# HdfR/MaoP pathway of *Escherichia coli* regulates flagella expression and growth in aquatic environment

DISSERTATION

zur Erlangung des Grades eines Doktors
der Naturwissenschaften (Dr. rer.nat.)

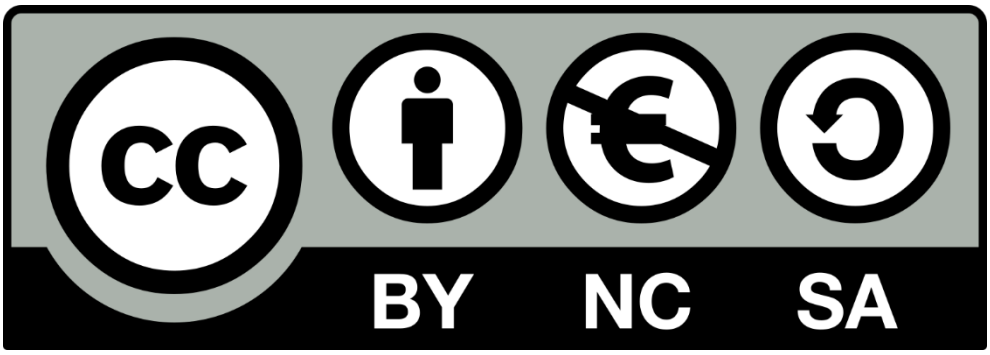
des Fachbereichs Biologie der Philipps-
Universität Marburg

Vorgelegt von

**Nataliya Teteneva**

aus Murmansk, Russland

Marburg, 2023



Die vorliegende Dissertation wurde von Oktober 2018 bis Mai 2023 am Max-Planck-Institut für Terrestrische Mikrobiologie unter Leitung von Prof. Dr. Victor Sourjik angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am **24.08.2023**

Erstgutachter: Prof. Dr. Victor Sourjik

Zweitgutachter: Prof. Dr. Robert Junker

Tag der Disputation: **20.09.2023**

Teteneva NA, Mart'yanov SV, Esteban-Lopez M, Kahnt J, Glatter T, Netrusov AI, Plakunov VK & Sourjik V (2020) Multiple Drug-Induced Stress Responses Inhibit Formation of *Escherichia coli* Biofilms. *Appl Environ Microbiol* **86**.

*The data may not contain the answer. The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data.*

*- John Tukey*

## *Acknowledgements*

First, I cordially thank Prof. Victor Sourjik for giving me a chance to work in his group, for the interesting research idea, and for his undying optimism. Without him and his guidance, I would not be able to go through this journey. I would also like to thank the members of my TAC, Prof. Hannes Link and Prof. Knut Drescher for their advice and discussion.

I am very thankful to all the current and former members of the research group for nice friendly atmosphere and amazing working experience. I thank Dr. Jing Yuan, Dr. Seán Murray for fruitful discussions and endless support. I am deeply grateful to Ismath Sadhir for his strains and his time and efforts dedicated to the microfluidics experiment and data analysis. I hug Ananda Medeiros for constructing the transposon mutant library, and for always laughing at my jokes. I thank Camila Cilveti for reading the final version of this thesis and correcting it. I thank Isa Pérez López, Ira Lisevich, Ina Biazruchka, Giovanni Scarinci and all the others for all the good time we spent together as a team. And of course, a very special thanks goes to our wonderful technicians Melissa, Sarah, Basti, Claudia and Inka.

I am immensely grateful to the entire Proteomics Facility of the MPI. The help and support of Dr. Timo Glatter and Jörg Kahnt were vast and extremely important for me and my project. I also thank Gabriele Malengo and Silvia Gonzáles Sierra for teaching me how to not be afraid of the complicated machinery.

I thank the EMBL GeneCore facility and the NGS Core facility of the MPI for Biochemistry for doing the DNA sequencing for us. I also thank GWDG for providing me the access to their high performance computational cluster.

Last but not least, I thank my family for always believing in me even against all odds, and their endless love and support.

## *List of abbreviations*

(p)ppGpp – guanosine penta- or tetraphospate

BioID – biotinylation-based proximity labelling

c-di-GMP – bis-(3′,5′)-cyclic dimeric guanosine monophosphate

CFU – colony forming unit

ChIP-seq – chromatin immunoprecipitation sequencing

Co-IP – co-immunoprecipitation

CV – crystal violet

DGC - diguanylate cyclase

ECA - enterobacterial common antigen

eGFP – enhanced green fluorescent protein

EPM – extracellular polymer matrix

FB – fluorescence-based screen

GASP – growth advantage in the stationary phase

IPTG – isopropyl- β-D-thiogalactopyranosid

LB – Luria-Bertrani medium

NER – nucleotide excision repair

PBS – phosphate buffered saline

PC – principal component

PCA – principal component analysis

PDE – phosphodiesterase

POI – protein of interest

qPCR – quantitative polymerase chain reaction

RNAP – RNA polymerase, core unit

sfGFP – superfolder green fluorescent protein

SOB – Super Optimal Broth

TB – tryptone broth

TBS – Tris buffered saline

TnSeq – transposon insertion sequencing

WT – wild type strain

## *Zusammenfassung*

Der Lebenszyklus von *Escherichia coli* ist zweiphasig. Die primäre Umgebung dieses Organismus ist der untere Darmtrakt von Warmblütern und Reptilien. Sobald dieses Bakterium von einem tierischen Wirt ausgeschieden wird, gelangt es in seinen sekundären Lebensraum: Die äußere Umgebung wie Wasser, Sand oder Erde. Früher wurde davon ausgegangen, dass die lebensfähigen Darmbakterien außerhalb des Darms nicht lange überleben können, wobei das Vorhandensein von *E. coli* im Boden oder Wasser lediglich auf eine kürzlich erfolgte Fäkalienkontamination zurückzuführen ist. Es gibt jedoch zunehmend Hinweise darauf, dass trotz aller Umweltstressfaktoren zumindest einige *E. coli*-Stämme über längere Zeiträume in der sekundären Umgebung überleben und sich vermehren können.

Während sich frühere Forschungen hauptsächlich auf externe Faktoren konzentrierten, die das Überleben von *E. coli* außerhalb des Wirts beeinflussen, wie Nährstoffknappheit, osmotische Verschiebungen, Prädation usw., spielen bekanntermaßen auch genetische Determinanten wie der Stressreaktionsweg eine wichtige Rolle. In dieser Arbeit wollten wir anhand des Seewasser-Modells systematisch Gene identifizieren, die für das Überleben von *E. coli* in seiner sekundären Umgebung wichtig sind. Unsere Ergebnisse belegen und erweitern die Bedeutung und den Dualismus des Stressreaktionswegs für das Wachstum von *E. coli* außerhalb des Darms, mit dem überraschenden Ergebnis, dass die Mutationen im allgemeinen Stressreaktions-Sigmafaktor $\sigma^S$ für das Wachstum im Seewasser von Vorteil sind. Darüber hinaus haben wir gezeigt, dass Defekte in der Membranintegrität schädlich für das Wachstum sind. Außerdem identifizierten wir Mutationen in zwei benachbarten, kaum charakterisierten Genen, *hdfR* und *maoP*, die sich positiv auf das Wachstum *von E. coli* im Seewasser auswirken. Wir haben gezeigt, dass die Mutationen in *hdfR* $\sigma^S$-abhängige Funktionen, wie z. B. die Biofilmbildung signifikant hemmen.

Die anschließende Analyse des HdfR- und MaoP-Paares zeigte, dass diese ein neuartiges Regulierungssystem bilden, das mehrere zelluläre Funktionen steuert, darunter Flagellenexpression, Plasmidkopienzahl und Chromosomenorganisation. Obwohl HdfR zuvor als negativer Transkriptionsregulator des Flagellen-Masteroperons *flhDC* beschrieben wurde, konnten wir zeigen, dass seine Wirkung auf die *flhDC*-Expression nicht direkt ist, sondern vielmehr durch MaoP vermittelt wird. MaoP wird selbst durch HdfR transkriptionell reguliert. Darüber hinaus beeinflusst dieses System die Kopienzahl des ColE1-Plasmids und die Organisation des *E. coli*-Chromosoms.

## *Summary*

The lifecycle of *Escherichia coli* is biphasic. The primary environment of this organism is the lower intestinal tract of warm-blooded animals and reptiles. Once excreted from an animal host, this bacterium enters its secondary habitat: external environment, such as water, sand or soil. It was traditionally assumed that the viable enteric bacteria cannot last long extraintestinally, with the presence of *E. coli* in soil or water being only a result of a recent fecal contamination. However, there is increasing evidence that despite all environmental stress factors at least some strains of *E. coli* can survive and reproduce in the secondary environment for prolonged periods of time.

While previous research mostly focused on external factors influencing *E. coli* persistence outside the host, such as nutrient scarcity, osmotic shifts, predation *etc*, genetic determinants such as the stress response pathway are also known to play an important role. In this work we aimed to systematically identify other genes crucial for *E. coli* to survive in its secondary environment, using the lake water as a model. Our results demonstrate and expand the importance and dualism of stress response pathway for *E. coli* growth outside the host, with the surprising finding that the mutations in the general stress response sigma factor $\sigma^S$ being beneficial for growth in the lake water. On top of this, we demonstrated that the defects in membrane integrity are detrimental for growth. Finally, we identified mutations in two neighboring poorly characterized genes, *hdfR* and *maoP*, that are beneficial for *E. coli* growth in the lake water. We demonstrated that the mutations in *hdfR* significantly inhibit $\sigma^S$-dependent functions, such as biofilm formation.

Subsequent analysis of the HdfR and MaoP pair showed that they form a novel regulatory system that controls several cellular functions, including flagella expression, plasmid copy number and chromosome organization. HdfR was previously described as a negative transcriptional regulator of flagellar master operon *flhDC*, but we demonstrated that its effect on *flhDC* expression is not direct but rather mediated by MaoP that is itself transcriptionally regulated by HdfR. Besides, this system affects the copy number of ColE1 plasmid and influences the organization of *E. coli* chromosome.

## *Table of contents*

# Chapter I. Introduction

## 1.1. *Escherichia coli* as an extraintestinal bacterium

In the year 1885, the German physician Theodor Escherich [1] described a rod-shaped bacterium he had previously isolated from feces. He did not know yet that this organism would become one of the most studied bacteria on Earth and immortalize his name.

*Escherichia coli* is a Gram-negative, facultative anaerobic non-sporulating coliform bacterium commonly found in the lower intestine of warm-blooded animals [2, 3], making up to 0.1% of intestinal microbiota [4]. It is a mostly commensal bacterium although the pathogenic variants do exist [5-7] and pose a great concern for the public health. As a prominent member of a gut microbiota, it is widely used as a marker of a recent fecal contamination in soil and both drinking and recreational water [8]. Due to its ease of handling and fast doubling, it is used routinely in molecular biology and microbiology both as a model organism and a tool to manipulate the other bacteria. On top of this, recombinant *E. coli* strains are widely used in biotechnology to mass-produce biopharmaceuticals such as human insulins, interferons, and many more [9].

The lifecycle of *E coli* is biphasic [10, 11] and consists of primary and secondary habitats. The primary habitat of this bacterium is the lower intestinal tract of warm-blooded animals and reptiles [2]. Once excreted from an animal host, *E. coli* enters its secondary habitat: external environment (**Figure 1**). Approximately one half of a total *E. coli* biomass is considered to reside in the external environment [12], but this number of bacteria is thought to be maintained due to the constant transfer of the bacterial matter from human and animal hosts. While *E. coli* is thoroughly studied under laboratory conditions and inside the large intestine, relatively little is known about the ecophysiology of its populations (and their survival) in the outer environment [10].

In its secondary environment, *E. coli* finds itself battling for survival, facing limited nutrient availability, osmotic shifts, large variations in temperature and pH, UV radiation, predation and competition for the scarce resources. It is traditionally assumed that the viable enteric bacteria cannot last long extraintestinally, and the presence of *E. coli* in soil or water can only be a result of a recent fecal contamination.

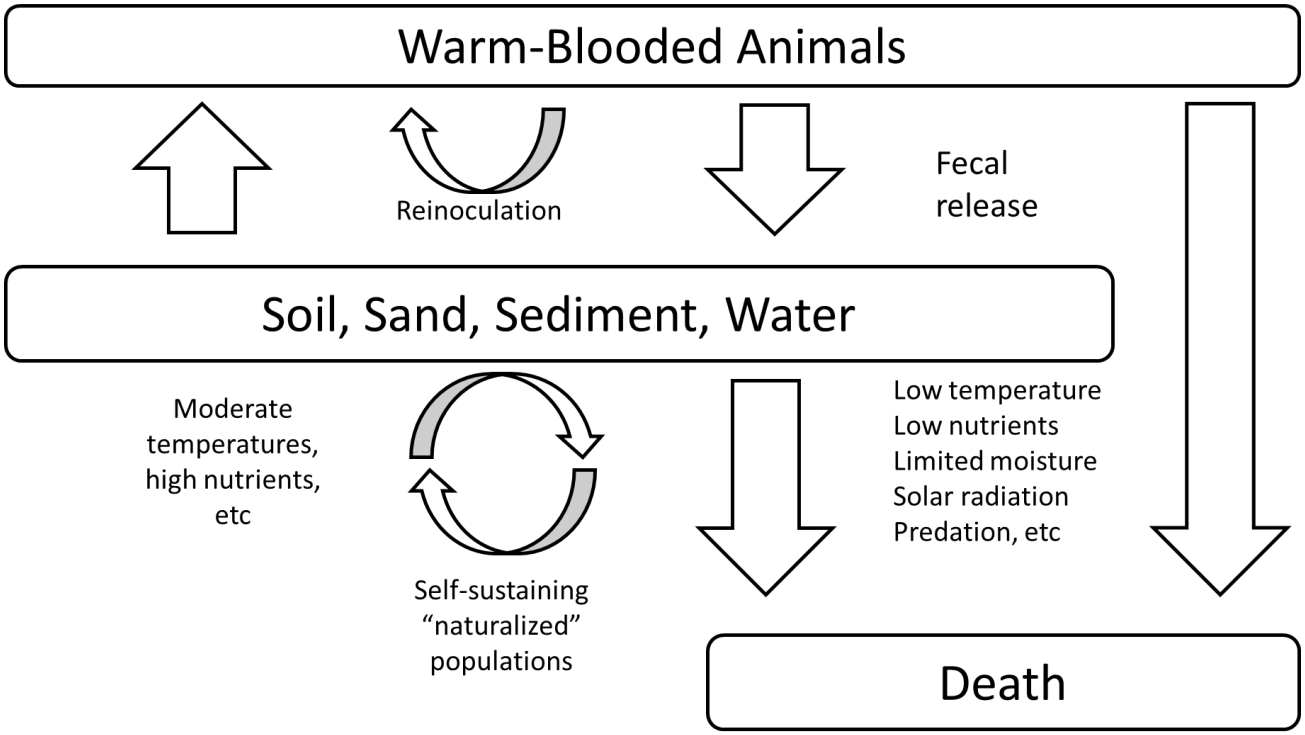



**Figure 1. The lifecycle of *E. coli.*** Adapted from [11].

However, there is an increasing number of evidences that despite all these factors at least some strains of *E. coli* are robust enough to survive outside of the host for prolonged periods of time and reproduce in soil, sand, seawater, sediment, growing plants, soil invertebrates and other environments [10, 11, 13-18]. Some strains are reported to be capable to form long-lasting naturalized populations [11, 15, 19] that apparently have been thriving in their respective environment for months and even years. Moreover, many tropical soil and freshwater samples contained a significant number of viable coliforms including *E. coli* despite absence of known fecal source [12]. For example, this study from Puerto Rico demonstrated presence of *E. coli* in bromeliad water [20] in a pristine rainforest. The authors decline the hypothesis that these bacteria come from a recent fecal contamination, since the studied bromelia plants were elevated up to 6 meters above the ground, and Puerto Rico has a relatively low population of birds and tree-climbing mammals who might have contaminated these plants. In tropical climates, the temperature and humidity are constantly high and the nutrients are usually quite abundant, thus enabling *E. coli* to maintain its population. However, there are reports about long-lasting *E. coli* populations from colder habitats, such as the shores of the Great Lakes [19], soil in Ireland [13] or pastures of French Alps [21].

Soil environments are essential reservoirs for various antibiotic resistance genes [22]. Sub-inhibitory concentrations of antibiotics raise the background rates of mutation, recombination and lateral gene transfer among microbial populations. These effects indicate that previously susceptible soil microbes can acquire antibiotic resistance via mutation or lateral gene transfer from co-polluting resistant bacteria [23]. Previous studies have reported multidrug-resistant *E. coli* strains to be found in the environment [15]. Therefore, *E. coli* might be both a donor and an acceptor of the antibiotic resistance genes, and both of these scenarios are concerning for the public health.

Since *E. coli* is associated with many foodborne infection outbreaks worldwide [24-26] and is widely used as a marker of a recent fecal contamination [8], it is crucial to be able to track its survival outside of the intestine. However, the fact that *E. coli* is able to reproduce outside the gut and especially the discovery of the naturalized populations seriously challenge the feasibility to use it to determine recent pollution. For example, *Salmonella enterica,* another common intestinal bacterium causing many cases of gastrointestinal diseases worldwide [27], is shown to persist a wider variety of environmental stresses than *E. coli,* and is more likely to be transmitted between the hosts [2]. In light of these data, neither presence of *E. coli* in a water or soil sample does not actually mean a recent fecal contamination, nor its absence guarantees safety.

## 1.2. External factors influencing *Escherichia coli* survival

Both growth and death rates are determined by the environmental conditions at the local scale and by the ability of the microorganism to cope with these local conditions [10]. We will briefly discuss the most influential conditions in this chapter.

### 1.2.1. Abiotic factors

Temperature is probably one of the most important factors influencing *E. coli* survival and growth. In the intestinal tract the temperature ranges between 36–40°C, and temperature in external environment is usually low (<30°C). However, the optimal temperature for survival is not necessarily the same as the one for growth. *E. coli* has the ability to survive long-term under temperature lower than 30°C [14, 19]. Both division and death rates are decreased if the temperature is lower [10], and most reports indicate enhanced stability of *E. coli* at lower temperatures [18]. One of the potential factors enhancing its population might be that the lower temperatures favor biofilm formation by *E. coli* [28-30].

Additionally, temperature in the open environments tends to fluctuate in accordance to both circadian and annual cycles, as opposed to maintained and strictly regulated temperature of the host. Some tropical soil-borne strains adapted to temperature fluctuations, demonstrating better growth at fluctuating temperatures than under constant warmth [14]. Some soil strains isolated in colder climate (shores of Lake Superior) were even shown to withstand repeated freeze-thaw cycles [19].

Desiccation is one of the most common stresses to bacteria in external environments [31]. Growth of *E. coli* in the soil environment was negatively influenced by soil desiccation; while *E. coli* survival rates were not different between dried and wet soils [14]. Upon rehydration, *E. coli* population in the dried soil was resuscitated, indicating that water availability is limiting for *E. coli* to grow but not to survive.

Growth and survival of *E. coli* in open environments are often restricted by the availability of nutrients and energy sources. In a growing culture, starvation will ensue at a given moment due to a limitation of particular carbon or other substrates [10]. External environments are generally low in available nutrients, especially compared to the intestinal tract. However, *E. coli* is quite versatile in energy acquisition, as it can use oxygen, nitrate, nitrite and fumarate as the acceptors in its respiratory chain, that allows timely response to environmental changes [32]. Under the low-nutrient conditions, different alternative catabolic functions and binding proteins will become de-repressed [10]. *E. coli* can consume a wide range of carbon sources, including sugars [33], amino acids [34], short chain fatty acids [35] and other compounds. Such versatility and flexibility in energy and carbon acquisition should help *E. coli* survive and grow in the environment [11, 36].

When released into the sea, enteric bacteria are subjected to an immediate osmotic upshift. It was demonstrated that both distilled water and undiluted seawater were detrimental for *E. coli* survival [18, 37]. The optimal salinity was 25% seawater which approximately corresponds to 0.9% salt. Similar results were obtained with NaCl solution instead of seawater [37].

The survival of *E. coli* in soil was shown to be related to the local pH, and in particular soil acidity was detrimental for most of the strains [15]. However, enterohemorrhagic *E. coli* serotype O157:H7 was shown to survive extreme acid exposure (pH 3) better than its non-O157 counterparts [38]. In seawater, slightly acidic environment (pH 5) was found to be most

favorable for *E. coli* survival [18, 37], although the pH of seawater is usually slightly basic (pH 7.5-8.5)

Solar radiation is an often overlooked but very efficient abiotic factor causing death of enteric bacteria in environmental waters. In *E. coli*, sunlight causes a rapid decrease of the colony forming ability [18]. In seawater, 90% of the fecal coliforms were inactivated within 30 to 90 minutes of sunlight exposure [39]. The most lethal portion of the sunlight is UV-B (280-320 nm) that directly causes DNA damage and oxidation of cellular contents [40], as well as cause it indirectly via generation of reactive oxygen species. The effect of sunlight on the *E. coli* survival may vary by insolation time, cloud cover, season, turbidity of the water environment, presence of autotrophic bacteria or protists, latitude, depth of water body, type of its bottom and potentially some other factors.

### 1.2.2. Biotic factors

*Escherichia coli* is never alone in its both primary and secondary habitats. It can be predated by protozoa and lysed by phages [41], and needs to compete with indigenous microorganisms for limited nutrient sources [15]. Some organisms could be direct competitors by occupying the same niche, whereas others might have antagonistic or predatory activities.

In the marine environment, the primary predators are protozoa [18, 42]. It was shown that bacterial competition, antagonism, and bacterial predation were relatively unimportant in coliform removal [18]. However, the competition between *E. coli* and the native marine bacteria still occurs, and *E. coli* was found to be a successful competitor in rich media but a very poor one under the low nutrient concentrations characterizing natural seawater [43]. Bacteriophages were effective in reducing *E. coli* population sizes only under nutrient-rich conditions, suggesting a very minor role for bacteriophages under external conditions [44].

In soil environment, the presence of autochthonic microbiota is important as well. *Escherichia coli* populations grew better in sterile *vs* nonsterile soils, highlighting that the autochthonic microbiota has a crucial effect on *E. coli* survival [14]. When introduced into soil microbial communities of varying diversity, *E. coli* survived worse in more diverse communities, suggesting that complex communities are more difficult to invade [10]. Conversely, there is a chance that *E. coli* would integrate into existing microbial community and establish a cross-feeding relationship with the autochthonic microbiota. For instance, cellulose- and lignin-degrading organisms might provide it with easily available carbon

sources. Another interesting aspect is that sometimes the bacterial cells might establish a relationship with the small planktonic crustaceans, using them as a transitional host [45].

In *E. coli* survival research, most of the studies focus on the external factors listed above. In this work, however, we would like to focus on genetic variability of *E. coli* and genetic determinants of its survival.

## 1.2. Protective role of the biofilm formation

The well-known persistence factor of bacteria is biofilm formation. Within the mature biofilm, bacteria are more protected from environmental conditions [46]. It was demonstrated, that the *E. coli* strains isolated from soil form thicker biofilms than the laboratory strains both in rich and minimal medium, and both at low and high temperature [47].

Biofilms are microbial communities growing at the interphase and submerged into self-produced extracellular polymer matrix (EPM) [48, 49]. Mature biofilms protect the bacteria from hostile environmental conditions such as UV radiation, desiccation, protozoan predators, and chemicals including antibiotics and disinfectants [46], and there is a multitude of factors involved in this process.

### 1.2.1. Extracellular polymer matrix: functions and composition

In *E. coli,* the extracellular polymer matrix (EPM) mostly consists of amyloid fibers named curli. They are encoded by the *csgBA* operon and regulated by the master regulator *csgD*, which is in turn under the regulation of the stress response σ factor RpoS [50, 51].

The other components are type I fimbria, antigen 43, poly-β-1,6-N-acetyl-D-glucosamine, cellulose and colanic acid [51]. They are the minor components of the biofilm and their content greatly varies depending on strain and growth conditions [51, 52].

The EPM offers a constantly hydrated viscous layer that physically protects the embedded cells from desiccation, host immune system, protozoan grazers and phage infection [53]. It also plays a significant protective role as a diffusion barrier for both toxic molecules (by preventing them from entering the mature biofilm) and useful molecules such as nutrients, signal molecules or enzymes (by preventing their escape from the biofilm and maintaining the local favorable concentrations of them) [53].

In addition to its protective role, one of the major functions of the matrix is a structural one. The adhesive properties of the matrix enable the bacteria to remain bound to the surface and in the close proximity to each other. These conditions are very favorable for establishing the horizontal gene transfer [54] and establishing the metabolic interactions, such as cross-feeding, between the members of the community. Due to the spacial organization provided by the matrix, the chemical gradients are established, allowing the formation of small ecological niches within the biofilm [55].

Due to the yet unknown regulatory cascades, the expression of the curli master regulator *csgD* is induced at lower temperature (<30°C) [28-30]. This further highlights the importance of biofilm formation by *E. coli* in its secondary environments and not in the primary one. All these aspects of the putative roles of the matrix could contribute to *E. coli* survival especially in its secondary environments.

## 1.3. Genetic and phenotypic variability within *Escherichia coli* species

*Escherichia coli* is a highly diverse species, both phenotypically and genotypically. Its core genome is only about 2000 genes [10, 56]. The pan genome, however, is much larger and is estimated to be up to 55 000 genes due to many strain- or group-specific genes [57]. An average genome size in *E. coli* strains is estimated at about 4700 genes [10]. Such a variability within the species allows it to populate various environments. The average *E. coli* genome is shaped by a multitude of evolutionary forces derived both from its primary and secondary habitats, in which both biotic and abiotic pressures are present.

### 1.3.1. Phylogenetic groups of *Escherichia coli*

Most of *Escherichia coli* strains can be divided into eight main phylogenetic groups (**Figure 2**) named A, B1, B2, C, D, E, F and G [58, 59]. To differentiate between the phylogenetic groups, the triple and quadruple Clermont PCR assays were developed [60-62]. Phylogroup determination is important because the strains from different phylogroups can be attributed to the different specific hosts and lifestyles. Even the genome size differs among the phylogroups, with A and B1 strains having smaller genomes than B2 or D strains [63].

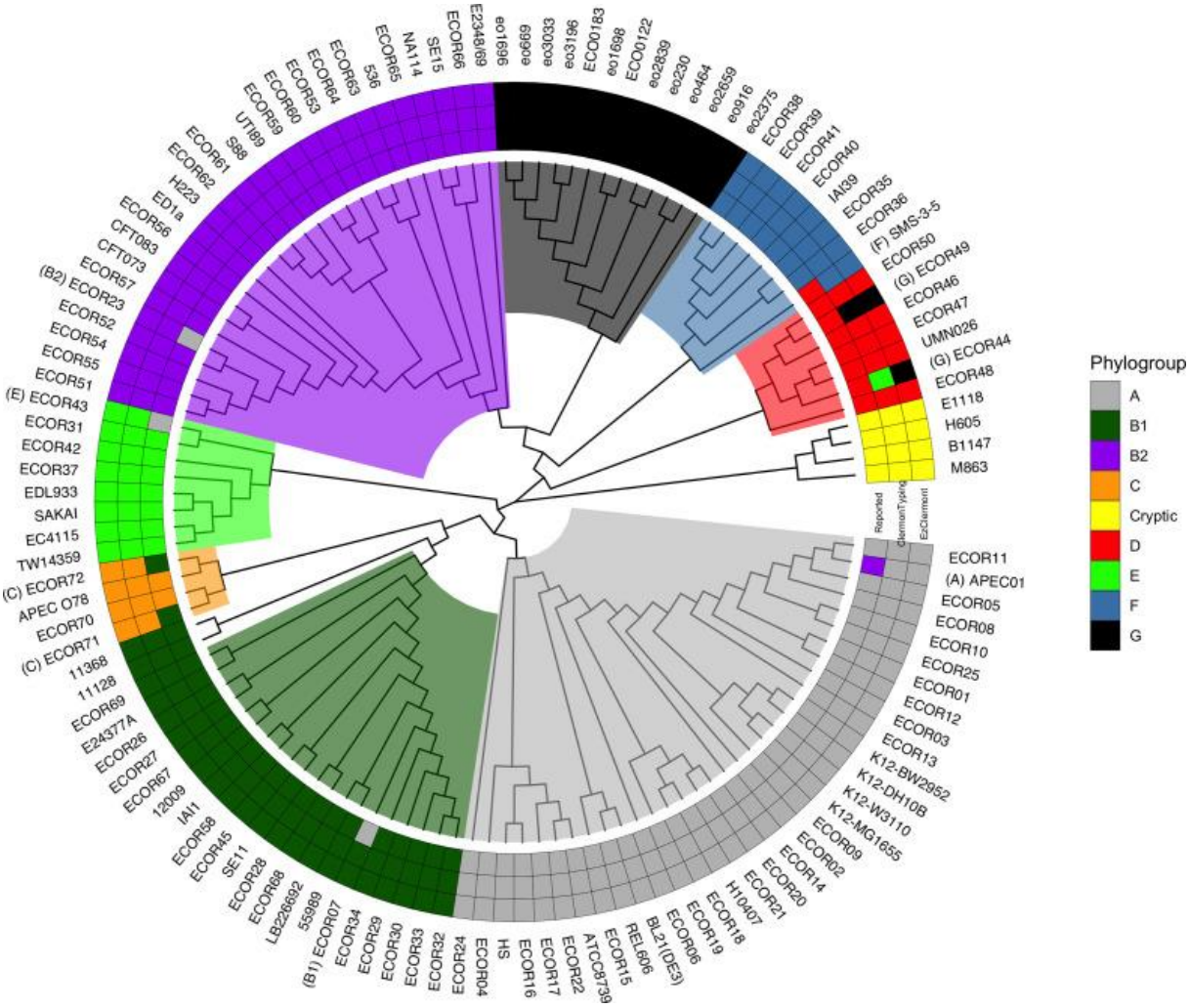



**Figure 2. Phylogenetic groups of *E. coli*.** Adapted from [58].

The distribution of phylogroups among *E. coli* sources is nonrandom. There is a difference between its primary and secondary environments. It was shown that in human gut (primary habitat of *E. coli*) A is usually the most abundant group, followed by B2. In the farm animals, B1 is the most abundant group [3]. On the opposite side, B1 phylogroup is more likely to be isolated from soils and sands (secondary habitat), followed by A and D [64-66]. Moreover, in a longitude experiment in sand, B1 phylotype has outcompeted the other phylotypes [67]. The virulent extra-intestinal strains belong mainly to the group B2 and to a lesser extent to groups D and F [6, 7]. Phylogroup E was initially associated with the enterohaemorrhagic *E. coli*, especially famous for including the O157:H7 pandemic lineage, but was later described as a highly diverse group including commensal and environmental strains [68]. The most recently described phylogroup G is associated with poultry [69], and has high virulence and antibiotic-

resistance potential. This data indicates that some phylotypes are associated with particular hosts and niches.

### 1.3.2. Genetic variability of *Escherichia coli* contributes to its ability to survive

This vast diversity of different *E. coli* strains inevitably influences its ability to survive in the secondary environment. Indeed, the strains that were isolated from the secondary environment are often quite different from their commensal counterparts. For example, the study by Fiona Brennan shows that persistent soil strains of *E. coli* are able to utilize more various carbon sources and at lower temperatures than the laboratory strains [36]. Another example is that the B1 phylogroup outcompeted the other phylogroups when grown together in the sand [67].

Environmental isolates tend to have smaller genomes and less mobile elements than commensal or pathogenic isolates independently of the phylogenetic group. They also have fewer virulence factors and antibiotic resistance genes [56]. These data confirm that many of the environmental strains did not appear because of recent fecal contamination but are rather the result of long-term adaptation to the outer environment.

In general, attempts to understand the genetic determinants of *E. coli* survival are quite rare as most of the studies focus on environmental factors discussed in the previous chapters [15].

However, it is demonstrated that mutations in the certain genes (*rpoS*, *otsA*, *relA*, *spoT*, *ompC*, *ompF*) significantly and negatively influence *E. coli* survival in seawater [18]. The most prominent effect was caused by mutations in the *rpoS* gene, which encodes a general stress response σ-factor in *E. coli*. A mutation in *rpoS* decreased the survival rate of *E. coli* cells in seawater by 3 logs over 8 days [70]. A similar effect was caused by the absence of two guanosine penta- or tetraphospate ((p)ppGpp) synthases *relA* and *spoT,* which are both under RpoS regulation and upregulate the expression of *rpoS* via (p)ppGpp synthesis.

The importance of RpoS and its regulon was further highlighted by experiments on survival of *E. coli* in soil. In the study by Somorin et al [41] it was shown that the strains of *E. coli* isolated from soil were more resistant to predation by protozoa than the commensal or laboratory strains. On top of this, isogenic Δ*rpoS* mutants of the studied strains were predated significantly faster than the wild types. The authors prove that it happens due to the diminished production of proteinaceous fibers curli by the Δ*rpoS* mutants. Earlier the same

group has shown that RpoS activity is enhanced in the soil-persistent strains [47], highlighting the importance of this regulatory pathway in the outer environments. On the opposite side, inactivation of RpoS is a frequent event in the laboratory setting [71] and some commonly used strains (i.e. *E. coli* K-12 W3110 [72]) have a non-functional RpoS.

One would expect such an important regulatory pathway to be well conserved in the soil, and there are studies that support it [47]. But on the opposite side, environmental populations of *E. coli* and *Salmonella* contain a high frequency of polymorphisms in *rpoS* [73]. For example, in a study by Chiang et al [74] the *rpoS* sequence was determined for 45 environmental isolates, and of these, six isolates were confirmed as mutants with the complete loss of RpoS function. They have also demonstrated that the *rpoS* mutations quickly arise if the bacteria are grown on poor media and/or non-preferred carbon source such as succinate [74, 75]. In the environmental conditions, lack of nutrients is a common factor that bacteria might face and they have to adapt to it.

## 1.4. RpoS and general stress response in *Escherichia coli*

Bacteria have to adapt to a variety of abiotic stresses, such as suboptimal temperature, pH, osmolarity, nutrient limitation, reactive oxygen species, toxic metal ions and predation. One strategy to counter these is to induce highly specific stress responses that deal with the stress-induced damage in a targeted way (such as DNA repair after the UV illumination). On the other hand, many different stress conditions or entry into the stationary phase induce the complex general stress response, which makes the bacteria more resistant to diverse stress conditions, and therefore the damage is prevented rather than has to be repaired [76].

RpoS is a global regulator of general stress response. It is an alternative σ factor of RNA polymerase and it regulates (directly or indirectly) approximately 10% of *E. coli* genes [77]. $\sigma^S$ is closely related to the vegetative σ factor $\sigma^{70}$, and they recognize similar but not identical promoter sequences [76]. As an adaptive regulator, $\sigma^S$ is not essential for the core metabolism but controls a wide network of genes and has the second largest regulon after $\sigma^{70}$ [78].

### 1.4.1. Regulation of RpoS

In the rapidly growing cells, the levels of $\sigma^S$ are relatively low, but it accumulates upon entering the stationary phase and/or upon exposure to the environmental stresses [79, 80]. The highly fine-tuned gradual accumulation of $\sigma^S$ during the transition from exponential to

stationary phase mostly relies on the control of $\sigma^S$ expression and combines mechanisms of transcriptional and translational regulation [76, 81]. On the opposite side, rapid accumulation of $\sigma^S$ in response to life-threatening stress conditions is mainly dependent on inhibition of $\sigma^S$ proteolysis [76, 80] (**Figure 3**).

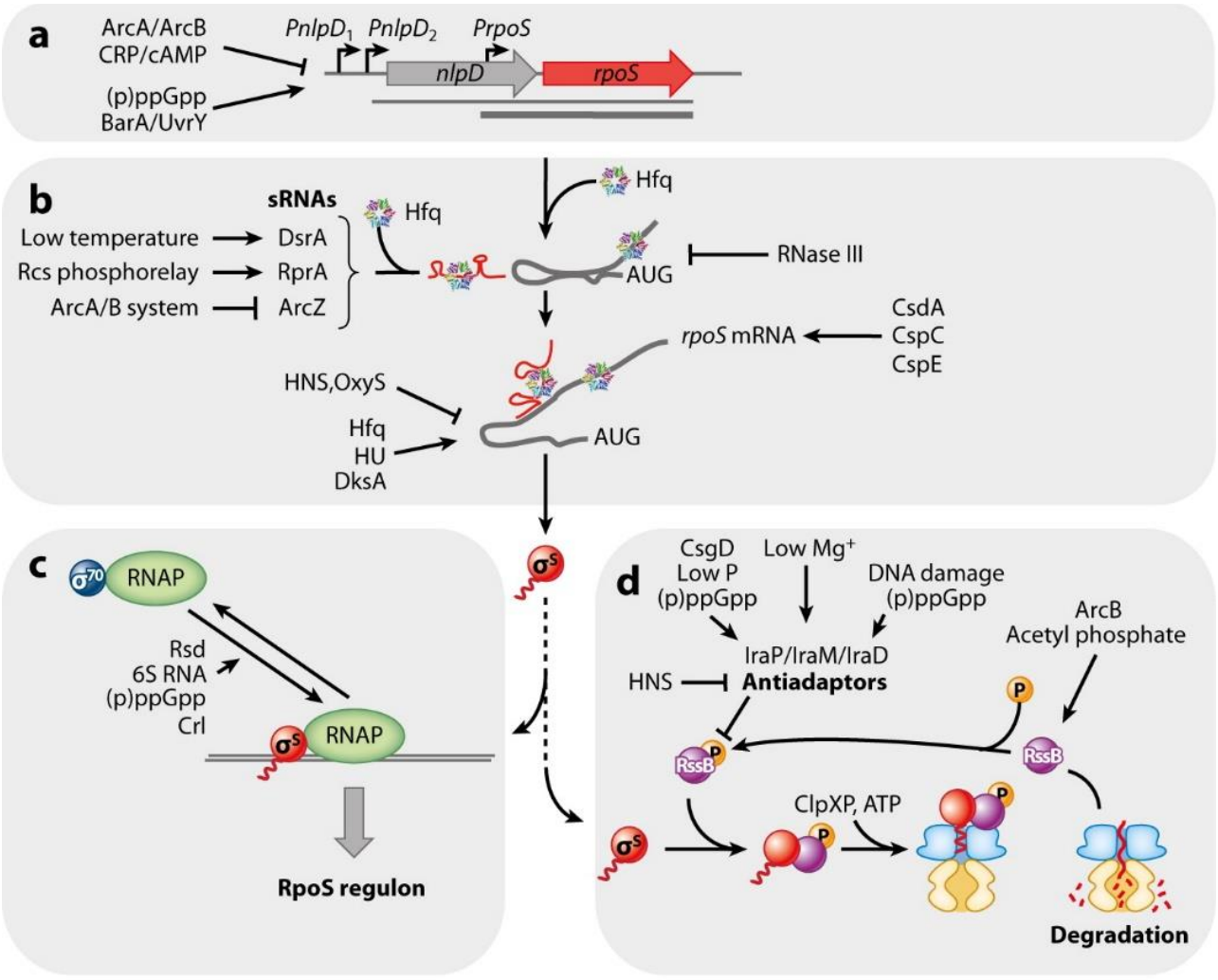



**Figure 3. Regulation of RpoS is fine-tuned and happens at the multiple levels**. The levels shown in the boxes are (A) transcriptional regulation, (B) post-transcriptional regulation, (C) σ-factor competition, (D) proteolytic regulation. See the text for the details. Adapted from [79].

At the transcriptional level, *rpoS* is regulated by the several transcriptional factors, including ArcA, Crp and BarA/UvrY two-component system [76, 79]. On top of this, a very important physiological input is provided by the alarmone (p)ppGpp which accumulates in cells exposed to starvation and other stresses [76, 82]. (p)ppGpp-less mutant Δ*relA*Δ*spoT* show strongly reduced $\sigma^S$ levels under all conditions [76, 79, 83] and the transcription activity from the *rpoS* promoter was reduced [83]. In addition, many RpoS-regulated genes also require

(p)ppGpp for efficient expression. However, regulation at the level of transcription is not dramatic compared with effects on translation and protein turnover [81].

*rpoS* mRNA has a very long (567 nucleotides) 5`- untranslated region which is important for the regulation at the translational level [79]. It folds into a stem-loop that occludes the ribosome-binding site and minimizes the translation [76, 79, 84]. To unwind this loop and initiate the translation, a small regulatory RNA and a chaperon Hfq acting as a scaffold are required [79, 84]. Four small regulatory RNAs, DsrA, RprA, ArcZ, and OxyS, have been found to affect *rpoS* translation. DsrA expression increases and stimulates *rpoS* mRNA translation at low temperature [85]. RprA is required for response to the osmotic shock [86]. ArcZ stimulates *rpoS* translation while being itself repressed by the ArcB/ArcA two-component system (which is activated by low-oxygen/high-energy supply conditions) [87]. OxyS, unlike three other RNAs, represses [88] *rpoS* translation, likely by titrating Hfq. OxyR responds to the presence of reactive oxygen species such as $H_2O_2$ [76, 79]. In summary, these small regulatory RNAs provide a connection between the specific stress responses and the general stress response governed by RpoS.

In the actively growing cells, RpoS is barely detectable due to efficient proteolytic degradation [80]. RpoS is degraded by the ATP-dependent ClpXP protease. However, the protease cannot target and cleave RpoS without it being bound to the response regulator RssB, which effectively acts as an anti-$\sigma^S$ factor [76, 80]. *rssB* is under $\sigma^S$ transcriptional control, and thus a negative feedback loop is closed [76, 80]. Other important players in this process are RssB antagonists that bind it in a non-productive complex and prevent its binding to RpoS and consequent RpoS degradation. To date, three of these proteins are identified and termed Ira. IraP and IraM are induced in by phosphate starvation [89] and magnesium starvation [90] respectively, and IraD is induced by DNA damage or hydrogen peroxide [91].

In summary, diverse stresses as well as the nutrient starvation increase the levels of RpoS in the cells via the multitude of general and specific pathways.

### 1.4.2. Stress response in *Escherichia coli*

$\sigma^S$ controls not only a regulon but rather a regulatory network with a complex intrinsic hierarchical and modular structure. Quite a large number of $\sigma^S$-dependent genes encode regulatory proteins [77] which can be expected not only to affect the expression of subsets of

$\sigma^S$-dependent genes but also to serve as additional signal integrators. Moreover, $\sigma^S$ itself is under complex transcriptional and posttranscriptional control [76, 80].

Because of the large size of the $\sigma^S$ regulon and its internal complexity, it is often difficult to pinpoint the specific genes required for particular $\sigma^S$-controlled phenotypes. Consequently, complex phenotypes are still not understood well, whereas, on the other hand, many $\sigma^S$-controlled genes remain functionally uncharacterized [76]. In this chapter, we will only briefly discuss how the cells respond to some stresses.

Acid resistance network is an example of $\sigma^S$-controlled pathway. $\sigma^S$ is induced in response to pH downshift, and it is required for the transcription of the regulatory genes *gadX*, *gadW*, *gadY*, and *gadE*. The small regulatory GadY RNA controls the cellular level of GadX and GadW by stabilizing their polycistronic mRNA and preventing its degradation [92, 93]. They are in turn required for the expression of the core acid resistance genes (including *gadA*, *gadBC*, *hdeAB*, *hdeD*). Positive autoregulation of GadE allows very strong expression of these target genes [76, 77, 92].

$\sigma^S$ is induced upon osmotic upshift in both growing and stationary cells and serves as a major osmoregulator [77]. $\sigma^S$ activates the *otsBA* operon, which encodes the enzymes that produce osmoprotectant trehalose, as well as the genes for the trehalases TreA and TreF which are involved in degrading trehalose upon osmotic downshift [94]. Other genes required for osmotolerance include *osmY*, *osmF,* and *osmC* [76].

$\sigma^S$ is involved in resistance to the oxidative stress. The genes induced by it in response to both endo- and exogenous oxidative stress include *dps, xthA, uspB, katE* and *katG, gor, sodC, sufABCDSE* operon*, bfr,* and *hem* [76]. Dps and Bfr scavenge iron and thereby prevent the oxidative reactions in a close proximity to DNA [95, 96]. XthA (exonuclease III) and UspB are involved in DNA repair [97, 98]. The catalases KatE and KatG, as well as glutathione reductase Gor and periplasmic superoxide dismutase SodC inactivate the reactive oxygen species and prevent the further DNA damage [77].

### 1.4.3. RpoS as the second housekeeping σ in slow-growing cells

Around 20% of the genes in the $\sigma^S$ regulon have metabolic functions [76, 77, 99]. On top of this, during slow growth, $\sigma^S$ can also take over the expression of some housekeeping genes that are activated by $\sigma^{70}$ during the exponential phase. These include very crucial pathways, such

as ribosomal genes [100] and glycolytic genes [101]. In a way, $\sigma^S$ thus becomes a second housekeeping σ-factor in slowly growing or stationary-phase cells.

For example, a variety of ABC-transport systems for amino acids, oligo- or dipeptides, phosphate sources or polyamines (encoded by the *art*, *dpp*, *opp*, *ugp*, and *pot* operons, respectively) are under positive $\sigma^S$ control [77, 102]. On the other hand, $\sigma^S$ seems to collectively downregulate uptake and/or metabolism of certain alternative carbon sources. As a consequence, Δ*rpoS* mutations improve growth on low-concentrated [103] or non-optimal carbon sources such as succinate [74, 75] or arginine [99, 104].

Importantly, the anti-$\sigma^{70}$ factor Rsd is under control of $\sigma^S$ which helps to transition from the rapid growth in the exponential phase to slow residual growth in the stationary phase [77].

### 1.4.4. Growth advantage in the stationary phase (GASP)

Upon having reached the stationary phase and the maximal cell density, the cells eventually begin to lose viability, marking the transition from stationary phase into death phase. It is marked by loss of viability of majority of the cells [105]. But even after the death phase, *E. coli* can be maintained in batch culture for long periods of time without the addition of nutrients. This period is called long-term stationary phase. Unlike early stationary phase, long-term stationary phase is a highly dynamic period in which the birth and death rates are balanced. In order for one cell to divide, the other has to die [105].

During the long-term stationary phase, some cells start to express the growth advantage in stationary phase (GASP) phenotype. This phenotype is defined by the ability of cells aged in long-term batch cultures to outcompete cells from younger cultures [105, 106]. This competitive advantage of cells expressing the GASP phenotype is shown to be genetically determined. The best-characterized and most common GASP mutations are in *rpoS*. The mutations that reduce, but do not eliminate, $\sigma^S$ activity are frequently associated with the expression of GASP [106, 107]. Phenotypically, GASP mutations in *rpoS* increase the ability to use alanine, arginine, aspartate, glutamate, glutamine, serine, threonine and proline as sole sources of carbon and energy [108].

There is a considerable degree of genotypic diversity during long-term batch culture. It is believed that this diversity is partially introduces via SOS-response pathways [109]. DNA might be damaged so badly that during the replication DNA polymerase III (the main DNA

polymerase of *E. coli*) cannot copy through. Frequently when this happens the SOS response is activated, inducing the genes encoding the alternative DNA polymerases polymerase II, polymerase IV and polymerase V [109, 110]. These polymerases are more error-prone (especially polymerase IV and polymerase V) than the standard polymerase III, and introduce mutations at the regular basis.

### 1.4.5. RpoS and mutagenesis

Obviously, the main stress that induces the SOS response is direct DNA damage, such as UV illumination or exposure to toxic chemicals. The SOS-inducing signal is single-stranded DNA which is generated during double and single breaks [111]. However, other stress conditions that do not directly damage DNA such as high hydrostatic pressure or starvation, can trigger SOS-response as well [112, 113]. In this case, the double-strand breaks are caused by a cryptic endonuclease encoded on the *E. coli* chromosome, and processing of the double-strand breaks by RecBCD produces the SOS-inducing single-stranded DNA [111].

The general stress response is linked to two main mechanisms that can increase mutation rates: induction of polymerase IV and downregulation of mismatch repair. The induction of polymerase IV is shown to be independent from the SOS-response. Instead, its gene *dinB* is regulated directly by $\sigma^S$ [114]. Δ*dinB* strains do not display the GASP phenotype, highlighting the importance of this particular gene for genetic variability in the long-term stationary phase [111]. The levels of two mismatch repair proteins, the mismatch binding protein MutS and the endonuclease MutH, decline via an RpoS-dependent mechanism [115].

These two mechanisms contribute vastly to the genetic diversity in the stationary phase and upon the stress conditions. The mutation rate is increased compared to the exponentially growing unstressed cells [111]. This leads to the increased probability to acquire a beneficial mutation that will give an advantage in the harsh conditions. Attenuation of these two mechanisms via the reduced $\sigma^S$ activity (but importantly, not a complete abolishment of them) in the GASP cells is probably a result of the stabilizing selection.

## 1.5. σ factors competition

As important as RpoS is, there are multiple evidences that its loss might be beneficial in certain conditions. In the laboratory setting, loss of RpoS activity is a rather mundane event [71] but it is usually explained by the lack of stresses that the bacteria face as they grow in the rich

medium at the optimal temperature. But even in the commensal, pathogenic or environmental strains of *E. coli* the *rpoS* locus remains highly polymorphic [73-75, 116] and sometimes these strains even lose its function completely. The RpoS inactivation happens despite the highly competitive environments these bacteria have to live in. These seemingly contradictory results might be explained by the theory of σ-factors competition. According to this theory, the number of core RNA polymerase subunits (RNAPs) in the cell is constant, but the number of the σ-factor molecules fluctuates in response to the external conditions. Therefore, the σ-factors have to compete to each other for the limited amount of RNAP complexes to initiate the transcription [73, 117, 118].

As a result of competition, any increase in activity or amount of one σ-factor indirectly represses binding of other σ-factors to core RNAP and thus transcription of the genes they control. Consequently, a decrease of absence of a one σ- factor would increase the transcription of the genes governed by the other σ- factors [119]. To date, one housekeeping σ-factor $\sigma^{70}$ (*rpoD*) and six alternative σ-factors $\sigma^{N}$, $\sigma^{S}$, $\sigma^{H}$, $\sigma^{F}$, $\sigma^{E}$, and $\sigma^{I}$ (*rpoN*, *rpoS*, *rpoH*, *rpoF*, *rpoE* and *fecI*; governing nitrogen metabolism, stationary phase and general stress response, heat shock response, motility, envelope stress, and iron metabolism respectively) have been described in *Escherichia coli*. The biggest regulon is $\sigma^{70}$, followed by a plethora of genes under the joint regulation of both $\sigma^{70}$ and $\sigma^{S}$ [78, 120]. All σ factors but $\sigma^{N}$ belong to the $\sigma^{70}$ family while $\sigma^{N}$ differs substantially from the others in sequence and structure [104].

There are several parameters that modulate the σ- factor competition. First and foremost, different σ- factors select different promoters to start the transcription from [76-78, 119]. Second, there are proteins named anti-σ-factors, which bind to the σ-factor and prevent the formation of a productive complex with RNAP [119]. Third, the amount of the alternative σ-factors is not constant and changes in response to the external conditions.

### 1.5.1. RpoS: friend or foe?

It was shown in the evolutionary experiments on *E. coli* that *rpoS*-negative mutants often appear and rapidly spread in low-carbon and low-nitrogen conditions [118]. The authors explain it with competition between stress response σ-factor $\sigma^{S}$ and housekeeping σ-factor $\sigma^{70}$ for the RNA polymerases. For example, vast majority of genes (i.e. all the genes involved in glucose uptake) are under $\sigma^{70}$ regulation [121]. Reducing the number of $\sigma^{S}$ molecules leads to the higher availability of the polymerases for $\sigma^{70}$, and therefore expression of all genes involved in

glucose uptake is increased. Consequently, glucose uptake is increased and this mutation becomes advantageous in the conditions of the carbon limitation. Most of the *rpoS* mutants that appear in the starvation conditions have little or no residual RpoS protein [103].

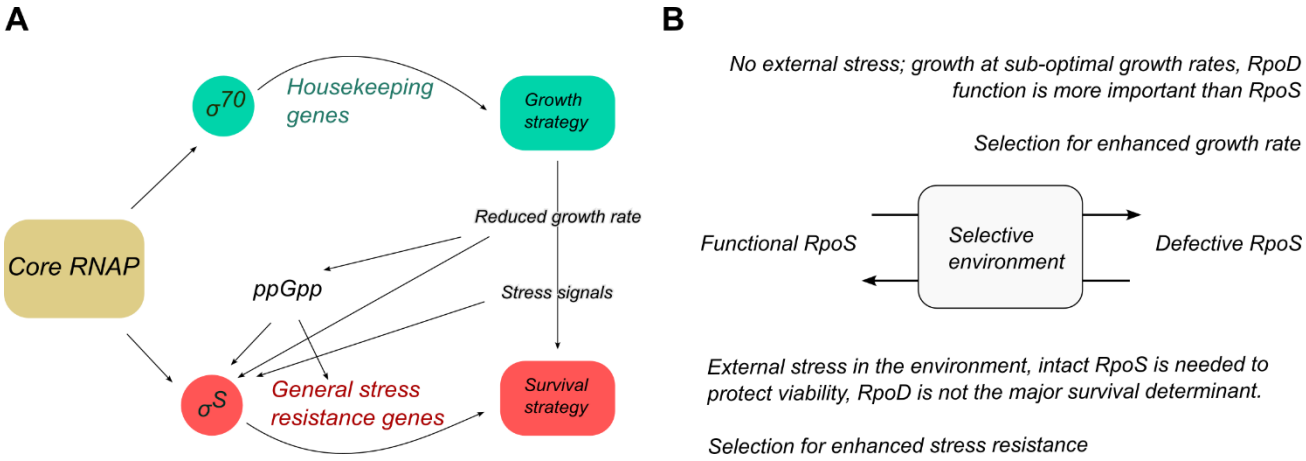



**Figure 4. Competition between $\sigma^{70}$ and $\sigma^{S}$.** (A) Decision making between vegetative growth and survival strategies in *Escherichia coli*. (B) Antagonism in the *rpoS* polymorphism cycle. Adapted from [73].

When neither extreme of vegetative or stress-induced gene expression is appropriate, a slow-growth dilemma is imposed (**Figure 4**A). This is a major limitation in environments that are sub-optimal for growth-rate but require high housekeeping gene expression, such as low-nutrient conditions. In this respect, the $\sigma^{70}$–$\sigma^{S}$ switch is not well designed for dealing with environments that do not exert external stress and do not correspond to the extreme feast or famine scenarios. Therefore, escape from the burden imposed by stress-gene expression provides selection pressure for the common occurrence of *rpoS* mutations in populations of *E. coli*. The pressure can be extremely strong, and lead to takeover by *rpoS* mutants within 10 generations of growth under glucose-limited chemostat culture conditions [103].

The situation changes, however, if the secondary stress is present. When a mild external stress (such as lowered pH) is combined with the carbon limitation, *rpoS* mutants are still enriched, but are predominantly partial mutants capable of limited stress resistance rather than null mutants, which by contrast are common in the absence of physical stress [103]. Hence the trade-off between stress resistance and vegetative growth is sensitive to the environment and can also contribute to the diversity of alleles found in natural populations [73]. These results match to the GASP phenotype [105] described above.

Strains of *E. coli* that have naturally higher levels of $\sigma^S$ are more likely to lose the activity of it in the conditions discussed above [118]. This happens due to the stress-induced mutagenesis which is dependent on $\sigma^S$ and is more intensive in high- $\sigma^S$ strains [111]. Complementary to this, *rpoS* mutants have a selective advantage in mixed culture by functioning as "cheaters". They benefit for the products produced by the wild type but do not have to pay the high metabolic cost of expressing the large RpoS regulon [122].

### 1.5.2. Other σ factors participate in this competition too

*rpoS* mutants are highly motile in comparison with wild type strains [99, 104]. This is attributed to the effect of $\sigma^{70}$ and $\sigma^F$ as these are known to direct flagellar gene transcription [123]. On top of this, flagellar gene transcription was shown to depend on $\sigma^N$ on a minimal medium [104, 124], as Δ*rpoN* mutants were less motile, and the motility was completely abolished in the Δ*rpoN*Δ*rpoS* mutants. On top of that, it was observed that in the minimal media, RpoS mutants utilize nitrogen more efficiently and more sources including sub-optimal ones like arginine possibly because $\sigma^N$ level is higher [99, 104]. This implies σ-factor competition not only between the $\sigma^{70}$ and $\sigma^S$, but also between the other σ.

