



Skolkovo Institute of Science and Technology

ESCAPE MECHANISMS OF MOBILE GENETIC ELEMENTS AGAINST CRISPR-CAS SYSTEM
AND DIVERSITY IN MICROBIAL COMMUNITIES

Doctoral Thesis

by

VIKTOR MAMONTOV

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor
Professor Konstantin Severinov

Moscow - 2023

© Viktor Mamontov 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Viktor Mamontov (Candidate)

Prof. Konstantin Severinov (Supervisor)

Abstract

The arms race between prokaryotes and MGEs has driven the evolution of new strategies for prokaryotes to prevent phage infection and for MGEs to overcome these defense mechanisms. The CRISPR-Cas system carries out an adaptive RNA-guided immune response against foreign nucleic acids in prokaryotes. MGEs possess several strategies to evade CRISPR interference, a process for preventing the proliferation of foreign nucleic acids in a host cell. Numerous escape mechanisms including escape mutations and Acr have been described. Using classical microbiological experiments, fluorescent microscopy, and microfluidics in combination with mathematical modeling, we described an additional mechanism based on the dynamics of plasmid replication. The mechanism allows the plasmids targeted by CRISPR interference to persist in the *Escherichia coli* population for an extended period without genetic modifications.

MGEs often bear antibiotic resistance genes and facilitate the spread of antimicrobial resistance among bacteria. Marine microbiota emerges as a promising reservoir for novel antimicrobial drugs. At the same time, a comprehensive understanding of the true composition of microbial communities is essential for fundamental marine ecological and microbiological studies. We assessed a set of popular DNA extraction kits for 16S rRNA amplicon sequencing. To evaluate the efficacy of the kits we utilized metagenomes sourced from soil, water, and gut of a marine invertebrate *Magallana gigas*. We ranked the kits based on contamination level, alpha diversity, the ratio of 18S/16S rRNA after extraction, and the presence of PCR inhibitors.

While microbes tend to defend themselves against MGEs, the presence of MGEs in the microbial population supports genetic diversity and can provide advantageous traits. We investigated a set of *E. coli* strains isolated from healthy farm animals using the long-read sequencing technology and subsequent *in silico* analysis of genome assemblies. We identified two *E. coli* strains with probiotic features that harbored several plasmids with bacteriocin gene clusters. These bacteriocins inhibit the growth of pathogenic strains *in vitro* and *in vivo*.

In conclusion, the results presented in this thesis extend our understanding of the interaction between MGEs and microbes and introduce novel methodologies for studying these interactions.

Publications

1. **Mamontov, V.**, Martynov, A., Morozova, N., Bukatin, A., Staroverov, D. B., Lukyanov, K. A., Ispolatov, Y., Semenova, E., & Severinov, K. (2022). Persistence of plasmids targeted by CRISPR interference in bacterial populations. *Proceedings of the National Academy of Sciences*, 119(15). DOI: <https://doi.org/10.1073/pnas.2114905119>.
2. Mihailovskaya, V., Sutormin, D., Karipova, M., Trofimova, A., **Mamontov, V.**, Severinov, K., Kuznetsova, M. (2023). Bacteriocin-Producing *Escherichia coli* Q5 and C41 with Potential Probiotic Properties: *In Silico*, *In Vitro*, and *In Vivo* Studies. *International Journal of Molecular Sciences*. 2023; 24(16):12636. DOI: <https://doi.org/10.3390/ijms241612636>.
3. Demkina, A., Slonova, D., **Mamontov, V.**, Konovalova, O, Yurikova, D, Rogozhin, D., Belova, V., Korostin, D., Sutormin, D., Severinov, K., & Isaev, A. (2023). Benchmarking DNA isolation methods for marine metagenomics. *Sci Rep* 13, 22138 2023. <https://doi.org/10.1038/s41598-023-48804-z>.

Conferences

1. The stochastic nature of the interaction between CRISPR interference and replication events provides the coexistence of virus and bacteria populations. **V. Mamontov**, Y. Ispolatov, E. Semenova, K. Severinov. Skolkovo Institute of Science and Technology, MCCMB 2021, July 30 – August 2, 2021. Poster presentation.
2. Stochastic molecular processes create heterogeneity in the population of genetically identical *E. coli* cells mounting type I-E CRISPR-Cas interference. **V. Mamontov**, A. Martynov, N. Morozova, A. Bukatin, D.B. Staroverov, K.A. Lukyanov, Y. Ispolatov, E. Semenova, K. Severinov. Institute Pasteur, CRISPR 2021, June 1-10, 2021. Poster presentation.

Acknowledgements

My PhD journey has been an enlightening experience, and I owe a debt of gratitude to many who have supported me along the way.

I am profoundly thankful to my supervisor, Prof. Konstantin Severinov, whose guidance, patience, and expertise have been instrumental in shaping my research. Your unwavering belief in my potential has been a driving force behind my endeavors. I extend my heartfelt appreciation to my committee members Dr. Yaroslav Ispalotov and Dr. Konstantin Lukyanov for their constructive feedback, encouragement and invaluable contribution in my research project. I am grateful to Ekaterina Semenova for providing her inestimable insights into planning experimental work. I would like to express my deepest gratitude to Alexander Martynov, Natalia Morozova, Anton Bukatin and Dmitry Staroverov for their irreplaceable contribution and help in the execution of my PhD research project. I appreciate Aleksandra Strotskaya for providing all the necessary plasmids for the experimental work. I am also deeply grateful to my co-authors for their invaluable contributions and collaboration.

I sincerely appreciate the dedicated members of the Laboratory for the Regulation of gene expression of prokaryotic mobile elements Evgeniy Klimuk and Olga Musharova for their companionship, shared insights, and unwavering support. Your feedback and friendship made this journey enjoyable and fulfilling. I would like to specially acknowledge Ekaterina Savitskaya, who unfortunately is no longer with us, for her invaluable assistance and for imparting her knowledge at the initial stages of my academic journey. I want to express my gratitude to Anna Trofimova for preparing DNA libraries for long-read sequencing. I am deeply appreciative of my colleagues and co-authors Dmitry Sutormin and Artem Isaev. Working with you has been a delightful experience, and I hope to continue our productive collaboration in future studies.

Above all, my heartfelt gratitude is for my friends and family, who have believed in me, even in moments of self-doubt. This achievement is borne not only of my efforts but also of the collective belief and support of each one of you. My deepest thanks to all.

Table of Contents

Abstract.....	3
Publications	4
Conferences	4
Acknowledgements	5
Table of Contents.....	6
Abbreviations.....	10
Preface	12
Chapter 1: Introduction.....	14
1.1 Interaction of mobile genetic elements and prokaryotes	14
1.2 Plasmids and transposons	17
1.3 Diversity of CRISPR-Cas systems: a short introduction	20
1.3.1 Class 1 CRISPR-Cas systems	22
1.3.2 Class 2 CRISPR-Cas systems	25
1.3.3 CRISPR-interference on the example of type I-E CRISPR-Cas system	27
1.3.4 Escape mechanisms of mobile genetic elements against CRISPR-Cas systems	28
1.4 Role of MGEs in antibiotic resistance of microbes	30
1.5 Studies of diversity in microbial communities	31
Chapter 2: Persistence of plasmids targeted by CRISPR interference in bacterial populations.....	33
2.1 Introduction.....	33
2.2 Methods.....	34
2.2.1 Strains and plasmids	34
2.2.2 CRISPR interference assay	35
2.2.3 Quantitative PCR assay of plasmids	36
2.2.4 Replating of transformants.....	36
2.2.5 Antibiotic-free cultivation of CRISPR ON cells and the growth curves	37
2.2.6 Flow cytometry	38
2.2.7 Microscopy assay.....	38
2.2.8 Single cell microscopy in microfluidic device.....	38
2.2.9 Dynamics of replication and degradation of plasmids.....	39
2.2.10 Redistribution of plasmids during cell division	40

2.2.11 Simulation procedure	41
2.2.12 Transcriptome analysis (<i>for unpublished results</i>).....	42
2.2.13 Analysis of escape phages M13 (<i>for unpublished results</i>)	43
2.3 Results.....	44
2.3.1 Colonies formed after transformation of protospacer plasmids into cells mounting CRISPR interference contain cells with active CRISPR-Cas and unchanged plasmids ..	44
2.3.2 Cells from CRISPR ON colonies contain fewer plasmids than CRISPR OFF colony cells	47
2.3.3 Direct observation of plasmid-bearing cells in CRISPR ON colonies	50
2.3.4 Direct real time observation of plasmid loss due to CRISPR interference.....	53
2.3.5 Discussion and theoretical analysis	56
2.3.6 Transcriptome assay of CRISPR ON cells and the nature of acquisition escape mutation by M13 phages against CRISPR-Cas system (<i>unpublished results</i>)	62
Chapter 3: Benchmarking DNA Isolation Methods for Marine Metagenomics.	66
3.1 Introduction.....	66
3.2 Methods.....	71
3.2.1 DNA purification kits	71
3.2.2 Sample collection.....	72
3.2.3 DNA extraction.....	73
3.2.4 Additional purification procedures	74
3.2.5 DNA quantification and quality assessment.....	75
3.2.6 PCR Amplification.....	76
3.2.7 Quantitative real-time PCR.....	76
3.2.8 16S rRNA libraries sequencing	77
3.2.9 16S data analysis.....	77
3.2.10 Shotgun sequencing	78
3.3 Results.....	78
3.3.1 Sample collection and processing.....	78
3.3.2 DNA yield.....	79
3.3.3 DNA fragmentation	80
3.3.4 DNA purity	81
3.3.5 Presence of PCR inhibitors	82
3.3.6 Admixture of eukaryotic DNA	84

3.3.7 Contamination levels and determination of “kitomes”	87
3.3.8 Technical reproducibility of DNA extraction kits	89
3.3.9 Alpha diversity of microbial communities	92
3.3.10 Effects of DNA extraction kits on the composition of bacterial communities	92
3.4 Discussion	95
Chapter 4: Bacteriocin-Producing <i>Escherichia coli</i> Q5 and C41 with Potential Probiotic Properties: <i>In Silico</i> , <i>In Vitro</i> , and <i>In Vivo</i> Studies.....	99
4.1 Introduction.....	99
4.2 Results.....	100
4.2.1 <i>In silico</i> Analysis of <i>E. coli</i> Q5 and C41 Genomes	100
4.2.1.1 General Genome Features.....	100
4.2.1.2 Antimicrobial Resistance and Virulence-Associated Genes.....	102
4.2.1.3 Adhesion-Related Genes.....	102
4.2.1.4 Bacteriocin Gene Clusters.....	103
4.2.2 <i>In vitro</i> Analysis of <i>E. coli</i> Q5 and C41 Potential as Possible Probiotics	104
4.2.2.1 Antimicrobial Activity in Spent Media.....	104
4.2.2.2 Adhesion Ability.....	105
4.2.2.3 Antimicrobial Susceptibility and Lysogeny.....	106
4.2.3 <i>In vivo</i> Analysis of Probiotic Properties of <i>E. coli</i> Q5 and C41	106
4.2.3.1 The Effect of <i>E. coli</i> Q5 and C41 on the Physiological Parameters of Rats.....	106
4.2.3.2 Composition of Rat Intestinal Microbiota after Administration of <i>E. coli</i> Q5 and C41 and upon Experimental Infection with Toxigenic <i>E. coli</i> C55 after Preliminary Administration of <i>E. coli</i> Q5 and C41	107
4.2.3.3 Hematological and Biochemical Parameters of Rats after Administration of <i>E. coli</i> Q5 and C41 and during Experimental Infection with Toxigenic <i>E. coli</i> C55 after Preliminary Administration of <i>E. coli</i> Q5 and C41	109
4.2.3.4 Histological Analysis of Small Intestine, Peyer’s Patches, Spleen, and Liver Morphology of Rats in Experimental Infection with Toxigenic <i>E. coli</i> C55 after Preliminary Administration of the <i>E. coli</i> Q5 and C41 Strains	111
4.3 Discussion	114
4.4 Materials and Methods.....	117
4.4.1 Bacterial Strains	117
4.4.2 Genome Sequencing and Assembly.....	118
4.4.3 Genome Annotation and Analysis	119

4.4.4 Data Deposition	119
4.4.5 Antimicrobial Activity of Cell-Free Supernatants of <i>E. coli</i> Strains	120
4.4.6 Nonspecific Adhesion of <i>E. coli</i> Strains	120
4.4.7 Specific Adhesion of <i>E. coli</i> Strains	121
4.4.8 Antimicrobial Susceptibility	121
4.4.9 Bacteriophage Induction	122
4.4.10 Probiotic and Pathogenic Inocula Preparation	122
4.4.11 Experimental Design <i>In vivo</i>	122
4.4.12 Analysis of the Composition of the Intestinal Microbiota	124
4.4.13 Hematological and Biochemical Blood Analysis	125
4.4.14 Histologic Analysis	125
4.4.15 Statistical Analysis	125
4.5 Conclusions	125
Chapter 5: Conclusions	127
Bibliography	133
Supplementary Information	154
For Chapter 2	154
For Chapter 3	163
For Chapter 4	174

Abbreviations

A, T, G, C – adenine, cytosine, guanine, thymine, respectively
AAI – average adhesion index
Abi – abortive infection
AC – adhesion coefficient
Acr – anti-CRISPR protein
ALP – alkaline phosphatase
ALT – alanine aminotransferase
AMR – antimicrobial resistance
ARGs – antibiotic resistance genes
ASV – amplicon sequence variant
ATP – adenosine triphosphate
BGCs – biosynthetic gene clusters
bp – base pairs
BREX – bacteriophage exclusion
CDT – cytolethal distending toxin
CFU – colony forming units
CRISPR – clustered regularly interspaced short palindromic repeats
Cas – CRISPR associated
crRNA – CRISPR-RNA
CTD – C-terminal domain
DGRs – diversity-generating retroelements
DIN – DNA integrity
DNA – deoxyribonucleic acid
DNase – deoxyribonuclease
dNTP – deoxynucleotide
dsDNA – double-stranded DNA
EOT – efficiency of transformation
EPS – extracellular polymeric substance
GFP – green fluorescent protein
HEPN – higher eukaryotes and prokaryotes nucleotide-binding
HGT – horizontal gene transfer
Ht – hematocrit
IAM – index adhesiveness of microorganisms
IPTG – isopropyl- β -D-thiogalactopyranosid
IS – insertion sequence
LB – Luria-Bertani
MCH – mean corpuscular hemoglobin
MCHC – mean corpuscular hemoglobin concentration
MCV – mean corpuscular volume
MDR – multidrug resistance
MGE – mobile genetic elements

ml – milliliter
mM – millimole
ng – nanogram
NGS – next-generation sequencing
NMDS – non-metric multidimensional scaling
NTS – non-target-strand
OTU – operational taxonomic unit
PAM – protospacer adjacent motif
PAPS – 3'-phosphoadenosine-5'-phosphosulfate
PBS – phosphate-buffered saline
PCN – plasmid copy number
PCR – polymerase chain reaction
PDMS – polydimethylsiloxane
Pgl – phage growth limitation
PLT – platelet count
pre-crRNA – precursor CRISPR RNA
qPCR – quantitative PCR
RBC – red blood cells
RiPP – post-translationally modified peptide
RM – restriction-modification
RNA – ribonucleic acid
RNase – ribonuclease
rRNA – ribosomal RNA
RTn – retrotransposon
ssDNA – single stranded DNA
TA – toxin-antitoxin
Tase – transposase
TEs – Transposable Elements
Tn – transposon
tracrRNA – trans-activating CRISPR RNA
TS – target-strand
UV – ultraviolet
VAGs – virulence-associated genes
WBC – white blood cell count
 μg – microgram
 μl – microliter
 μm – micrometer

Preface

The research results in this thesis were conducted at the Severinov Laboratory and the Skoltech Laboratory of Metagenome Analysis as a part of my Ph.D. research project in the Life Sciences program at the Skolkovo Institute of Science and Technology. Much of the work was done in collaboration with other scientists, as noticed in the Acknowledgements section. The research chapters of this thesis are based on content from my first-authored publication or publications to which I have made significant contributions. Details of these publications are provided in the Publications section. The Introduction and Conclusion chapters are my original work, authored exclusively by me. Below, I detail my contributions to the research process for each chapter.

Chapter 2

Chapter 2 is based on the following publication:

Mamontov, V., Martynov, A., Morozova, N., Bukatin, A., Staroverov, D. B., Lukyanov, K. A., Ispolatov, Y., Semenova, E., & Severinov, K. (2022). Persistence of plasmids targeted by CRISPR interference in bacterial populations. *Proceedings of the National Academy of Sciences*, 119(15).

I was involved in all stages of this research including design of research, experiments, and analysis of the results. The main part of the experiments was carried out solely by me. Co-author Alexander Martynov conducted the mathematical modeling. I have also incorporated unpublished results in Chapter 2 to enhance the findings from the original publication.

Chapter 3

Chapter 3 is based on the following publication:

Demkina, A., Slonova, D., **Mamontov, V.**, Konovalova, O, Yurikova, D, Rogozhin, D., Belova, V., Korostin, D., Sutormin, D., Severinov, K., & Isaev, A. (2023). Benchmarking DNA isolation methods for marine metagenomics. *Sci Rep* 13, 22138 2023. <https://doi.org/10.1038/s41598-023-48804-z>.

In this work, I was responsible for the analysis of 16S rRNA sequencing data and for the design of subsequent bioinformatic and statistical analyses of these sequencing results. I played a significant

role in obtaining the main results, which are illustrated in Figures 3.7, 3.8, 3.9, and 3.10, and have included these findings in this chapter.

Chapter 4

Chapter 4 is based on the following publication:

Mihailovskaya, V., Sutormin, D., Karipova, M., Trofimova, A., **Mamontov, V.**, Severinov, K., Kuznetsova, M. Bacteriocin-Producing *Escherichia coli* Q5 and C41 with Potential Probiotic Properties: *In Silico*, *In Vitro*, and *In Vivo* Studies. *International Journal of Molecular Sciences*. 2023; 24(16):12636. DOI: <https://doi.org/10.3390/ijms241612636>. Supporting information is available at: <https://www.mdpi.com/article/10.3390/ijms241612636/s1>.

For this chapter, I was responsible for computational analysis, including installing the necessary software and configuring the computational server for Nanopore sequencing. I processed the long-read sequencing data for *de novo* assembly and developed a comprehensive pipeline for assembling and annotating the sequencing data. Additionally, I made substantial contributions to the *in silico* analysis of the sequenced strains detailed in this publication.

Chapter 1: Introduction

Chapter 1 was written solely by me.

