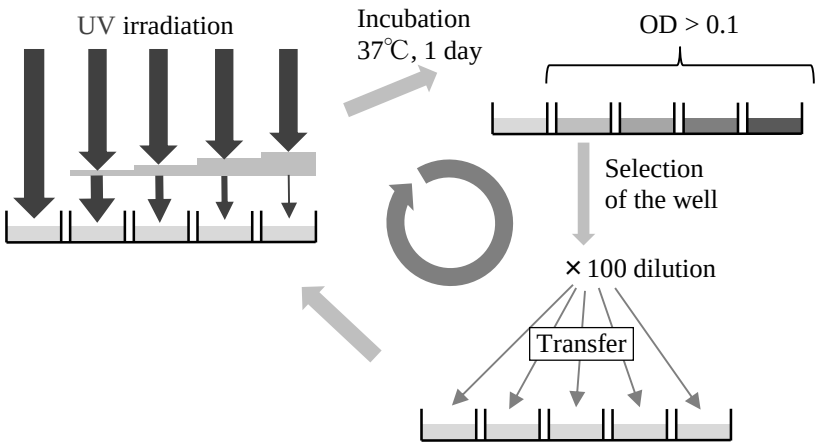


Fig. 2.1 Cycle of the culture system with UV-exposure

The operation of each round consisted of an OD measurement (upper right), dilution (lower right), and ultraviolet irradiation (left). First, in the OD measurement, the well with the largest amount of UV irradiation in the last round among ODs exceeding the reference value was selected. Then, the sampled cell culture was diluted 100-fold and poured into 5 wells of a new 96-well plate. These five wells were irradiated with different amounts of UV light and cultured for 1 day.

A conceptual diagram of subculture cycle without UV irradiation is shown in Fig. 2.2. Each round of operation consisted of an OD measurement and dilution. The OD_{595} of the 96-well microplate that had been cultured for 1 day from the previous round was measured using a plate reader (Infinite F200 PRO). On the plate, there were five sets of wells diluted with different dilution ratios in the previous round. The well with the largest dilution rate in the previous round satisfying $OD_{595} > 0.1$ was selected, and a 10 µl aliquot of the culture was sampled. I prepared 5-fold diluted cell culture at different rates and poured 100 µl each into 5 wells on a new 96-well microplate. The plate was covered with a lid and sealed, followed by culturing at 37 ° C with shaking at 550 rpm for 1 day. The dilution rates of the 5 wells differed by 10-fold intervals. In the example in Fig. 2.2, the rates were 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}. I made a 15% glycerol stock every 7 rounds and stored it at –80 ºC. The highest dilution ratios of wells were 10^{-6} to 10^{-9}. For each of the three ancestral strains (WT, ΔS, ΔHSB), 6 lineage evolution experiments were performed simultaneously using one 96-well plate.

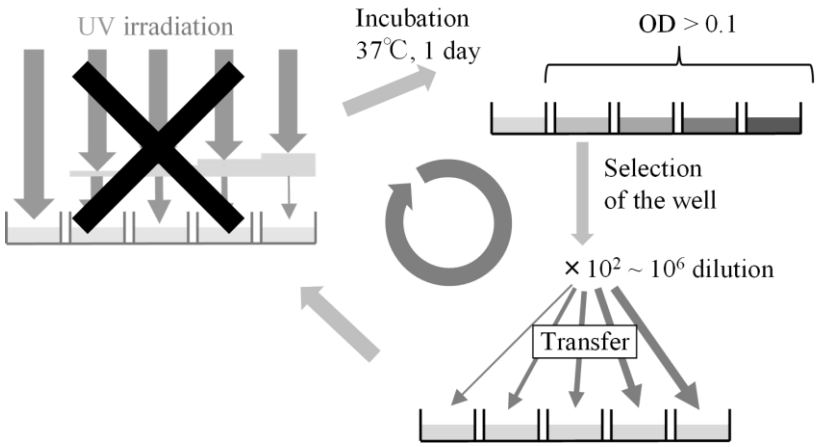


Fig. 2.2 Cycle of the culture system without UV-exposure

The operation of each round consisted of an OD measurement (upper right) and dilution (lower right). UV irradiation (left) was not performed. The well with the highest dilution ratio of the previous round was selected based on the OD exceeding the reference value. The cell culture sampled from the well was then diluted at 5 different dilution rates, poured into 5 wells of a new 96-well plate, and cultured for 1 day.

2.2.5 Whole-genome sequencing

Genomic DNA was extracted using a Wizard Genomic DNA Purification kit (Promega, Madison, WI, USA). Multiplex analysis (usually 6 plex) was performed as pair-end sequencing (251 bp) at 500 cycles using MiSeq Reagent Kit v 2 (Illumina, San Diego, CA, USA). I used Burrows-Wheeler Aligner (BWA, http://bio-bwa.sourceforge.net/) to map the obtained reads onto the reference genome sequence of *E. coli* MDS42 (accession number: AP012306, the origin of NC_020518, GI: 471332236). I used SAMtools[22] to call base-pair substitutions (BPSs) and small insertions or deletions (InDels) with the maximum read depth set to 500 (-D option) and the other parameters set to default. To ensure the quality of mutation calling, called mutations with Phread quality scores less than 100 were eliminated[23,24]. Additionally, the BPSs with reads less than 90% were eliminated. I then supposed that the remaining mutations were dominant in the sequenced populations.

2.2.6 Measurement of base-pair substitution rate during the evolution experiment

I calculated the rate of base-pair substitution by the following equation:

$$\rho = \frac{N_{syn} \cdot L}{F(syn) \cdot L_{CDS} \cdot D} \quad (\text{genome}^{-1}\ \text{day}^{-1}) \qquad (1)$$

where ρ, N_{syn}, L_{CDS}, L, and D are the rate of base pair substitution, the number of synonymous substitutions, the length of total coding DNA sequences per genome, genome size (3.98 Mbp for the MDS42), and the number of days of the evolution experiments, respectively. $F_{(syn)}$ is the probability of a synonymous substitution and was calculated based on the codon usage and the probability of each substitution.

2.2.7 Calculating dN/dS values

The dN/dS value was calculated by dividing the number of nonsynonymous substitutions per nonsynonymous site, dN, with the number of synonymous substitutions per synonymous site, dS. Hence, the following equation gives the value:

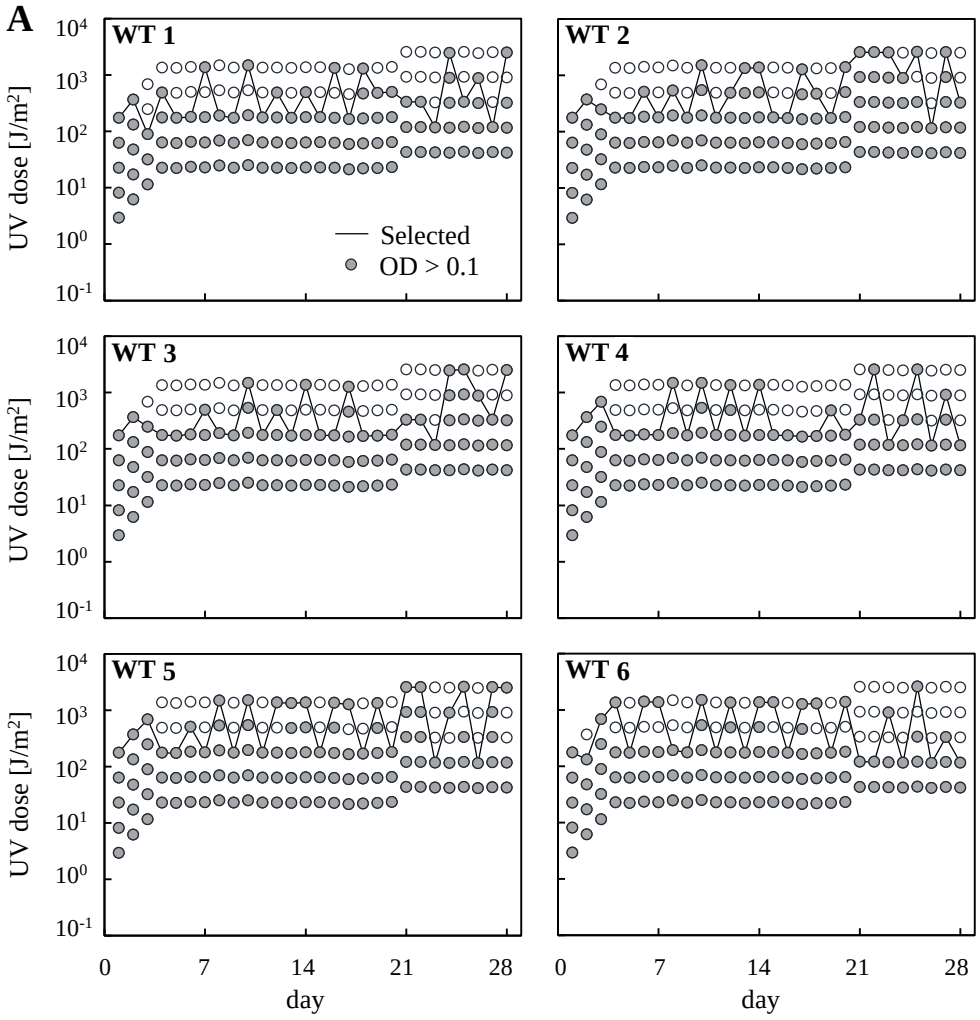
$$dN/dS = \frac{N_{nsyn}}{N_{syn}} \cdot \frac{F(syn)}{1 - F(syn)} \qquad (2)$$

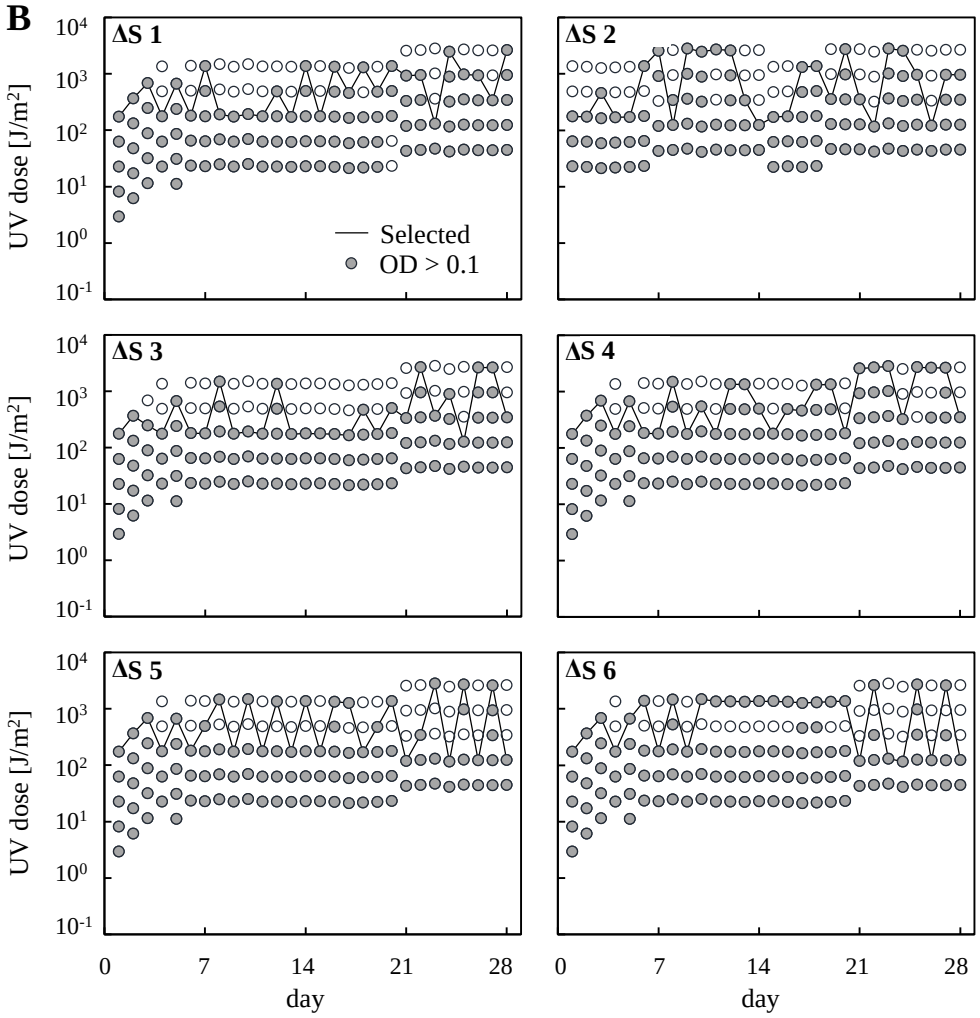
where N_{syn} and N_{nsyn} represent the number of synonymous and nonsynonymous substitutions, respectively. $F_{(syn)}$ is the probability that a substitution is synonymous (as described above).

2.3 Results

2.3.1 Mutation accumulation experiment

I performed short-term culture experiments with and without UV radiation exposure in 6 lineages for each of the 3 *E.coli* strains. In each lineage with UV radiation exposure, I cultured the cell in 5 wells with different UV light doses. I measured the level of the UV light dose every day and plotted it (Fig. 2.3). The symbols in Fig. 2.3 represent the optical density of the cells on the following day (closed circle if $OD_{595} > 0.1$ and open circle if $OD_{595} \leq 0.1$). I selected the well with the maximum UV light dosage in which cells sufficiently grew and transferred to a culture sample to 5 new wells with fresh medium. The data of the selected wells are connected with lines.




B
ΔS 1
ΔS 2
ΔS 3
ΔS 4
ΔS 5
ΔS 6
Selected
OD > 0.1
UV dose [J/m²]
day
0
7
14
21
28

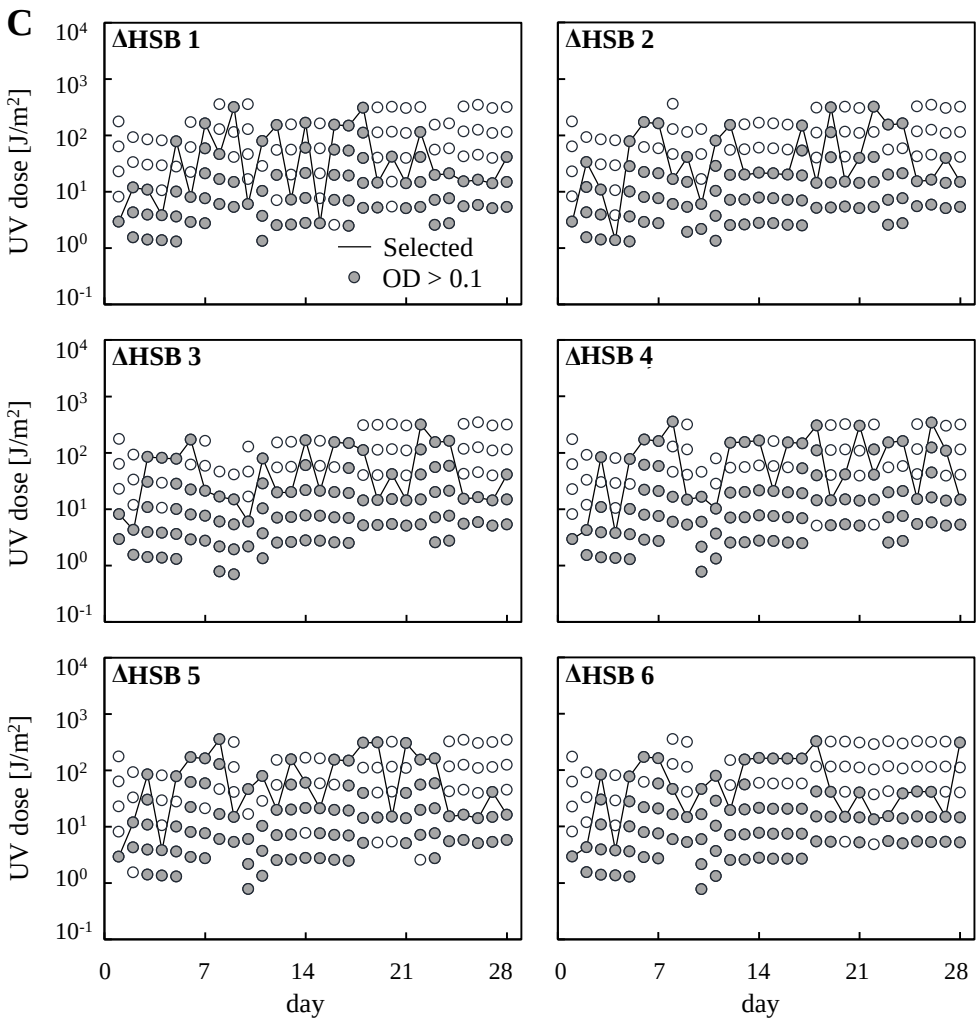


Fig. 2.3 Changes in UV radiation dosage during experiment with UV light exposure
The daily UV light doses and population growth of WT (A), ΔS (B), and ΔHSB (C) strains for 28 days are shown. The horizontal axis represents the passage number in days and the vertical axis represents the radiation dose applied to the well. Closed circles indicate that the OD of the well was 0.1 or more, and open circles indicate that it was less than 0.1. The solid line connecting circles shows the propagated well. (A) Lineages with WT as ancestral strain. (B): Lineages with ΔS as ancestral strain. (C): Lineages with ΔHSB as ancestral strain.