Whole-transcriptome analysis of Δ*rpoN* and Δ*rpoS* strains reveals two distinct groups of genes. One group is positively regulated by RpoS but negatively by RpoN. The second group, on the opposite side, is positively regulated by RpoN but negatively by RpoS [104]. The former group involves, for example, the multidrug efflux pump *mdtEF* and *gadABCE* acid resistance system. The latter group prominently involves the flagellar regulon. The effect of $\sigma^N$ on flagella, however, might be limited to the *E. coli* K-12 MG1655 strain because it has an additional $\sigma^N$-dependent promoter of the flagellar master regulator *flhD* brought there by a transposon IS1 [72, 124]. This transposon insertion is unique for the MG1655, even a closely related W3110 strain does not have it [72].

## 1.6. Motility and flagellar regulation in *Escherichia coli*

Another major regulon in *E. coli* is a motility and chemotaxis system. Its regulation is tightly interweaved with the $\sigma^S$-regulon [125].

### 1.6.1. Composition and assembly of *Escherichia coli* flagella

The motility and chemotaxis of *E. coli* is a complicated system which requires tight regulated expression of around 50 genes. In addition to the flagellum (a very complex machine *per se*),

motility requires a secretion apparatus, receptors, chaperones, numerous components of signal transduction systems and several transcriptional factors, which together lead to a tightly regulated and coordinated system [126].

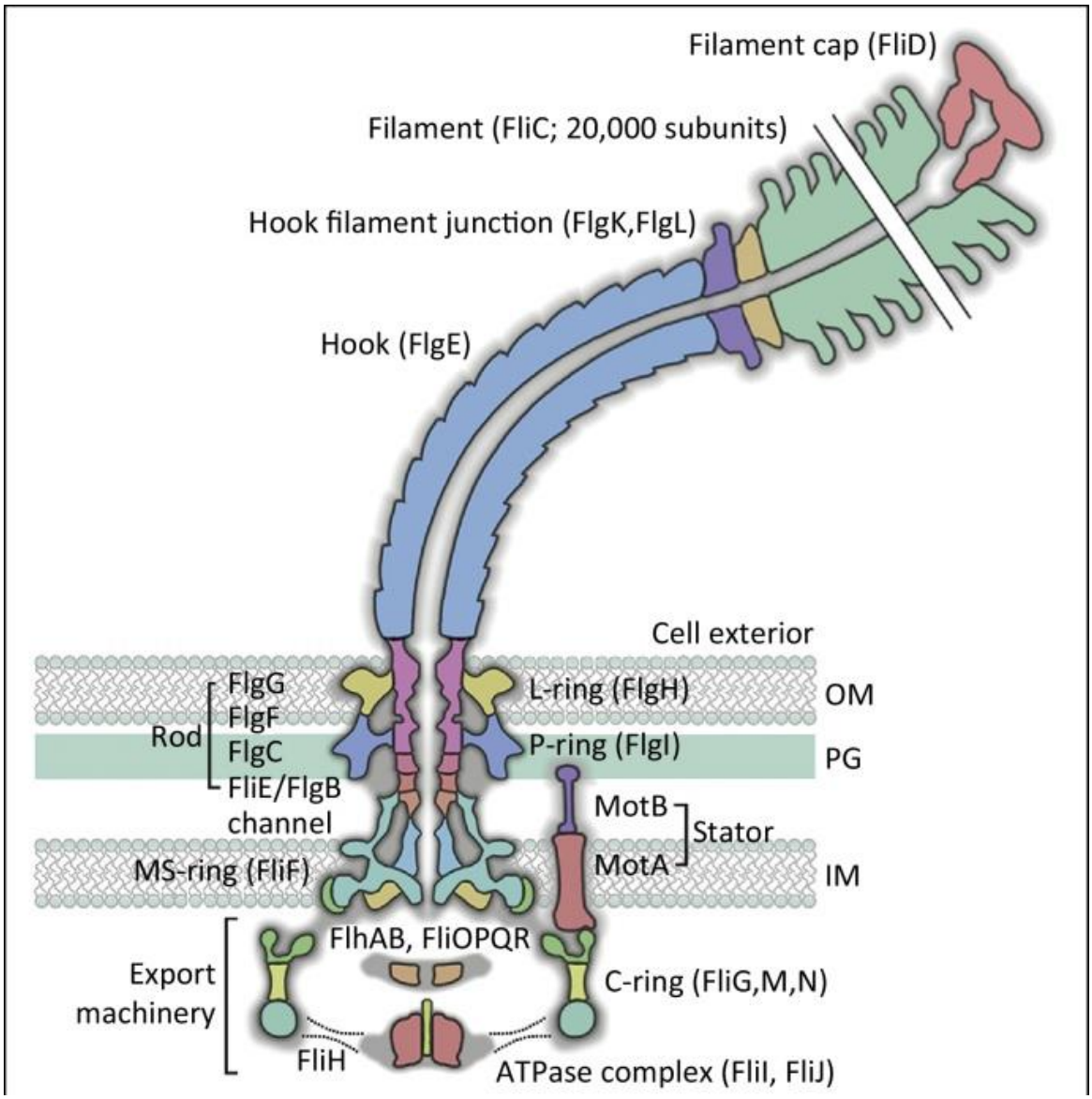



**Figure 5. Structure of a mature bacterial flagellum.** Adapted from [127]. See the text for the details.

The basal part of the flagellum is formed by several ring structures that are anchored in the membranes. These are cytoplasmic C ring, inner membrane-embedded membrane–supramembrane (MS) ring, peptigoglycane (P) ring, outer-membrane-embedded lipopolysaccharide (L) ring, distal and proximal rod structures, type three secretion system and, finally, the motor components [126, 127](**Figure 5**).

The early stage of flagellar assembly involves the formation of the inner rings (MS and C, composed of FliG, FliM and FliN), and the assembly of the type three secretion system. The secretion system is composed of the membrane proteins (FlhA, FlhB, FliF, FliO, FliP, FliQ, FliR) and components of the ATPase complex: ATPase FliI, negative regulator FliH and chaperone FliJ [128]. This system secrets the external parts of the basal body (FlgB, FlgC, FlgE, FlgG, FlgJ) as well as the filament itself. The components of the MS and P rings (FliF, FlgI) are secreted through the Sec system. Finally, MotA and MotB integrate into the membrane and form stator, while C-ring components function as the rotor [129, 130].

The flagellar filament is a long tube which consists of the flexible hook and the rigid propeller. The hook is formed by the polymerization of the hook filaments FlgE. FlgK and FlgL control the length of the hook and later form hook-filament junction on the outer end of the hook [131]. The formation of the propeller starts with an assembly of the cap structure FliD which moves outward as the monomers of flagellin polymerize and form the filament [126, 129].

### 1.6.2. Regulation of flagellar operon

In *E. coli* and closely related *Salmonella* species, around 50 motility and chemotaxis genes are hierarchically organized in three classes of expression: class I, class II, and class III [126]. The general logic behind the organization of this operon is that the genes are expressed in the same order in which they will be required for assembly of the flagellum (**Figure 6**).

The master regulator of the flagella and chemotaxis operons is a transcriptional activator FlhDC. It consists of the products of *flhC* and *flhD* genes. Both genes are expressed from the single $\sigma^{70}$ -dependent class I promotor and function as a heterohexamer $FlhD_4C_2$ [126, 132]. FlhDC expression and stability is tightly regulated on different levels because the flagellar synthesis and operation are very costly processes [133]. Some regulators such as H-NS, CRP and QseB have a positive effect on its expression, while the Rcs and Omp system act as repressors [134-138]. The expression of *flhDC* is inhibited by high temperature, possibly via the existence of recently discovered antisense-*flhD* RNA [139] which is expressed from a $\sigma^E$-dependent promoter.

In turn, FlhDC induces expression of class II genes that encode components of the basal body and a flagellar σ factor $\sigma^F$ (FliA). After the complete assembly of the basal body and the type three secretion system, $\sigma^F$ binds to the RNA polymerase and induces expression of the class

III genes, which form the hook, filament, motor and the components of the chemotaxis sensory and signal transduction network [126].

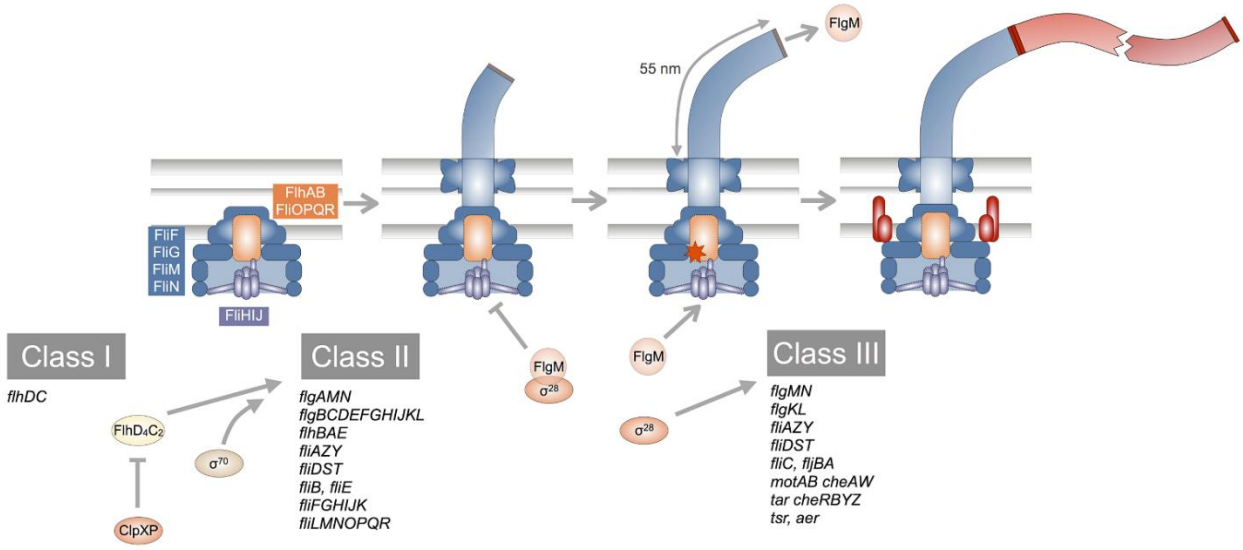



**Figure 6. Overview of the flagella regulation cascade.** Adapted from [127]. See the text for the details.

One of the class III genes is FlgM that functions as anti-σ factor, binding $\sigma^F$ and thus inhibiting the expression of the class III genes. The ratio of FlgM to $\sigma^F$ is maintained via secretion of FlgM through the type three secretion system. The excess FlgM is secreted to the outside of the cell, thus allowing the expression of $\sigma^F$-dependent genes [140]. FlgM plays the major role in inhibiting the assembly of flagella at higher temperature [141].

### 1.6.3. Biofilm development in *Escherichia coli* is regulated by both RpoS and flagella

Biofilm formation commonly involves several developmental steps (**Figure 7**). It is typically initiated with surface attachment of motile planktonic bacteria, which occurs in two stages – reversible (or transient) and then irreversible attachment [142]. Reversible attachment is typically unstable, characterized by cells attaching to a surface and returning to the planktonic state. No phenotypic changes occur at this stage. After initial reversible attachment, bacteria start to strengthen the contact with the surface by producing the ECM components and adhesines, and form microcolonies. At this stage, they often lose flagella. Then the biofilm matures into a complex three-dimensional structure. At the final stage of the biofilm development, the dispersion phase, the bacteria are released from the biofilm, ready to establish the new community elsewhere [46, 48, 49, 51].

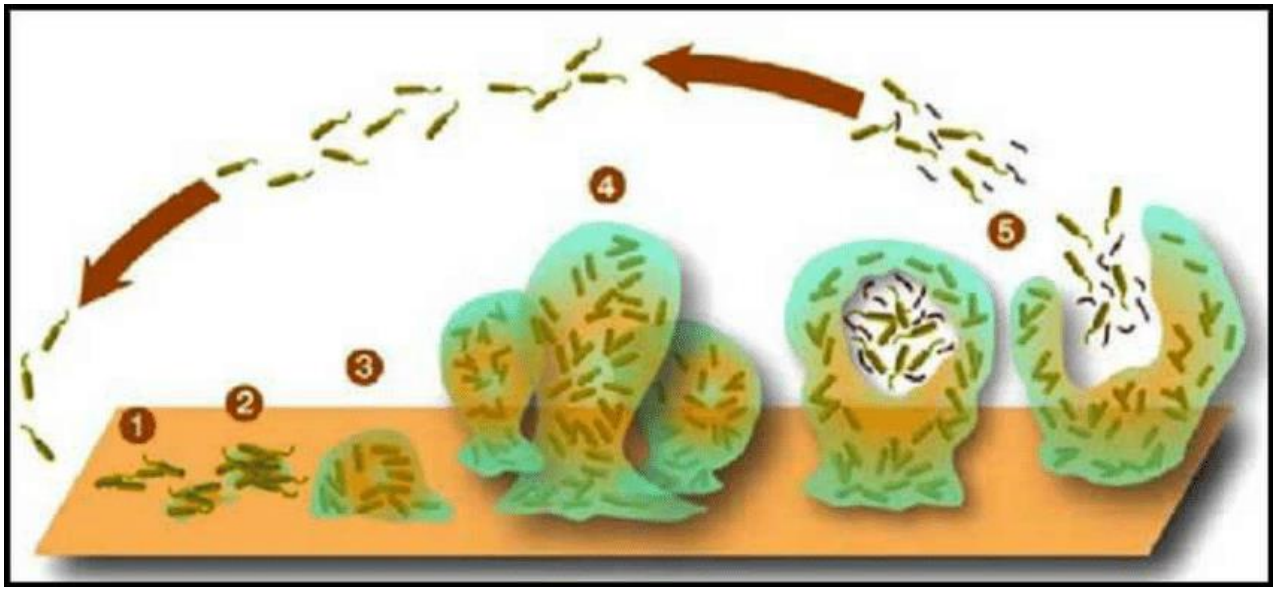



**Figure 7. Stages of biofilm development.** 1. Initial reversible attachment. 2. Irreversible attachment. 3. Start of ECM formation and formation of microcolonies. 4. Mature biofilm. 5. Dispersion of the biofilm. Adapted from [142]. See the text for the details.

All the stages of the biofilm formation are tightly regulated in time and space. In shortest terms, the transition from motility to biofilm lifestyle requires both downregulation of flagellar and activation of $\sigma^S$ regulatory cascades. The activation of these two systems is mutually exclusive and both inhibit each other on several levels [125]. The scheme of this regulation is shown at the **Figure 8**.

The first level of regulation occurs on the mutual repression of FlhDC- and $\sigma^S$-regulated operons at the transcription level. This happens via the σ-factor competition between the $\sigma^S$ and $\sigma^F$ [125].

The major player in this system is CsgD, an important activator of biofilm formation [28, 52, 125]. $\sigma^S$ both directly and indirectly controls *csgD* transcription, and many of the secondary regulators are themselves under the $\sigma^S$ control. These include the MerR-like transcription factor MlrA, several pairs of diguanylate cyclases (DGC) and phosphodiesterases (PDE) [77, 125]. Diguanylate cyclases synthesize cyclic dimeric guanosine monophosphate (c-di-GMP) while phosphodiesterases hydrolyze it. Thus, these antagonistic pairs of enzymes control and maintain the levels of c-di-GMP which is an important second messenger [80, 143]. Two such DGC/PDE pairs (YdaM/YciR and YegE/YhjH, with the DGC YedQ acting as a backup for YegE) increase *csgD* transcription in a nonadditive manner [77, 125].

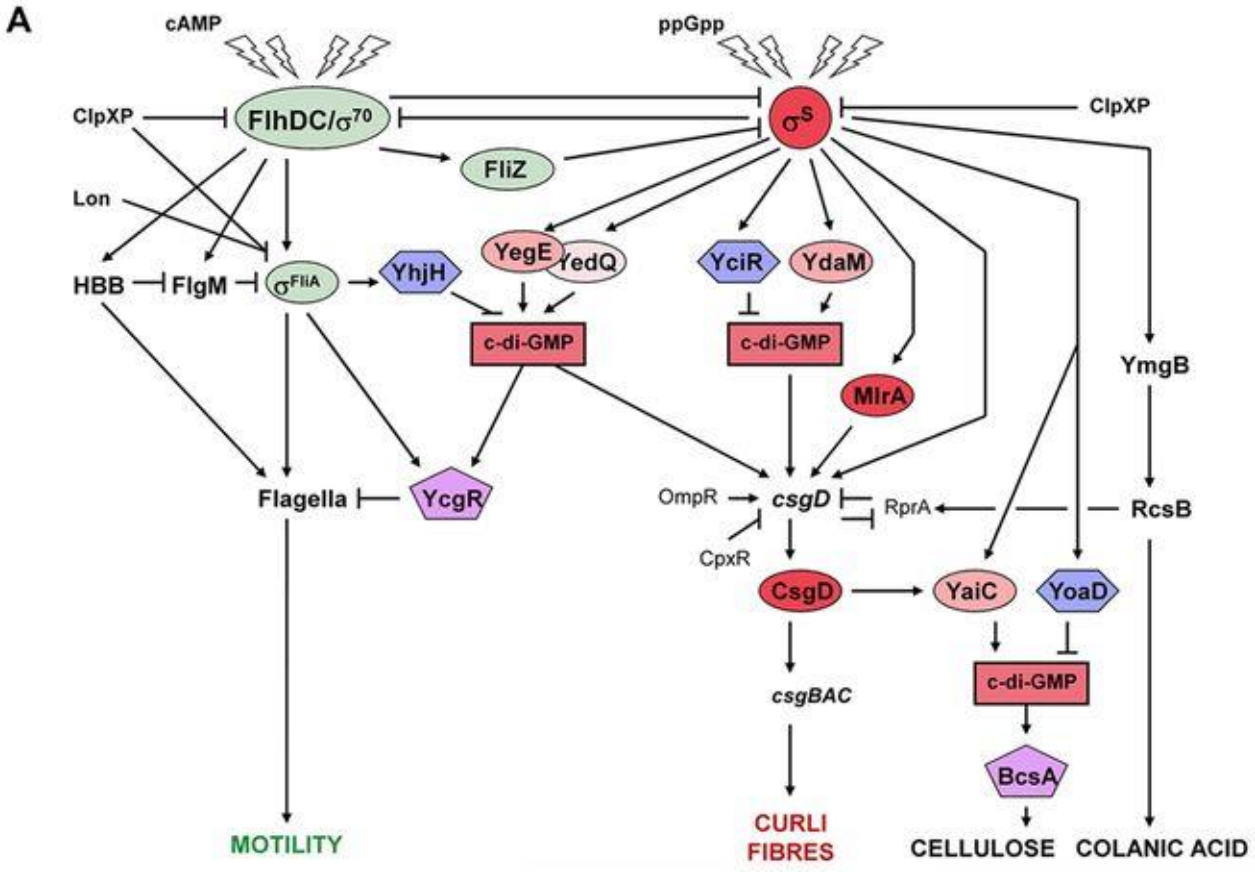



**Figure 8. Cross-talk between curli regulation and flagella regulation in *E. coli*.** Adapted from [76, 125].

At this level, the second layer of mutual inhibition takes place. All the DGC/PDE genes listed above are under $\sigma^S$ control except *yhjH* which belongs to the class III flagellar cascade gene [77, 126]. As mentioned earlier, $\sigma^S$ is involved in downregulating FlhDC and all the downstream genes of the flagellar cascade via σ-factor competition. This downregulation of the PDE YhjH is the triggering signal that allows the DGCs YegE and YedQ to accumulate sufficient c-di-GMP to activate *csgD* [125]. On top of this, FlhDC regulon contains phosphodiesterase YcgR, which counteracts YegE and YedQ and allow rotation of flagellum [126].

Once CsgD accumulates, it activates the expression of the curli operon *csgBAC* and thus the curli formation. Additionally, the cellulose biosynthesis is triggered via the activation of another DGC *yiaC*. This plays a role later during biofilm formation and the EPM biosynthesis, which mainly consists of curli and cellulose [30, 76, 125].

The third level of mutual inhibition is a protein FliZ which is identified as a factor that suppresses the expression of multiple $\sigma^S$-dependent genes. The gene *fliZ* belongs to the class II of the flagellar cascade and is expressed from the same operon as the $\sigma^F$. FliZ acts as a

global repressor that antagonizes $\sigma^S$ by recognizing a sequence of $\sigma^S$-dependent promoters, thus not allowing $\sigma^S$ to bind to DNA. Physiologically, FliZ gives priority to the expression of flagellar genes and motility as the Δ*fliZ* strains enter the stationary phase earlier than the wild type [76, 125]. Moreover, cells reduce their swimming speed, via a c-di-GMP-dependent flagellar brake YcgR, and finally become nonmotile later in the stationary phase [76, 125].

The highly hierarchical and modular structure of this system allows rapid response to the environmental cues, and more importantly, it allows the possibility of switching between motile and sessile lifestyle.

## 1.7. HdfR: another potential link in RpoS-flagella cross-talk

### 1.7.1. HdfR as a negative regulator of motility

HdfR was described as an H-NS-dependent negative transcriptional regulator of the flagellar master operon *flhDC*. The disruption of *hdfR* enhances the activity of *flhDC* [134], and the strains of *E. coli* that have enhanced activity of *flhDC* are more motile [135]. Moreover, the binding of HdfR to *flhD* promoter was demonstrated in the gel-shift assay [134]. The functions of HdfR are not fully studied yet. Besides the effect on flagella regulation, it is reported to be involved in glutamate biosynthesis and acid stress resistance by activation the transcription of *gltBD* operon [144]. It was also shown that a point mutation in *hdfR* (E40K) had reduced sensitivity to the redox stress and increased NADPH/$NADP^+$ ratio compared to the wild type strains [145] in a minimal medium. Additionally, it was reported that overexpression of HdfR enabled the cells to withstand higher concentrations of cadmium in the medium [146], while its depletion leads to higher sensitivity to it. In *Salmonella enterica*, a homologue of HdfR is necessary to induce one of its fimbria operons [147] via methylation-dependent bistable switch. According to the bioinformatic analysis, *hdfR* has two promoters, one on which is under RpoS regulation [148] and the other one is regulated by FliA [149]).

### 1.7.2. MaoP: HdfR's neighbor with unclear function

*hdfR* has an adjacent divergently transcribed gene named *maoP*. These two genes share the same intergenic region, although the transcripts unlikely overlap. Such a close neighborhood implies that these genes might influence each other transcription. Indeed, *maoP* was demonstrated as a regulatory target of HdfR, as its transcription was completely abolished in the Δ*hdfR* strain [150].

The available information on MaoP is scarce and disconnected. It was first described in 2006 [151] in a bioinformatic screening of *dam* (DNA adenine methyltransferase) dependent genes. Dam-dependent methylation of DNA is actively recognized during the course of the cell cycle, and is crucial to perform correct replication initiation, chromosome partition, cell division and mismatch repair [152]. Brézellec et al. have shown that in γ-proteobacteria, 18 genes were co-conserved and strictly confined to genomes that contain a resident *dam* gene. This set includes a number of genes whose products are directly involved in DNA maintenance and/or sensitive to DNA methylation status, and a number of genes with unknown function including *maoP*. The authors suggested that these genes they have highlighted were potential targets to study their involvement in chromosome organization and their association with DNA methylation status. Interestingly, a homologue of HdfR in *S. enterica* has tight connection to Dam-dependent DNA methylation as well [147].

In 2012, a comprehensive screen for discovery of genes involved in stress-induced mutagenesis under starvation conditions was performed [153]. Among other results, it was shown that both Δ*hdfR* and Δ*maoP* strains demonstrate lower rate of stress-induced mutations of both frameshift and base substitution type. Additionally, these knockouts were more sensitive to UV radiation. Therefore, *hdfR* and *maoP* are both important to induce mutagenesis during the starvation stress.

Later in 2016, MaoP was indeed described as a protein involved in chromosome organization and more specifically in organization of Ori macrodomain [154]. In this paper, the name change from *yifE* to *maoP* (**Ma**crodomain **O**ri **P**rotein) was proposed. Previously, it was shown that movement of DNA loci in the cell was highly constrained, and the distribution of these loci correlates with macrodomain organization [155]. However, inactivation of *maoP* lifted some of the constraints on DNA mobility in the Ori macrodomain and allowed long-range interaction between Ori and Right macrodomains. The other important part of this system, sequence *maoS*, is situated in the intergenic region between *hdfR* and *maoP*. Both *maoP* and *maoS* were required to constrain DNA mobility, and *maoS* had to be *in cis* on the chromosome. The authors claim that *maoS* is likely a binding target sequence of MaoP.

In 2021, two more papers mentioning *maoP* were published. According to one of them *maoP* mutants tend to overproduce putrescine in the rich medium [156] by interfering with the pathways of amino acid biosynthesis. *maoP* deletion increased the transcription of *hdfR* and consequently *gltB* genes thus increasing the available levels of glutamate in the cell. In the

other paper, the whole-genome screen of essential genes for biofilm formation was performed [157]. *ΔmaoP* had lower curli production and lower biofilm biomass compared to the wild type strain. It is also reported that a homologue to *maoP* in *Yersinia pestis* might have a role in adhesion and early infection [158], and therefore the function of *maoP* in biofilm formation was assessed as related to adhesion and EMP biosynthesis.

## 1.8. Aims of the current work

Previous studies [13, 18, 41, 47] demonstrate that the stress response pathway is extremely important for *E. coli* to thrive in the secondary environment. The primary aim of this project was to determine the other potential genetic features of *E. coli* growth and survival in lake water. To achieve these results, we employed two complementary approaches. First, we used the analysis of a transposon mutant library that allows quick comparison of growth of different mutants in different conditions using the next-generation sequencing. Second, we have performed whole-genome fluorescence-based screen of growth and survival of the Keio collection mutants labelled with GFP.

In the second part of the work, we investigated the functions of HdfR/MaoP regulatory system that was determined as a novel determinant of *E. coli* proliferation in the lake water. HdfR is a poorly studied transcription factor described as a negative regulator of motility. Therefore, the secondary aim of the project was to study the mechanism of the interplay between HdfR, MaoP, flagellar regulation cascade and the stress response cascade. We studied motility, biofilm formation and expression of the array of genes, as well as chromosome organization, protein-protein interaction and protein-DNA interaction.

# Chapter II. Whole-genome screen of *Escherichia coli* persistence in the lake water

To assess *E. coli*'s ability to persist, we employed two approaches. First, we used the analysis of a transposon mutant library (5.2. Methods). This is a powerful tool to obtain a high-throughput data of bacterial growth and survival in various conditions. In this approach, a bacterium on interest is mutated by means of random barcoded transposon insertions, resulting in a pool of barcode-labelled mutants. Each barcode is flanked by the common PCR-binding sites U1 and U2, enabling the amplification of the barcode sequence. The comprehensive table of all the barcode sequences and their respective positions in the genome is then generated using Illumina sequencing and bioinformatics analysis. After the position of each barcode is defined, it becomes possible to track the mutants and measure their abundance in various conditions by amplifying the barcode region with the transposon-specific primers and then sequencing of the resulting amplicon [159]. In this setup, the bacteria are pooled together and have to compete to each other in order to grow.

Complementary to this approach, we have performed a whole-genome comprehensive fluorescence-based (FB) screen of growth and survival of different mutants, based on the Keio collection of all the non-essential knockouts in *E. coli* [160, 161]. We have transformed the entire Keio collection with a GFP-carrying plasmid and cultivated each mutant separately in the lake water with the daily fluorescence measurements. Unlike in the transposon library, where all the various mutants are pooled together, in this setup we cultivated the strains one-by-one. This allowed us to eliminate certain factors such as cross-feeding, competition between different strains and formation of co-dependent communities.

The lake water for all the experiments was taken between mid-April and mid-September from the lake near Marburg, Germany. Prior to sampling, we measured temperature of air, temperature of water, salinity, and pH (**Supplementary table 1**). The cultures of *E. coli* were grown at 37°C on tryptone broth (TB) medium in a rotary shaker, and then inoculated into plates filled with native (taken from the lake without any additional treatment) or filter-sterilized lake water. We did not add any antibiotics to the water to avoid harm to the native microbiota of the lake. Then the cultivation took place at room temperature. In both the experimental setups, the cells were not washed prior to inoculation to mimic the excretion from the host.

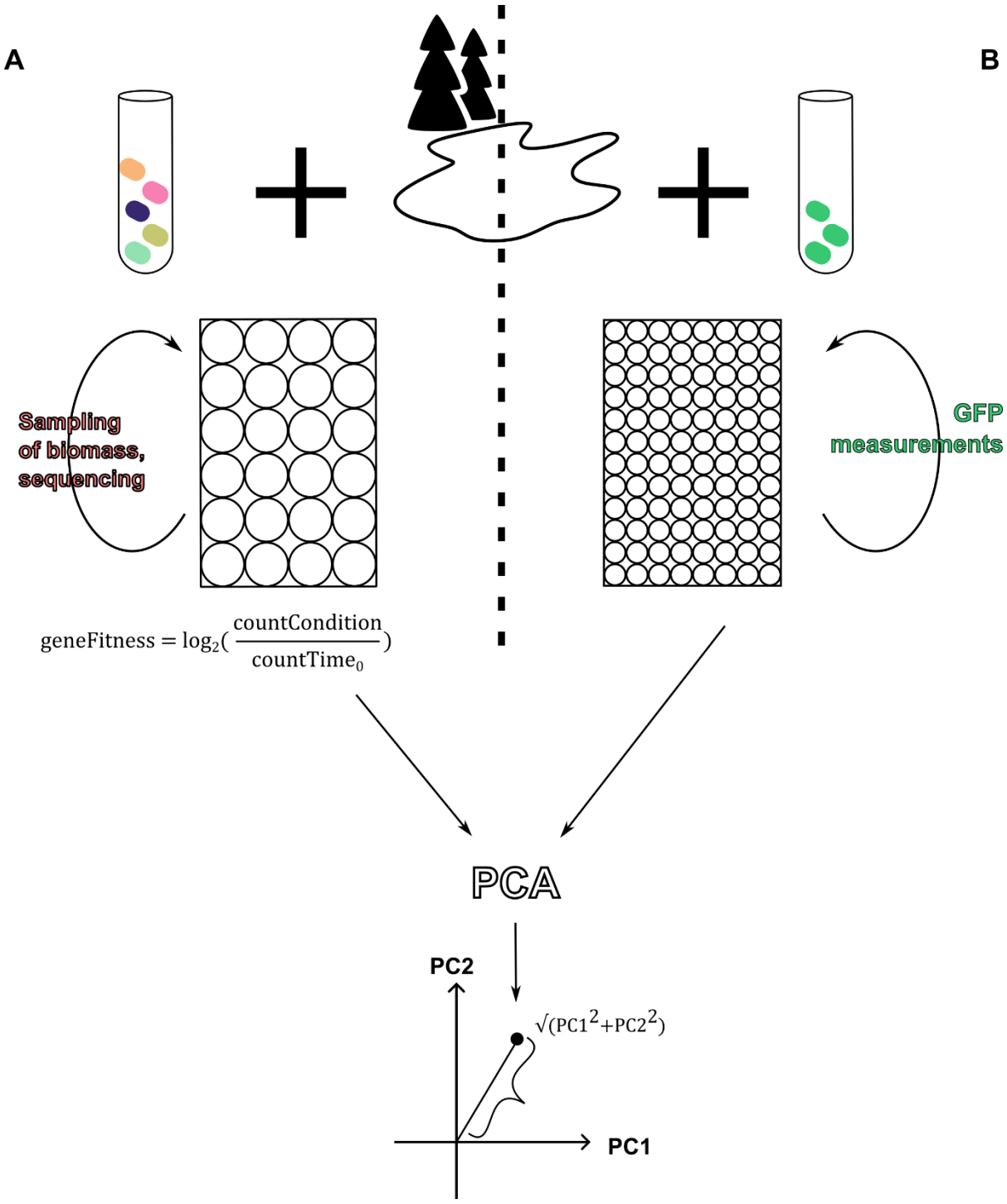



**Figure 9. Experimental design for transposon mutant sequencing (A) and fluorescence-based (B) screens.**

## 2.1. Transposon mutant library screening

To test the *E. coli* persistence in lake water we first used the transposon insertion sequencing (TnSeq) approach. We used *E. coli* K-12 W3110 RpoS$^+$, a well-studied and well-annotated laboratory strain, widely used in *E. coli* biofilm research [29, 52, 162]. The lineage of W3110 that

is stored in our laboratory has some differences from the reference genome (**Table 12**) [72]. Most notably, in the *rpoS* gene, the reference genome has a truncation at the 33$^{d}$ codon, rendering it a non-functional pseudogene. In our lineage, this stop codon is spontaneously replaced by tyrosine, allowing the full-length and functional *rpoS* to be expressed.

To start the experiment, we grew the library in TB supplemented with kanamycin to $OD_{600}$ ~ 1 and then inoculated the cultures into the filtered or non-filtered lake water. The samples of the bacterial biomass for DNA extraction and further amplification were taken at days 2, 4 and 8. The starting culture was used as the control (**Figure 9**A).

### 2.1.1. Growth of *Escherichia coli* W3110 RpoS$^{+}$ in the lake water

To analyze the results of the sequencing, we calculated the fitness of each mutant in the library. To do so, we counted the abundance of each barcode in the Illumina library in each condition. Strain fitness was then calculated as a logarithm of the fold change of barcode abundance in the condition of interest compared to the control (the starting culture). Therefore, fitness above zero means that the strain grows better compared to the average strain in the library, and *vice versa*.

First, we have compared the gene fitness values to 0 using one-sample Student t-test, and presented the results as volcano plots (**Figure 10**). As expected, some knockout strains were underrepresented in both filtered and non-filtered lake water, which means that they cannot successfully compete to the other strains and autochthonic microbiota (if it was present). To our surprise, the better-surviving knockout strains were also found. In this case, the fitness advantage is observed only in the filtered lake water.

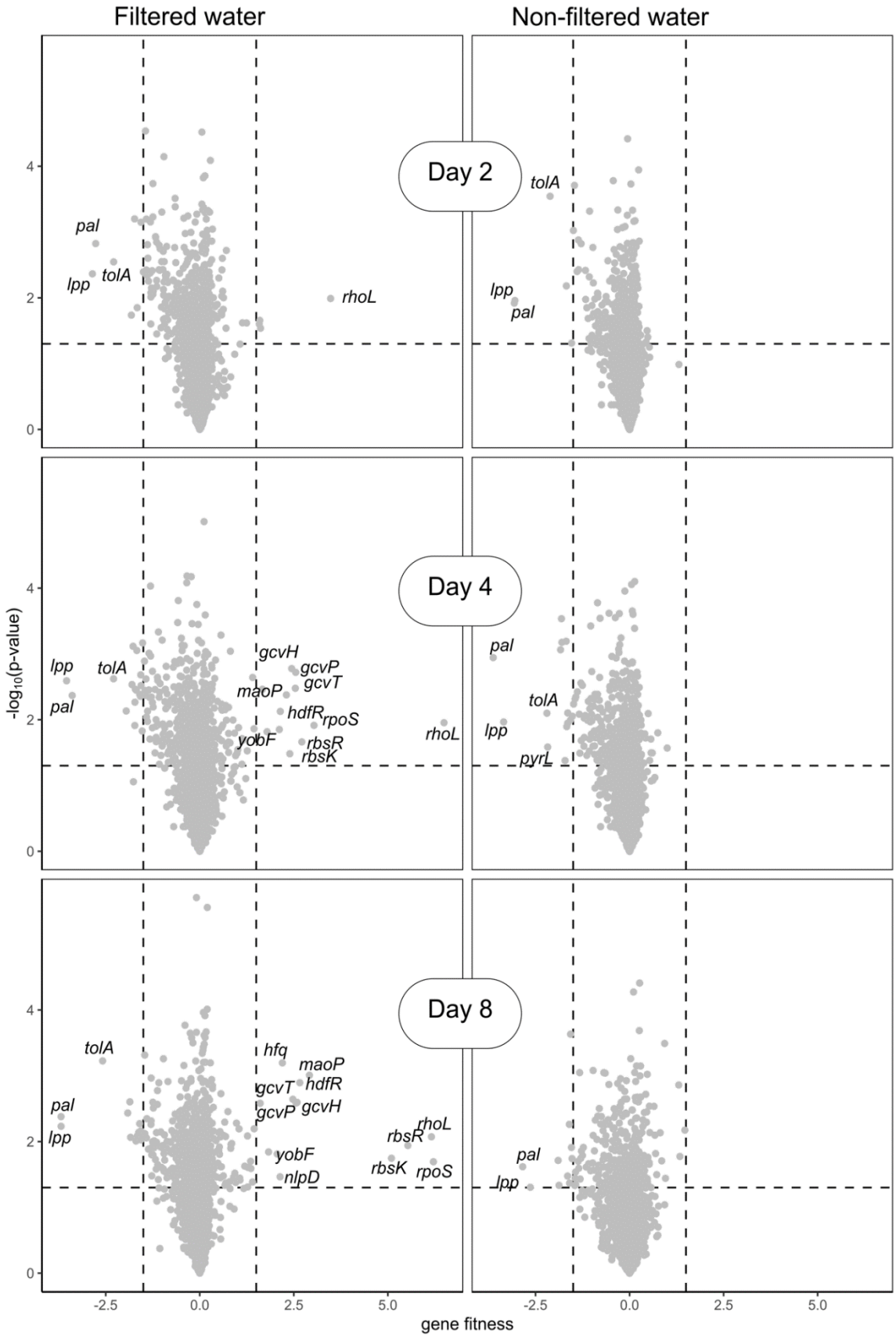

Filtered water
Non-filtered water
Day 2
Day 4
Day 8
-log10(p-value)
gene fitness
pal
lpp
tolA
rhoL
gcvH
gcvP
gcvT
maoP
hdfR
rpoS
yobF
rbsR
rbsK
pyrL
hfq
nlpD

**Figure 10. Changes in gene fitness in *E. coli* W3110 RpoS$^+$ in filtered and non-filtered lake water over time.** Shown are gene fitness (X axis) and $-\log_{10}$(p-value) of one-sample Student t-test (Y-axis). The results are based on three biological replicates.

To further study the dataset, we applied principal component analysis to the resulting data. Principal component analysis (PCA) is widely used in the analysis of multidimensional data. This approach allows not only reducing the number of dimensions in the dataset with the minimal loss of information, but also removing redundancy in the data, improving the visualization and facilitating further analysis of the dataset. Another advantage of PCA is that it does not have distributional requirements and we could not guarantee that our data had a normal distribution [163].

In this approach, the first principal component (PC) is a direction in the dataset that describes the greatest fraction of the variance. The second PC is always orthogonal to the first one and captures the second most of variance in the dataset. The third and every consecutive PCs are created using the same principle. Then, the new dataset is created via transposing the data to the space formed by the PCs. The number of the PC is always equal to the number of the variables in the initial dataset but only the first 2 components are used for the further analysis as they capture more than 70% of the variance in the dataset (**Supplementary figure 1**).

Each biological replicate was analyzed separately. On the biplots, we can see that the results are reproducible between the different samples of water, therefore the effect of the most over- and underrepresented mutations must be stronger than the effect of the lake water. In this library, vast majority of the mutations introduced during the Tn5 mutagenesis vast majority of the mutations, resulting in fitness values close to 0 and forming a dense cloud of points. The mutations in these genes were therefore not important for survival in the lake water. However, we were mostly interested in the strains that lie outside of that cloud as they resulted in significant change of fitness in the experimental conditions. We have calculated the Euclidean distance between the origin and the position of the respective gene in the space formed by the first two PCs, and used this value as a cutoff to determine the most pronounced differences.

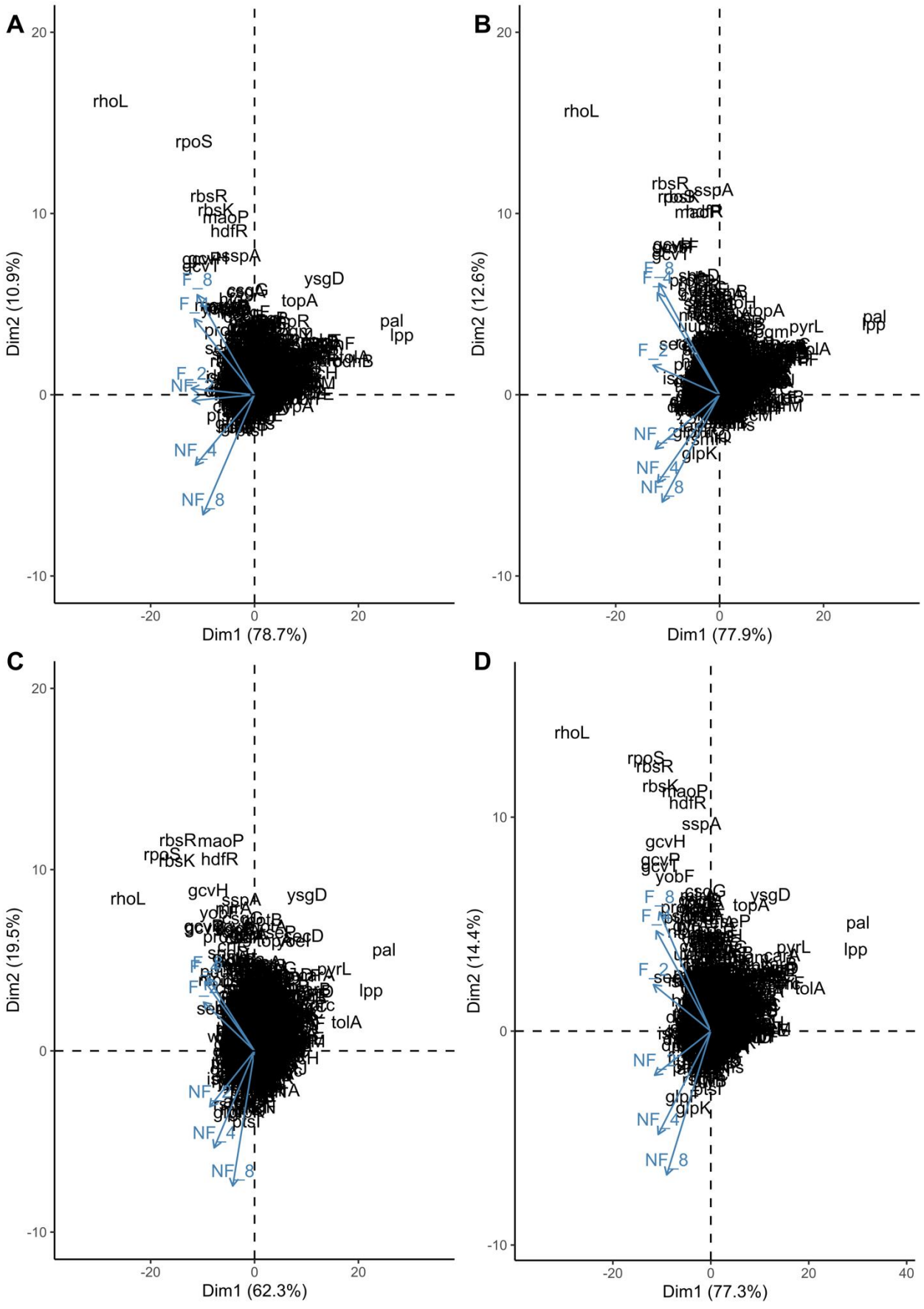

A
rhoL
rpoS
rbsR
rbsK
maoP
hdfR
sspA
gcvH
ysgD
topA
pal
lpp
NF_4
NF_8
Dim2 (10.9%)
Dim1 (78.7%)
B
rhoL
rbsR
sspA
pyrL
pal
lpp
glpK
NF_4
NF_8
Dim2 (12.6%)
Dim1 (77.9%)
C
rbsR
maoP
rpoS
rbsK
hdfR
rhoL
gcvH
sspA
ysgD
pal
pyrL
lpp
tolA
NF_4
NF_8
Dim2 (19.5%)
Dim1 (62.3%)
D
rhoL
rpoS
rbsR
rbsK
maoP
hdfR
sspA
gcvH
yobF
ysgD
topA
pal
pyrL
lpp
tolA
glpK
NF_4
NF_8
Dim2 (14.4%)
Dim1 (77.3%)

**Figure 11. PCA for *E. coli* W3110 RpoS$^+$.** Fitness data is scaled and centered prior to the analysis. X-axis represents the first PC, Y-axis represents the second PC, and blue arrows show the projections of the variables from the original dataset (combination of state of water and time in days). Shown are (A-C) fitness values for three independent biological replicates incubated in the different batches of lake water, and (D) Fitness values averaged from three independent biological replicates.

An interesting feature is the temporal component of the PCA analyses. The variables describing the filtered and non-filtered water at the day 2 are well-correlated to each other, while the correlation becomes less prominent as the time passes. Neither time nor type of lake water were the explicit variables in the dataset, and the appearance of the temporal components on the biplots confirmed the validity of our approach.

The PCA allowed us to reveal two groups of knockout strains that were different from the majority of the strains. First, we could see the group that was outcompeted in both filtered and non-filtered water. These strains clearly carry the mutations that are very detrimental for growth and survival in aquatic environments. Secondly, there was a group of strains that is reproducibly overrepresented but only in the filtered lake water.

**Table 1. Gene fitness with of *E. coli* W3110 RpoS$^+$ mutant strains in filtered and non-filtered lake water.**

| Gene name | Filtered water | | | Non-filtered water | | | PC1 | PC2 | Euclidean distance | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|
| | Day 2 | Day 4 | Day 8 | Day 2 | Day 4 | Day 8 | | | | |
| *pal* | -2.8 | -3.4 | -3.7 | -3.1 | -3.6 | -2.8 | 27.1 | 3.8 | 27.4 | Peptidoglycan-associated lipoprotein |
| *lpp* | -2.9 | -3.5 | -3.7 | -3.0 | -3.3 | -2.6 | 26.8 | 2.7 | 27.0 | Major outer membrane lipoprotein Lpp |
| *tolA* | -2.3 | -2.3 | -2.6 | -2.1 | -2.2 | -1.6 | 18.5 | 1.2 | 18.5 | Tol-Pal system protein TolA |
| *pyrL* | -1.7 | -2.0 | -1.9 | -1.5 | -2.2 | -1.9 | 15.2 | 2.7 | 15.5 | *pyr* operon leader peptide |
| *prc* | -1.7 | -1.8 | -1.9 | -1.7 | -1.8 | -1.3 | 14.3 | 1.4 | 14.4 | Tail-specific protease |
| *arnF* | -1.8 | -1.7 | -1.8 | -1.5 | -1.6 | -1.4 | 13.9 | 1.3 | 14.0 | Probable 4-amino-4-deoxy-L-arabinose-phosphoundecaprenol flippase subunit ArnF |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *carA* | -1.4 | -1.7 | -1.6 | -1.4 | -1.8 | -1.6 | 13.0 | 2.4 | 13.3 | Carbamoyl-phosphate synthase small chain |
| *pyrD* | -1.4 | -1.7 | -1.6 | -1.3 | -1.7 | -1.5 | 12.7 | 1.9 | 12.9 | Dihydroorotate dehydrogenase |
| *pyrE* | -1.4 | -1.6 | -1.5 | -1.5 | -1.6 | -1.3 | 12.6 | 1.9 | 12.7 | Orotate phosphoribosyl-transferase |
| *carB* | -1.4 | -1.6 | -1.5 | -1.4 | -1.8 | -1.4 | 12.5 | 2.1 | 12.7 | Carbamoyl-phosphate synthase large chain |
| *wecE* | -1.4 | -1.6 | -1.6 | -1.1 | -1.6 | -1.4 | 11.6 | 1.6 | 11.8 | dTDP-4-dehydro-6-deoxy-D-glucose transaminase |
| *wecF* | -1.4 | -1.7 | -1.6 | -1.1 | -1.3 | -1.5 | 11.7 | 1.0 | 11.8 | TDP-N-acetylfucosamine lipid II N-acetylfucosaminyl-transferase |
| *wecB* | -1.3 | -1.5 | -1.6 | -1.1 | -1.4 | -1.5 | 11.5 | 1.8 | 11.6 | UDP-N-acetylglucosamine 2-epimerase |
| *wecC* | -1.3 | -1.5 | -1.6 | -1.1 | -1.5 | -1.4 | 11.4 | 1.6 | 11.5 | UDP-N-acetyl-D-mannosamine dehydrogenase |
| *pyrC* | -1.3 | -1.5 | -1.4 | -1.2 | -1.5 | -1.3 | 11.3 | 1.6 | 11.4 | Dihydroorotase |
| *purM* | -1.6 | -1.7 | -1.7 | -1.1 | -1.0 | -1.0 | 11.2 | -0.6 | 11.2 | Phosphoribosylformyl-glycinamidine cycloligase |
| *purE* | -1.5 | -1.8 | -1.7 | -1.2 | -1.0 | -1.0 | 11.1 | -0.6 | 11.1 | N5-carboxyaminoimidazole ribonucleotide mutase |
| *wzyE* | -1.1 | -1.2 | -1.5 | -0.8 | -1.2 | -1.6 | 9.9 | 1.7 | 10.0 | Enterobacterial common antigen (ECA) polymerase |
| *pyrF* | -1.2 | -1.3 | -1.2 | -1.0 | -1.3 | -1.2 | 9.6 | 1.5 | 9.9 | Orotidine 5'-phosphate decarboxylase |
| *wecA* | -1.1 | -1.3 | -1.2 | -1.0 | -1.2 | -1.0 | 9.5 | 1.0 | 9.6 | Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase |
| *rhoL* | 3.5 | 6.5 | 6.2 | 1.3 | 1.0 | 1.3 | -26.2 | 12.6 | 29.2 | *rho* operon leader peptide |
| *rpoS* | 0.7 | 3.0 | 6.2 | 0.3 | -0.2 | 0.4 | -12.6 | 11.1 | 17.0 | RNA polymerase sigma factor RpoS |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *rbsR* | 0.8 | 2.7 | 5.5 | 0.0 | -0.3 | 0.2 | -11.0 | 10.6 | 15.4 | Ribose operon repressor |
| *rbsK* | 0.7 | 2.4 | 5.1 | 0.1 | -0.3 | 0.1 | -9.9 | 9.7 | 14.1 | Ribokinase |
| *gcvP* | 1.6 | 2.6 | 2.6 | 0.3 | 0.0 | 0.2 | -9.7 | 6.7 | 11.8 | Glycine cleavage system P-protein |
| *gcvT* | 1.6 | 2.5 | 2.5 | 0.4 | 0.1 | 0.1 | -9.8 | 6.4 | 11.7 | Glycine cleavage system T-protein |
| *gcvH* | 1.6 | 2.4 | 2.6 | 0.2 | -0.2 | 0.0 | -8.9 | 7.4 | 11.6 | Glycine cleavage system H protein |
| *maoP* | 0.7 | 2.3 | 2.9 | -0.1 | -0.8 | -0.7 | -5.4 | 9.7 | 11.1 | Macrodomain Ori protein |
| *hdfR* | 0.7 | 2.1 | 2.7 | -0.1 | -0.8 | -0.6 | -4.8 | 9.1 | 10.4 | HTH-type transcriptional regulator HdfR |
| *yobF* | 1.1 | 2.1 | 2.1 | 0.1 | -0.1 | -0.1 | -7.0 | 6.0 | 9.3 | Uncharacterized small protein YobF |
| *sspA* | 0.5 | 1.8 | 1.8 | -0.4 | -1.1 | -0.7 | -2.1 | 8.3 | 8.6 | Stringent starvation protein A |
| *proQ* | 1.2 | 1.7 | 1.4 | 0.1 | -0.2 | 0.1 | -6.0 | 4.6 | 7.6 | RNA chaperone ProQ |
| *seqA* | 1.1 | 1.4 | 1.0 | 0.4 | 0.2 | 0.6 | -6.9 | 1.8 | 7.1 | Negative modulator of initiation of replication |
| *puuR* | 0.7 | 1.4 | 1.6 | 0.2 | 0.0 | -0.1 | -5.1 | 4.3 | 6.8 | HTH-type transcriptional regulator PuuR |
| *iscR* | 0.6 | 1.0 | 1.0 | 0.5 | 0.8 | 0.8 | -6.5 | -0.7 | 6.7 | HTH-type transcriptional regulator IscR |
| *nlpD* | 0.3 | 1.0 | 2.1 | 0.2 | 0.1 | 0.1 | -4.9 | 3.6 | 6.2 | Murein hydrolase activator NlpD |
| *glpF* | 0.4 | 0.4 | 0.3 | 0.3 | 0.7 | 1.5 | -4.8 | -3.1 | 5.8 | Glycerol uptake facilitator protein |
| *dgcE* | 0.6 | 1.0 | 1.2 | 0.0 | -0.3 | -0.6 | -2.6 | 4.8 | 5.5 | Diguanylate cyclase DgcE |
| *uup* | 0.7 | 1.1 | 1.0 | 0.2 | 0.0 | 0.1 | -4.3 | 2.7 | 5.1 | ATP-binding protein Uup |
| *tsaC* | 0.6 | 0.6 | 0.6 | 0.4 | 0.6 | 0.5 | -4.7 | -0.7 | 4.8 | Threonylcarbamoyl-AMP synthase |

Shown are the average fitness values across three biological replicates in the respective points, as well as the values of the principal components and the Euclidean distance from the center. Only top-

20 genes from each group (see above) with the highest value of the distance are shown, sorted by the Euclidean distance. The full version of the table is **Supplementary table 2**.

In the first group, two subgroups were significantly enriched. Many mutants in nucleoside biosynthesis pathway have been outcompeted in both filtered and non-filtered water (**Figure 12**A, purple). Lake water is unlikely to contain copious amounts of dissolved nucleosides, and the bacteria have to rely on their own machinery to synthesize them. It is not surprising that the strains with impaired nucleoside biosynthesis have a disadvantage during the growth in lake water.

Another subgroup of mutations that was underrepresented in the lake water contains either the structural membrane proteins (**Figure 12**A, green) or the proteins involved in the biosynthesis of membrane structures (**Figure 12**A, yellow). The structural proteins are represented by TolA, Pal and Lpp proteins. All three of them are outer membrane proteins, with Lpp being one the most abundant (lipo)proteins in the cell [164]. It anchors the outer membrane to the peptidoglycan layer and defines the distance between the membranes, thus playing the key part in maintaining the outer membrane integrity. Surprisingly, the cells lacking Lpp are viable and have no major growth defect under laboratory conditions. However, they are more sensitive to toxic compounds such as EDTA, antibiotics and detergents, and are known to release the periplasmic proteins into the outer medium [165]. TolA and Pal are the members of Tol-Pal system which is as well important for maintaining outer membrane integrity and takes part in cell division [166]. Mutations in Tol-Pal system lead to the similar phenotype as Lpp mutations [167].

On top of this, the fitness decrease was observed in some strains that are involved in biosynthesis of the enterobacterial common antigen (ECA, **Figure 12**A, yellow). This polymer consists of three amino sugars, N-acetyl-glucosamine, N-acetyl-D-mannosaminuronic acid, and 4-acetamido-4,6-dideoxy-D-galactose [168]. Among other functions of this compound (i.e. virulence in *S. typhimurum* [169], resistance to toxins [170] etc), its involvement in cell membrane integrity was shown. It was demonstrated not only in *E. coli* [171] but also in *Serratia marcescens* [172]. In many strains of *E. coli*, primarily the pathogenic ones, the ECA carries the O-antigen which is important to determine the serotype of the strain [173].

To sum this up, membrane stability is an important factor contributing to cells' viability and ability to proliferate. Loss of the membrane integrity is one of the widely accepted biomarkers for the dead cells [174], and clearly this group of mutants has more fragile and vulnerable membranes. This explains why they were outcompeted and underrepresented in both filtered and non-filtered water.