1.1 Interaction of mobile genetic elements and prokaryotes

The coexistence of mobile genetic elements (MGEs) and microbes has ancient origins and their interactions are multifaceted [1, 2]. MGEs enrich genetic diversity within microbial communities and are a driving force behind genetic changes in prokaryotic cells [3]. In general, mobile genetic elements include viruses, plasmids, and transposons. This review focuses exclusively on viruses of prokaryotes, known as bacteriophages or phages. Phages are the most abundant living organisms on the Earth, with an estimated population of up to 10^{31} viral particles that exceed microbes by an order of magnitude [4, 5]. Phages inhabit all environments where microbes reside, including oceans, soil, hot springs, and symbiotic microbial communities associated with humans, animals, or plants [6]. Due to their vast population and ubiquity, phages are characterized by significant genetic diversity, forms, and infection strategies. As intracellular parasites of microbes, phages require bacterial or archaeal cells to reproduce. Each bacterial strain is believed to be infected by at least one type of phage, often more [7], subjecting bacteria to constant phage attacks. Generally, phages are categorized by their infection type – lytic, temperate, and chronic; though the actual mechanisms of infection are more diverse and complex [8, 9]. The lytic phages are characterized by an aggressive infection strategy with a short period between the injection of their genetic material into microbial cells and the release of new phage particles by lysing the bacterial cell [10]. In contrast, temperate phages can integrate their genetic material into the host's chromosome as a prophage and activate the prophage state in harmful conditions, producing new virions (phage particles) [8]. Owing to continual phage infections of microbes, phages act as significant drivers of microbial evolution. Temperate phages promote genetic diversity in microbial communities and facilitate horizontal gene transfer (HGT) by capturing and transferring parts of the host genome DNA to other strains [11, 12]. Lytic phages permanently drive microbes to evolve and refine their defense strategies to avoid elimination [13]. Numerous microbial defense systems and molecular mechanisms appear to overcome phage infection during the ongoing arms race. The evolutionary arms race between phages and prokaryotes has led to the accumulation of defense islands within microbial genomes, comprising genes that confer resistance to

phage infection [14-16]. Thus, the interaction of MGEs and prokaryotes can be considered from the viewpoint of the permanent arms race: improving prokaryotic defense mechanisms and overcoming these defenses by MGEs. A short review of the most studied defense systems and mechanisms is followed in this chapter.

To entrench a bacterial cell and inject genetic material into the cell, phages recognize specific molecules on the bacterium's surface, known as receptors [17-18]. After adsorbing onto the cell surface, the phage injects its genetic material into the microbial cell [17-18]. Bacterial phenotypes with modified receptors can have a selective advantage over those with normal receptors, a process known as "surface modification," which prevents phage attachment [19]. However, the surface modification can only prevent infection by some phages, as not all surface receptors can be altered or masked. Moreover, phages can evolve to recognize modified receptors [20]. To defend against the introduction of phage genetic material, bacteria employ various defense systems. One of the earliest discovered systems is restriction-modification (RM) systems [21, 22]. There are several types of RM systems, but all of them consist of two main proteins: restrictase and methylase [23, 24]. The restrictase provides cleavage of specific DNA sites termed restriction sites, while the methylase adds methyl group on the restriction sites. The methylation of restriction sites prevents cleavage by the restriction enzyme, thus protecting the host DNA. As foreign DNA lacks these modifications, it is vulnerable to cleavage by the restriction enzyme [23, 24].

Phage growth limitation (Pgl) systems offer defense mechanisms that contrast with RM systems by being based on the modification of phage genomes [25, 26]. The Pgl system comprises genes encoding four proteins: adenine-methyltransferase PglX, ATPase PglY, protein kinase PglW, and alkaline phosphatase PglZ [26, 27]. The mechanism of the Pgl system from *Streptomyces coelicolor* has been experimentally described *in vitro* [26]. PglX and PglZ proteins form the toxin and antitoxin pair, correspondingly. The Pgl⁺ cell sacrifices itself by modifying the phage genome that enables other Pgl⁺ cells to recognize and defend against the modified phage in subsequent infection cycles [26]. The primary advantage of the Pgl system is its capacity to prevent phages from mutating to bypass the genomic modification, thereby securing survival for Pgl⁺ cells within the bacterial community [26].

The BREX (bacteriophage exclusion) system is a relatively recently discovered system including a *pglZ*-like gene, ATPase, and a SAM-dependent methyltransferase [14, 29]. The type IV BREX system, instead of the methyltransferase, features a PAPS reductase that facilitates DNA phosphorothioation [30]. The defense mechanism of the BREX system is based on epigenetic DNA modification, which, while analogous to RM systems in modifying the host DNA, targets unmodified phage DNA by different means [31, 32]. BREX methyltransferase modifies specific DNA sites only at one strand but the exact restriction mechanism remains to be elucidated. Phages have evolved ways to circumvent the BREX defense, such as employing DNA mimic proteins, exemplified by the phage T7 protein Ocr [29].

Abortive infection (Abi) is a defense mechanism that limits the spread of phage infections by altering cellular metabolism. This defense response either arrests host metabolic processes or leads to cell death, thereby reducing the production of new phage particles [33]. Unlike active defense systems such as CRISPR-Cas or RM systems, which act early in infection without destroying the host cell, the Abi response is triggered in the middle or late stages of a phage infection cycle [34, 35]. To limit phage propagation, the Abi system employs a sensor module that detects phage infection and an effector module that disrupts cellular processes [36]. The biological rationale for the self-destruction of infected cells is to ensure the survival of the remaining bacterial population [37]. Although Abi systems act at the population level, they are primarily found in prophages and plasmids. The Abi system in *E. coli*, known as Rex system, is encoded by lambda prophage region [34, 38]. The Rex system limits plaque formation of T4, T5, and T7 phages [34, 39]. The multifarious mechanisms of Abi systems were described in the lactic acid bacteria [40]. These mechanisms include the limitation of DNA replication by proteins such as AbiA, AbiK, and AbiF [41], the degradation of mRNA by AbiB and AbiQ [42, 43], the interaction with the phage holin and lysin for cell lysis by AbiZ system [44]. While the exact sensor mechanisms remain unclear, the *E. coli* RexA protein is thought to detect protein-DNA complexes indicative of phage replication [45]. Another *E. coli* Abi system PifA, encoded on the F plasmid, is activated by the T7 capsid protein gp10 [46]. The primary localization of Abi systems in MGEs provides their relatively effortless spread among bacterial communities through HGT [47]. Thus, Abi systems not only provide the defense at the bacterial population level but also limit the spread of competing MGEs.

Toxin-antitoxin (TA) systems are widespread in prokaryotes. TA systems are two-gene modules including a stable toxin molecule and a less stable antitoxin. Under normal conditions, antitoxin inhibits toxin activity, but during harmful conditions, the balance between toxin and antitoxin molecules is disrupted, leading to changes in cellular processes [48, 49]. Many conjugative plasmids harbor TA systems that induce dependency in the host cell on the plasmid's presence [47, 50]. The prevalent diversity of toxins is proteins while antitoxins can be either RNAs or proteins that suppress toxin activity [51-53]. TA systems are classified based on the mechanism of cooperation toxin and antitoxin components. Toxins are primarily proteins in type I to type VII TA systems, except for type VIII TA systems where a small RNA acts as the toxin [54]. Antitoxins can be small noncoding RNAs in type I, type III, and type VIII TA, or proteins in type II, type IV, type V, type VI, and type VII TA. The small noncoding RNA in type I TA inhibits the transcript of the cognate toxin [55]. In type II systems, antitoxin forms a complex with the toxin, neutralizing its toxicity [56]. The first time the TA module was discovered on F conjugative plasmid in *E. coli* was type II *ccdAB*, and it was responsible for plasmid maintenance [57]. In type III systems, the RNA antitoxin binds to the toxin protein [58]. Type IV TA systems feature toxin and antitoxin proteins that compete to bind the same cellular target [59]. In type V systems, the antitoxin protein selectively degrades the mRNA of the toxin gene [60]. The TA system is widespread in microbial communities and can provide Abi response against phage infection [58, 61, 62]. It can be expected that conjugative plasmids have acquired TA modules to maintain their persistence in host cells over multiple divisions and to compete with other MGEs within a prokaryotic population.

The central part of this thesis delves into the interaction between plasmids and the CRISPR-Cas system and studies of natural microbial communities. Therefore, in the following sections, I emphasize the description of CRISPR-Cas systems and the role of MGEs in the ecology of microbes.

1.2 Plasmids and transposons

Plasmids are mobile self-replicating dsDNA elements and are typically transmitted in vertical gene transfer, from parent cell to offspring by simple binary fusion [65]. However, prokaryotes can also exchange plasmids, particularly conjugative plasmids, through a process known as conjugation [66], or

capture extracellular DNA via transformation [67, 68]. Conjugative plasmids, which can be up to 200 kb, are extra-chromosomal genetic elements that can carry genes for both the conjugation mechanism and antimicrobial resistance [69, 70]. Conjugative plasmids significantly enhance the genetic diversity of the bacteria population. In *Bacteroides fragilis*, for example, conjugative plasmids can replace the host type VI secretion systems (T6SS) with a more effective antagonistic secretion system thereby enhancing the host's fitness in intraspecies competition [71]. Recent studies have also illustrated that conjugative play a crucial role in the spread of antibiotic resistance among clinical strains [72, 73].

Transposable elements (TEs), also termed transposons (Tns), are another group of MGEs. The first transposon was discovered by Barbara McClintock in the 1940s [74, 75]. The main hallmark of Tns is the transposase (Tase) enzyme, which facilitates their transposition, the process of moving the transposon from one location to another within the same or different DNA molecules [76]. Depending on the mechanism of transposition Tns are classified into two main groups: retrotransposons (class I) and DNA transposons (class II). Class I Tns are primarily found in eukaryotes [77] while class II Tns are widespread in both eukaryotes and prokaryotes [77, 78]. The transposition of Tns often leads to mutations in the bacteria genome. Tns also can include antibiotic resistance genes. Conjugative Tns facilitate the transmission of resistance genes among bacterial communities. It was demonstrated that the conjugative Tn916 family is responsible for bacterial antibiotic resistance in *Enterococcus* species [79, 80].

Class I Tns are also called retroelements (RTns). The reverse transcriptase enzyme conducts the transposition of RTns [81, 82]. The reverse transcriptase carries out reverse transcription from an RNA template, synthesizing cDNA that is then inserted at a new locus. RTns are divided into autonomous and non-autonomous RTns groups. The main difference is that autonomous RTns use their own reverse transcriptase enzyme while non-autonomous retroelements borrow the proteins from the first [83]. Another class of retroelements, the diversity-generating retroelements (DGRs) were found in proteobacteria, cyanobacteria and archaea [84]. DGRs have not shown mobility but have been implicated in introducing genetic variations into target genes through an error-prone reverse transcriptase mechanism [85]. Phages can use DGRs to modify their genes responsible for recognition of new bacterial receptors [86]. Bacterial

TEs can be involved in antimicrobial resistance; for example, Tns have been found in *Salmonella* genome islands locus with antibiotic resistance genes [87, 88].

Class II Tns, or DNA Tns, contain the Tase enzyme and inverted repeats (IRs) [89]. The IRs serve as sites for recognition by the transposase. The transposition of DNA Tns is conducted through the cut-and-paste mechanism [89]. DNA Tns in bacteria include four groups: insertion sequence (IS), composite Tns, non-composite Tns (Tn3 family), and the transposable phage Mu. ISs are the smallest (up to 2.5 kbp) independent MGEs responsible for mutations in bacterial genomes [90, 91]. Thus, ISs are also applicable for random mutagenesis in bacterial cells [92]. It has been demonstrated that the IS element *ISPa133* in *Pseudomonas aeruginosa* produces the PrD protein which provides resistance to carbapenems [93]. Composite Tns are DNA loci flanked by ISs [94]. These loci between ISs can bear antibiotic resistance genes and genes facilitating the transposition between bacteria cells [95 – 97]. For instance, *E. coli* strains can bear Tn5, Tn9, Tn10, and Tn903 composite Tns associated with antibiotic resistance [98]. Non-composite Tns, or Tn3 family, are another group that does not include ISs but is also flanked by IRs [99]. Non-composite Tns also participate in mutagenesis and antibiotic resistance in bacteria [99, 100].

Phage Mu, a phage from the family *Myoviridae*, combines features of both phages and transposable elements [101]. Mu phage infects *E. coli* and can be inserted into the chromosome by transposition. Thus, Mu phage chooses strategies between the prophage form and the lytic cycle. The Mu phage bears *attR* and *attL* sites recognized by its Tase enzyme [102]. Replication of Mu phage is based on replicative transposition, a process in which many copies of Mu DNA are produced [103]. As a result, the Mu DNA occupies about half of the total DNA in the bacterial cell [104]. The phage Mu minimizes the risk of self-integration using its MuA-transposase and MuB ATP-dependent nonspecific DNA-binding protein [105]. In the presence of ATP MuB proteins form the oligomeric clusters on DNA. These clusters indicate the nearby DNA as a target for integration [106]. MuA proteins provide transposition immunity by binding to Mu ends sequences to protect local Mu DNA [107]. The balance between MuA and MuB is essential for transposition immunity since a high concentration of MuB can overcome MuA protection effect [107]. After the isolation of phage Mu, a series of Mu-like phages were discovered in *P. aeruginosa* [108], *Rhodobacter capsulatus* [109] and *Haemophilus parasuis* [110].

1.3 Diversity of CRISPR-Cas systems: a short introduction

The CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats and CRISPR-associated genes) system is a prokaryotic adaptive immune system that provides bacteria and archaea with sequence-specific immunity against foreign mobile genetic elements, such as phages and plasmids [111, 112]. According to bioinformatics studies, the CRISPR-Cas system is widespread among 40% of bacterial and 85% of archaea genomes [113, 114]. This system is characterized by its ability to acquire, store, and utilize information about past infections of invading genetic elements to fend off future invasions through short nucleotide sequences. The system's components include CRISPR arrays and *cas* genes. The CRISPR array is composed of short repeated sequences interspaced with unique sequences derived from foreign genetic elements [115, 116]. These unique sequences, known as spacers, are the key elements of the CRISPR-Cas system in the recognition process of foreign genetic elements that the prokaryotic cell has previously encountered. The *cas* genes are responsible for coding Cas proteins, which process the CRISPR array and destroy foreign genetic elements bearing sequences that are entirely or partially complementary to the spacers. The destruction of foreign genetic elements is carried out by specific Cas proteins, known as Cas effectors. To date, the diversity of CRISPR-Cas systems identified is vast, encompassing two classes, six types, and numerous subtypes [114]. The classification is based on the signature Cas effectors, the organization of the CRISPR locus, and the mechanisms of interference and adaptation. The two classes are distinguished by the number of effector molecules involved in the interference stage, with Class 1 (types I, III, IV) systems utilizing multi-protein complexes and Class 2 (types II, V, and VI) systems relying on a single protein [114]. Furthermore, there is no significant evidence of the evolutionary relationship between CRISPR effector proteins of Class 1 and Class 2 CRISPR-Cas systems [117].

The first CRISPR array in *Escherichia coli* strain was described in 1987 [118] but detailed studies were started by Francisco Mojica in the 1990s [119, 120]. Over the past decades, a vast number of studies have been dedicated to CRISPR-Cas system that shed light on the mechanisms and diversity of CRISPR-Cas systems. It was discovered that the presence of a short specific nucleotide motif next to the protospacer region is necessary for CRISPR interference [121, 122]. This motif, called protospacer adjacent motif or PAM [123], is specific for particular types of CRISPR-Cas systems. The seed region proximal to PAM is also essential for recognition by crRNA [124]. These discoveries demonstrated the importance of perfect

matching between spacer and protospacer for successful CRISPR interference. In other words, MGEs are able to escape the CRISPR-Cas system if they acquire mutations in the protospacer region. In the seed region of the protospacer, it is enough to acquire a point mutation in one nucleotide position, so-called escape mutations [124]. To overcome escape mutations in MGEs CRISPR-Cas systems can adapt via the acquisition of new spacers in the CRISPR array [125]. Recently, the CRISPR-Cas system has been actively used for gene editing and gene regulation for experimental purposes [126, 127]. For this aim, it was essential to understand the appropriate PAMs and protospacer in the target region. For example, the different PAMs of the type I-E CRISPR-Cas system have varying degrees of recognition strength by crRNA-guided complex [128].

Notwithstanding the diversity of CRISPR-Cas systems, the mechanisms of CRISPR-Cas actions are similar except for particular details and include three stages: adaptation, expression, and interference (**Fig. 1.1**). During the adaptation segments of DNA (called protospacers) of invading mobile genetic elements are captured by the adaptation complex comprising Cas 1 - Cas 2 proteins and integrated in the first position of CRISPR array as spacers [125, 129]. The second phase includes the expression of the CRISPR array and subsequent processing of the transcript to crRNA (CRISPR RNA) [116]. Depending on the type of CRISPR-Cas system these crRNAs interact with one or more Cas proteins to form crRNA-guided protein complex for recognition of foreign DNA fragments. In addition to crRNA, in type II of Class 2 CRISPR-Cas systems the *trans*-activating CRISPR RNA (tracrRNA) is essential for immunity [126, 127]. After recognition of appropriate DNA fragment due to complementary matching to crRNA, the Cas effector complex degrades foreign DNA during a process known as CRISPR interference [111, 116]. The detailed descriptions of each type CRISPR-Cas system are discussed below.

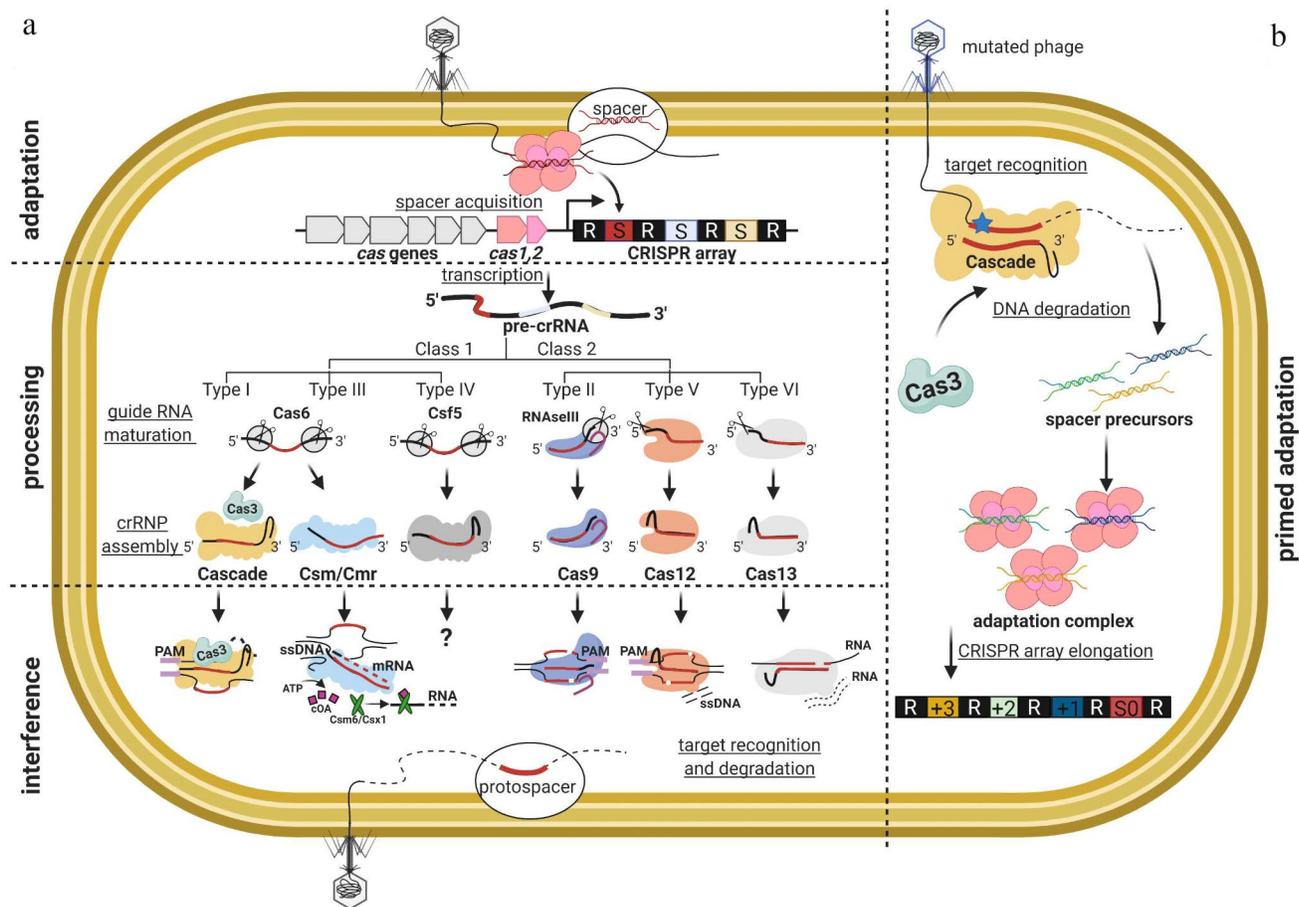


Figure 1.1. The diversity of CRISPR-Cas system and the mechanisms of CRISPR-Cas defense (adapted with permission from Isaev, A., Musharova, O. & Severinov, K., 2021).

a, The demonstration of Class I and II CRISPR-Cas immunity against phages. Three main steps are shown: spacer acquisition during adaptation, pre-crRNA processing and target recognition and degradation during CRISPR interference. **b**, The scheme of primed adaptation during mutant phage infection in type I-E CRISPR-Cas system.