Changes in the daily growth rates during the short-term culture experiments without UV light are shown in Fig. 2.4. The daily growth rates were calculated by dividing the logarithm of the inversed dilution rates by the incubation time (24 h). Only the growth rates of the selected wells are plotted in

the figure.

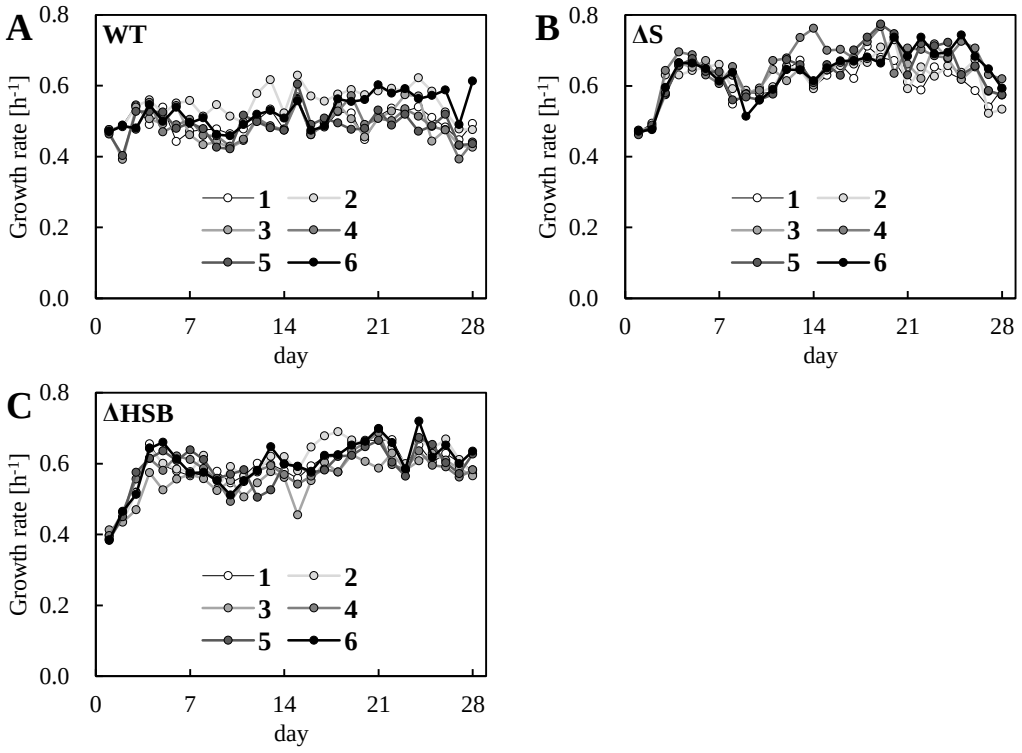


Fig. 2.4 Changes in growth rates during experiments without UV exposure

The growth rates of the WT (A), ΔS (B), and ΔHSB (C) strains were measured daily during the short-term culture experiments.

2.3.2 Changes in maximal growth rate

During the mutation accumulation experiments, it was expected that the cells were under selective pressure to grow quickly. Therefore, high-mutagenicity conditions could have accelerated evolution of the growth rate. In order to determine this, I measured the maximal growth rates of glycerol-stocked cells of ancestral strains (anc) and evolved lineages at the 28th day (evo) following the method described in 2.2.3. The maximum growth rates of each sample are shown in Fig. 2.5. For WT, the growth rates seen in the ancestral strain were improved in the evolved strains under UV light conditions. On the other hand, in evolved WT strains grown without UV light, there was no clear trend in the change in growth rate. For ΔS, the evolved strains showed larger growth rates than

did the ancestral strain, regardless of the presence or absence of UV irradiation. For ΔHSB, the ancestral strain and evolved strains showed similar growth rates, regardless of the presence or absence of UV irradiation. These results suggested that evolutions of the WT strain with UV irradiation and the ΔmutS strains with and without UV irradiation were accelerated.

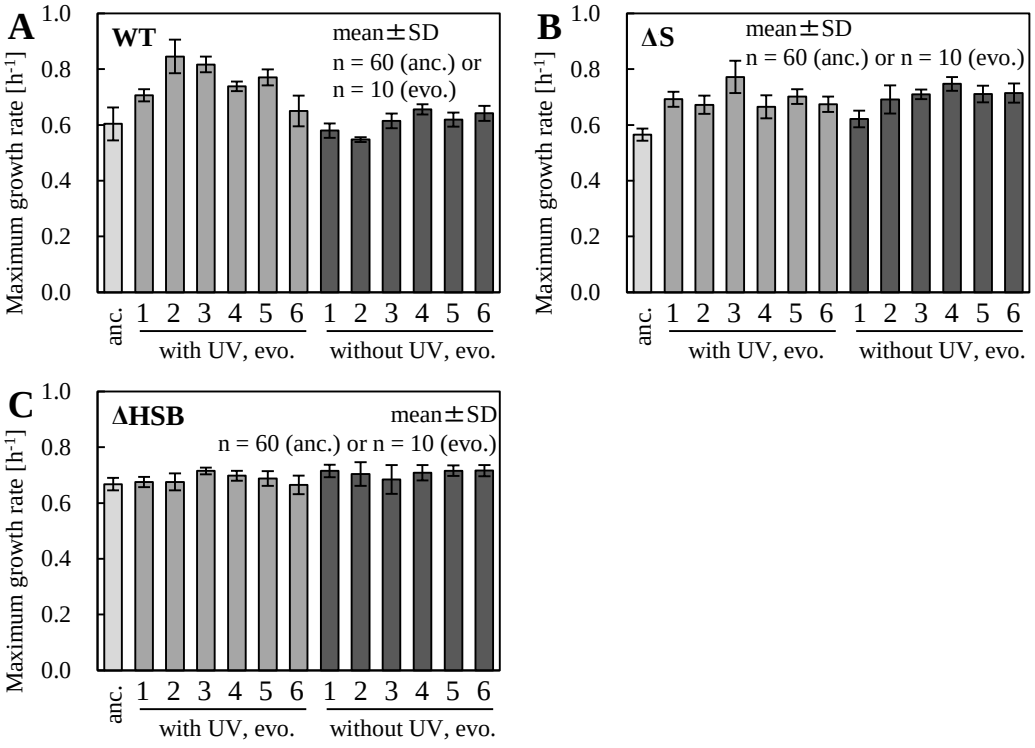


Fig. 2.5 Maximal growth rates changed before and after mutation accumulation experiments

The maximum growth rates of the ancestral and evolved strains are shown. The heights of the bars indicate average values, and error bars indicate standard deviations. Bar color indicates ancestral strains, evolved strains with UV, and evolved strains without UV. (A): WT, (B): ΔS, (C): ΔHSB.

2.3.3 Rate and spectrum of accumulated base-pair substitutions

The spectrum of BPSs by UV radiation is thought to determine the possible evolutionary pathways of the amino acid sequence. The spectrum is expected to vary depending on the mutagen[30,31]. For example, lack of the *mutS* gene, an intrinsic mutagen, causes only 2 types of BPS among the possible 6 BPS patterns[9] (the horizontal axis in Fig. 2.6). I considered BPS spectrum to be one of the most important components of UV radiation-induced mutational profiles. In order to examine the spectrum, the frequency of BPSs on the genomic sequence of the evolved strain for each of the conditions, with and without UV irradiation, were measured (Fig. 2.6, top). The spectra of the spontaneous substitutions were previously explored for a *mutS*-defective mutator strain and a wild type (Fig. 2.6, middle and bottom)[9,7]. Contrary to the wide distribution for the wild type, the mutator strain exhibited two peaks at A:T to G:C and G:C to A:T and very small frequencies for the other substitutions (Fig. 2.6, bottom). This transition-biased spectrum of the mutator strain was similar to that of other mutator strains with deficient mismatch-repair or proofreading processes. Compared with the typical spectrum of mutators, the spectrum of the UV-induced synonymous substitutions was more various. The fraction of G:C to A:T was still high, consistently with previous research using only *lacI*[39], while the fractions of A:T to T:A, A:T to G:C, and G:C to T:A were at levels comparable to those in the wild type. That is, the UV radiation-induced substitutions included not only transitions but also transversions similar to the spontaneous substitutions in the wild type.

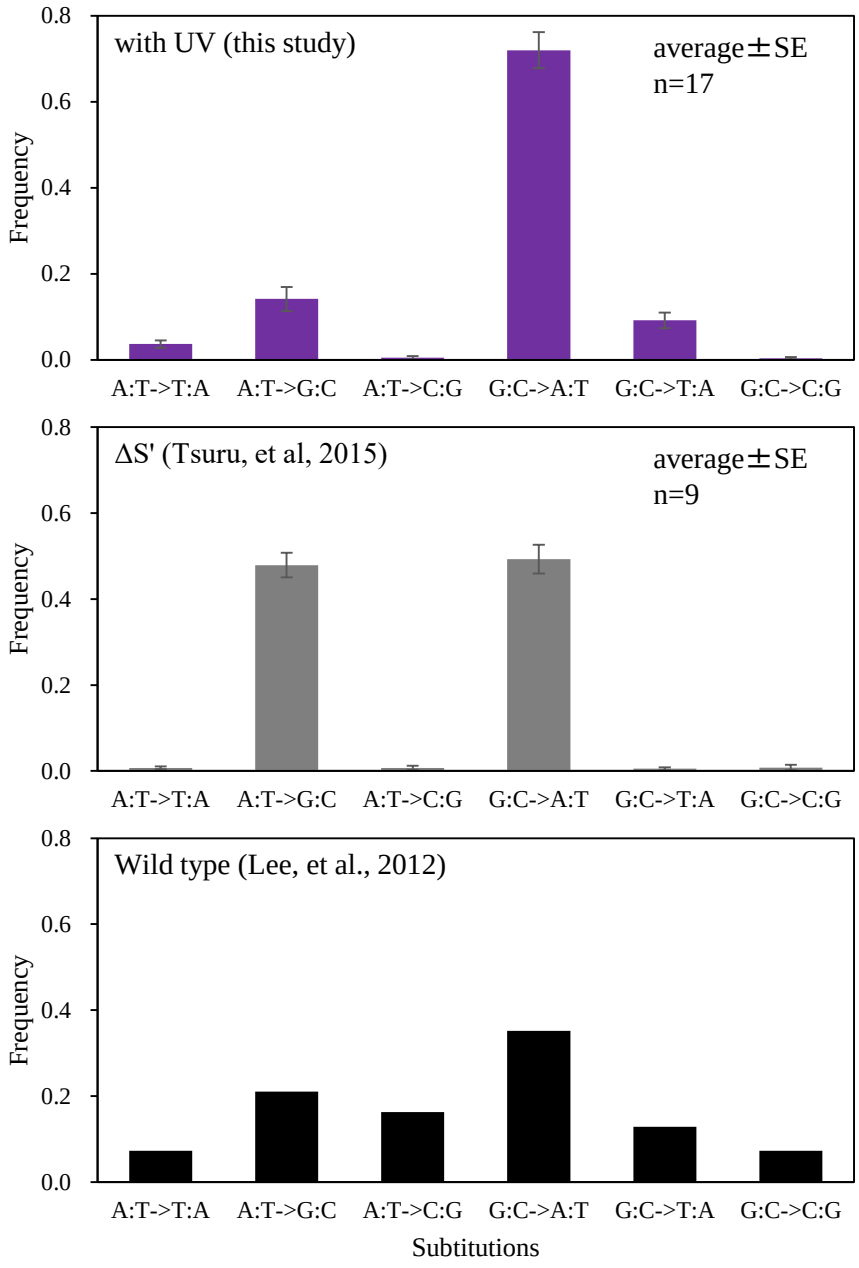


Fig. 2.6 Base-pair substitution spectrum

Bar color correspond to the ancestral strains. The error bars indicate the standard deviations. The y-axis indicates the frequency of each type of synonymous substitution. The spectrum for UV-induced mutations was calculated by combining all strains (top). The spectra of spontaneous mutations for the wild-type (bottom) and mutS-defective strains (middle), denoted as ΔS', were obtained from previous studies[9,7].

The BPS rates were calculated according to eq. (1) using the number of synonymous mutations and the probability of a substitution on the genome by UV light to be synonymous (*F(syn)*). *F(syn)* was calculated using codon usage of *E. coli* and the BPS spectrum obtained in the experiment (Fig. 2.6). First, under UV light conditions, *F(syn)* values were almost equal (0.31 to 0.32) for WT, ΔS, and ΔHSB. On the other hand, in evolved strains non-UV light conditions, *F(syn)* values were 0.26 in WT, 0.35 in ΔS, and 0.34 in ΔHSB. BPS rates were then calculated according to eq. (1) and are shown in Fig. 2.7. Under UV light conditions, BPS rates were similar for WT, ΔS, and ΔHSB ($5.5 \times 10^{-8} \sim 7.1 \times 10^{-8}$ [mutations $\cdot$ bp^{-1} $\cdot$ generation^{-1}]. On the other hand, BPS rates for non-UV light conditions greatly varied depending on the ancestral strains. For WT, the rate was indefinite, because I did not detect any synonymous substitutions. The rates were 3.9×10^{-9} and 1.3×10^{-8} [mutations $\cdot$ bp^{-1} $\cdot$ generation^{-1}] for ΔS and ΔHSB, respectively. BPS rates were obviously lower under non-UV light than under UV light conditions. The BPS rate under UV irradiation conditions was approximately 5 times greater than that of ΔHSB with the highest mutation rate under non-UV irradiation conditions.

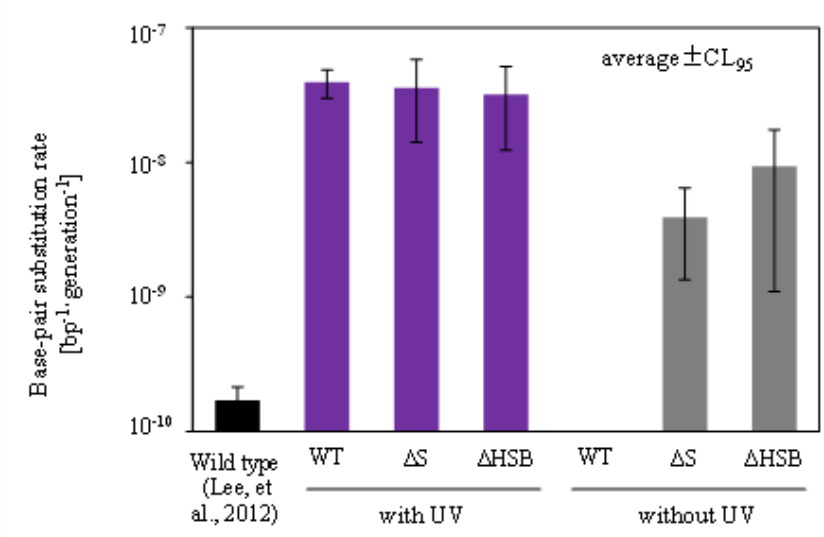


Fig. 2.7 Base-pair substitution rates during mutation accumulation experiment

The error bar indicates the 95% confidence interval. Bar color corresponds to the presence or absence of UV irradiation in subculture and literature value. Although information on synonymous substitution number is necessary for calculating the substitution rate, synonymous substitution was not found for WT under non-UV light conditions, so it was indefinite. The literature value for the wild type (black bar) was obtained from a previous study[7].