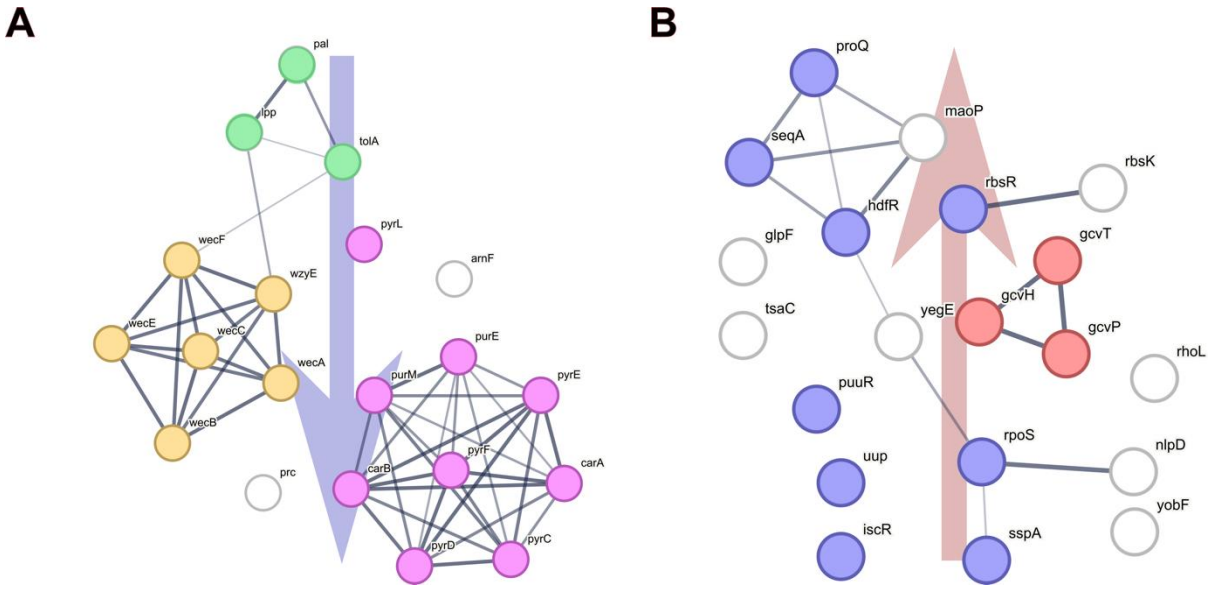



**Figure 12. Top-20 of underrepresented mutations in both filtered and non-filtered lake water (A) and overrepresented in the filtered lake water (B) mutations.** Colored are: structural membrane proteins (green), ECA biosynthesis pathway (yellow), nucleotide biosynthesis pathway (purple), regulatory proteins (blue) and glycine cleavage system (red). The networks are constructed using the STRING database [175]. Thickness of the lines indicates the strength of data support in STRING.

On the opposite side, we were able to identify the group of strains that was reproducibly overrepresented in our experimental results. This effect, however, was only present in the filtered lake water. Therefore, these strains carry mutations that help them to adapt well to the nutrient-poor conditions but do not help them to compete against autochthonic microbiota in the non-filtered lake water.

Among these mutations, we can note three out of four genes of the glycine cleavage system (**Figure 12**B, red). The fourth member of the group, *lpd*, is placed in the different operon and involved in other pathways in *E. coli* and it is probably the reason why it did not appear in our data. This group of proteins catalyzes the oxidation of glycine to carbon dioxide and ammonia, yielding a methylene group accepted by tetrahydrofolate and one NADH. This reaction allows the bacteria to use glycine as a carbon source and not as a proteinogenic

amino acid. In *E. coli*, glycine cleavage is one of the ways to provide one-carbon units for methylation and other biosynthetic reactions [176].

Another major group among the better growing strains were the mutants of the regulatory genes (**Figure 12**B, blue). For instance, the mutation in RNA chaperone *proQ* is beneficial. The regulatory target of *proQ* is protein ProP which acts as a transporter of osmolytes and protects the cells from the hyperosmotic shock [177]. In our experimental setup, the cells rather underwent the hypoosmotic shock and the activity of this transporter might be undesirable.

*rbsR* is a repressor of the ribose catabolism operon. It also suppresses the *de novo* synthesis of purine nucleotides from D-ribose-5-phosphate [178]. With this pathway de-repressed, the Δ*rbsR* strains would have an advantage in nucleotide biosynthesis. This corresponds to the growth defect of the strains with impaired nucleoside biosynthesis. Presence of the ribokinase *rbsK* in the data is likely caused by the polar effect as *rbsK* and *rbsR* are in the same operon.

*puuR* is a repressor of the putrescine degradation pathway. *E. coli* is able to use putrescine and other polyamines as carbon and nitrogen source [179], and its presence in lake water is highly likely. Being able to consume putrescine might be advantageous.

Another example of the beneficial mutation is Δ*hdfR*. It was reported to be a repressor of a flagellar operon, and it is demonstrated that Δ*hdfR* strains have higher expression of flagellar genes and are more motile [134]. One might speculate that in nutrient-poor conditions enhanced motility and chemotaxis might be an advantage. In addition, it was recently found that loss of HdfR function appears to increase NADPH availability in *E. coli* [145], which might as well improve its surviving potential. However, as we demonstrate in detail in the next chapter (3.2. HdfR and MaoP regulate the levels and activity of RpoS), the growth advantage of this strain is due to the lower level of RpoS but not its complete absence.

Surprisingly, Δ*rpoS* strain (lacking the stress response master regulator) was overrepresented in our lake water samples, indicating that it actually persisted better than the majority of strains. This comes in a conflict to the literature data [18, 41], where its knockouts survived worse compared to the wild type (WT) strains.

We need to note here that some of both under- and overrepresented mutations that we identified are situated very close to the origin of replication on the chromosome (**Supplementary figure 3**). For example, the operons of ECA biosynthesis, *hdfR*, *maoP* are

all found in the narrow 33 kb window. It is theoretically possible that these mutations are enriched in the dataset simply due to their proximity to the origin, since the ongoing replication doubles the *ori* region first. However, we think it is unlikely for the following reasons: (1) Both over- and underrepresented mutations are found in the *ori* region. (2) The growth rate in the studied conditions is very low and the replication does not happen often, therefore its contribution is negligible. (3) The fitness calculation algorithm includes normalization to the chromosomal region (5.2. Methods) to counter this problem.

## 2.2. Fluorescence-based screen of the Keio collection

Complementary to our Tn5 approach, we have performed a whole-genome fluorescence-based (FB) screening of growth and survival of *E. coli* based on the Keio collection of all the viable knockouts [160]. To do so, we have transformed all the strains in the Keio collection with a plasmid pTrc99a carrying enhanced green fluorescent protein (eGFP). This plasmid has an artificial isopropyl- β-D-thiogalactopyranosid (IPTG)-inducible pTrc promoter and ColE1 origin of replication [180]. This allowed us to track growth of individual knockout strains in the lake water using a plate reader, and differentiate them from the autochthonic microbiota at the same time.

The cultures of *E. coli* were grown overnight at 37°C on TB medium in a rotary shaker in the 96-deepwell plates. Then they were inoculated in 96-well plates filled with native (taken from the lake without any additional treatment) or filter-sterilized lake water supplemented with 50 μM of IPTG. We did not add the antibiotic to the water to avoid harm to the native microbiota of the lake. Then the cultivation took place at room temperature with daily measurements of turbidity and intensity of GFP fluorescence.

### 2.2.1. Proof of concept

To start with, we have performed this experiment on just one plate from the Keio collection to confirm the validity of the method (**Figure 13**). This allowed us to confirm several assumptions that we had before starting this part of the work.

First, this method indeed allows us to track the bacteria over time and observe the differences both between the strains and between the filtered and non-filtered water. Second, we observed that the same strain performs better in filtered than in non-filtered water. This corresponds to the literature data, where *E. coli* grew better in sterile soil [14] compared to non-

sterile. Third, as confirmed with the flow cytometry measurements, the loss of the plasmid does not occur at the noticeable rate. The absence of an antibiotic combined with the presence of an inductor makes the favorable condition for the plasmid loss. Due to the low growth rate, the bacteria were able to maintain the stable level of GFP fluorescence over the course of the cultivation. These results allowed us to scale the experiment up to the whole collection.

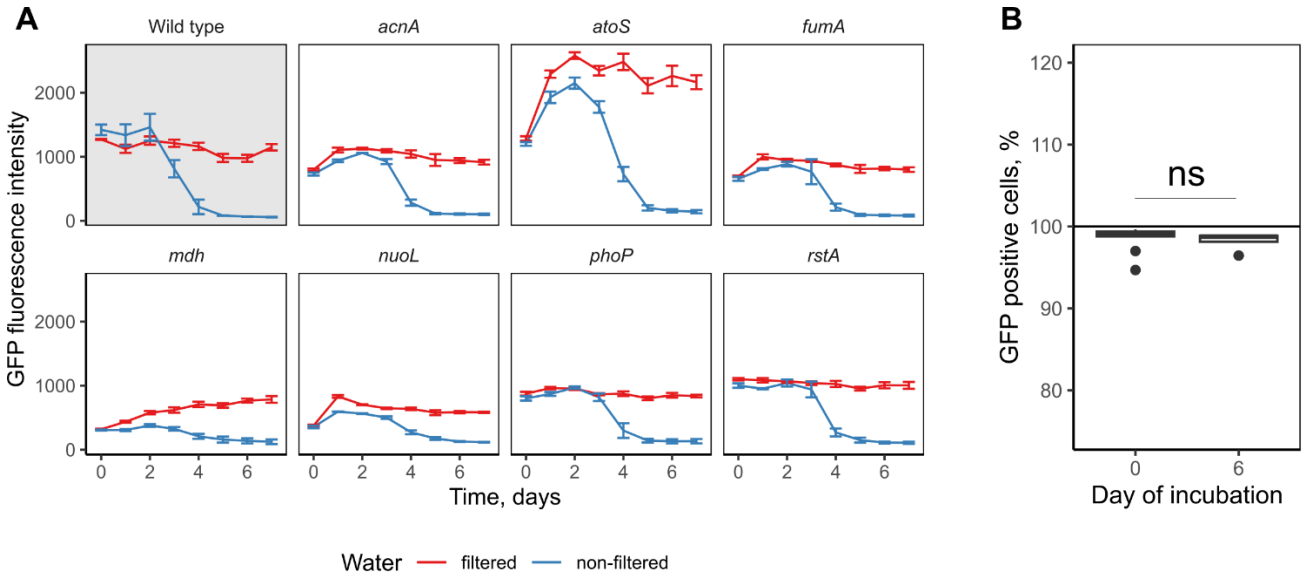



**Figure 13. Pilot experiment, plate 3 of Keio collection.** (A) Shown are absolute values of GFP fluorescence intensity (mean and standard error of mean) measured over the course of 8 days in technical triplicates. Only the representative strains are shown, the complete version of the plot is **Supplementary figure 4**. (B) The percentage of GFP-positive cells in the WT in the filtered lake water in the beginning (day 0) and the end (day 6) of the incubation. The percentage is determined by flow cytometry. Shown are the boxplot based on 7 biological replicates.

### 2.2.2. Global screen of the Keio collection

We have then performed a genome-wide screening of *E. coli* persistence in the lake water. The Keio collection was split in three equal parts and every part was screened twice. In the resulting dataset, we have normalized the GFP fluorescence values of each strain to its respective initial point. On top of that, we discarded the strains that grew notably worse than the WT strain (=yielding less than 30% of the WT's fluorescence at the time point 0).

The bacteria tend to die out in the non-filtered water but maintain their population in the filtered water. On top of that, there is variability of growth or decline between the samples of water (**Figure 14**). This can be explained by the fact that the water samples were taken at the different time points and both chemical composition of the water and its community of bacteria, fungi and protists vastly changed.

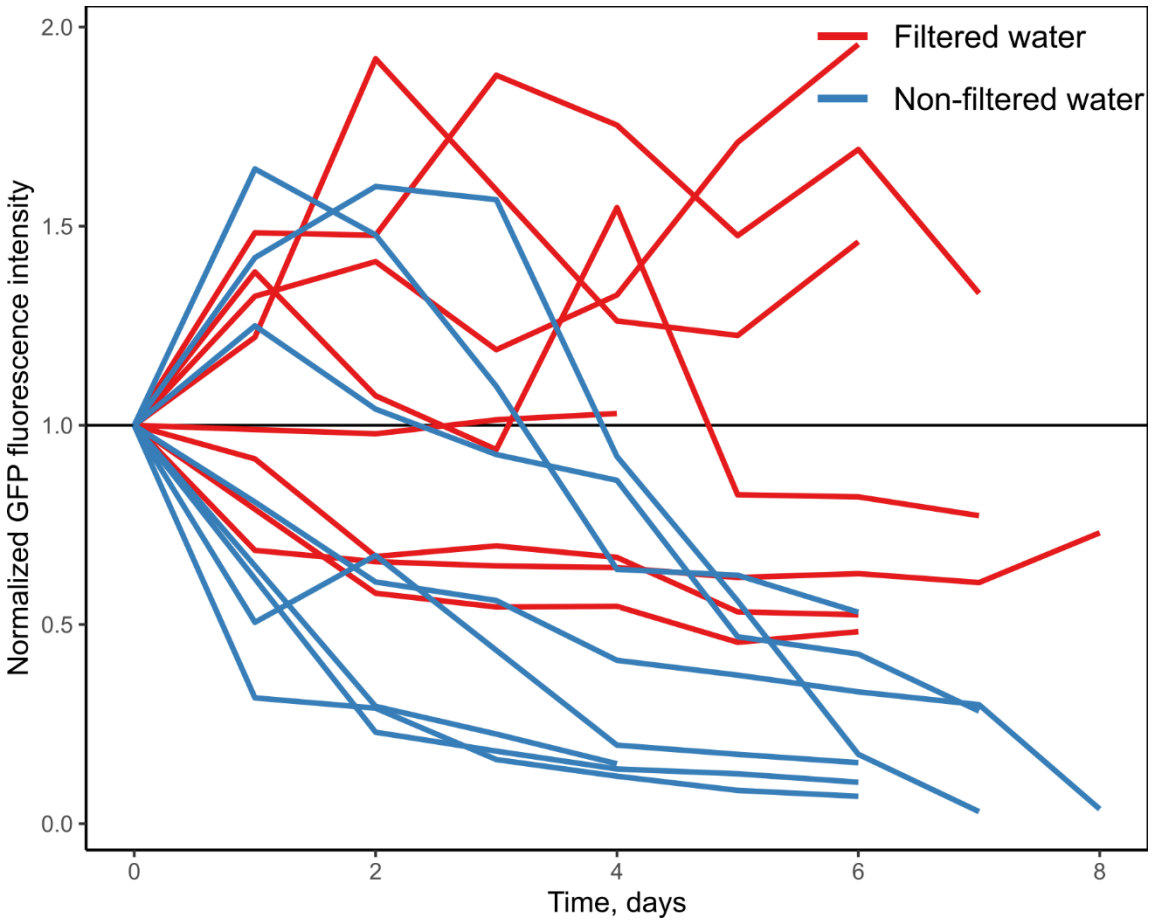



**Figure 14. Growth of the wild type Keio strain in different samples of non-filtered and filtered water**. The GFP fluorescence data is normalized to the initial point. Different lines represent different biological replicates (=different samples of water). See **Supplementary figure 5** for more details.

The bacteria die out faster in non-filtered water if it was taken at warmer temperature (**Supplementary figure 5**), and this applies to both water temperature and air temperature. This is probably caused by more favorable conditions for the autochthonic microbiota in the warmer water.

After the normalization of the GFP fluorescence values to the initial point of each respective strain, we have then normalized the data to the average value in each time point, rescaled it and applied the principal component analysis to the resulting dataset. The approach of principal component analysis is described in greater detail in the previous chapter (2.1. Transposon mutant library screening).

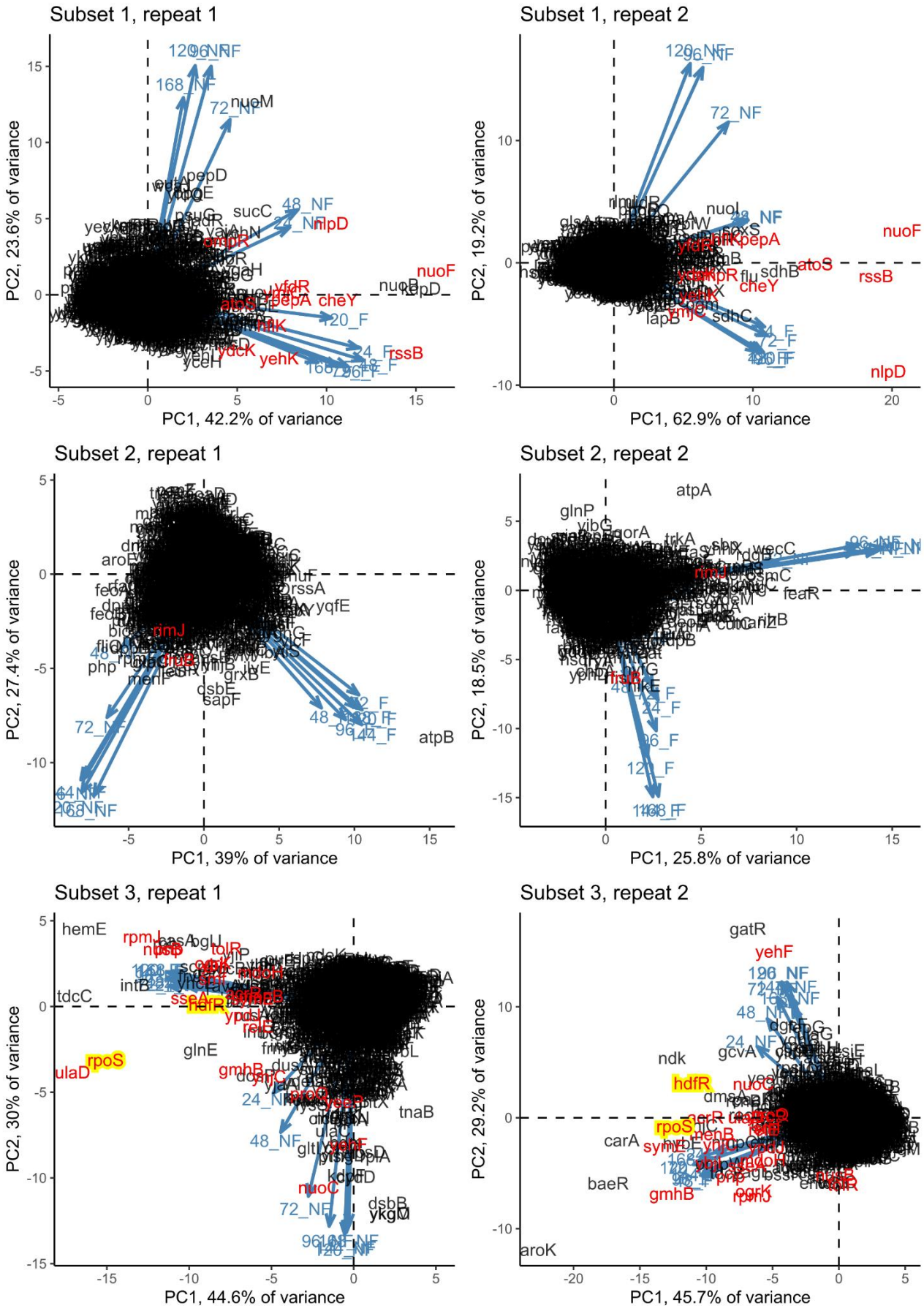

Subset 1, repeat 1
PC2, 23.6% of variance
PC1, 42.2% of variance
Subset 1, repeat 2
PC2, 19.2% of variance
PC1, 62.9% of variance
Subset 2, repeat 1
PC2, 27.4% of variance
PC1, 39% of variance
Subset 2, repeat 2
PC2, 18.5% of variance
PC1, 25.8% of variance
Subset 3, repeat 1
PC2, 30% of variance
PC1, 44.6% of variance
Subset 3, repeat 2
PC2, 29.2% of variance
PC1, 45.7% of variance

**Figure 15. PCA results for the FB screen.** Data is scaled and centered prior to the analysis. X-axis represents the first PC, Y-axis represents the second PC, and blue arrows show the projections of the variables from the original dataset (combination of time in hours and state of water). Knockouts selected for the follow-up screen are highlighted in red and knockouts selected for the in-depth screen are highlighted in yellow.

Similarly to the transposon mutagenesis data, most of the mutations in the Keio collection did not result in better or worse growth in the lake water. In fact, surprisingly, the WT strain actually grew slightly worse than the average strain in the collection. On top of that, some strains actually performed better than the wild type strain and the majority of the strains (**Supplementary figure 6**). This might partially be explained by filtering out the strains that grew poorly in the rich medium.

In our PCA results, the correlation between the variables describing the filtered water is very high (**Figure 15**). These variables also mostly contribute to the first principal component, and therefore most of the variability of the dataset is explained by the filtered water data. On the other hand, the variables describing the non-filtered water are less correlated to each other and the temporal component is visible at the plots, similar to the PCA results for TnSeq. The "non-filtered" variables contribute to the second principal component. The important aspect is that the variables describing the filtered water and the variables describing the non-filtered water appear nearly orthogonal at the biplots and this result is reproducible between all the experiments.

The first three PCs combined capture 70-80% of the variation between the samples (**Figure 15**, **Supplementary figure 2**). Therefore, for the further analysis the rest of the components can be omitted. As a cutoff we used the Euclidean distance between the origin and the position of the respective gene in the space formed by the first three PCs. After that, we selected 5% of the furthest strains in every experiment, and selected those appearing in the both repeats of the same subset of the Keio collection.

Surprisingly, only the strains that performed better than the WT strain were able to pass the cutoff. We were not able to identify the strains that grew worse than the majority of the strains.

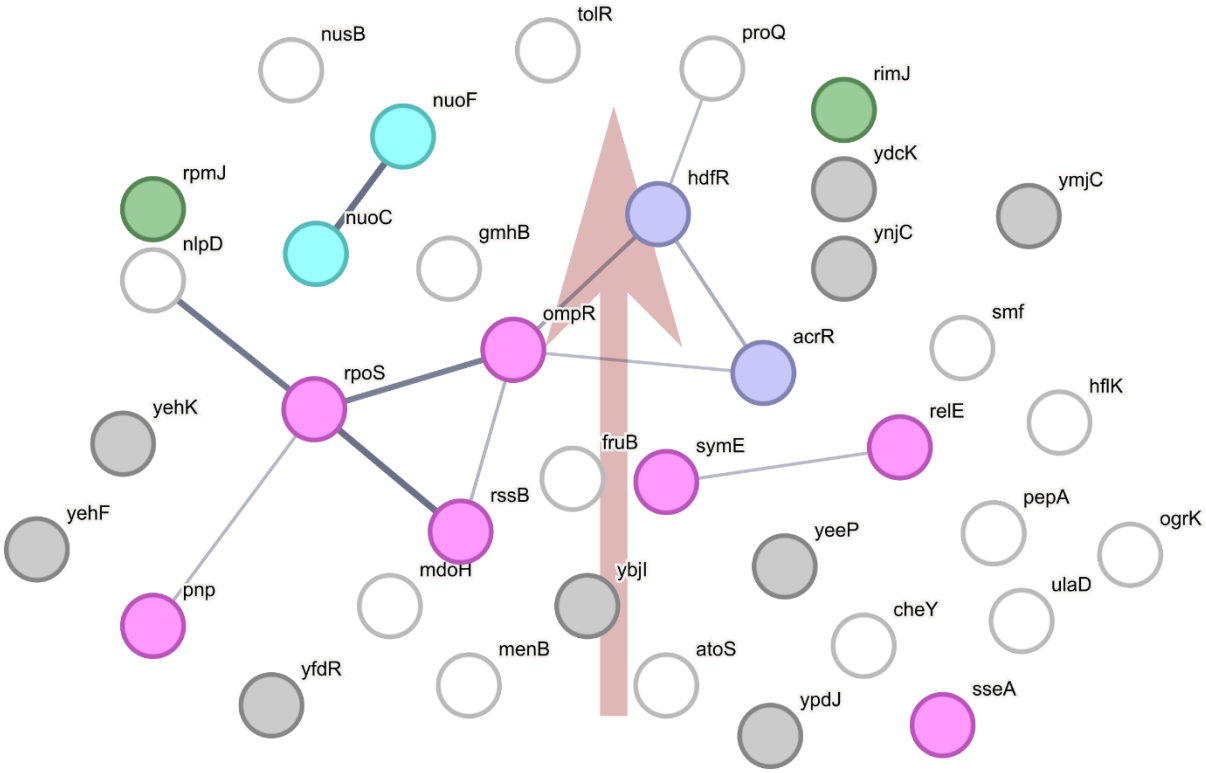



**Figure 16. STRING diagram for the mutants persisting better in the lake water.** Colored are: ribosome (green), respiratory chain (cyan), stress response (purple), motility enhancers (blue) and genes with unknown function (gray). The networks are constructed using the STRING database [175]. Thickness of the lines indicates the strength of data support in STRING.

**Table 2. The selected genes based on the distance and occurrence in the both biological replicates**.

| Gene | Description | Euclidean distance, first experiment | Euclidean distance, second experiment | Average Euclidean distance |
|---|---|---|---|---|
| *nuoF* | NADH-quinone oxidoreductase subunit F | 16.4 | 20.9 | 18.8 |
| ***nlpD*** | **Murein hydrolase activator NlpD** | **11.5** | **21.7** | **17.3** |
| *rssB* | Negative regulator of RpoS | 15.0 | 18.9 | 17.0 |
| ***rpoS*** | **RNA polymerase sigma factor RpoS** | **17.2** | **13.2** | **15.3** |
| *ulaD* | 3-keto-L-gulonate-6-phosphate decarboxylase UlaD | 18.0 | 7.0 | 13.7 |
| *yehF* | Uncharacterized protein YehF | 8.1 | 15.9 | 12.6 |
| *rpmJ* | 50S ribosomal protein L36 | 13.9 | 10.0 | 12.1 |

| | | | | |
|---|---|---|---|---|
| *gmhB* | D-glycero-beta-D-manno-heptose-1,7-bisphosphate 7-phosphatase | 7.8 | 14.4 | 11.6 |
| *cheY* | Chemotaxis protein CheY | 10.8 | 11.6 | 11.2 |
| *pnp* | Polyribonucleotide nucleotidyltransferase | 12.1 | 9.8 | 11.0 |
| *atoS* | Signal transduction histidine-protein kinase AtoS | 5.6 | 14.5 | 11.0 |
| *symE* | Toxin-like protein of the SOS response | 6.3 | 14.0 | 10.8 |
| ***hdfR*** | **Potential negative transcriptional regulator of *flhDC*** | **9.0** | **11.7** | **10.4** |
| *ybjI* | 5-amino-6-(5-phospho-D-ribitylamino)uracil phosphatase YbjI | 8.8 | 10.7 | 9.8 |
| *nuoC* | NADH-quinone oxidoreductase subunit C/D | 11.1 | 8.2 | 9.8 |
| *pepA* | Cytosol aminopeptidase and DNA-binding transcriptional repressor | 8.6 | 10.7 | 9.7 |
| *nusB* | Transcription antitermination protein NusB | 12.3 | 5.7 | 9.6 |
| *ogrK* | Prophage late control protein OgrK | 8.9 | 9.6 | 9.3 |
| *sseA* | 3-mercaptopyruvate sulfurtransferase | 10.1 | 8.2 | 9.2 |
| *acrR* | HTH-type transcriptional repressor AcrR | 7.2 | 10.1 | 8.8 |
| *ynjC* | Putative inner membrane ABC transporter permease protein YnjC | 6.8 | 9.3 | 8.1 |
| *menB* | 1,4-dihydroxy-2-naphthoyl-CoA synthase | 5.7 | 9.9 | 8.1 |
| *yehK* | Uncharacterized protein YehK | 9.2 | 6.7 | 8.1 |
| *tolR* | Tol-Pal system protein TolR | 8.7 | 7.0 | 7.9 |
| *hflK* | Modulator of FtsH protease HflK | 7.3 | 8.3 | 7.8 |
| *smf* | Uncharacterized protein Smf | 8.8 | 5.5 | 7.3 |
| *yfdR* | CPS-53 (KpLE1) prophage, uncharacterized protein YfdR | 8.2 | 6.0 | 7.2 |
| *ymjC* | Uncharacterized protein YmjC | 7.7 | 6.6 | 7.2 |

| | | | | |
|---|---|---|---|---|
| *ompR* | DNA-binding dual transcriptional regulator OmpR | 5.9 | 7.4 | 6.7 |
| *relE* | Qin prophage gene, involved in starvation stress | 6.1 | 7.1 | 6.6 |
| *mdoH* | Glucans biosynthesis glucosyltransferase H | 6.1 | 6.9 | 6.5 |
| *ypdJ* | CPS-53 (KpLE1) prophage, putative uncharacterized protein YpdJ | 6.9 | 6.1 | 6.5 |
| *ydcK* | Uncharacterized acetyltransferase YdcK | 6.2 | 6.4 | 6.3 |
| *fruB* | Multiphosphoryl transfer protein | 5.6 | 6.5 | 6.1 |
| ***proQ*** | **RNA chaperone ProQ** | **6.2** | **5.5** | **5.9** |
| *rimJ* | [Ribosomal protein S5]-alanine N-acetyltransferase | 5.8 | 5.8 | 5.8 |
| *yeeP* | CP4-44 prophage, putative uncharacterized protein YeeP | 5.6 | 5.9 | 5.8 |

The distance cutoff is top 5% for each of the experiments, and the intersection was applied. The strains appearing in the TnSeq results are highlighted in bold. Colored are: ribosome (green), respiratory chain (cyan), stress response (purple), motility enhancers (blue).

Many of the knockout strains that we have identified here belong to the various regulatory pathways [181], including stress response [80]. The fact that the strains with impaired stress response are capable persist in the lake water conditions better than the WT strain is counterintuitive and contradicts to the literature data [41, 47]. However, there are reports [118] that Δ*rpoS* mutants arise in the evolutionary experiments on *E. coli* in low-carbon and low-nitrogen environments and these mutations are often beneficial in these conditions due to the sigma factor competition. Interestingly, growth and GFP production of Δ*rssB* is similar to those of Δ*rpoS,* despite RssB being the negative regulator of RpoS activity. However, it is common for Δ*rssB* mutants to lose RpoS activity as well because too high levels of RpoS are very detrimental for the cell [80].

Some specific stress response genes are represented in the resulting list. For instance, *ompR* is a master regulator of osmoregulation and plays the key role in response to both acid and osmotic stress responses [182], and *sseA* protects the cell from the buildup of ROS by generating $H_2S$ which binds $Fe^{2+}$ [183].

On top of that, it is interesting to see here several genes with an unclear function (i.e. *yphD, ogrK, yphD* etc.). Some of these genes are actually prophage genes (*yeeP*, *ypdJ, relE*). The role of prophages in the *E. coli* stress response is quite controversial. On one hand, certain prophages are able to detect host's response to various stresses and use these signals to trigger their own lytic cycle [184]. On the other hand, the cryptic prophages of *E. coli* are shown to enhance its resistance to oxidative stress, antibiotics, pH shifts and other environmental stresses [185]. In our results, however, we rather saw that the prophage knockouts grow better than the WT strain, which supports the hypothesis that prophages might kill their host under the stress conditions.

Another peculiar feature of this dataset is the gene *rpmJ* encoding 50S ribosomal protein L36. The ribosomal genes are the housekeeping genes, and it is quite unusual for them to be able to grow better than the WT strain. However, as it was shown, this particular protein is replaced by its paralog encoded by *ykgO* in the late stationary phase. There is the tendency towards YkgO having more contacts with the rest of the ribosome than RpmJ [186]. Consequently, this replacement might make the ribosome more stable in the late stationary phase and under the stress conditions. We have also seen Δ*ykgO* mutant growing better in non-filtered water in one of the experiments (**Figure 15**, subset 2, repeat 1) but this result was not reproducible.

Interestingly, there are some members of the complex I of the respiratory chain in this list, namely *nuoF* and *nuoC.* Null-mutations of all the *nuo*-genes lead to the growth defect under aerobic conditions in rich medium because complex I was unable to assemble properly and therefore inactive [187]. This implies that aerobic respiration was not the limiting factor of survival in our setup, possibly due to the low growth rate.

AtoS is a sensor part of AtoSC two-component system [181, 188]. This system regulates metabolism of short chain fatty acids in response to the presence of acetoacetate, and it positively mediates biosynthesis of poly-(R)-3-hydroxybutyrate in response to the short-chain fatty acids. Poly-(R)-3-hydroxybutyrate is the major *E. coli* energy storage compound, and AtoSC mutants produce less of it [188]. We observe very high level of GFP fluorescence in the Δ*atoS* strain (**Figure 13**). Potentially, this might happen if the resources normally spent on the biosynthesis of poly-(R)-3-hydroxybutyrate are spend to produce GFP instead.

HdfR was described as a negative regulator of the flagellar master operon *flhDC* [134]. The disruption of *hdfR* enhances the activity of *flhDC,* and the strains of *E. coli* that have enhanced activity of *flhDC* are more motile. In the second chapter of this thesis, we will describe the mutants in this locus in the greater detail, and demonstrate that their growth advantage is a consequence of lower RpoS levels.

To verify our results, we have created the knockouts of the selected genes in BW25113 Δ*luxS* background strain, further referred as the WT, and repeated the experiment with them and Keio collection mutants. LuxS is a key enzyme of the biosynthesis of autoinducer 2, the only quorum sensing signal of *E. coli* [189]. BW25113 Δ*luxS* was chosen because it was motile (unlike the parental strain) and absence of LuxS did not have significant effects on its growth in lake water. By using the Δ*luxS* strain, we wanted to reduce the autoaggregation to facilitate the flow cytometry measurements as this behavior is quorum sensing-dependent in *E. coli* [190].

We have selected two knockouts for the further study, namely Δ*hdfR* and Δ*rpoS*. These two strains both grew better in the filtered lake water according to both of our screens. Therefore, these two genes are both especially important in the context of the lake water and deserve the deeper investigation.

Despite the difference from the WT strain and appearance in both screens, we decided to omit the *ΔnlpD* strain because of its genomic context. Genes encoding *nlpD* and *rpoS* are located on the chromosome right next to each other, and in fact one of *rpoS*'s promoters is located in the middle of *nlpD* [76, 191]. To check if expression of *rpoS* is influenced by *nlpD* deletion we have performed a proteomic analysis that has shown that *ΔnlpD* is in fact a double knockout of *rpoS* and *nlpD* (**Supplementary figure 8**). We also omitted the *ΔproQ* strain because the effect of this deletion was mild in comparison to Δ*hdfR* and Δ*rpoS* in the initial screen, and there were no differences from the wild type in the confirmatory screen we performed (**Supplementary figure 7**).

Since *hdfR* is reported to be under the $\sigma^S$-dependent promoter [148], we have tested whether the influence of *hdfR* and *rpoS* deletions is epistatic (**Figure 17**). To do so, we have tracked the GFP fluorescence of the double knockout in filtered lake water to the respective single knockouts. Indeed, the fluorescence of the double knockout was higher compared to the WT strain and to the both single knockouts, and the effect of these mutations is rather additive.

Therefore, the effects of *hdfR* and *rpoS* on growth and persistence in the nutrient-depleted conditions have to follow the different pathways.

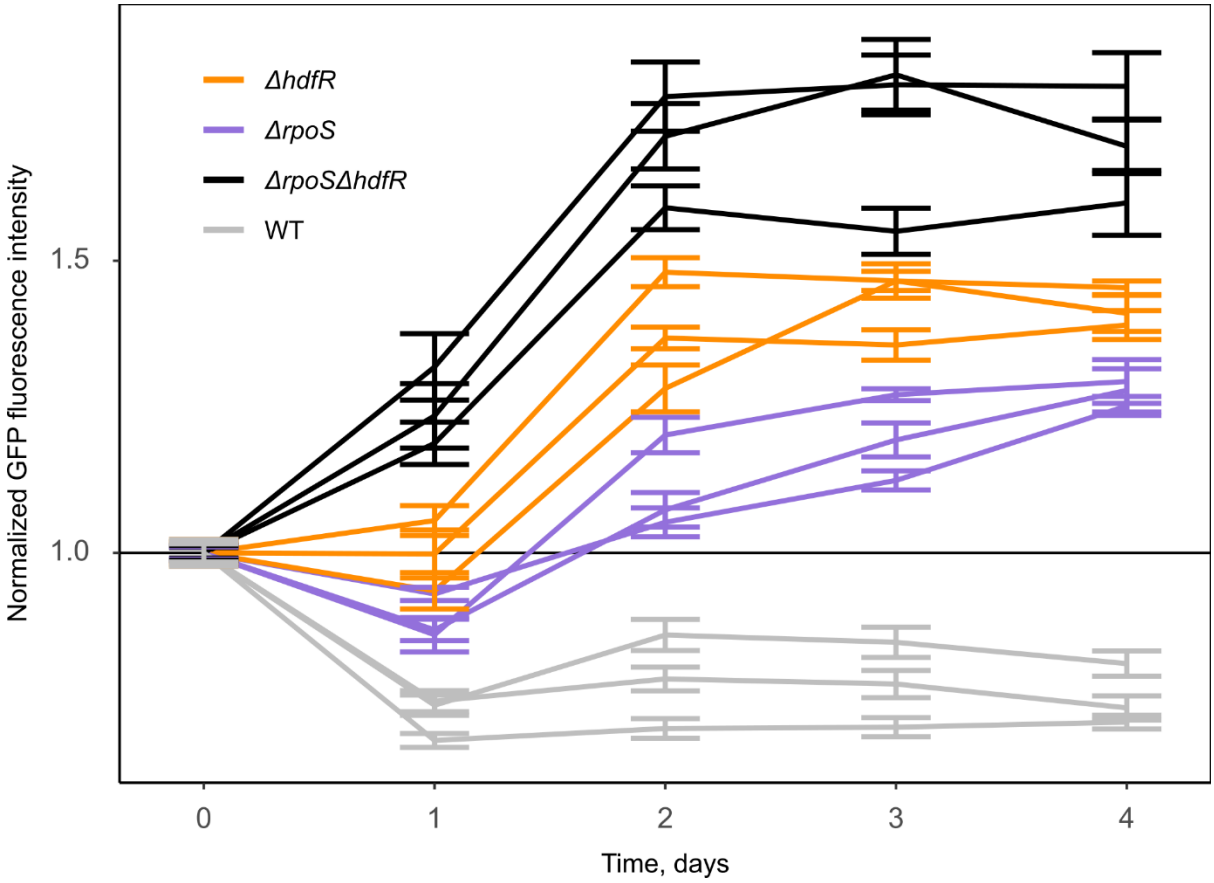



**Figure 17. Effect of *hdfR* deletion in combination with *rpoS* deletion**. The GFP fluorescence data is normalized to the initial point. Lines represent different samples of lake water. Error bars represent the standard error of the mean.

### 2.2.3. The mutant strains produce more GFP per cell than the wild type

To collect more information on growth of these strains, we have performed the flow cytometry measurements in the filtered lake water. The unwashed overnight cultures of strains of interest grown in TB at 37°C were inoculated to the lake water in 24-well plates. We have measured total count of cells per ml, count of GFP-positive cells per ml, GFP fluorescence of the single cell, colony forming units (CFU) per ml and a copy number of plasmids per cell. These parameters were measured daily except for CFU (at the last day of cultivation) and the copy number of plasmids per cell (at the beginning and end of the cultivation). All these experiments were performed in the filtered lake water. As a control, some measurements were additionally performed in a minimal medium M9 without carbon source (M9H) and in a minimal medium M9 + 0.04% sodium succinate (M9S).

We have immediately noticed a dramatic difference between the WT strain and the selected knockout strains. Both of the strains selected for the in-depth screen have higher GFP fluorescence per cell than the WT. Over the course of the incubation, the GFP amount per cell slightly decreases in the WT strain and in the Δ*hdfR* strain but it tends to increase in the Δ*rpoS.*

The cell counts measured with the flow cytometer, however, do not show a striking difference between the knockout strains and the WT strain. Therefore, the increase in GFP fluorescence that was observed during the initial screen is explained not only by the higher quantities of cells but primarily by higher GFP expression. The counts of the GFP-positive cells follow the same pattern as the total counts, and more than 90% of the cells remain fluorescent to the end of the experiment (**Figure 13**, **Figure 18**).

We observed that the number of cells in the filtered lake water actually slightly increases despite the medium being very poor. The possible explanation is presence of the residual amino acids transferred from the overnight culture. However, we tested the persistence of the strains in the defined medium without carbon source (M9H) and discovered that the cell counts were slightly lower in these conditions. This data indicates that the lake water contains a carbon source beside the transferred amino acids that supported the growth of *E. coli* in our experimental conditions.

If the cells are allowed to grow and divide in the presence of 0.04% sodium succinate (in M9S), the cell counts increased approximately tenfold. There were no observable differences between the strains. GFP fluorescence rapidly dropped after just one day of the incubation and remained very low until the end of the experiment. The cells remained fluorescent, and therefore this decrease cannot be explained by the plasmid loss but rather by adaptation to the non-optimal carbon source.

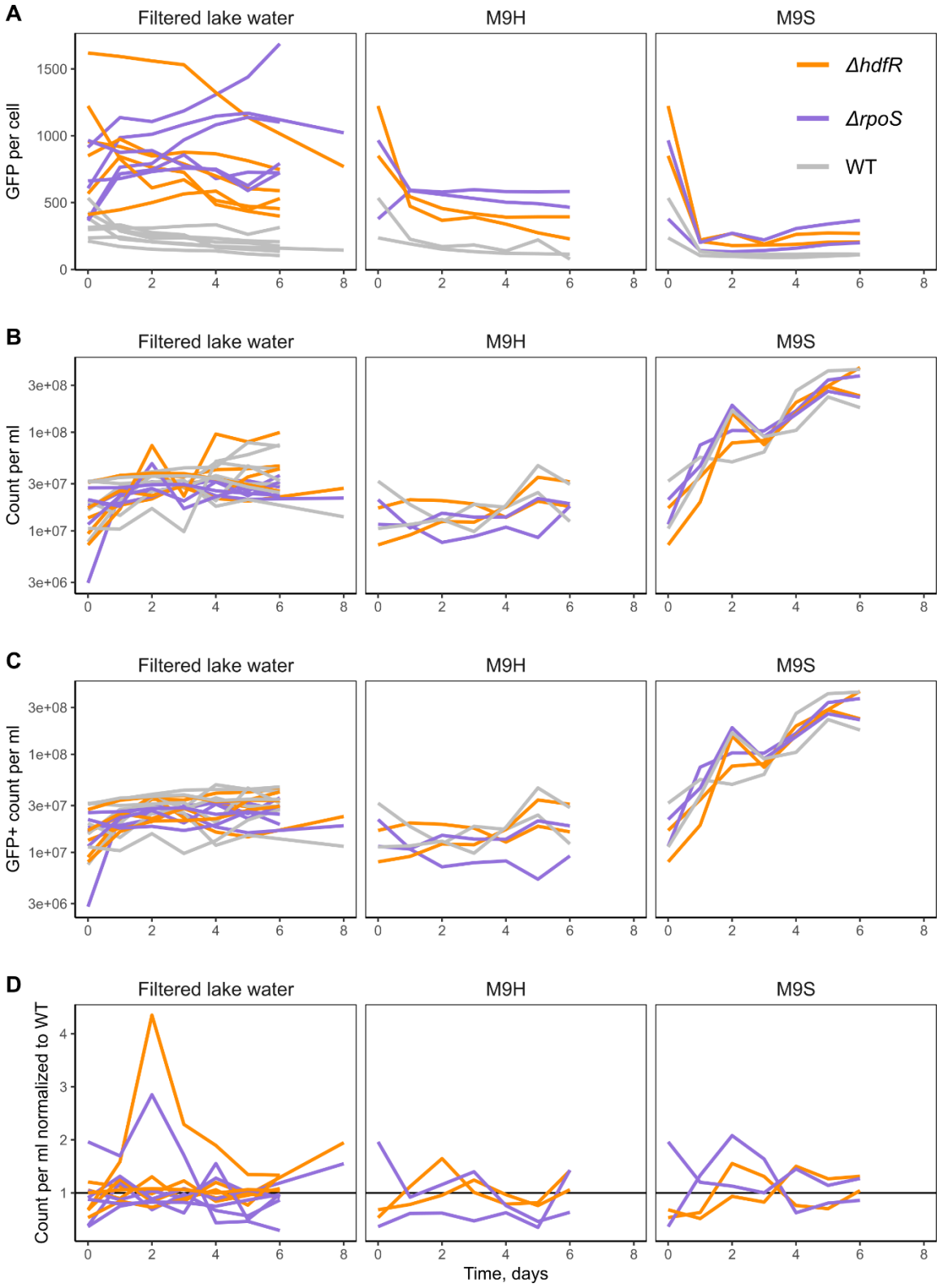



**Figure 18**. **The selected mutants produce more GFP per cell but do not have a significant growth advantage.** (A). Median GFP fluorescence of the selected strains measured over time in the filtered lake water and defined minimal medium. (B). Total counts of cells in 50 µl of $10^{-1}$ dilution of sample. (C). Counts of GFP-positive cells in 50 µl of $10^{-1}$ dilution of sample. (D). Total counts of mutant cells normalized to the values of the WT. Lines represent differenf repeats of the experiment.

### 2.2.4. The increased GFP production is caused by the elevated plasmid copy number

To find out the cause of the higher GFP production observed earlier, we have performed the qualitative PCR (qPCR) assay to assess the copy number of plasmids per cell in the exponential phase in TB, stationary phase in TB and after the 7 days of incubation in the lake water (**Figure 19**, **Table 3.**). Additionally, we compared these results to GFP per cell measurements performed on flow cytometer. As internal controls, we used the same strains without plasmid and with a low-copy plasmid pUA66 carrying a GFP-reporter of promoter activity of ribosomal gene *rplN* [192]. This is a strong promoter and GFP production per cell is comparable to our experimental conditions.

To measure the plasmid copy number, we have performed a real-time qPCR analysis (5.2. Methods). Briefly, each sample was boiled to disrupt the cells and diluted, and then used as a template for amplification. We used two pairs of primers, to either plasmid-borne GFP or a chromosomal gene *dxs.* To determine the number of plasmids per chromosome, we used the formula $plasmid\_copy = 2^{-(Cq_{gfp} - Cq_{dxs})}$, where Cq is a quantification cycle for *dxs* and *gfp* respectively [193].

In case of pTrc99a backbone, the copy number is high and highly variable both between the strains and the experimental conditions, as it is a high-copy plasmid with relaxed regulation [194]. Both the selected strains have more pTrc99a plasmids per cell than the WT especially outside of the exponential phase. The copy number of pUA66 is more robust as expected from a tightly-regulated low-copy plasmid [195].

The two strains we studied responded differently to the two plasmids. In case of Δ*rpoS,* the copy number of both plasmids was elevated, and the production of GFP per plasmid was also elevated in case of pUA66 and remained comparable to the WT strain in case of pTrc99a. The increased amount of GFP per cell therefore cannot be solely explained by the increased number of plasmids. We can conclude that in the case of Δ*rpoS* the expression of our GFP promoter was higher than in the WT strain or Δ*hdfR.* The expression of GFP from both plasmids is governed by $\sigma^{70}$ and the absence of $\sigma^{S}$ might result in the observed increase of the GFP production.

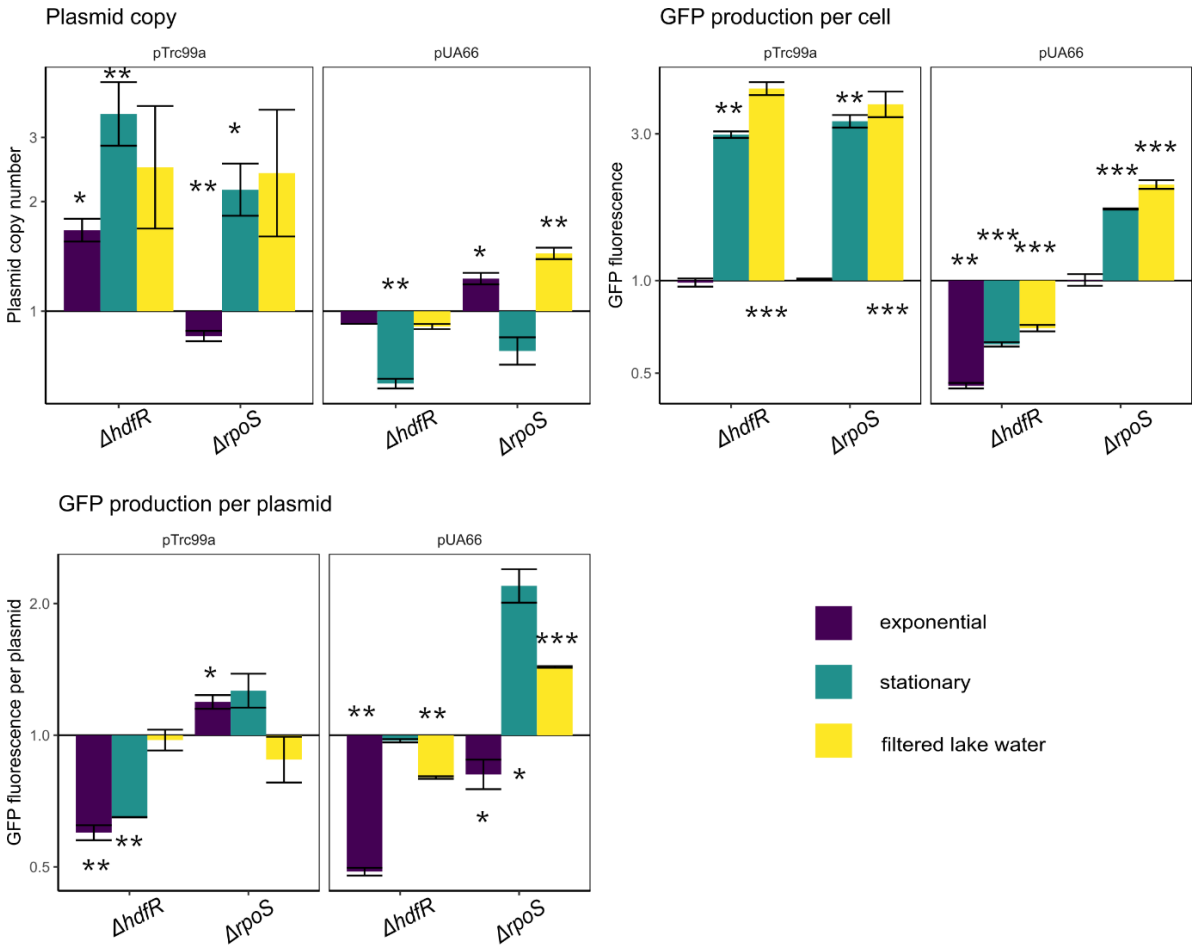



**Figure 19. Plasmid copy number and GFP production per cell and per plasmid normalized to the respective values of the wild type strain.** pTrc99a carries GFP under the Trc promoter, pUA66 carries GFP under the *rplN* promoter. Both exponential and stationary cultures were grown in TB, and the medium was supplemented with 50 µM IPTG for the pTrc99a plasmid. The Y-axis is $\log_{10}$. The error bars represent the standard error of mean. The absolute values are in the table below.

**Table 3. Plasmid copy in various strains and conditions.**

| Sample | strain | Plasmid copy | | GFP per plasmid, AU | | GFP per cell, AU | |
|---|---|---|---|---|---|---|---|
| | | pTrc99a | pUA66 | pTrc99a | pUA66 | pTrc99a | pUA66 |
| Exponential phase | Δ*hdfR* | 48±4.2 | 4±0 | 3.5±0.3 | 56.2±0.4 | 160.5±2 | 225±1.7 |
| | Δ*rpoS* | 24.3±0.4 | 5.3±0.2 | 6.8±0 | 93.9±4.7 | 164.5±2 | 496.5±7.8 |
| | WT | 28.5±0.5 | 4.3±0 | 5.7±0.2 | 115.5±3.2 | 163±2.9 | 495±13.9 |
| Stationary phase | Δ*hdfR* | 116.7±2.3 | 2±0 | 5±0.2 | 125.8±2.2 | 570±13.9 | 251.5±4.3 |
| | Δ*rpoS* | 73±4 | 2.5±0.3 | 9.6±0.5 | 288.4±32.4 | 631±29.4 | 693±2.3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | WT | 36.7±7.7 | 3.2±0.1 | 7.6±0.3 | 129.5±3.3 | 191±0 | 405.5±0.3 |
| Filtered lake water | Δ*hdfR* | 258.2±8.9 | 4.8±0.1 | 3±0.16 | 68±1 | 745±7.5 | 315.5±3.8 |
| | Δ*rpoS* | 247.5±5.5 | 7.7±0 | 2.7±0 | 121.7±2.2 | 661±24.3 | 925±19 |
| | WT | 141.2±50 | 5.3±0.2 | 3.1±0.3 | 85±1.9 | 178±10.4 | 450.5±5.5 |

Shown are the average values and standard errors of mean.

In case of Δ*hdfR* strain, the GFP level per cell is only higher if the plasmid present is a high-copy pTrc99a. The copy number of pTrc99a was elevated, while the copy number of pUA66 was decreased or remained the same as the WT. For both the plasmids, the production of GFP per plasmid was lower than the WT. These results lead us to the conclusion, that Δ*hdfR* produced more GFP per cell in our setup because it had higher number of pTrc99a. We speculate that HdfR might be involved in plasmid replication and segregation, although such a function of this protein is not reported.

### 2.2.5. The Δ*rpoS* strain is less viable than the wild type strain

To assess the actual viability of the selected strains, we have performed the CFU measurements of the samples at the end of the cultivation in the lake water, and used the total counts from the flow cytometer to calculate the percentage of alive cells. The results from CFU are highly variable between the water samples but despite this variability the knockout strains reproducibly survive less or at the same level with the WT. This is especially visible in Δ*rpoS* (**Figure 20**). These results correspond to the data from the literature that Δ*rpoS* mutant survived less in the soil [18, 47] due to the downregulation of the stress response pathways, and to the other results that the total counts don't change significantly unlike the numbers of the viable cells [18].

To confirm these results, we have switched to live-dead staining and consequent flow cytometry measurements, as it is less laborious method of viability assessment. As internal controls, we used the same strains without plasmid and with a low-copy plasmid pUA66 carrying a GFP-reporter of promoter activity of ribosomal gene *rplN* [192]. This is a strong promoter and GFP production per cell is comparable to our experimental conditions. Moreover, the experiment was performed in the exponential phase in TB, stationary phase in TB and at the end of the cultivation in the lake water.

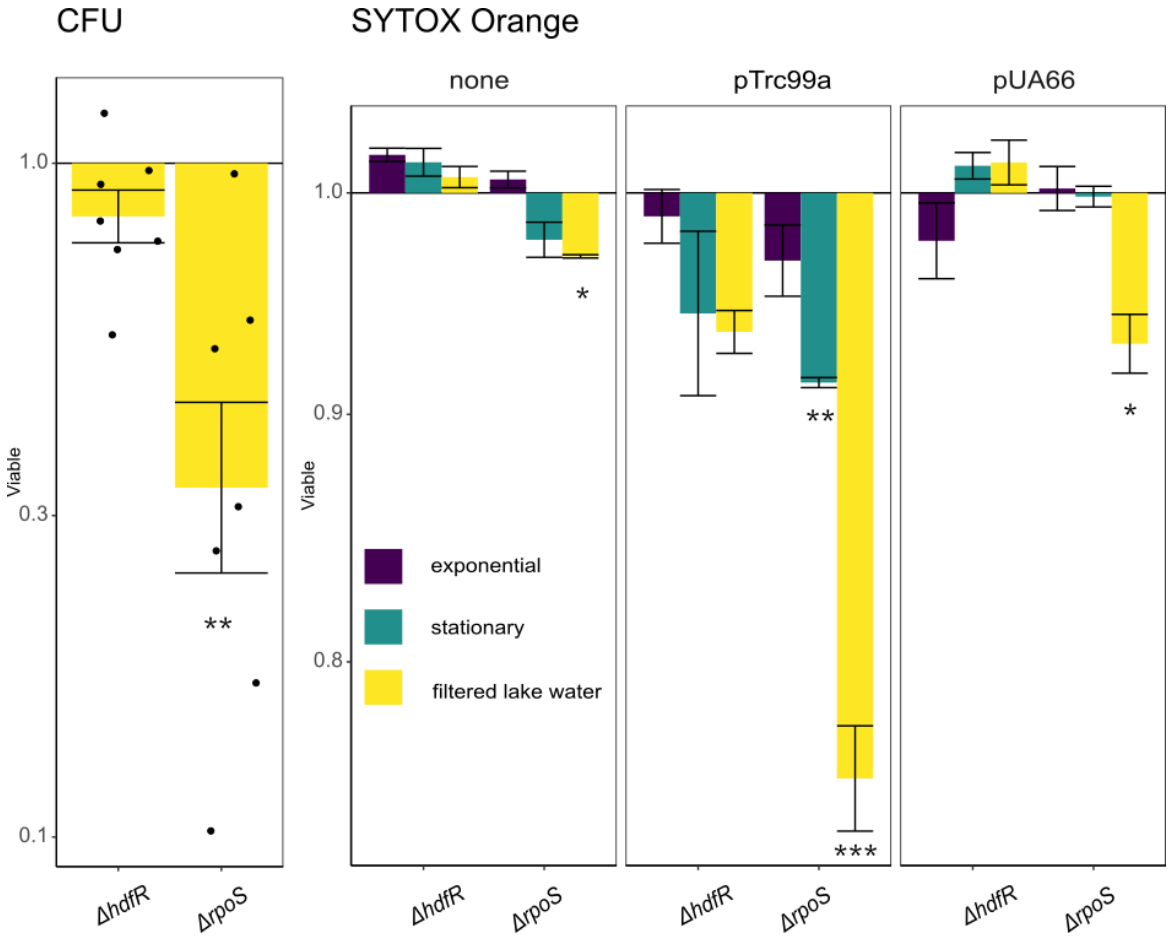



**Figure 20. Viability of the selected strains measured by CFU at the end of the cultivation in the lake water (A) or SYTOX Orange staining in various conditions (B).** Shown is the proportion of viable cells in each strain normalized to the WT in the respective conditions. pTrc99a carries GFP under the Trc promoter, pUA66 carries GFP under the *rplN* promoter. The CFU measurements were performed with the strains carrying pTrc99a-GFP plasmid in presence of 50 uM of IPTG, in 2 technical replicates and 7 biological replicates, each point represents the average of a biological replicate. The SYTOX Orange measurements were performed in biological triplicates. Error bars indicate standard errors. P values were calculated using a Student's t-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, ns, $P>0.05$). The Y-axis is $\log_{10}$.