1.3.1 Class 1 CRISPR-Cas systems

Nowadays, the class 1 CRISPR-Cas systems consist of three different types: type I, III, and IV. Cas3 nuclease and multi-subunit crRNA-binding Cascade complex are the characteristic components of the type I CRISPR-Cas [113]. Cas3 nuclease includes the helicase domain with two RecA motifs and the N-terminal HD nuclease domain [130, 131]. These domains enable ATP-dependent ssDNA translocation and degradation. In addition to the main domains, Cas3 possesses an accessory C-terminal domain (CTD) [131]. The deletion of CTD decreases the affinity of Cas3 for Cascade and inhibits CRISPR interference

in vitro [132]. The experiments with similarly structured helicases have demonstrated that the CTD facilitates the unwinding DNA strands by the helicase domain [133]. According to the crystal structure, the various metal ions such as Co^{2+} , Mn^{2+} , and Ni^{2+} facilitate the functionality of Cas3 HD-domain, but the nuclease activity is not activated by Ca^{2+} or Mg^{2+} [134]. A popular model bacterium for molecular biology studies, the *E. coli* K-12 strain was engineered to harbour an inducible type I-E system, with IPTG- and arabinose-inducible promoters, that is one of the best-studied CRISPR-Cas systems [124, 125, 128]. The ribonucleoprotein complex Cascade consists of one Cse1, two Cse2, one Cas5, six Cas7, and one Cas6 protein and a 61 nt crRNA [113]. Although the *E. coli* K-12 strain is actively used for studies of type I-E CRISPR-Cas system, no active CRISPR-Cas systems have been identified in natural *E. coli* isolates to date [113, 135, 136], yet functional type I-E CRISPR-Cas systems are found in natural *Klebsiella* spp. strains [137]. The study conducted by Swarts. *et al* has demonstrated that the activated type I-E CRISPR-Cas cured *E. coli* K12 Δhns cells from high-copy plasmids in non-selective prolonged cultivation with the acquisition of new spacers against the plasmids into the CRISPR array [138]. Cas1 and Cas2 proteins form the CRISPR adaptation complex responsible for the acquisition of new spacers into the CRISPR array. Two different types of CRISPR adaptation are considered: naive and primed. Overexpressed Cas1 and Cas2 proteins can acquire new spacers during naive adaptation without other Cas proteins [125, 129]. In contrast to the naive adaptation, the primed adaption in type I CRISPR-Cas systems requires all Cas machinery [125, 139]. Type I-E primed adaptation provides a quick selection of spacers from the *cis*-strand of DNA in relation to the protospacer. It was also demonstrated the acquisition of spacers is enhanced with primed adaptation if there is imperfect matching with the corresponding crRNA spacer [139]. Thus, the primed adaptation allows the quick update of the CRISPR array with new spacers to overcome escaper mutations in invading genetic elements. The seed region for Cascade complex guided crRNA during CRISPR interference is located on PAM and 5'-end of protospacer [124, 139].

Type III CRISPR-Cas systems illustrate a diverse composition and organization. The main signature of the type III CRISPR-Cas system is Cas10 protein including the HD and Palm (RNA-recognizing) domains, which both participate in the recognition of target nucleic acids [140]. Type III CRISPR-Cas systems are divided into two main types III-A and III-B. Type III-A system harbors a signature Csm2 protein and usually Cas1, Cas2, and Cas6 proteins. Cmr protein is a signature of type III-B systems, but many type

III-B systems lack Cas1, Cas2, and Cas6 proteins [113]. In addition to the main subtypes, two single variants were differentiated in type III-C and type III-D systems, which do not possess Cas1 and Cas2 proteins. The main characteristic of the type III-C system is a Cas10 with an inactive cyclase-like domain while the type III-D system coded Cas10 protein without the HD domain [113, 140]. It was shown that transcription of the target DNA is required for type III CRISPR interference [141]. According to the model of the type III CRISPR interference, the type III multi-subunit effector complex recognizes and cleavages the target ssRNA transcript complementary to crRNA at UA sites [142]. The Cas7 family proteins, Csm3 in type III-A and Cmr4 in type III-B possess RNase activity [143, 144] while the Cas10 HD domain is responsible for ssDNA cleavage [140]. The Palm domain of the Cas10 protein plays a key role in the activation of Csm6 RNase activity through the synthesis of cyclic oligoadenylate messenger [145]. The activated HEPN domain of the Csm6 protein provides collateral RNA degradation to rapid clearance of invading mobile genetic elements [146]. With a lack of target RNA transcript for the type III effector, the synthesis of cyclic oligoadenylate by the Cas10 subunit decreases subsequently regulating the Csm6 RNase activity [147].

Recent studies have expanded the diversity of type IV CRISPR-Cas systems by describing new subtypes. Along with already known types IV-A and IV-B [113, 148], new subtypes such as types IV-C, IV-D, and IV-E were described [149]. Type IV systems are primarily found in prokaryotic MGEs such as large antibiotic resistance plasmids. Type IV systems encode a signature *csf2* gene similar to the *cas7* gene [113, 147]. The main subtype-specific genes for types IV-A, IV-B, and IV-C are *dinG*, *cysH*-like, and *cas10-like* genes [113]. Type IV-A is characterized by the three type-specific *csf1*, *csf2*, and *csf3* genes, an endoribonuclease *cas6/csf5* gene, a helicase *dinG* gene [150], and a CRISPR array [148, 151]. Type IV-B systems contain the three type-specific genes, the *cas11-like* gene, and do not include a CRISPR array [148]. Type IV-C harbors *csf2*, *csf3*, and *cas10-like* genes. Some type IV-C systems additionally contain the *cas11-like* gene and CRISPR array [114, 152]. Bioinformatics analysis showed that spacers of type IV systems are complementary to protospacers from plasmids [148]. Clinical *Klebsiella pneumoniae* strains bear plasmids with antibiotic resistance genes and type IV CRISPR-Cas systems [153]. A type IV-A system from *P. aeruginosa* demonstrated defense against plasmids [154]. Thus, it is suggested that

plasmids bearing type IV CRISPR-Cas system limit a horizontal transfer of other plasmids. At the same time, the mechanism of type IV CRISPR interference requires additional studies.

1.3.2 Class 2 CRISPR-Cas systems

Class 2 CRISPR-Cas systems are characterized by a single-unit Cas effector and includes types II, V, and VI [114]. A single multidomain RNA-guided Cas9 effector is the hallmark of type II CRISPR-Cas systems. Besides the *cas9* gene, a typical composition of type II CRISPR-Cas system contains *cas1* and *cas2* genes and two RNAs: crRNA and trans-encoded small RNA (tracrRNA). The tracrRNA contains a sequence complementary to repeats in CRISPR array and is required for the maturation of pre-crRNA [155] and CRISPR interference [126, 127]. A specific host ribonuclease RNase III is responsible for crRNA maturation. It was shown that RNAase III cuts a duplex of tracrRNA and pre-crRNA. The final RNA hybrid contains a 75-nucleotide tracrRNA and a 42-nucleotide crRNA and participates in guiding Cas9 effector to target protospacer. The canonical PAM motif recognized by Cas9 effector is sequence 5'-NGG-3' [155]. However, subsequent studies have demonstrated PAM recognition depends on the type of Cas9 [156]. Cas9 protein contains two metal-ion-dependent nuclease domains, termed RuvC and HNH, that are responsible for target dsDNA cleavage [126]. The HNH domain is responsible for cutting the target DNA strand while the non-target DNA strand is cleaved by the RuvC domain [126]. During CRISPR interference Cas9 generates blunt ends within the PAM-proximal site [157]. Due to the simplicity of type II composition, RNA-programmed Cas9 protein became a popular tool for gene editing [157]. The Nobel Prize in Chemistry 2020 was awarded to Emmanuelle Charpentier and Jennifer A. Doudna for discovering new gene editing technology based on Cas9 effector so-called CRISPR/Cas9 genetic scissors [158]. Nowadays, gene manipulation based on various modifications of the Cas9 protein provides new approaches in gene editing [160] and to track molecular events within cells [161, 162]. Nuclease-inactive Cas9 proteins (dCas9) are applied for gene knockdown [163] and as well as for gene expression activation experiments [164].

Type V CRISPR-Cas systems possess Cas12 effector protein and its homologs, also known as Cpf1 in type V-A and C2c1 in type V-B, guided by a single crRNA [114, 165]. RuvC nuclease domain is explicitly detected in Cas12 amino acid sequence [114, 166]. The crystal structure of crRNA-Cpf1 effector complex

has demonstrated the presence of REC and NUC lobes in the complex [167, 168]. The Nuc domain identified in NUC lobe possesses the similar functionality in the cleavage of the target strand of DNA with HNH domain of Cas9 effector [167]. It was demonstrated that both RuvC and Nuc domains are essential in the cleavage of dsDNA [167]. In contrast to type II systems in type V systems RNase III is not required for pre-crRNA processing provided by the Cas12 effector [169]. *Acidaminococcus* sp Cas12 effector complex recognizes the 5'-TTTN-3' PAM and provides cleavage of dsDNA with staggered DNA double-strand breaks each 4-5 nucleotide long [168]. The cryoelectron microscopy structure of a miniature (422–603 amino acids) effector complex Cas12f was obtained and described [170]. It was shown that Cas12f and sgRNA form an asymmetric dimer (Cas12f)₂-sgRNA complex providing CRISPR interference. Cas12f nicks DNA in 22 nt in TS and 22 nt in NTS positions upstream of the PAM [170]. Another PAM motif and the cleavage mechanism of target dsDNA compared to the Cas9-based gene editing tools potentially enhance the application of Cas12 effector complex for genome editing. The successful gene editing applications for mutagenesis in plants were demonstrated in recent experimental studies [171].

Type VI systems are distinguished by a signature Cas13 RNA-guided RNA endonuclease, formerly known as C2c2, in types VI-A, VI-B, VI-C, and VI-D [114, 165]. Cas13 protein contains two HEPN (higher eukaryotes and prokaryotes nucleotide-binding) nuclease domains [165, 172]. Remarkably, type VI systems were initially predicted in bioinformatics searches via the identification of new effector proteins with two HEPN domains (HEPN1 and HEPN2) next to CRISPR arrays [165]. HEPN1 domain is located in the center of Cas13a, Cas13c, and Cas13d, or at the N-terminus of Cas13b, while HEPN2 domain is generally located at the C-terminus of Cas13 structure. Cas13 enzyme exhibits two separate RNase activities for pre-crRNA maturity for target RNA cleavage and recognition [173]. Cas13 (C2c2) crystal structure demonstrated the presence of a crRNA recognition (REC) lobe and a nuclease (NUC) lobe [174]. The NUC lobe consists of both HEPN domains and carries out the recognition and cleavage of target RNA transcripts while the REC lobe includes Helical-1 domain providing pre-crRNA cleavage [174]. Besides the HEPN domains, Cas13 protein possesses an additional metal-ion-independent RNA hydrolytic activity for processing pre-crRNA [174]. The composition of crRNA includes direct repeats in both sides of the spacer sequence which are responsible for high-affinity to Cas13 protein [174]. The matured crRNA remains bound to Cas13 protein, forming the RNA-guided effector complex that can recognize ssRNA

complementary to the crRNA spacer sequence [174]. The high-resolution structure of Cas13b was also described [175]. The Cas13 effector complex has a significant potential application for RNA manipulation. Cas13 effectors have been applied for RNA interference [176] and RNA detection [177].

1.3.3 CRISPR-interference on the example of type I-E CRISPR-Cas system

To implement anti-phage protection by CRISPR-Cas system a bacterial cell provides an expression of *cas* genes. In the type I-E system of *E. coli* K-12, it was shown that approximately 60 Cascade complexes provide around 87.5% of defense efficiency [178]. Moreover, the Cascade complex requires a significant time to find a target MGE. For instance, it takes approximately 9 minutes for 10 Cascade complexes, and about 1 minute for 90 Cascade complexes to find a target MGE, respectively [178]. The concentration of Cascade complexes in highly induced model strains reaches up to about 130 copies per cell [178]. However, this high concentration for Cascade complexes is not achievable under natural conditions for several reasons. Firstly, the high expression of *cas* genes requires a high metabolic cost for a single bacteria cell [179, 180]. Secondly, the risk of autoimmunity increases with a high concentration of CRISPR effector complexes [181, 182]. The spatial distribution of Cascade represents a Gaussian distribution among the cell cytoplasm with the highest concentration in the nucleoid [178] which potentially notes a space-dependent search for target MGEs. Cascade complexes permanently scan DNA for protospacers matched to the crRNA component [178]. Target foreign DNA recognition depends on unidirectional R-loop formation provided by base-pair complementarity [183]. The presence of PAM is crucial for CRISPR interference, allowing discrimination between host and invading DNA [123, 124]. Mutations in the PAM-proximal region have been demonstrated to cause unstable binding by Cascade [184] which in turn provides for phages to overcome CRISPR immunity [125]. Otherwise, the complete R-loop across the protospacer region forms a highly stable Cascade and target DNA complex [184]. The high affinity triggers DNA cleavage by the CRISPR effector. In type I-E CRISPR-Cas the DNA cleavage occurs through the recruitment of the helicase/nuclease Cas3 protein [185]. The Cas3 protein can provide collateral ssDNA cleavage [186]. If the target DNA bears tolerated mutations in the seed region, the Cascade complex forms an incomplete R-loop that triggers the collateral *trans* ssDNase activity of Cas3 protein [186]. This collateral activity of Cas3 protein can play an essential role in primed adaptation [187]. During the formation of a stable R-loop in perfectly matched target DNA, the Cas3 helicase/nuclease performs nick

of non-target strand in *cis* position [186]. The Cse1 subunit of Cascade complex is responsible for interaction with PAM, facilitating an initial step for R-loop formation [188]. The stable R-loop binding with Cascade provides activation of DNase and ATP-dependent helicase activity of Cas3 proteins [189]. Firstly, Cas3 nicks the non-target strand of DNA in *cis* position and then unwinds double-stranded DNA by ATP-dependent helicase activity [186]. Then, Cas3 provides cleavage of the target strand of DNA in *trans* position and keeps repetitive cycles of cleavage of non-target DNA in *cis* position for final dsDNA degradation [186].

1.3.4 Escape mechanisms of mobile genetic elements against CRISPR-Cas systems

Mobile genetic elements are able to overcome CRISPR-Cas systems due to various mechanisms. The most common way to avoid CRISPR interference is random mutagenesis in the target protospacer and PAM. During the replication process, bacterial or viral DNA polymerase might make mistakes with a low likelihood of approximately 10^{-9} errors per nucleotide for *E. coli* DNA polymerase III per replication cycle [190]. It was described earlier that the escape mutations in protospacer region can inhibit or prevent CRISPR interference. The seed region of protospacer are the more sensitive for mutations to prevent CRISPR interference compared to the distal to PAM protospacer site [124]. The CRISPR-Cas system can also be broken due to mutations that make bacteria cell vulnerable for phage infection [191-194]. IS elements have been shown to break the CRISPR immunity via transposition in CRISPR-Cas loci [195].

In addition to random mutagenesis, phages and plasmids can encode specific proteins termed as anti-CRISPR (Acr) proteins that inhibit CRISPR-Cas activity [196]. Acr proteins were first identified in experiments involving phages of *P. aeruginosa* [197]. Acrs use various mechanisms for inhibiting different stages of CRISPR-Cas immunity. For instance, both AcrIF1 and AcrIF2 block target DNA binding by Cascade but use different molecular mechanisms. AcrIF1 binds with Cascade complex and blocks crRNA-DNA hybridization [197]. At the same time, AcrIF2 binds with another site of Cascade and prevents PAM recognition [197]. AcrIF3 prevents DNA cleavage by Cas3 endonuclease/helicase in type I-F CRISPR-Cas system [197]. AcrIF9 provides non-specific DNA binding by Cascade [198]. AcrIE1 dimer inactivates type I-E CRISPR interference blocking Cas3 [199]. Today, various Acr types have been identified [200]. Acr classification is based on the inhibition of the type CRISPR-Cas system [201]. The

low homology among different Acr proteins makes the search for new Acr types challenging [202, 203]. Experimental studies have revealed the diverse molecular activities of Acr proteins in preventing CRISPR-Cas action, which includes interference with crRNA, inhibition of protospacer recognition, inhibition of Cas effectors, and CRISPR interference [204]. One particular Acr protein can inhibit different types of CRISPR-Cas systems. AcrIF2 have been demonstrated inhibited activity against both type I-F and type I-C CRISPR-Cas systems in *P. aeruginosa* [205]. At the same time, AcrIC1 and AcrIC3 are able to limit only type I-E CRISPR-Cas systems [205]. AcrIF2 blocks binding Cas effector complex with target DNA but AcrIC1 and AcrIC3 inhibit Cas3 nuclease [205].

Acrs are characterized by a relatively short sequence length (50 - 200 aa) [206], and their lack of conserved domains and common structural features complicates their prediction using traditional bioinformatics methods. However, *acr* genes often cluster together and are regulated by expression factor genes, termed Acr-associated (*aca*) genes, which possess a single conserved HTH motif [207]. These clusters are often harbored in prophage DNA or plasmids. The Aca proteins serve as a significant marker for predicting potential Acr candidates [203, 207]. The target protospacers located in the bacterial genome bearing a potential active CRISPR-Cas system, or *self-targeting*, might indicate the presence of potential Acrs [207]. In addition to traditional bioinformatics methods, machine learning and deep learning approaches are also employed to search for new Acrs [207, 208]. The deep learning models have demonstrated that Acr features like molecular weight, isoelectric point, location in prophage area, and count of positive and negative charge residues might be essential for the prediction [208]. Despite high level of prediction significance by the models, the experimental validation of Acr candidates might be complicated. The predicted AcrIVA proteins demonstrated no inhibition activity against Cas13a effector complex from *Leptotrichia wadei* strain but inhibited Cas13a nuclease in *Leptotrichia shahii* [208]. This high selectivity can demonstrate the specialization of Acr protein on the CRISPR-Cas system of the particular strain. The diversity of Acr proteins sheds light on how phages and plasmids bearing several protospacers can overcome CRISPR-Cas systems. Many Acrs have been found in antibiotic resistant conjugative plasmids in clinic isolates that facilitate fast HGT of these plasmids in bacterial populations [209]. From an applicable point of view, Acrs proteins represent useful tools for the accurate regulation of CRISPR-Cas gene editing [210].