2.3.4 dN/dS values

The dN/dS value is an index showing the degree of selective pressure for amino acid substitution mutation. If the dN/dS value is more than 1, it implies that fixation of amino acid substitutions are promoted, while if the value is less than one, it implies that amino acid substitutions are excluded. This index is important in discussing the selective advantages of mutations produced by UV radiation. The dN/dS ratio of each sample was calculated from the patterns of BPSs, including synonymous and non-synonymous mutations. The results are shown in Fig. 2.8. Under UV conditions, the dN/dS ratio was near 1.0, indicating that mutations were accumulated under near-neutral selective pressure.

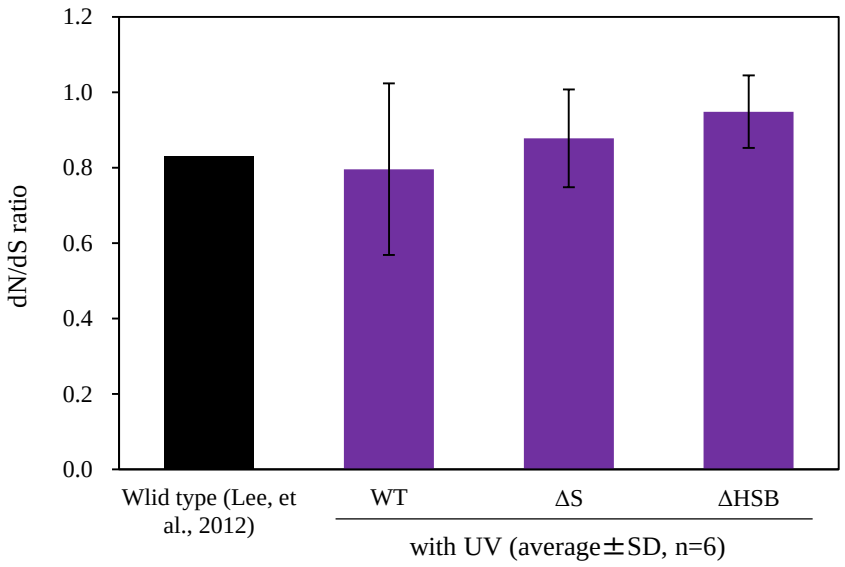


Fig. 2.8 dN/dS values during mutation accumulation experiments

The error bars indicate standard deviations. The literature value for the wild type (black bar) was obtained from a previous study[7].

2.4 Discussion

In the experiments detailed in this chapter, I investigated the profile of mutations caused by UV irradiation on the *E. coli* genome and identified the mutational spectrum, mutation rate, and dN/dS values with and without UV radiation exposure. The results supported a considerable contribution of UV radiation as an extrinsic source of diverse mutations. These results form a basis for the studies in the following chapters and also provide interesting findings as discussed below.

2.4.1 Did the genetic background influence UV mutagenesis?

As noted above, I used the three stains, WT, ΔS, and ΔHSB, to examine how UV radiation affected the inherent mutability, spontaneous mutation rate in the absence of UV exposure, and UV sensitivity of *E. coli*. I found almost the same BPS rates in the presence of UV exposure for all strains (Fig. 2.7, ANOVA, $F(2,14)=0.15$, $p=0.86$). These accumulation rates were higher than those in the absence of UV exposure (26-fold and 3.4-fold increase for ΔS and ΔHSB, respectively). The similar rates of mutation accumulation among the different genetic backgrounds implied a limit to mutation production per generation. The majority of the accumulated mutations in the lineages exposed to UV radiation were caused by this exposure. It is interesting that the rates of mutation accumulation in these lineages were quite similar even though the applied doses varied among the lineages, implying that the UV sensitivity of ΔHSB allowed more harmful mutations with smaller UV doses compared to those applied to the WT. These factors limited the allowable UV dose during the culture experiments.

2.4.2 Comparing UV-induced mutations and internally emerging mutations

The BPS rates with UV irradiation were 5.8×10^{-8}, 5.5×10^{-8}, and 7.1×10^{-8} [$\text{mutations} \cdot \text{bp}^{-1} \cdot \text{generation}^{-1}$] for WT, ΔS, and ΔHSB, respectively. On the other hand, for WT under non-UV conditions, the BPS rate was indefinite, because I did not find any mutations in the experiment, possibly because the BPS rate was very low. Additionally, for ΔS and ΔHSB under non-UV conditions, the rates were 3.9×10^{-9} and 1.3×10^{-8} [$\text{mutations} \cdot \text{bp}^{-1} \cdot \text{generation}^{-1}$], respectively, which were obviously lower values than those observed under UV conditions. This result suggested that the mutations induced by UV radiation had a dominant effect during the evolution experiments compared to those emerging spontaneously.

2.4.3 Determinants of mutation accumulation rate

In order to think about how UV irradiation causes mutations while killing cells, I constructed a deductive model. I hypothesized that UV irradiation simultaneously causes two types of damage in the cell population. The first includes both lethal and non-lethal mutations. Please note that, for simplicity, beneficial mutations were not included in this model. I supposed this to be appropriate, because I mainly considered the process of cell death, and because beneficial mutations are known to be rare[5]. According to conventional assumptions, the number of mutations per cell, *m*, follows the Poisson distribution of the rate parameter λ. For *m* mutations, lethality follows the binomial distribution of the lethal mutation rate (*l*). The second type of damage includes deleterious physiological damage, such as unrecoverable double-strand breaks in genomic DNA. I regarded such physiological damage as side effects of UV-induced mutations. The instances of physiological damage per cell are represented as *p*, which follows the Poisson distribution of the rate parameter γ.

The probability distribution of a cell with *p* instances of lethal physiological damage and *q* lethal mutations among *m* mutations can be calculated as follows:

$$P(m,q,p)=\frac{\lambda^m e^{-\lambda}}{m!}\binom{m}{q}l^q(1-l)^{m-q}\cdot\frac{\gamma^p e^{-\gamma}}{p!}. \quad (3)$$

Here, I assume that non-lethal mutations are neutral, i.e., they do not affect the fitness of the cells. Then, the survival rate, S, of the population after UV irradiation was calculated as

$$S=\sum_{m=0}^{\infty}P(m,q=0,p=0)=e^{-(\lambda l+\gamma)}. \quad (4)$$

Here, the ratio of λl to the power exponent, i.e., $\lambda l/(\lambda l+\gamma)$, indicates the mutagenesis efficiency of UV radiation against the death rate in the population. λ can be estimated by considering a relationship between λ and the average number of mutations per surviving cell (ρ), which can be measured by whole-genome sequencing of the evolved samples. The average number of non-lethal mutations with non-lethal physiological damage per surviving subpopulation divided by the number of surviving cells is given as follows:

$$\rho = \frac{N_{tot}\sum_{m=0}^{\infty} m \cdot P(m, q=0, p=0)}{N_{tot}S} = \lambda(1-l). \qquad (5)$$

N_{tot} represents the total number of cells in the population. Therefore, ρ is equivalent to the non-lethal fraction, $1-l$, of λ. In this experiment I assumed $S = 10^{-4}$ according to the growth rate (Fig. 2.5) and culturing time. l was set as 0.1 because the dN/dS values were roughly 0.8–0.9 (Fig. 2.8), which suggests that 10–20 % of mutations were excluded by selection. Additionally, ρ was 3.13 [bps/genome/day] (Fig. 2.7, WT). Introducing these parameters into eq. (4) and (5) gives λl = 0.35 [lethal mutations/genome/day] and γ = 8.86 [lethal side effects/genome/day]. Then, the efficiency of mutagenesis was obtained as $\lambda l/(\lambda l + \gamma)$ = 0.038. This implied that the dominant cause of cell death due to UV irradiation was not deleterious genetic mutations, but physiological side effects.

In order to discuss the general efficiency of UV radiation as a mutagen, I performed a similar estimate for spontaneous mutations as follows. Previous studies have reported that the increase in the spontaneous mutation rate caused by deleting genes related to error correction also reduced growth[9,10]. Therefore, I also calculated the efficiency of mutation production due to the lack of error-correcting genes from the growth defects in the hyper-mutable strains. Additionally, I assumed that the lack of these genes also caused two types of lethal damage: lethal mutations and some lethal side effects. For simplicity, the mutations and the deleterious side effects occurred every replication event in this case. That is, equation (3) is also applicable, and S in equation (4) indicates the rate of success for each self-replication event in this situation. The growth rate of the hyper-mutable strains (Δmut) is represented as follows:

$$\mu_{mut} = \mu_{WT} S = \mu_{WT} e^{-(\lambda l + \gamma)}, \qquad (6)$$

where μ_{WT} represents the growth rate of the wild-type strain with a sufficiently low mutation rate. Equation (6) was fitted to the corresponding data indicating that the growth rate decreased with the mutation rate (Fig. 2 in Ishizawa et al.[10]). Assuming that γ is proportional to λ and l is 0.1, I obtained γ / λ =4.85±2.72 (average ± standard error). Then, the efficiency of mutation production for the lack of error-correcting genes was calculated as $\lambda l/(\lambda l + \gamma)$=0.020 (standard error interval was 0.013–0.045). Interestingly, this value is comparable to that of UV exposure, suggesting that the low efficiency of mutation production by UV exposure is not a specific property of UV light.

2.5 Summary

In this chapter, I examined the mutational profiles induced by UV light on the *E. coli* genome. Specifically, I focused on (1) the BPS spectrum of UV light, which affects the possible evolutionary pathways of amino acid sequences of the genome; (2) BPS rate per generation, which regulates the speed of evolution; and (3) dN/dS values, which indicate how UV-induced mutations are fixed in the population. I conducted short-term mutation accumulation experiments on the *E. coli* genome with UV irradiation. Based on the mutations detected in the evolved samples, I showed various spectra and the high upper limit of the rate of UV-induced BPSs. Additionally, I showed that the dN/dS value was near 1, implying that most UV-induced mutations were not excluded by selective pressure in the experiment. These results supported a considerable contribution of UV radiation as an extrinsic source of mutations to generate diverse mutants in both natural and artificial conditions.

According to previous studies, the growth rate of *E. coli* decreases by half with about 1000 base-pair substitutions on the genome in the absence of selective pressure for growth rate[28,29]. The results of the present study showed that the number of occurrences of base-pair substitution in the UV irradiation environment was approximately 3 per day. In order to accumulate 1000 substitutions and decrease the growth rate, it was necessary to continue the experiment for a year. Performing this experiment would require manipulations for several hours at a fixed time every day. In order to conduct these experiments, I thought it necessary to automate this process and designed the system described in the following chapter.

Chapter 3: Development of an automated UV-irradiated culture system

(1) The results of Chapter 2 implied that a year-long culture experiment was required to achieve the purpose of Chapter 4.

(2) I developed a culture system that automatically measures cell growth and applies a corresponding UV radiation dose to conduct such long-term experiments stably.

(3) The developed system was able to apply the maximum dose of UV radiation while preventing extinction of the cell population from excessive irradiation.

3.1 Introduction

In the experiment described in Chapter 2, the UV light dose per hour was manually adjusted, and several hours of experimental manipulations were performed each day at regular intervals to prevent both the extinction and saturation of the *E. coli* populations. However, to achieve the purposes of Chapter 4, it was necessary to continue the experiments for a year. Therefore, in this chapter, I describe the development of a culture system that automatically measures cell growth and applies a corresponding UV radiation dose to conduct such long-term experiments stably. Specifically, I developed a device that can avoid cell saturation and extinction with manual operation only once every several days.

In general, intense UV doses are required for UV mutagenesis, because the amount of mutations is thought to be positively correlated with the UV dose. Intense UV irradiation, however, is accompanied by massive killing of *E. coli* along with induction of mutations, resulting in extinction and the end of cultivation (Fig. 3.1). Therefore, mild and frequent UV irradiation is appropriate for longer-term cultivation. Presently, it is technically difficult to manually perform irradiation with a small amount of ultraviolet rays frequently (e.g., once every several hours). Therefore, I first developed a device that can measure cell culture concentration in real time and apply UV light to the culture. I also developed a system to operate and control the device and automated the experimental operation.

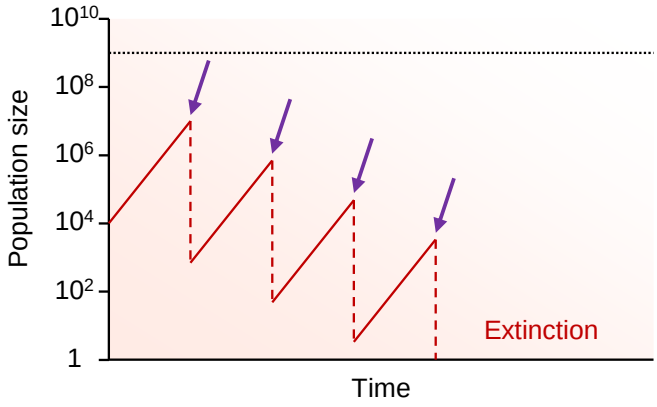


Fig. 3.1 Cyclic exposure of excess UV leads to population extinction

3.2 Methods

3.2.1 Strain and media

I used *E. coli* MDS42 and a chemically defined medium, mM63, which contained 62 mM K_2HPO_4, 39 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 2 μM $FeSO_4 \cdot 7H_2O$, 15 μM thiamine hydrochloride, 203 μM $MgSO_4 \cdot 7\ H_2O$, and 22 mM glucose. Cells were cultured at 37 ºC with shaking at 200 rpm. To inhibit growth on the walls of the tubes, the shaking speed was raised to 300 rpm for 1 min every 6 h.

3.2.2 Automated device

The automated UV-irradiating culture device was constructed from a fused quartz test tube and a socket with a turbidimeter and an UV light emitter (Fig. 3.2). The turbidimeter consisted of a visible LED and a photo transistor (typical wave lengths were 590 and 560 nm, respectively). They were positioned at a 135º angle to maximize the detection of scattered light[16,17]. For the UV light emitter, I used a Deep-UV LED module (UV-EC910ZA, Panasonic Photo & Lighting, Osaka, Japan). The UV LED was placed such that the tube was irradiated from its bottom. A 3D-printed ABS resin housing held the electronic elements in appropriate positions. The electronic components were connected in a circuit (Fig. 3.3) and controlled by an Arduino Uno (Arduino, Somerville, MA, USA) driven by a laptop computer.