**Table 4. Viability of the strains.**

| **sample** | **strain** | **Percentage of the living cells, %** | | | **compared to WT** | | |
|---|---|---|---|---|---|---|---|
| | | **No plasmid** | **pTrc99a** | **pUA66** | **No plasmid** | **pTrc99a** | **pUA66** |
| Filtered lake water (assessed by CFU counting) | WT | | 65.8±12.67 | | | | |
| | Δ*hdfR* | | 57±15.8 | | | 0.85 | |
| | Δ*rpoS* | | 22.8±5.3 | | | 0.42 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Filtered lake water | WT | 95.9±0.7 | 96±0.84 | 96±1.27 | | | |
| | Δ*hdfR* | 96.6±0.37 | 89.8±0.12 | 97.4±0.23 | 1.01 | 0.94 | 1.01 |
| | Δ*rpoS* | 93.1±0.62 | 72.7±1.18 | 89.3±0.09 | 0.97 | 0.76 | 0.93 |
| Exponential phase | WT | 94.1±0.55 | 95.5±1.24 | 97.8±0.55 | | | |
| | Δ*hdfR* | 95.9±0.72 | 94.4±0 | 95.6±1.18 | 1.02 | 0.99 | 0.98 |
| | Δ*rpoS* | 94.8±0.91 | 92.4±0.38 | 98±0.46 | 1.01 | 0.97 | 1.00 |
| Stationary phase | WT | 96.7±0.7 | 91.6±1.07 | 97.7±0.78 | | | |
| | Δ*hdfR* | 98.2±0.12 | 86.5±2.37 | 99±0.17 | 1.01 | 0.95 | 1.01 |
| | Δ*rpoS* | 94.6±0.26 | 83.7±0.78 | 97.6±0.32 | 0.98 | 0.91 | 1.00 |

Shown are the average values and the standard errors of mean.

Viability estimates via CFU counting tend to be lower than the viability estimates via live-dead staining. In traditional microbiology methods, like CFU counting, viability is equated with cultivability [174], and this method tends to overestimate the number of dead cells. On the other hand, most of the live-dead (including SYTOX Orange that we used) act by forming stable fluorescent complexes with double-stranded DNA, and they can only reach the DNA molecule if the cell membrane is compromised. Therefore, live-dead staining refers to a different definition of death in bacteria, specifically the loss of membrane integrity. In this approach, the number of dead cells is often underestimated [196].

Despite these caveats, the pattern in the normalized data obtained by these two different methods tend to match pretty well. Δ*rpoS* strain was less viable in our screening conditions (pTrc99a plasmid, lake water) than the WT strain regardless of the method used. Δ*hdfR* had a marginal viability defect compared to the WT strain, but it was not significant.

However, the results become different if we compare the viability of the strains that do not carry any plasmid at all or carry a low-copy plasmid pUA66. In this setup, both of our target strains were only slightly less viable than the WT strain, and only Δ*rpoS* survived significantly lower. Since the GFP levels per cell were comparable for both plasmids, we consider specifically pTrc99a and not GFP to be the main burden. The effect of Δ*rpoS* appears regardless of the GFP level and therefore we confirm the importance of its mutation in the conditions of lake water.

# Chapter III. HdfR-MaoP system: suppression of *Escherichia coli* motility

After the completion of the lake water study, we have determined that the mutants Δ*hdfR* and Δ*rpoS* have the growth advantage in the filtered lake water. Functions, regulatory cascades and role of *rpoS* in various conditions are widely described in the literature [76, 77, 80, 103, 118, 122, 125]. On the opposite side, *hdfR* is poorly studied and the information about it is sparse. To unravel its role in growth and survival of *E. coli*, we decided to study its deletion strain in greater detail.

HdfR was described as an H-NS-dependent negative transcriptional regulator of the flagellar master operon *flhDC*. The disruption of *hdfR* enhances the activity of *flhDC* [134]. Besides the effect on flagella regulation, it is reported to activate the transcription of *gltBD* operon [144] and thus facilitate the glutamate biosynthesis and acid stress resistance. According to the bioinformatic analysis, *hdfR* has two promoters, one on which is under RpoS regulation [148] and the other one is regulated by FliA [149]).

Next to *hdfR* there is an adjacent divergently transcribed gene named *maoP* sharing the intergenic region with *hdfR* (**Figure 21**A). MaoP was described as a protein involved in chromosome organization and more specifically in organization of Ori macrodomain [154] but no clear mechanism was proposed. Inactivation of *maoP* lifted some of the constraints on DNA mobility in the Ori macrodomain and allowed long-range interaction and recombination between Ori and Right macrodomains. The other important part of this system, sequence *maoS*, is situated in the shared intergenic region, and both *maoP* and *maoS* were required to constrain DNA mobility [154]. Since Δ*maoP,* similarly to Δ*hdfR,* had an advantage in the lake water in the TnSeq results, we studied them together as a system.

## 3.1. Flagellar regulation by HdfR-MaoP tandem

Since the role of HdfR in motility has already been reported, we started by assessing the motility of its mutant and measuring the expression of the flagellar genes. *hdfR* is reported to have a $\sigma^S$-dependent promoter, and therefore the measurements of expression were performed in the stationary phase.

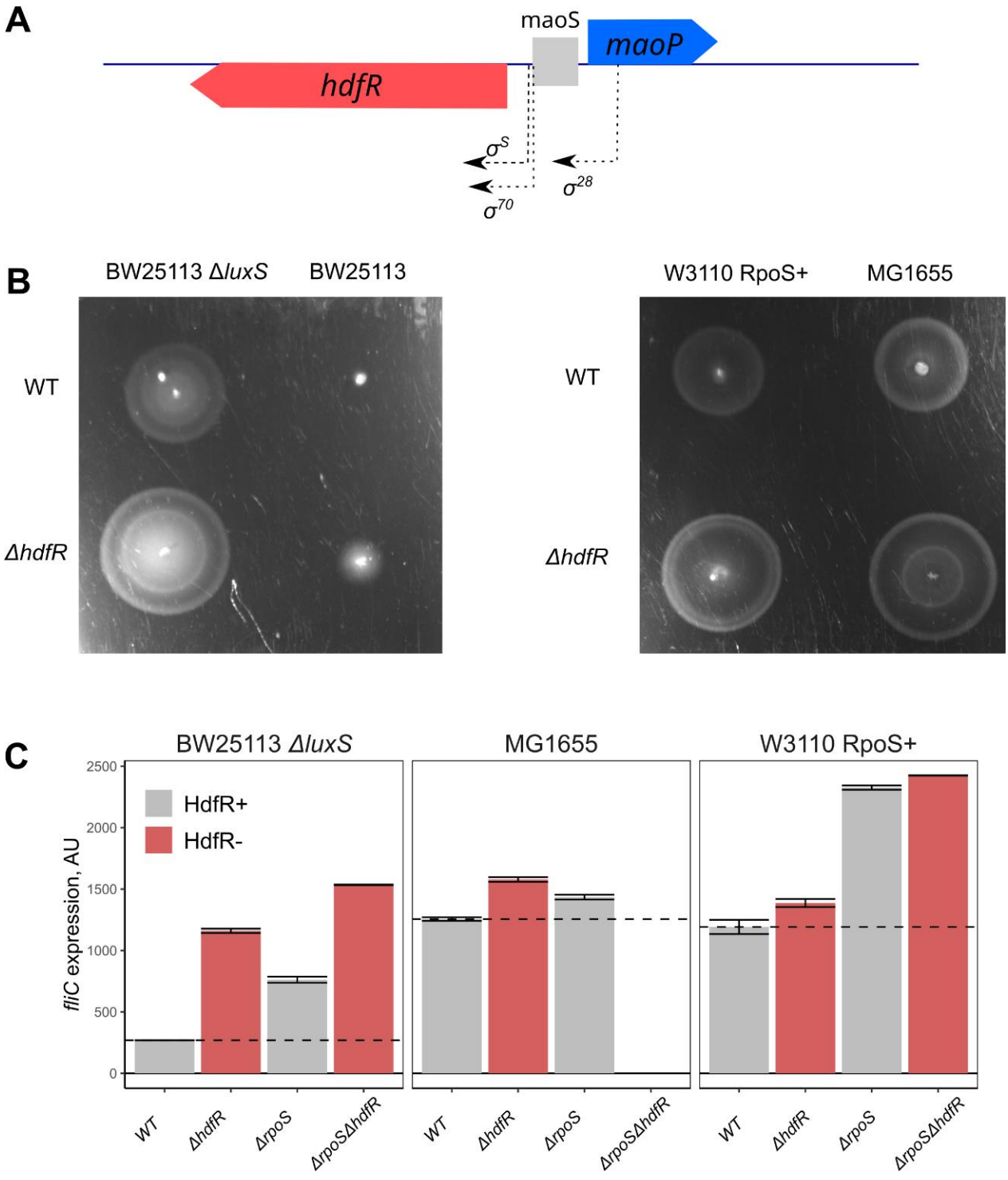



**Figure 21. Deletion of *hdfR* increases motility in *E. coli.*** (A) Genomic context of *hdfR-maoP* system. (B). On the soft agar, Δ*hdfR* strains spread better than the WT strains in both motile and poorly motile BW25113 backgrounds well as in W3110 RpoS$^{+}$ and MG1655 backgrounds. Shown are representative plates. (C). Changes in expression of *fliC* in various deletion strains in three background strains of *E. coli*. Samples were taken from the overnight cultures. The dashed line corresponds to the average value of the WT. All measurements were performed in triplicates. Error bars indicate standard errors.

We have tested whether the deletion of *hdfR* increases spreading in *E coli* on the soft agar plates (**Figure 21**B) on three different lineages of *E. coli* K-12 (W3110 RpoS+ [52], MG1655 [72] and BW25113 [160]). In the case of BW25113, we have additionally confirmed the effect on its isogenic Δ*luxS* strain as it is motile unlike the parental strain due to the spontaneous transposon insertion upstream of *flhDC* (**Table 12**).

Similar results were achieved by measuring the expression of *fliC* (as a representative of *flhDC* regulon) in the stationary phase of growth. The most pronounced effect was in the BW25113 Δ*luxS* strain, and therefore it was chosen for the further work (**Figure 21**C). The expression of *fliC* was increased in all three tested backgrounds. Since *hdfR* is reported to have a $\sigma^S$-dependent promoter [148], we also tested the activity of *fliC* promoter in the Δ*rpoS* and the Δ*rpoS*Δ*hdfR* deletion strains, and the effect of these two mutations was additive. This made us doubt whether the $\sigma^S$-dependent promoter of *hdfR* is the main promoter.

Then we have measured the expression of the genes of flagellar regulon at all three levels (*flhD* as the class I gene, *fliA* as the class II gene, *fliC* as the class III gene [126, 127]) in WT, Δ*hdfR* and Δ*maoP* strains. In both the knockout strains, the expression of all three representative genes was elevated compared to the WT, and there was no significant difference between Δ*hdfR* and Δ*maoP* strains (**Figure 22**A).

We have additionally measured the expression of *maoP* in the WT strain and Δ*hdfR*. It was completely abolished in the Δ*hdfR* strain (**Figure 22**C). To rule out the hypothesis that this effect is caused by disruption of *maoP* promoter in Δ*hdfR*, we performed Sanger sequencing of this region in the mutant and observed that the intergenic region between *maoP* and *hdfR* was intact. Moreover, as shown later (**Figure 23**), the phenotype of Δ*hdfR* can be complemented from the plasmid confirming the intactness of the promoter. This led us to the conclusion that *maoP* expression was strongly dependent on presence of HdfR in the cell, which is confirmed by the literature data [150].

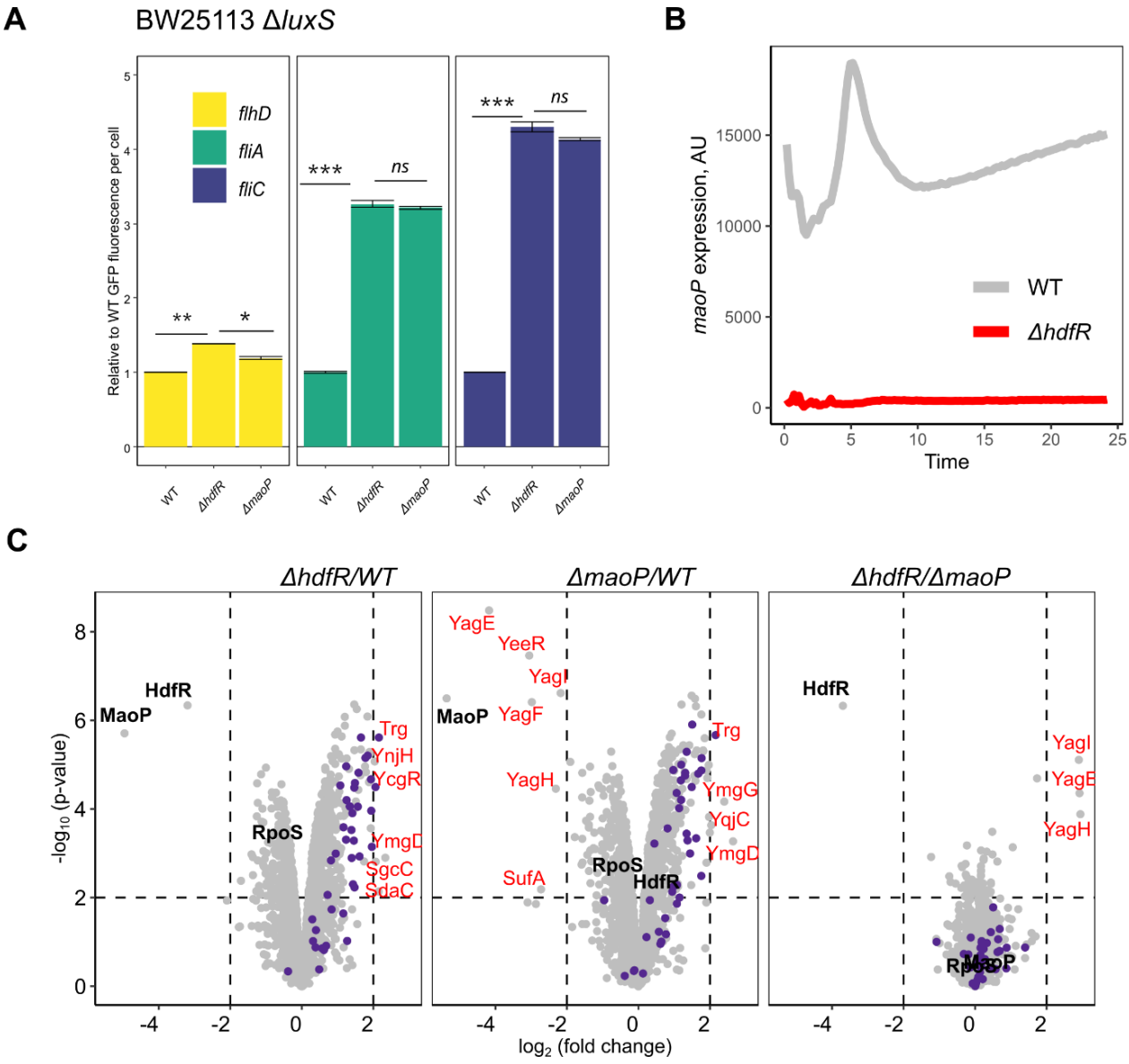



**Figure 22. Initial characterization of Δ*maoP* mutant. (A).** Changes in expression of *flhD*, *fliA* and *fliC* in Δ*hdfR* and Δ*maoP* strains in the stationary phase. The values are normalized to the value of the WT strain BW25113 Δ*luxS* carrying the respective reporter. All measurements were performed in triplicates. Error bars indicate standard errors. P values were calculated using a Student's t-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$, ns, $P>0.05$). (B) The activity of *maoP* promoter in the WT strain BW25113 Δ*luxS* and its isogenic Δ*hdfR* knockout. The measurements are performed in duplicates and corrected for non-growth control and non-GFP control. Each line represents an average of two wells in the plate reader. (C). Global changes in proteome in Δ*hdfR* and Δ*maoP* strains. Protein changes were analyzed using mass spectrometry. Proteins of *flhDC* regulon are highlighted in purple, and the most up- and dowregulated proteins are highlighted in red. For more information, see **Supplementary table 3** and **Supplementary figure 10**.

In an attempt to discover other effects of *hdfR* deletion, we have performed a whole-proteome analysis. Most prominently, MaoP was found very downregulated in Δ*hdfR* strain. The level of HdfR in Δ*maoP* was comparable to the WT strain, if not slightly higher (**Figure 22**C). However, the differences in whole proteomes of Δ*hdfR* and Δ*maoP* are extremely small and there is no difference in flagellar proteins. Complementary to the gene expression measurements, the abundance of FlhD, FliA, FliC is elevated similarly in both the knockout strains. These results imply that HdfR does not inhibits the flagellar master regulon directly, and show greater MaoP involvement in this process. Moreover, *maoP* is probably the only regulatory target of HdfR at least in the tested conditions. Another important finding from this dataset is that the level of $\sigma^S$ is reduced in both mutants.

To verify our results and confirm the leading role of MaoP in flagellar regulation, we have expressed both *hdfR* and *maoP* from the plasmids and checked whether these plasmids complement the mutant phenotype. Expression of *fliC* was measured over the course of incubation of the WT strain, both knockouts and overexpression/complementation strains (**Figure 23**). The effect of *hdfR* and *maoP* deletions is prominent in the late exponential/early stationary phase. We demonstrate that in the WT, it is typical for *fliC* expression to have a sharp peak in the mid-exponential phase that quickly declines by early stationary phase. In Δ*maoP* and Δ*hdfR,* the expression of *fliC* before this peak is similar to the WT, but the peak itself is higher and is shifted to the later timepoint, and the steady stationary phase level is higher. We were able to show that the WT phenotype can be restored in Δ*hdfR* by either *maoP*- or *hdfR*-expressing plasmid. However, in Δ*maoP* WT phenotype is restored only by *maoP*-expressing plasmid and not by *hdfR*-expressing one. This experiment leads us to three conclusions. First, MaoP is a negative regulator of *flhDC.* Second, it is positively regulated by HdfR. Third, contrary to the previous literature reports [134], HdfR does not have a direct negative influence on flagellar expression.

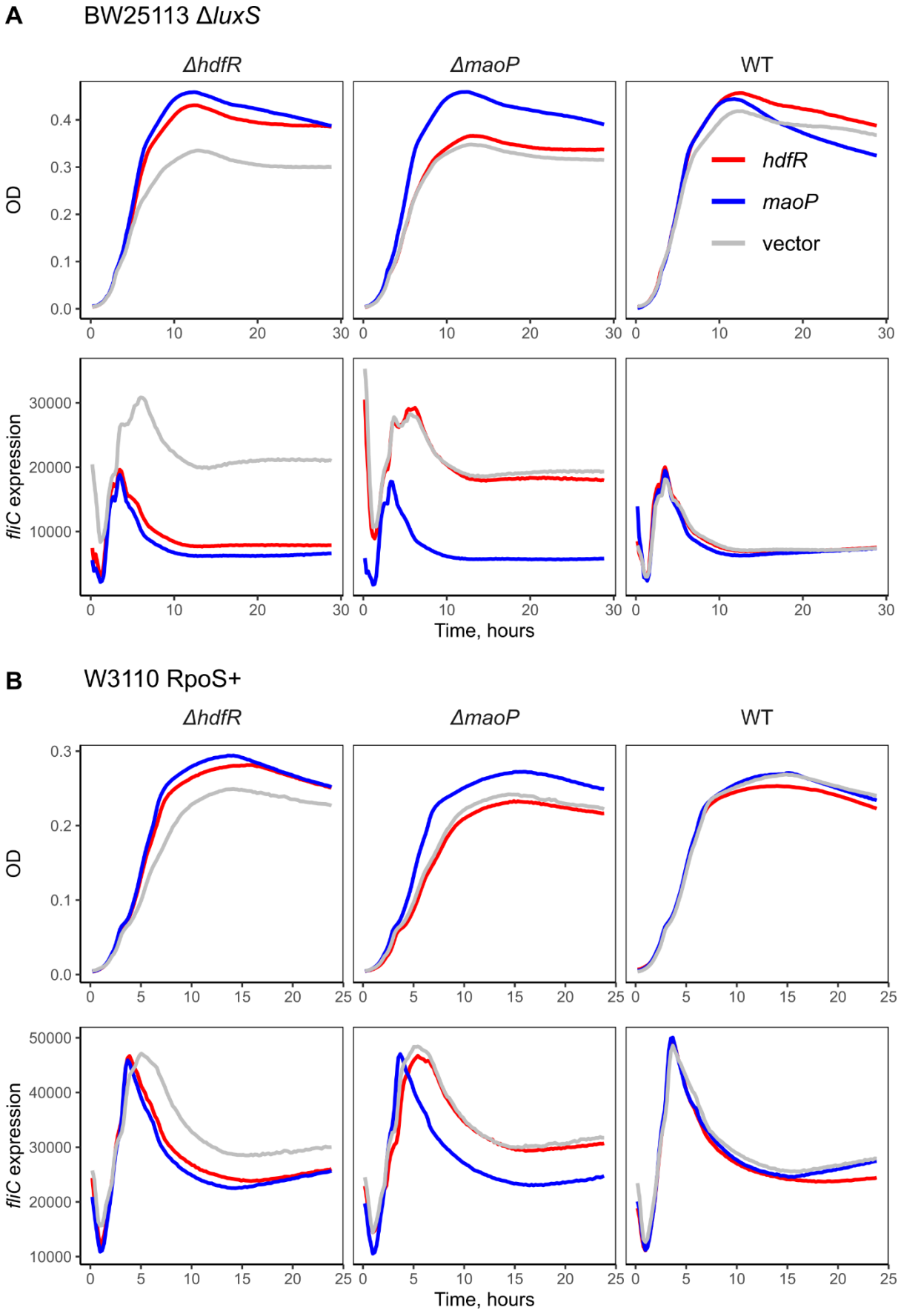

A
BW25113 ΔluxS
ΔhdfR
ΔmaoP
WT
OD
fliC expression
hdfR
maoP
vector
Time, hours
B
W3110 RpoS+
ΔhdfR
ΔmaoP
WT
OD
fliC expression
Time, hours

**Figure 23. Complementation of Δ*hdfR* and Δ*maoP* mutant phenotypes by trans-expression**. (A) Growth curves of BW25113 Δ*luxS* and its isogenic knockout strains Δ*hdfR* and Δ*maoP* with and without complementation and activity of *fliC* promoter reporter normalized to the optical density of the culture. (B) Same experiment performed in W3110 RpoS$^+$ background. All measurements are performed in duplicates and corrected for non-growth control and non-GFP control. Each line represents an average of two wells in the plate reader.

The same result, although less prominent, was observed in *E. coli* K-12 W3110 RpoS$^+$ background strain (**Figure 23**B). The main difference between two strains that we have studied is higher activity of stress response σ factor *rpoS* in the W3110 RpoS$^+$ strain.

Based on our experimental data, we propose a new scheme of flagellar regulation in the early stationary phase (**Figure 24**). Instead of direct inhibition of *flhDC* transcription, HdfR regulates it indirectly via MaoP. FlhDC, in turn, activates the transcription of the entire flagellar regulon by initiating the transcription of flagellar σ factor *fliA* and some other genes. To sum this up, the interplay of HdfR and MaoP is crucial for maintaining low expression of flagellar genes in the stationary phase.

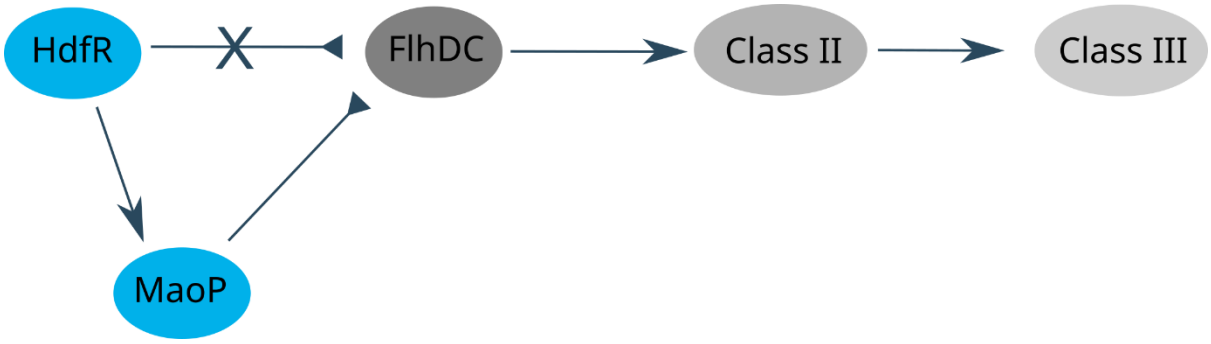



**Figure 24. Schematic of proposed flagella regulation.**

To unravel other potential functions of HdfR and MaoP, we have overexpressed them beyond their respective physiological levels, and tested how it would influence both growth and expression of *fliC*. We have shown that HdfR overexpression is very toxic to the cell, leading to the significant growth defect (**Figure 25**A). However, flagella expression does not decrease. Overexpression of MaoP, on the contrary, does not result in any growth defects (**Figure 25**A). Interestingly, the effect on *fliC* expression is milder than we expected. Even very high levels of induction reduce *fliC* expression just slightly.

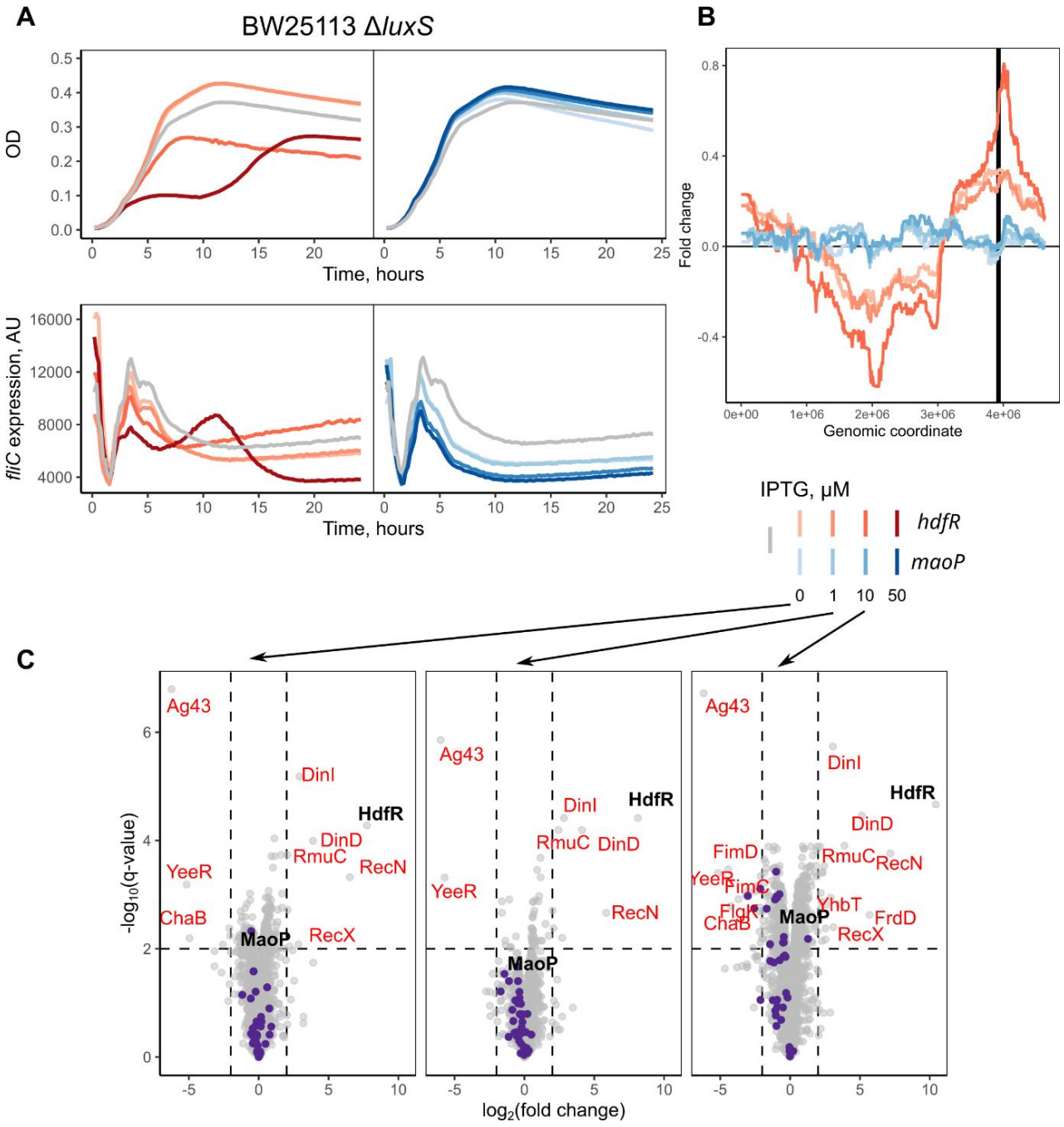



**Figure 25. Overexpression of *hdfR* is toxic to the cell, unlike overexpression of *maoP*.** (A) Growth curves and *fliC* expression of BW25113 Δ*luxS* expressing *hdfR* or *maoP* in *trans* in presence of increasing concentrations of IPTG. Gray lines represent the strains carrying the empty vector plasmid. (B) Running median (window width 101) of fold changes of the proteins across the genome in the whole-proteome data. The thick black line represents the origin of replication. (C) A number of proteins involved in SOS-response are upregulated if HdfR is in excess. Note that this effect is visible even at the level of induction where there is still no growth defect. The purple points represent the proteins of the flagellar regulon. The whole proteome study is performed in TB on the overnight cultures. For more information, see **Supplementary table 4**.

We have then conducted the whole proteome analysis of the BW25113 Δ*luxS* strains overexpressing *hdfR* and *maoP* with varying concentrations of IPTG as an inductor. Besides the effect on flagella, overexpression of HdfR caused significant upregulation of the proteins involved in DNA repair and SOS-response (**Figure 25**C), such as recombination repair protein RpoN, RecA inhibitor RecX, DNA damage-inducible proteins DinI and DinD. This matches to the growth-impairing effect of overexpressed HdfR.

Strikingly, the response to HdfR overexpression seems to be dependent on the chromosome coordinate. The genes that are situated close to the origin of replication, are upregulated (**Figure 25**B). Moreover, this effect is induction-dependent. Overexpression of MaoP, on the other hand, does not cause such an effect. This effect might be caused by the elevated number of the replication forks. This, in turn, might be caused by either increased frequency of replication initiation or excess HdfR binding causing the stalling of the replisome. Given the toxicity of HdfR overexpression, we tend to accept the latter hypothesis.

Unfortunately, overexpression of MaoP did not give us more hints about its functions and potential influences (**Supplementary table 4**, **Supplementary figure 11**). The abundance of only few proteins changed significantly, and most of them are uncharacterized.

## 3.2. HdfR and MaoP regulate the levels and activity of RpoS

In our proteomic results, $\sigma^S$ level was lower in both Δ*hdfR* and Δ*maoP* stains, Additionally, we observe that the HdfR/MaoP tandem had a weaker effect on flagellar expression in *E. coli* K-12 W3110 RpoS$^+$ than in *E. coli* K-12 BW25113 (**Figure 23**). One of the differences between these backgrounds is a point mutation in the *rpoS* gene that results in a different amino acid at the codon 33 (glutamine in BW25113 and tyrosine in W3110 RpoS$^+$). This position is a mutational hotspot [116, 197], and the resulting RpoS activity might differ among the strains. On top of that, HdfR is reported to have a $\sigma^S$-dependent promoter [148].

Moreover, it was described in detail [125] that flagellar regulon and $\sigma^S$ regulon are tightly interweaved, and mutually suppress each other at several levels. For example, *flhDC*-regulated FliZ suppresses the expression of the $\sigma^S$-dependent genes by competitively binding to their promoters. All these results and data give a strong hint that the potential link between RpoS and HdfR/MaoP exists.

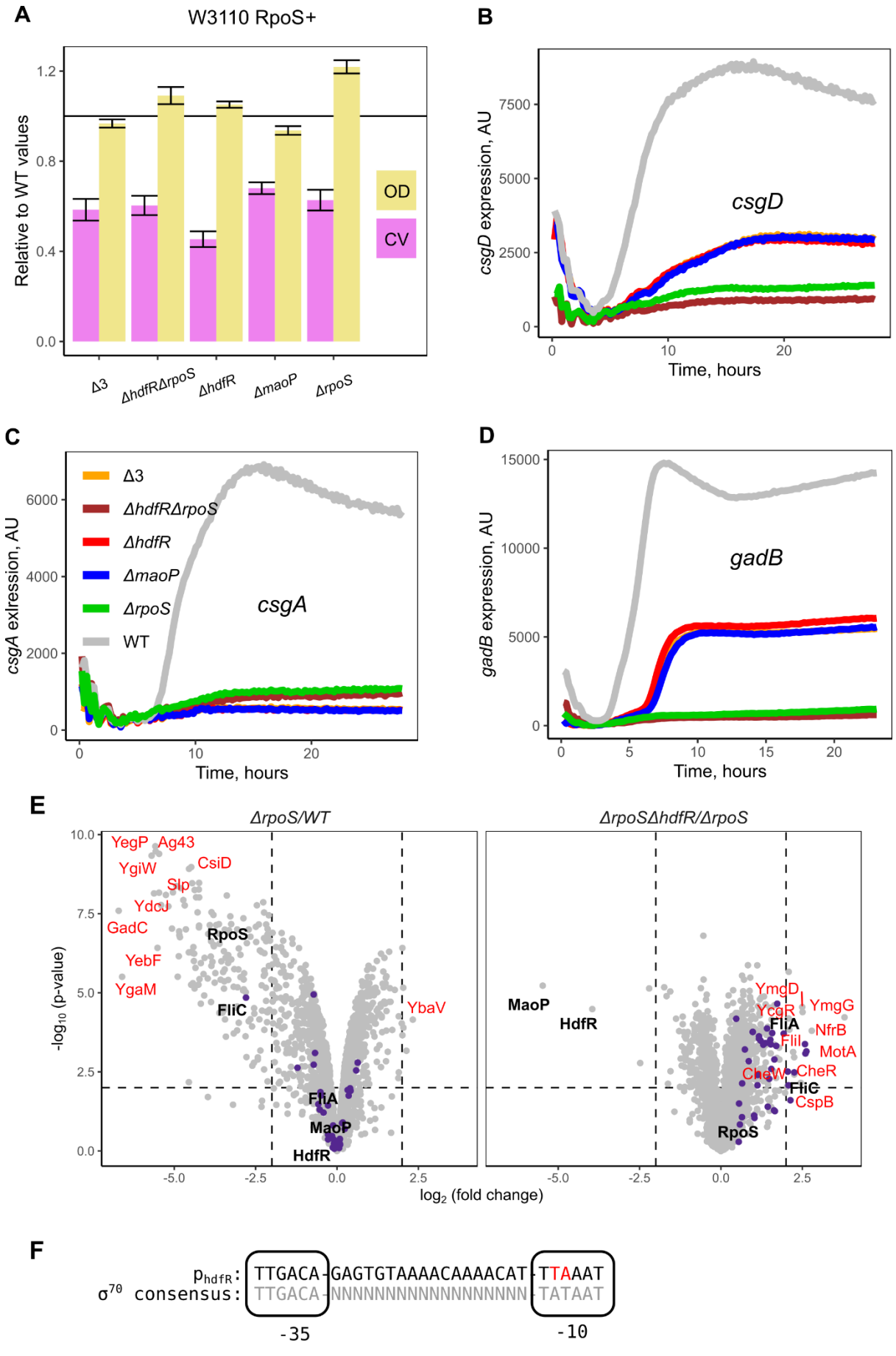

A
W3110 RpoS+
Relative to WT values
OD
CV
Δ3
ΔhdfRΔrpoS
ΔhdfR
ΔmaoP
ΔrpoS
B
csgD expression, AU
csgD
Time, hours
C
csgA exlression, AU
Δ3
ΔhdfRΔrpoS
ΔhdfR
ΔmaoP
ΔrpoS
WT
csgA
Time, hours
D
gadB expression, AU
gadB
Time, hours
E
ΔrpoS/WT
ΔrpoSΔhdfR/ΔrpoS
-log10 (p-value)
log2 (fold change)
YegP
Ag43
YgiW
CsiD
Slp
YdcJ
GadC
RpoS
YebF
YgaM
FliC
YbaV
FliA
MaoP
HdfR
YmgD
YcgR
YmgG
FliA
NfrB
FliI
MotA
CheW
CheR
FliC
CspB
F
phdfR: TTGACA-GAGTGTAAAACAAAACAT-TTAAAT
σ70 consensus: TTGACA-NNNNNNNNNNNNNNNNNN-TATAAT
-35
-10

**Figure 26. RpoS-dependent functions are inhibited in Δ*hdfR* strains** (A) Biofilm formation by Δ*hdfR* deletion strains. Shown are crystal violet (CV) staining results and $OD_{600}$ measurements of the planktonic cultures, normalized to the WT strain. The measurements are performed in three biological replicates and three technical replicates. Error bars represent the standard error of the mean. (B-D) Activity of *csgA*, *csgD* and *gadB* promoter reporters normalized to the optical density of the culture. All measurements are performed in duplicates and corrected for non-growth control and non-GFP control. Each line represents an average of the two wells in the plate reader. (E) Global changes in proteome in caused by *hdfR* deletion in Δ*rpoS*. Protein changes were analyzed using mass spectrometry. Proteins of *flhDC* regulon are highlighted in purple, and the most up- and downregulated proteins are highlighted in red. (F) Putative $\sigma^{70}$-dependent promoter of *hdfR* discovered by us. The consensus sequence is taken from [122].

Therefore, we studied the RpoS-dependent activity in the Δ*hdfR,* Δ*maoP* and Δ*rpoS* strains in W3110 $RpoS^+$ background (**Figure 26**). We also performed the same measurements in the strain we named Δ3, where both of these genes were removed together with their intergenic region, and in the Δ*hdfR*Δ*rpoS* double deletion strain.

First, biofilm formation was severely disrupted by the mutations in the HdfR/MaoP system (**Figure 26**A). Surprisingly, the defect in the biofilm formation was comparably severe in all the tested mutants. The formation of the biofilm strongly depends on extracellular polymer matrix which is (in the case of *E. coli*) mostly composed from the amyloid curli fibers [30, 53]. We have therefore measured the expression of main curlin subunit *csgA* and its positive transcriptional regulator *csgD* (**Figure 26**B, C). These genes are primarily expressed from the *rpoS*-dependent promoters [29]. Expression of both these genes drops dramatically if HdfR/MaoP interplay is disrupted. In case of *csgA* expression, the suppression is even stronger than in the Δ*rpoS* mutant strain, implying more specific regulation of *csgA* by HdfR/MaoP. This explains why the biofilm formation defect was comparable to one in Δ*rpoS* strain.

However, the expression of *csgA* and *csgD* is in addition negatively regulated by the proteins of the flagellar regulon, such as YhjH and FliZ [125], and the flagellar expression is increased in the HdfR/MaoP mutants. The effect we observe here might be caused by the increased expression of YhjH or FliZ. To isolate the $\sigma^S$-dependent activity, we measured the activity of *gadB* promoter, which is primary transcribed from the $\sigma^S$-dependent promoter [77, 198] and was not shown to be specifically suppressed by the genes of flagellar cascade. In all three tested strains, *gadB* activity was significantly reduced but not to the level of the Δ*rpoS* strain (**Figure 26**D). On top of that, the start of *gadB* expression was delayed in the deletion strains, implying the later onset of the stationary phase.

When comparing the whole proteome of Δ*rpoS* and Δ*rpoS*Δ*hdfR* strains, we noticed that the mutation in *hdfR* still elevates the abundance of flagellar proteins in Δ*rpoS* strain. Therefore, the effects of HdfR/MaoP on $\sigma^S$ regulon and flagellar regulons are independent.

Interestingly, despite *hdfR* was reported to have a $\sigma^S$-dependent promoter, the abundance of HdfR and MaoP remained unchanged in the Δ*rpoS* mutant. Moreover, we could not find a $\sigma^S$-dependent promoter of *hdfR.* Instead, we have found a sequence that rather resembled a $\sigma^{70}$-dependent promoter (**Figure 26**F) [122].

Taken together, these results tell that the $\sigma^S$ regulon is vastly suppressed if HdfR or MaoP are not present in the cells. It could be explained by the fact that the level of $\sigma^S$ is reduced in both Δ*hdfR* and Δ*maoP* mutants (**Figure 22**C). We can conclude then, that HdfR/MaoP system not only suppresses flagella, but might positively regulate *rpoS* and thus take part in transition to the stationary phase.

## 3.3. Effect of MaoP on flagella is methylation-independent

In the previous chapters, we have demonstrated that MaoP suppresses the expression of flagellar master regulator *flhDC* and increases the amount of $\sigma^S$, but the exact mechanism is not obvious. In the previously reported results [151, 153, 154, 156, 157], nothing was mentioned about any possible links between MaoP and regulation of motility or stress response. MaoP is not a canonical transcription factor and its ability to bind DNA was suggested but not proven [154]. Protein structure and domain prediction suggested by AlphaFold [199] and Pfam [200] respectively did not give any valuable input because MaoP lacks homology to well characterized proteins. On top of this, there are experimental results demonstrating that HdfR does bind to *flhDC* promoter region [134], which contradicts our data.

To explore the link between DNA-methylation and MaoP described earlier [151], we have performed several experiments on *Δdam, ΔdamΔhdfR* and *ΔdamΔmaoP* strains. We were interested in measuring the activity of *fliC* promoter in these strains (**Figure 27**). For this experiment, the BW25113 background was chosen because MaoP deletion had a higher effect on flagella. The *fliC* expression in Δ*dam* strain turned out to be lower than in WT strain but retaining the shape of the expression curve typical to the WT with a sharp mid-exponential peak (the effect is opposite to *maoP/hdfR* deletion, where the shape of the curve was altered). Dam-methylation can possibly modify the recognition sequence of transcription factors or σ factors, therefore limiting the protein-DNA interaction and influencing the

expression of the downregulated genes [201]. Dam-recognition site GATC is very abundant in *E. coli* genome and is found upstream of many genes including the ones that belong to the flagellar regulon. It was also shown that Δ*dam* strains are poorly motile in several *E. coli* background strains [201], and therefore this result was expected. However, the expression curves of the double mutants have the shape typical for *maoP/hdfR* deletion and lower levels of expression observed for *dam* deletion. The additive effect means that the effects of these mutations are independent.

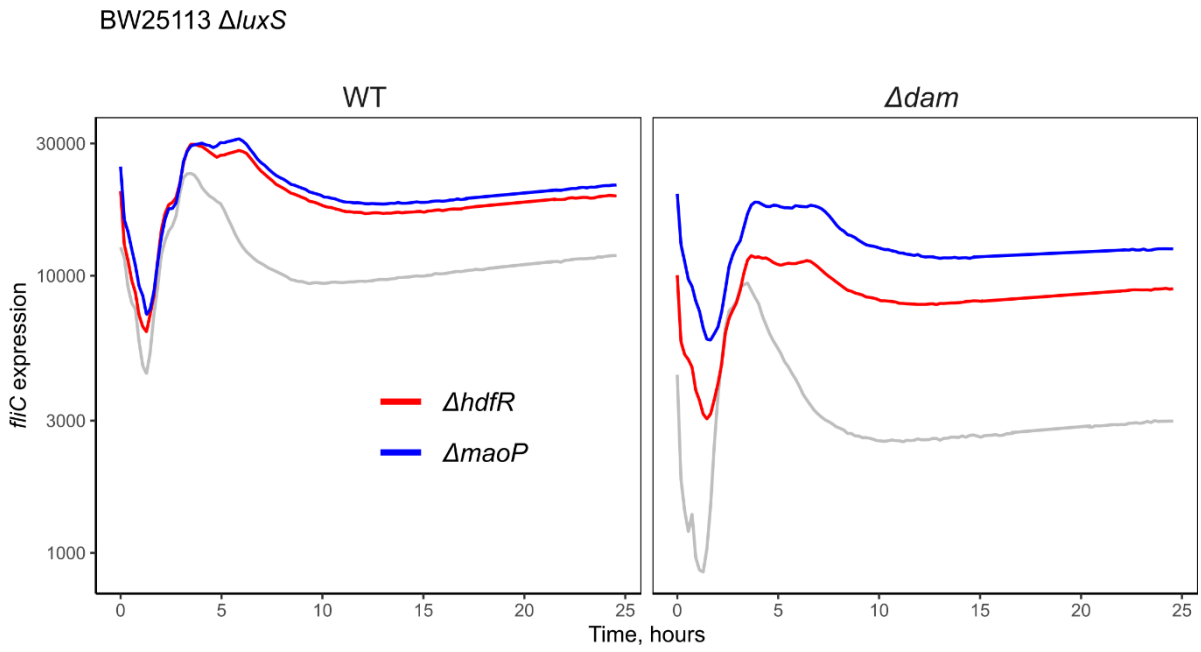



**Figure 27. Activity of *fliC* promoter in Δ*dam* and the double knockouts normalized to the optical density.** All measurements are performed in duplicates and corrected for non-growth control and non-GFP control. Each line represents an average of two wells in the plate reader. The Y-axis is $\log_{10}$.

## 3.4. MaoP might interact with RNA polymerase

We have then fused HdfR and MaoP with the fluorescent label (superfolder GFP, sfGFP) to use them as a bait to study the protein-protein interactions. We used the C-terminal fusion of HdfR because it was better expressed, and the N-terminal fusion of MaoP because it was able to complement the mutant phenotype on the soft agar plates, unlike the C-terminal one.

To explore the interaction partners of MaoP, we have performed the co-immunoprecipitation assays (co-IP). To cross-validate our results, we used two complementary approaches. First, we performed the standard co-IP. Briefly, formaldehyde was added to the cells to fixate them and create the strong covalent bond between the interacting partners. After that, the cells were

lyzed in the mild buffer and the protein of interest (POI) is recovered from the lysate using the protein-specific antibodies on magnetic beads [202]. The second approach we used was biotinylation-based (BioID) proximity labelling. In this setup, the POI is fused to the biotin-ligase (we used the TurboID enzyme) and the bacterial cells are cultivated in presence of excess biotin. The biotinylase attached to the POI ligates the excess biotin to the surrounding proteins thus labelling them *in cellulo*. The cultures are then harvested and the cell lysates are prepared, and the labelled proteins can then be recovered from the lysate using the streptavidin-coated beads [203]. In both approaches, the proteins recovered on the beads are likely to be the interactors of the POI.

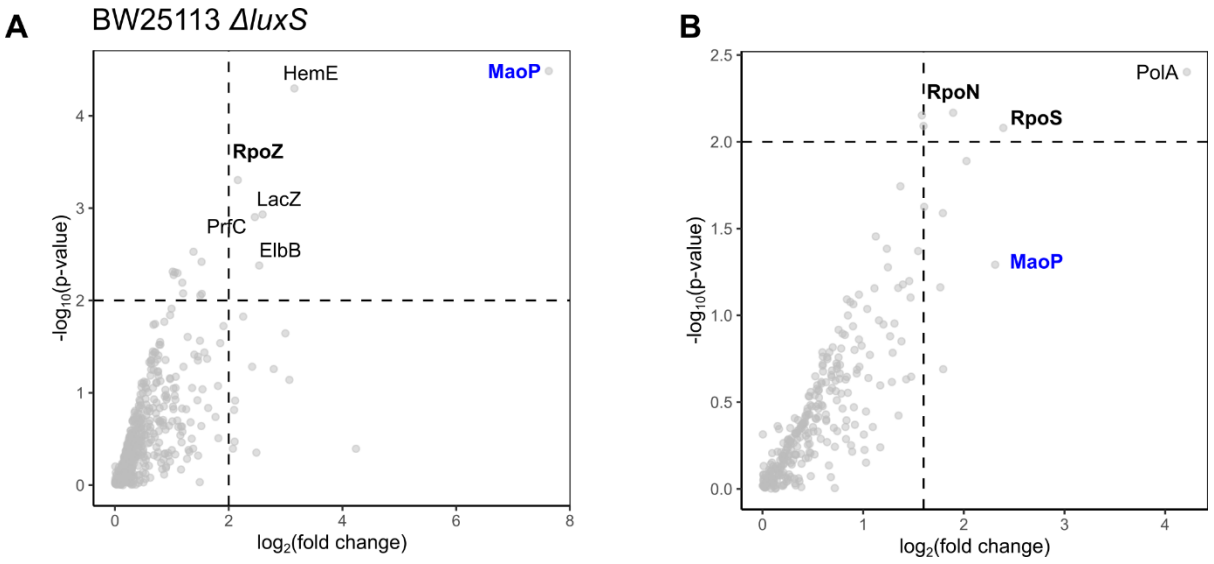



**Figure 28. Protein-protein interaction of MaoP.** (A) The results of a classical pulldown with formaldehyde cross-linking and anti-GFP beads. (B) The results of the BioID proximity labelling. For both approaches, the culture was collected at the OD=1 in TB medium (early stationary phase). The subunits of RNAP are highlighted in bold. See 5.2. Methods for more details.

In both approaches, we found that certain subunits of RNAP were enriched in the interactor fraction (**Figure 28**A, B). First, in the classic pulldown RpoZ, an ω-subunit of RNAP, was determined as an interactor. This small protein is not required for transcription but it acts as a chaperone to stabilize and ensure the correct folding of the β`-subunit [204]. The ω-subunit is involved in recruitment of σ factors and regulation of promoter selectivity. For instance, the Δ*rpoZ* mutant favored the formation of $\sigma^S$ RNAP holoenzyme and the entire $\sigma^S$ regulon was upregulated [205]. The ω-less RNAP is reported to recruit the alternative σ factors with higher efficiency than the one that has the ω-subunit. The hypothesized role for the ω-subunit is to make the RNAP core more selective to the $\sigma^{70}$ [204].

Complementary to this, two alternative σ factors ($\sigma^S$ and $\sigma^N$) were found among the interactors of MaoP in the biotinylation-based proximity labelling assay. Thus, the subunits of RNAP were discovered in the interactome of MaoP mapped with two different methods. This strongly suggests that MaoP physically interacts with RNAP and might be involved in global regulation of transcription.

Another interesting hit from the BioID assay was DNA polymerase I (PolA). It is involved in several pathways of DNA repair, including nucleotide excision repair (NER) [206, 207] and post-replication repair [208]. The link between transcription and NER was observed long ago [209], when it was observed that the active running transcription increases the rate of DNA repair. However, it was considered the minor subpathway of DNA-repair. This idea is currently contested by the results [210] showing that in fact coupling with the transcription is absolutely necessary for the successful NER. However, another reason for PolA to appear in this dataset is that it is involved in the replication of ColE1 plasmids [211] and we relied on a ColE1 plasmid to produce the fusions.

In the BioID assay, all the proteins in close proximity to POI are labelled with the biotin tag. However, proximity does not equate interaction, since the proteins might just share the same cell compartment. Since PolA is a key enzyme of NER and the coupling of NER and transcription is proven to be essential, its close proximity to the RNAP is expected. The most probable reason why it appeared in the MaoP interactome is that they both interact to RNAP.

Additionally, we have performed the same assay on HdfR-sfGFP fusion (**Supplementary figure 12**). In case of the classic co-IP, some proteins of internal part of flagellar apparatus were determined as interactors. However, these results were not verified with the BioID approach.

Due to the limitations of the BioID assay discussed above, we found the classic pulldown assay to be more reliable. The potential interaction of MaoP and RpoZ had to be investigated.

We have, therefore, measured the activity of both *fliC* and *maoP* promoters in Δ*maoP*, Δ*rpoZ* and Δ*rpoZ*Δ*maoP* deletion strains. The activity of the *maoP* promoter was virtually unaffected by the Δ*rpoZ* deletion (**Figure 29**A). This was not the case for the *fliC* promoter activity, which was higher in the Δ*rpoZ* strain. The possible reason for this is the upregulation of entire $\sigma^F$ regulon due to higher specificity of RpoZ-less RNAP for the alternative σ factors. The effect of Δ*maoP* and Δ*rpoZ* on *fliC* expression seems to be additive,

suggesting that a different pathway is involved in MaoP downregulating flagella (**Figure 29**B).

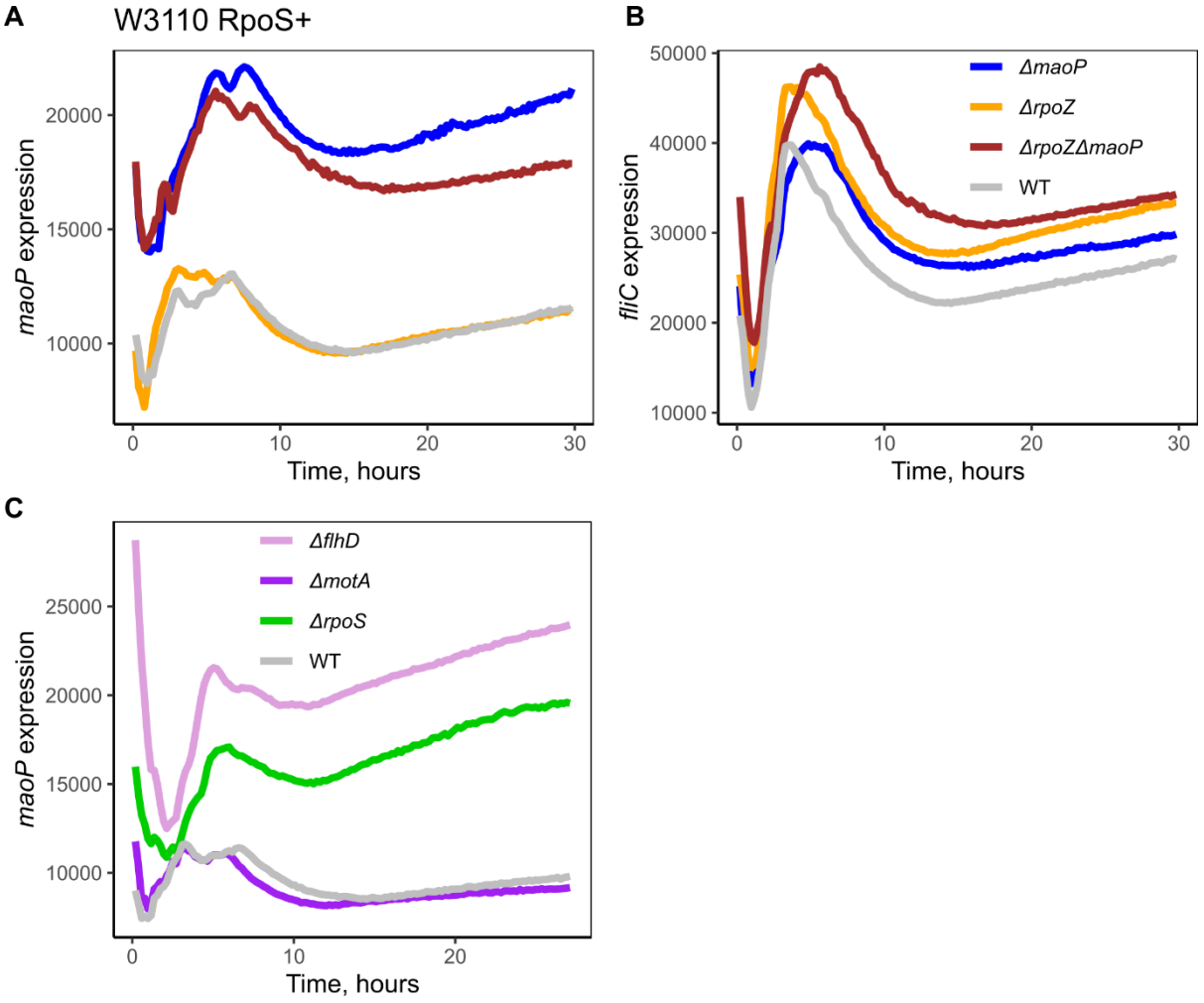



**Figure 29. MaoP acts independently from RpoZ.** (A) Activity of *maoP* promoter. (B) Activity of *fliC* promoter. (C) Activity of *maoP* promoter in the strains that lack sigma-factors is elevated. All measurements are performed in duplicates and corrected for non-growth control and non-GFP control. Each line represents an average of two wells in the plate reader

Since the role of RpoZ in the σ-factor competition had been hypothesized [204, 205], we measured the activity of *maoP* promoter in some strains that lack a single σ-factor, namely Δ*flhD* and Δ*rpoS.* In both of these strains, the activity of *maoP* promoter is elevated compared to the WT strain, especially in the stationary phase. On the contrary, the deletion of *motA,* the class III gene of the flagellar regulatory cascade, did not result in any change in *maoP* expression. This result is particularly puzzling in the Δ*flhD* strain, as we did not expect such a high increase in the *maoP* expression to be caused by elimination of a relatively small $\sigma^F$ regulon.