1.4 Role of MGEs in antibiotic resistance of microbes

Antibiotic resistance genes in microbes are generally associated with MGEs, thus, I provide a short review of the problem related to antimicrobial resistance (AMR). The spread of AMR, the phenomenon in which bacteria become less sensitive to antibiotics, poses a global health challenge [211]. Many pathogenic strains acquire multidrug resistance (MDR) that significantly complicates antimicrobial treatment during infection. Microbes with MDR have been evolving to overcome the efficacy of previously discovered antibiotics [212]. Microbes possessing MDR are associated with 4.95 million deaths annually and are directly responsible for at least 1.27 million deaths yearly [213]. The current estimation of the situation forecasts up to 10 million deaths per year, caused by antibiotic resistant pathogens, by 2050 [213]. It was also shown that the spread of AMR occurs more rapidly compared to chromosomal mutations [214]. The rapid spread of AMR is facilitated with HGT of conjugative plasmids that carry antimicrobial resistance genes [215]. Moreover, the AMR plasmids bear genes that facilitate both HGT and plasmid sustainability in bacterial populations. These genes are often clustered in pathogenicity islands, the sets of genes providing ecological fitness in changed conditions [216, 217]. For example, the genes encoding AcrIE9.2 proteins were identified in AMR plasmids of *Klebsiella pneumoniae* clinical isolates [209]. Thus, the AcrIE9.2 blocks the type I-E CRISPR-Cas system in *K. pneumoniae* strain that facilitates AMR plasmid spreading. AMR plasmids frequently bear TA systems that increase the maintenance of plasmids by a bacterial cell and complicate the elimination of AMR genes from bacterial populations [219]. Many microorganisms form spatial heterogeneous structures: colonies or biofilms; to overcome environmental stress conditions including antibiotic treatments [215, 217]. The species composition of biofilms can consist of one microbial strain [220] or multiple different strains [221]. Microbes produce various bioactive compounds forming extracellular polymeric substance (EPS) matrix [221]. It was demonstrated that biofilms provide resistance to antibiotics from 10 to 1000 times and facilitate the appearance of persistent cells [222]. Chapter 2 clearly illustrates that even a small fraction of persistent *E. coli* cells can ensure survival for the entire colony against antibiotic treatment. The problem of spreading antibiotic resistance forces the search for a new source of antimicrobial compounds and methods to overcome antibiotic resistance. Phage therapy demonstrates a potential application to threaten pathogenic microbes

for clinical purposes [223]. It was also demonstrated that poor-studied natural microbial communities of oceans are the valuable source of secondary metabolites including compounds with antimicrobial properties [224].

1.5 Studies of diversity in microbial communities

Natural microbial communities are overwhelmingly abundant in the biosphere. Bacteria and archaea constitute approximately 15% and 1.2% of the total living biomass correspondingly while animals represent only 0.36% [225]. Microbial communities are heterogeneous ecological systems, engaging in complex metabolic interactions that produce a vast array of metabolites [224]. The diversity of Earth's prokaryotes is estimated in 2.2-4.3 million species [226], yet only about 25 thousand prokaryotic species have been genomically described (~0.5-1%) [227]. This underscores the immense potential for discovering new taxa and biologically active compounds, including antimicrobial molecules [224]. The isolation of new strains from natural microbial communities provides an investigation of new biosynthetic clusters. However, the impossibility of cultivating most bacteria and archaea in laboratory conditions significantly complicates the investigation of microbial diversity [228]. To overcome this limitation, next-generation sequencing (NGS) has emerged as a powerful tool, providing new possibilities for researchers to study microbial communities. Depending on the study's aim, two approaches are employed: metagenomic sequencing of the entire microbial community with subsequent bioinformatic analysis of metagenomes [229] and 16S ribosomal RNA sequencing to identify the taxonomic composition of the microbial community [230]. For whole metagenomics sequencing both short-read (Illumina, BGI Genomics) and long-read (Oxford Nanopore Technologies, Pacific Biosciences) sequencing technologies are applied [231, 232]. The annotation of new metagenomes enables the discovery of novel biosynthetic clusters, enzymes, and biologically active molecules [233].

The natural environment, including plants and animals, contains abundant and diverse bacterial communities [234, 235]. The commensal gut microbiota of humans plays significant role in human health [236]. At the same time, soil microbiota also provides a huge impact on the health of human, animals and plants [237]. The ecological drift is suggested to influence the microbiome composition of different organisms [238]. It is estimated that up to 80% of natural microbial communities exist in biofilms and

other spatial structures [239]. This spatial organization creates microenvironments that enable various strains to exchange energy and nutrients more efficiently [239, 240]. The interactions within such communities are diverse and often highly sophisticated. For instance, the sophisticated commensal interactions between bacterial communities and *Arabidopsis thaliana* have been shown on leaf and root samples [241]. An analogous spatial organization has been demonstrated in aquatic communities between bacteria and phytoplankton [242]. The diversity of microbial communities in the ocean environments and marine sediments represents significant interest for studies as less-investigated communities [243, 244]. The marine microbiota is a valuable source of new biologically active molecules, antibiotics, new enzymes for bioengineering purposes, and chemical compounds for biotechnological applications [245]. A recent study of ocean microbiome data has discovered new biosynthetic gene clusters including post-translationally modified peptide (RiPP) pathways [245]. New RiPPs potentially represent a new source of antimicrobial active compounds [246]. Phages in marine communities also play an important role in ocean ecology. Every day, bacteriophages infect about 20-30% of bacteria in oceans, significantly influencing the global carbon cycle [247]. Therefore, the marine microbiota can provide the additional source of novel phages.

In conclusion, the diversity of interactions between phages and bacteria in natural environments provides heightened interest in fundamental studies and searching for new biologically active compounds. The methodology of the study of microbial communities is described in Chapter 3. In Chapter 4, I contributed bioinformatics analysis for characterizing *E. coli* isolates and searching for potential valuable metabolites in their genomes.

Chapter 2: Persistence of plasmids targeted by CRISPR interference in bacterial populations

The results of Chapter 2 are published in:

Mamontov, V., Martynov, A., Morozova, N., Bukatin, A., Staroverov, D. B., Lukyanov, K. A., Ispolatov, Y., Semenova, E., & Severinov, K. (2022). Persistence of plasmids targeted by CRISPR interference in bacterial populations. *Proceedings of the National Academy of Sciences*, 119(15).

2.1 Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR associated genes) is a widespread form of adaptive immunity in prokaryotes [114]. CRISPR-Cas systems are able to recognize and destroy nucleic acids with sequences complementary to spacers stored in CRISPR arrays [111-112]. In the array, spacers are separated by the repeat sequences. CRISPR array transcripts are processed into individual CRISPR RNAs (crRNAs) containing spacer sequences with flanking repeat fragments. Individual crRNAs bind to Cas proteins forming an effector complex, which can recognize protospacers — sequences complementary to the crRNA spacer part. For CRISPR-Cas effectors targeting DNA, the recognition requires, in addition to full or partial complementarity between crRNA spacer and the protospacer, the presence of PAM, a protospacer adjacent motif, that is recognized by the protein part of the effector complex [123].

Multiple examples of protection of prokaryotic cells by different CRISPR-Cas systems acting through the CRISPR interference mechanism described above from infection by DNA and RNA viruses and transformation by plasmids have been documented [124, 248, 249]. Depending on the virus and the type of CRISPR-Cas system, a cell mounting the interference response can clear the infection and survive or die in the course of abortive infection. In the latter case, the population as a whole benefits because the appearance of progeny viruses is prevented [250]. Viruses respond to the pressure from CRISPR-Cas by acquiring point mutations in protospacers targeted by crRNAs or in their PAMs [251]. In turn, cells respond to such viral escapers by updating their CRISPR memory by acquiring additional viral-derived spacers [125].

During plasmid transformation/conjugation experiments CRISPR interference results in decreased efficiencies of DNA uptake. Mutations in targeted protospacers or their PAMs restore transformation

efficiencies [124]. In experiments where cells are forced to keep a plasmid targeted by CRISPR-Cas by inclusion of an appropriate antibiotic in the medium, mutations inactivating CRISPR-Cas system components are observed [188-190].

Given a considerable interest in potential use of CRISPR-Cas targeting antibiotic-resistance plasmids as means to reduce antibiotic resistance spread, we here undertook a study of the interaction of the well-studied *E. coli* type I-E CRISPR-Cas system [125, 128, 252] with plasmids carrying protospacers recognized by the Cascade effector complex. We were specifically interested in colonies formed on antibiotic-containing selective media by cells with an active CRISPR-Cas system transformed with plasmids carrying protospacers targeted by the effector. We report that only a small fraction of resulting colonies is formed by cells with inactivated CRISPR-Cas. Most colonies have an active CRISPR-Cas system and unaltered plasmids which are subject to CRISPR interference. Using a combination of microbiological, microscopic and microfluidics experiments we show that cells in such colonies are heterogeneous, with most cells having little or no plasmid. Apparently, these colonies are formed due to the presence of a minor fraction of cells that manage to keep the plasmid at conditions of ongoing CRISPR interference. We use mathematical modeling to show how plasmids persist in generations of such cells due to a balance of CRISPR interference and plasmid replication rates. Our results show that potentially beneficial plasmids can be stably maintained in bacterial populations even while being targeted by CRISPR, allowing rapid expansion of plasmid-bearing subpopulations when conditions demand.

2.2 Methods

2.2.1 Strains and plasmids

E. coli strain KD263 (K-12 F+, *lacUV5-cas3* ara Bp8-*cseI*, CRISPR I: repeat-spacer g8-repeat, CRISPR II deleted) has been described [253]. The pG8 plasmid carrying a 209-bp phage M13 fragment with the g8 protospacer has been described [125]. The pRSFG8 plasmid carrying the 209-bp phage M13 fragment with the g8 protospacer has been constructed previously [254]. The pG8-GFP plasmid was derived from pG8 by cloning the *TagGFP2* gene (Evrogen) following the Gibson assembly protocol (NEB). Primers

used for DNA amplification are listed in **Table S2.2**. *E. coli* cells were grown at 37 °C in LB medium (per 1 L: 5 g NaCl, 10 g tryptone, and 5 g yeast extract) or on LB-agar plates containing 1.5% agar.

2.2.2 CRISPR interference assay

E. coli strain KD263 overnight culture was diluted 100 times into 5 mL of LB. The cells were grown in the presence (CRISPR ON) or in the absence (CRISPR OFF) of 1 mM arabinose and 1 mM IPTG for *cas* genes expression until cultures OD₆₀₀ reached 0.6. The electrocompetent cells were prepared following a standard protocol [255] and transformed with 5 ng of the protospacer plasmid (pG8 plasmid or pRSFG8). Next, the transformed cells were grown in 1 ml of LB supplemented with 1 mM arabinose and 1 mM IPTG for CRISPR ON cultures and 1 ml of LB for CRISPR OFF cultures for 1 h. The 50 µl aliquots of serial dilutions of the transformation mixtures were plated onto LB agar plates containing 100 µg/ml ampicillin (for pG8 plasmid transformation) or 50 µg/ml kanamycin (for pRSFG8 plasmid transformation) (CRISPR ON) or without (CRISPR OFF) inducers. The plates were incubated at 37°C overnight. The efficiency of transformation (EOT) was determined as a number of colony forming units (CFU) per 1 µg of plasmid DNA (**Fig. 2.1a,b**). Each transformation was performed in triplicate. PCR assay of CRISPR array for investigating CRISPR adaptation in transformant colonies was performed as described previously [254].

To test the condition of the protospacer plasmids in CRISPR ON transformants, plasmid DNA from ten randomly chosen and pooled CRISPR ON colonies was isolated using GeneJET Plasmid Miniprep Kit (Thermo scientific) and retransformed into fresh prepared CRISPR ON and CRISPR OFF cells (**Fig. 2.1c**). To test the functionality of CRISPR-Cas system in CRISPR ON transformants, retransformation of cells derived from ten randomly chosen individual CRISPR ON colonies was carried out with the second plasmid bearing the compatible origin and the g8 protospacer: the cells initially transformed with pG8 plasmid received pRSFG8 plasmid and vice versa (**Fig. 2.1d**). The efficiency of transformations was determined as described above.

2.2.3 Quantitative PCR assay of plasmids

qPCR was performed using DTlite4 Real-Time PCR System (DNA-Technology). Reactions were carried out in triplicate (technical repeats) in a 20 μ l reaction volume supplemented with 0.8 units of HS Taq DNA polymerase, 2 μ l 10x Taq Turbo buffer (Evrogen), 0.25 mM dNTPs, 0.2 μ l Tween 20, 0.1 μ l of SYTOTM 13 intercalating dye (LifeTechnology), 1 μ l of sample and appropriate primers at 5 pM. The primers for qPCR are listed in **Table S2.2**. Three randomly chosen colonies were suspended in 20 μ l distilled water. The results of qPCR with plasmid-specific primers were normalized to genomic DNA with regard to the efficiency of the primers (Dataset).

The efficiency of the primers was calibrated following a standard curve [256]. To calculate the standard curve for the primers, three random individual colonies of the transformants were chosen and suspended in 20 μ l distilled water. Next, three 10-fold serial dilutions of the suspended samples were assayed with qPCR using three technical replicates for each sample. We used only the results of qPCR with deviation less than 0.1 Δ Ct (cycle threshold) among technical repeats of one dilution. The efficiency of primers was calculated as an average slope of the plot of logarithmic concentration per dilutions vs. Δ Ct (Dataset). Three repeats for each group of the primers were carried out. The following efficiencies of the primers were obtained: 2.0 for Bla_dir and Bla_rev (amplification efficiency 100%), 1.94 for GyrA_dir and GyrA_rev (amplification efficiency 94%) and 2.08 for pRSF_ori_dir and pRSF_ori_rev (amplification efficiency 108%). Mean PCN was estimated as a ratio of genomic to plasmid Δ Ct values considering the efficiency of the primers (Dataset).

2.2.4 Replating of transformants

Four randomly chosen individual colonies of the CRISPR ON and CRISPR OFF transformants were replated on three types of selective media: LB supplemented with appropriate antibiotic (Ab) and inducers (Ind) to maintain the CRISPR-Cas activity, LB supplemented with appropriate Ab only (to determine the number of plasmid-bearing cells) and LB (to determine the total number of cells) (**Fig. 2.2a**). Each colony was suspended in 500 μ l of LB and eight 4-fold serial dilutions of the suspended cells were prepared. Next, 5 μ l of each dilution was plated on the selective media. The CFUs were counted on each plate. The colonies

from plates with Ab/Ind were used for the subsequent replating. Each replating was repeated at least three times.

2.2.5 Antibiotic-free cultivation of CRISPR ON cells and the growth curves

Three randomly chosen CRISPR ON colonies obtained after transformation with the plasmid pG8 were resuspended in 500 μ l of fresh LB. 50 μ l of each culture were transferred in antibiotic-free and antibiotic-supplemented 5 ml LB media for parallel cultivation. The antibiotic-supplemented LB media was prepared with 100 μ g/ml ampicillin, 1 mM arabinose and 1 mM IPTG, the antibiotic-free LB media was prepared with 1 mM arabinose and 1 mM IPTG. After 24 h cultivation at 37 °C the cultures from each tube were plated on three types of selective media using serial dilutions (see “Replating transformants”) and corresponding CFU were counted for each plate. Three CRISPR ON colonies obtained on plates, supplemented with 100 μ g/ml ampicillin, 1 mM arabinose and 1 mM IPTG, were randomly picked to test the conditions of the protospacer g8 and CRISPR-Cas system. The tests of the colonies were performed as described above in the subsection “CRISPR interference assay”.

The growth curves: Overnight culture of plasmid-free *E. coli* KD263 cells was resuspended 1:100 in fresh LB to estimate the growth rate of plasmid-free cells. To obtain overnight plasmid-bearing culture, a colony of *E. coli* KD263 cells bearing the plasmid pG8 was cultivated in LB media supplemented with 100 μ g/ml ampicillin. The overnight plasmid-bearing culture was resuspended 1:100 in fresh LB to estimate the growth rate of plasmid-bearing cells. The cell cultures were cultivated at 37 °C until cells reached the stationary phase. OD₆₀₀ readings were taken every hour using Thermo Scientific Helios Omega UV-VIS Spectrophotometer. The growth rate was calculated as the slope of the log of the growth curve [257]. Each measurement was performed in triplicates.

72 h parallel cultivation: CRISPR ON colonies were resuspended in 5 ml of LB media supplemented with 1 mM arabinose and 1 mM IPTG, CRISPR OFF colonies were resuspended in 5 ml of LB media. Cells were cultivated at 37 °C. Every 12 h of cultivation each culture was resuspended 1:100 in fresh appropriate LB media. Every 24 h of cultivation the cultures were plated on three types of selective media using serial dilutions (see “Replating transformants”) and plasmid-bearing fraction was counted for each culture. After 72 h cultivation 100 μ l of CRISPR ON culture was additionally plated on media supplemented with appropriate antibiotic and *cas* genes inducers. Obtained CRISPR ON colonies were tested on the presence

of functional CRISPR-Cas system and unchanged plasmid as described above in the subsection “CRISPR interference assay”. Each cultivation was performed in triplicates.

2.2.6 Flow cytometry

Several colonies of *E. coli* CRISPR ON and OFF cells bearing the pG8-GFP plasmid were suspended in PBS (phosphate-buffered saline) and passed through 100- μ m filters. Samples were investigated using FACSaria III (BD Biosciences); the flow cytometry protocol was customized for bacterial cells. Forward versus side scatter (FSC vs SSC) plots were used to gate the area of single cells; 2×10^5 events per sample in the gate was collected. TagGFP2 fluorescence was excited with 488-nm laser and detected with 530/30 filter. Three biological replicates for each sample were done. The data were analysed by FCSalyzer and Flowing Software.

2.2.7 Microscopy assay

Fluorescence imaging microscopy of CRISPR OFF and ON colonies was performed using Leica AF6000 LX system based on a DMI 6000 B inverted microscope equipped with HCX PL APO lbd. BL 63x 1.4NA oil objective and Photometrics CoolSNAP HQ CCD camera. GFP filter cube (excitation BP470/40 and emission BP525/50) was used to visualize TagGFP2. LB-agar fragments containing colonies were cut, placed on glass-bottom dishes so that the colonies were adjacent to the glass bottom. The colonies were observed in fluorescence and transmitted light channels. Colonies of *E. coli* KD263 cells transformed with pG8 were used as a negative control (no fluorescent protein). CRISPR OFF and ON colonies were also visualized with laser scanning confocal microscope DMIRE2 TCS SP2 (Leica) with HCX PL APO lbd. BL 63x 1.4 NA oil objective. The green fluorescent signal was acquired at 488-nm excitation and detected at 500- to 530-nm wavelength range.