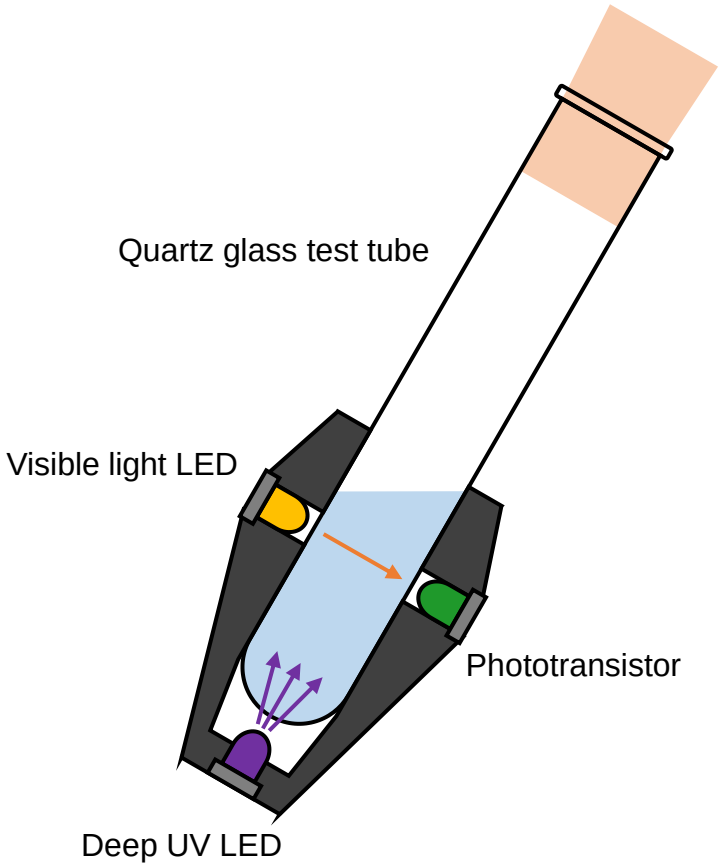


Fig. 3.2 Schematic of the structure of the device

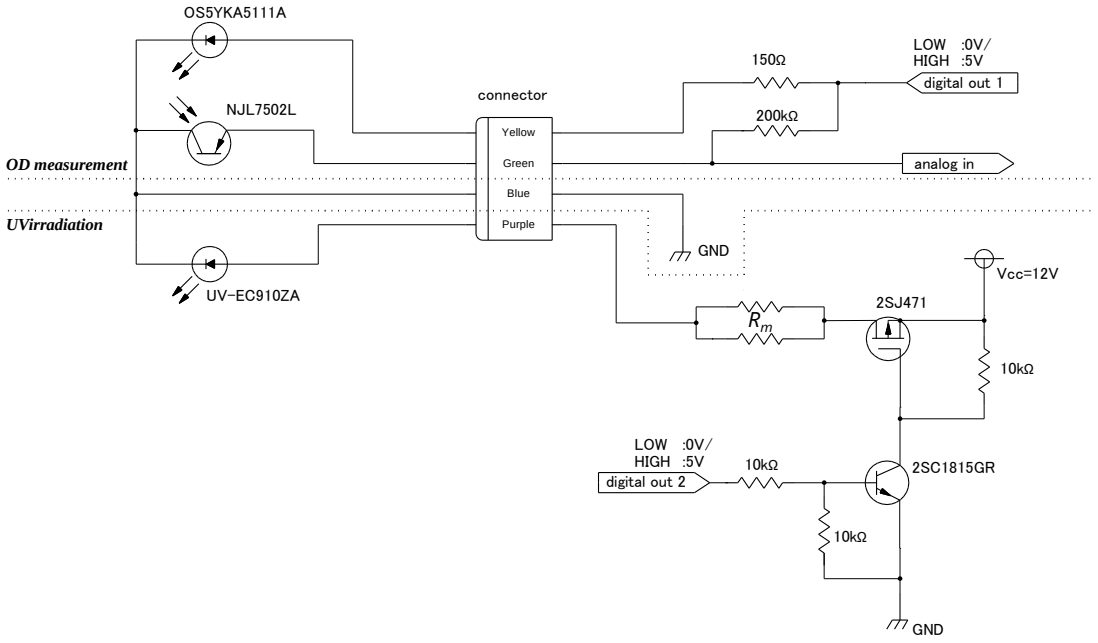


Fig. 3.3 Circuit diagram of the device

3.2.3 Operation procedure

All operations were controlled by programs that I developed in Arduino IDE and Visual Studio 2013 (Microsoft, Redmond, WA, USA). The process flow of the method is shown in Fig. 3.4A. Optical density (OD) measurement and UV irradiation are implemented in repeated cycles. In each cycle, cells were incubated for a time period (3 min), throughout which the OD is recorded. UV radiation was applied when OD values were greater than a set threshold (OD_{THR}) in two consecutive cycles. The OD_{THR}, which was set as $OD_{THR} = OD_{STEP}$ initially, was then set to $OD_{THR} + OD_{STEP}$, so that the next irradiation was conducted when the cell population had recovered to a density corresponding to OD_{STEP} (Fig. 3.4B). In this way, the cells continued to grow exponentially as OD increased linearly, enabling cells to avoid saturation for a long time. In this study, OD_{STEP} was set to 0.0015.

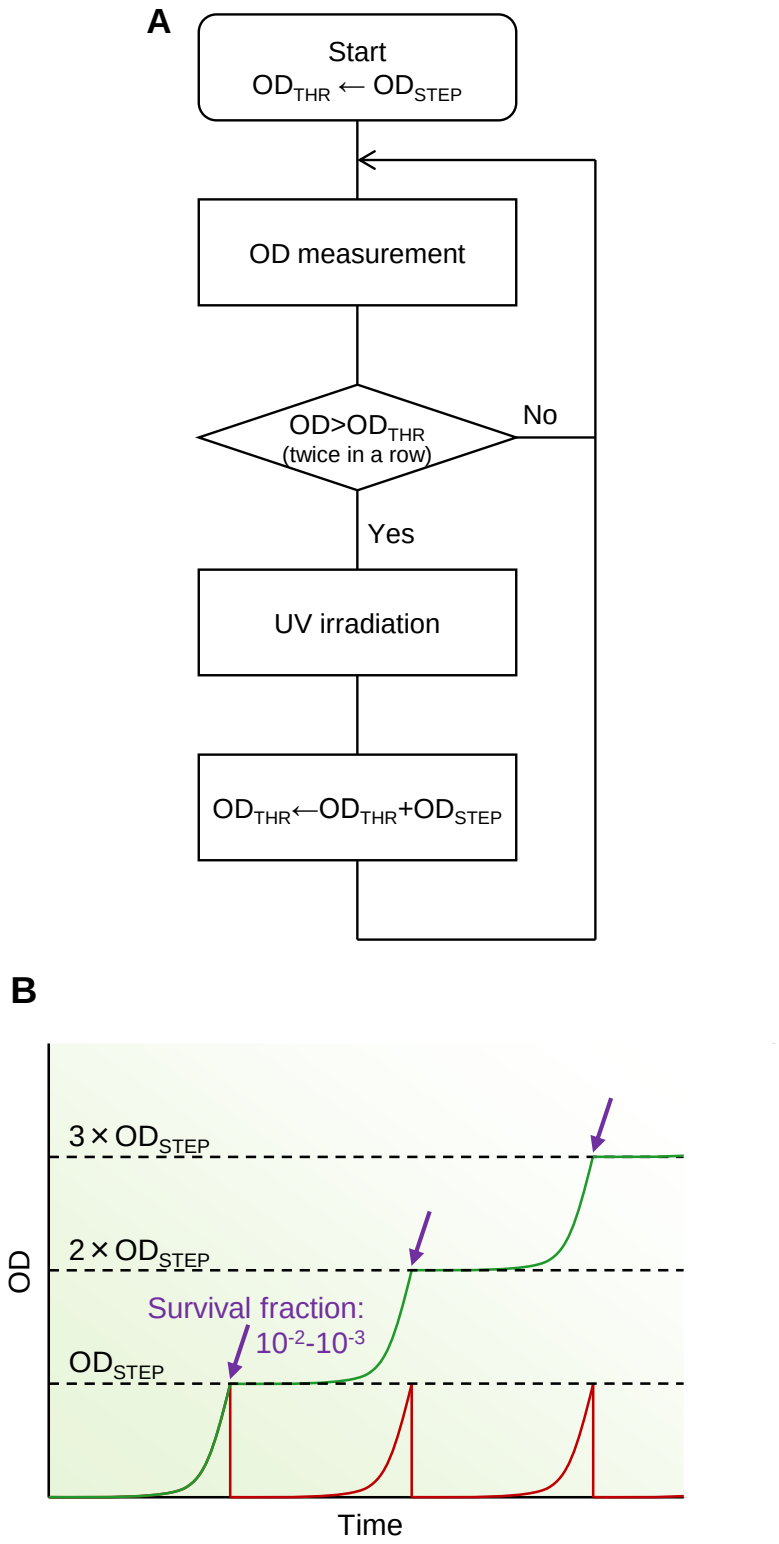


Fig. 3.4 Flowchart of operational procedure and ideal growth curve

3.2.4 Culturing various UV doses per exposure

The cells were cultured according to the operational procedure described above. Three cultures were used, and their UV exposure conditions were varied as 10, 20, or 30 s UV exposure / step growth. The culture experiments were continued for 4 days.

3.2.5 Mutation accumulation experiment

I conducted subculture experiments for 56 days, in which one round corresponds of 4 days. The time duration of one UV irradiation was determined such that the survival fraction of the wild-type cells was in the range of 10^{-2}–10^{-3}. The cells were serially transferred every 4 days and stocked at the same time. The experiment was replicated 6 times and continued for 56 days. Five replicates of evolution experiments were performed simultaneously.

3.2.6 Whole-genome sequencing

Genomic DNA was extracted using a Wizard Genomic DNA Purification kit (Promega). Multiplex analysis (usually 6 plex) was performed as pair end sequencing (251 bp) at 500 cycles using MiSeq Reagent Kit v 2 (Illumina). I used Burrows-Wheeler Aligner (BWA) to map the obtained reads onto the reference genome sequence of *E. coli* MDS42 (accession number: AP012306, the origin of NC_020518, GI: 471332236). I used SAMtools[22] to call BPSs and small InDels with the maximum read depth set to 500 (-D option) and the rest of parameters set to default. To ensure the quality of mutation calling, called mutations with Phread quality scores less than 100 were eliminated[23,24]. Additionally, BPSs with reads less than 90% were eliminated. I then supposed that the remaining mutations were dominant in the sequenced populations.

3.2.7 Measurement of base-pair substitution rate during the evolution experiments

I calculated the rate of base pair substitution by the following equation:

$$\rho = \frac{N_{syn} \cdot L}{F(syn) \cdot L_{CDS} \cdot D} \quad \text{(genome}^{-1}\ \text{day}^{-1}\text{)} \qquad (7)$$

where ρ, N_{syn}, L_{CDS}, L, and D are the rate of base pair substitution, the number of synonymous substitutions, the length of total coding DNA sequences per genome, genome size (3.98 Mbp for the MDS42), and the number of days of the evolution experiments, respectively. $F_{(syn)}$ is the probability of a synonymous substitution and was calculated based on the codon usage and the probability of

each substitution.

3.2.8 Calculating dN/dS values

The dN/dS value was calculated by dividing the number of nonsynonymous substitutions per nonsynonymous site, dN, with the number of synonymous substitutions per synonymous site, dS. Hence, the following equation gives the value:

$$dN / dS = \frac{N_{nsyn}}{N_{syn}} \cdot \frac{F(syn)}{1 - F(syn)}, \quad (8)$$

where N_{syn} and N_{nsyn} represent the number of synonymous and nonsynonymous substitutions, respectively. $F_{(syn)}$ is the probability that a substitution is synonymous (as described above).

3.3 Results

3.3.1 Calibration curve for optical density measurement

OD measurement is one of the two fundamental functions of the device (the other is UV irradiation, described later). To verify the reliable range of OD measurement, I drew a calibration curve of the device with a commercial spectrophotometer. I prepared serial dilutions of cell culture and measured their OD using the device and the spectrophotometer. The obtained calibration curve is shown in Fig. 3.5. The horizontal axis represents the measured value of the developed device and the vertical axis shows the output of the commercial spectrophotometer. As a result, the calibration curve was obtained as OD_{660} = 0.148 × (*Voltage*) in the OD range from 0.001 to 0.175, indicating that OD values in this range obtained with the device are reliable. This range depends on the resistance in the control circuit (*Rm* in Fig. 3.3), such that I can use a different range with small modifications to the system.

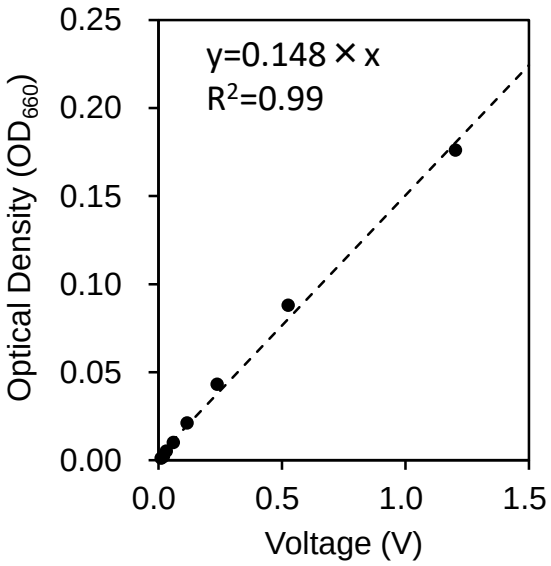


Fig. 3.5 Calibration curve of measured voltage and known cell density

3.3.2 UV killing curve

The relationship between UV dose and cell death is essential information for designing further experiments using the device. To obtain this information, I plotted a killing curve of UV irradiation. I prepared cell cultures of *E. coli* and exposed them to UV light over several time durations. The

cultures were then spread on an agar medium and cultured to determine the killing rate as the change in number of colony forming units (CFUs). As a result, I observed that cell death increased exponentially with time of UV irradiation. The fitted curve was log_{10} (*Survival fraction*) = –0.123 × (*UV irradiation time*). This result enabled selection of the UV dose in each irradiation event with consideration of its killing rate. This killing curve was used during the culture experiments described in this chapter and in Chapter 4 to set appropriate UV irradiation parameters.

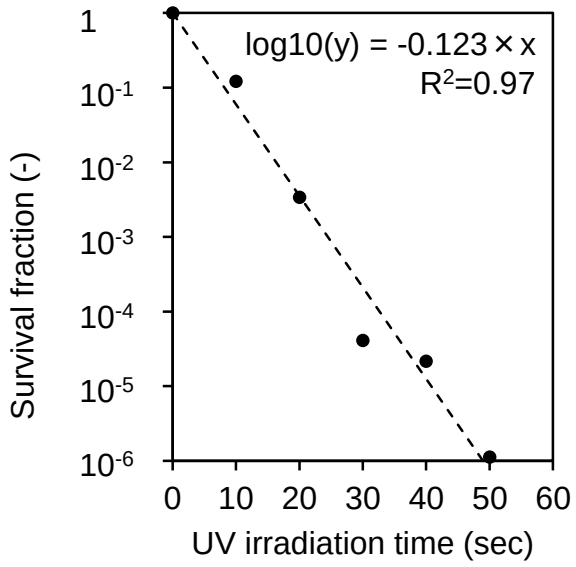


Fig. 3.6 UV killing curve of the device

3.3.3 Culturing various UV doses per exposure

I then checked whether feedback control from cell growth to UV irradiation worked correctly. In general, *E. coli* cell grows exponentially with time; on the other hand, UV light killed *E. coli* exponentially as dosage increased. Therefore, if the feedback control for irradiating with the maximum UV dose that maintains cell growth worked correctly, the total UV dose to the population per unit time should be the same regardless of irradiation frequency or size of a single dose. To confirm this, I conducted 3 culture experiments with various UV doses per step growth. I observed time-proportional increases in OD values (Fig. 3.7A), which was similar to the ideal concept of my control scheme (Fig. 3.4B). The speed of increase was larger for more frequent, smaller doses per exposure; on the other hand, the total UV doses were almost the same for all conditions (Fig. 3.7B), indicating that the feedback control worked appropriately.