These results imply that *maoP* and *hdfR* promoters are neither $\sigma^F$ nor $\sigma^S$ dependent, and are highly likely governed by $\sigma^{70}$. This is further supported by the putative $\sigma^{70}$ dependent promoter of *hdfR* we discovered (**Figure 26**E).

## 3.5. ChIP-Seq revealed another potential target of HdfR

According to our proteomic results, the difference between Δ*hdfR* and Δ*maoP* strains is extremely small. This tells us that in the conditions we studied, the only function of HdfR as a transcription regulator was to initiate the transcription of *maoP.* However, we wanted to determine the other regulatory targets of HdfR in hope that they will tell us more about its functions, end employed the chromatin immunoprecipitation sequencing (ChIP-Seq) technique.

To do so, we have expressed the HdfR-sfGFP fusion in the BW25113 Δ*luxS* strain and grew the culture to early stationary (OD=1) phase. Then the cells were crosslinked with formaldehyde, lysed and sonicated to shear the DNA. The DNA fragments attached to the HdfR-sfGFP fusion were collected from the lysate using the anti-GFP magnetic beads, and the Illumina library was then prepared. The same steps were performed with the cells expressing sfGFP as the control.

After alignment of the resulting reads to the reference genome and peak calling, several peaks with the pileup fold change greater than 3 were revealed (**Table 5**). Out of 9 discovered peaks, only one (upstream of *yhjR*) was not intergenic and was situated on *bcsE* gene. Some peaks correspond to two genes because they are situated between two divergently transcribed genes.

First, we were able to confirm the binding of HdfR to the upstream region of *maoP.* We have already demonstrated that the presence of HdfR in the cells is absolutely necessary for them to express *maoP,* and now we demonstrate that HdfR indeed acts as a transcriptional factor.

Second, we have checked the binding sites of HdfR reported in the literature. HdfR was reported to bind to *flhDC* promoter [134] and to *gltB* promoter [144]. Both of these studies employed *in vitro* gel-shift assays to assess the binding. In our results, we did not see a significant increase in coverage in these loci.

**Table 5. Results of HdfR ChIP-Seq.**

| Target gene | Pileup fold change | -log$_{10}$ (q-value) | Proteomic log$_2$ (fold change) | Proteomic -log$_{10}$ (q-value) | Summit position, bp | Distance to start codon, bp | σ |
|---|---|---|---|---|---|---|---|
| *kdgT* | 7.82 | 9321.2 | | | 4091475 | 142 | ? |
| *kdpF* | 5.99 | 4889.4 | | | 724467 | 190 | **70** |
| *ybfA* | | | | | | 123 | F, E |
| *hdfR* | 5.55 | 5152.5 | -3.2 | 6.3 | 3941344 | 27 | 70, S, F |
| *maoP* | | | -5 | 3.3 | | 102 | ? |
| *kduI* | 4.92 | 3345.8 | 0.7 | 2.3 | 2977592 | 108 | 70 |
| *panM* | 4.88 | 4164.6 | 0.4 | 1.6 | 3591147 | 196 | S, E |
| *livK* | | | 0.3 | 0.6 | | 227 | **70** |
| *dhaR* | 4.86 | 3254.8 | -0.3 | 0.2 | 1246432 | 90 | S |
| *dhaK* | | | -0.2 | 1.8 | | 138 | S |
| *uxaB* | 3.44 | 2702.4 | 0.3 | 0.8 | 1605015 | 78 | **70** |
| *yhjR* | 3.34 | 1723.6 | | | 3690258 | 713 | **70**, E |
| *ymiA* | 3.30 | 1393.8 | | | 1329299 | 211 | **70** |

The cultures were collected at the OD=1 in TB medium (early stationary phase). The fold change and q-value are reported with comparison of HdfR-sfGFP to sfGFP. The experiment was performed in biological duplicates. Only the peaks with the pileup fold change > 3 are shown. The proteomic results are reported from the comparison of Δ*hdfR* to the WT strain. Regulatory targets of HdfR are highlighted in blue. Sigma factors with high-quality experimental evidences according to EcoCyc database [191] are highlighted in bold and underlined. For more information, see **Supplementary figure 13**.

Then, we compared the ChIP-Seq results to the whole proteome comparison of BW25113 Δ*luxS* and its isogenic Δ*hdfR* mutant (**Figure 22**C). Unfortunately, the proteins corresponding to two highest peaks were not found in the proteome comparison, and we could not confirm or reject these finding. Among the proteins represented in the whole-proteome data, only KduI levels were significantly changed. This is an isomerase from the hexuronate metabolism. This protein was slightly elevated in the Δ*hdfR* mutant, making HdfR its transcriptional repressor. However, the very small value of fold change implies that there are

other pathways involved in *kduI* regulation. Indeed, the expression of *kduI* was found elevated in presence of glucoronate in the medium [212].

## 3.6. Effect of HdfR/MaoP on cell division and chromosome dynamics

MaoP was first described in the literature as a protein involved in the chromosomal organization and more specifically in organization of Ori macrodomain [154]. In their work, inactivation of MaoP allowed long-range interactions between Ori and Right macrodomains, and increased the mobility of the DNA loci. It was demonstrated that both *maoP* and the short sequence *maoS* situated in the intergenic region between *hdfR* and *maoP* were necessary to constrain DNA mobility.

To explore the link between chromosome organization and MaoP, we used an *E. coli* MG1655 strain with labelled *ori* and *ter*. To label the *ter*, MatP-Ypet fusion was used, as MatP specifically binds to the *ter* [213]. To label the *ori*, the *parS* sequence is inserted close to the *ori,* and the mTurquoise2-ParB fusion was expressed from the plasmid, as ParB binds specifically to *parS* [155]. This labelling allowed to see the localization of *ori* and *ter* as foci in the cell. In this part of the work, we grew the bacteria using a microfluidic chip. This chip, known as mother machine, consists of a main channel through which nutrient media flows and narrow growth channels in which cells are trapped and allowed to divide [214] (**Figure 30**A). We have tracked the division of the labelled strain and its isogenic mutant Δ3 (where both *hdfR* and *maoP* were removed together with their intergenic region) over 72 hours in the M9 medium+0.4% glycerol, imaging every 10 minutes.

Over the course of the experiment, we observed 9945 cell cycles for the wild type and 10788 cell cycles for the Δ3. We then calculated the growth rate (**Figure 30**B) and the cell lengths at birth and at division (**Figure 30**C). There was no major difference in growth rates between the two strains, and the cell length did not globally differ either.

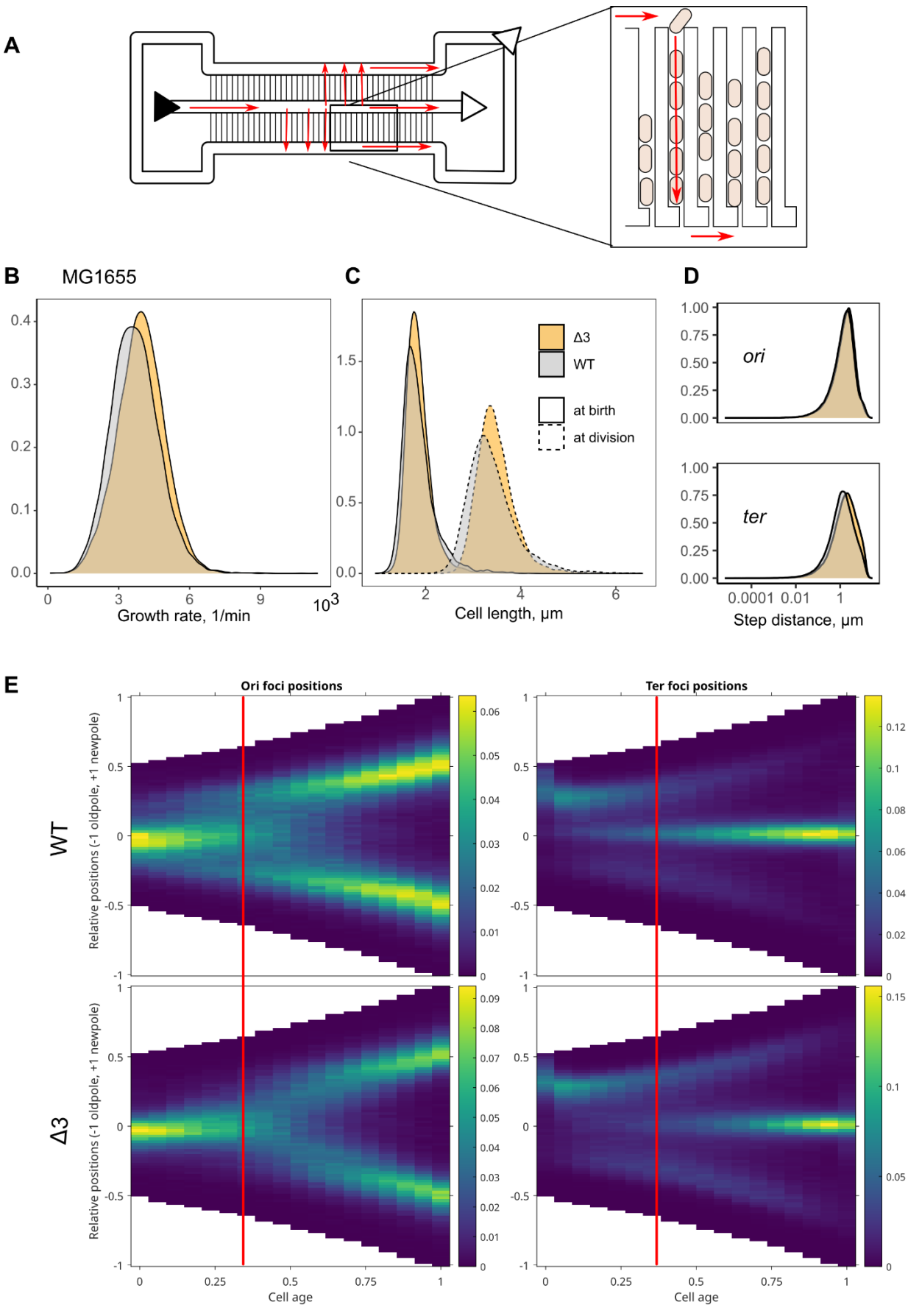

A
B
MG1655
Growth rate, 1/min
10^3
C
Δ3
WT
at birth
at division
Cell length, μm
D
ori
ter
Step distance, μm
E
Ori foci positions
Ter foci positions
WT
Δ3
Relative positions (-1 oldpole, +1 newpole)
Cell age

**Figure 30. Chromosome organization is altered upon disruption of HdfR/MaoP system.** (A). Simplified scheme of the mother machine chip. The red arrows indicate the direcrion of the medium flow. The filled triangle is an inlet, and the emtpy triangles are outlets. (B). Growth rate. (C). Cell lengths at birth and division. (D). Step distances (distances travelled between the frames) for *ori* (top) and *ter* (bottom). (E). Population kymograph of *ori* and *ter* foci positions along the long axis of the cell. Data from different cell cycles are combined by using cell age (0 is birth, 1 is division) and positions relative to an exponentially increasing normalised length. The values in the color scale for the kymographs represent the frequency of occurrence of foci positions normalized to the number of cell cycles at each cell age. The red lines indicate the time of the *ori* focus splitting and *ter* focus appearing in the WT. All the results are calculated based on 9945 cell cycles for the wild type and 10788 cell cycles for the Δ3.

Then, we analyzed the localization of the *ori* and *ter* foci in the cells. The newborn cells normally have a single *ori* focus in the middle of the cell. Later in the cell cycle, the *ori* focus duplicates and segregates outwards to the quarter positions as described [215]. The *ter* focus is close to the pole in the newborn cell but later it moves to the midcell where it remains for the rest of the cell cycle. It was demonstrated that *ter* centralization requires *ori* splitting to happen and coincides with it [215].

We indeed observed the central location and splitting of the *ori* in both our strains. However, the localization of *ori* was more tight and defined in the Δ3, and the segregation happened later than in the wild type (**Figure 30**E). The mobility of both *ori* and *ter* foci was not affected by the mutation (**Figure 30**D). Delayed *ori* splitting in turn caused the delayed centralization of *ter*. Thus, elimination of the HdfR/MaoP system caused perturbation in a normal cell cycle. Interestingly, these alterations generally do not result in growth defects.

# Chapter IV. Discussion

## 4.1. Growth in the lake water is genetically determined

We have conducted the comprehensive whole-genome screening of *E. coli* survival in the lake water using two complementary approaches. First, we used the TnSeq approach to assess which strains from the pooled diverse population will have a growth advantage or disadvantage. Second, we used fluorescence-based screen to study the strains one-by-one and eliminate certain factors such as cross-feeding, competition between different strains and formation of co-dependent communities. This allowed us to reveal that mutations in several pathways improve or decrease *E. coli*'s ability to proliferate in these conditions.

As expected, many mutations were detrimental for *E. coli* growth, and we were able to identify them using the TnSeq approach. In the FB approach, however, we could not identify any detrimental mutations. Potentially, it might happen due to the absence of competition, as the growth defect were not as obvious in the pure culture as opposed to the pooled library. To our surprise, some mutations were beneficial for *E. coli* to grow but only in the filtered lake water. These mutations were identified in both approaches.

### 4.1.1. Mutations that cause the cell envelope defects are the most detrimental

First, we have demonstrated that the cell envelope integrity was crucial for *E. coli* to thrive in the lake water conditions (**Figure 12**). Cell membrane and the surface structures form a border between the cell and the outer world, and their role is hard to underestimate. Loss of the membrane integrity is one of the widely accepted biomarkers for the dead cells [174]

In the TnSeq results, mutations in *tolA*, *pal* and *lpp* genes were the top-3 most detrimental for the cells. The bacteria that carry these mutations were unable persist in both filtered and non-filtered water. Lpp functions as an anchor connecting peptidoglycan and the outer membrane under normal growth conditions, preventing this membrane to tear apart from the cell [165]. TolA and Pal are the members of Tol-Pal system, which colocalizes to the divisome and is required for correct outer membrane invagination and septal peptidoglycan processing [166] during the cell division. Mutations in the tail-specific protease Prc, which is involved in regulation of peptidoglycan biosynthesis [216], also resulted in the strong fitness reduction. Mutants of *tolA*, *pal* and *lpp* are viable and even do not have a major growth defect under

most of laboratory conditions [165, 167]. However, they are more sensitive to toxins (such as detergents and EDTA) and are known to release the periplasmic content into the outer medium [165], potentially due to the membrane leakiness. In the lake water setting, this might lead to (1) easier penetration of the toxic substances to the cell and (2) leakage of nutrients and other molecules from the periplasm.

Mutations that affect the capsule were also detrimental for the cells but not to a degree of Tol-Pal mutations. We observe that the mutations in the enterobacterial common antigen (ECA) biosynthesis decrease the cell's ability to grow in the lake water. This polysaccharide moiety is usually attached to the outer leaflet of the outer membrane, and is conserved in the family of *Enterobacteriaceae*. Three forms of ECA were described: bound to the outer membrane lipopolysaccharide, bound to phospholipids, and cyclic. The former two are membrane-associated and exposed to the outer environment and the cyclic ECA is primarily found in the periplasm [168]. The cyclic ECA was shown to play a role in maintaining the outer membrane permeability [171]. The mutants without the periplasmic cyclic ECA had less permeable membranes, and were less sensitive to vancomycin than the WT strain. In the conditions of lake water, it might be important to scavenge for the nutrients and be able to absorb them.

To sum this up, membrane stability and ability to act as a selective barrier is an important factor contributing to cells' viability and ability to proliferate. Both too leaky (Tol-Pal, Lpp mutations) and too tight (ECA biosynthesis mutations) membranes are risk factors that do not allow the *E. coli* cells to thrive in the outer environment.

### 4.1.2. The cells have to rely on their own machinery to produce the nucleotides

We have shown that the mutants in nucleoside biosynthesis pathways have been outcompeted in both filtered and non-filtered water (**Figure 12**). Disruptions in both purine and pyrimidine biosynthesis were detrimental for growth and resulted in lower abundance in the TnSeq data.

Moreover, we observed that the strain Δ*rbsR* that lacks the repressor of the ribose catabolism operon had an advantage over the majority of strains. RbsR, however, not only represses the ribose catabolism, but also the *de novo* synthesis of purine nucleotides from D-ribose 5-phosphate [178]. With this pathway de-repressed, the Δ*rbsR* strains would have an advantage in

nucleotide biosynthesis. This is an illustration how under the low-nutrient conditions, different alternative metabolic functions become de-repressed [10].

Thus, the importance of domestic biosynthesis of nucleotides is brought to our attention twice. First, by the growth defect of the strains unable to synthesize their own nucleotides. Second, by the growth advantage of the strain with the de-repressed biosynthesis pathway. Lake water is unlikely to contain copious amounts of dissolved nucleotides (and any organic matter), and the bacteria have to rely on their own machinery to synthesize them. Lack of nucleotides might inhibit replication, transcription, DNA repair and other cellular processes, and lead to the growth defect.

### 4.1.3. Several mutations give an advantage in the lake water

Several mutations that were beneficial for *E. coli* to proliferate in the lake water were identified in both approaches.

According to the TnSeq results, the mutations resulting in disruption of the glycine cleavage system are beneficial and allow their hosts to grow better in the filtered lake water (**Figure 12**). This group of proteins catalyzes the oxidation of glycine to carbon dioxide and ammonia, yielding a methylene group accepted by tetrahydrofolate and one NADH. This pathway is present in plants, animals and bacteria [176]. In plants, glycine cleavage acts as a salvage reaction during photosynthesis, allowing to recover some of the sugars cleaved by the ribulose 1,5-bisphosphate carboxylase/oxygenase [217]. In mammals, including humans, defects in the glycine cleavage system cause severe untreatable genetic disorder known as glycine encephalopathy [218], resulting in early death. Since most of the studies of glycine cleavage system focus on its importance in plants and mammals, and little is known about its significance in bacteria, we have to speculate why these mutants have a growth advantage in the lake water. We hypothesize it might be caused by the medium composition. The tryptone broth that we used to cultivate our library pre-seeding contains ~0.01% of glycine (~1% dry mass), as opposed to ~0.05% of serine (~5% dry mass) or 0.18% of glutamine (~18% dry mass). The cells potentially did not have enough glycine to produce proteins and could not waste it producing something else.

In both screens, the mutations with altered transcription termination were present and gave their hosts an advantage over the other strains. In the TnSeq screen, it was the Δ*rhoL* strain (**Figure 12**), and in the FB screen, it was the Δ*nusB* strain (**Table 2**). NusB is an

antitermination factor, particularly required for proper transcription of the ribosomal RNA genes. It was demonstrated that the overall transcript levels are lower in the Δ*nusB* strain, and there were more truncated transcripts [219] and these effects decreased at the lower temperature. *rhoL* is a regulatory sequence upstream of the *rho* gene which takes part in *rho* attenuation. The mutations in this region might elevate the expression of *rho* and result in premature transcriptional termination happening more frequently [220], similarly to the phenotype observed in Δ*nusB*. It was demonstrated that Rho-dependent premature termination was a global regulatory mechanism specifically suppressing the expression of the genes that have long 5` untranslated regions (for example, *rpoS*) [221]. Therefore, the expression of *rpoS* should be lower in the Δ*rhoL* and we later demonstrate the importance and the dual role of *rpoS* during the growth in the lake water.

Finally, two knockouts, Δ*rpoS* and Δ*hdfR,* reproducibly grew better in the filtered lake water in both TnSeq and FB screen (**Table 1**, **Table 2**). This result implies that these two genes (or their combination) are especially important for the cells to thrive the lake water conditions. We have selected both of them for an in-depth study, and attempted to describe the functions of *hdfR* and its adjacently transcribed neighbor *maoP*.

### 4.1.4. Dual role of *rpoS*

Most strikingly, mutations that impaired general stress response by disrupting *rpoS* might be beneficial for *E. coli* at least in the conditions of lake water. In both our approaches, Δ*rpoS* strain was selected among the best performing strains in the filtered lake water. This finding comes to a conflict with the existing literature data, because the importance of an intact and active $\sigma^S$ was demonstrated multiple times in both seawater [18, 70] and soil [36, 41, 47]. In the literature reports, complete or partial loss-of-function mutants of *rpoS* had lower viability in the external environments compared to their WT counterparts, and were predated faster by the protozoa in soil. Moreover, some strains isolated from soil had enhanced $\sigma^S$ levels and activity [47], highlighting the importance of this regulatory cascade in the highly competitive outer environments.

On the opposite side, environmental populations of *E. coli* and *Salmonella* contain a high frequency of polymorphisms in *rpoS* [73], including complete loss of its function [74]. The emergence of loss-of-function mutation is further facilitated by growth on poor media and/or non-preferred carbon source such as succinate [74, 75] or arginine [99, 104]. In the evolutionary

experiments, *E. coli* tends to completely lose the activity of general stress response σ-factor RpoS in carbon- and nitrogen-depleted conditions [103, 121]. When the secondary stress is present, such as lowered pH or increased temperature, the different variety of *rpoS* mutants arises. These mutants rather have their *rpoS* activity attenuated but not abolished and are still capable of the limited stress resistance [73, 103]. The mutations of this type are frequently associated with the expression of GASP [106, 107]. Phenotypically, GASP mutations increase the ability of the cells to use various amino acids as sole sources of carbon and energy [108].

The low-nutrient conditions (such as filtered lake water) impose the slow-growth dilemma before the cells. They are sub-optimal for growth, therefore causing accumulation of (p)ppGpp and start of expression of the $\sigma^S$-dependent genes. On the other hand, they still require the high levels of expression of the $\sigma^{70}$-dependent genes to scavenge for the scarce nutrients. In the absence of other stressors beside the carbon limitation, the $\sigma^S$-regulon becomes a burden that provides selective pressure for the occurrence of *rpoS* mutants.

In the conditions of the filtered lake water, the cells did not have any major stress beside nutrient limitation and slight hypoosmotic stress. These parameters of the medium, apparently, gave a growth advantage to the *rpoS* mutants from the TnSeq library, and to the Δ*rpoS* strain in the Keio collection. Basically, we observed these strains expressing the GASP-like phenotype in the filtered lake water. This advantage, however, disappears in the non-filtered lake water where *E. coli* has to deal with the other bacteria and resist predation.

### 4.1.5. Plasmid copy numbers and GFP production differ in several strains of the Keio collection

The fluorescence-based screening unfortunately had a drawback completely unexpected at the beginning of the work. We expected that the usage of the artificial pTrc promoter on the plasmid would result in the equal expression of the GFP across the Keio collection. This assumption was only partially true. The absolute majority of the Keio knockouts indeed had the GFP fluorescence comparable to the WT strain, but there were several outliers that produced more GFP. Thus, in the fluorescence-based screening, we have rather selected the strains favoring GFP production instead of the better growing/surviving ones.

Both Δ*hdfR* and Δ*rpoS* that we selected have higher GFP fluorescence per cell than the WT. We performed qPCR to assess the copy number of the high-copy pTrc99a plasmid we used

for the screening, and a low-copy pUA66 plasmid carrying GFP under the ribosomal *rplN* promoter; and compared these results to the flow cytometry measurements (**Figure 19**).

We determined that Δ*rpoS* mutant has more pTrc99a plasmids per cell and higher GFP levels per plasmid than the WT especially outside of the exponential phase. Similarly, Δ*rpoS* strain had more pUA66 plasmids and higher GFP levels per plasmid than the WT, meaning essentially that the GFP expression was higher in this strain regardless of the plasmid and promoter used. In absence of $\sigma^S$, all the $\sigma^{70}$-dependent genes become upregulated due to higher RNAP availability to the $\sigma^{70}$. The artificial Trc promoter is a hybrid between Trp and Lac promoters [180] both of which are $\sigma^{70}$-dependent, and the ribosomal gene *rplN* is $\sigma^{70}$-dependent too. Thus, these promoters might be upregulated in absence of the $\sigma^S$. To conclude, the increased levels of GFP production by the *ΔrpoS* strain were caused by both increased copy number and increased expression of GFP.

The deletion of *hdfR* only caused the elevation of the copy number in the pTrc99a plasmid. The copy number of pUA66 remained unchanged, making the effect plasmid-specific. The GFP production per plasmid was lower for both plasmids. These results lead us to the conclusion that Δ*hdfR* produced more GFP per cell in our setup simply because it had higher number of pTrc99a but not pUA66.

Another example might be the Δ*pepA* that demonstrated the higher GFP levels in the lake water. It was shown to be involved in proper division and separation of plasmids ColE1 and pSC101. Absence of PepA in the cells causes the formation of plasmid concatemers [222]. The pTrc99a plasmid that we used for screening has the same origin of replication as ColE1 [194]. We did not measure the concrete parameters of this particular strain but we hypothesize that the concatemer formation might have increased the production of GFP by it and therefore give us a false positive.

### 4.1.6. Definition of death in bacteria is a multi-faceted problem

To assess the viability of our experimental strains, we employed two approaches. At first, we used the standard CFU assay, but later switched to the live-dead staining. Both of these methods have their advantages and disadvantages.

The problem of definition of live/death in bacteria is widely discussed in the literature [174, 196, 223, 224], and different approaches measure different parameters. In traditional microbiology

methods, like CFU counting, viability is equated with cultivability. Thus, a number of cells that entered "viable but non-culturable" state would be considered dead, despite their metabolic activity. The viability assessment done by this method tends to underestimate the number of the viable cells.

The other reason behind CFU counting underestimating the actual counts of the viable bacteria is that the cells undergo the so-called nutrient shock upon plating. Often, the bacteria from low-nutrient environments like soil or freshwater fail to grow on the rich media but are able to grow on diluted rich media or on special low-nutrient formula [225-227]. Some protocols of isolation of the oligotrophic bacteria rely on slow and stepwise transition from the low-nutrient media to high-nutrient media [228] specifically to avoid the nutrient shock. This applies not only to the organisms than normally live in the low-nutrient environment but also to the commensals or pathogens temporarily residing there, like *Helicobacter pylori* [226]. For example, it was demonstrated that the colony counts of *E. coli* exposed to the tap water were higher on poor R2A medium than on rich TSA medium despite the plating was performed from the same dilutions [227].

On the other hand, most of the live-dead (including SYTOX Orange that we used) act by forming stable fluorescent complexes with double-stranded DNA, and they can only reach the DNA molecule if the cell membrane is compromised. Therefore, live-dead staining refers to a different definition of death in bacteria, specifically the loss of membrane integrity. In this approach, the number of dead cells is often underestimated [196].

We observed that the Δ*rpoS* strain was less viable in our screening conditions (pTrc99a plasmid, lake water) than the WT regardless of the method used, the plasmid that the cells carried, and the GFP level per cell. However, the discrepancy between the CFU counting and the live-dead staining was quite large (23% and 73% of viable cells respectively). We suspect that this discrepancy is at least partially explained by the nutrient shock, which Δ*rpoS* apparently handles worse than the WT strain and Δ*hdfR*.

Δ*hdfR* had a marginal viability defect compared to the WT strain, but it was not significant. The discrepancy between two different methods was not as dramatic (57% of the viable cells by CFU and 90% by the live-dead staining). Later in this chapter, we prove that Δ*hdfR* is a GASP-like mutation, as Δ*hdfR* mutant has lower levels of $\sigma^S$ and decreased expression of the stationary phase genes. Low but not zero level of $\sigma^S$ in the Δ*hdfR* strain allows it to take the

best of both worlds. This mutant has a competitive advantage in the low-carbon environment due to the lower level of $\sigma^S$, but it is still able to withstand the other stresses (i.e. nutrient shock) due to the still present and active $\sigma^S$.

The viability of the bacteria also depended on a plasmid they carried. We have shown that the cells having the low-copy pUA66a plasmid were in terms of viability comparable to the cells that did not carry any plasmid at all. However, the high-copy pTrc99a plasmid was a major burden, and the effect of the Δ*hdfR* and Δ*rpoS* further amplifies its effect (**Figure 20**).

## 4.2. HdfR/MaoP is a global regulatory system

Like Δr*poS*, Δ*hdfR* strain had a growth advantage in the lake water. Unlike well-studied and well-described *rpoS, hdfR* is studied very poorly and most of the information about it comes from the high-throughput screens. To unravel its role in growth of *E. coli*, we decided to systematically study its deletion strain and the deletion strain of its divergently transcribed neighbor *maoP* in greater detail.

### 4.2.1. HdfR/MaoP as a novel regulator of flagella expression

HdfR was described as a negative transcriptional regulator of the flagellar master operon *flhDC.* To discover the effect of *hdfR/maoP* mutations on flagellar regulation, we tracked the expression of *fliC* in a multitude of strains. We showed that the mutations in this system not only increase the overall level of flagellar gene expression but also alter the shape of the expression curve (**Figure 23**). In both Δ*hdfR* and Δ*maoP* mutants, the period of maximal *fliC* expression was prolonged into the early stationary phase compared to the WT.

By the complementation of the mutant phenotype using *hdfR* and *maoP* expressed from the plasmids, we were able to show that these two genes indeed form a regulatory system that in turn regulates the flagella. We demonstrate that HdfR does not directly repress the flagellar master regulator *flhDC* as described in the literature [134]. Instead, it activated the transcription of *maoP*, which suppresses the *flhDC* activity (**Figure 24**). Moreover, in our ChIP-Seq experiment, we did not observe the binding of HdfR to the *flhDC* promoter region (**Table 5**).

We have performed the tracking of *fliC* expression in two strains of K-12 lineage. One lineage, BW25113 *ΔluxS,* is a derivate of a poorly motile BW25113 Keio collection parental strain. However, due to the spontaneous transposon insertion upstream of *flhDC* it is motile (**Table 12**). This strain has a relatively low baseline expression of *fliC*, and the deletion of

*hdfR* or *maoP* increased both its spreading motility and *fliC* expression dramatically. The other lineage, W3110 RpoS$^{+}$, has a higher baseline expression of *fliC,* and the deletion of *hdfR* or *maoP* does not strongly increase it. However, these mutations still alter the shape of the expression profile (**Figure 23**), expanding the period of the maximal expression into the stationary phase, and result in increased spreading.

The overexpression of *maoP* does decrease the expression of *fliC* but only slightly and without changing the expression profile. Even very high levels of induction do not result in complete loss of *fliC* expression. Apparently, it is not possible to render the cells non-motile by simply overexpressing *maoP* (**Figure 25**). These results suggest that despite the effect of MaoP on flagellar expression, it is not a specific flagellar regulator and has other functions to be discovered.

According to our proteomic data, *maoP* is the sole regulatory target of HdfR at least in our experimental conditions. We can conclude this because the difference in whole proteome between Δ*hdfR* and Δ*maoP* is extremely small, implying that no other targets could be detected at in our setup (**Figure 22**).

### 4.2.2. HdfR/MaoP as a part of σ$^{S}$ regulatory cascade

In our proteomic results, σ$^{S}$ level was lower in both Δ*hdfR* and Δ*maoP* stains. Moreover, it was described in detail [125] that flagellar regulon and σ$^{S}$ regulon are tightly interweaved, and mutually suppress each other at several levels. We have therefore assessed the σ$^{S}$-related processes in the Δ*hdfR* and Δ*maoP* stains, using the Δ*rpoS* as an internal control (**Figure 26**).

First, we measured the expression of *gadB*, which is primary transcribed from the σ$^{S}$-dependent promoter [77, 198]. The expression profile of *gadB* reflects the σ$^{S}$-activity. In the Δ*rpoS* strain, the expression of *gadB* is completely abolished. HdfR/MaoP mutations cause severe reduction in *gadB* expression, combined with the delayed start of it. However, unlike in the Δ*rpoS* strain, *gadB* is still expressed in the Δ*hdfR* and Δ*maoP* mutants.

We have then assessed the biofilm formation by these strains, and the expression of two more σ$^{S}$-dependent genes involved in the biofilm formation (*csgA* and *csgD*). These processes depend on flagella, σ$^{S}$, and the interaction of these regulatory cascades [125]. In all the experiments performed, the deletions in HdfR/MaoP system resulted in severe defect. The expression of *csgD* was decreased in the Δ*hdfR* and Δ*maoP,* but it was not reduced to the

Δ*rpoS* level. The decrease in *csgD* expression was stronger than in *gadB* expression, possibly due to the mutual inhibition of flagellar and $\sigma^S$ regulons [125]. The expression of *csgA* was surprisingly even lower in Δ*hdfR* and Δ*maoP* than in Δ*rpoS.* This result implies more specific regulation of *csgA* by MaoP potentially at CsgD level [125]. Finally, all Δ*rpoS*, Δ*hdfR* and Δ*maoP* had a severe defect in the biofilm formation compared to the WT strain. Biofilm formation is a tightly regulated complex process that requires both downregulation of flagellar and activation of $\sigma^S$ regulatory cascades. Therefore, the severity of the biofilm formation defect might be caused by both lack of $\sigma^S$ and excess of flagellar regulators.

Taken together, these results tell that the RpoS level and/or activity are reduced but not completely abolished if HdfR or MaoP are not present in the cells. The other evidence we could find in the literature is that both Δ*hdfR* and Δ*maoP* strains demonstrate lower rate of stress-induced mutations of both frameshift and base substitution type, and this reduction is comparable to the Δ*rpoS* strain [153]. The most probable cause for reduced $\sigma^S$-dependent activity is the downregulation of $\sigma^S$ itself in the Δ*hdfR* and Δ*maoP* strains, as observed in our proteomics data.

There are no determined promoters for *maoP* yet. However, two promoters are predicted for *hdfR,* one on which is under $\sigma^S$ regulation [148] and the other one is regulated by $\sigma^F$ [149]). However, in our proteomic data we could not see any decrease in HdfR levels in both Δ*rpoS* (**Figure 26**) and Δ*flhD* (**Supplementary figure 9**) strains, suggesting that both of these promoters are not as active. We examined the upstream sequence of *hdfR* and found a sequence that rather resembled a $\sigma^{70}$-dependent promoter. Based on the timing of the peak expression of *maoP* and the presence of the $\sigma^{70}$-dependent promoter of *hdfR*, we argue that the transcription of *hdfR* and consequently *maoP* is dependent on the $\sigma^{70}$.

We can conclude then, that HdfR/MaoP system not only suppresses flagella, but also positively regulates *rpoS*.

### 4.2.3. HdfR and MaoP might play a role in transition to the stationary phase

The growth phases of *E. coli,* as well as transition between them and the regulatory cascades involved in this process are beautifully described in the great detail in the seminal review by Regine Hengge [76]. Briefly, she defines three main phases of growth. In the Phase I, or the exponential phase, the cell density is low and the resources are still abundant. The cells grow rapidly and devote most of their resources to the synthesis of ribosomes.

In the Phase II, the resources are still abundant but the cell density is higher, and the cells start to enter the state of carbon limitation. The expression of numerous genes is activated, including most prominently the flagellar master regulator *flhDC* and the other genes of the flagellar cascade [125]. At this point, the genes governed by *flhDC* and $\sigma^F$ have the highest rate of expression. During this phase, (p)ppGpp starts to accumulate, causing gradual reduction in the ribosomal gene expression and increasing availability of RNAP for alternative σ-factors, such as $\sigma^F$ and $\sigma^N$ . The importance of $\sigma^S$ is low at this point, because the activity of $\sigma^S$-dependent genes is suppressed by the regulatory factor FliZ belonging to the flagellar gene hierarchy.

In the Phase III, or the stationary phase, $\sigma^S$ accumulates in the cells and activates the $\sigma^S$-dependent genes. The levels of $\sigma^{70}$, $\sigma^N$ and $\sigma^F$ decrease due to the (p)ppGpp accumulation and the activity of their respective anti-σ-factors. As a result, the expression of various nutrient-scavenging systems is diminished, and cells become nonmotile. The remaining resources of the cell are now invested in maintenance and survival [77].

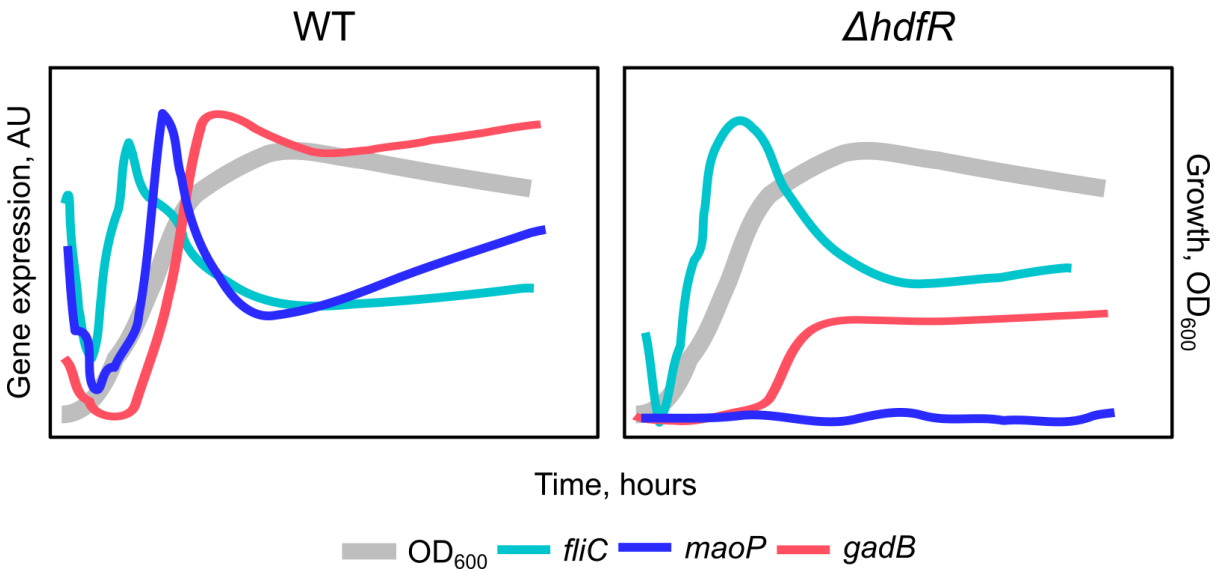



**Figure 31. Absence of *hdfR* and *maoP* leads to perturbations when entering the stationary phase.** Shown are the schematic expression profiles of *fliC, maoP* and *gadB* overlayed with a typical growth curve.

In our work, we measured the expression of multiple genes over the course of incubation. They are expressed in the different phases of growth. For example, it is typical for *fliC* expression to have a sharp peak that quickly declines by early stationary phase. On the

opposite side, the expression of *gadB* is low during the active growth and then sharply increases, reflecting the accumulation of $\sigma^S$, and remains high and steady.

In the Δ*hdfR* mutant, that does not express *maoP,* period of the maximal *fliC* expression is expanded further into the stationary phase (**Figure 31**B). Moreover, depending on the background strain, the peak itself might be higher than the WT (**Figure 23**). The start of *gadB* expression is delayed compared to the WT strain, and the steady level is lower. These results imply that the Δ*hdfR* and Δ*maoP* strains struggle to properly enter the stationary phase, and the described cascade of the regulatory signals is disrupted.

Given that, we conclude that HdfR/MaoP system might be the players in transition to the stationary phase. The Δ*hdfR* and Δ*maoP* strains have reduced but not completely abolished activity of the $\sigma^S$, making them similar to the strains exhibiting the GASP phenotype [106]. Indeed, truncation of *hdfR* was once described as a GASP mutation [229]. This explains why the Δ*hdfR* and Δ*maoP* strains were overrepresented in the filtered lake water. Their GASP-like phenotype gave them the growth advantage.

### 4.2.4. MaoP interactome: RNAP and DNA polymerase I

We have explored the protein-protein interaction of MaoP using two complementary approaches: a standard co-IP [202] and a BioID proximity labelling [203]. In both of these experiments, we observed that non-essential subunits of the RNAP were found in the interactor fraction of MaoP (**Figure 28**). These subunits were two alternative σ factors ($\sigma^S$ and $\sigma^N$) and an ω-subunit (RpoZ) of RNAP. Discovery of the RNAP subunits by two different methods suggests that MaoP physically interacts with RNAP. Absence of the essential subunits in our data is probably explained by the order of RNAP core assembly ($2\alpha > \alpha2 > \alpha2\beta > \alpha2\beta\beta'\omega$ [120]).

The functions of the ω-subunit are not fully described yet. It is not required for transcription but it might act as a chaperone to stabilize and ensure the correct folding of its interactor β`-subunit [204]. The ω-less RNAP is reported to recruit $\sigma^S$ with higher efficiency [205]. The other reported phenotypes of *rpoZ* deletion include relaxation of DNA supercoiling and upregulation of $\sigma^S$ regulon [205]; reduced affinity of RNAP to (p)ppGpp and defect in biofilm formation [230]; and increased binding of RNAP to the prophage genes promoters [231].

The hypothesized role for the ω-subunit is to make the RNAP core more selective to the $\sigma^{70}$ [204]. Our results show that the *fliC* expression is elevated in the Δ*rpoZ* strain but this effect is independent from the Δ*maoP* mutation, making us reject the hypothesis that these two proteins interact (**Figure 29**). Observed physical interactions between RNAP and MaoP hint that it is a global regulator of transcription. We hypothesize that it might increase the selectivity of RNAP to $\sigma^S$, or help to deliver $\sigma^S/\sigma^N$ to the RNAP core.

Another interesting interaction partner of MaoP was DNA polymerase I (PolA). It combines a DNA polymerase activity, a 5'→3' exonuclease activity and a 3'→5' proofreading exonuclease activity making it extremely multifunctional enzyme. It is involved in several pathways of DNA repair, including nucleotide excision repair (NER) [206, 207] and post-replication repair [208]. The link between transcription and NER was observed long ago [209], when it was observed that the active running transcription increases the rate of DNA repair. However, it was considered the minor subpathway of DNA repair. This idea is currently contested by the results [210] showing that in fact coupling with the transcription is absolutely necessary for the successful NER. PolA is shown to be involved in initiation of replication of ColE1 plasmids [211]. Earlier, we have observed the elevation of the copy number of pTrc99a plasmid that belongs to the ColE1 group in the Δ*hdfR* deletion strain. Potentially, the increase in the copy number might be caused by absence of interaction between MaoP and PolA.

However, the most important and well-known function of PolA is maturation of the Okazaki fragments during the chromosome replication by replacing RNA primers with DNA [232, 233]. This is possible due to unique 5′→3′ exonuclease activity of PolA small subunit [234]. This activity is essential for the cells, and the Δ*polA* mutants have to be supplemented with the 5′→3′ exonuclease *in trans* in order to remain viable [235].

### 4.2.5. MaoP deletions influence chromosome segregation

MaoP was first described in the literature as a protein that limits the mobility of the chromosome in the *ori* region [154]. The movement of the DNA loci in the cell is highly constrained, and the distribution of these loci correlates with macrodomain organization of the chromosome [155]. It was reported that inactivation of *maoP* lifted some of the constraints on DNA mobility in the Ori macrodomain and allowed long-range interaction between Ori and Right macrodomains.

Moreover, our own results indicate the possible involvement of MaoP and/or HdfR in DNA replication (i.e. potential presence of PolA in the MaoP interactome (**Figure 28**), increased copy numbers of the ColE1 plasmid in the Δ*hdfR* strain (**Figure 19**) and massive position-dependent upregulation of the genes situated close to the *ori* upon *hdfR* overexpression (**Figure 25**). Based on this and the literature data mentioned above, we used the microfluidics approach to study the cell division of the WT and the Δ3 mutant lacking the entire HdfR/MaoP system including their intergenic region. On top of that, we tracked the number and position of labelled *ori* and *ter* foci.

In comparison to the WT, the *ori* splitting of the Δ3 happened later in the cell cycle (**Figure 30**). Since *ori* splitting and *ter* centralization happen at the same time [215], the *ter* centralization was delayed in the Δ3 too. Surprisingly, these changes in the chromosome dynamics did not result in reduced growth rate, cell elongation or other signs of stress. We will still have to study the effects of HdfR and MaoP on the chromosome dynamic in the greater detail but our data suggests its importance for proper chromosome organization.

We need to note here that in the study by Valens et al. [154] they observed the increase in mobility of the DNA loci situated close to the *ori* in the Δ*maoP* strain. In our results, we did not observe such an increase, and rather saw tighter and more defined localization of *ori* in the Δ3 strain. However, Valens et al. used a low-copy P1 plasmid to express the ParB fusion and label the *ori* [236], and we used the high-copy ColE1-derived one. This might have caused an artifact in our results.

## 4.3. Conclusions and future prospective

There is increasing evidence that at least some strains of *E. coli* are robust enough to survive and reproduce outside of the host for prolonged periods of time, and even integrate into the local microbial community. Due to *E. coli* importance to public health as both contamination marker and a facultative pathogen, it is crucial to be able to track and predict its growth and survival outside of the intestine. In this work, we attempted to discover the genetic determinants of its growth in the lake water besides the well-described *rpoS*.

First, we discovered major types of mutations that greatly reduce the growth of *E. coli* in the outside environments. These are mutations that cause the cell envelope defects, and the mutations in the nucleotide biosynthesis pathways.

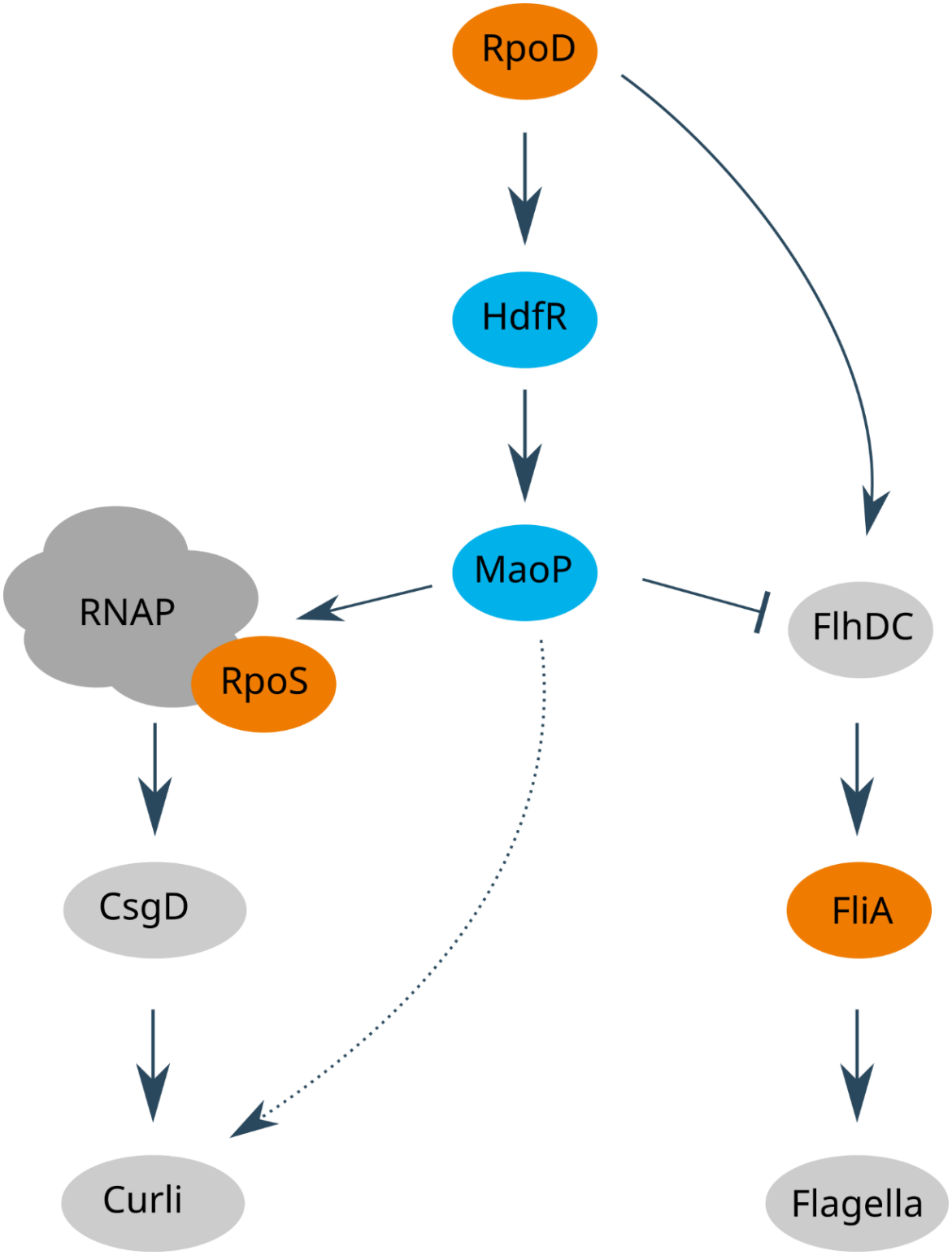



**Figure 32. Place of HdfR and MaoP in the cross-talk between flagella and $\sigma^S$.** Based on [125]. The potential but not proven interaction is shown as the dashed line. Sigma factors are shown in orange.

Second, we determined that the role of *rpoS* was in fact dual, and absence of it might be advantageous. We demonstrate that the *E. coli* strains demonstrating GASP-like phenotype (like Δ*rpoS* or Δ*hdfR*) have a growth advantage in the lake water environment in case there is no other stress present.

Third, we discovered and described the new regulatory pathway involved into transition to the stationary phase. According to our data, the HdfR/MaoP system acts upstream of both flagellar and $\sigma^S$ regulatory cascades, inhibiting the former and activating the latter, and thus guides the cell through the $\sigma^{70} – \sigma^S$ transition (**Figure 32**).

There are, however, still several unresolved questions. We still do not know how MaoP suppresses flagella and activates RpoS. There are hints that it happens at the transcriptional level via the physical interaction to the RNA polymerase but the exact mechanism is yet to be deciphered. Another potential target for the research is to determine the signals that activate and inhibit the HdfR/MaoP system. The timing of the *maoP* expression peak is quite remarkable, as it happens after the peak expression of the *flhDC*-dependent genes but before the expression of the $\sigma^S$-dependent genes. Deciphering the regulatory cascade behind its activation and suppression might shed more light on transition of *E. coli* into the stationary phase. Last but not least, the link between the chromosome organization and the transition to the stationary phase needs to be investigated.

# Materials and methods

## 5.1. Materials

**Table 6. Devices**

| Name | Manufacturer |
|---|---|
| MicroPulser Electroporator | BioRad |
| MSC-Advantage™ Class II Biological Safety Cabinet | ThermoFisher Scientific |
| VIAFLO 96/384-well pipette | Integra |
| Digital salinity meter/thermometer | Hanna Instruments |
| BD LSRFortessa SORP cell analyzer | BD Biosciences |
| Zeiss Axio Observer Z1 widefield fluorescence microscope | Zeiss |
| Nicon Ti microscope | Nicon |
| CFX384 Touch Real-Time PCR Detection System | BioRad |
| PeqStar Thermocycler | VWR |
| Tecan Infinite M200 Pro plate reader | Tecan |
| Tecan Infinite M1000 Pro plate reader | Tecan |
| Bioruptor Standard sonicator | Diagenode |
| Digital tube revolver | ThermoFisher Scientific |
| Mini-Protean Tetra SDS-PAGE system | BioRad |
| PerfectBlue 'Semi-Dry' Electro Blotter | VWR |
| Agilent Bioanalyser 2100 | Agilent |
| Precellys Evolution cell homogenizer | Bertin Technologies |

**Table 7. Consumables**

| Name | Manufacturer |
|---|---|
| Anti-FLAG primary antibodies, produced in mouse | Sigma Aldrich |
| Anti-GFP primary antibodies, produced in mouse | Clontech |
| IRDye® 800CW Goat anti-Mouse IgG Secondary Antibody | Licor |
| IRDye® 800CW Streptavidin | Licor |
| 96-deepwell plates (MegaBlocks) | Sarstedt |
| 96-well plates, 48-well plates, 24-well plates | Greiner Bio-one |
| Bottle-top filters (500 ml, polyethersulfone membrane, pore size 0.22 µm) | Merck |
| SYTOX Orange live-dead dye 5 mM | Invitrogen |
| GeneRuler 1 kb DNA-Ladder, GeneRuler 1 kb Plus DNA-Ladder | ThermoFisher Scientific |
| Gibson Assembly Master Mix | New England Biolabs |
| Proteinase K | ThermoFisher Scientific |
| T4 ligase | New England Biolabs |
| DpnI | New England Biolabs |
| SalI | New England Biolabs |
| XbaI | New England Biolabs |
| 8-well µ-slides | ibidi |
| High Precision microscopy slides #1.5 24x60 mm | Marienfeld |
| Poly-L-lysine, 0.1% | Sigma Aldrich |
| Dream Taq Polymerase | ThermoFisher Scientific |
| Q5 polymerase | New England Biolabs |
| 384-well plates for qPCR | BioRad |
| 96-well plates for PCR | Sarstedt |
| KAPA SYBR Fast qPCR Master Mix (2X) | KAPA Biosystems |
| Precision Blue™ Real-Time PCR Dye | BioRad |
| PD-10 desalting columns | Cytiva |

| **Name** | **Manufacturer** |
|---|---|
| Proteinase inhibitors | ThermoFisher Scientific |
| Dynabeads™ M-280 Streptavidin | Invitrogen |
| Binding Control Magnetic Agarose Beads | ChromoTek |
| GFP-Trap® Magnetic Agarose beads | ChromoTek |
| Mini-PROTEAN TGX Precast Gels | BioRad |
| PageRuler™ Prestained Protein Ladder | ThermoFisher Scientific |
| Roti-Load 4x Loading dye | Roth |
| Amersham Hybond P 0.2 µm PVDF Western blotting membrane | Cytiva |
| AMPure Beads | Beckman |
| Glycogen 20 mg/ml | ThermoFisher Scientific |
| Phenol-chlorophorm-isoamyl alcohol | Roth |

**Table 8. Reaction kits**

| **Name** | **Manufacturer** |
|---|---|
| Bioanalyzer High Sensitivity dsDNA Kit | Agilent |
| Qubit High Sensitivity dsDNA Kit | Invitrogen |
| NucleoSpin Microbial DNA Kit | Macherey Nagel |
| NEBNext Ultra™ II DNA Library Prep | New England Biolabs |
| NEBNext® Multiplex Oligos for Illumina® (96 Index Primers) | New England Biolabs |
| GeneJET PCR Purification Kit | ThermoFisher Scientific |
| GeneJET Plasmid Midiprep Kit | ThermoFisher Scientific |
| GeneJET Plasmid Miniprep Kit | ThermoFisher Scientific |

The kits were used according to the guidelines given by the manufacturers.

**Table 9. Media**

| **Medium** | **Composition per liter** |
|---|---|
| Luria-Bertrani (LB) medium | 10 g bacto tryptone, 5 g yeast extract, 8.55 mM NaCl<br>Adjust pH 7 with NaOH |
| Luria-Bertrani (LB) agar plates | 1.5 % agar in LB medium<br>autoclaved |
| LB cryo storage medium | 20% glycerol in LB medium |
| M9 minimal medium | 48 mM $Na_2HPO_4$ x $7H_2O$, 22 mM $KH_2PO_4$, 8.55 mM NaCl, 18.7 mM $NH_4Cl$, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4 % necessary carbon source |
| Super Optimal Broth (SOB) medium | 20 g bacto tryptone, 5 g bacto yeast extract, 8.55 mM NaCl<br>Adjust pH 7 with 1 M NaOH<br>Before inoculation add 10 mM $MgSO_4$ and 10 mM $MgCl_2$ |
| Tryptone broth (TB) medium | 10 g bacto tryptone, 8.55 mM NaCl, $H_2O$<br>Adjust pH 7 with 1 M NaOH |
| Swarming agar | 0.27 % agar in TB medium |

All media except M9 were autoclaved at 120°C for 20 minutes. The M9 medium was sterilized by filtration through the filter membrane with the pore size 0.22 µm.