2.2.8 Single cell microscopy in microfluidic device

Design of microfluidic device: The device was designed using AutoCAD® (AUTODESK®) and the Metafluidics database and fabricated from PDMS following standard soft lithography technique [258]. It

includes four major trenches of 100 μm width and 40 μm depth each, along which the growth medium is passed, and 1000 growth chambers with the depth of about 1 μm and the length of 20 μm on the front side that adjacents to the major trench and 80 μm on the lateral side. The inlet of the device contained a 25 μm filter to prevent clogging. To make the device a double layer mold was fabricated using SU - 8 2025 photoresist (Kayaku Advanced Materials, Inc) spin-coated onto a silicon wafer and exposed by contact photolithography with two chromium masks. For the first layer SU - 8 2025 was diluted by SU - 8 T thinner to achieve the thickness of the layer about 1 μm . After that the PDMS prepolymer and the curing agent (Sylgrad 184, Dow Corning) were mixed in a ratio of 10:1 w/w, degassed, poured into the mold, and cured at 65 $^{\circ}\text{C}$ for 4 h in an oven. Then the PDMS replica was detached from the mold, inlet and outlet holes were made by a 1 mm biopsy puncher. Finally, the replica was bonded with a cover glass slide after oxygen plasma treatment.

Single cell microscopy: Several colonies of *E. coli* CRISPR OFF cells bearing pG8-GFP plasmid were resuspended in LB medium with 1 mM arabinose and 1 mM IPTG inducers for *cas* genes expression and loaded to the microfluidic device. Single fluorescent cells caught in the growth chambers were tracked using a Nikon Eclipse Ti-E inverted epifluorescence microscope. Cells were cultivated in the growth chambers overnight at 37 $^{\circ}\text{C}$ on LB medium supplemented with *cas* genes inducers. The images were captured every 15 minutes for generating a time-lapse movie in transmitted light for all cell observation and in the green channel for fluorescence detection. The tagGFP2 fluorescence signal was detected using the Semrock filter set YFP-2427B. All images were obtained using Zyla 4.2 sCMOS camera (Andor). Fluorescence intensity from single cells was analysed using ImageJ software.

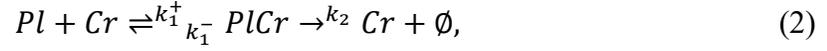
2.2.9 Dynamics of replication and degradation of plasmids (*conducted by A. Martynov*)

The dynamics of plasmid replication can be quite complex, yet it has two universal limits: For a few plasmids, the replication rate is proportional to the number of plasmids (i.e. the replication rate per plasmid is constant), while for the target (target) concentration of plasmids $[PI]_{st}$, the replication rate is zero. As it is often done [259], we approximate such dynamics by the Logistic model,

$$\left. \frac{d[PI]}{dt} \right|_{\text{replication}} = k_d [PI] \left(1 - \frac{[PI]}{[PI]_{st}} \right). \quad (1)$$

The coefficient k_d is the per capita rate of plasmid replication in the low concentration limit. The symbol $[x]$ indicates the concentration of a substance x . Assuming that the volume of a cell stays approximately constant, we define a concentration as the number of molecules per cell, and in the following we use the terms “concentration” and “copy number” interchangeably.

As a catalytic process, the interaction of CRISPR-Cas complexes Cr with plasmids Pl ,



is assumed to be well-described by the Michaelis-Menten kinetics,

$$\frac{d[Pl]}{dt} \Big|_{cutting} = -k_2[PlCr] = -k_2 \frac{[Pl]_0 + [Cr]_0 + \chi - \sqrt{([Pl]_0 + [Cr]_0 + \chi)^2 - 4[Pl]_0[Cr]_0}}{2}. \quad (3)$$

Here, as in the standard Michaelis-Menten derivation, the stationarity of concentration of the CRISPR-Cas-plasmid complex is assumed, the generalized dissociation constant χ is defined as

$$\chi \equiv \frac{k_1^- + k_2}{k_1^+}, \quad (4)$$

and no assumption is made about overabundance of the catalyst (CRISPR-Cas) or the substrate (plasmid). The total (bound in the Pl - Cr complex plus free) concentrations of plasmids and CRISPR-Cas complexes are denoted as $[Pl]_0$ and $[Cr]_0$.

Assuming that replication only increases the plasmid concentration so that $[Pl]$ in (Eq. 1) never exceeds $[Pl]_{st}$, we define a one-step birth-death process [260] for the population of plasmids. The probabilities of increasing or decreasing the population of plasmids by one $\beta_{[Pl]}$ and $\delta_{[Pl]}$ are given by $d[Pl]/dt|_{replication}$ (Eq. 1) and $d[Pl]/dt|_{cutting}$ (Eq. 3). The master equation that describes the temporal evolution of probability $P_{[Pl]}(t)$ to find a cell having $[Pl]$ plasmids at time t [260] reads

$$\frac{dP_{[Pl]}(t)}{dt} = \beta_{[Pl]-1}P_{[Pl]-1}(t) + \delta_{[Pl]+1}P_{[Pl]+1}(t) - (\beta_{[Pl]} + \delta_{[Pl]})P_{[Pl]}(t). \quad (5)$$

2.2.10 Redistribution of plasmids during cell division (conducted by *A. Martynov*)

In addition to cutting and replication of plasmids, the per cell PCN is also affected by cell division, which on average happens every $\tau \approx 20$ min. A conservative estimate would be that the redistribution of plasmids between daughter cells is completely random (in reality it is biased towards equal or half and half distribution). Assuming also that the act of cell division happens fast (instantaneous) compared to the replication and cutting of plasmids, the outcome of the redistribution process can be described by the

binomial distribution with the probability for each plasmid to go into any of two daughter cells equal to $1/2$. If a cell before the division had j plasmids, then the probability B_{ij} to find $0 \leq i \leq j$ plasmids in one of the daughter cells is

$$B_{ij} = \frac{j!}{i!(j-i)!} \left(\frac{1}{2}\right)^j. \quad (6)$$

2.2.11 Simulation procedure (conducted by A. Martynov)

As presented above, the temporal dynamics of plasmid copy number in a cell is approximated by a sequence of periods of continuous evolution, described by the master equation (Eq. 5), each followed by the instantaneous redistribution between daughter cells, described by the binomial distribution (Eq. 6). To estimate the distribution of plasmids in cells in CRISPR ON colonies after several hours of growth, we implement the following numerical procedure:

- For a given set of plasmid replication and CRISPR interference parameters k_d , $[PI]_{st}$, k_2 , χ , and $[Cr]_0$, we tabulate the replication and cutting rates $\beta_{[PI]}$ and $\delta_{[PI]}$ for all possible plasmid copy numbers, $1 \leq [PI] \leq [PI]_{st}$.
- We numerically integrate the master equation (5) till the cell cycle time τ , starting from every possible initial number of plasmids j , $0 \leq j \leq [PI]_{st}$. Naturally, the solution with zero initial plasmids will always be zero plasmids with probability one.
- The probabilities C_{ij} for a cell to end up with i plasmids at time τ after starting with j plasmids at $t = 0$,

$$C_{ij} \equiv P_i(\tau), P_k(0) = \delta_{k,j}, \quad (7)$$

are collected into the matrix \hat{C} . Another matrix \hat{B} is composed of binomial probabilities B_{ij} (Eq. 6).

- The probability to find k plasmids after time t is given by the $k + 1$ th element of the $[PI]_{st} + 1$ -dimensional vector \vec{P} ,

$$P^{\rightarrow} = \hat{C}^N (\hat{B}^N \hat{C}^N)^N P^{\rightarrow}(0), \quad (8)$$

where N (equal to the integer part of t/τ) is the number of cell cycles and the initial condition $\vec{P}(0)^T$ indicates how many plasmids were in each cell when the CRISPR-Cas system was activated. Here we assumed that the number of plasmids in a cell is assessed at the final stage of cell cycle just before cell division, thus an extra multiplication by \hat{C} . (Alternatively, when the number of generations is

not very large, this probability can be computed more efficiently by direct solution of the master equation (Eq. 5) for periods of time between cell divisions, alternated with binomial redistribution of plasmids between daughter cells according to (Eq. 6). In such a case we do not need to compute the matrix \hat{C} .

The evolution of the probability density $P_k(t)$ for the replication and interference rates (1) and (3) plotted in Fig. 6d is shown in Fig. 6e for cells initially having 1 plasmid, ($P_k(0) = \delta_{k,1}$ being the typical initial condition in a CRISPR-ON experiment) and in Fig. 6f for cells initially having the target number of plasmids, ($P_k(0) = \delta_{k,[PI]_{st}}$ being the initial condition for replating the CRISPR OFF cells on plates with inductor). The plots in Fig. 6 were computed using the following parameters $k_d = 0.3$, $[PI]_{st} = 100$, $k_2 = 0.5$, $\chi = 1$, and $[Cr]_0 = 10$.

As many birth-death processes, this stochastic process of plasmid replication, cutting, and redistribution has the unique convergent steady state $\vec{P}(\infty)^T = (1, 0, \dots, 0)$, corresponding to the extinction of all plasmids. However, after a few cell cycles, while the component $P_0(t)$ that corresponds to the fraction of cells with no plasmids steadily grows, other components $P_k(t)$, $k = 1, \dots, k_{st}$ that correspond to the probability to have a non-zero number of plasmids approach a steady state scaling form,

$$P_k(t) = f(t)P^{\sim}_k, k = 1, \dots, [Pl]_{st}, \quad (9)$$

shown in *SI Appendix*, Fig. 8. The slowly-decaying function $f(t)$ represents a universal convergence to the absorbing state $\vec{P}(\infty)^T = (1, 0, \dots, 0)$.

2.2.12 Transcriptome analysis (for unpublished results)

CRISPR ON and CRISPR OFF colonies were preliminarily obtained by transformation using the plasmid pRSFG8 as described for the CRISPR interference assay in *section 2.2.2*. The obtained CRISPR ON and CRISPR OFF *E. coli* colonies were used for overnight cultivation in LB media supplemented with 50 $\mu\text{g/ml}$ kanamycin, 1 mM arabinose and 1 mM IPTG for the CRISPR ON overnight culture, and in LB media with only 50 $\mu\text{g/ml}$ kanamycin for the CRISPR OFF overnight culture. The CRISPR ON and CRISPR OFF overnight cultures were then diluted 100 times into 5 mL of LB. The cells were grown in the presence (CRISPR ON) or in the absence (CRISPR OFF) of 1 mM arabinose and 1 mM IPTG in LB with 50 $\mu\text{g/ml}$ kanamycin until the optical density at 600 nm (OD 600) reaches a value of 0.6. Total RNA

was extracted from the harvested cells using the GeneJET RNA Purification Kit (Thermo Scientific) following the manufacturer's protocol. Contaminating DNA was removed using the TURBO DNA-free™ Kit (Thermo Scientific). In total, 3 CRISPR ON samples were prepared and 1 CRISPR OFF sample was as a control sample. The sequence was performed using 75 bp paired end Illumina technology sequencing with 10 million read per sample. All samples were sequenced in three technical repeats.

Sequencing data was converted to the fastq format and processed using trim-galore. The quality of the sequencing was checked with FastQC both before and after trim-galore processing. Read mapping was carried out using HISAT2. The gene expression level was determined as the number of reads covering the sequence of a given gene. Normalization of gene expression against a control sample and statistical analysis was performed in R using the DESeq2 package. For subsequent analysis, only genes that pass the statistical significance test with a p-value < 0.05 will be considered. The function of the genes will be determined using the R package GOFfunction. The UniProt and EcoCyc databases will be used to describe the set of genes with increased and decreased regulation. All obtained results of differential gene expression were analyzed for statistical significance. All necessary statistical analyses and visualizations were performed in R using the following packages: ggplot2, clusterProfiler, EnhancedVolcano, pheatmap, and GOFfunction.

2.2.13 Analysis of escape phages M13 (for unpublished results)

E. coli CRISPR ON and CRISPR OFF competent cells were prepared as described in 2.2.2. *CRISPR interference assay* section. The M13mp18 replicative form bearing g8 protospacer, β -lactamase gene and *gfp* gene was provided by Aleksandra Strotskaya. The replicative form was used for transformation of the competent cells. The efficiency of transformation (EOT) was determined as a number of colony forming units (CFU) per 1 μ g of dsDNA replicative form of phage M13. Each transformation was performed in triplicate. PCR assay of CRISPR array for investigating CRISPR adaptation in transformant colonies was performed as described previously [254]. The reseeded experiments were performed as described in 2.2.4 *Replating of transformants* section. To test the escape mutations, phage DNA from 9 randomly chosen colonies was isolated using GeneJET Plasmid Miniprep Kit (Thermo scientific) and sequenced with Illumina. The replicative form of phage M13 from initial mixture was used as a control sample. The

sequenced reads were trimmed using Trimmomatic v. 0.39 and mapped to the reference of the M13mp18 genome. The escape mutations were visualized with IGV browser [261] and R packages.

2.3 Results

2.3.1 Colonies formed after transformation of protospacer plasmids into cells mounting CRISPR interference contain cells with active CRISPR-Cas and unchanged plasmids

The *E. coli* KD263 cells that contain inducible *cas* genes and a CRISPR array with a single g8 spacer [253, 254] were grown in the presence or in the absence of *cas* gene expression inducers and transformed with pG8, a pUC-based plasmid containing the g8 protospacer with an interference-proficient ATG PAM [125]. After transformation, cells without *cas* gene induction (CRISPR OFF) were plated on a medium supplemented with ampicillin to select colonies formed by plasmid-bearing cells. Pre-induced CRISPR ON cells were plated on a medium that contained both ampicillin and inducers of *cas* genes expression (**Fig. 2.1a**). Compared to CRISPR OFF cells, approximately 200 times less ampicillin-resistant colonies were formed after the transformation of CRISPR ON cells (**Fig. 2.1b**). No difference in the number of transformants was observed when a plasmid vector without the g8 protospacer was used for transformation, indicating that the drop in transformation efficiency was due to CRISPR interference mounted when the g8 protospacer in pG8 was recognized by the Cascade effector charged with the g8 crRNA. A similar experiment with another plasmid, pRSFG8, which provides resistance to kanamycin, showed similar results: ~50-80 times less transformant colonies were formed by CRISPR ON cells compared to CRISPR OFF cells (**Fig. 2.1b**). With both plasmids, antibiotic-resistant colonies formed by CRISPR ON cells were visually indistinguishable from CRISPR OFF cells colonies.

To test whether plasmids in CRISPR ON colonies escaped interference by accumulating mutations, plasmids from ten randomly chosen individual CRISPR ON colonies were purified and retransformed into CRISPR ON and CRISPR OFF competent cells. In every case, a drop in transformation efficiency into

induced cells was the same as that observed during the original transformation experiment (Fig. 2.1c). We therefore conclude that plasmids present in CRISPR ON colonies are subject to interference by CRISPR effector charged with g8 spacer crRNA and in this respect are identical to plasmids used in the original experiment. Consistently, sequencing of the protospacer region from plasmid prepared from pooled CRISPR ON colonies did not reveal differences from the pG8 sequence (**Fig. S2.1a**). PCR analysis of CRISPR adaptation in CRISPR ON colonies showed no acquisition of new spacers in the CRISPR array (**Fig. S2.1b**), which is also consistent with the absence of escape mutations in g8 protospacer/PAM, that would be expected to stimulate primed adaptation [128, 254].

To determine whether CRISPR ON cells forming colonies on selective medium contain a functional CRISPR-Cas system, competent cells were prepared from CRISPR ON transformants and transformed with compatible plasmids carrying the g8 protospacer and a different antibiotic resistance marker. Cells derived from pG8-transformed CRISPR ON colonies interfered with pRSFG8 transformation as efficiently as induced control plasmid-less KD263 cells (**Fig. 2.1d**). The same situation was observed when competent cells derived from pRSFG8-transformed CRISPR ON colonies were transformed with pG8 (**Fig. 2.1d**). We therefore conclude that CRISPR ON colonies transformed with plasmids carrying a protospacer matching crRNA spacer are formed by cells with a functional CRISPR-Cas system.

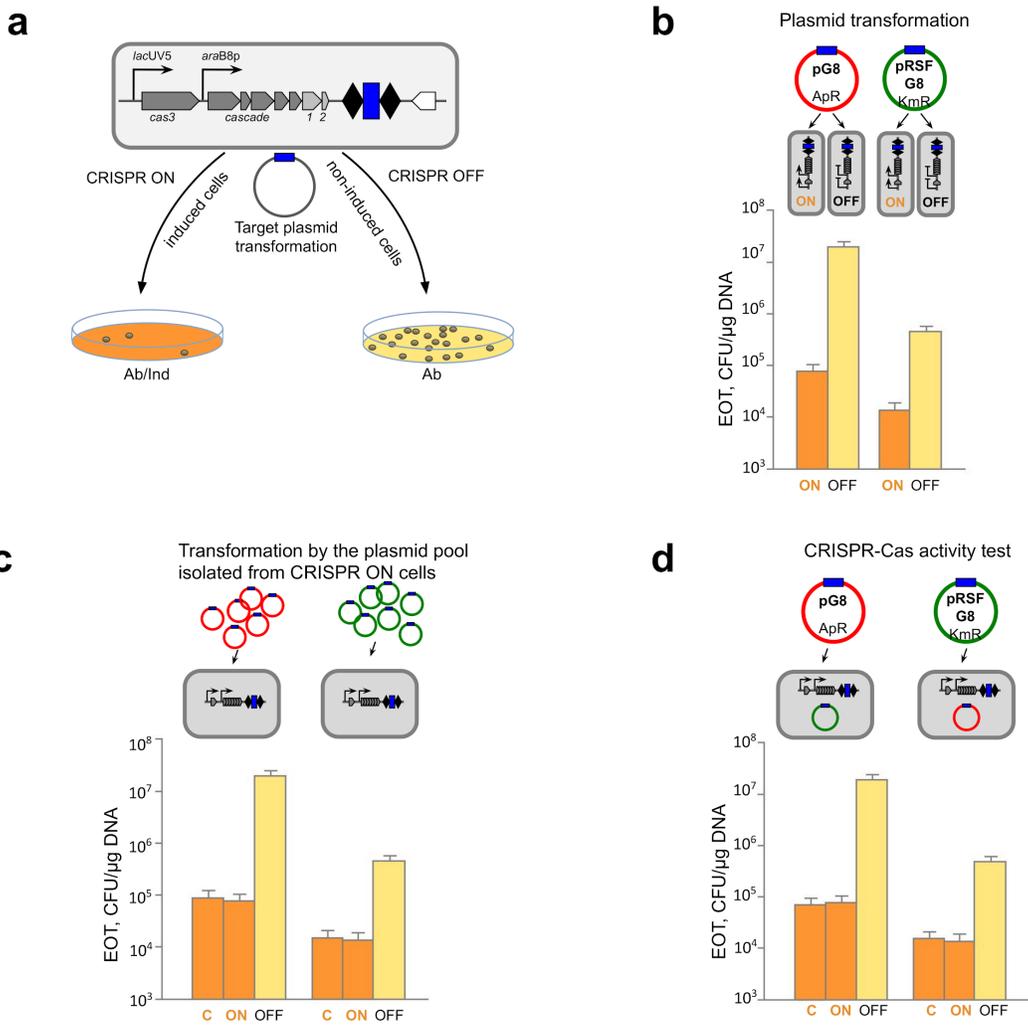


Figure 2.1. Cells forming colonies on selective media under CRISPR ON conditions contain plasmids that are subject to interference by the functional CRISPR-Cas system.

a, An *E. coli* KD263 cell harboring *cas* genes controlled by inducible promoters and a CRISPR array with two repeats (black rhombi) and a single *g8* spacer (blue rectangle) is schematically shown at the top. Cells are grown in the presence or in the absence of *cas* gene expression inducers to prepare, correspondingly, CRISPR ON and CRISPR OFF competent cells, which are transformed with a plasmid bearing the *g8* protospacer (shown as a blue rectangle on a circle representing the plasmid, fully matches the *g8* spacer) with a functional PAM. After transformation, CRISPR ON cells are plated on a medium supplemented with *cas* genes inducers and an appropriate antibiotic (Ab/Ind); CRISPR OFF cells are plated on a medium containing only the antibiotic (Ab).

b, CRISPR ON and CRISPR OFF cells were transformed with ampicillin-resistant pG8 or kanamycin-resistant pRSFG8 plasmids bearing the *g8* protospacer (blue rectangle) and efficiency of transformation (EOT) was determined as CFUs per μ g of plasmid DNA. Bars show mean EOTs from three independent experiments. Standard deviations of the mean are indicated.

c, Plasmids purified from CRISPR ON colonies transformed with either pG8 or pRSFG8 were retransformed into CRISPR ON or CRISPR OFF competent cells and EOT was determined. Transformation of CRISPR ON cells with initial pG8 and pRSFG8 plasmids was used as a control (“C”). Bars show mean EOTs from three independent experiments. Standard deviations of the mean are indicated.

d, Competent cells prepared from cells from CRISPR ON colonies transformed with pG8 or pRSFG8 were transformed with compatible g8 protospacer plasmids (cells bearing pG8 were transformed with pRSFG8 and vice versa). As a control, transformation of plasmid-less CRISPR ON cells with pG8 and pRSFG8 plasmids was performed (“C”). Bars show mean EOTs from three independent experiments. Standard deviations of the mean are indicated.