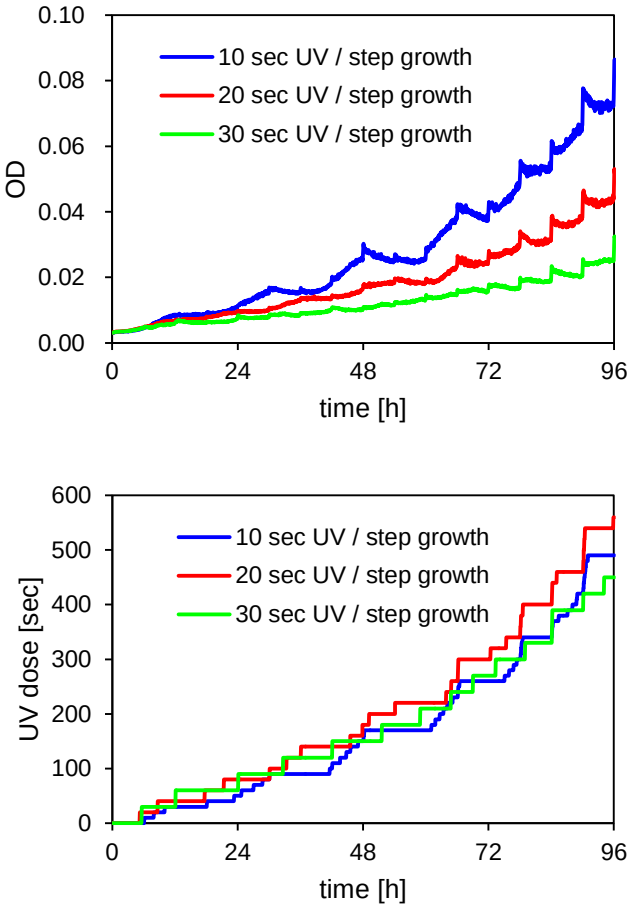


Fig. 3.7 Time-proportional growth curve controlled with the system

3.3.4 Mutation accumulation experiment

I conducted a mutation accumulation experiment for 56 days using the device to examine the mutagenesis performance. Evolution experiments on 5 lineages were performed simultaneously. The results are shown in Fig. 3.8A. The x-axis represents culturing time and y-axis represents OD value of the cell culture. Cell cultures were serially transferred into fresh media in new tubes, as described in 3.2.5. Fig. 3.8B and C are enlarged views of the beginning and ending rounds, respectively in Fig. 3.8A. Fig. 3.8D and E show the accumulated UV doses for the beginning and ending rounds, respectively. At the beginning of the experiment, the five lineages showed very similar behavior (Fig. 3.8B and D). On the other hand, at the end of the experiment, these 5 lineages showed greatly divergent behavior (Fig. 3.8C and E). These results suggested that the lineages adapted to UV radiation by various evolutionary pathways.

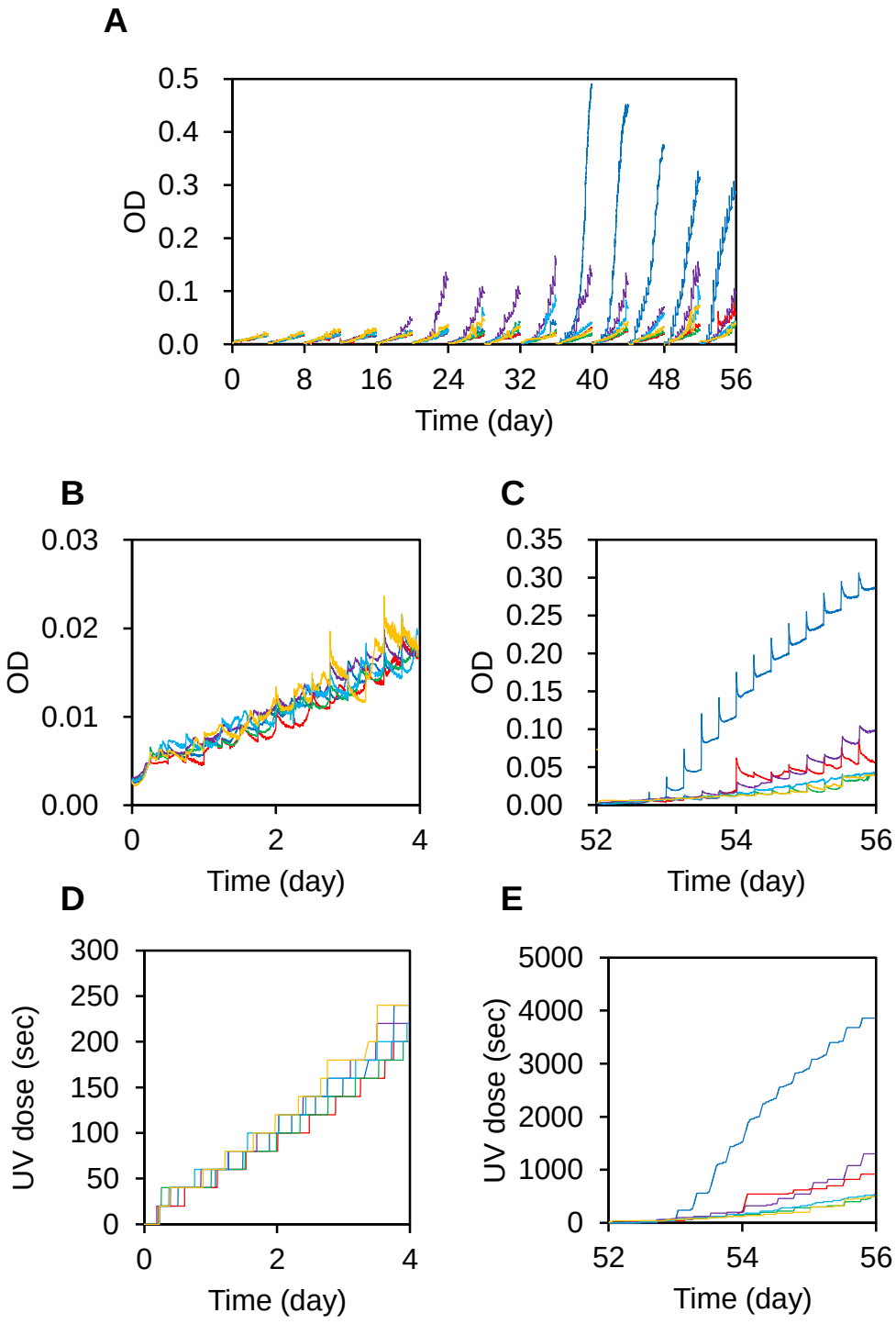
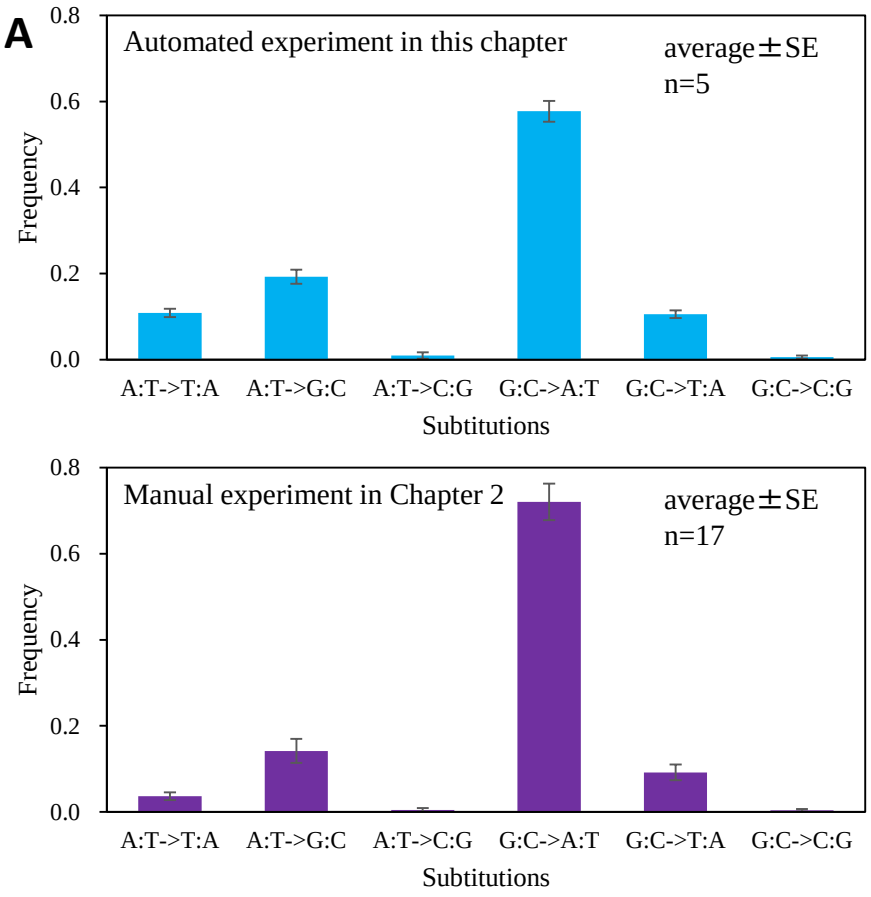


Fig. 3.8 Transition of growth and UV dose during the mutation accumulation experiments

3.3.5 Detection of accumulated mutations

In order to verify the mutagenesis performance of the developed device, genomic analysis of the samples obtained by the mutation accumulation experiment was performed. Whole-genome sequencing of the cells at the end of the experiment (56th day) revealed the number of genomic mutations accumulated in the populations. The rates of BPSs during the evolution experiments were calculated based on the spectra and numbers of synonymous mutations. As a result, the accumulation rate was much higher than that in the absence of UV reported by Drake et al.[17] (Fig. 3.9B). Additionally, the BPS spectrum, the BPS rate, and dN/dS values were comparable to the results described in Chapter 2 from manual application of UV radiation as a mutagen (Fig. 3.9 and Fig. 2.6, 2.7, 2.8).



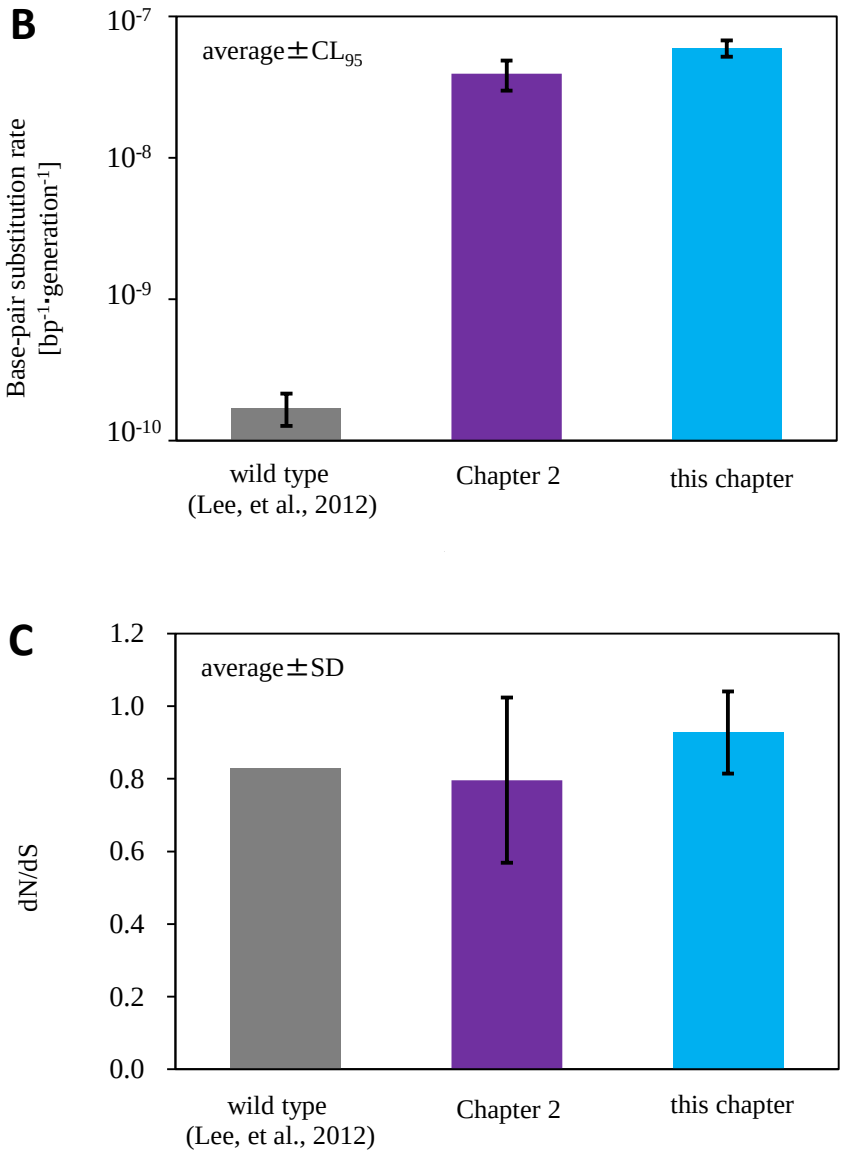


Fig. 3.9 Mutational profiles induced by UV irradiation of cell cultures with the developed device

3.4 Discussion

In this chapter, I developed a culture device that automatically irradiates cell cultures according to OD measurements. The basic abilities of the device to measure cell concentration and apply UV light were confirmed. In addition, a mutation accumulation experiment was carried out, and its results were compared with those of the manual experiment described in Chapter 2.

3.4.1 Does feedback control work?

In this chapter, to realize the automation of a UV irradiation cell culture experiment, I designed a device that can measure the concentration of bacterial cells and apply ultraviolet light according to the concentration. The concentration was measured by changes in light intensity emitted from an LED arranged on the side of a test tube and received by a light receiving element arranged on the opposite side. UV irradiation was performed from the bottom of the test tube using a deep UV LED. If feedback control worked correctly, the UV dose per unit time would converge regardless of the UV dose per unit growth (3.3.3). In order to confirm this, I conducted the experiment with various UV doses per unit cell growth. The results showed that the UV doses per unit time converged (Fig. 3.7B), suggesting that the feedback control worked correctly. The UV dose accumulation curves were skewed upward over 4 days, and the radiation dose per unit time increased. There are two possible explanations. The first is that the UV tolerance of the cell population evolved genetically in 4 days. The second is the appearance of a cell population that temporarily escaped from UV light by clustering on the wall of the test tube. Although these phenomena are not exclusive, the latter could inhibit the uniformity of the population structure and prohibit the increase in mutation rates. One possible solution is to transfer cells to a new tube before they can form clusters on the tube walls.

3.4.2 Changes in the tolerable UV dose via mutation accumulation experiments

The accumulated UV doses of the 5 replicates were almost the same in the first few rounds (Fig. 3.8D), though the doses greatly diverged at the 14^{th} round (52^{nd}–56^{th} day) of the experiment (Fig. 3.8E). I assumed that this was because each lineage independently acquired UV tolerance at different times. As a result, the total UV dose for each lineage should have varied. However, genomic analysis revealed that the mutation rates did not differ much between lineages (Fig. 3.9B). This could be because the enhanced growth due to improved UV tolerance was counterbalanced by the increase in UV irradiation due to the feedback control, resulting in a similar generation and thus similar mutations for all lineages. This implied that the designed function of the device, automatically

applying the maximum UV irradiation tolerable to the population, was realized.

3.4.3 Mutational profile of our automated UV-irradiating culture device

One possible concern related to the automation of these experiments is lower mutational efficiency than that in manual experiments, e.g., lower mutation rates or narrower mutational spectra. However, we found that the mutational profiles in the automated experiments remained almost the same as those in the manual experiment (3.3.5). Therefore, the experimental system has been successfully improved by the development of the device.

3.5 Summary

In this chapter, I described the design of an automated culture device that can measure bacterial cell concentration and apply UV light according to that concentration. Cell concentration was measured by the change in light intensity emitted from an LED and detected by a receiving element arranged on opposite sides of a test tube. UV irradiation was performed from the bottom of the test tube using a deep UV LED. The housing was manufactured using a 3D printer. The device was controlled using programs running on an Arduino Uno microcontroller driven by a laptop computer. I conducted a short-term mutation accumulation experiment using the developed system and analyzed the genome sequence of *E. coli* after the experiment with a next-generation sequencer. The results confirmed that the mutation rate was approximately 100 times higher than that without UV irradiation, similar to the experiment in which UV irradiation was performed manually. I succeeded in the development of a device that observes the state of an *E. coli* population and automatically performs appropriate UV irradiation. This achievement enabled longer-term evolution experiments under stable, highly mutagenic environments.