**Table 10. Buffers**

| **Buffer** | **Composition per liter** |
|---|---|
| Phosphate buffered saline (PBS) 10x stock | 1.4 M NaCl, 27 mM KCl, 15 mM $KH_2PO_4$, 81 mM $Na_2HPO_4$<br>Adjust pH with HCl to 7.4, autoclave |
| TAE buffer | 40 mM Tris base, 1 mM EDTA, 20 mM glacial acetic acid, pH 8 |
| Tethering buffer | 10 mM KPO4, 0.1 mM EDTA, 1 M methionine, 10 mM lactic acid,<br>pH 7.0, sterile filtered |
| ChIP lysis buffer | 20 mM Tris pH 8.1, 150 mM NaCl, 2 mM EDTA, 0.1% SDS, 1% Triton X-100<br>High salt variation: same composition except 500 mM NaCl |
| ChIP wash buffer | 10 mM Tris pH 8.1, 0.25 M LiCl, 1 mM EDTA,1% IGEPAL CA-630, 1% SDC |
| ChIP elution buffer | 10 mM Tris pH 8, 300 mM NaCl, 1% SDS |
| HNN lysis buffer | 50mM HEPES pH 7.5, 150mM NaCl, 50mM NaF, 0,5% IGEPAL CA-630<br>Add IGEPAL CA-630 immediately prior to use |
| RIPA buffer | 50 mM Tris pH 7, 150 mM NaCl, 1% Triton, 0.5% SDC, 0.2% SDS |
| SDS-PAGE running buffer | 25 mM Tris, 192 mM glycine, 0.1% SDS |
| Western blot transfer buffer | 25 mM Tris, 192 mM glycine, 20% methanol |
| Tris-buffered saline with Tween 20 (TBS-T) | 150 mM NaCl, 20 mM Tris, adjust pH to 7.6, autoclave. add 0.1% Tween 20 prior to use |

All buffers were sterilized by filtration through the filter membrane with the pore size 0.22 μm.

**Table 11. Stock solutions**

| **Solution** | **Concentration** |
|---|---|
| Ampicillin 1000x | 100 mg/ml |
| Chloramphenicol 1000x (in ethanol) | 34 mg/ml |
| Kanamycin 1000x | 50 mg/ml |
| L-arabinose | 10%, 1% |
| Isopropyl-β-D-thiogalactopyranosid (IPTG) | 100 mM, 10 mM |
| Biotin | 50 mM |
| Glucose | 20% |
| Glycerol | 10%, 87% |
| $CaCl_2$ | 1M, 0.1 M |
| Formaldehyde | 16 % in PBS, prepared from paraformaldehyde |
| Glycine | 2.5 M |

All stock solutions were sterilized by filtration through the filter membrane with the pore size 0.22 µm.

**Table 12. Differences between the reference *E. coli* K-12 strains and the strains used in the work.**

| Strain | Mutation | Position | Effect on protein | Gene | Description |
|---|---|---|---|---|---|
| W3110 RpoS$^+$ [52] | G→A | 228,602 | noncoding | Y75_RS01035 | 23S ribosomal RNA |
| | A→G | 547,694 | pseudogene | *ylbE* | DUF1116 domain-containing protein |
| | +G | 547,835 | Potentially restored function | *ylbE* | DUF1116 domain-containing protein |
| | A→T | 556,858 | L36Q | *folD* | bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase |
| | C→T | 1,065,700 | R415C | *yccE* | YccE family protein |
| | T→C | 1,093,686 | V130A | *dgcT* | Diguanylate cyclase |
| | A→G | 1,303,712 | S273G | *oppA* | oligopeptide ABC transporter substrate-binding protein OppA |
| | A→G | 1,739,298 | intergenic | *ynhF/purR* | cytochrome bd-I accessory subunit CydH/HTH-type transcriptional repressor PurR |
| | C→A | 2,866,109 | stop33Y (restored function) | *rpoS* | RNA polymerase sigma factor RpoS |
| | C→G | 2,906,975 | D450H | *pyrG* | CTP synthase (glutamine hydrolyzing) |
| BW25113 Δ*luxS* [160] | IS5 insertion | 166,147 | disruption | *fhuA* | ferrichrome outer membrane transporter |
| | A→G | 799,201 | E81E | *ybhJ* | putative hydratase |
| | G→T | 906,459 | A16E | *poxB* | pyruvate dehydrogenase (pyruvate oxidase), thiamine triphosphate-binding, FAD-binding |
| | IS1 | 1,973,088 | intergenic | *flhD/uspC* | flagellar master regulator/universal stress protein |
| | T→C | 2,719,426 | intergenic | *kgtP/rrfG* | alpha-ketoglutarate transporter/5S ribosomal RNA of rrnG operon |

| | | | | | |
|---|---|---|---|---|---|
| MG1655 [72] | IS1 | 1,873,031 | disruption | *dgcJ* | putative diguanylate cyclase DgcJ |
| | +GC | 4,296,381 | intergenic | *gltP/yjcO* | glutamate/aspartate : H(+) symporter GltP/Sel1 repeat-containing protein YjcO |

For W3110 RpoS$^{+}$, the mutation in *rpoS* restoring its function is highlighted in blue. For BW25113 Δ*luxS*, the mutation restoring its motility is highlighted in orange. The mutation list is created using BRESEQ software [237].

The complete list of the strains can be found in the **Supplementary table 5**.

**Table 13. Plasmids**

| Name | Backbone | Resistance | Insert | Reference |
|---|---|---|---|---|
| *csgA* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *csgD* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *flhD* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *fliA* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *fliC* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *gadB* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| *rplN* reporter | pUA66 | Kanamycin | GFP-mut2 | [192] |
| pSIJ8 | pKD46 | Ampicillin | lambda Red recombineering genes and flippase gene in independent operons | [238] |
| pNB1 | pTrc99a [180] | Ampicillin | sfGFP | [239] |
| pNT11 | pTrc99a | Ampicillin | *hdfR* | |
| pNT12 | pTrc99a | Ampicillin | *maoP* | |
| pNT13 | pBAD33 [240] | Chloramphenicol | *hdfR* | |
| pNT14 | pBAD33 | Chloramphenicol | *maoP* | |
| pNT16 | pTrc99a | Ampicillin | HdfR-sfGFP translational fusion | |
| pNT20 | pTrc99a | Ampicillin | MaoP-sfGFP translational fusion | |

| Name | Backbone | Resistance | Insert | Reference |
|---|---|---|---|---|
| pNT23 | pTrc99a | Ampicillin | sfGFP-MaoP translational fusion | |
| pNT24 | pTrc99a | Ampicillin | TurboID-sfGFP translational fusion | 241 |
| pNT25 | pTrc99a | Ampicillin | TurboID-HdfR translational fusion | |
| pNT26 | pTrc99a | Ampicillin | TurboID-MaoP translational fusion | |
| pNT26 | pTrc99a | Ampicillin | MaoP-TurboID translational fusion | |
| pNT27 | pTrc99a | Ampicillin | HdfR-TurboID translational fusion | |
| pNT32 | pBAD33 | Chloramphenicol | HdfR-sfGFP translational fusion | |
| pFHCP1-mTurquoise2 | pFHCP1 | Ampicillin | ParB- mTurquoise2 translational fusion | Constructed by Ismath Sadhir |
| pFHCP1-mScarlet-I | pFHCP1 | Ampicillin | ParB- mScarlet-I translational fusion | Constructed by Ismath Sadhir |
| pVS1515 | pTrc99a | Ampicillin | eGFP | |

The complete list of the oligonucleotides can be found in **Supplementary table 6**.

## 5.2. Methods

**Cloning.** The knockouts of the selected genes were constructed using λ-red mediated recombination, and kanamycin resistance cassette was removed via FLP-recombination as described [242]. The parental strains were transformed with a heat-sensitive plasmid carrying both λ-red recombinase and FLP-recombinase [238]. The DNA fragment containing the flanking regions and kanamycin resistance cassette was PCR-amplified from the respective Keio collection knockout strain with Q5 polymerase and then purified. The cells were grown in SOB medium in presence of 0.15% L-arabinose at 30°C to $OD_{600}$ = 0.4. The cells were then washed three times with ice-cold 10% glycerol, and the PCR fragment was electroporated into the cells. Cells were then incubated in SOB + 0.8% glucose for 2-3 hours at 30°C, and plated onto selective plates to grow overnight 30°C. To remove the resistance cassette, the culture was incubated overnight at 30°C in presence of 50 mM rhamnose. Both recombinations were PCR-confirmed. The plasmid was then removed via passaging the

culture several times at 45°C. Plasmids used in the study were constructed using either Gibson assembly [243] or standard restriction-ligation protocol according to the manufacturer's guidelines.

**High-throughput transformation.** To transform the Keio collection [160, 161] with the GFP-carrying plasmid, deepwell plates and 96-channel pipette (Integra Viaflo 96/384) were used. Bacteria were grown overnight in the deepwell plates in the rotary shaker at 37°C on LB medium with 50 µg/ml kanamycin. After that, they were inoculated to the fresh medium and grown to the $OD_{600}$ of 0.6 - 0.8. Cells were washed with ice-cold 0.1 M $CaCl_2$ and resuspended in 0.1 M $CaCl_2$. After that, the plasmid was introduced and the cells underwent heat shock (2 minutes at 42°C). To develop the resistance, cells were incubated in LB for 1.5 hour at 37°C. Then, they were inoculated to the selective medium (LB, 100 µg/ml ampicillin, 50 µM IPTG). After the overnight incubation, glycerol stocks were prepared.

**Water sampling.** The lake water was taken between mid-April and mid-September from the lake near Marburg, Germany (GPS coordinates: 50.830471, 8.780292). We carefully avoided taking samples shortly after the rain. Prior to sampling, we measured temperature of air, temperature of water, salinity, and pH (**Supplementary table 1**). Approximately 1.5 L of water was taken each time. This water was later filter-sterilized through the filter membrane with the pore size 0.22 µm or used as is in its native state. To work during the winter time, we took approximately 4 liters of water, filtered it and stored at 4°C.

**Plate reader screening.** To perform the fluorescence measurements, cultures of *E. coli* were grown overnight at 37°C on TB medium in a rotary shaker and then inoculated in 96-well plates (Greiner) filled with native (taken from the lake without any additional treatment) or filter-sterilized lake water. The water was supplemented with 50 µM of IPTG. The cultivation took place at the room temperature. The measurements of $OD_{600}$ and GFP fluorescence (excitation 485 nm, emission 520 nm) were performed on TECAN Infinite M1000 Pro right after the inoculation and then every 24 hours for 8 days.

**Growth and gene expression assessment.** Growth curves and gene expression were assessed in 24- or 48-well plates using a plate reader. Briefly, cultures of *E. coli* were grown overnight in TB in a rotary shaker at 37°C, supplemented with antibiotics if necessary. The $OD_{600}$ was then measured, and the cultures were diluted in fresh medium to an $OD_{600}$ of 0.02 and loaded to the well plates (1 ml per well for 24-well plates or 0.4 ml per well for 48-well

plates). If the measurements had to be done in the minimal medium, the culture was washed 2 times in PBS before loading. The measurements of $OD_{600}$ and GFP fluorescence (excitation 485 nm, emission 520 nm) were performed on TECAN Infinite M1000 Pro plate reader at 37° or 30°C every 10 minutes with a combination of orbital and linear shaking in between.

**Transposon mutant libraries: construction.** The transposon mutant library of W3110 $RpoS^+$ was kindly gifted by Ananda Medeiros. It was constructed via delivery of randomly barcoded Tn5 system to the cells, which is turn randomly inserted the labelled barcodes into the genome. Each barcode is flanked by the common PCR-binding sites U1 and U2, enabling the amplification of the barcode sequence. To create the comprehensive map of the library, we have extracted the genomic DNA from the library samples, sonicated them to the average size of ~300 bp, and used the NEBNext Ultra™ II DNA Library Prep Kit and NEBNext® Multiplex Oligos to prepare the sequencing library. To enrich the barcode sequences in the library, the primer to U1 was used at the last step of the protocol. The sequences and positions of the barcodes were then determined by deep Illumina MiniSeq sequencing, using the script written on Python version 3.9.12 with Biopython version 1.78 [244]. Briefly, every read was checked for presence of both U1 and U2, and the sequence between them underwent the quality check (length is exactly 20 bp and Q>30 for each base). The reads containing the barcodes were then aligned to the *E. coli* K-12 W3110 reference genome using *blastn* from the *ncbi-blast* toolkit [245]. The sequence and the position of each barcode were then combined in a single csv file.

**Transposon mutant libraries: experiment.** To start the work with the library we thawed the glycerol stocks, inoculated the whole content of the stock into 50 ml TB medium and grew at 37°C in the rotary shaker until $OD_{600}$ reached ~1. Then we inoculated these cultures into the lake water in the 24-well plates and took samples at days 2, 4, and 8 of the cultivation. The starter culture was used as the control. We extracted the genomic DNA with Macherey-Nagel Microbial DNA kit and PCR-amplified the barcode sequences using Q5 polymerase supplemented with GC-enhancer and barcoded tru-seq derived primers to U1 and U2 sequences (200 nM of each, forward primers carried an individual barcode to allow multiplexing and the reverse primer was common). The amount of the template genomic DNA was 150-200 ng. The cycling conditions were 98°C for 4 min followed by 25 cycles of 30 s at 98°C, 30 s at 55°C, and 30 s at 72°C, followed by a final extension at 72°C for 5 min.

The PCR-reactions were pooled, purified with 0.9X of Beckman AMPure beads and sequenced using Illumina NextSeq 500 High platform.

**Transposon mutant libraries: data analysis.** To analyse the results of the transposon sequencing, we used a custom script written on Python version 3.9.12 with Biopython version 1.78 [244]. This script determined the barcode sequence from each read and counted the number of each barcode in each sample. The fitness calculations were performed in R. The fitness of each knockout strain was calculated as the normalized $\log_2$ ratio of counts in experimental and control conditions [159]. Gene fitness is the weighted average of the strain fitness, where strains with more reads are weighted more highly. To eliminate the potential bias of DNA extractions, the fitness data were normalized to the chromosome region. For this, we use the running median (window=251 gene) of the gene fitness values across the chromosome and subtract it from the fitness values. Please refer to [159] for more details.

**Flow cytometry.** Total number of cells, number of GFP-positive cells per milliliter of lake water and intensity of GFP fluorescence were measured using BD LSRFortessa SORP cell analyzer (BD Biosciences, Germany) with the high-throughput system for 96-well plates. The cultures were well mixed using a pipette and then diluted in phosphate buffered saline. For excitation of GFP we used a 488 nm laser, for excitation of mCherry and SYTOX Orange we used a 561 nm laser. To analyze the flow cytometry results, we used FlowJo software (BD, USA) version 10.

**SYTOX Orange live-dead staining.** Cells were washed once with PBS. Then, the SYTOX Orange dye was added to the cell suspension in a final concentration of 0.2 µM, and the mixture was incubated at 37°C for 15 minutes with vigorous shaking. After staining, the cells were washed with PBS 2-3 times again and prepared for the flow cytometry.

**Widefield fluorescent microscopy.** To attach the cells, wells of 8-well µ-slides were treated for 10 minutes with 0.1% poly-L-lysine solution [246] and then washed with the tethering buffer. Then, the washed cells were loaded to the wells and allowed to attach for 10 minutes. Free-floating cells were then discarded, and the wells were then washed thoroughly with the tethering buffer. Alternatively, we prepared an agarose pad (1% agarose in tethering buffer) using several glass slides stacked together. After solidifying, we cut the pad into ~1x1 cm squares. 3-5 µL of washed bacterial cells were then pipetted onto a high precision microscopy slide and covered with the agarose pad [247]. Imaging took place on Zeiss Axio Observer Z1

widefield fluorescence microscope with 100X oil DIC 1.46 NA objective. The excitation/emission wavelengths are listed in the **Table 14**.

**Table 14. Fluorophores used for microscopy**

| **Fluorophore** | **Excitation wavelength, nm** | **Emission wavelength, nm** |
|---|---|---|
| mTurqoise2 | 434 | 474 |
| sfGFP | 488 | 509 |
| YPet | 517 | 530 |

**Microfluidics.** For *ori* and *ter* tracking, we used the strains kindly given by Ismath Sadhir. They produced MatP-Ypet fusion to label the *ter* region of the chromosome, and mTurquoise2-ParB fusion to label the *ori* region. To enable the ParB labelling, the *parS* sequence, a binding target of ParB, was introduced close to the *ori* [248]. The strains were grown overnight at 30°C in M9 minimal medium supplemented with 0.4% glycerol. Three hours prior to the experiment, 50 µM IPTG was added to induce the expression of mTurquoise2-ParB. The microfluidics chip (mother machine) was prepared according to the protocol described in [248]. The design consists of a main channel through which nutrient media flows and narrow growth-channels in which cells are trapped. Briefly, polydimethylsiloxane (PDMS) mixture composed of a ratio of 1:7 (curing agent:base) was poured over the silicon wafer with the design, and allowed to degas at the low pressure. The chip was then incubated overnight at 80°C. Before imaging, the chip is bonded to a glass slide using a plasma generator (30 s at 75 W) [214]. After loading, images were acquired every 10 minutes on a Nikon Ti microscope with a 100 x/1.45 oil objective and a Hamamatsu Photonics camera using phase contrast and two fluorescence channels. Segmentation and analysis of the acquired images were performed using *Mothersegger* software [248].

**CFU estimation.** Samples containing bacteria were serially diluted tenfold in PBS and then 50 µl of 3 dilutions were plated onto LB agar in duplicates using glass beads. The dilutions (typically $3^{d}$-$5^{th}$) were selected based on flow cytometry results. After the overnight incubation at 37°C the resulting colonies were counted either manually or using a custom-written ImageJ macro [249]. Plates containing less than 100 or more than 500 colonies were considered unreliable. The number of CFU per milliliter was determined using a formula

$CFU = \frac{N*10^d}{V}$ where N is a number of colonies per plate, d is a dilution and V is a volume of this dilution plated in milliliters.

**qPCR.** For qPCR we incubated the bacterial cultures at 95°C for 10 minutes to disrupt the cells, followed by freezing at -20°C [250]. These samples were diluted and used as a template for amplification. Real-time qPCR reactions were performed in triplicate using 10 µL mixtures. The BioRad CFX384 instrument was used for amplification and detection. For each reaction, we used KAPA SYBR FAST 2x master mix, 1-4 µl of diluted cell lysate, 200 nM of each primer to either GFP or a chromosomal gene *dxs*. The cycling conditions were the following: 3 minutes at 95°C, followed by 40 cycles of 10 seconds at 95°C and 30 seconds at 60°C. After the 40$^{th}$ cycle, the melting curves were assessed by increasing temperature from 65°C to 95°C with increment 0.5°C. To determine the number of plasmids, we used the formula $plasmid\_copy = 2^{-(Cq_{gfp}-Cq_{dxs})}$, where Cq is a quantification cycle for *dxs* and *gfp* respectively [193]. For analysis of qPCR data, we used BioRad CFX manager software.

**Spreading motility assay.** Bacterial cultures were grown overnight at 37°C in TB supplemented with an antibiotic if necessary. The swarming agar medium (TB+0.27% agar) was melted and allowed to cool down to approximately 50-55°C, poured into square 10x10 cm Petri dishes and allowed to solidify in the sterile cabinet. After that, 1-2 µL of an overnight culture was inoculated onto the agar, making sure to scrape the surface. Up to 16 cultures could be inoculated onto the same dish. After inoculation, the dishes were incubated at 37°C for 4-5 hours, and then they were imaged using the gel transilluminator. To measure the diameter of the spreading disk, we used ImageJ software [249].

**Crystal violet biofilm quantification.** Biofilms were quantified using a standard crystal violet (CV) assay on microtiter plates [251], with modifications. Briefly, cultures of *E. coli* were grown overnight in TB in a rotary shaker at 30°C were diluted 1:100 into fresh TB medium and grown at 220 rpm to the mid-exponential phase ($OD_{600}$ = 0.5) at 30°C. The culture was diluted in fresh TB medium to an $OD_{600}$ of 0.05, and 300 µl was loaded into a 96-well plate. The $OD_{600}$ of the planktonic culture was measured after 20 to 24 h of stationary incubation at 30°C, and the liquid culture was then removed from the wells. The wells were washed once with PBS, and the biofilms were then fixed with 300 µl of 96% ethanol. After 20 min, ethanol was removed, and the plates were left to dry completely under a fume hood for 40 min and then stained with 300 µl of a 0.1% CV solution for 15 min. After the incubation, CV was

removed, and biofilms were washed twice with PBS. The remaining CV stain in biofilms was extracted by adding 300 µl of 96% ethanol for 35 min, and the $OD_{595}$ was measured. All the measurements were performed with a Tecan Infinite 200 Pro multimode plate reader.

**Western blot.** The samples (cell suspensions, cell lysates or proteins attached to the beads) were mixed with the Roti-Load 4x Loading dye and incubated at 95°C for 10 minutes. After that they were loaded to the 10 % polyacrylamide gels and the electrophoresis was run using Mini-Protean Tetra system (Bio-Rad) in the SDS-PAGE running buffer at 90 V for 15 minutes, followed by 45-50 minutes at 140V. The proteins were then transferred to polyvinylidene fluoride membranes using a PerfectBlue 'Semi-Dry' Electro Blotter (VWR) according to the manufacturer's instructions. The membranes were blocked in 5% milk in TBS-T for 1 hour at room temperature with gentle shaking, followed by staining with the primary antibodies overnight at 4°C. After several washed with TBS-T, the membranes were stained with the secondary antibodies for 1 hour at room temperature, followed by thorough washing. All antibodies were diluted 1:10000 in 5% milk. In case of anti-biotin western blot, blocking of the membranes was performed in 1% milk for 1 hour [241], and after thorough washing, the membrane was incubated in 0.3 µg/ml streptavidin-dye conjugate (Licor Biosciences) dissolved in 3% bovine serum albumin solution in TBS-T. Stained membranes were imaged using Licor Odyssey infrared imager.

**Whole proteome analysis.** To perform the whole proteome analysis, the cultures were grown overnight in TB medium at 37°C, and then harvested by centrifugation. After that, the cells were washed twice with ice-cold PBS and stored at -80°C before submission to the proteomic facility.

**Protein co-immunoprecipitation.** Translational fusions of MaoP and HdfR with GFP were constructed using Gibson assembly. Both C- and N-terminal fusions of MaoP were created, as well as C-terminal fusion for HdfR. Integrity of the constructs was confirmed with anti-GFP Western blots. The cultures overexpressing these fusions grown to early stationary phase ($OD_{600}$ = 1) in TB medium at 37°C in a rotary shaker. After the cultivation, the cells were washed once with PBS and treated with formaldehyde (1% final concentration) at room temperature for 10 min to achieve crosslinking. After quenching with glycine (0.25 M final concentration), incubation for an additional 30 min on ice and three washes in PBS, the cells were harvested by centrifugation and stored at -80°C until further use. For cell lysis, the cell pellets were resuspended in 10 ml HNN lysis buffer supplemented with proteinase inhibitors.

Cell suspension were then pressurized in a French press, and the resulting lysates were then cleared by centrifugation. The cleared lysates were then incubated with the GFP-trap magnetic agarose beads at 4°C with gentle agitation for 1 hour. After the incubation, we harvested the beads with a magnetic rack, washed them multiple times (twice with HNN lysis buffer and thrice with 0.1 M ammonium bicarbonate solution), and stored at -20°C before submission to the proteomics facility. Solubility of the baits and their recovery on beads were confirmed with anti-GFP Western blots [202, 252].

**BioID proximity labelling.** Translational fusions of MaoP, HdfR and GFP with TurboID biotinylase [203, 241] were constructed using Gibson assembly. Both C- and N-terminal fusions were created. Integrity and functionality of the constructs were confirmed with anti-FLAG and anti-biotin Western blots respectively. The cultures overexpressing these fusions were grown to early stationary phase ($OD_{600}$ = 1) or to late exponential phase ($OD_{600}$ = 0.6) in TB medium supplemented with 50 µM biotin at 37°C in a rotary shaker. Cells were collected via centrifugation and washed three times with ice-cold PBS, and then stored at -80°C. Cell pellets were resuspended in 2 ml of RIPA buffer and incubated for 30 minutes at 37°C to complete the lysis. Cell lysates were briefly sonicated and then desalted using PD-10 desalting columns according to the manufacturer's instructions (we have chosen the gravity protocol) to remove free biotin from the solution. Desalted cell lysates were incubated with streptavidin-coated magnetic beads (Dynabeads) for 1 hour at 4°C with gentle rotation. The beads were then collected using a magnetic rack, washed multiple times (twice with RIPA buffer, twice with 1M NaCl and thrice with 50 mM Tris-HCl pH 7) and stored at -20°C before submission to the proteomics facility. Solubility of the baits and their recovery on beads were confirmed with anti-FLAG and anti-biotin Western blots.

**ChIP-Seq.** The same set of GFP fusions (HdfR-sfGFP, MaoP-sfGFP, sfGFP-MaoP) was used as for the co-immunoprecipitation. The sfGFP was used as the control mock IP. The protocol was taken from [253] with minor modifications. The cultures overexpressing these fusions were grown to early stationary phase ($OD_{600}$ = 1) in TB medium at 37°C in a rotary shaker. After the cultivation, the cells were treated with formaldehyde (1% final concentration) at room temperature for 10 min to achieve crosslinking. After quenching with glycine (0.5 M final concentration), incubation for an additional 30 min on ice and three washes in PBS, the cells were harvested by centrifugation and stored at -80°C until further use. For cell lysis, the pellets were resuspended in 1 ml ChIP lysis buffer supplemented with

proteinase inhibitors and incubated for 30 minutes at 37°C. The lysates were then sonicated (Bioruptor Standard, Diagenode) in 1.5 ml Ependorf tubes at 4°C using 15 bursts of 30 s to shear DNA fragments to an average length of approximately 350 bp and cleared by centrifugation. The cleared lysates were then incubated with the binding control magnetic agarose beads at 4°C with gentle agitation for 1 hour to reduce non-specific binding. The beads were then collected with a magnetic stand and discarded, and the lysates were incubated with GFP-Trap magnetic agarose beads overnight. The harvested beads were then washed multiple times (twice with ChIP lysis buffer, once with ChIP lysis buffer 500 mM NaCl, once with ChIP wash buffer and twice with 10 mM Tris/HCl pH 8.1, 1 mM EDTA). After the washing, the beads were resuspended in 300 µl ChIP elution buffer and incubated overnight at 65°C to reverse the crosslinking, followed by 2 hours on-beads digest with 20 µg of proteinase K at 45°C. DNA was then extracted using phenol:chloroform:isoamyl alcohol mixture (25:24:1), ethanol-precipitated using 40 mg of glycogen as a carrier and resuspended in 30 mL of nuclease-free water [253, 254]. The libraries for Illumina sequencing were prepared using NEBNext II Ultra kit according to the manufacturer's instructions. Paired-end runs with 110 cycles were performed on a NovaSeq (Illumina) next-generation DNA sequencing instrument, yielding several million reads per sequenced samples.

**ChIP-Seq analysis pipeline.** The quality of the reads was confirmed by the *FastaQC* software. The paired reads were then aligned to the reference genome using *bowtie2* [255], and the alignment with the quality lower than 20 were filtered out using *samtools* toolkit [256]. Then, the peaks were called using *MACS2* software [257], with the mock IP as a control. To view the alignments and the peaks, IGV browser was used [258]. To annotate the datasets and select the genomic intervals, the *bedtools* toolkit was used [259].

**Data analysis.** Data was collected from different files and then analyzed using the scripts written on R version 4.2 [260]. External packages such as *ggplot2* [261] with its extensions (*ggpubr*, *ggforce*, *ggarrange* etc.), *tidyverse* [262], *data.table, plater* [263] and some others were used. To count the abundance of barcodes in the TnSeq data we used a custom script written on Python version 3.9.12 with Biopython version 1.78 [244], the set of Python tools for computational molecular biology. Later analysis was performed in R. To assess the significance of the differences between the groups, the two-sample Student's t-test was routinely employed.

**Principal component analysis (PCA)**. The main approach used to analyze high-dimentional data (i.e. whole-genome screens, proteomics etc) was principal component analysis. We used the standard PCA implemented in R with centering and scaling the data prior to the analysis. Briefly, the data from the original dataset is used to calculate new uncorrelated variables (principal components, PCs). To achieve this, the covariance matrix ($\begin{pmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \cdots & c_{n,n} \end{pmatrix}$) for all the variables in the dataset is calculated. Then eigenvectors and their respective eigenvalues ($M * \vec{v} = \lambda\vec{v}$) are found for this matrix. The number of eigenvectors is always equal to the number of the variables in the initial dataset. The resulting eigenvectors are then sorted descendingly by their eigenvalues, and the transformation matrix is formed $\begin{pmatrix} eigenvector_1 \\ \vdots \\ eigenvector_n \end{pmatrix}$. This matrix is then multiplied by the transposed initital dataset, resulting in change of basis. As the result, the new dataset is obtained, where eigenvectors are axes and these axes are sorted by the variance explained in the decreasing order. This approach allows to reduce the number of parameters while minimizing the loss of information [163].

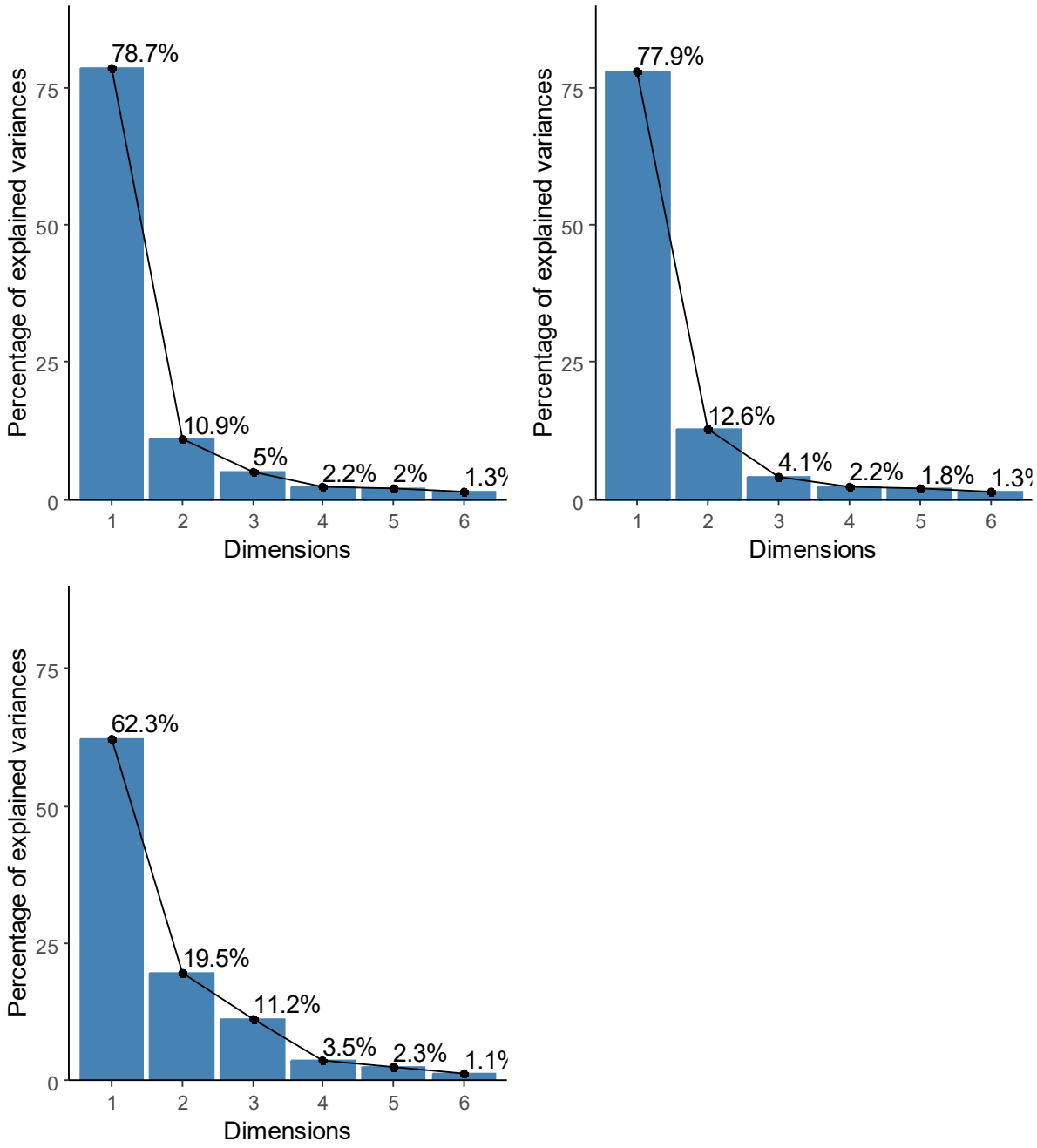
# Appendix

## 6.1. Supplementary data for the Chapter II

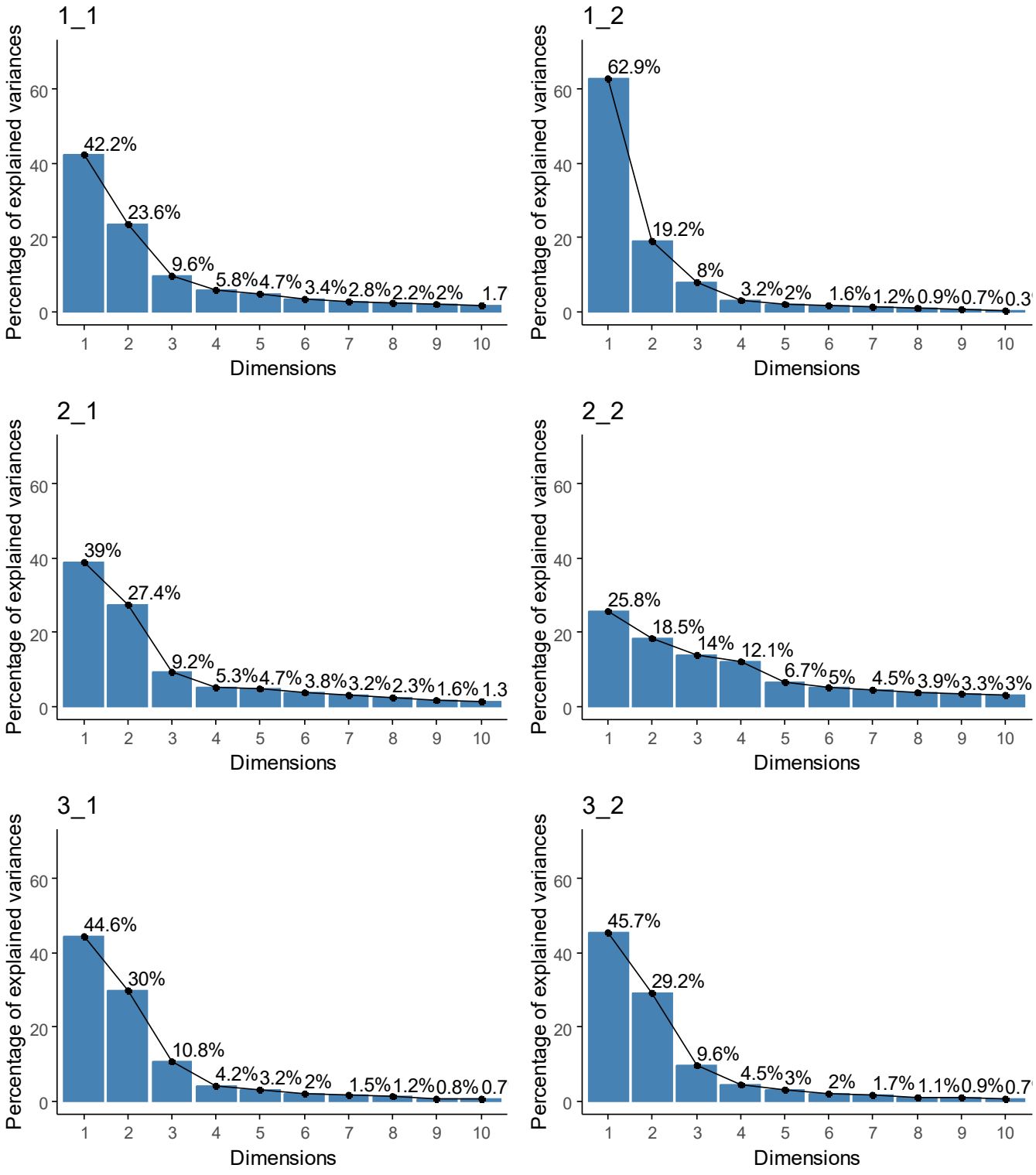
**Supplementary table 1. Lake water parameters.**

| Date of sampling | Purpose of sampling | pH | Air temperature, °C | Water temperature, °C |
|---|---|---|---|---|
| 24.05.2019 | Keio, plates 1-29 | 6.7 | 21 | 13 |
| 8.06.2019 | Keio, plates 1-29 | 6.4 | 26 | 20 |
| 19.06.2019 | Keio, plates 31-59 | 6.7 | 21 | 17 |
| 6.07.2019 | Keio, plates 61-89 | 6.7 | 22 | 18 |
| 1.09.2019 | Keio, plates 31-59 | 6.6 | 16 | 18 |
| 15.09.2019 | Keio, plates 61-89 | 6.5 | 28 | 12.5 |
| 2.04.2021 | PCA cutoffs | 6.5 | 8 | 8.8 |
| 2.06.2021 | PCA cutoffs | 7 | 19 | 14.2 |
| 22.07.2021 | PCA cutoffs | 7 | 22 | 17.4 |
| 16.08.2021 | PCA cutoffs + samples 1 and 2 | 7 | 19 | 16.3 |
| 15.09.2021 | Samples 3 and 4 | 7 | 15 | 12.1 |
| 29.09.2021 | PCA cutoffs | 7.5 | 12 | 11.9 |
| 9.05.2022 | Sample 5 | 7 | 13 | 12.4 |
| 19.06.2022 | Sample 6+TnSeq | 6.5 | 29 | 18.4 |
| 6.07.2022 | TnSeq | 6.5 | 17 | 16.1 |
| 4.08.2022 | Sample 7+TnSeq | 7 | 26 | 18,1 |

Water salinity is 0 ppt in all the samples. Samples labelled as “PCA cutoffs” were used for a confirmatory screening of the strains selected from the whole-genome screen with the PCA analysis. Samples 1-7 were used for the in-depth screen of Δ*hdfR* and Δ*rpoS* strains.




**Supplementary figure 1. Scree plots for the TnSeq results.** Shown is the percentage of the variance explained by each PC.




**Supplementary figure 2. Scree plots for the FB screen.** Shown is the percentage of the variance explained by each PC.

**Supplementary table 2. Gene fitness with of *E. coli* W3110 RpoS+ transposon mutants in filtered and non-filtered lake water**

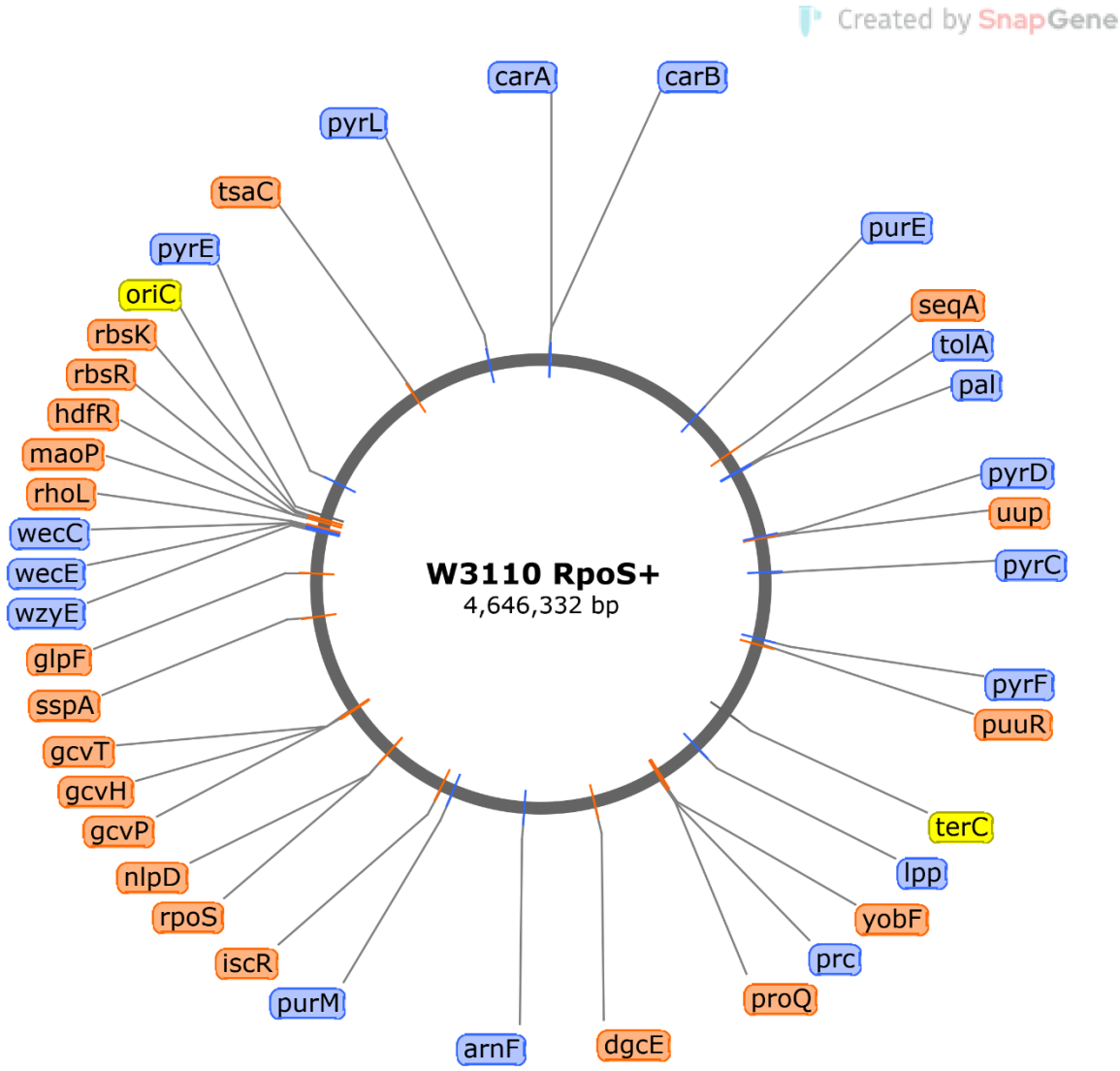
| Gene name | Filtered water | | | Non-filtered water | | | PC1 | PC2 | Euclidean distance | Protein description |
|---|---|---|---|---|---|---|---|---|---|---|
| | Day 2 | Day 4 | Day 8 | Day 2 | Day 4 | Day 8 | | | | |
| *arnF* | -2.2 | -1.6 | -1.9 | -1.9 | -1.9 | -1.8 | 15.9 | 2.0 | 16.0 | Probable 4-amino-4-deoxy-L-arabinose-phosphoundecaprenol flippase subunit ArnF |
| | -1.9 | -1.9 | -2.1 | -1.9 | -1.7 | -1.7 | 15.7 | 1.1 | 15.7 | |
| | -1.3 | -1.6 | -1.5 | -0.8 | -1.3 | -0.8 | 10.2 | 0.7 | 10.2 | |
| *carA* | -1.4 | -1.8 | -1.5 | -1.5 | -1.9 | -1.6 | 13.5 | 1.9 | 13.6 | Carbamoyl-phosphate synthase small chain |
| | -1.5 | -1.6 | -1.8 | -1.5 | -1.8 | -1.8 | 13.6 | 1.8 | 13.8 | |
| | -1.4 | -1.6 | -1.3 | -1.2 | -1.8 | -1.4 | 12.0 | 3.4 | 12.4 | |
| *carB* | -1.3 | -1.5 | -1.5 | -1.4 | -1.9 | -1.5 | 12.7 | 2.2 | 12.9 | Carbamoyl-phosphate synthase large chain |
| | -1.5 | -1.7 | -1.8 | -1.5 | -1.8 | -1.6 | 13.6 | 1.5 | 13.7 | |
| | -1.4 | -1.5 | -1.3 | -1.2 | -1.7 | -1.0 | 11.2 | 2.6 | 11.5 | |
| *dgcE* | 0.4 | 0.7 | 0.8 | 0.2 | -0.1 | -0.6 | -2.1 | 3.5 | 4.1 | Diguanylate cyclase |
| | 0.4 | 0.9 | 1.3 | 0.0 | -0.3 | -0.7 | -2.0 | 5.0 | 5.4 | |
| | 1.0 | 1.3 | 1.4 | -0.2 | -0.7 | -0.6 | -3.7 | 5.9 | 7.0 | |
| *gcvH* | 1.3 | 2.4 | 2.7 | 0.4 | -0.1 | 0.1 | -8.9 | 6.7 | 11.2 | Glycine cleavage system H protein |
| | 1.4 | 2.6 | 2.7 | 0.1 | 0.0 | 0.3 | -9.0 | 7.5 | 11.7 | |
| | 2.2 | 2.3 | 2.3 | 0.1 | -0.6 | -0.5 | -8.9 | 8.0 | 12.0 | |
| *gcvP* | 1.3 | 2.7 | 2.7 | 0.6 | 0.1 | 0.2 | -10.1 | 6.7 | 12.1 | Glycine cleavage system P-protein |
| | 1.4 | 2.6 | 2.8 | 0.2 | 0.1 | 0.2 | -9.2 | 7.3 | 11.8 | |
| | 2.1 | 2.3 | 2.3 | 0.1 | -0.1 | 0.1 | -9.8 | 6.1 | 11.5 | |
| *gcvT* | 1.2 | 2.8 | 2.6 | 0.6 | 0.1 | 0.3 | -10.2 | 6.3 | 12.0 | Glycine cleavage system T protein |
| | 1.4 | 2.6 | 2.6 | 0.3 | 0.2 | 0.1 | -9.3 | 6.9 | 11.6 | |
| | 2.1 | 2.3 | 2.2 | 0.2 | -0.1 | 0.0 | -9.7 | 6.0 | 11.4 | |
| *glpF* | 0.3 | 0.5 | 0.3 | 0.3 | 0.3 | 1.3 | -4.2 | -2.4 | 4.8 | Glycerol uptake facilitator protein |
| | 0.3 | 0.5 | 0.4 | 0.5 | 0.8 | 1.5 | -5.7 | -2.9 | 6.4 | |
| | 0.4 | 0.3 | 0.4 | 0.2 | 1.0 | 1.7 | -4.6 | -4.2 | 6.3 | |
| *hdfR* | 0.3 | 2.0 | 2.5 | 0.3 | -0.5 | -0.6 | -4.9 | 8.2 | 9.6 | HTH-type transcriptional regulator HdfR |
| | 0.3 | 2.0 | 2.6 | -0.2 | -0.8 | -0.9 | -2.9 | 9.3 | 9.8 | |
| | 1.5 | 2.5 | 2.8 | -0.5 | -1.2 | -0.4 | -6.7 | 9.8 | 11.8 | |
| *iscR* | 0.7 | 1.1 | 0.9 | 0.2 | 0.6 | 0.7 | -6.1 | 0.1 | 6.1 | HTH-type transcriptional regulator IscR |
| | 0.6 | 1.1 | 1.3 | 0.6 | 0.8 | 1.0 | -7.6 | 0.1 | 7.6 | |
| | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 0.7 | -5.9 | -2.4 | 6.3 | |
| *lpp* | -3.2 | -3.6 | -3.9 | -2.8 | -3.6 | -3.2 | 28.4 | 2.5 | 28.5 | Major outer |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | -2.8 | -3.8 | -4.0 | -3.7 | -3.8 | -3.3 | 29.7 | 3.1 | 29.8 | membrane lipoprotein Lpp |
| | -2.6 | -3.2 | -3.1 | -2.7 | -2.7 | -1.4 | 22.4 | 2.5 | 22.6 | |
| *maoP* | 0.3 | 2.1 | 2.8 | 0.4 | -0.3 | -0.9 | -5.5 | 9.0 | 10.5 | Macrodomain Ori protein |
| | 0.3 | 2.2 | 2.9 | -0.2 | -0.6 | -0.7 | -4.2 | 9.3 | 10.2 | |
| | 1.5 | 2.6 | 3.1 | -0.6 | -1.4 | -0.5 | -6.5 | 10.8 | 12.6 | |
| *nlpD* | 0.1 | 1.0 | 1.8 | 0.4 | 0.1 | -0.1 | -4.3 | 3.9 | 5.9 | Murein hydrolase activator NlpD |
| | 0.1 | 0.8 | 1.6 | 0.0 | 0.1 | -0.2 | -2.9 | 3.6 | 4.6 | |
| | 0.7 | 1.2 | 2.9 | 0.2 | 0.1 | 0.5 | -7.3 | 3.1 | 8.0 | |
| *pal* | -3.0 | -3.2 | -3.3 | -2.8 | -3.4 | -3.1 | 26.4 | 3.3 | 26.6 | Peptidoglycan-associated lipoprotein |
| | -2.6 | -3.8 | -4.1 | -3.7 | -3.8 | -3.5 | 29.9 | 3.5 | 30.1 | |
| | -2.7 | -3.1 | -3.7 | -2.7 | -3.6 | -2.0 | 24.9 | 4.8 | 25.4 | |
| *prc* | -1.8 | -1.8 | -2.0 | -1.5 | -1.9 | -1.5 | 14.5 | 1.1 | 14.6 | Tail-specific protease |
| | -1.7 | -2.0 | -1.9 | -1.9 | -1.9 | -1.6 | 15.4 | 1.5 | 15.5 | |
| | -1.7 | -1.7 | -1.7 | -1.6 | -1.7 | -0.6 | 13.1 | 1.7 | 13.2 | |
| *proQ* | 0.9 | 1.5 | 1.2 | 0.4 | -0.1 | 0.4 | -5.9 | 2.8 | 6.5 | RNA chaperone ProQ |
| | 1.3 | 1.7 | 1.6 | 0.1 | -0.2 | -0.1 | -5.9 | 5.5 | 8.0 | |
| | 1.6 | 1.8 | 1.5 | -0.2 | -0.3 | -0.1 | -6.1 | 5.5 | 8.2 | |
| *purE* | -1.5 | -1.7 | -1.9 | -1.3 | -1.1 | -1.1 | 11.8 | -0.8 | 11.8 | N5-carboxyaminoimidazole ribonucleotide mutase |
| | -1.7 | -1.8 | -1.7 | -1.3 | -1.0 | -1.1 | 11.6 | -1.0 | 11.6 | |
| | -1.3 | -1.8 | -1.4 | -0.9 | -1.0 | -0.6 | 9.9 | -0.1 | 9.9 | |
| *purM* | -1.5 | -1.6 | -1.7 | -1.3 | -1.0 | -1.3 | 11.7 | -0.2 | 11.7 | Phosphoribosylformylglycinamidine cyclo-ligase |
| | -1.6 | -2.0 | -2.0 | -1.2 | -1.1 | -1.2 | 12.1 | -1.4 | 12.2 | |
| | -1.5 | -1.6 | -1.4 | -0.9 | -1.0 | -0.5 | 9.7 | -0.3 | 9.7 | |
| *puuR* | 0.3 | 1.3 | 1.6 | 0.4 | 0.1 | -0.2 | -4.9 | 4.0 | 6.3 | HTH-type transcriptional regulator PuuR |
| | 0.5 | 1.4 | 1.7 | -0.1 | -0.2 | -0.2 | -4.0 | 5.2 | 6.6 | |
| | 1.2 | 1.5 | 1.5 | 0.2 | 0.0 | 0.0 | -6.4 | 3.6 | 7.4 | |
| *pyrC* | -1.2 | -1.3 | -1.3 | -1.1 | -1.5 | -1.4 | 10.7 | 1.9 | 10.9 | Dihydroorotase |
| | -1.4 | -1.6 | -1.7 | -1.3 | -1.8 | -1.5 | 12.7 | 1.4 | 12.8 | |
| | -1.4 | -1.5 | -1.3 | -1.1 | -1.3 | -1.0 | 10.4 | 1.7 | 10.6 | |
| *pyrD* | -1.4 | -1.6 | -1.5 | -1.2 | -1.6 | -1.7 | 12.7 | 2.1 | 12.9 | Dihydroorotate dehydrogenase |
| | -1.5 | -1.9 | -1.9 | -1.4 | -1.8 | -1.8 | 13.8 | 1.2 | 13.8 | |
| | -1.4 | -1.6 | -1.4 | -1.2 | -1.6 | -1.1 | 11.7 | 2.4 | 11.9 | |
| *pyrE* | -1.5 | -1.6 | -1.5 | -1.4 | -1.9 | -1.6 | 13.3 | 2.2 | 13.5 | Orotate phosphoribosyltransferase |
| | -1.5 | -1.8 | -1.7 | -1.6 | -1.7 | -1.6 | 13.6 | 1.3 | 13.6 | |
| | -1.2 | -1.4 | -1.2 | -1.5 | -1.4 | -0.8 | 10.9 | 2.2 | 11.1 | |
| *pyrF* | -1.1 | -1.2 | -1.2 | -0.9 | -1.2 | -1.2 | 9.3 | 1.2 | 9.4 | Orotidine 5'-phosphate |
| | -1.3 | -1.5 | -1.4 | -1.1 | -1.2 | -1.0 | 10.2 | 0.2 | 10.2 | |