2.3.2 Cells from CRISPR ON colonies contain fewer plasmids than CRISPR OFF colony cells

Quantitative PCR with plasmid-specific primers was used to determine the plasmid copy number (PCN) in cells from CRISPR ON and CRISPR OFF colonies. Amplification of the chromosomal *gyrA* gene was used for normalization (see Methods). On average, there were 233 ± 46 copies of pG8 per cell in CRISPR OFF colonies (**Fig. 2.2b**, top row panel), which is consistent with PCN values for the pUC vector on which pG8 is based [262]. For pRSFG8, an average value of 119 ± 21 copies per CRISPR OFF colony cell was calculated (**Fig. 2.2c**, top row panel), which is also consistent with published data [262, 263]. In contrast, cells from CRISPR ON colonies had an average PCN of 0.18 ± 0.06 (for pG8) and 0.71 ± 0.27 (for pRSFG8) (**Figs. 2.2b** and **2.2c**, top row panels).

The below 1 PCN value indicates that many cells in CRISPR ON colonies are plasmid-free and the colonies should thus be heterogeneous. To determine the ratio of plasmid-bearing and plasmid-free subpopulations in colonies formed at different conditions, we replated cells from randomly chosen CRISPR ON and CRISPR OFF colonies transformed with pG8 or pRSFG8 on three types of media (**Fig. 2.2a**). Plating on a medium with no *cas* gene expression inducers and without an antibiotic determined the total number of viable cells. Plating on a medium supplemented with an appropriate antibiotic determined the number of viable plasmid-bearing cells. Plating on a medium supplemented with *cas* gene expression inducers and an appropriate antibiotic allowed us to determine whether cells from CRISPR ON colonies that carried a plasmid were losing it during growth under conditions of continued CRISPR interference.

For CRISPR OFF transformants, the number of colonies formed upon reseeded on plates with and without antibiotics was the same (**Figs. 2.2b** and **2.2c**, right panels in the second row) indicating that both pG8 and pRSFG8 are stably maintained in the absence of antibiotic selection. In contrast, only one out of a few thousand reseeded cells from CRISPR ON colonies grew on antibiotic-containing plates (**Figs. 2.2b**

and **2.2c**, left panels in the second row). Thus, most cells in CRISPR ON colonies indeed lost the plasmid and must have survived due to the presence of a minor fraction of plasmid-bearing cells that decreased antibiotic concentration within the colony.

The number of colonies formed by cells from CRISPR ON colonies on plates supplemented with both *cas* gene expression inducers and an antibiotic was further decreased 5-10-fold compared to the number of colonies grown on plates with antibiotic only (compare **Figs. 2.2b** and **2.2c**, left panels in the second row). This indicates that CRISPR interference continues to purge plasmids from CRISPR ON plasmid-bearing cells, albeit at an efficiency that is considerably lower than that observed during plasmid transformation.

The number of colonies observed after reseeded cells from CRISPR OFF colonies on a medium containing both the *cas* genes inducers and an antibiotic was the same as that on the medium with antibiotic only or without any additions (**Figs. 2.2b** and **2.2c**, right panels in the second row). This result seems to indicate that interference against an established plasmid is inefficient. Yet, quantitative PCR showed that per cell PCN values for colonies formed upon reseeded of original CRISPR OFF colonies on media with *cas* genes inducers and an antibiotic were as low as those for initial CRISPR ON transformants (**Figs. 2.2b** and **2.2c**, left panels in the second row). In contrast, PCN per cell for colonies formed on plates containing antibiotic only was as high as in the corresponding CRISPR OFF colonies, implying that PCN restores to normal levels in the absence of CRISPR interference.

The second round of reseeded confirmed that most cells in colonies originating from CRISPR OFF colonies have lost plasmids after growth at conditions of *cas* genes expression induction (**Figs. 2.2b** and **2.2c**, right panels in the third row). In addition, the proportion of plasmid-bearing cells was further decreased when cells from a CRISPR ON colony formed after the first reseed were replated in the presence of *cas* genes inducers and an antibiotic. The effect was ~100 fold for pG8 colonies and less pronounced for pRSFG8 colonies. The third reseed demonstrated the same proportions of plasmid-bearing cells as in the second reseed, indicating that plasmid bearing cells persist at CRISPR ON conditions for many generations (**Figs. 2.2b** and **2.2c**, bottom row).

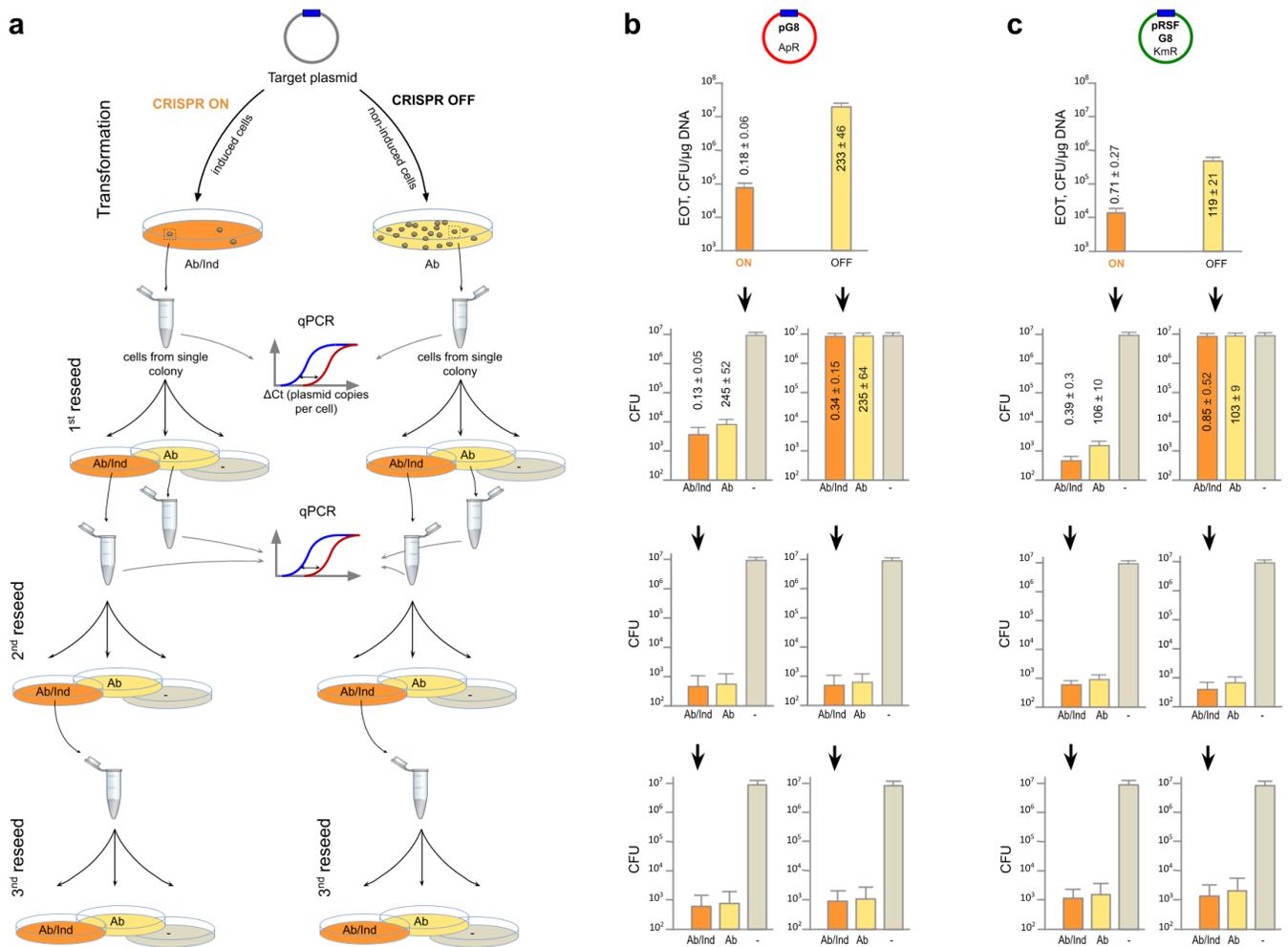


Figure 2.2. Colonies formed at CRISPR ON conditions mostly contain plasmid-less cells.

a, Cells from CRISPR ON and CRISPR OFF colonies obtained as in Fig. 2.1 are reseeded on media supplemented with antibiotic and inducers (“Ab/Ind”, orange), antibiotic only (“Ab”, yellow), or plates with neither inducers nor antibiotics (“-”, grey). Cells from colonies formed on Ab/Ind plates after the first reseed are reseeded the second time, and cells from colonies formed on Ab/Ind plates after the second reseed are reseeded the third time. Real-time PCR is used to determine PCN per cell in original transformants and in colonies formed after the first reseed. **b** and **c**, Experiment was done as outlined in panel **a** using the pG8 and pRSFG8 plasmids, correspondingly. The bars show mean colony numbers formed at each indicated condition from three independent experiments. Standard deviations of the mean are indicated. Numbers within bars for original transformants and after the first reseed show mean PCN values/SD deviations calculated based on real-time PCR measurements for three independent colonies taken from each plate.

To demonstrate that there is no significant influence of antibiotic on plasmid persistence, we performed parallel cultivation of cells from the same CRISPR ON colony bearing the pG8 plasmid in liquid medium supplemented with both *cas*-gene expression inducers and the appropriate antibiotic or with *cas*-gene expression inducers only. After 24 h cultivation, the fraction of plasmid-bearing cells was estimated by plating on media with and without antibiotic (**Fig. S2.2a**). The ratio of plasmid-bearing cells (i.e., cells able to form colonies on antibiotic-supplemented plates) was ca. 10-fold less in induced cultures grown in the absence of antibiotic compared to cultures grown in its presence (**Fig. S2.2a**). Several individual colonies formed on plates supplemented with *cas* genes inducers and antibiotic were randomly picked and analyzed for the integrity of the plasmid and the CRISPR-Cas system as was done in experiment shown in **Figure 2.2**. The results showed that plasmids extracted from CRISPR ON colonies formed after growth with or without antibiotic selecting for plasmid bearing cells retained wild-type protospacers (**Fig. S2.2b**) and that CRISPR-Cas system remained functional in cells from these colonies (**Fig. S2.2c**). Thus, plasmids persist in a small fraction of cells under ongoing CRISPR interference even in the absence of selection for plasmid maintenance. The ~10-fold decrease in the number of plasmid-bearing cells in induced cultures grown in the absence of antibiotic is likely due to decreased fitness of plasmid-bearing cells at these conditions. To test this hypothesis we determined growth rates of plasmid-free and plasmid-bearing *E. coli* cells (**Fig. S2.2d**). Indeed, we observed that plasmid-bearing cells grew slower than plasmid-free cells in the absence of antibiotic (**Fig. S2.2d**). In the presence of antibiotic the growth of plasmid-free cells is suppressed, which increases the probability to detect plasmid-bearing cells in a population.

2.3.3 Direct observation of plasmid-bearing cells in CRISPR ON colonies

Plasmid pG8-GFP, a derivative of pG8 carrying a constitutively expressed green fluorescent protein TagGFP2 gene, was created to allow direct observation of plasmid-bearing cells. Similar to pG8, the pG8-GFP plasmid was subject to CRISPR interference as evidenced by a ~200-fold decrease in the number of colonies formed at CRISPR ON conditions compared to CRISPR OFF conditions. While all CRISPR OFF colonies transformed with pG8-GFP were highly fluorescent when irradiated with a handheld UV lamp, most CRISPR ON colonies were dim. Only 1-3% of all CRISPR ON colonies fluoresced (**Fig. 2.3a**). Retransformation experiments conducted as in **Fig. 2.1c** revealed that plasmids from these rare colonies

did not contain escape mutations. However, whole genome sequencing of DNA extracted from two randomly chosen fluorescent CRISPR ON colonies revealed frame-shift mutations in the *cseI* gene and/or in the *araBp8* promoter from which the *cas* operon is transcribed (Table S2.1).

The individual dim CRISPR ON and CRISPR OFF colonies were analysed using confocal and wide-field fluorescence microscopy at high magnification (Fig. 2.3b,c and Fig. 5.1). In the CRISPR OFF colonies, all cells were brightly fluorescent. In contrast, only a small fraction of cells possessed detectable fluorescence in the CRISPR ON colonies. Radially extended serpentine-shaped rows of fluorescent cells on the background of non-fluorescent plasmid-less cells that are clearly seen at the colony edges are consistent with inherited maintenance of plasmids in some lineages within the colony. The absence of expanding fluorescent sectors seems to suggest that most cells in such lineages lose the plasmid with time (Fig. 2.3b).

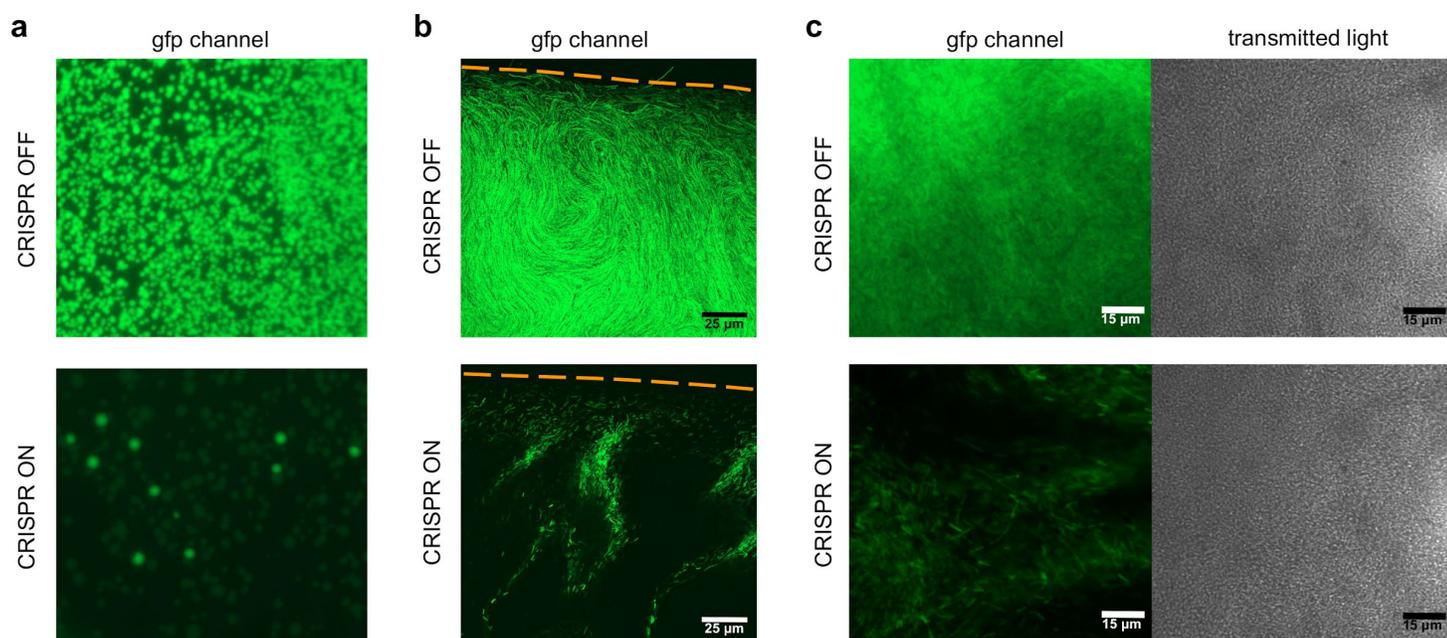


Figure 2.3. Fluorescence microscopy of *E. coli* KD263 colonies.

a, Images of fragments of plates containing CRISPR ON and CRISPR OFF transformant colonies. Note that the majority of CRISPR ON colonies are weakly fluorescent; their internal structure at a higher magnification is shown in panels **b** and **c**. Highly fluorescent CRISPR ON colonies have a non-functional CRISPR-Cas system (Table S2.1). **b**, Images of a CRISPR OFF and a dim CRISPR ON colony periphery obtained using confocal microscopy in the gfp channel. Orange dashed lines show colony edges. **c**, Wide-field fluorescence microscopy of CRISPR OFF and dim CRISPR ON colonies.

Our results show that most CRISPR ON colonies (each derived from a single founder cell transformed with the plasmid bearing the target protospacer) are heterogeneous and most cells in such colonies are either completely or nearly plasmid-less. Clearly, for continued colony growth on selective medium there must be at least one uninterrupted line of plasmid-bearing cells that persists through multiple generations. To study the distribution of plasmid-bearing fluorescent cells we analysed cells from dim CRISPR ON colonies transformed with pG8-GFP using flow cytometry (**Fig. 2.4a**). Cells transformed with the pG8 plasmid bearing no fluorescent marker were used to define the level of autofluorescence. Cells derived from pG8-GFP transformed CRISPR OFF colonies were used as a positive control. The results showed that about 10% of cells from CRISPR ON colonies were fluorescent (**Figs. 2.4b,c**). However, the mean intensity of fluorescence of such cells was ~7 times less than that of positive control cells. Detailed statistics demonstrated that PCN (assumed here to be directly proportional to fluorescence) in plasmid-bearing fraction of CRISPR ON cells reached an upper limit corresponding to 0.3-0.4 of that in CRISPR OFF cells (**Fig. 2.4c**).