Chapter 4: Long-term evolution experiment with UV exposure

(1) A UV-irradiated culture experiment was carried out for 700 days using the system developed in Chapter 3.

(2) The base-pair substitution rate was maintained approximately 100 times higher than that in a typical environment without UV radiation for one year, after which time the UV tolerance of E. coli improved.

(3) These achievements will contribute to the planning of laboratory evolution experiments that require large numbers of mutations.

4.1 Introduction

The purpose of the experiment described in this chapter was to clarify whether the growth rate of *E. coli* and highly mutagenic conditions can be maintained even after mutation accumulation by UV mutagenesis. The results in Chapter 2 indicated that about one year of irradiation was necessary to accumulate the number of mutations capable of decreasing the growth rate as reported in previous studies[28,29]. With the equipment developed in Chapter 3, it was possible to conduct experiments over such a long period.

In these experiments, I performed *E. coli* evolution experiments with UV irradiation for 700 days using the device designed in Chapter 3 and analyzed the genetic variation that emerged on the *E. coli* genomes. The accumulation of mutations can cause defects in *E. coli* cell systems. Thus, such accumulation over time can gradually decrease the proliferation ability of the cells[28,29], and the experiment cannot continue. Therefore, it was not self-evident that such experiments were possible on a yearly scale. Thus, conducting the evolution experiment was a challenge in itself. Additionally, I showed how the mutational profiles induced by UV radiation on the *E. coli* genome changed over a longer time span than that of the experiment in Chapter 2. As the experiment continued, the *E. coli* adaptation process reduced the mutation rate achievable with UV light. In addition, in an attempt to predict genetic evolution from existing data, I compared the information of each gene obtained from the database with the mutational patterns generated in the experiment.

4.2 Methods

4.2.1 Strain and media

I used *E. coli* MDS42 and a chemically defined medium, mM63, which contained 62 mM K_2HPO_4, 39 mM KH_2PO_4, 15 mM $(NH_4)_2SO_4$, 2 μM $FeSO_4 \cdot 7H_2O$, 15 μM thiamine hydrochloride, 203 μM $MgSO_4 \cdot 7\ H_2O$, and 22 mM glucose. Cell cultures were performed at 37 ºC with shaking at 200 rpm. To inhibit growth on the walls of tubes, the shaking speed was raised to 300 rpm for 1 min in every 6 h.

4.2.2 Evolution experiment

I conducted subculture experiments for 700 days, in which one round corresponded to 4 days. The time duration for one UV irradiation was set such that the survival fraction of wild-type cells was 10^{-2}–10^{-3}. The cells were serially transferred every 4 days and stocked at the same time. The experiment was replicated 6 times.

4.2.3 Measurement of maximal growth rate

The frozen, stocked cells were inoculated into 100 μl of mM63 broth on a 96-well microplate and cultured at 37 °C with shaking at 550 rpm for 12 h. The cell culture was then diluted 100 times with fresh mM63 broth and poured into wells in a microplate such that the culture volume was 100 μl/well (10 wells for each evolved strain and 60 wells for each ancestral strain). The microplate was incubated at 37 °C with shaking in a plate reader (Infinite F200 PRO, Tecan). The plate reader measured OD_{595} every 15 min. I calculated the maximum growth rate [h^{-1}] by fitting the slopes of the growth curves during the exponential growth phase (OD_{595} = 0.01–0.06) to the standard Malthusian growth model.

4.2.4 Whole-genome sequencing

Genomic DNA was extracted using a Wizard Genomic DNA Purification kit (Promega). Multiplex analysis (usually 6 plex) was performed as pair-end sequencing (251 bp) at 500 cycles using MiSeq Reagent Kit v 2 (Illumina). I used Burrows-Wheeler Aligner (BWA) to map the obtained reads onto the reference genome sequence of *E. coli* MDS42 (accession number: AP012306, the origin of NC_020518, GI: 471332236). I used SAMtools[22] to call BPSs and small InDels with the maximum

read depth set to 500 (-D option) and the rest of parameters set to default. To ensure the quality of mutation calling, called mutations with Phread quality scores less than 100 were eliminated[23,24]. Additionally, BPSs with reads less than 90% were eliminated. I then considered the remaining mutations to be dominant in the sequenced populations.

4.2.5 Measurement of base-pair substitution rate during the evolution experiment

I calculated the rate of base-pair substitution by the following equation:

$$\rho = \frac{N_{syn} \cdot L}{F(syn) \cdot L_{CDS} \cdot D} \quad \text{(genome}^{-1}\text{ day}^{-1}\text{)} \qquad (9)$$

where ρ, N_{syn}, L_{CDS}, L, and D are the rate of base pair substitution, the number of synonymous substitutions, the length of total coding DNA sequences per genome, genome size (3.98 Mbp for the MDS42), and the number of days of the evolution experiments, respectively. $F_{(syn)}$ is the probability of a synonymous substitution and was calculated based on the codon usage and the probability of each substitution.

4.2.6 Calculating dN/dS values

The dN/dS value was calculated by dividing the number of nonsynonymous substitutions per nonsynonymous site, dN, with the number of synonymous substitutions per synonymous site, dS. Hence, the following equation gives the value:

$$dN/dS = \frac{N_{nsyn}}{N_{syn}} \cdot \frac{F(syn)}{1 - F(syn)}, \qquad (10)$$

where N_{syn} and N_{nsyn} represent the number of synonymous and nonsynonymous substitutions, respectively. $F_{(syn)}$ is the probability that a substitution is synonymous.

4.2.7 Comparing mutational pattern and essentiality of each gene in *E. coli*

Genes were classified by these criteria: (1) genes with at least one nonsynonymous BPS among 6 lineages and (2) genes with at least one inactivating mutation among 6 lineages. The essentiality of each gene of *E. coli* was taken from the Profiling of *E. coli* Chromosome database[19]. The essential genes were defined by a previous study that conducted deletion experiments on each gene[49]. This information was queried for each gene, and whether essentiality can be used to predict mutation accumulation was considered in the present experiments.

4.2.8 Gene ontology enrichment analysis for mutated genes

I performed gene ontology (GO) enrichment analysis using the R Bioconductor package GOstats[20] for Biological Process (BP), Cellular Components (CC), and Molecular Functions (MF). The significance level was set such that the false discovery rate (FDR) cutoff was 0.05 using the R qvalue package[21]. The gene sets used for GO enrichment analysis were prepared by extracting genes in which the following three types of mutation occurred at least once among 6 lineages: (1) synonymous BPSs; (2) nonsynonymous BPSs; and (3) inactivating mutations, which include nonsense mutations and insertions or deletions in gene coding regions. In this study, I counted nonsense mutations, which place stop codons in incorrect wrong positions, and short insertions and deletions, which cause frame shifts in codon reading. These mutations reportedly have deleterious effects on gene function[50].

4.3 Results

4.3.1 Evolution experiment

It was not self-evident that evolution experiments were possible on a yearly scale, because accumulation of mutations can cause defects in *E. coli* cell systems that result in the loss of proliferation ability, ending the experiments prematurely. Fig. 4.1 shows changes in the final OD values on each day during the evolution experiments and changes in the UV doses in each round. I succeeded in culturing cells for 700 days in all 6 lineages. Considering that the UV dose per round represents the UV tolerance of the *E. coli*, this result indicated that a rise in UV tolerance occurred approximately 100 days from the start of the experiment. This suggested that UV-irradiated culture experiments could be continued on *E. coli* for several hundred days.

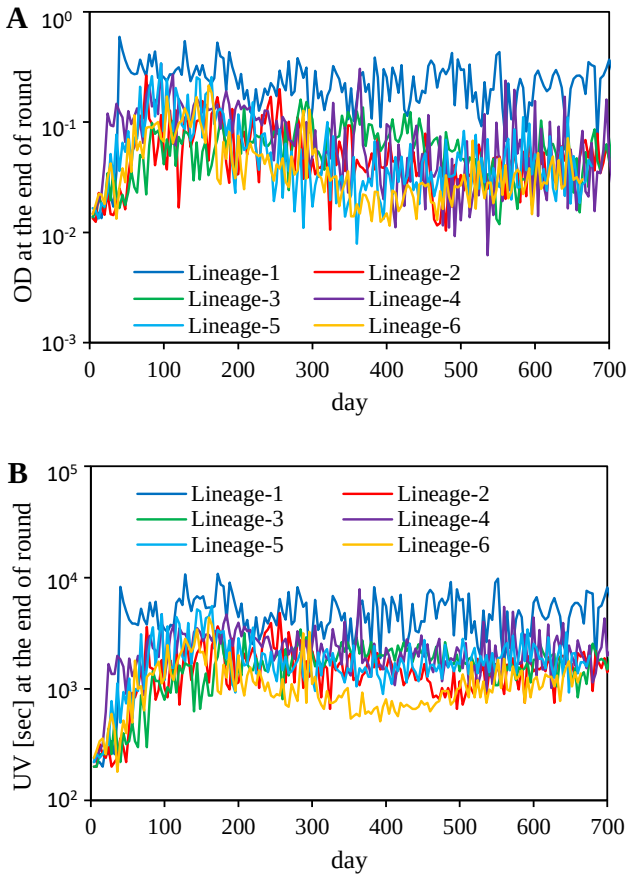


Fig. 4.1 Time records of the 700-day evolution experiments

The final OD value (A) and the UV dose per round (B) are shown.

4.3.2 Change in maximal growth rates

The *E. coli* populations were considered to be under selective pressure for faster growth during the experiment; on the other hand, extremely high BPS rates produce many deleterious mutations that can disrupt growth ability. Since the experiment will fail if the growth rate continuously decreases, it is important to know the tendency of the genetic evolution of the growth rate. I measured the maximal growth rates of the stocked cells every 7 rounds (28 days) following the method described in 4.2.3. There were temporary drops in growth rates on the 308th day. However, this is likely an experimental artifact, because it occurred simultaneously in multiple lineages and recovered soon after. The changes in the maximal growth rates of the stocked cells are shown in Fig. 4.2. In the Lineage-1, no tendency in the change in growth rate was detected because of large measurement errors. In Lineage-3, the maximum growth rate seemed to gradually decrease. However, in the other 4 lines, the growth rates exhibited no tendency to decrease with time.

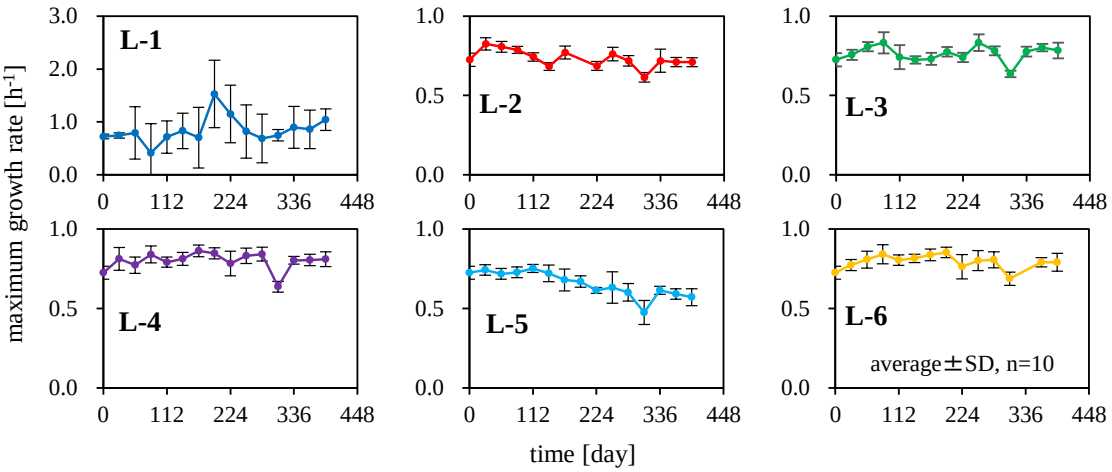


Fig. 4.2 Change in maximal growth rates in evolution experiments

4.3.3 Rate and spectrum of accumulated base-pair substitutions

The genome sequences of the evolving lineages were analyzed using cells stocked on the 56th, 168th, and 336th day and next-generation sequencing (4.2.4). The changes in the mutation rates estimated by analyzing the evolved strain genomes are shown in Fig. 4.3. There was no significant difference between the values in the early and late stages (56th and 336th day, paired t-test, n=6, p=0.91). In Lineage-1, the mutation rate tended to decrease greatly, but the mutation rates remained unchanged in the other 5 lineages.

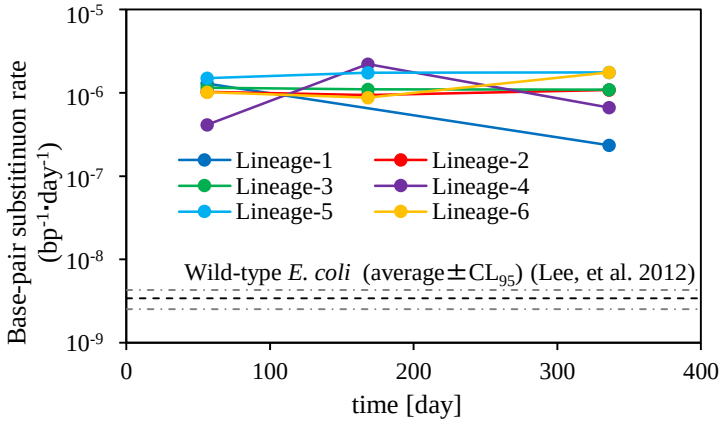


Fig. 4.3 Base-pair substitution rates during the evolution experiment with UV irradiation

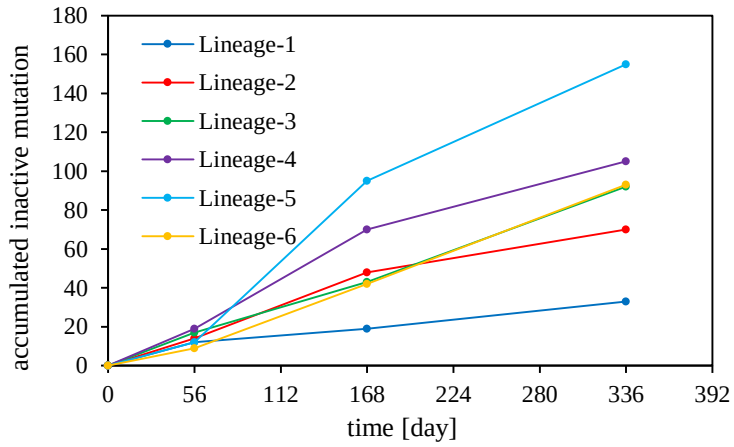


Fig. 4.4 Accumulation of gene-inactivating mutations in the genomes of each lineage

4.3.4 dN/dS values in the evolution experiment

The sequenced genomes from the 56th, 168th and 336th days (described above) of the evolved lineages were further analyzed. The changes in the dN/dS values obtained by analyzing the genomes of the evolved strains are shown in Fig. 4.5. There was no significant difference between the values in the early and late stages (56th and 336th day, paired t-test, n=6, p=0.90). As a result, the values remained at approximately 0.8 with the exception of that of Lineage-1. This indicated that only approximately 10–20% of the amino acid substitutions were retained before being excluded by selection in the period of about one year.