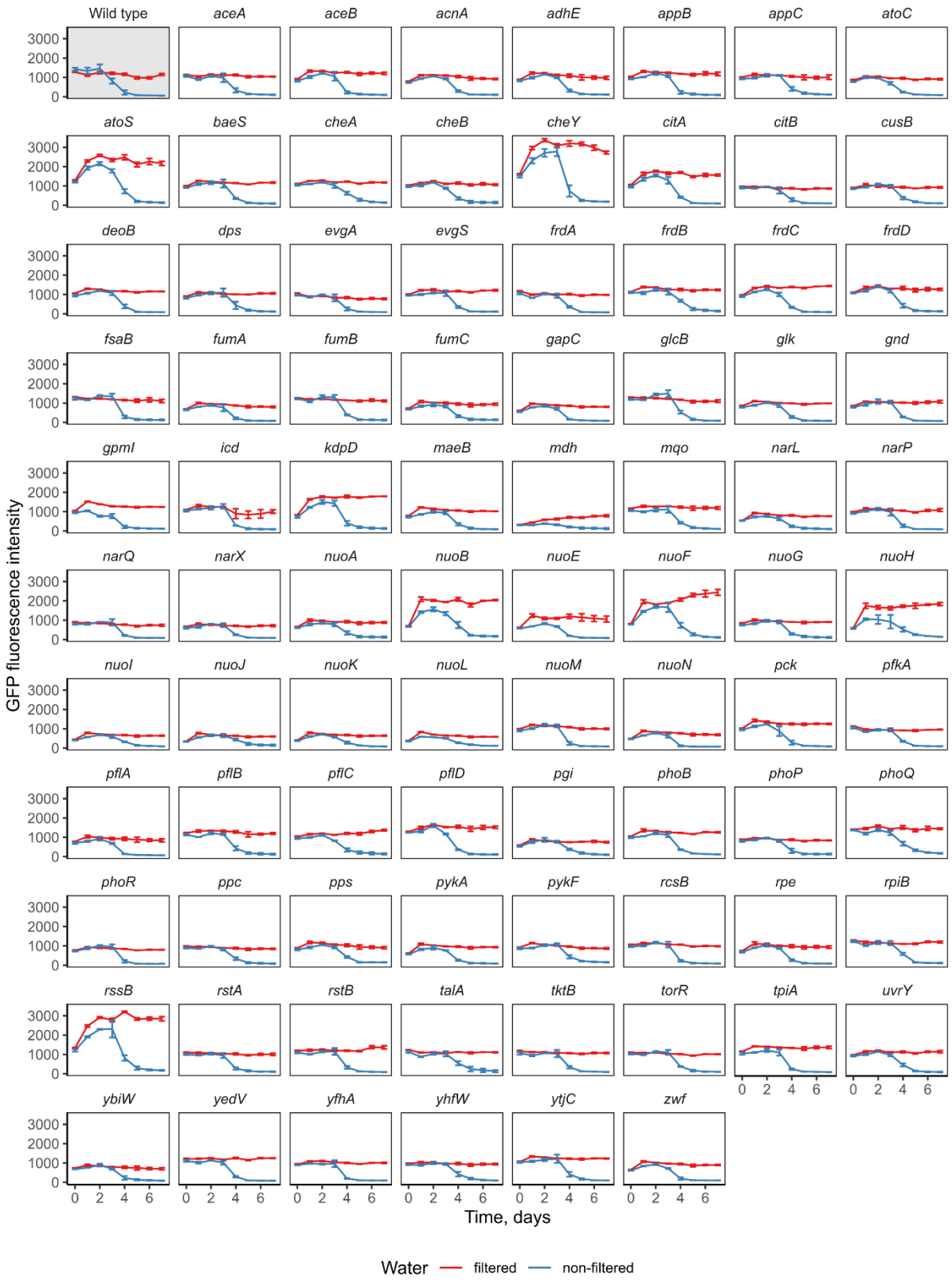
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | -1.2 | -1.2 | -1.0 | -0.9 | -1.4 | -1.5 | 9.3 | 3.3 | 9.9 | decarboxylase |
| *pyrL* | -1.3 | -2.0 | -1.7 | -1.5 | -1.5 | -1.8 | 13.5 | 1.4 | 13.6 | *pyr* operon leader peptide |
| | -1.6 | -2.2 | -2.0 | -1.5 | -2.6 | -2.4 | 16.7 | 2.9 | 17.0 | |
| | -2.0 | -1.6 | -2.1 | -1.4 | -2.5 | -1.5 | 15.4 | 3.7 | 15.8 | |
| *rbsK* | 0.2 | 1.7 | 4.4 | 0.3 | -0.2 | -0.3 | -7.4 | 9.4 | 12.0 | Ribokinase |
| | 0.5 | 2.2 | 4.4 | -0.1 | -0.3 | -0.1 | -7.4 | 10.1 | 12.5 | |
| | 1.6 | 3.2 | 6.5 | 0.0 | -0.4 | 0.8 | -14.9 | 9.7 | 17.8 | |
| *rbsR* | 0.3 | 2.0 | 5.1 | 0.3 | -0.3 | 0.0 | -8.8 | 10.1 | 13.4 | Ribose operon repressor |
| | 0.7 | 2.8 | 4.8 | 0.0 | -0.1 | 0.0 | -9.5 | 10.8 | 14.4 | |
| | 1.5 | 3.4 | 6.7 | -0.2 | -0.5 | 0.7 | -14.8 | 10.8 | 18.3 | |
| *rhoL* | 3.0 | 7.0 | 7.0 | 2.2 | 0.7 | 1.0 | -27.7 | 15.4 | 31.7 | *rho* operon leader peptide |
| | 3.2 | 7.4 | 6.4 | 1.0 | 1.2 | 1.6 | -26.6 | 14.8 | 30.4 | |
| | 4.2 | 5.1 | 5.1 | 0.7 | 1.2 | 1.4 | -24.3 | 7.6 | 25.5 | |
| *rpoS* | 0.1 | 2.9 | 6.1 | 0.7 | -0.1 | -0.3 | -11.7 | 13.2 | 17.6 | RNA polymerase sigma factor RpoS |
| | 0.2 | 2.5 | 4.7 | 0.0 | -0.2 | 0.1 | -8.3 | 10.1 | 13.1 | |
| | 1.6 | 3.7 | 7.8 | 0.1 | -0.5 | 1.5 | -17.8 | 10.0 | 20.4 | |
| *seqA* | 0.8 | 1.1 | 0.9 | 0.6 | 0.2 | 0.3 | -5.7 | 1.8 | 5.9 | Negative modulator of initiation of replication |
| | 1.2 | 1.7 | 1.0 | 0.5 | 0.3 | 0.8 | -7.7 | 2.0 | 8.0 | |
| | 1.4 | 1.5 | 1.2 | 0.2 | 0.2 | 0.8 | -7.3 | 1.5 | 7.4 | |
| *sspA* | 0.3 | 1.4 | 1.8 | 0.1 | -0.8 | -0.6 | -2.6 | 6.9 | 7.3 | Stringent starvation protein A |
| | 0.2 | 2.2 | 2.2 | -0.7 | -1.0 | -1.2 | -1.1 | 10.5 | 10.5 | |
| | 1.0 | 1.7 | 1.5 | -0.7 | -1.3 | -0.2 | -2.5 | 7.5 | 7.9 | |
| *tolA* | -2.4 | -2.1 | -2.7 | -2.1 | -2.3 | -2.0 | 19.2 | 1.3 | 19.3 | Tol-Pal system protein TolA |
| | -2.0 | -2.2 | -2.5 | -2.2 | -2.5 | -1.9 | 18.5 | 1.7 | 18.5 | |
| | -2.4 | -2.5 | -2.5 | -2.1 | -1.8 | -0.9 | 17.7 | 0.8 | 17.7 | |
| *tsaC* | 0.6 | 0.7 | 0.4 | 0.2 | 0.4 | 0.5 | -4.0 | -0.2 | 4.0 | Threonylcarbamoyl-AMP synthase |
| | 0.7 | 0.7 | 0.6 | 0.4 | 0.7 | 0.6 | -5.2 | -0.4 | 5.2 | |
| | 0.6 | 0.3 | 0.7 | 0.5 | 0.7 | 0.5 | -4.8 | -1.5 | 5.0 | |
| *uup* | 0.5 | 0.9 | 0.9 | 0.3 | 0.2 | 0.0 | -4.0 | 2.1 | 4.5 | ATP-binding protein Uup |
| | 0.7 | 1.4 | 1.1 | 0.3 | 0.2 | 0.1 | -5.2 | 3.0 | 6.0 | |
| | 0.9 | 1.1 | 0.9 | 0.0 | -0.4 | 0.2 | -3.8 | 3.0 | 4.8 | |
| *wecA* | -1.2 | -1.3 | -1.2 | -1.0 | -1.2 | -1.2 | 9.8 | 1.1 | 9.8 | Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase |
| | -1.0 | -1.4 | -1.4 | -1.2 | -1.4 | -1.4 | 10.6 | 1.4 | 10.7 | |
| | -1.1 | -1.3 | -1.2 | -0.8 | -1.0 | -0.5 | 8.2 | 0.5 | 8.2 | |
| *wecB* | -1.5 | -1.5 | -1.7 | -1.2 | -1.5 | -1.6 | 12.2 | 1.4 | 12.3 | UDP-N-acetylglucosamine 2-epimerase |
| | -1.2 | -1.5 | -1.8 | -1.4 | -1.6 | -1.8 | 12.6 | 1.9 | 12.8 | |
| | -1.2 | -1.4 | -1.4 | -0.9 | -1.3 | -1.2 | 9.7 | 2.1 | 9.9 | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| *wecC* | -1.4 | -1.6 | -1.6 | -1.1 | -1.4 | -1.6 | 12.0 | 1.4 | 12.1 | UDP-N-acetyl-D-mannosamine dehydrogenase |
| | -1.4 | -1.5 | -1.8 | -1.5 | -1.7 | -1.8 | 13.3 | 1.9 | 13.4 | |
| | -1.1 | -1.4 | -1.3 | -0.7 | -1.3 | -0.9 | 9.0 | 1.3 | 9.1 | |
| *wecE* | -1.5 | -1.6 | -1.5 | -1.0 | -1.5 | -1.6 | 11.9 | 1.6 | 12.0 | dTDP-4-dehydro-6-deoxy-D-glucose transaminase |
| | -1.4 | -1.7 | -1.9 | -1.6 | -1.9 | -1.7 | 14.0 | 1.8 | 14.1 | |
| | -1.2 | -1.4 | -1.3 | -0.6 | -1.4 | -0.9 | 9.1 | 1.4 | 9.2 | |
| *wecF* | -1.4 | -1.9 | -1.7 | -1.1 | -1.3 | -1.8 | 12.6 | 1.2 | 12.7 | TDP-N-acetylfucosamine:lipid II N-acetylfucosaminyltransferase |
| | -1.5 | -1.9 | -1.8 | -1.6 | -1.6 | -1.7 | 13.8 | 1.1 | 13.9 | |
| | -1.2 | -1.3 | -1.4 | -0.6 | -1.1 | -0.8 | 8.6 | 0.6 | 8.7 | |
| *wzyE* | -1.3 | -1.4 | -1.4 | -0.9 | -1.2 | -1.6 | 10.6 | 1.6 | 10.8 | Enterobacterial common antigen polymerase |
| | -1.1 | -1.3 | -1.5 | -0.9 | -1.5 | -1.5 | 10.7 | 1.3 | 10.8 | |
| | -1.0 | -1.0 | -1.4 | -0.7 | -0.9 | -1.6 | 8.2 | 2.2 | 8.5 | |
| *yobF* | 0.7 | 1.7 | 1.7 | 0.5 | 0.1 | 0.2 | -6.5 | 3.9 | 7.6 | Uncharacterized small protein YobF |
| | 1.0 | 2.6 | 2.6 | 0.2 | 0.0 | 0.0 | -7.8 | 7.3 | 10.7 | |
| | 1.5 | 2.1 | 1.9 | -0.2 | -0.4 | -0.4 | -6.8 | 6.8 | 9.6 | |

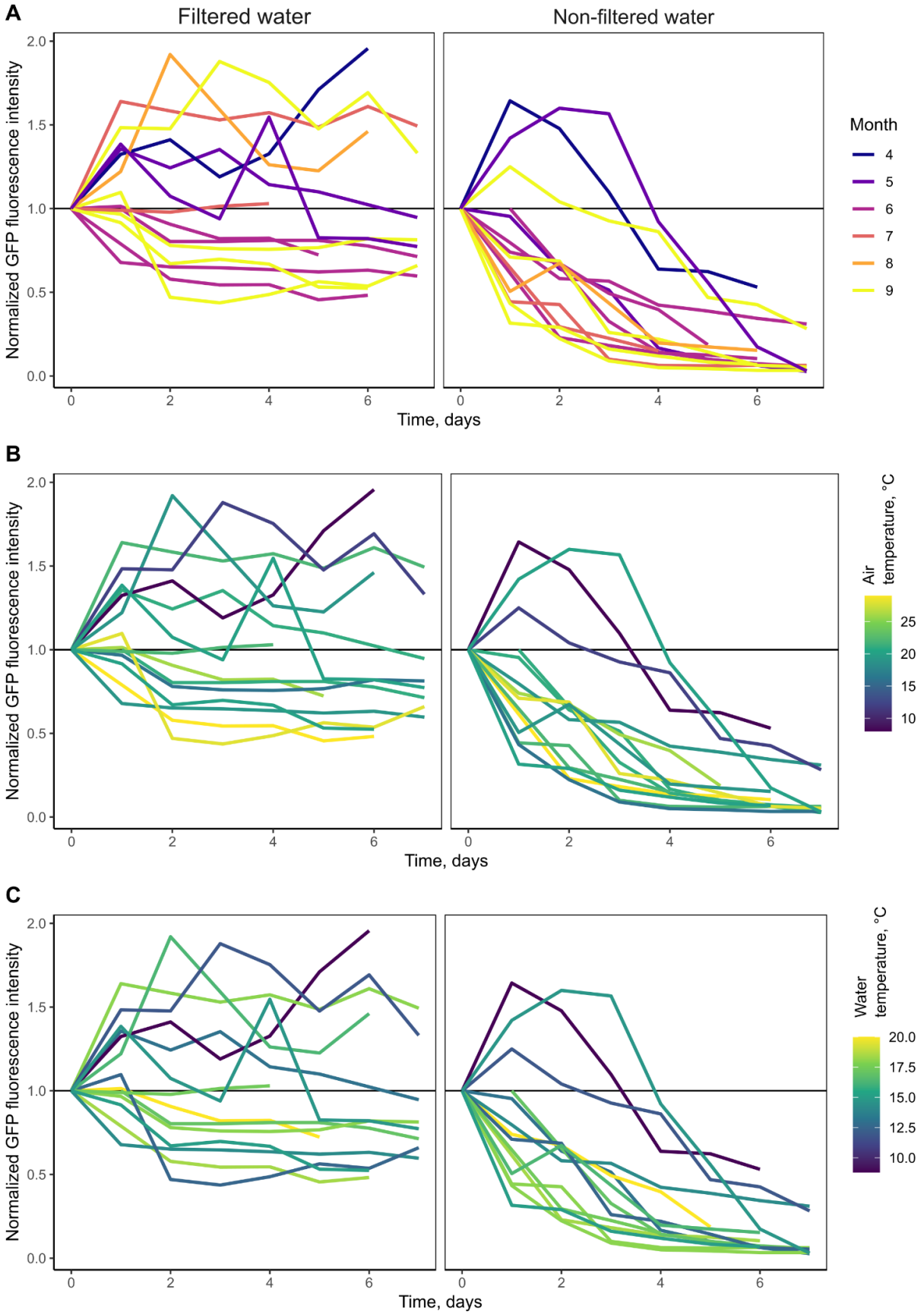
Only the genes that reproducibly demonstrate the absolute fitness value above 1.5 are shown. Three independent biological replicates are presented. This table is an expanded version of the **Table 1** in the main text.



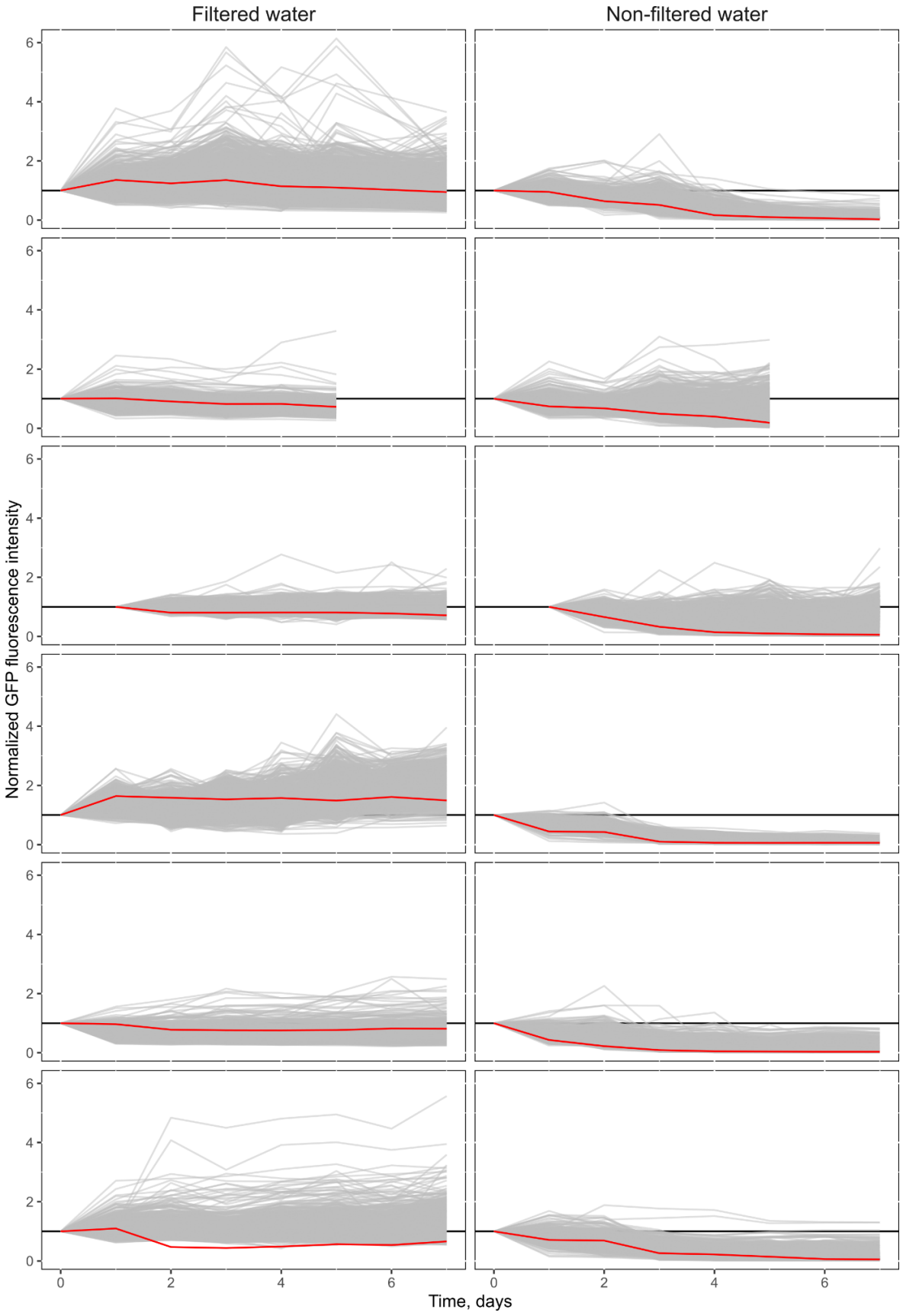

**Supplementary figure 3. Over- (orange) and under- (blue) represented mutations in the TnSeq experiment plotted on the chromosome map.** *Ori* and *ter* sites are shown in yellow.



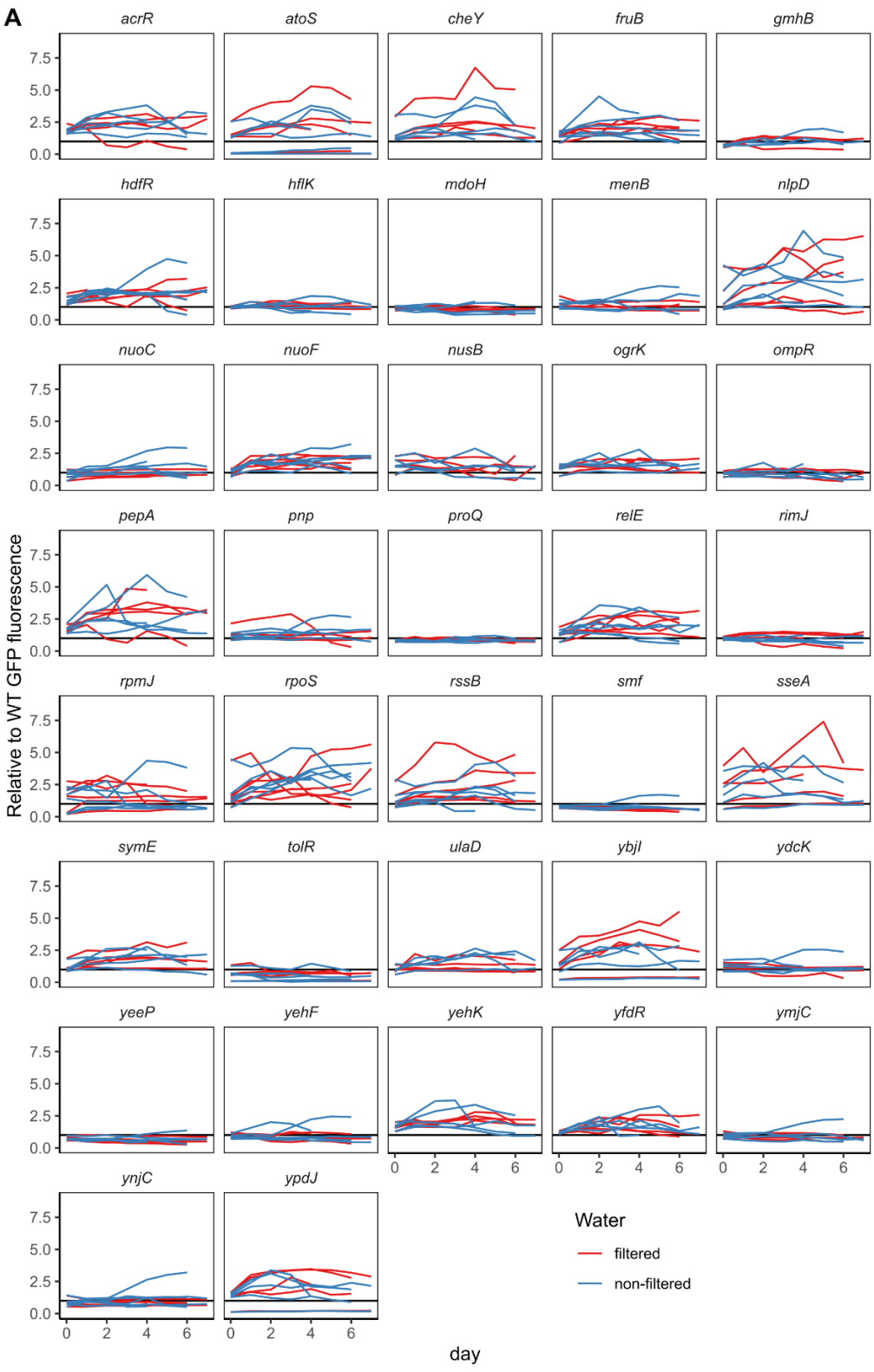

**Supplementary figure 4. Growth of the strains in the Keio collection plate Nr 3 in the lake water.** This is the expanded version of the **Figure 13** in the main text.



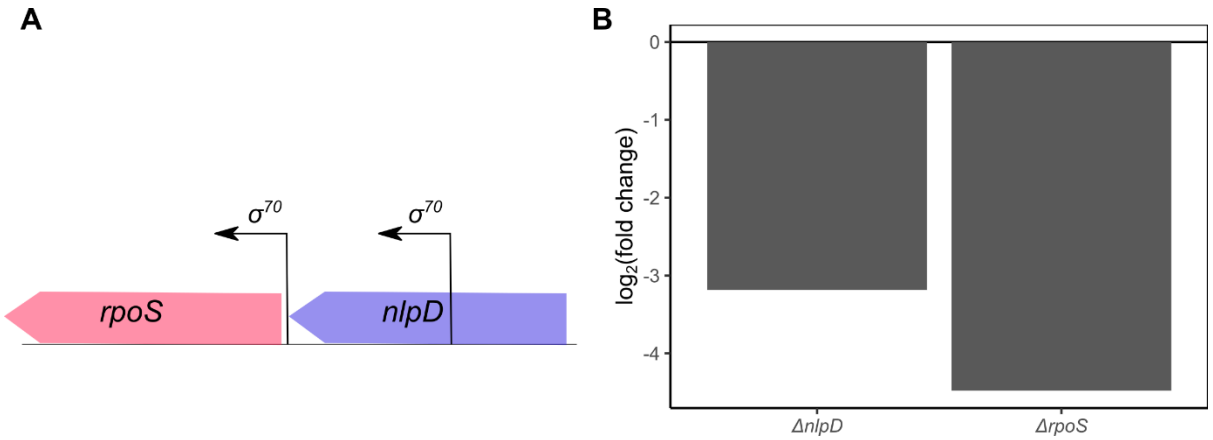

**Supplementary figure 5. Effect of lake water parameters on Keio WT (BW25113) growth.** This is the expanded version of **Figure 14** in the main text.




**Supplementary figure 6. Growth of the Keio collection in the 6 different samples of lake water. The WT strain is highlighed in red.**

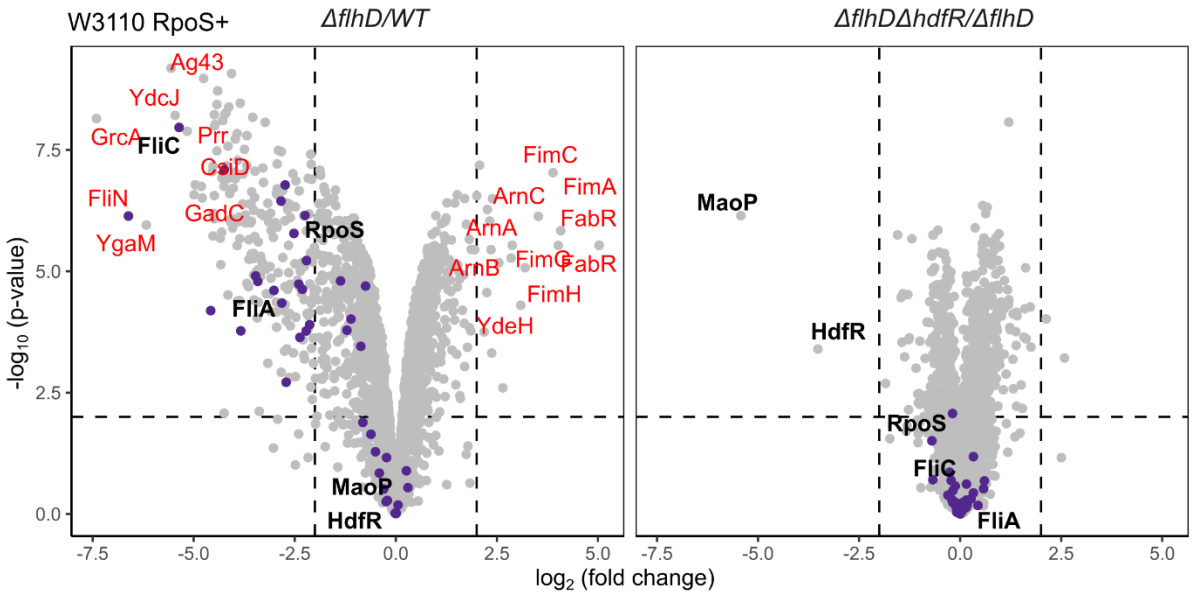



**Supplementary figure 7. Growth of the strains selected by PCA in the Keio collection background**. The reporter used is GFP fluorescence. The data is normalized to the respective point of the WT. Red color indicates filter-sterilized water, while blue color indicates non-filtered water. Different lines represent repeats of the experiment, where each line is an average of 3 technical replicates and 2 biological replicates incubated in the same sample of lake water.




**Supplementary figure 8. Δ*nlpD* does not produce RpoS.** (A) Genomic context of *nlpD*. (B) fold changes of RpoS in Δ*nlpD* and Δ*rpoS* knockout strains

## 6.2. Supplementary data for the Chapter III




**Supplementary figure 9. Global changes in proteome in Δ*flhD* and Δ*hdfR*Δ*flhD* strains.** Protein changes were analyzed using mass spectrometry. Proteins of *flhDC* regulon are highlighted in purple.

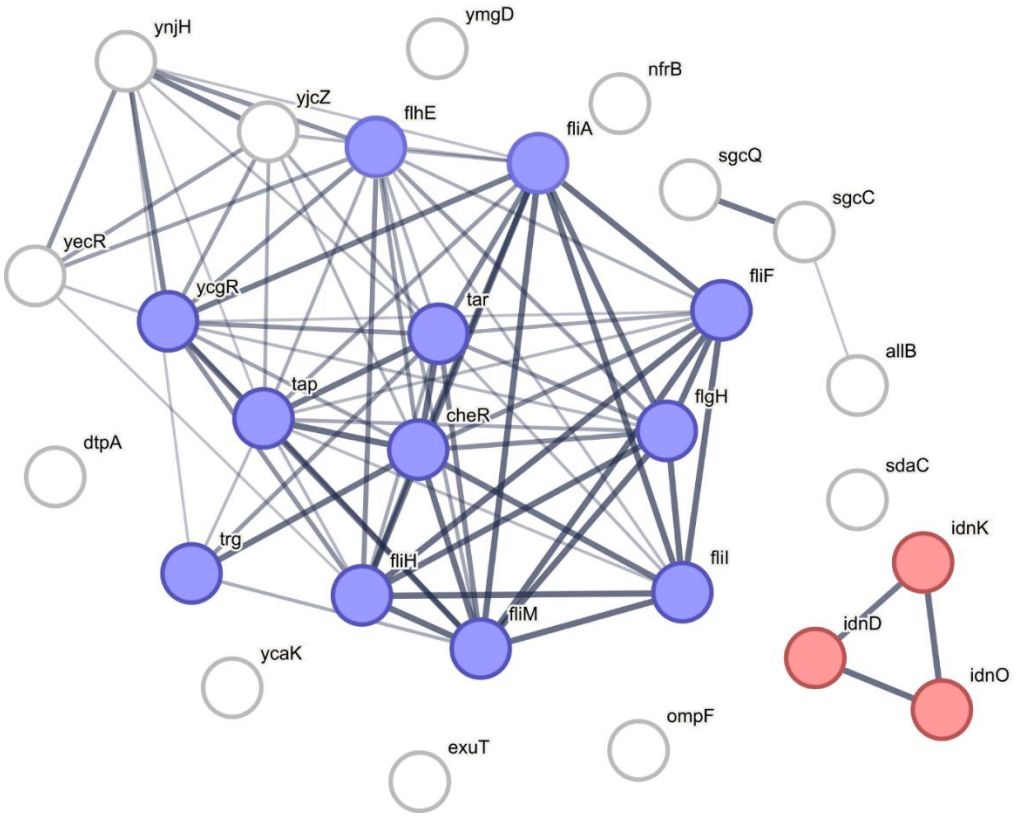
**Supplementary table 3. The most up- and downregulated proteins in Δ*hdfR* and Δ*maoP* strains.**

| comparison | Protein name | Fold change | -log$_{10}$(p-value) | Protein description |
|---|---|---|---|---|
| Δ*hdfR* to WT | MaoP | -4.96 | 5.71 | Macrodomain Ori Protein |
| | HdfR | -3.20 | 6.34 | HTH-type transcriptional regulator |
| | Bfd | -1.70 | 2.38 | Bacterioferritin-associated ferredoxin |
| | YecR | 1.51 | 4.17 | Uncharacterized protein |
| | IdnO | 1.54 | 6.26 | Gluconate 5-dehydrogenase |
| | FliA | 1.57 | 4.05 | RNA polymerase sigma factor 28 |
| | Tap | 1.58 | 4.43 | Methyl-accepting chemotaxis protein IV |
| | FliM | 1.59 | 4.82 | Flagellar motor switch protein |
| | FliH | 1.61 | 2.93 | Flagellar assembly protein |
| | ExuT | 1.64 | 4.76 | Hexuronate transporter |
| | OmpF | 1.65 | 5.34 | Outer membrane protein F |
| | FliF | 1.66 | 5.61 | Flagellar M-ring protein |
| | IdnK | 1.74 | 2.82 | Thermosensitive gluconokinase |
| | AllB | 1.77 | 4.96 | Allantoinase |
| | Tar | 1.78 | 5.15 | Methyl-accepting chemotaxis protein II |
| | IdnD | 1.80 | 4.83 | L-idonate 5-dehydrogenase (NAD(P)(+)) |
| | YcaK | 1.84 | 6.09 | Uncharacterized NAD(P)H oxidoreductase |
| | FlgH | 1.84 | 5.20 | Flagellar L-ring protein |
| | DtpA | 1.88 | 4.61 | Dipeptide and tripeptide permease |
| | YjcZ | 1.89 | 5.29 | Uncharacterized protein |
| | NfrB | 1.90 | 5.25 | Bacteriophage N4 adsorption protein B |
| | SgcQ | 1.93 | 3.57 | Putative nucleoside triphosphatase |
| | FliI | 1.94 | 4.67 | Flagellum-specific ATP synthase |
| | FlhE | 1.94 | 3.96 | Flagellar protein FlhE |
| | CheR | 1.96 | 3.15 | Chemotaxis protein methyltransferase |

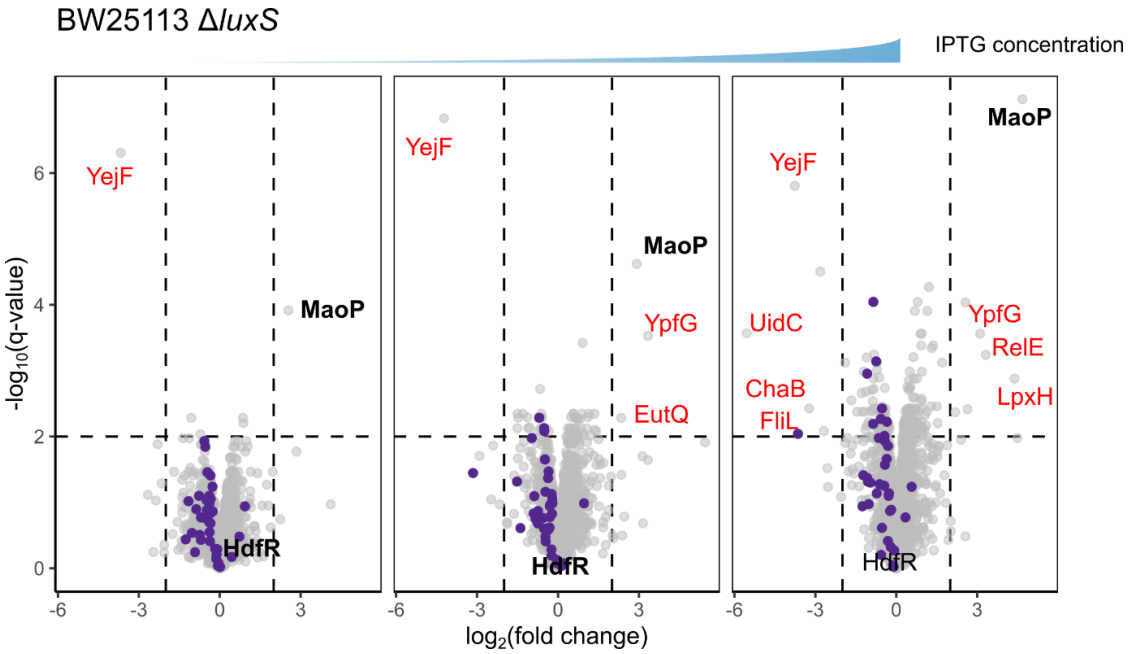
| | | | | |
|---|---|---|---|---|
| | YnjH | 2.04 | 5.07 | Uncharacterized protein |
| | YcgR | 2.06 | 4.50 | Flagellar brake protein |
| | SgcC | 2.07 | 2.81 | Putative permease IIC component |
| | SdaC | 2.15 | 2.14 | Serine transporter |
| | Trg | 2.16 | 5.61 | Methyl-accepting chemotaxis protein III |
| | YmgD | 2.34 | 2.90 | Uncharacterized protein |
| Δ*maoP* to WT | MaoP | -5.37 | 6.50 | Macrodomain Ori Protein |
| | YagE | -4.18 | 8.48 | Probable 2-keto-3-deoxy-galactonate aldolase YagE |
| | YeeR | -3.05 | 7.46 | Uncharacterized protein |
| | YagF | -2.98 | 6.41 | Uncharacterized protein |
| | SufA | -2.73 | 2.19 | Iron-sulfur cluster insertion protein |
| | YagH | -2.31 | 4.46 | Putative beta-xylosidase |
| | YagI | -2.18 | 6.62 | Uncharacterized HTH-type transcriptional regulator |
| | YeaC | -1.91 | 5.06 | Uncharacterized protein |
| | YgaM | -1.78 | 3.37 | Uncharacterized protein |
| | DppB | -1.59 | 2.74 | Dipeptide transport system permease |
| | GadE | -1.57 | 4.26 | Transcriptional activator of acid stress response |
| | HdeB | -1.57 | 4.82 | Acid stress chaperone |
| | YjiN | -1.57 | 3.33 | Uncharacterized protein |
| | PncB | -1.56 | 2.79 | Nicotinate phosphoribosyltransferase |
| | FliF | 1.50 | 5.90 | Flagellar M-ring protein |
| | XylG | 1.56 | 4.75 | Xylose import ATP-binding protein |
| | OmpF | 1.57 | 5.24 | Outer membrane protein F |
| | SgcX | 1.58 | 6.49 | Putative aminopeptidase |
| | YnjH | 1.59 | 4.64 | Uncharacterized protein |
| | IdnO | 1.61 | 6.32 | Gluconate 5-dehydrogenase |
| | FlhE | 1.61 | 3.34 | Flagellar protein FlhE |
| | Tap | 1.63 | 5.32 | Methyl-accepting chemotaxis protein IV |

| | | | | |
|---|---|---|---|---|
| | Tar | 1.67 | 4.80 | Methyl-accepting chemotaxis protein II |
| | FliE | 1.76 | 2.49 | Flagellar hook-basal body complex protein |
| | FliI | 1.76 | 4.87 | Flagellum-specific ATP synthase |
| | YcaK | 1.76 | 6.13 | Uncharacterized NAD(P)H oxidoreductase |
| | FlgH | 1.77 | 5.15 | Flagellar L-ring protein |
| | AllB | 1.81 | 5.32 | Allantoinase |
| | IdnD | 1.81 | 5.31 | L-idonate 5-dehydrogenase (NAD(P)(+)) |
| | YjcZ | 1.85 | 5.60 | Uncharacterized protein |
| | SgcC | 1.90 | 2.79 | Putative permease IIC component |
| | DtpA | 1.91 | 4.87 | Dipeptide and tripeptide permease |
| | YcgR | 1.93 | 4.70 | Flagellar brake protein |
| | NfrB | 1.97 | 4.99 | Bacteriophage N4 adsorption protein B |
| | IdnK | 1.97 | 3.82 | Thermosensitive gluconokinase |
| | SgcQ | 2.00 | 3.47 | Putative nucleoside triphosphatase |
| | YqjC | 2.04 | 3.63 | Uncharacterized protein |
| | Trg | 2.16 | 5.67 | Methyl-accepting chemotaxis protein III |
| | YmgG | 2.40 | 4.17 | Uncharacterized protein |
| | YmgD | 2.65 | 3.27 | Uncharacterized protein |
| Δ*hdfR* to Δ*maoP* | HdfR | -3.70 | 6.33 | HTH-type transcriptional regulator |
| | YagF | 1.73 | 4.69 | Uncharacterized protein |
| | YagI | 2.90 | 5.11 | Uncharacterized HTH-type transcriptional regulator |
| | YagE | 2.92 | 4.36 | Probable 2-keto-3-deoxy-galactonate aldolase YagE |
| | YagH | 2.95 | 3.88 | Putative beta-xylosidase |

Proteins belonging to the flagellar regulon are highlighted in blue. Only the proteins with absolute fold change greater than 1.5 are shown.




**Supplementary figure 10. STRING diagram for the proteins upregulated in the Δ*hdfR* knockout.** Colored are: flagellar and motility network (blue), L-idonate metabolism (red). The networks are constructed using the STRING database [175]. Thickness of the lines indicates the strength of data support in STRING.



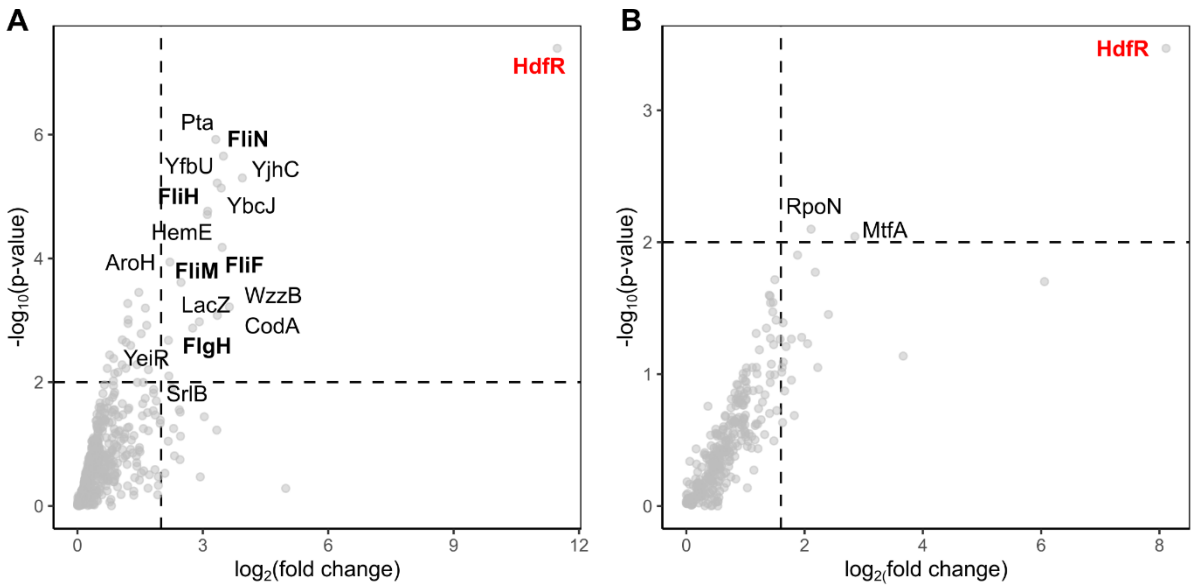

**Supplementary figure 11. Effect of MaoP overexpression.** The proteins of flagellar regulon are highlighted in purple.

**Supplementary table 4. Global changes in whole proteome upon overexpression of HdfR or MaoP.**

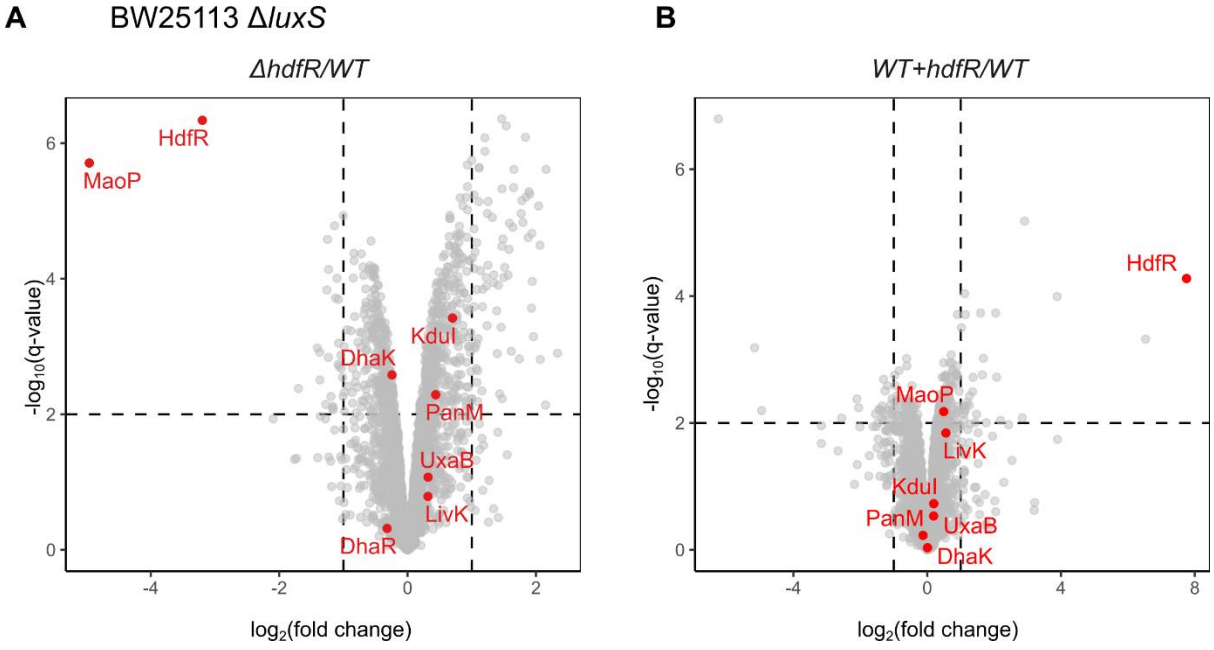
| Overexpressed protein | induction | Protein name | Fold change | -log$_{10}$(p-value) | Protein description |
|---|---|---|---|---|---|
| **HdfR** | 0 | RecN | 6.53 | 3.32 | DNA repair protein RecN |
| | 0 | DinD | 3.88 | 3.99 | DNA-damage-inducible protein D |
| | 0 | DinI | 2.91 | 5.18 | DNA-damage-inducible protein I |
| | 0 | RecX | 2.84 | 2.08 | Regulatory protein RecX |
| | 0 | YobB | -2.56 | 2.08 | Uncharacterized protein YobB |
| | 0 | ChaB | -4.95 | 2.19 | Cation transport regulator ChaB |
| | 0 | YeeR | -5.16 | 3.19 | Inner membrane protein YeeR |
| | 0 | Flu | -6.24 | 6.79 | Antigen 43 |
| | 1 | RecN | 5.86 | 2.66 | DNA repair protein RecN |
| | 1 | DinD | 4.12 | 4.19 | DNA-damage-inducible protein D |
| | 1 | DinI | 2.83 | 4.41 | DNA-damage-inducible protein I |
| | 1 | HisI | 2.70 | 2.20 | Histidine biosynthesis bifunctional protein HisIE |
| | 1 | YeeR | -5.71 | 3.32 | Inner membrane protein YeeR |
| | 1 | Flu | -6.00 | 5.86 | Antigen 43 |
| | 10 | RecN | 7.18 | 3.76 | DNA repair protein RecN |
| | 10 | FrdD | 5.69 | 2.62 | Fumarate reductase subunit D |
| | 10 | DinD | 5.13 | 4.46 | DNA-damage-inducible protein D |
| | 10 | RmuC | 3.89 | 3.90 | DNA recombination protein RmuC |
| | 10 | YhbT | 3.53 | 2.82 | Uncharacterized protein YhbT |
| | 10 | RecX | 3.10 | 2.40 | Regulatory protein RecX |
| | 10 | DinI | 3.06 | 5.74 | DNA-damage-inducible protein I |
| | 10 | MdtC | 2.90 | 2.94 | Multidrug resistance protein MdtC |
| | 10 | HisI | 2.57 | 2.82 | Histidine biosynthesis bifunctional protein HisIE |
| | 10 | RecQ | 2.53 | 2.37 | ATP-dependent DNA helicase RecQ |

| | | | | | |
|---|---|---|---|---|---|
| | 10 | SbmC | 2.52 | 2.17 | DNA gyrase inhibitor |
| | 10 | UidA | -2.51 | 3.05 | Beta-glucuronidase |
| | 10 | FliH | -2.58 | 2.75 | Flagellar assembly protein FliH |
| | 10 | IntS | -2.66 | 2.01 | Putative prophage CPS-53 integrase |
| | 10 | FimC | -3.01 | 3.07 | Chaperone protein FimC |
| | 10 | FlgK | -3.02 | 2.98 | Flagellar hook-associated protein 1 |
| | 10 | YgcE | -3.17 | 2.98 | Uncharacterized sugar kinase YgcE |
| | 10 | ChaB | -3.70 | 2.92 | Cation transport regulator ChaB |
| | 10 | YciE | -4.23 | 2.79 | Protein YciE |
| | 10 | FimD | -4.44 | 3.47 | Outer membrane usher protein FimD |
| | 10 | YeeR | -5.16 | 3.40 | Inner membrane protein YeeR |
| | 10 | Flu | -6.19 | 6.72 | Antigen 43 |
| **MaoP** | 0 | YejF | -3.67 | 6.30 | Uncharacterized ABC transporter ATP-binding protein YejF |
| | 1 | YpfG | 3.34 | 3.53 | Uncharacterized protein YpfG |
| | 1 | YejF | -4.23 | 6.83 | Uncharacterized ABC transporter ATP-binding protein YejF |
| | 10 | LpxH | 4.39 | 2.88 | UDP-2,3-diacylglucosamine hydrolase |
| | 10 | RelE | 3.32 | 3.24 | mRNA interferase RelE |
| | 10 | YpfG | 3.11 | 3.56 | Uncharacterized protein YpfG |
| | 10 | HisI | 2.64 | 2.41 | Histidine biosynthesis bifunctional protein HisIE |
| | 10 | YjgL | 2.56 | 4.04 | Uncharacterized protein YjgL |
| | 10 | BioF | -2.69 | 2.08 | 8-amino-7-oxononanoate synthase |
| | 10 | UidA | -2.82 | 4.50 | Beta-glucuronidase |
| | 10 | ChaB | -3.23 | 2.42 | Cation transport regulator ChaB |
| | 10 | FliL | -3.65 | 2.04 | Flagellar protein FliL |
| | 10 | YejF | -3.76 | 5.80 | Uncharacterized ABC transporter ATP-binding protein YejF |
| | 10 | UidC | -5.55 | 3.57 | Membrane-associated protein UidC |

Proteins involved in SOS-response are highlighted in pink. Only the proteins with absolute fold change greater than 2 are shown.




**Supplementary figure 12. Co-IP (A) and proximity labelling (B) results for HdfR.** The flagellar apparatus genes are highlighted in bold




**Supplementary figure 13. Potential regulatory targets of HdfR as determined by ChIP-Seq in the knockout (A) and the overexpression (B) strains.**

## 6.3. Supplementary data for Materials and methods

**Supplementary table 5. Strains**

| Name | Background | Relevant genotype | Source |
|---|---|---|---|
| W3110 RpoS$^+$ | | wild type | 52 |
| MG1655 | | wild type | 72 |
| Keio collection | BW11325 | full collection of the knockouts | 160, 161 |
| NT6 | W3110 RpoS$^+$ | *Δaer* | |
| NT7 | W3110 RpoS$^+$ | *Δtap* | |
| NT8 | W3110 RpoS$^+$ | *Δtrg* | |
| NT9 | W3110 RpoS$^+$ | *Δtar* | |
| NT10 | W3110 RpoS$^+$ | *Δtsr* | |
| VS604 | W3110 RpoS$^+$ | *ΔcsgA* | Constructed by Verena Suchanek |
| VS646 | W3110 RpoS$^+$ | *ΔmotA* | Constructed by Verena Suchanek |
| BW25113 | BW25113 | wild type | 160, 161 |
| BW25113 *ΔluxS* | BW25113 | *ΔluxS* | 160 |
| VS644 | W3110 RpoS$^+$ | *ΔcheA* | Constructed by Verena Suchanek |
| VS599 | W3110 RpoS$^+$ | *ΔcheR* | Constructed by Verena Suchanek |
| VS596 | W3110 RpoS$^+$ | *ΔcheY* | Constructed by Verena Suchanek |
| VS578 | W3110 RpoS$^+$ | *ΔfliC* | Constructed by Verena Suchanek |
| OB302 | W3110 RpoS$^+$ | *ΔfimH* | Constructed by Olga Lamprecht |
| OB277 | W3110 RpoS$^+$ | *ΔompR* | Constructed by Olga Lamprecht |
| ME69 | W3110 RpoS$^+$ | *ΔfimA* | Constructed by Maria Esteban-Lopez |
| SH180 | W3110 RpoS$^+$ | *ΔcheA* | Constructed by Sarah Hoch |
| VM329 | W3110 RpoS$^+$ | *ΔmotA* | Constructed by Verena Suchanek |
| OB283 | W3110 RpoS$^+$ | *ΔflhD* | Constructed by Olga Lamprecht |
| NT134 | BW25113 *ΔluxS* | *ΔrpoS* | |

| **Name** | **Background** | **Relevant genotype** | **Source** |
|---|---|---|---|
| NT136 | BW25113 *ΔluxS* | *ΔhdfR* | |
| NT137 | W3110 RpoS$^+$ | *ΔflhD* | |
| NT241 | W3110 RpoS$^+$ | *ΔfliCΔhdfR* | |
| NT243 | W3110 RpoS$^+$ | *ΔmotAΔhdfR* | |
| NT278 | W3110 RpoS$^+$ | *ΔmaoP* | |
| NT280 | W3110 RpoS$^+$ | *ΔflhDΔhdfR* | |
| NT282 | BW25113 *ΔluxS* | *ΔhdfRΔrpoS* | |
| NT331 | BW25113 *ΔluxS* | *Δdam* | |
| NT339 | BW25113 *ΔluxS* | *ΔdamΔhdfR* | |
| NT340 | BW25113 *ΔluxS* | *ΔdamΔmaoP* | |
| IS123 | MG1655 | *hupA-mCh::frt::kan::frt* | Constructed by Ismath Sadhir |
| ASM76 | W3110 RpoS$^+$ | transposon mutant library | Constructed by Ananda Medeiros |
| NT379 | W3110 RpoS$^+$ | *ΔhdfR* | |
| NT380 | W3110 RpoS$^+$ | *ΔmaoP* | |
| NT381 | VS558 | *ΔrpoS* | |
| NT382 | MG1655 | *ΔhdfR* | |
| NT383 | MG1655 | *ΔmaoP* | |
| NT384 | MG1655 | *ΔrpoS* | |
| NT386 | W3110 RpoS$^+$ | *ΔrpoSΔhdfR* | |
| NT389 | W3110 RpoS$^+$ | *ΔmaoS* | |
| NT390 | MG1655 | *ΔmaoS* | |
| IS 169 | MG1655 | *matP-ypet::frt oriC-parS::frt::cat::frt* | Constructed by Ismath Sadhir |
| NT393 | W3110 RpoS$^+$ | *ΔrpoZ* | |
| NT395 | W3110 RpoS$^+$ | *ΔrpoZ ΔmaoP* | |

| Name | Background | Relevant genotype | Source |
|---|---|---|---|
| NT399 | MG1655 | *matP-ypet::frt oriC-parS::frt::cat::frt ΔhdfR* | |
| NT400 | MG1655 | *matP-ypet::frt oriC-parS::frt::cat::frt ΔmaoP KanS, CAMS* | |
| NT401 | MG1655 | *matP-ypet::frt oriC-parS::frt::cat::frt ΔmaoS* | |

All listed strains are Biosafety level 1. The strains are constructed by Nataliya Teteneva unless specified otherwise. Bacteria were routinely grown in a rotary shaker at 37°C on Luria-Bertrani (LB) or Tryptone broth (TB) medium supplemented with the relevant antibiotic (50 µg/ml of kanamycin, 100 µg/ml of ampicillin, 34 µg/ml of chloramphenicol) if necessary. All the strains were stored at -80°C on LB medium supplemented with 20% glycerol.