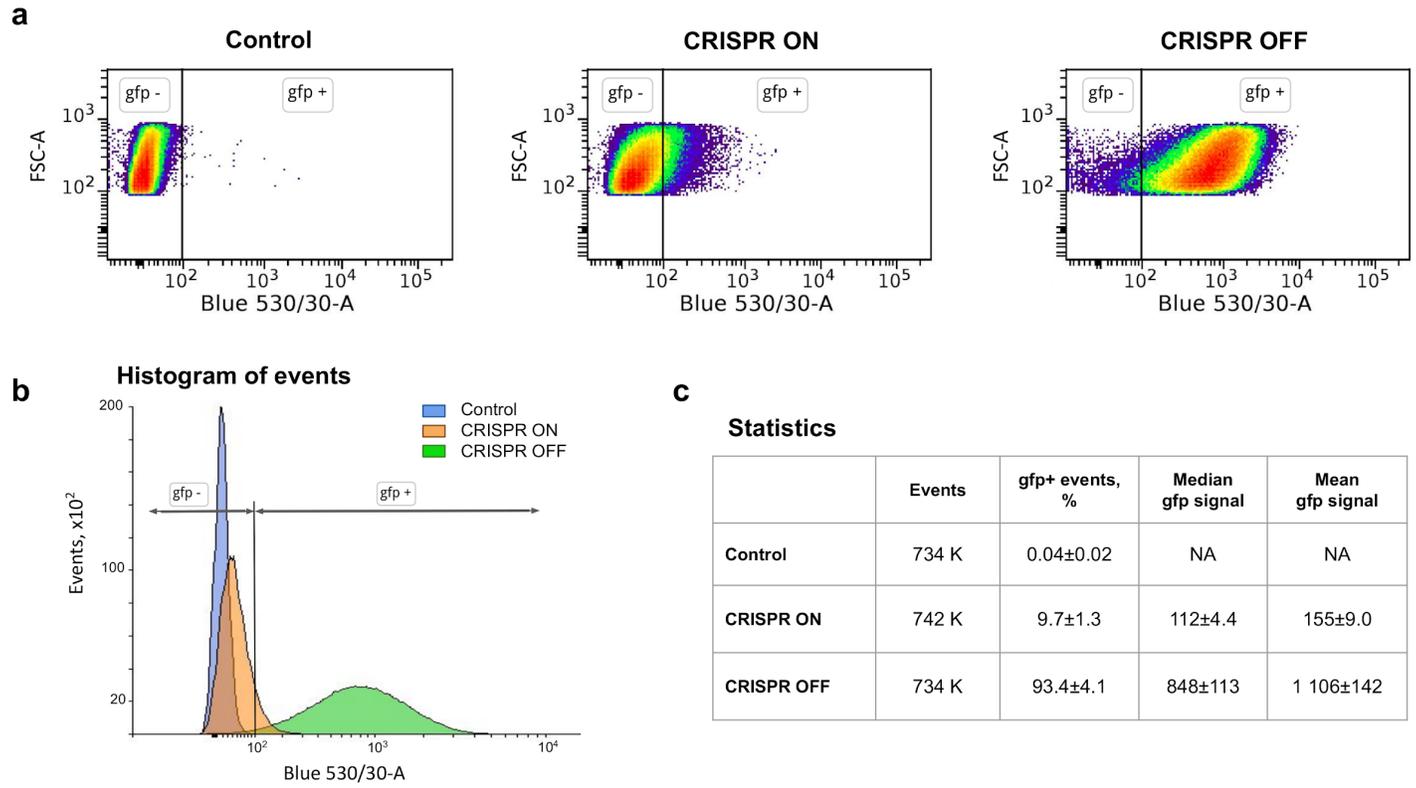


Figure 2.4. Flow cytometry analysis of cells from CRISPR ON and CRISPR OFF colonies. **a**, Density plots in forward scattering (FSC-A) and green fluorescence (Blue 530/30-A) channels. Vertical black lines in all panels show borders between gfp- gate consisting of non-fluorescent events and gfp+ gate consisting of fluorescent events. The control panel shows cytometry results of cells bearing the pG8 plasmid (without GFP) as a negative control. **b**, Histogram of the distribution of fluorescence levels in cells from indicated colonies. **c**, Statistics of flow cytometry results.

2.3.4 Direct real time observation of plasmid loss due to CRISPR interference

To observe plasmid loss caused by CRISPR interference in real time, cells from CRISPR OFF colonies transformed with pG8-GFP were used to seed microfluidic growth chambers (Fig. S2.4). Growth chambers seeded with single cells were observed for 7 hours in the presence or in the absence of *cas* gene expression inducers in the medium flowing through the main channel of the microfluidic device. No antibiotic was added. As expected, in the absence of *cas* genes expression inducers, the founder cells divided and all progeny remained highly fluorescent (Fig. 2.5a and Figs. S2.5, S2.6). In contrast, in the presence of inducers, progeny cells remained fluorescent only until the fifth division (Fig. 2.5b). Afterwards, most

cells ceased to fluoresce, presumably due to the earlier loss of plasmid caused by CRISPR interference and dilution of the GFP protein present in the founder cell and its immediate descendants. However, some cells retained fluorescence. Though some descendants of such fluorescent cells subsequently lost fluorescence, others formed lineages of fluorescent cells that persisted during the time of the experiment. The schematic tree of cell divisions illustrates that one single cell is able to generate a branch of fluorescent cells, as well as multiple branches of offspring that lost fluorescence (**Fig. 2.5c**).

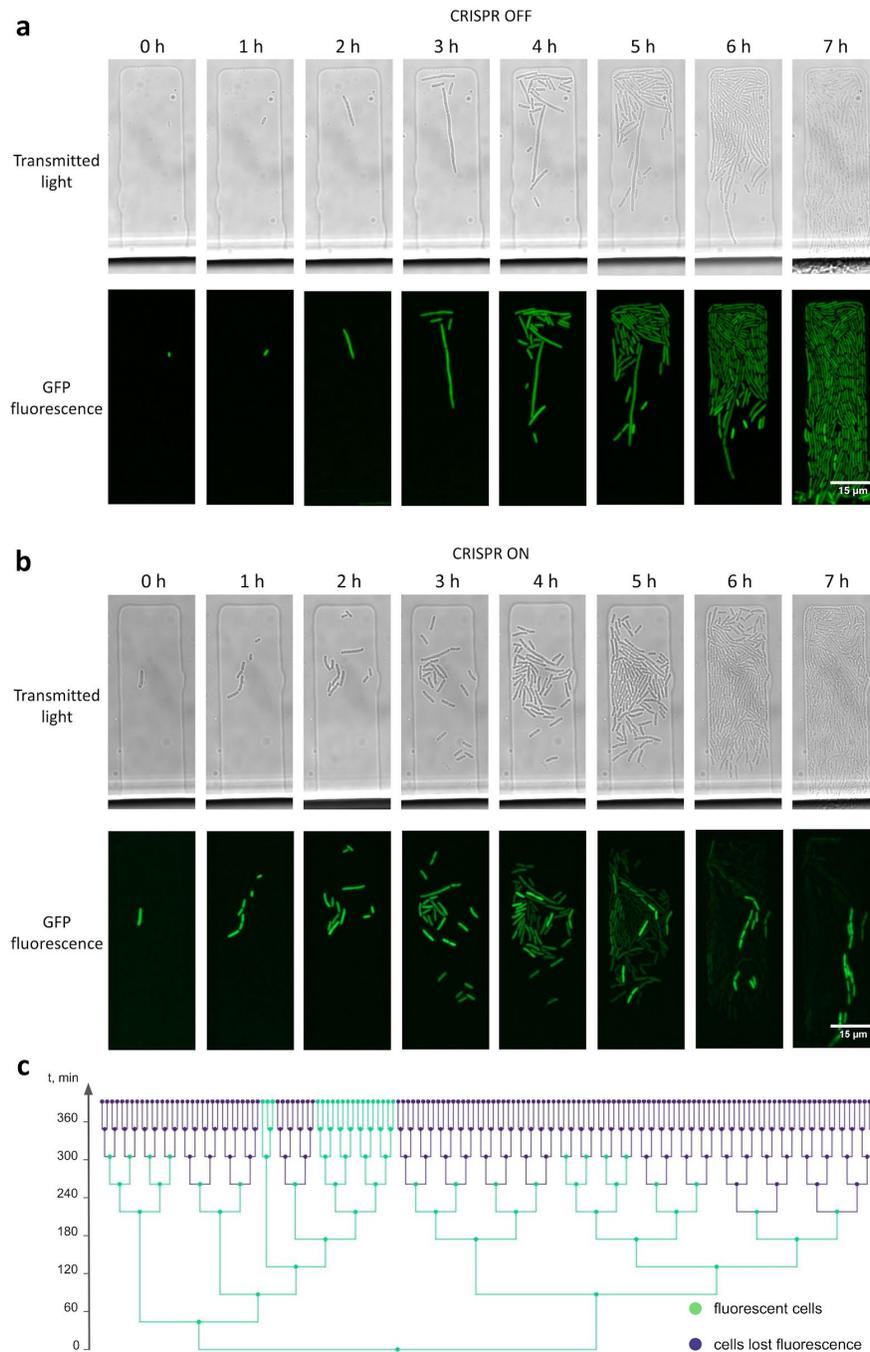


Figure 2.5. Live fluorescence microscopy of cells in a microfluidics device.

a, The divisions of a single CRISPR OFF cell bearing the pG8-GFP plasmid over time in a microfluidics chamber supplemented with LB medium. **b**, As in **a** but in a medium containing *cas* gene expression inducers. **c**, A tree of cell divisions depicts the loss of fluorescence through cell generations in a microfluidics chamber shown in panel **b**. The length of the branches schematically illustrates the time between subsequent divisions.

2.3.5 Discussion and theoretical analysis (*in collaboration with A. Martynov*)

In this work, we show that plasmids can survive in cells mounting an active CRISPR interference response against them. We experimentally ruled out possibilities such as inactivation of CRISPR-Cas system or accumulation of escape mutations in such cells. While only a small fraction of cells retains plasmids under the ongoing pressure from the CRISPR-Cas system, plasmids remain in these cells and their descendants for many generations, such that apparently healthy colonies are formed on selective media containing antibiotics in concentrations sufficient to completely prevent growth of cells without plasmids.

Plasmids providing resistance to ampicillin rely on β -lactamase secreted in the periplasm to degrade the antibiotic outside the cell [254, 265]. The phenomenon of indirect resistance where ampicillin-resistant colonies decrease the concentration of ampicillin in the medium and support the growth of susceptible satellite colonies is well known [266, 267]. In our experiments with pG8 and pG8-GFP plasmids, such indirect resistance is apparently responsible for the observed CRISPR ON colonies heterogeneity with a small number of plasmid-bearing cells supporting the growth of a much larger number of cells that have lost the plasmid. At the same time, previous studies demonstrated the absence of indirect resistance for kanamycin [267]. Thus, the mechanism of growth of CRISPR ON colonies with pRSFG8 remains unclear and may involve specific three-dimensional arrangement of cells within a colony [266, 268].

While all our experiments clearly demonstrate that the majority of cells in CRISPR ON colonies consist of cells completely devoid of plasmids and the average PCN in rare plasmid-bearing cells is few-fold less than that in CRISPR OFF cells, the measurements made by replating, flow cytometry, and in microfluidic device agree with each other only qualitatively. The possible reasons for such discrepancies could lie in difference in time elapsed after transformation, varying levels of *cas* gene induction, and other more specific distinctions between experimental setups. These differences notwithstanding, we below suggest an explanation to the apparently probabilistic and history-dependent response of PCN to CRISPR interference based on a simple stochastic model of plasmid replication and interference inside a cell.

For multicopy plasmids used here, when only one or a few plasmids are present in the cell, both the interference and replication kinetics should be limited by plasmid concentration. In this case the per plasmid rates of both processes are constant, i.e. independent of PCN, while the corresponding per plasmid population rates depend on PCN linearly. Yet when the PCN is large and close to the stationary number

of plasmids $[PI]_{st}$, the replication rate should approach zero. Similarly, for all reasonable forms of CRISPR interference kinetics, an increase in the number of plasmids should result in a progressively smaller increase in the interference rate and its eventual saturation to a constant when the concentration of plasmids becomes high. Three scenarios that satisfy these general constraints are possible:

- The replication rate is always lower than the interference rate (**Fig. 2.6a**).
- The replication rate is higher than the interference rate until a certain threshold plasmid copy number is reached (**Fig. 2.6b**).
- There is a range of PCN values for which the replication rate exceeds the interference rate; beyond this range the interference rate is higher (**Fig. 2.6c**).

The first scenario leads to quick loss of plasmids in all cells; the second results in survival of plasmids in the majority of cells at an equilibrium PCN $[PI]_{eq}$ that is less than that in the absence of interference. The third scenario explains prolonged survival of plasmids in a small fraction of cells observed in our experiments. Since all transformed cells initially have just one copy of a plasmid, most lose it since at low PCN the interference rate is higher than the replication rate. However, due to an intrinsic stochasticity of interference and replication events, there is a small but finite possibility that in some transformed cells plasmid replication events happen more often than interference events. If such a favourable (for plasmid) situation occurs, the PCN may go over a “bifurcation threshold” (marked as $[PI]_{bif}$ in **Fig. 2.6c**), above which the replication rate exceeds the interference rate. From this point on, plasmids will likely survive and continue to expand deterministically until reaching $[PI]_{eq}$. In the Methods section, we outline quantitative analysis of this survival scenario, which confirms that the qualitative arguments presented above indeed explains the persistence of plasmids in some cell lineages despite the ongoing CRISPR interference. Our analysis is based on a numerical solution of the master equation that describes time evolution of the probability $P_n(t)$ for a cell to have n plasmids at time t . The master equation accounts for plasmid replication and interference processes, which are assumed to follow the Logistic dynamics and Michaelis-Menten kinetics, and the binomial partition of plasmids between daughter cells upon mother cell division. The results of the master equation solution for two initial conditions, a cell with a plasmid initially present in a single copy or a cell with a stationary PCN $[PI]_{st} = 100$ are presented in **Fig. 2.6e,f**. The competition between interference and plasmid replication produces two cell subpopulations, one having a substantial number of plasmids distributed around $[PI]_{eq}$, and another completely devoid of

plasmids. The probability for a cell to retain plasmids quickly drops in the first few generations and levels after 5-10 generations. It follows from our model that the fraction of cells that lose (and, reciprocally, retain) plasmids after this initial transitory period depends on the initial plasmid number in a cell with an active CRISPR system. However, the distribution of PCN in cells that retain plasmids converges after ~ 10 generations to a universal form, which does not depend on the initial number of plasmids and is determined solely by the kinetics of interference and replication. The universal distribution is shown by pale blue lines in **Fig. 2.6e** for a single initial plasmid per cell and in **Fig. 2.6f** for $[PI]_{st} = 100$ plasmids per cell (the fraction of cells with plasmids is much larger for the case with multiple initial plasmids (**Fig. 2.6f**) than when there is a single plasmid at the beginning of the process (**Fig. 2.6e**), hence the pale blue line in **Fig. 2.6f** is higher than that in **Fig. 2.6e**). Note that the subsequent fate of the plasmid-free cells and their resulting stationary concentration depend on many environmental factors and metabolic costs of maintaining the plasmids (**Fig. S2.2**) and is not considered in this model.

Since the average PCN converges to a rather large number $[PI]_{eq}$ (which is independent of the initial conditions), the subsequent probability to lose all plasmids becomes quite low, and such cells with plasmids form colonies that survive indefinitely on antibiotic medium. Thus, our model recapitulates three key experimental observations:

- Under pressure from CRISPR-Cas initially clonal cellular populations become bimodal, consisting of the subpopulations with and without plasmids.
- The fraction of cells that retain plasmids is affected by the distribution of plasmids that exists before CRISPR interference commences.
- The distribution of plasmids under pressure from CRISPR-Cas depends solely on the nature of CRISPR-Cas and plasmids, rather than on the initial plasmid distribution.

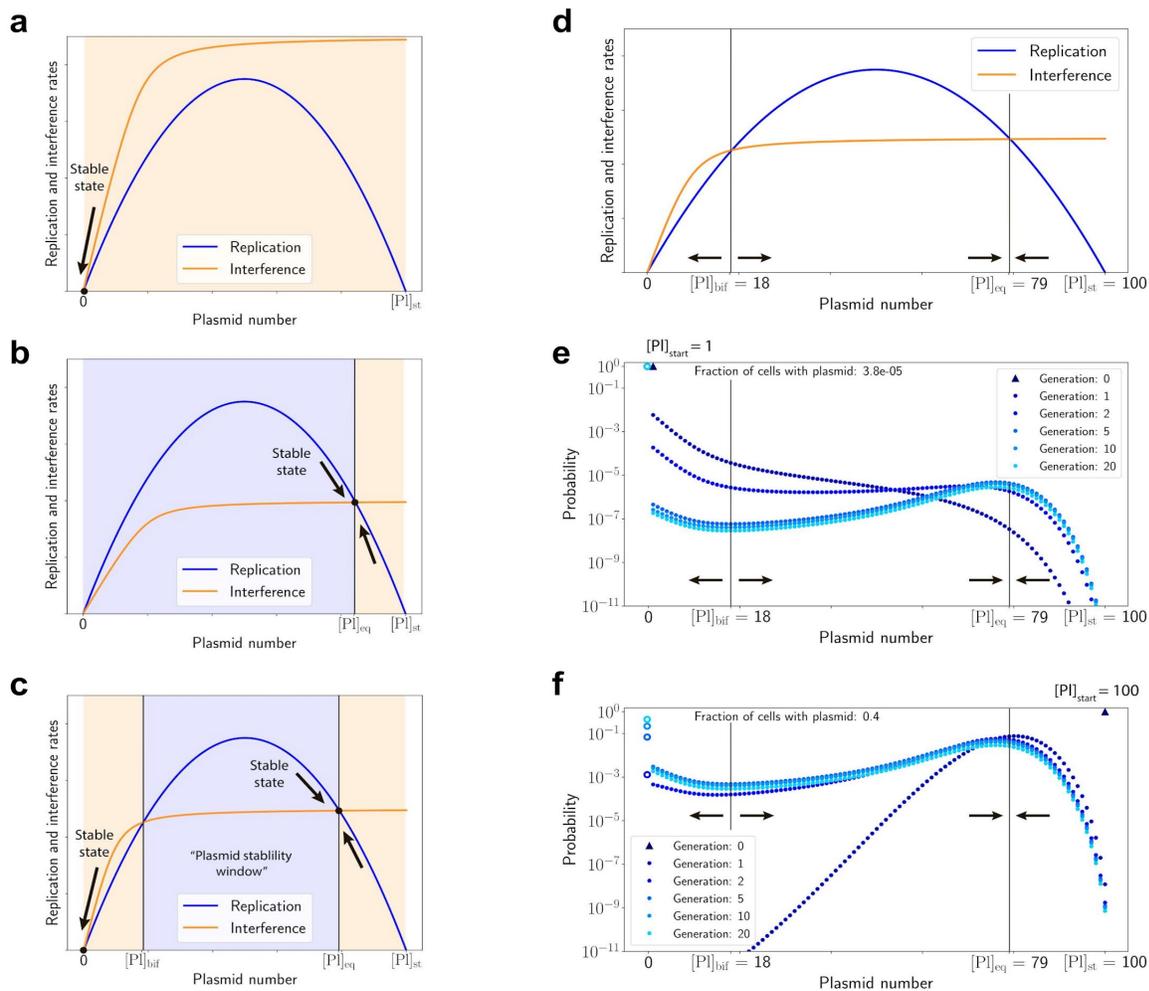


Figure 2.6. Plasmid copy number dynamics in a CRISPR ON cell.

a-c, Three possible scenarios between plasmid replication (blue lines) and CRISPR interference (orange lines) rates: **a**, The interference rate is higher than the replication rate for any PCN, plasmids quickly become extinct. **b**, The replication rate is higher than the interference rate, PCN quickly reaches an equilibrium point $[PI]_{eq}$. **c**, There exists an intermediate range of PCN values, $[PI]_{bif} < [PI] < [PI]_{eq}$, where the replication rate is higher than the interference rate; beyond this range the interference dominates. Ranges of PCN where replication or interference rates dominate are shown by blue or orange shading, respectively. **d**, The replication (blue line) and interference (orange line) rates used in the solution of the master equation (5), parametrized as the Logistic (Eq. (1)) and Michaelis-Menten (Eq. (3)) kinetics. **e**, The probability $P_n(t)$ for a cell to have n plasmids just before their partition between two daughter cells after 1, 2, 5, 10, and 20 generations shown by dots of varying shades of blue. Initially, a single plasmid was introduced into a cell, which is marked by a triangle in the upper left corner. Empty circles, also marked by shades of blue of the corresponding generations, show the fraction of cells that lost all plasmids. **f**, Same as in **e**, but for the initial number of plasmids equal to $[PI]_{st}$, marked by a triangle in the upper right corner. After ≈ 10 generations, the probability distribution $P_n(t)$ converges to the universal form, shown by a pale blue line in panels (e) and (f). The parameters used in these solutions are listed in the Methods section. Vertical black lines in panels **d-f** show stable and unstable fixed points of PCN dynamics, their stability is shown by converging and diverging arrows.