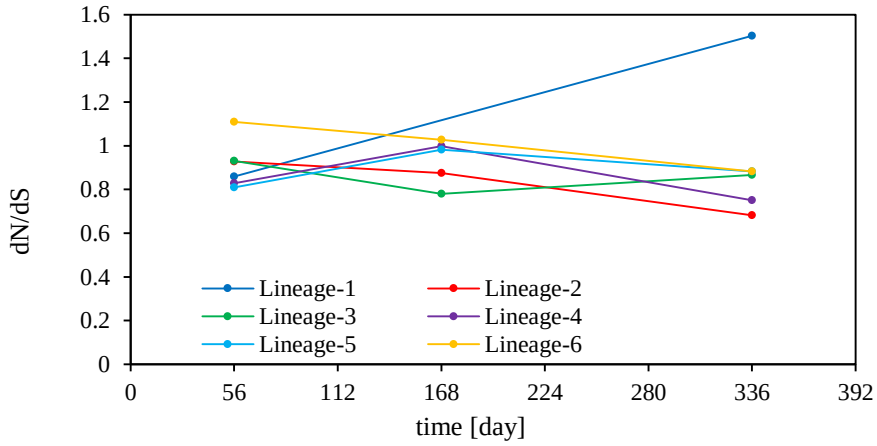


Fig 4.5 dN/dS values in the evolution experiment

4.3.5 Essentiality of *E. coli* genes

I examined the predictability of genetic evolution by comparing data on essentiality of each gene with the mutational pattern on each gene obtained in this study. *E. coli* genes were classified into those with and without nonsynonymous and/or inactivating mutations through the evolution experiment. The essentiality of each gene was taken from the Profiling of E. coli Chromosome database. The mutational pattern and essentiality of the genes were combined and shown in Fig. 4.8. The vertical axis shows the frequency of essential or nonessential genes. The number near the top of a bar represents the number of genes in each category. As a result, I demonstrated that inactivating mutations rarely occurred on the essential genes. There was one gene, *aspS*, which is essential and had one inactivating mutation. However, the *aspS* amino acid sequence length was 590, and the mutation was a nonsense mutation in which the 582nd amino acid was changed to a stop codon. Therefore, as most of the *aspS* protein remained unchanged, its essential function may not be lost. These results suggest that, basically, inactivating mutations are not fixed in genes that are considered essential and that the mutation pattern of evolution experiments can be at least partially predicted by essentiality.

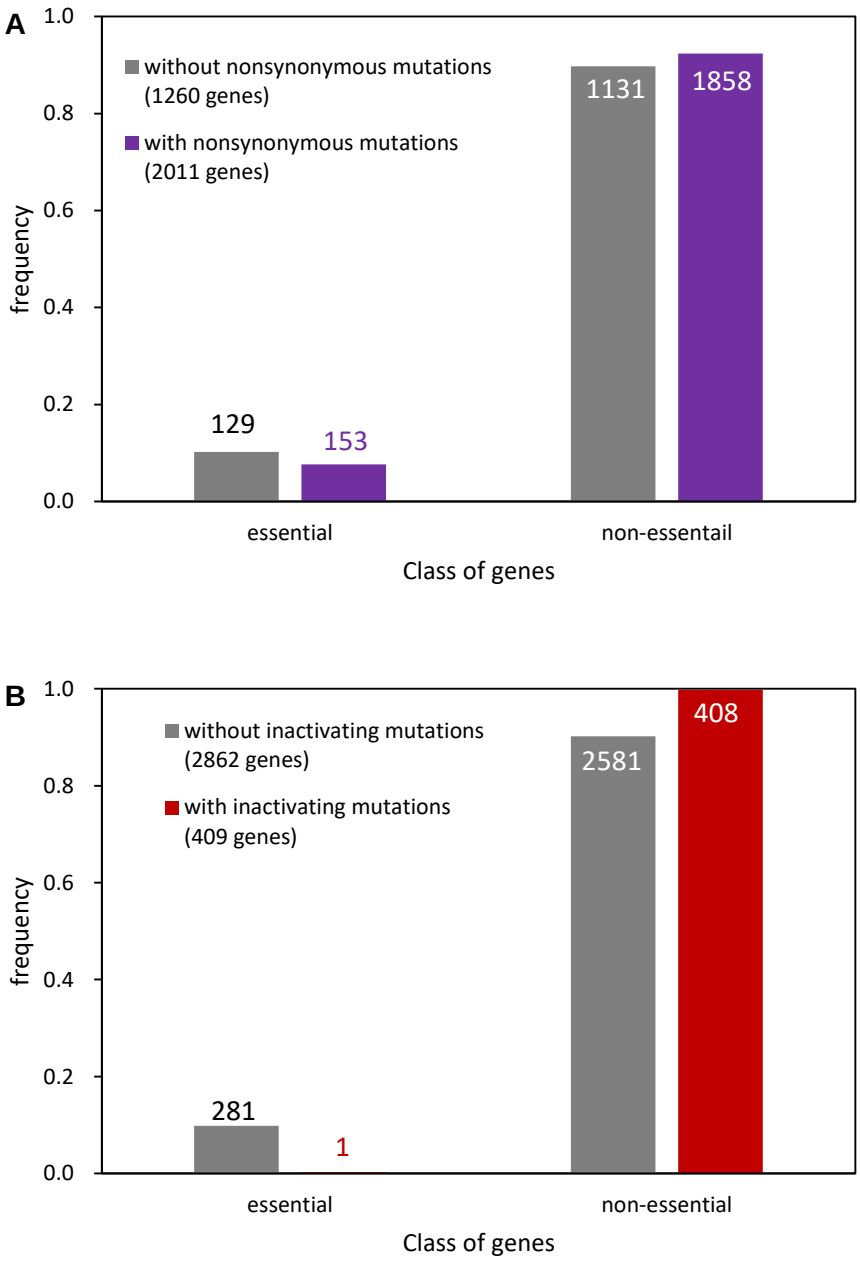


Fig. 4.8 Relationship between mutational patterns and essentiality of genes

4.4 Discussion

In this study, I investigated two aspects of UV mutagenesis, the occurrence and accumulation of mutations, and focused on them in short-term and long-term experiments, respectively. The aim of the study was to clarify how short-term trends in UV mutagenesis would change as mutations accumulated on *E. coli* genomes. Specifically, I examined (1) changes in UV-induced mutation rate and (2) changes in their growth ability.

(1) Previous studies have shown that most mutations have harmful effects on *E. coli*[5]. It is also known that *E. coli* easily evolves such that its mutation rate decreases[42-47]. In the experiments of this study, it was expected that *E. coli* would evolve to reduce the mutagenic effect of UV radiation on its genome as mutation accumulated. With the system designed in Chapter 3, the UV dose was automatically increased as UV tolerance increased. However, it was unclear whether this would also maintain a high mutation rate. In the experiments discussed in this chapter, I analyzed *E. coli* genomes obtained at multiple time points in the long-term experiments; the results indicated that the high mutation rate caused by UV irradiation was approximately 100 times that of non-irradiated cell genomes and remained this high for one year.

(2) It is also known that accumulation of mutations on microbial genomes causes growth inhibition. In previous studies, it has been suggested that when approximately 1000 mutations accumulate on the *E. coli* genome without selective pressure on the growth rate, the rate will decrease by half[28,29]. Other studies have reported microbial species[48] that have evolved such that their mutation rates rise as their growth rates decrease, even in the nature, which is expected to put selective pressure on growth rates. Though there was selective pressure on growth rate in the experiments of this study, the growth rate of *E. coli* could still decrease after the accumulation of more than 1000 mutations. Thus, decreased growth rates threaten the sustainability of culture experiments. The experiments described in this chapter demonstrated that the mutation and growth rates of *E. coli* can be maintained for more than one year in experiments with UV irradiation, even when the bacterial genomes accumulate more than 1000 mutations. The knowledge alone that such a high mutation rate by UV mutagenesis can be sustained for one year in laboratory evolution experiments is a significant result.

In the experiments described in this chapter, I demonstrated that UV-irradiated cell culture experiments could be continued for years, while maintaining high growth ability in most cases. Additionally, BPS rates and dN/dS values were similar to those in the short-term experiments

described in the previous chapters. These results demonstrated that evolutionary experiments with constant mutation profiles can be performed for several hundred days. Furthermore, I succeeded in finding significantly rich GO terms in the mutated genes. In addition, the use of gene essentiality to predict mutational patterns was promising.

4.4.1 How did *E. coli* adapt to its UV-irradiated environment?

Through evolution experiments, *E. coli* adapted to its UV-irradiated environment (Fig. 4.1). To determine which genes contributed to the adaptive genetic evolution to UV-irradiated environments, I conducted GO enrichment analysis of the mutated genes. The results of the analysis are shown in Table 1. The Mut column indicates the type of mutation (S: synonymous BPS; N: nonsynonymous BPS; I: inactivating mutations). The Cat column indicates the GO term category (BP: Biological process; CC: Cellular components: MF: Molecular functions). The significance level was adjusted such that FDR < 0.05. GO terms related to the cell membrane (i.e., transporters, channels, and others) were found to be related to function-inactivating mutations in Lineages 1–5, while Lineage-6 had no significant GO terms. These results were consistent with the lineages (1–5) in which *E. coli* evolved to aggregate strongly, suggesting that changes in membrane properties may have contributed to the acquisition of cohesiveness and UV tolerance. I concluded that *E. coli* acquired the property of aggregating, thereby creating a structure that protects the inner cells from UV radiation. Aggregates of cells are frequently observed in cultures of evolved strains. Since aggregation of cells is thought to be related to a change in cell membrane properties, it is consistent with the result that many mutations were found among cell membrane-related genes in GO analysis. Please note that in order to verify these hypotheses, it will be necessary to conduct a similar analysis on the accumulation pattern of mutations under non-UV irradiation conditions and to examine any differences in the detected GO terms.

Table 1 GO enrichment analysis for mutated genes

GO enrichment analysis				
Lineage	**Mut**	**Cat**	**ID**	**Term**
1	S	BP	GO:0002099	tRNA wobble guanine modification
1	S	BP	GO:0006522	alanine metabolic process
1	S	BP	GO:0006523	alanine biosynthetic process
1	S	BP	GO:0008616	queuosine biosynthetic process
1	S	BP	GO:0009078	pyruvate family amino acid metabolic process
1	S	BP	GO:0009079	pyruvate family amino acid biosynthetic process
1	S	BP	GO:0015942	formate metabolic process
1	S	BP	GO:0015944	formate oxidation
1	S	BP	GO:0016259	selenocysteine metabolic process
1	S	BP	GO:0016260	selenocysteine biosynthetic process
1	S	BP	GO:0030632	D-alanine biosynthetic process
1	S	BP	GO:0034660	ncRNA metabolic process
1	S	BP	GO:0042851	L-alanine metabolic process
1	S	BP	GO:0042852	L-alanine biosynthetic process
1	S	BP	GO:0043213	bacteriocin transport
1	S	BP	GO:0046116	queuosine metabolic process
1	S	BP	GO:0046144	D-alanine family amino acid metabolic process
1	S	BP	GO:0046145	D-alanine family amino acid biosynthetic process
1	S	BP	GO:0046416	D-amino acid metabolic process
1	S	BP	GO:0046436	D-alanine metabolic process
1	S	BP	GO:0046437	D-amino acid biosynthetic process
1	S	MF	GO:0004872	receptor activity
1	S	MF	GO:0008713	ADP-heptose-lipopolysaccharide heptosyltransferase
1	S	MF	GO:0008920	lipopolysaccharide heptosyltransferase activity
1	S	MF	GO:0015343	siderophore transmembrane transporter activity
1	S	MF	GO:0016740	transferase activity
1	S	MF	GO:0022885	bacteriocin transmembrane transporter activity
1	S	MF	GO:0030151	molybdenum ion binding
1	S	MF	GO:0042912	colicin transmembrane transporter activity
1	S	MF	GO:0042927	siderophore transporter activity
1	S	MF	GO:0060089	molecular transducer activity
1	I	MF	GO:0022892	substrate-specific transporter activity
1	I	MF	GO:0042936	dipeptide transporter activity
2	S	MF	GO:0003678	DNA helicase activity
2	S	MF	GO:0019842	vitamin binding
2	S	MF	GO:0030170	pyridoxal phosphate binding
2	S	MF	GO:0030554	adenyl nucleotide binding
2	S	MF	GO:0036094	small molecule binding

GO enrichment analysis

Lineag	Mut	Cat	ID	Term
2	S	MF	GO:0043167	ion binding
2	S	MF	GO:0043168	anion binding
2	S	MF	GO:0048037	cofactor binding
2	S	MF	GO:0070279	vitamin B6 binding
2	N	BP	GO:0018106	peptidyl-histidine phosphorylation
2	N	BP	GO:0018202	peptidyl-histidine modification
2	N	BP	GO:0023014	signal transduction by protein phosphorylation
2	N	MF	GO:0000155	phosphorelay sensor kinase activity
2	N	MF	GO:0004672	protein kinase activity
2	N	MF	GO:0004673	protein histidine kinase activity
2	N	MF	GO:0004872	receptor activity
2	N	MF	GO:0009927	histidine phosphotransfer kinase activity
2	N	MF	GO:0015267	channel activity
2	N	MF	GO:0016775	phosphotransferase activity, nitrogenous group as acceptor
2	N	MF	GO:0022803	passive transmembrane transporter activity
2	N	MF	GO:0038023	signaling receptor activity
2	N	MF	GO:0060089	molecular transducer activity
2	I	MF	GO:0005216	ion channel activity
2	I	MF	GO:0015267	channel activity
2	I	MF	GO:0016787	hydrolase activity
2	I	MF	GO:0022803	passive transmembrane transporter activity
2	I	MF	GO:0022833	mechanically gated channel activity
2	I	MF	GO:0022836	gated channel activity
2	I	MF	GO:0022857	transmembrane transporter activity
3	N	MF	GO:0016798	hydrolase activity, acting on glycosyl bonds
3	N	MF	GO:0030151	molybdenum ion binding
4	N	CC	GO:0016020	membrane
4	N	CC	GO:0071944	cell periphery
4	I	MF	GO:0004180	carboxypeptidase activity
4	I	MF	GO:0015175	neutral amino acid transmembrane transporter activity
4	I	MF	GO:0015190	L-leucine transmembrane transporter activity
4	I	MF	GO:0015291	secondary active transmembrane transporter activity
4	I	MF	GO:0015293	symporter activity
5	I	CC	GO:0016020	membrane
5	I	MF	GO:0005215	transporter activity
5	I	MF	GO:0008713	ADP-heptose-lipopolysaccharide heptosyltransferase
5	I	MF	GO:0008920	lipopolysaccharide heptosyltransferase activity
5	I	MF	GO:0022857	transmembrane transporter activity
5	I	MF	GO:0022891	substrate-specific transmembrane transporter activity

4.4.2 *E. coli* genes are redundant in the laboratory environment

If most *E. coli* genes are important to cellular systems, the accumulation rate of gene-inactivating mutations should decrease with time, or the maximum growth rate should decrease with time. However, inactivating mutations were accumulating almost linearly over time (4.3.3), and the growth rates did not decrease in most lineages (4.3.2). These results implied that the functions of the genes of *E. coli* are highly redundant in current laboratory environments. Because of this property, the proposed artificial genome reduction of *E. coli* towards a minimal genome factory and/or a minimal cell model[32-35] would be promising.

4.5 Summary

In order to accumulate mutations in the genome of *E. coli*, evolution experiments at high mutation rates were carried out for 700 days. The experiments were conducted in parallel with 6 lineages. Then, using a next-generation sequencer, the genome sequences of the evolved strains at time points of 2, 6, and 12 months were analyzed and mutations were detected. I found that mutation rates on the *E. coli* genome were maintained at levels approximately 100 times higher than those under non-irradiation conditions. Furthermore, mutations patterns of each gene obtained in the experiment were analyzed. GO enrichment analysis indicated that cell membrane-related genes were significantly mutated, suggesting that changes in membrane properties contributed to the observed rise of UV tolerance. Additionally, I found that inactivating mutations are not fixed in genes considered essential, implying the effectiveness of integrating knowledge from multiple databases. The above results suggested that the method of this study was useful for evolutionary engineering, and that it could contribute to the understanding of the adaptive mechanisms of organisms.