**Supplementary table 6. Oligonucleotides**

| Name | Gene/Purpose | Sequence |
|---|---|---|
| BarSeq common | TnSeq | aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctnnnnngtcgacctgcagcgtacg |
| BarSeq 1 | TnSeq | caagcagaagacggcatacgagatcgtgatgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 2 | TnSeq | caagcagaagacggcatacgagatacatcggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 3 | TnSeq | caagcagaagacggcatacgagatgcctaagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 4 | TnSeq | caagcagaagacggcatacgagattggtcagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 5 | TnSeq | caagcagaagacggcatacgagatcactgtgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 6 | TnSeq | caagcagaagacggcatacgagatattggcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 7 | TnSeq | caagcagaagacggcatacgagatgatctggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 8 | TnSeq | caagcagaagacggcatacgagattcaagtgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 9 | TnSeq | caagcagaagacggcatacgagatctgatcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |

| Name | Gene/Purpose | Sequence |
|---|---|---|
| BarSeq 10 | TnSeq | caagcagaagacggcatacgagatctgatcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 11 | TnSeq | caagcagaagacggcatacgagatgtagccgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 12 | TnSeq | caagcagaagacggcatacgagattacaaggtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 13 | TnSeq | caagcagaagacggcatacgagatttgactgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 14 | TnSeq | caagcagaagacggcatacgagatggaactgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 15 | TnSeq | caagcagaagacggcatacgagattgacatgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 16 | TnSeq | caagcagaagacggcatacgagatggacgggtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 17 | TnSeq | caagcagaagacggcatacgagatctctacgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 18 | TnSeq | caagcagaagacggcatacgagatgcggacgtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 19 | TnSeq | caagcagaagacggcatacgagattttcacgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 20 | TnSeq | caagcagaagacggcatacgagatggccacgtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 21 | TnSeq | caagcagaagacggcatacgagatcgaaacgtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 22 | TnSeq | caagcagaagacggcatacgagatcgtacggtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 23 | TnSeq | caagcagaagacggcatacgagatccactcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 24 | TnSeq | caagcagaagacggcatacgagatgctaccgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 25 | TnSeq | caagcagaagacggcatacgagatatcagtgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 26 | TnSeq | caagcagaagacggcatacgagatgctcatgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 27 | TnSeq | caagcagaagacggcatacgagataggaatgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |

| Name | Gene/Purpose | Sequence |
|---|---|---|
| BarSeq 28 | TnSeq | caagcagaagacggcatacgagatcttttggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 29 | TnSeq | caagcagaagacggcatacgagattagttggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 30 | TnSeq | caagcagaagacggcatacgagatccggtggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 31 | TnSeq | caagcagaagacggcatacgagatatcgtggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 32 | TnSeq | caagcagaagacggcatacgagattgagtggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 33 | TnSeq | caagcagaagacggcatacgagatcgcctggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 34 | TnSeq | caagcagaagacggcatacgagatgccatggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 35 | TnSeq | caagcagaagacggcatacgagataaaatggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 36 | TnSeq | caagcagaagacggcatacgagattgttgggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 37 | TnSeq | caagcagaagacggcatacgagatattccggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 38 | TnSeq | caagcagaagacggcatacgagatagctaggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 39 | TnSeq | caagcagaagacggcatacgagatgtataggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 40 | TnSeq | caagcagaagacggcatacgagattctgaggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 41 | TnSeq | caagcagaagacggcatacgagatgtcgtcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 42 | TnSeq | caagcagaagacggcatacgagatcgattagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 43 | TnSeq | caagcagaagacggcatacgagatgctgtagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 44 | TnSeq | caagcagaagacggcatacgagatattatagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 45 | TnSeq | caagcagaagacggcatacgagatgaatgagtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |

| Name | Gene/Purpose | Sequence |
|---|---|---|
| BarSeq 46 | TnSeq | caagcagaagacggcatacgagattcgggagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 47 | TnSeq | caagcagaagacggcatacgagatcttcgagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 48 | TnSeq | caagcagaagacggcatacgagattgccgagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 49 | TnSeq | caagcagaagacggcatacgagatatgtttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 50 | TnSeq | caagcagaagacggcatacgagattgctttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 51 | TnSeq | caagcagaagacggcatacgagatgcatttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 52 | TnSeq | caagcagaagacggcatacgagattttgttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 53 | TnSeq | caagcagaagacggcatacgagatcaagttgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 54 | TnSeq | caagcagaagacggcatacgagatagtcttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 55 | TnSeq | caagcagaagacggcatacgagattcgcttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 56 | TnSeq | caagcagaagacggcatacgagatgtccttgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 57 | TnSeq | caagcagaagacggcatacgagatcctattgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 58 | TnSeq | caagcagaagacggcatacgagatgtttgtgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 59 | TnSeq | caagcagaagacggcatacgagatagatgtgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 60 | TnSeq | caagcagaagacggcatacgagatctgggtgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 61 | TnSeq | caagcagaagacggcatacgagatgccggtgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 62 | TnSeq | caagcagaagacggcatacgagattatcgtgtgactggagttcagacgtgtgctcttccgatctgatgtccacga ggtctct |
| BarSeq 63 | TnSeq | caagcagaagacggcatacgagatgagagtgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |

| Name | Gene/Purpose | Sequence |
|---|---|---|
| BarSeq 64 | TnSeq | caagcagaagacggcatacgagattcttctgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 65 | TnSeq | caagcagaagacggcatacgagatctatctgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 66 | TnSeq | caagcagaagacggcatacgagatgatgctgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 67 | TnSeq | caagcagaagacggcatacgagatagcgctgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 68 | TnSeq | caagcagaagacggcatacgagatcggcctgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 69 | TnSeq | caagcagaagacggcatacgagatccgtatgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 70 | TnSeq | caagcagaagacggcatacgagattaggatgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 71 | TnSeq | caagcagaagacggcatacgagatatagatgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 72 | TnSeq | caagcagaagacggcatacgagatgcgtgggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 73 | TnSeq | caagcagaagacggcatacgagatcatggggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 74 | TnSeq | caagcagaagacggcatacgagatttgcgggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 75 | TnSeq | caagcagaagacggcatacgagatctaagggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 76 | TnSeq | caagcagaagacggcatacgagatttctcggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 77 | TnSeq | caagcagaagacggcatacgagatcagcaggtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 78 | TnSeq | caagcagaagacggcatacgagatggtttcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 79 | TnSeq | caagcagaagacggcatacgagatttattcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 80 | TnSeq | caagcagaagacggcatacgagattccgtcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |
| BarSeq 81 | TnSeq | caagcagaagacggcatacgagattatatcgtgactggagttcagacgtgtgctcttccgatctgatgtccacgaggtctct |

| **Name** | **Gene/Purpose** | **Sequence** |
|---|---|---|
| BarSeq 82 | TnSeq | caagcagaagacggcatacgagatagcatcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 83 | TnSeq | caagcagaagacggcatacgagatccttgcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 84 | TnSeq | caagcagaagacggcatacgagataagtgcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 85 | TnSeq | caagcagaagacggcatacgagattaaggcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 86 | TnSeq | caagcagaagacggcatacgagattggagcgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 87 | TnSeq | caagcagaagacggcatacgagattgtgccgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 88 | TnSeq | caagcagaagacggcatacgagatcaggccgtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 89 | TnSeq | caagcagaagacggcatacgagatggtagagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 90 | TnSeq | caagcagaagacggcatacgagatcattcagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 91 | TnSeq | caagcagaagacggcatacgagatatggcagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 92 | TnSeq | caagcagaagacggcatacgagatccagcagtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 93 | TnSeq | caagcagaagacggcatacgagatgcgccagtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 94 | TnSeq | caagcagaagacggcatacgagatttcgaagtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| BarSeq 95 | TnSeq | caagcagaagacggcatacgagatggagaagtgactggagttcagacgtgtgctcttccgatctgatgtccac gaggtctct |
| BarSeq 96 | TnSeq | caagcagaagacggcatacgagataaacctgtgactggagttcagacgtgtgctcttccgatctgatgtccacg aggtctct |
| rpoS fw | *rpoS* | tgatcgaaacctttggcgcttc |
| rpoS rv | *rpoS* | gagcttacaacacaccagcgac |
| hdfR fw | *hdfR* | agatgcgatgtggttggaagga |
| hdfR rv | *hdfR* | ccgaactcagaagtgaaacgcc |
| SacI_hdfR_fw | cloning *hdfR* into plasmid | gatccagagctcacgacaaataattttgcggaga |

| **Name** | **Gene/Purpose** | **Sequence** |
|---|---|---|
| XbaI_hdfR_rv | cloning *hdfR* into plasmid | caggtatctagacctctccgaagtaaatccttct |
| SacI_maoP_fw | cloning *maoP* into plasmid | gatccagagctcagcgtaactaatactccgcgcc |
| XbaI_maoP_rv | cloning *maoP* into plasmid | attctgtctagaagatgcgatgtggttggaagga |
| qPCR_dxs_fw | qPCR for *dxs* | aaggcccgcagttcctgcat |
| qPCR_dxs_rv | qPCR for *dxs* | ggcaaaccgccgctacttttc |
| NT_255_maoP_RBS_fw | *maoP* fusion with native RBS | actagtgaaggagtgtgccatgggtagcgtaactaatactccgcgccata |
| maoP_fus_rv_fx | Gibson GFP fusion | caccccttggagctacctccgcccccatcgtcagaatcggtgtagtcttca |
| hdfR_fus_fw | Gibson GFP fusion | actagtgaaggagtgtgccatgggtgtggatacggaattgttaaaaactt |
| hdfR_fus_rv_fx | Gibson GFP fusion | caccccttggagctacctccgcccccatacacttcatccagcacgttaatt |
| NBp1 | pNB1 linearization | acccatggcacactccttcactag |
| NBp68 | pNB1 linearization | gggggcggaggtagctccaagggtgaagagctatttactg |
| NBp5 | pTrc99a insert check | ttgacaattaatcatccggctcg |
| NBp6 | pTrc99a insert check | acagccaagcttgcatgcct |
| NBp18 | pTrc99a seq primer | gatttaatctgtatcagg |
| dam_fw | *dam* | aggttatctcccgcaatggttt |
| dam_rw | *dam* | catcaggtgtacgtcgataggg |
| qPCR_GFP4_fw | qPCR for eGFP | cttgttccatggccaaca |
| qPCR_GFP4_rv | qPCR for eGFP | acttcagcacgtgtcttgt |
| linearN_fw | linearization of pNB1 for N-term fusions | tctagagtcgacctgcaggc |

| **Name** | **Gene/Purpose** | **Sequence** |
|---|---|---|
| linearN_rv | linearization of pNB1 for N-term fusions | gctacctccgcccccccgacccctttataaagctcgtcc |
| maoPN_fw | *maoP* N term fusion | taaggggtcgggggggcggaggtagcgcggaaagctttacgacgactaatc |
| maoPN_rv | *maoP* N term fusion | tgcatgcctgcaggtcgactctagattaatcgtcagaatcggtgtagtct |
| NT_271_PL1 | TurboID | aggagtgtgccatgggtatgatcccgctcctgaacgcgaa |
| NT_272_PL2 | TurboID | gctacctccgcccccccttctcggcgctgcgcaggc |
| NT_273_PL3 | TurboID-GFP | gccgagaaggggggcggaggtagctccaagggtgaagagctatt |
| NT_274_PL4 | GFP-FLAG-vector | gcctgcaggtcgactctagattacttgtcgtcgtcgtccttgtagtccgacccctttataaagctcgt |
| NT_275_PL5 | TurboID-*hdfR* | cgccgagaaggggggcggaggtagcgatacggaattgttaaaaactttc |
| NT_276_PL6 | *hdfR*-FLAG-vector | aagcttgcatgcctgcaggtcgactctagattacttgtcgtcgtcgtccttgtagtcatacacttcatccagcacgt |
| NT_277_PL7 | TurboID-*maoP* | cgccgagaaggggggcggaggtagcgcggaaagctttacgacgac |
| NT_278_PL8 | *maoP*-FLAG-vector | gcttgcatgcctgcaggtcgactctagattacttgtcgtcgtcgtccttgtagtcatcgtcagaatcggtgtagt |
| NT_279_PL9 | Linker-*hdfR* | ggagcgggatgctacctccgcccccatacacttcatccagcacg |
| NT_280_PL10 | Linker-*maoP* | gggatgctacctccgcccccatcgtcagaatcggtgtagt |
| NT_281_PL11 | Linker-TurboID | gggggcggaggtagcatcccgctcctgaacgcgaa |
| NT_282_PL12 | Vector-TurboID | tgcatgcctgcaggtcgactctagattacttgtcgtcgtcgtccttgt |
| OB183 | *flhD* | ggtggttctgcttattgcagc |
| OB184 | *flhD* | cttccactgttgaccatgacag |
| flhD_seqfw | *flhD* seq | cctgaagggaaagctgcacg |
| oriC1_fw | qPCR *oriC* | ttatccacagggcagtgcg |
| oriC1_rv | qPCR *oriC* | cgggccgtggattctactc |
| terC2_fw | qPCR *terC* | aatcagcacgttgagcgga |
| terC2_fv | qPCR *terC* | tctcgacaccgctgaacac |

| Name | Gene/Purpose | Sequence |
|---|---|---|
| maoS_l_fw | remove both *maoP* and *hdfR* | cataaaaaaagggcatttcgccctttttattaatcgtcagaatcggtgtgtgtaggctggagctgcttc |
| maoS_l_rv | remove both *maoP* and *hdfR* | aagatcaagctggcgttcatgcagctgttttaccagagactgccgttgggctgtcaaacatgagaattaa |
| rpoZ_fw | *rpoZ* | ttaccgccgtccaaaattgaac |
| rpoZ_rv | *rpoZ* | tcggcaagtttgatgaggatga |
| IB28_pBAD33 insert F | pBAD33 insert check | cactttgctatgccatag |
| IB29_pBAD33 insert F | pBAD33 insert check | gatttaatctgtatcaggctg |

## 6.4. List of figures

## 6.5. List of tables

# References

1. Escherich T. Die Darmbacterien des Neugeborenen und Säuglings. Fortschritte der Medicin. 1885;3:515-22.

2. Winfield MD, Groisman EA. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. Appl Environ Microbiol. 2003;69(7):3687-94.

3. Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol. 2010;8(3):207-17.

4. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the human intestinal microbial flora. Science. 2005;308(5728):1635-8.

5. Kaper JB, Nataro JP, Mobley HL. Pathogenic *Escherichia coli*. Nat Rev Microbiol. 2004;2(2):123-40.

6. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic *Escherichia coli*. Nat Rev Microbiol. 2021;19(1):37-54.

7. Dale AP, Woodford N. Extra-intestinal pathogenic *Escherichia coli* (ExPEC): Disease, carriage and clones. J Infect. 2015;71(6):615-26.

8. Khan FM, Gupta R, editors. *Escherichia coli (E. coli)* as an Indicator of Fecal Contamination in Groundwater: A Review. Sustainable Development of Water and Environment; 2020 2020//; Cham: Springer International Publishing.

9. Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MM, Ramadan HA, Saini KS, et al. Production of Biopharmaceuticals in *E. coli*: Current Scenario and Future Perspectives. J Microbiol Biotechnol. 2015;25(7):953-62.

10. van Elsas JD, Semenov AV, Costa R, Trevors JT. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. ISME J. 2011;5(2):173-83.

11. Ishii S, Sadowsky MJ. *Escherichia coli* in the Environment: Implications for Water Quality and Human Health. Microbes Environ. 2008;23(2):101-8.

12. Solo-Gabriele HM, Wolfert MA, Desmarais TR, Palmer CJ. Sources of *Escherichia coli* in a coastal subtropical environment. Appl Environ Microbiol. 2000;66(1):230-7.

13. Brennan FP, O'Flaherty V, Kramers G, Grant J, Richards KG. Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. Appl Environ Microbiol. 2010;76(5):1449-55.

14. Ishii S, Yan T, Vu H, Hansen DL, Hicks RE, Sadowsky MJ. Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. Microbes Environ. 2010;25(1):8-14.

15. Jang J, Hur HG, Sadowsky MJ, Byappanahalli MN, Yan T, Ishii S. Environmental *Escherichia coli:* ecology and public health implications - a review. J Appl Microbiol. 2017;123(3):570-81.

16. Wright KM, Crozier L, Marshall J, Merget B, Holmes A, Holden NJ. Differences in internalization and growth of *Escherichia coli* O157:H7 within the apoplast of edible plants, spinach and lettuce, compared with the model species *Nicotiana benthamiana*. Microb Biotechnol. 2017;10(3):555-69.

17. Williams AP, Roberts P, Avery LM, Killham K, Jones DL. Earthworms as vectors of *Escherichia coli* O157:H7 in soil and vermicomposts. FEMS Microbiol Ecol. 2006;58(1):54-64.

18. Rozen Y, Belkin S. Survival of enteric bacteria in seawater. FEMS Microbiol Rev. 2001;25(5):513-29.

19. Ishii S, Ksoll WB, Hicks RE, Sadowsky MJ. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl Environ Microbiol. 2006;72(1):612-21.

20. Rivera SC, Hazen TC, Toranzos GA. Isolation of fecal coliforms from pristine sites in a tropical rain forest. Appl Environ Microbiol. 1988;54(2):513-7.

21. Texier S, Prigent-Combaret C, Gourdon MH, Poirier MA, Faivre P, Dorioz JM, et al. Persistence of culturable *Escherichia coli* fecal contaminants in dairy alpine grassland soils. J Environ Qual. 2008;37(6):2299-310.

22. Zhu YG, Zhao Y, Zhu D, Gillings M, Penuelas J, Ok YS, et al. Soil biota, antimicrobial resistance and planetary health. Environ Int. 2019;131:105059.

23. Knopp M, Andersson DI. Predictable Phenotypes of Antibiotic Resistance Mutations. mBio. 2018;9(3).

24. Carstens CK, Salazar JK, Darkoh C. Multistate Outbreaks of Foodborne Illness in the United States Associated With Fresh Produce From 2010 to 2017. Front Microbiol. 2019;10:2667.

25. Sarno E, Pezzutto D, Rossi M, Liebana E, Rizzi V. A Review of Significant European Foodborne Outbreaks in the Last Decade. J Food Prot. 2021;84(12):2059-70.

26. Gambushe SM, Zishiri OT, El Zowalaty ME. Review of *Escherichia coli* O157:H7 Prevalence, Pathogenicity, Heavy Metal and Antimicrobial Resistance, African Perspective. Infect Drug Resist. 2022;15:4645-73.

27. Popa GL, Papa MI. *Salmonella* spp. infection - a continuous threat worldwide. Germs. 2021;11(1):88-96.

28. Barnhart MM, Chapman MR. Curli biogenesis and function. Annu Rev Microbiol. 2006;60:131-47.

29. Besharova O, Suchanek VM, Hartmann R, Drescher K, Sourjik V. Diversification of Gene Expression during Formation of Static Submerged Biofilms by *Escherichia coli*. Front Microbiol. 2016;7:1568.

30. Hufnagel DA, Depas WH, Chapman MR. The Biology of the *Escherichia coli* Extracellular Matrix. Microbiol Spectr. 2015;3(3).

31. Evans SE, Wallenstein MD. Soil microbial community response to drying and rewetting stress: does historical precipitation regime matter? Biogeochemistry. 2012;109(1):101-16.

32. Unden G, Bongaerts J. Alternative respiratory pathways of *Escherichia coli:* energetics and transcriptional regulation in response to electron acceptors. Biochim Biophys Acta. 1997;1320(3):217-34.

33. Ammar EM, Wang X, Rao CV. Regulation of metabolism in *Escherichia coli* during growth on mixtures of the non-glucose sugars: arabinose, lactose, and xylose. Sci Rep. 2018;8(1):609.

34. Maser A, Peebo K, Vilu R, Nahku R. Amino acids are key substrates to *Escherichia coli* BW25113 for achieving high specific growth rate. Res Microbiol. 2020;171(5-6):185-93.

35. Salanitro JP, Wegener WS. Growth of *Escherichia coli* on short-chain fatty acids: nature of the uptake system. J Bacteriol. 1971;108(2):893-901.

36. Brennan FP, Grant J, Botting CH, O'Flaherty V, Richards KG, Abram F. Insights into the low-temperature adaptation and nutritional flexibility of a soil-persistent *Escherichia coli*. FEMS Microbiol Ecol. 2013;84(1):75-85.

37. Carlucci AF, Pramer D. An evaluation of factors affecting the survival of *Escherichia coli* in sea water. II. Salinity, pH, and nutrients. Appl Microbiol. 1960;8(4):247-50.

38. Lin J, Smith MP, Chapin KC, Baik HS, Bennett GN, Foster JW. Mechanisms of acid resistance in enterohemorrhagic *Escherichia coli*. Appl Environ Microbiol. 1996;62(9):3094-100.

39. Fujioka RS, Hashimoto HH, Siwak EB, Young RH. Effect of sunlight on survival of indicator bacteria in seawater. Appl Environ Microbiol. 1981;41(3):690-6.

40. Whitman RL, Nevers MB, Korinek GC, Byappanahalli MN. Solar and temporal effects on *Escherichia coli* concentration at a Lake Michigan swimming beach. Appl Environ Microbiol. 2004;70(7):4276-85.

41. Somorin Y, Bouchard G, Gallagher J, Abram F, Brennan F, O'Byrne C. Roles for RpoS in survival of *Escherichia coli* during protozoan predation and in reduced moisture conditions highlight its importance in soil environments. FEMS Microbiol Lett. 2017;364(19).

42. Enzinger RM, Cooper RC. Role of bacteria and protozoa in the removal of *Escherichia coli* from estuarine waters. Appl Environ Microbiol. 1976;31(5):758-63.

43. Jannasch HW. Competitive elimination of *Enterobacteriaceae* from seawater. Appl Microbiol. 1968;16(10):1616-8.

44. Carlucci AF, Pramer D. An evaluation of factors affecting the survival of *Escherichia coli* in sea water. IV. Bacteriophages. Appl Microbiol. 1960;8(4):254-6.

45. Di Cesare A, Riva F, Colinas N, Borgomaneiro G, Borin S, Cabello-Yeves PJ, et al. Zooplankton as a Transitional Host for *Escherichia coli* in Freshwater. Appl Environ Microbiol. 2022;88(9):e0252221.

46. McDougald D, Rice SA, Barraud N, Steinberg PD, Kjelleberg S. Should we stay or should we go: mechanisms and ecological consequences for biofilm dispersal. Nat Rev Microbiol. 2011;10(1):39-50.

47. Somorin Y, Abram F, Brennan F, O'Byrne C. The General Stress Response Is Conserved in Long-Term Soil-Persistent Strains of *Escherichia coli*. Appl Environ Microbiol. 2016;82(15):4628-40.

48. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol. 2004;2(2):95-108.

49. O'Toole G, Kaplan HB, Kolter R. Biofilm formation as microbial development. Annu Rev Microbiol. 2000;54:49-79.

50. Evans ML, Chapman MR. Curli biogenesis: order out of disorder. Biochim Biophys Acta. 2014;1843(8):1551-8.

51. Beloin C, Roux A, Ghigo JM. *Escherichia coli* biofilms. Curr Top Microbiol Immunol. 2008;322:249-89.

52. Serra DO, Richter AM, Klauck G, Mika F, Hengge R. Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. mBio. 2013;4(2):e00103-13.

53. Flemming HC, Wingender J. The biofilm matrix. Nat Rev Microbiol. 2010;8(9):623-33.

54. Molin S, Tolker-Nielsen T. Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr Opin Biotechnol. 2003;14(3):255-61.

55. Liu W, Roder HL, Madsen JS, Bjarnsholt T, Sorensen SJ, Burmolle M. Interspecific Bacterial Interactions are Reflected in Multispecies Biofilm Spatial Organization. Front Microbiol. 2016;7:1366.

56. Touchon M, Perrin A, de Sousa JAM, Vangchhia B, Burn S, O'Brien CL, et al. Phylogenetic background and habitat drive the genetic diversification of *Escherichia coli*. PLoS Genet. 2020;16(6):e1008866.

57. Horesh G, Blackwell GA, Tonkin-Hill G, Corander J, Heinz E, Thomson NR. A comprehensive and high-quality collection of *Escherichia coli* genomes and their genes. Microb Genom. 2021;7(2).

58. Waters NR, Abram F, Brennan F, Holmes A, Pritchard L. Easy phylotyping of *Escherichia coli* via the *EzClermont* web app and command-line tool. Access Microbiol. 2020;2(9):acmi000143.

59. Cummins EA, Hall RJ, Connor C, McInerney JO, McNally A. Distinct evolutionary trajectories in the *Escherichia coli* pangenome occur within sequence types. Microb Genom. 2022;8(11).

60. Clermont O, Bonacorsi S, Bingen E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol. 2000;66(10):4555-8.

61. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013;5(1):58-65.

62. Chaudhuri RR, Henderson IR. The evolution of the *Escherichia coli* phylogeny. Infect Genet Evol. 2012;12(2):214-26.

63. Bergthorsson U, Ochman H. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol Biol Evol. 1998;15(1):6-16.

64. Walk ST, Alm EW, Calhoun LM, Mladonicky JM, Whittam TS. Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. Environ Microbiol. 2007;9(9):2274-88.

65. Berthe T, Ratajczak M, Clermont O, Denamur E, Petit F. Evidence for coexistence of distinct *Escherichia coli* populations in various aquatic environments and their survival in estuary water. Appl Environ Microbiol. 2013;79(15):4684-93.

66. Ratajczak M, Laroche E, Berthe T, Clermont O, Pawlak B, Denamur E, et al. Influence of hydrological conditions on the *Escherichia coli* population structure in the water of a creek on a rural watershed. BMC Microbiol. 2010;10:222.

67. Rumball NA, Mayer HC, McLellan SL. Selective survival of *Escherichia coli* phylotypes in freshwater beach sand. Appl Environ Microbiol. 2020;87(4).

68. Clermont O, Condamine B, Dion S, Gordon DM, Denamur E. The E phylogroup of *Escherichia coli* is highly diverse and mimics the whole *E. coli* species population structure. Environ Microbiol. 2021;23(11):7139-51.

69. Clermont O, Dixit OVA, Vangchhia B, Condamine B, Dion S, Bridier-Nahmias A, et al. Characterization and rapid identification of phylogroup G in *Escherichia coli*, a lineage with high virulence and antibiotic resistance potential. Environ Microbiol. 2019;21(8):3107-17.

70. Munro PM, Flatau GN, Clement RL, Gauthier MJ. Influence of the RpoS (KatF) sigma factor on maintenance of viability and culturability of *Escherichia coli* and *Salmonella typhimurium* in seawater. Appl Environ Microbiol. 1995;61(5):1853-8.

71. Bleibtreu A, Clermont O, Darlu P, Glodt J, Branger C, Picard B, et al. The *rpoS* gene is predominantly inactivated during laboratory storage and undergoes source-sink evolution in *Escherichia coli* species. J Bacteriol. 2014;196(24):4276-84.

72. Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, et al. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol. 2006;2:2006 0007.

73. Ferenci T. What is driving the acquisition of *mutS* and *rpoS* polymorphisms in *Escherichia coli*? Trends Microbiol. 2003;11(10):457-61.

74. Chiang SM, Dong T, Edge TA, Schellhorn HE. Phenotypic diversity caused by differential RpoS activity among environmental *Escherichia coli* isolates. Appl Environ Microbiol. 2011;77(22):7915-23.

75. Dong T, Chiang SM, Joyce C, Yu R, Schellhorn HE. Polymorphism and selection of *rpoS* in pathogenic *Escherichia coli*. BMC Microbiol. 2009;9:118.

76. Hengge R. Stationary-Phase Gene Regulation in *Escherichia coli*. EcoSal Plus. 2011;4(2).

77. Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigma$^S$-dependent genes, promoters, and sigma factor selectivity. J Bacteriol. 2005;187(5):1591-603.

78. Cho BK, Kim D, Knight EM, Zengler K, Palsson BO. Genome-scale reconstruction of the sigma factor network in *Escherichia coli:* topology and functional states. BMC Biol. 2014;12:4.

79. Battesti A, Majdalani N, Gottesman S. The RpoS-Mediated General Stress Response in *Escherichia coli*. Annual Review of Microbiology. 2011;65(1):189-213.

80. Hengge R. Proteolysis of sigma$^S$ (RpoS) and the general stress response in *Escherichia coli*. Res Microbiol. 2009;160(9):667-76.

81. Lange R, Hengge-Aronis R. The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. Genes Dev. 1994;8(13):1600-12.

82. Traxler MF, Zacharia VM, Marquardt S, Summers SM, Nguyen HT, Stark SE, et al. Discretely calibrated regulatory loops controlled by ppGpp partition gene induction across the 'feast to famine' gradient in *Escherichia coli*. Mol Microbiol. 2011;79(4):830-45.

83. Hirsch M, Elliott T. Role of ppGpp in *rpoS* stationary-phase regulation in *Escherichia coli*. J Bacteriol. 2002;184(18):5077-87.

84. Muffler A, Fischer D, Hengge-Aronis R. The RNA-binding protein HF-I, known as a host factor for phage Qbeta RNA replication, is essential for *rpoS* translation in *Escherichia coli*. Genes Dev. 1996;10(9):1143-51.

85. Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. Proc Natl Acad Sci U S A. 1998;95(21):12462-7.

86. Majdalani N, Chen S, Murrow J, St John K, Gottesman S. Regulation of RpoS by a novel small RNA: the characterization of RprA. Mol Microbiol. 2001;39(5):1382-94.

87. Mandin P, Gottesman S. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. 2010;29(18):3094-107.

88. Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, Storz G. The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. EMBO J. 1998;17(20):6061-8.

89. Bougdour A, Wickner S, Gottesman S. Modulating RssB activity: IraP, a novel regulator of sigma$^S$ stability in *Escherichia coli*. Genes Dev. 2006;20(7):884-97.

90. Eguchi Y, Ishii E, Hata K, Utsumi R. Regulation of acid resistance by connectors of two-component signal transduction systems in *Escherichia coli*. J Bacteriol. 2011;193(5):1222-8.

91. Merrikh H, Ferrazzoli AE, Bougdour A, Olivier-Mason A, Lovett ST. A DNA damage response in *Escherichia coli* involving the alternative sigma factor, RpoS. Proc Natl Acad Sci U S A. 2009;106(2):611-6.

92. Foster JW. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat Rev Microbiol. 2004;2(11):898-907.

93. Opdyke JA, Kang JG, Storz G. GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. J Bacteriol. 2004;186(20):6698-705.

94. Kaasen I, McDougall J, Strom AR. Analysis of the *otsBA* operon for osmoregulatory trehalose synthesis in *Escherichia coli* and homology of the OtsA and OtsB proteins to the yeast trehalose-6-phosphate synthase/phosphatase complex. Gene. 1994;145(1):9-15.

95. Bou-Abdallah F, Lewin AC, Le Brun NE, Moore GR, Chasteen ND. Iron detoxification properties of *Escherichia coli* bacterioferritin. Attenuation of oxyradical chemistry. J Biol Chem. 2002;277(40):37064-9.

96. Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, Chiancone E, et al. Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of *Escherichia coli*. J Biol Chem. 2002;277(31):27689-96.

97. Eisenstark A. Bacterial genes involved in response to near-ultraviolet radiation. Adv Genet. 1989;26:99-147.

98. Persson O, Nystrom T, Farewell A. UspB, a member of the sigma$^S$ regulon, facilitates RuvC resolvase function. DNA Repair (Amst). 2010;9(11):1162-9.

99. Dong T, Schellhorn HE. Control of RpoS in global gene expression of *Escherichia coli* in minimal media. Mol Genet Genomics. 2009;281(1):19-33.

100. Raffaelle M, Kanin EI, Vogt J, Burgess RR, Ansari AZ. Holoenzyme switching and stochastic release of sigma factors from RNA polymerase in vivo. Mol Cell. 2005;20(3):357-66.

101. Olvera L, Mendoza-Vargas A, Flores N, Olvera M, Sigala JC, Gosset G, et al. Transcription analysis of central metabolism genes in *Escherichia coli*. Possible roles of sigma$^{38}$ in their expression, as a response to carbon limitation. PLoS One. 2009;4(10):e7466.

102. Patten CL, Kirchhof MG, Schertzberg MR, Morton RA, Schellhorn HE. Microarray analysis of RpoS-mediated gene expression in Escherichia coli K-12. Mol Genet Genomics. 2004;272(5):580-91.

103. Notley-McRobb L, King T, Ferenci T. *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. J Bacteriol. 2002;184(3):806-11.

104. Dong T, Yu R, Schellhorn H. Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*. Mol Microbiol. 2011;79(2):375-86.

105. Finkel SE. Long-term survival during stationary phase: evolution and the GASP phenotype. Nat Rev Microbiol. 2006;4(2):113-20.

106. Zambrano MM, Kolter R. GASPing for life in stationary phase. Cell. 1996;86(2):181-4.

107. Farrell MJ, Finkel SE. The growth advantage in stationary-phase phenotype conferred by *rpoS* mutations is dependent on the pH and nutrient environment. J Bacteriol. 2003;185(24):7044-52.

108. Zinser ER, Kolter R. *Escherichia coli* evolution during stationary phase. Res Microbiol. 2004;155(5):328-36.

109. Yeiser B, Pepper ED, Goodman MF, Finkel SE. SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. Proc Natl Acad Sci U S A. 2002;99(13):8737-41.

110. Tippin B, Pham P, Goodman MF. Error-prone replication for better or worse. Trends Microbiol. 2004;12(6):288-95.

111. Williams AB, Foster PL. Stress-Induced Mutagenesis. EcoSal Plus. 2012;5(1).

112. Aertsen A, Van Houdt R, Vanoirbeek K, Michiels CW. An SOS response induced by high pressure in *Escherichia coli*. J Bacteriol. 2004;186(18):6133-41.

113. Janion C, Sikora A, Nowosielska A, Grzesiuk E. Induction of the SOS response in starved *Escherichia coli*. Environ Mol Mutagen. 2002;40(2):129-33.

114. Layton JC, Foster PL. Error-prone DNA polymerase IV is controlled by the stress-response sigma factor, RpoS, in *Escherichia coli*. Mol Microbiol. 2003;50(2):549-61.

115. Tsui HC, Feng G, Winkler ME. Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. J Bacteriol. 1997;179(23):7476-87.

116. Atlung T, Nielsen HV, Hansen FG. Characterisation of the allelic variation in the *rpoS* gene in thirteen K12 and six other non-pathogenic *Escherichia coli* strains. Mol Genet Genomics. 2002;266(5):873-81.

117. Farewell A, Kvint K, Nystrom T. Negative regulation by RpoS: a case of sigma factor competition. Mol Microbiol. 1998;29(4):1039-51.

118. Ferenci T. The spread of a beneficial mutation in experimental bacterial populations: the influence of the environment and genotype on the fixation of *rpoS* mutations. Heredity (Edinb). 2008;100(5):446-52.

119. Mauri M, Klumpp S. A model for sigma factor competition in bacterial cells. PLoS Comput Biol. 2014;10(10):e1003845.

120. Ishihama A. Functional modulation of *Escherichia coli* RNA polymerase. Annu Rev Microbiol. 2000;54:499-518.

121. Seeto S, Notley-McRobb L, Ferenci T. The multifactorial influences of RpoS, Mlc and cAMP on *ptsG* expression under glucose-limited and anaerobic conditions. Res Microbiol. 2004;155(3):211-5.

122. Schellhorn HE. Function, Evolution, and Composition of the RpoS Regulon in *Escherichia coli*. Front Microbiol. 2020;11:560099.

123. Fitzgerald DM, Bonocora RP, Wade JT. Comprehensive mapping of the *Escherichia coli* flagellar regulatory network. PLoS Genet. 2014;10(10):e1004649.

124. Zhao K, Liu M, Burgess RR. Promoter and regulon analysis of nitrogen assimilation factor, sigma$^{54}$, reveal alternative strategy for *E. coli* MG1655 flagellar biosynthesis. Nucleic Acids Res. 2010;38(4):1273-83.

125. Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, Mehlis A, et al. Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. Genes Dev. 2008;22(17):2434-46.

126. Chevance FF, Hughes KT. Coordinating assembly of a bacterial macromolecular machine. Nat Rev Microbiol. 2008;6(6):455-65.

127. Evans LD, Hughes C, Fraser GM. Building a flagellum outside the bacterial cell. Trends Microbiol. 2014;22(10):566-72.

128. Minamino T, Kinoshita M, Inoue Y, Morimoto YV, Ihara K, Koya S, et al. FliH and FliI ensure efficient energy coupling of flagellar type III protein export in *Salmonella*. Microbiologyopen. 2016;5(3):424-35.

129. Macnab RM. Type III flagellar protein export and flagellar assembly. Biochim Biophys Acta. 2004;1694(1-3):207-17.

130. Terashima H, Kawamoto A, Tatsumi C, Namba K, Minamino T, Imada K. In Vitro Reconstitution of Functional Type III Protein Export and Insights into Flagellar Assembly. mBio. 2018;9(3).

131. Anderson JK, Smith TG, Hoover TR. Sense and sensibility: flagellum-mediated gene regulation. Trends Microbiol. 2010;18(1):30-7.

132. Wang S, Fleming RT, Westbrook EM, Matsumura P, McKay DB. Structure of the *Escherichia coli* FlhDC complex, a prokaryotic heteromeric regulator of transcription. J Mol Biol. 2006;355(4):798-808.

133. Ni B, Colin R, Link H, Endres RG, Sourjik V. Growth-rate dependent resource investment in bacterial motile behavior quantitatively follows potential benefit of chemotaxis. Proc Natl Acad Sci U S A. 2020;117(1):595-601.

134. Ko M, Park C. H-NS-Dependent regulation of flagellar synthesis is mediated by a LysR family protein. J Bacteriol. 2000;182(16):4670-2.

135. Fahrner KA, Berg HC. Mutations That Stimulate *flhDC* Expression in *Escherichia coli* K-12. J Bacteriol. 2015;197(19):3087-96.

136. Sperandio V, Torres AG, Kaper JB. Quorum sensing *Escherichia coli* regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in *E. coli*. Mol Microbiol. 2002;43(3):809-21.

137. Francez-Charlot A, Laugel B, Van Gemert A, Dubarry N, Wiorowski F, Castanie-Cornet MP, et al. RcsCDB His-Asp phosphorelay system negatively regulates the *flhDC* operon in *Escherichia coli*. Mol Microbiol. 2003;49(3):823-32.

138. Shin S, Park C. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. J Bacteriol. 1995;177(16):4696-702.

139. Lejars M, Caillet J, Solchaga-Flores E, Guillier M, Plumbridge J, Hajnsdorf E. Regulatory Interplay between RNase III and Antisense RNAs in *E. coli:* the Case of AsflhD and FlhD, Component of the Master Regulator of Motility. mBio. 2022;13(5):e0098122.

140. Chilcott GS, Hughes KT. Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar *typhimurium* and *Escherichia coli*. Microbiol Mol Biol Rev. 2000;64(4):694-708.

141. Rudenko I, Ni B, Glatter T, Sourjik V. Inefficient Secretion of Anti-sigma Factor FlgM Inhibits Bacterial Motility at High Temperature. iScience. 2019;16:145-54.

142. Sandle T. Bacterial Adhesion: an Introduction. Journal of Validation Technology. 2013;17.

143. Romling U, Amikam D. Cyclic di-GMP as a second messenger. Curr Opin Microbiol. 2006;9(2):218-28.

144. Krin E, Danchin A, Soutourina O. Decrypting the H-NS-dependent regulatory cascade of acid stress resistance in *Escherichia coli*. BMC Microbiol. 2010;10:273.

145. Reynolds TS, Courtney CM, Erickson KE, Wolfe LM, Chatterjee A, Nagpal P, et al. ROS mediated selection for increased NADPH availability in *Escherichia coli*. Biotechnol Bioeng. 2017;114(11):2685-9.

146. Hoegler KJ, Hecht MH. Artificial Gene Amplification in *Escherichia coli* Reveals Numerous Determinants for Resistance to Metal Toxicity. J Mol Evol. 2018;86(2):103-10.

147. Garcia-Pastor L, Sanchez-Romero MA, Jakomin M, Puerta-Fernandez E, Casadesus J. Regulation of bistability in the std fimbrial operon of *Salmonella enterica* by DNA adenine methylation and transcription factors HdfR, StdE and StdF. Nucleic Acids Res. 2019;47(15):7929-41.

148. Gama-Castro S, Jimenez-Jacinto V, Peralta-Gil M, Santos-Zavaleta A, Penaloza-Spinola MI, Contreras-Moreira B, et al. RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res. 2008;36(Database issue):D120-4.

149. Huerta AM, Collado-Vides J. Sigma$^{70}$ promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. J Mol Biol. 2003;333(2):261-78.

150. Bartoli J, My L, Belmudes L, Coute Y, Viala JP, Bouveret E. The Long Hunt for *pssR*-Looking for a Phospholipid Synthesis Transcriptional Regulator, Finding the Ribosome. J Bacteriol. 2017;199(14).

151. Brezellec P, Hoebeke M, Hiet MS, Pasek S, Ferat JL. DomainSieve: a protein domain-based screen that led to the identification of *dam*-associated genes with potential link to DNA maintenance. Bioinformatics. 2006;22(16):1935-41.

152. Lobner-Olesen A, Skovgaard O, Marinus MG. Dam methylation: coordinating cellular processes. Curr Opin Microbiol. 2005;8(2):154-60.

153. Al Mamun AA, Lombardo MJ, Shee C, Lisewski AM, Gonzalez C, Lin D, et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. Science. 2012;338(6112):1344-8.

154. Valens M, Thiel A, Boccard F. The MaoP/*maoS* Site-Specific System Organizes the Ori Region of the *E. coli* Chromosome into a Macrodomain. PLoS Genet. 2016;12(9):e1006309.

155. Espeli O, Mercier R, Boccard F. DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. Mol Microbiol. 2008;68(6):1418-27.

156. Thongbhubate K, Irie K, Sakai Y, Itoh A, Suzuki H. Improvement of putrescine production through the arginine decarboxylase pathway in *Escherichia coli* K-12. AMB Express. 2021;11(1):168.

157. Holden ER, Yasir M, Turner AK, Wain J, Charles IG, Webber MA. Massively parallel transposon mutagenesis identifies temporally essential genes for biofilm formation in *Escherichia coli*. Microb Genom. 2021;7(11).

158. Eichelberger KR, Sepulveda VE, Ford J, Selitsky SR, Mieczkowski PA, Parker JS, et al. Tn-Seq Analysis Identifies Genes Important for *Yersinia pestis* Adherence during Primary Pneumonic Plague. mSphere. 2020;5(4).

159. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. mBio. 2015;6(3):e00306-15.

160. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2:2006 0008.

161. Yamamoto N, Nakahigashi K, Nakamichi T, Yoshino M, Takai Y, Touda Y, et al. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. Mol Syst Biol. 2009;5:335.

162. Teteneva NA, Mart'yanov SV, Esteban-Lopez M, Kahnt J, Glatter T, Netrusov AI, et al. Multiple Drug-Induced Stress Responses Inhibit Formation of *Escherichia coli* Biofilms. Appl Environ Microbiol. 2020;86(21).

163. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.

164. Li GW, Burkhardt D, Gross C, Weissman JS. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell. 2014;157(3):624-35.

165. Asmar AT, Collet JF. Lpp, the Braun lipoprotein, turns 50-major achievements and remaining issues. FEMS Microbiol Lett. 2018;365(18).

166. Yakhnina AA, Bernhardt TG. The Tol-Pal system is required for peptidoglycan-cleaving enzymes to complete bacterial cell division. Proc Natl Acad Sci U S A. 2020;117(12):6777-83.

167. Cascales E, Bernadac A, Gavioli M, Lazzaroni JC, Lloubes R. Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. J Bacteriol. 2002;184(3):754-9.

168. Rai AK, Mitchell AM. Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic Molecule. mBio. 2020;11(4).

169. Makela PH, Jahkola M, Luderitz O. A new gene cluster *rfe* concerned with the biosynthesis of *Salmonella* lipopolysaccharide. J Gen Microbiol. 1970;60(1):91-106.

170. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, et al. Phenotypic landscape of a bacterial cell. Cell. 2011;144(1):143-56.

171. Mitchell AM, Srikumar T, Silhavy TJ. Cyclic Enterobacterial Common Antigen Maintains the Outer Membrane Permeability Barrier of *Escherichia coli* in a Manner Controlled by YhdP. mBio. 2018;9(4).

172. McMahon KJ, Castelli ME, Garcia Vescovi E, Feldman MF. Biogenesis of outer membrane vesicles in *Serratia marcescens* is thermoregulated and can be induced by activation of the Rcs phosphorelay system. J Bacteriol. 2012;194(12):3241-9.

173. Ørskov F, Ørskov I. Serotyping of *Escherichia coli*. In: Bergan T, editor. Methods in Microbiology. 14: Academic Press; 1984. p. 43-112.

174. Kumar SS, Ghosh AR. Assessment of bacterial viability: a comprehensive review on recent advances and challenges. Microbiology (Reading). 2019;165(6):593-610.

175. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1):D607-D13.

176. Kikuchi G, Motokawa Y, Yoshida T, Hiraga K. Glycine cleavage system: reaction mechanism, physiological significance, and hyperglycinemia. Proc Jpn Acad Ser B Phys Biol Sci. 2008;84(7):246-63.

177. Chaulk SG, Smith Frieday MN, Arthur DC, Culham DE, Edwards RA, Soo P, et al. ProQ is an RNA chaperone that controls ProP levels in *Escherichia coli*. Biochemistry. 2011;50(15):3095-106.

178. Shimada T, Kori A, Ishihama A. Involvement of the ribose operon repressor RbsR in regulation of purine nucleotide synthesis in *Escherichia coli*. FEMS Microbiol Lett. 2013;344(2):159-65.

179. Schneider BL, Reitzer L. Pathway and enzyme redundancy in putrescine catabolism in *Escherichia coli*. J Bacteriol. 2012;194(15):4080-8.

180. Amann E, Ochs B, Abel KJ. Tightly regulated *tac* promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. Gene. 1988;69(2):301-15.

181. Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A. Functional characterization in vitro of all two-component signal transduction systems from *Escherichia coli*. J Biol Chem. 2005;280(2):1448-56.

182. Chakraborty S, Kenney LJ. A New Role of OmpR in Acid and Osmotic Stress in *Salmonella* and *E. coli*. Front Microbiol. 2018;9:2656.

183. Mironov A, Seregina T, Nagornykh M, Luhachack LG, Korolkova N, Lopes LE, et al. Mechanism of H2S-mediated protection against oxidative stress in *Escherichia coli*. Proc Natl Acad Sci U S A. 2017;114(23):6022-7.

184. Panis G, Duverger Y, Courvoisier-Dezord E, Champ S, Talla E, Ansaldi M. Tight regulation of the *intS* gene of the KplE1 prophage: a new paradigm for integrase gene regulation. PLoS Genet. 2010;6(10):e1001149.

185. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, et al. Cryptic prophages help bacteria cope with adverse environments. Nat Commun. 2010;1(1):147.

186. Lilleorg S, Reier K, Pulk A, Liiv A, Tammsalu T, Peil L, et al. Bacterial ribosome heterogeneity: Changes in ribosomal protein composition during transition into stationary growth phase. Biochimie. 2019;156:169-80.

187. Erhardt H, Steimle S, Muders V, Pohl T, Walter J, Friedrich T. Disruption of individual *nuo*-genes leads to the formation of partially assembled NADH:ubiquinone oxidoreductase (complex I) in *Escherichia coli*. Biochim Biophys Acta. 2012;1817(6):863-71.

188. Theodorou EC, Theodorou MC, Kyriakidis DA. AtoSC two-component system is involved in cPHB biosynthesis through fatty acid metabolism in *E. coli*. Biochim Biophys Acta. 2011;1810(5):561-8.

189. Bassler BL. How bacteria talk to each other: regulation of gene expression by quorum sensing. Curr Opin Microbiol. 1999;2(6):582-7.

190. Laganenka L, Colin R, Sourjik V. Chemotaxis towards autoinducer 2 mediates autoaggregation in *Escherichia coli*. Nat Commun. 2016;7:12984.

191. Keseler IM, Gama-Castro S, Mackie A, Billington R, Bonavides-Martinez C, Caspi R, et al. The EcoCyc Database in 2021. Front Microbiol. 2021;12:711077.

192. Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat Methods. 2006;3(8):623-8.

193. Lee C, Kim J, Shin SG, Hwang S. Absolute and relative qPCR quantification of plasmid copy number in *Escherichia coli*. J Biotechnol. 2006;123(3):273-80.

194. del Solar G, Espinosa M. Plasmid copy number control: an ever-growing story. Mol Microbiol. 2000;37(3):492-500.

195. Thompson MG, Sedaghatian N, Barajas JF, Wehrs M, Bailey CB, Kaplan N, et al. Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. Sci Rep. 2018;8(1):1590.

196. Ericsson M, Hanstorp D, Hagberg P, Enger J, Nystrom T. Sorting out bacterial viability with optical tweezers. J Bacteriol. 2000;182(19):5551-5.

197. Subbarayan PR, Sarkar M. A comparative study of variation in codon 33 of the *rpoS* gene in *Escherichia coli* K12 stocks: implications for the synthesis of sigma$^{S}$. Mol Genet Genomics. 2004;270(6):533-8.

198. Waterman SR, Small PL. Transcriptional expression of *Escherichia coli* glutamate-dependent acid resistance genes *gadA* and *gadBC* in an *hns rpoS* mutant. J Bacteriol. 2003;185(15):4644-7.

199. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439-D44.

200. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2022;51(D1):D418-D27.

201. Oshima T, Wada C, Kawagoe Y, Ara T, Maeda M, Masuda Y, et al. Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. Mol Microbiol. 2002;45(3):673-95.

202. Lin JS, Lai EM. Protein-Protein Interactions: Co-Immunoprecipitation. Methods Mol Biol. 2017;1615:211-9.

203. Samavarchi-Tehrani P, Samson R, Gingras AC. Proximity Dependent Biotinylation: Key Enzymes and Adaptation to Proteomics Approaches. Mol Cell Proteomics. 2020;19(5):757-73.

204. Kurkela J, Fredman J, Salminen TA, Tyystjarvi T. Revealing secrets of the enigmatic ω-subunit of bacterial RNA polymerase. Mol Microbiol. 2021;115(1):1-11.

205. Geertz M, Travers A, Mehandziska S, Sobetzko P, Chandra-Janga S, Shimamoto N, et al. Structural coupling between RNA polymerase composition and DNA supercoiling in coordinating transcription: a global role for the omega subunit? mBio. 2011;2(4).

206. Orren DK, Selby CP, Hearst JE, Sancar A. Post-incision steps of nucleotide excision repair in *Escherichia coli.* Disassembly of the UvrBC-DNA complex by helicase II and DNA polymerase I. J Biol Chem. 1992;267(2):780-8.

207. Kisker C, Kuper J, Van Houten B. Prokaryotic nucleotide excision repair. Cold Spring Harb Perspect Biol. 2013;5(3):a012591.

208. Sharma RC, Smith KC. Role of DNA polymerase I in postreplication repair: a reexamination with *Escherichia coli* Δ*polA*. J Bacteriol. 1987;169(10):4559-64.

209. Mellon I, Hanawalt PC. Induction of the *Escherichia coli* lactose operon selectively increases repair of its transcribed DNA strand. Nature. 1989;342(6245):95-8.

210. Bharati BK, Gowder M, Zheng F, Alzoubi K, Svetlov V, Kamarthapu V, et al. Crucial role and mechanism of transcription-coupled DNA repair in bacteria. Nature. 2022;604(7904):152-9.

211. del Solar G, Giraldo R, Ruiz-Echevarria MJ, Espinosa M, Diaz-Orejas R. Replication and control of circular bacterial plasmids. Microbiol Mol Biol Rev. 1998;62(2):434-64.

212. Rothe M, Alpert C, Loh G, Blaut M. Novel insights into *E. coli*'s hexuronate metabolism: KduI facilitates the conversion of galacturonate and glucuronate under osmotic stress conditions. PLoS One. 2013;8(2):e56906.

213. Espeli O, Borne R, Dupaigne P, Thiel A, Gigant E, Mercier R, et al. A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. EMBO J. 2012;31(14):3198-211.

214. Baltekin O, Boucharin A, Tano E, Andersson DI, Elf J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. Proc Natl Acad Sci U S A. 2017;114(34):9170-5.

215. Sadhir I, Murray SM. Mid-cell migration of the chromosomal terminus is coupled to origin segregation in *Escherichia coli*. bioRxiv. 2023:2023.03.24.534095.

216. Su MY, Som N, Wu CY, Su SC, Kuo YT, Ke LC, et al. Structural basis of adaptor-mediated protein degradation by the tail-specific PDZ-protease Prc. Nat Commun. 2017;8(1):1516.

217. Artus NN, Somerville SC, Somerville CR, Lorimer GH. The biochemistry and cell biology of photorespiration. Critical Reviews in Plant Sciences. 1986;4(2):121-47.

218. Hoover-Fong JE, Shah S, Van Hove JL, Applegarth D, Toone J, Hamosh A. Natural history of nonketotic hyperglycinemia in 65 patients. Neurology. 2004;63(10):1847-53.

219. Miyashita T, Kano Y, Kuroki K, Ishii S, Imamoto F. In vivo evidence for *nusA* and *nusB* gene function in general transcription of the *Escherichia coli* genome. Biken J. 1982;25(3):121-30.

220. Matsumoto Y, Shigesada K, Hirano M, Imai M. Autogenous regulation of the gene for transcription termination factor rho in *Escherichia coli:* localization and function of its attenuators. J Bacteriol. 1986;166(3):945-58.

221. Sedlyarova N, Shamovsky I, Bharati BK, Epshtein V, Chen J, Gottesman S, et al. sRNA-Mediated Control of Transcription Termination in *E. coli*. Cell. 2016;167(1):111-21 e13.

222. Guhathakurta A, Summers D. Involvement of ArgR and PepA in the pairing of ColE1 dimer resolution sites. Microbiology (Reading). 1995;141 ( Pt 5):1163-71.

223. Bogosian G, Bourneuf EV. A matter of bacterial life and death. EMBO Rep. 2001;2(9):770-4.

224. Braissant O, Astasov-Frauenhoffer M, Waltimo T, Bonkat G. A Review of Methods to Determine Viability, Vitality, and Metabolic Rates in Microbiology. Front Microbiol. 2020;11:547458.

225. Hahn MW, Lunsdorf H, Wu Q, Schauer M, Hofle MG, Boenigk J, et al. Isolation of novel ultramicrobacteria classified as *Actinobacteria* from five freshwater habitats in Europe and Asia. Appl Environ Microbiol. 2003;69(3):1442-51.

226. Azevedo NF, Pacheco AP, Keevil CW, Vieira MJ. Nutrient shock and incubation atmosphere influence recovery of culturable *Helicobacter pylori* from water. Appl Environ Microbiol. 2004;70(1):490-3.

227. Azevedo NF, Braganca SM, Simoes LC, Cerqueira L, Almeida C, Keevil CW, et al. Proposal for a method to estimate nutrient shock effects in bacteria. BMC Res Notes. 2012;5:422.

228. Hahn MW, Stadler P, Wu QL, Pockl M. The filtration-acclimatization method for isolation of an important fraction of the not readily cultivable bacteria. J Microbiol Methods. 2004;57(3):379-90.

229. Chib S, Ali F, Seshasayee ASN. Genomewide Mutational Diversity in *Escherichia coli* Population Evolving in Prolonged Stationary Phase. mSphere. 2017;2(3).

230. Bhardwaj N, Syal K, Chatterji D. The role of omega-subunit of *Escherichia coli* RNA polymerase in stress response. Genes Cells. 2018;23(5):357-69.

231. Yamamoto K, Yamanaka Y, Shimada T, Sarkar P, Yoshida M, Bhardwaj N, et al. Altered Distribution of RNA Polymerase Lacking the Omega Subunit within the Prophages along the *Escherichia coli* K-12 Genome. mSystems. 2018;3(1).

232. Shlomai J, Polder L, Arai K, Kornberg A. Replication of phi X174 dna with purified enzymes. I. Conversion of viral DNA to a supercoiled, biologically active duplex. J Biol Chem. 1981;256(10):5233-8.

233. Makiela-Dzbenska K, Jaszczur M, Banach-Orlowska M, Jonczyk P, Schaaper RM, Fijalkowska IJ. Role of *Escherichia coli* DNA polymerase I in chromosomal DNA replication fidelity. Mol Microbiol. 2009;74(5):1114-27.

234. Okazaki R, Arisawa M, Sugino A. Slow joining of newly replicated DNA chains in DNA polymerase I-deficient *Escherichia coli* mutants. Proc Natl Acad Sci U S A. 1971;68(12):2954-7.

235. Joyce CM, Grindley ND. Method for determining whether a gene of *Escherichia coli* is essential: application to the *polA* gene. J Bacteriol. 1984;158(2):636-43.

236. Li Y, Austin S. The P1 plasmid in action: time-lapse photomicroscopy reveals some unexpected aspects of plasmid partition. Plasmid. 2002;48(3):174-8.

237. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151:165-88.

238. Jensen SI, Lennen RM, Herrgard MJ, Nielsen AT. Seven gene deletions in seven days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress. Sci Rep. 2015;5:17874.

239. Bellotto N, Agudo-Canalejo J, Colin R, Golestanian R, Malengo G, Sourjik V. Dependence of diffusion in *Escherichia coli* cytoplasm on protein size, environmental conditions, and cell growth. Elife. 2022;11.

240. Guzman LM, Belin D, Carson MJ, Beckwith J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose pBAD promoter. J Bacteriol. 1995;177(14):4121-30.

241. Branon TC, Bosch JA, Sanchez AD, Udeshi ND, Svinkina T, Carr SA, et al. Efficient proximity labeling in living cells and organisms with TurboID. Nat Biotechnol. 2018;36(9):880-7.

242. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97(12):6640-5.

243. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods. 2009;6(5):343-U41.

244. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25(11):1422-3.

245. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

246. Mazia D, Schatten G, Sale W. Adhesion of cells to surfaces coated with polylysine. Applications to electron microscopy. J Cell Biol. 1975;66(1):198-200.

247. Skinner SO, Sepulveda LA, Xu H, Golding I. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. Nat Protoc. 2013;8(6):1100-13.

248. Köhler R, Kaganovitch E, Murray SM. High-throughput imaging and quantitative analysis uncovers the nature of plasmid positioning by ParABS. Elife. 2022;11.

249. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012;9(7):676-82.

250. Skulj M, Okrslar V, Jalen S, Jevsevar S, Slanc P, Strukelj B, et al. Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. Microb Cell Fact. 2008;7:6.

251. O'Toole GA. Microtiter dish biofilm formation assay. J Vis Exp. 2011(47).

252. Ludwig N, Reissmann S, Schipper K, Gonzalez C, Assmann D, Glatter T, et al. A cell surface-exposed protein complex with an essential virulence function in *Ustilago maydis*. Nat Microbiol. 2021;6(6):722-30.

253. Osorio-Valeriano M, Altegoer F, Das CK, Steinchen W, Panis G, Connolley L, et al. The CTPase activity of ParB determines the size and dynamics of prokaryotic DNA partition complexes. Mol Cell. 2021;81(19):3992-4007 e10.

254. Bonocora RP, Wade JT. ChIP-seq for genome-scale analysis of bacterial DNA-binding proteins. Methods Mol Biol. 2015;1276:327-40.

255. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

256. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).

257. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

258. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

259. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

260. Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria2022.

261. Wickham H. *ggplot2*: Elegant Graphics for Data Analysis: Springer-Verlag New York; 2016.

262. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the *Tidyverse*. Journal of Open Source Software. 2019;4(43).

263. Hughes SM. plater: Read, Tidy, and Display Data from Microtiter Plates. The Journal of Open Source Software. 2016;1(7).

## *Erklärung*

Hiermit versichere ich, dass ich die vorliegende Dissertation mit dem Titel

“HdfR/MaoP pathway of *Escherichia coli* regulates flagella expression and growth in aquatic environment”

selbstständig verfasst, keine anderen als die im Text angegebenen Hilfsmittel verwendet und sämtliche Stellen, die dem Wortlaut oder dem Sinn nach aus anderen Werken entnommen sind, mit Quellenangaben kenntlich gemacht habe. Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinem anderen Prüfungszweck gedient.

Marburg, den 12.07.2023

Nataliya Teteneva