Chapter 5: Conclusion

5.1 Summary of this thesis

In this study, I aimed to clarify the types of mutations generated on the *E. coli* genome by UV irradiation. In addition, I investigated how the mutation profiles of *E. coli* shift in highly mutagenic environments. I examined two processes of UV mutagenesis: emergence and accumulation of mutations. Investigations related to these processes were discussed in Chapters 2 and 4, respectively. The latter process, accumulation of mutations, required relatively long culture experiments. Therefore, as described in Chapter 3, I developed a culture system that enables long and stable evolution experiments. The results demonstrated that the base-pair substitution rates of *E. coli* were approximately 100 times those measured in culture experiments without UV exposure. Additionally, the spectrum of base-pair substitution induced by UV radiation was examined shown in this research. These achievements will contribute to the planning of laboratory evolution experiments that require large numbers of mutations. Some applications based on these findings are discussed in 5.2.

5.1.1 Mutational profiles of UV exposure on the *E. coli* genome

In the experiments described in Chapter 2, I aimed to clarify the mutational profiles induced by UV irradiation on the *E. coli* genome. BPS spectrum, BPS rate per generation, and dN/dS values were focused on in this study. I conducted short-term mutation accumulation experiments on the *E. coli* genome with UV irradiation. Through next-generation sequencing analysis of the evolved genomes, I demonstrated 4 of the 6 types of substitution and a high upper limit to the rate of UV-induced BPSs. Additionally, the dN/dS value was near one, suggesting that most UV-induced amino acid-altering mutations were not excluded by selective pressure.

5.1.2 Development of an automated UV-irradiated culture system

As described in Chapter 3, I designed a device that measures the concentration of bacterial cells and applies UV irradiation according to the cell concentration to realize the automation of UV-irradiated cell culture experiments. The control scheme was designed such that the device applied the maximum UV dose while retaining cell populations. I then conducted a short-term mutation accumulation experiment using the developed system and analyzed the genome sequence of *E. coli* with a next-generation sequencer. The obtained mutational profile was similar to that in the manual irradiation experiment. Therefore, I succeeded in the development of a device which observes the

state of an *E. coli* population and automatically performs appropriate UV irradiation enabling longer-term evolution experiments under stable and highly mutagenic conditions.

5.1.3 Long-term evolution experiment with UV exposure

As described in Chapter 4, I conducted longer-term evolution experiments for 700 days to examine how the mutational profile of UV radiation changes over time. Using a next-generation sequencer, the genome sequences of the evolved strains at time points of 2, 6, and 12 months were analyzed, and mutations were detected. The results showed that the BPS rate on the *E. coli* genome was approximately 100 times that observed under non-irradiated conditions. Additionally, I found that inactivating mutations were not fixed in genes considered essential, implying that integrating knowledge from databases is useful in predicting mutation profiles. Further, the inactivating mutations accumulated over time without growth defects in most cases, suggesting that my system could be used for application as a novel genome reduction method.

5.2 Discussion

5.2.1 Application for high-mutagenicity evolution experiments

As presented in Chapter 2, the mutational profile of UV irradiation has high rate limit and produces 4 types of substitutions. The system described in Chapter 3 enabled maximization and stabilization of the UV-induced mutagenesis experiments. These findings will contribute to the progress of mutagenesis techniques used on microorganisms in laboratory. As an example of an application, I would like to mention the comparison of gene essentiality and the mutational pattern described in Chapter 4, which successfully showed consistency between knowledge obtained from database and the results of evolution experiments. I expect that, with such approaches, it will be possible to utilize our knowledge for prediction and understanding of genetic evolution. For example, the expression data of each *E. coli* gene is available for many experimental conditions[36]. The trend of genes with high expression levels to evolve more slowly has also been reported in other statistical and theoretical studies in evolutionary biology[19], and this trend is testable using my system. The findings of the present study will contribute to such research approaches by enabling rapid evolution experiments.

5.2.2 Genome reduction driven by random mutagenesis

As discussed in Chapter 4, inactivating mutations accumulated almost linearly over time, and growth rates did not decrease for most lineages. If the experiment had been continued further, it may have resulted in functional reduction of the *E. coli* genome through experimental evolution. As mentioned earlier, *E. coli* is a model organism with about 4000 genes[1], and has mutation repair mechanisms to maintain gene functions. Functional reduction of the *E. coli* genome would allow detailed observation of the changes that occur to gene functions in shrinking genetic networks, an attractive prospect that warrants future investigation.

References

1. Pertea, M., & Salzberg, S. L. (2010). Between a chicken and a grape: estimating the number of human genes. *Genome biology*, 11(5), 206.

2. Baumstark, R., Hänzelmann, S., Tsuru, S., Schaerli, Y., Francesconi, M., Mancuso, F. M., ... & Isalan, M. (2015). The propagation of perturbations in rewired bacterial gene networks. *Nature communications*, 6, 10105.

3. Wuchty, S., & Uetz, P. (2014). Protein-protein Interaction Networks of *E. coli* and *S. cerevisiae* are similar. *Scientific reports*, 4, 7187.

4. Peregrín-Alvarez, J. M., Xiong, X., Su, C., & Parkinson, J. (2009). The modular organization of protein interactions in *Escherichia coli*. PLoS computational biology, 5(10), e1000523.

5. Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), 610-618.

6. Barrick, J. E. and Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12), 827-839

7. Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences*, 109(41), E2774-E2783.

8. Couce, A., Guelfo, J. R., & Blázquez, J. (2013). Mutational spectrum drives the rise of mutator bacteria. *PLoS genetics*, 9(1), e1003167.

9. Tsuru, S., Ishizawa, Y., Shibai, A., Takahashi, Y., Motooka, D., Nakamura, S., & Yomo, T. (2015). Genomic confirmation of nutrient-dependent mutability of mutators in *Escherichia coli*. *Genes to Cells*, *20*(12), 972-981.

10. Ishizawa, Y., Ying, B. W., Tsuru, S., & Yomo, T. (2015). Nutrient-dependent growth defects and mutability of mutators in *Escherichia coli*. *Genes to Cells*, *20*(1), 68-76.

11. Goldman, R. P. and Travisano, M. (2011). Experimental evolution of ultraviolet radiation resistance in Escherichia coli. *Evolution*, 65(12):3486-3498

12. Doudney, C. O., & Rinaldi, C. N. (1984). Modification of UV-induced mutation frequency and

cell survival of *Escherichia coli* B/r WP2 trpE65 by treatment before irradiation. *Journal of bacteriology*, *160*(1), 233-238.

13. Krishna, S., Maslov, S., & Sneppen, K. (2007). UV-induced mutagenesis in *Escherichia coli* SOS response: a quantitative model. *PLoS computational biology*, *3*(3), e41.

14. Vermeulen, N., Keeler, W. J., Nandakumar, K., & Leung, K. T. (2008). The bactericidal effect of ultraviolet and visible light on *Escherichia coli*. *Biotechnology and bioengineering*, *99*(3), 550-556.

15. Setlow, P. (2001). Resistance of spores of Bacillus species to ultraviolet light. *Environmental and molecular mutagenesis*, *38*(2-3), 97-104.

16. Pósfai, G., Plunkett, G., Fehér, T., Frisch, D., Keil, G. M., Umenhoffer, K., ... & Burland, V. (2006). Emergent properties of reduced-genome *Escherichia coli*. *Science*, *312*(5776), 1044-1046.

17. Drake, J. W., Charlesworth, B., Charlesworth, D., & Crow, J. F. (1998). Rates of spontaneous mutation. *Genetics*, *148*(4), 1667-1686.

18. Zhang, J., & Yang, J. R. (2015). Determinants of the rate of protein sequence evolution. *Nature Reviews Genetics*, *16*(7), 409-420.

19. Hashimoto, M., Ichimura, T., Mizoguchi, H., Tanaka, K., Fujimitsu, K., Keyamura, K., ... & Katayama, T. (2005). Cell size and nucleoid organization of engineered *Escherichia coli* cells with a reduced genome. *Molecular microbiology*, *55*(1), 137-149.

20. Falcon, S., & Gentleman, R. (2006). Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*(2), 257-258.

21. Storey, J. D., & Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, *100*(16), 9440-9445.

22. Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.

23. Ewing, B., Hillier, L., Wendl, M. C., & Green, P. (1998). Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. *Genome research*, *8*(3), 175-185.

24. Cock, P. J., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2009). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic acids research*, *38*(6), 1767-1771.

25. Kashiwagi, A., Sakurai, T., Tsuru, S., Ying, B. W., Mori, K., & Yomo, T. (2009). Construction of *Escherichia coli* gene expression level perturbation collection. *Metabolic engineering*, *11*(1), 56-63.

26. Toprak, E., Veres, A., Michel, J. B., Chait, R., Hartl, D. L., & Kishony, R. (2012). Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature genetics*, *44*(1), 101-105.

27. Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., & Kishony, R. (2013). Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nature protocols*, *8*(3), 555-567.

28. Funchain, P., Yeung, A., Stewart, J. L., Lin, R., Slupska, M. M., & 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*(3), 959-970.

29. Trindade, S., Perfeito, L., & Gordo, I. (2010). Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *365*(1544), 1177-1186.

30. Couce, A., Guelfo, J. R., & Blázquez, J. (2013). Mutational spectrum drives the rise of mutator bacteria. *PLoS genetics*, *9*(1), e1003167.

31. Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences*, *109*(41), E2774-E2783.

32. Choe, D., Cho, S., Kim, S. C., & Cho, B. K. (2016). Minimal genome: worthwhile or worthless efforts toward being smaller?. *Biotechnology journal*, *11*(2), 199-211.

33. Mizoguchi, H., Mori, H., & Fujio, T. (2007). *Escherichia coli* minimum genome factory. *Biotechnology and Applied Biochemistry*, *46*(3), 157-167.

34. Mushegian, A. (1999). The minimal genome concept. *Current opinion in genetics & development*, *9*(6), 709-714.

35. Itaya, M. (1995). An estimation of minimal genome size required for life. *FEBS letters*, *362*(3), 257-260.

36. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, *30*(1), 207-210.

37. Ikehata, H., & Ono, T. (2011). The mechanisms of UV mutagenesis. *Journal of radiation research*, 52(2), 115-125.

38. Mundhada, H., Schneider, K., Christensen, H. B., & Nielsen, A. T. (2016). Engineering of high yield production of L-serine in *Escherichia coli*. *Biotechnology and bioengineering*, 113(4), 807-816.

39. Miller, J. H. (1985). Mutagenic specificity of ultraviolet light. *Journal of molecular biology*, 182(1), 45-65.

40. Lin, K., & Wang, A. (2001). UV mutagenesis in *Escherichia coli* K-12: Cell survival and mutation frequency of the chromosomal genes lacZ, rpoB, ompF, and ampA. *J Exp Microbiol Immunol*, 1, 32-46.

41. Schaaper, R. M. (1988). Mechanisms of mutagenesis in the *Escherichia coli* mutator mutD5: role of DNA mismatch repair. *Proceedings of the National Academy of Sciences*, 85(21), 8126-8130.

42. Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., Denamur, E., & Elion, J. (1997). Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science*, 277(5333), 1833-1834.

43. LeClerc, J. E., Li, B., Payne, W. L., & Cebula, T. A. (1996). High mutation frequencies among Escherichia coli and Salmonella pathogens. *Science*, 274(5290), 1208-1211.

44. Gross, M. D., & Siegel, E. C. (1981). Incidence of mutator strains in Escherichia coli and coliforms in nature. *Mutation Research Letters*, 91(2), 107-110.

45. Oliver, A., Cantón, R., Campo, P., Baquero, F., & Blázquez, J. (2000). High frequency of hypermutable Pseudomonas aeruginosa in cystic fibrosis lung infection. *Science*, 288(5469), 1251-1253.

46. Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., Gouyon, P. H., & Godelle, B. (1997).

Role of mutator alleles in adaptive evolution. *Nature*, 387(6634), 700-702.

47. Giraud, A., Matic, I., Tenaillon, O., Clara, A., Radman, M., Fons, M., & Taddei, F. (2001). Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science*, 291(5513), 2606-2608.

48. Viklund, J., Ettema, T. J., & Andersson, S. G. (2011). Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade. *Molecular biology and evolution*, 29(2), 599-615.

49. Juhas, M., Eberl, L., & Glass, J. I. (2011). Essence of life: essential genes of minimal genomes. *Trends in cell biology*, 21(10), 562-568.

50. Fontanesi, L., Beretti, F., Riggio, V., Dall'Olio, S., González, E. G., Finocchiaro, R., Davoli, R., Russo, V., & Portolano, B. (2009). Missense and nonsense mutations in melanocortin 1 receptor (MC1R) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. BMC genetics, 10(1), 47.

51. Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6), 1315-1341.

Acknowledgements

本研究を遂行するにあたって、そのすべてにわたりご指導をいただきました、指導教員である現東京大学の津留三良先生に心より深く感謝申し上げます。日々の研究におけるサポート、研究へのご助言、ご指導を頂きました市橋伯一先生に深く感謝申し上げます。同時に、研究環境や研究全体に対する貴重なご意見を頂いた、松田史生先生にも深く感謝申し上げます。

研究遂行のためのサポートや研究内容に対するディスカッションなど様々なご指導を賜りました代謝情報工学講座教授の清水浩先生に深く感謝いたします。情報科学の観点から研究全般に関するご意見を頂きました、人間情報工学講座教授の前田太郎先生、バイオシステム解析学講座教授の若宮直紀先生、ゲノム情報工学講座教授の松田秀雄先生に深く感謝いたします。

また、ゲノム配列解析において多大なご協力をいただいた大阪大学微生物病学研究所の中村昇太特任准教授、元岡大祐特任助教、理化学研究所の古澤力先生、堀之内貴明博士に厚く御礼申し上げます。生物情報データの扱いについて大変丁寧にご指導いただいたゲノム情報工学講座の瀬尾茂人先生に心からお礼申し上げます。

Imperial College London 教授の Mark Isalan 先生には共同研究としてイギリス滞在中に懇切丁寧にご指導賜りました。心から感謝申し上げます。

角南武志先生、辻岳志先生をはじめとする共生ネットワークデザイン学講座の皆様には、研究全般にわたってご指導頂くとともに、有益なディスカッションを頂きました。心から感謝申し上げます。特に、日々の研究活動をともに過ごした細胞チームの松本悠希博士、石澤裕佳博士、村上由衣博士、明野優也博士、高野壮太朗博士、山和馬氏、清水天馬氏、高橋佑輔氏、高見梨沙氏、東雄貴氏、伊藤忠弘氏、小森くん、柿園絵里氏、齊藤紘美氏には心から感謝しております。また一居哲夫博士、吉山先生には実験装置開発などにおいて多くのご支援を賜りました。御礼申し上げます。また実験科学や語学の面でご指導をいただいた Germond Arnaud 博士、組み換え株の作成において多大なるご助力を賜った津留奈津子氏に心より御礼申し上げます。

また，研究から生活まで多大なるご支援をいただいているヒューマンウェアイノベーション博士課程プログラムに携わっておられる方々、特に細田一史先生に厚く御礼申し上げます。またアドバイザリ委員として多岐にわたるご助言を賜りました大阪大学の倉光成紀先

生、和田成生先生、そして日本電信電話株式会社の芳賀恒之先生、瀬山倫子先生、岡田顕先生に厚く御礼申し上げます。

そして工学系の研究者としての基礎を授けていただいたにもかかわらず、生物系に移る時にも快く送り出していただいた奈良工業高等専門学校の飯田賢一教授に心から深く感謝いたします。またその進学に伴う分野替えにあたり、専門書の紹介や研究の方向性についての議論など様々なご支援を頂いた、奈良工業高等専門学校の堀内泰男先生、榊原和彦教授、新野康彦准教授にも厚く御礼申し上げます。また、寮食堂の店長として私に健康な心身を授けていただいた、現在舞鶴高専にいらっしゃる閏井康之氏に感謝いたします。

また本研究は文部科学省の日本学術振興会の研究者養成事業ならびに博士課程教育リーディングプログラムの支援を受けて遂行されました。感謝申し上げます。