Obviously, the model fidelity can be improved by utilizing experimentally-derived dependencies of rates vs. PCN, or fitting the Michaelis-Menten and Logistic constants to the experimental data. However, we believe that shape of the rate curves, illustrated in **Fig. 2.6d**, is universal. Since the stochastic survival of plasmids goes "against the odds" dictated by an excess of the CRISPR interference rate over the plasmid replication rate, the fraction of cells that retain plasmids falls dramatically with an increase of the rate-reversal threshold $[PI]_{\text{bif}}$ (shown in **Fig. 2.6d-f** by a left vertical line). So a fairly delicate balance between the Michaelis-Menten and plasmid replication rate constants, which determine $[PI]_{\text{bif}}$, is required to observe the reported plasmid survival in a small fraction of cells. Within our model, the outcome of the plasmid-CRISPR conflict at a single-cell level is purely random and all that can be predicted for a given cell is its probability to lose all or retain a certain number of plasmids. A possible determinant of fate of plasmids in a cell could be the level of Cas proteins and plasmid replication machinery enzymes, which themselves fluctuate inside a cell and vary between cells. Evidently, cells with above average concentration of plasmid replication machinery enzymes and below average concentration of Cas complexes will have a higher probability to retain plasmids, and vice versa. A "double-stochastic" model that takes into account not only the randomness of replication and interference events, but also fluctuations in levels of effector complexes and plasmid replication machinery components could provide even more realistic predictions. Furthermore, this model can serve as a building block in a more comprehensive multilevel description that predicts spatial structure and describes population dynamics of plasmid-bearing and plasmid-free cells in a realistic colony with or without antibiotics.

A common way to escape CRISPR interference by phages is the acquisition of mutations in the targeted protospacer or its PAM, which decreases and/or abolishes effector complex affinity [254, 269]. Indeed, phage plaques formed on lawns of cells identical to the ones used here with CRISPR targeting various phages are formed by such escaper phages [270]. Yet, no escape plasmids are found in colonies formed upon transformation under CRISPR ON conditions [252], even though escape plasmids created in the laboratory are efficiently transformed and are not subject to CRISPR interference [112, 254]. We hypothesize that the difference between phage and plasmid reactions to an ongoing CRISPR interference is due to the fact that the former but not the latter are cell-autonomous in a sense that they can be released from the infected cell and then reinfect surrounding cells. Thus, during formation of a plaque (a negative colony) multiple reinfections take place which allows rare phages that acquire escape mutations to take

over the population, such that the final plaque contains almost exclusively mutant phages [254]. In contrast, during formation of a bacterial colony by a founder cell transformed with a plasmid, plasmids are only passed vertically from parent to daughter cells. While an infected cell usually receives just one copy of the phage genome, a daughter cell inherits on average half of plasmids from its mother. Hence, if the probability for a phage to survive and replicate in a cell (which is quite low in the case of a single initial plasmid considered here), multiplied by the size of phage burst is much smaller than one, the infection would not propagate and we would simply register an apparent defeat of a phage by CRISPR-Cas.

A limited rate of CRISPR interference resulting in an incomplete extermination of phages and plasmids can be a consequence of simple evolutionary principles. Increasing the rate of interference costs the cell not only extra energy to produce additional copies of Cas proteins but can also lead to off-target DNA cleavage, causing autoimmunity. Thus, the evolutionary optimization of CRISPR-Cas interference rate would probably not go beyond some intermediate protection level, which eliminates foreign mobile genetic elements in most but not in every cell in the population. The stochasticity of CRISPR interference could allow conditionally favorable plasmids to take a hold in a subpopulation of cells and then proliferate when the environment selects [271]. Such effects can be especially prominent in structured environments offering specific niches to subpopulations differing in their genetic or physiological states [272, 273]. Bacterial colonies and biofilms are not homogeneous and contain micro-environments with complex spatial structures [274, 275] that may support, among other things, cooperative antibiotic [265, 266, 276] or phage resistance [277–279]. Some of such complex spatial structures may be similar to those microscopically observed in CRISPR ON colonies in our experiments and depend on a dynamic interplay between mobile genetic elements replication and host defence directed against them.

2.3.6 Transcriptome assay of CRISPR ON cells and the nature of acquisition escape mutation by M13 phages against CRISPR-Cas system (*unpublished results*)

We demonstrated that CRISPR ON colonies survive on antibiotic media and consist of plasmid-bearing and plasmid-free cells. To understand a potential mechanism providing survivability for the plasmid-free fraction of the CRISPR ON colony, we additionally performed transcriptome analysis of CRISPR ON and CRISPR OFF *E. coli* colonies obtained after transformation using the plasmid pRSFG8, as described in *section 2.2.12*. We randomly selected three CRISPR ON and three CRISPR OFF colonies and cultivated the selected colonies overnight in LB media containing both kanamycin and inducers for CRISPR ON cells, and kanamycin only for CRISPR OFF cells as a control. After overnight cultivation, we extracted RNA from the CRISPR ON and CRISPR OFF *E. coli* cells for sequencing, as described in *section 2.2.12*. We investigated the transcriptome results and underlined changes in gene expression (Supplementary Information, **Fig. S2.9**). We emphasized the up-regulated genes such as *ibpA*, *ibpB*, *marA*, *marB*, *mqsA*, *pspA*, and *pspB* genes. In *E. coli* cells the products of genes *ibpA* and *ibpB* stabilize and protect aggregated proteins from irreversible denaturation and extensive proteolysis during heat shock and oxidative stress [280]. MarA is a transcriptional activator of genes involved in the multiple antibiotic resistance (Mar) phenotype [281]. It can also activate genes such as *sodA*, *zwf* and *micF* [281]. MarB is an outer membrane-bounded periplasmic space protein that regulates the transcription rate of *marA* [282]. The protein MqsA is a component of the toxin-antitoxin system MqsR-MqsA, which controls biofilm formation and triggers programmed cell death in *E. coli* [283]. The phage shock protein (*psp*) operon (*pspABCDE*) can play a significant role in the competition for survival under nutrient- or energy-limited conditions [284]. PspA negatively regulates the expression of the *pspABCDE* promoter and of *pspG* through negative regulation of the *psp*-specific transcriptional activator PspF [285]. PspA is also required for membrane integrity, efficient translocation, and maintenance of the proton motive force. PspB is also involved in transcription regulation [284]. The most interesting down-regulated genes were *potD* and *potX*. PotD, or polyamine binding protein D, is required for the activity of the bacterial periplasmic transport system of putrescine and spermidine. [286]. ProX is a part of the ProU ABC transporter complex involved in glycine betaine and proline betaine uptake [287]. These changes in the expression profile can provide sustainability and a persistence phenotype for CRISPR ON cells under antibiotic conditions. In this experiment, we did not distinguish the plasmid-free fraction of CRISPR-ON population from the plasmid-bearing cells, but we

can approximate the transcriptome profile to the plasmid-free fraction due to its majority in CRISPR ON population. In addition to this, the control for the influence of inducers with respect to transcription of other genes was not included. Thus, the persistence mechanism remains unclear, so additional studies of the survivability of CRISPR ON plasmid-free cells under antibiotic conditions are required.

In contrast to plasmids, bacteriophages have the capacity for horizontal transfer of their genetic material among bacterial populations during infection cycles. The stochastic nature of the interaction between plasmids and CRISPR-Cas system, as described earlier, can similarly be considered within the context of phage infection of bacteria bearing a CRISPR-Cas system. From this perspective, the infection cycles can be considered to enhance the chances of phages survival within a bacteria population that harbors an active CRISPR-Cas system with spacers that match protospacers in the phage genome precisely. In the absence of escape mutations or other CRISPR interference escaping mechanisms, the stochasticity of phage genome replication and CRISPR-interference events may offer an ancillary survival mechanism for phages during infection cycles. Analogous to the plasmids, the phage replication dynamics might exhibit the stability window with respect to the CRISPR interference dynamics (**Fig. 2.6, C**). Within this window of stability, phages can produce new virions capable of infecting adjacent bacterial cells. According to this hypothesis, the persistence of the surviving bacteriophages should be more robust than that of plasmids due to the increased frequency of successful infections. This also suggests that the CRISPR immunity is important at the early stages of infection to prevent late infection stages [34].

To test this hypothesis, the experiments with transformation of *E. coli* KD263 cells with M13 replicative form bearing the native g8 protospacer, β -lactamase gene and a GFP fluorescent marker were performed. The transformants were plated on media supplemented with appropriate antibiotics for control and on media supplemented with both antibiotics and inducers for CRISPR interference assay as previously described for the experiments performed with plasmids. The results demonstrated a quick accumulation of M13 phage escapers in *E. coli* colonies after transformation with M13 replicative form (RF) under CRISPR interference conditions.

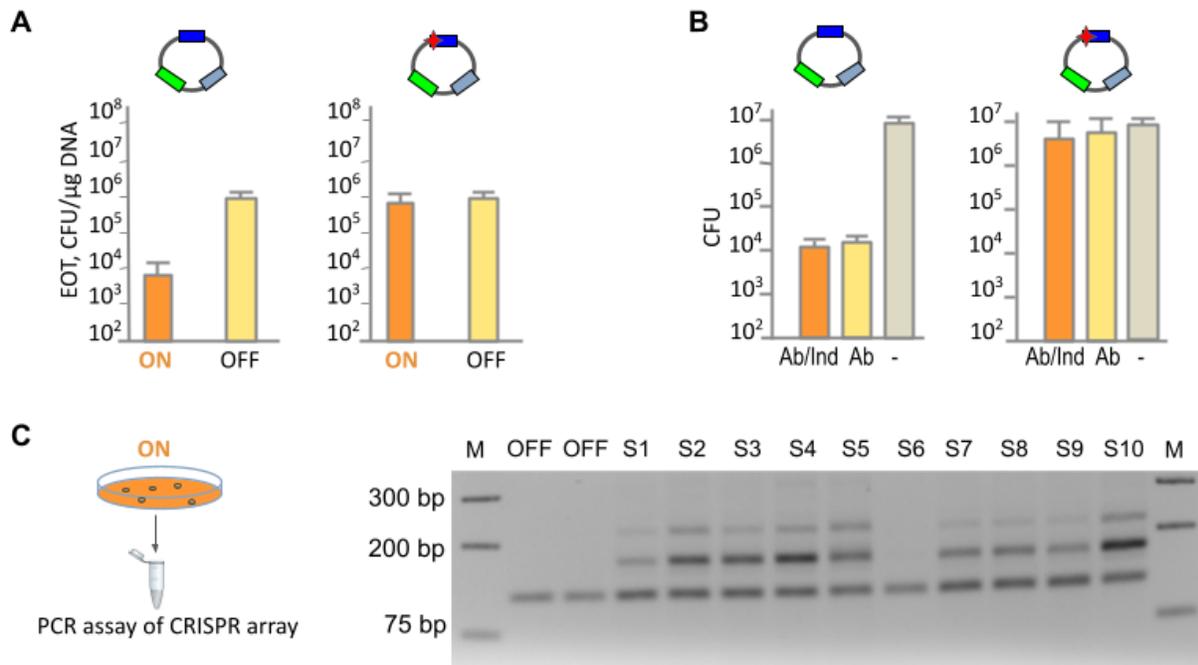


Figure 2.7. Results of transformation *E. coli* KD263 cells by M13 phage replicative form.

a, Efficiency of transformation of CRISPR ON and CRISPR OFF cells with a dsDNA replicative form of phage M13 bearing the wild-type protospacer (the left plot) and mutated protospacer (the right plot). All transformations were performed at least three times. **b**, The reseeding on selective media *E. coli* transformants bearing dsDNA replicative form of phage M13 with the wild-type protospacer (the left plot) and the protospacer with escape mutation (the right plot). **c**, The assay for CRISPR adaptation assay. The CRISPR OFF cells contain a non-extended CRISPR array while CRISPR ON samples acquired new spacers.

The reseeding experiments of *E. coli* transformants, bearing M13 phages, have demonstrated similar results to previous experiments with plasmids (**Fig. 2.7B**). To understand the nature of escape mutations in M13 genome, phage M13 genomes, isolated from 9 fluorescent colonies after transformation, were sequenced. The results have demonstrated that one *E. coli* colony can simultaneously possess various M13 escaper genotypes (**Fig. 2.8**). To explain this coexistence of various genotypes in the same transformant colony, escape mutation should arise independently during infection cycles. At the same time, the PCR test of CRISPR adaptation in CRISPR ON colonies has shown extended CRISPR arrays (**Fig. 2.7C**). The expanded CRISPR arrays and the various types of escape phages M13 indirectly illustrate the complex spatial structure of infected colonies, as shown previously [288].

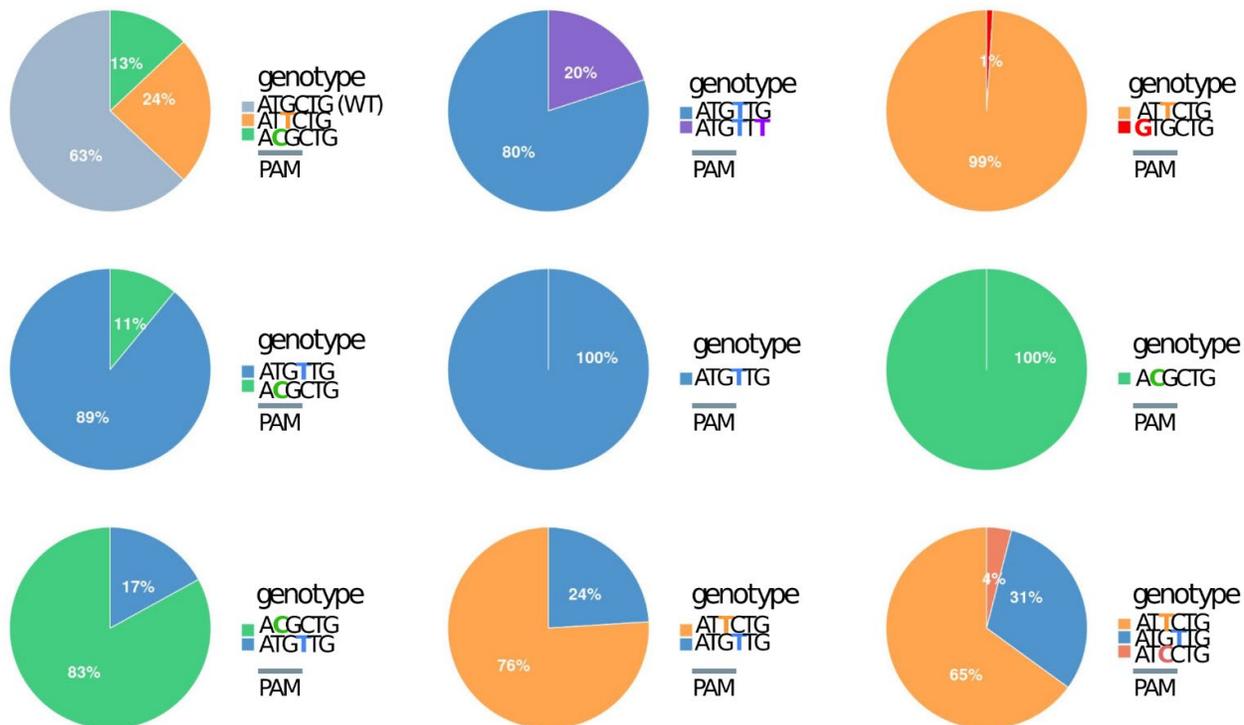


Figure 2.8. Genotypes of M13 phages isolated from infected *E. coli* colonies. The percentage corresponds to a fraction of reads bearing the corresponding genotype of the PAM and protospacer g8. Each genotype corresponds to separated Illumina reads.

In conclusion, these experimental results have demonstrated the fast proliferation of escape phages M13 in *E. coli* colonies. Escape phage M13 stimulated primed adaptation as previous studies have showed [125, 128]. The coexistence of escape phage M13 and *E. coli* cells with the active type I-E CRISPR-Cas system providing primed adaptation indicates more complex interaction than in the previous model system with the plasmids.

Chapter 3: Benchmarking DNA Isolation Methods for Marine Metagenomics.

The results of Chapter 3 are published in:

Demkina, A., Slonova, D., **Mamontov, V.**, Konovalova, O, Yurikova, D, Rogozhin, D., Belova, V., Korostin, D., Sutormin, D., Severinov, K., & Isaev, A. (2023). Benchmarking DNA isolation methods for marine metagenomics. *Sci Rep* 13, 22138 2023.

I conducted the bioinformatics analysis with the assistance of D. Sutormin. The collection of samples and DNA extraction were conducted by other authors. The text of the publication is presented only for the completeness of the storytelling.

3.1 Introduction

The Global Ocean is the planet's largest ecosystem. Yet, biodiversity of Earth's oceans remains understudied, partly due to the low accessibility of specific niches, such as ocean floor sediments or Arctic regions [289–293]. The problem is particularly poignant for prokaryotic communities, given that majority of bacteria are unculturable or require very specific growth conditions [294, 295]. In recent years, metagenomics, e.g., sequencing of total DNA isolated directly from collected samples allowed to estimate the real diversity and abundance of environmental microorganisms bypassing the isolation and cultivation steps [296–298]. Ambitious endeavors, such as the TARA Project [299–301], Earth Microbiome Project [302–304] or SEA-PHAGES [305, 306] aim to comprehensively describe microbial or viral diversity across Earth's marine and terrestrial ecosystems using large-scale metagenomic sequencing. Our project “Atlas of Microbial Communities of the Russian Federation” is conceived to cover one of the largest “white spots” on the Earth's microbial communities sampling map – the Arctic Ocean.

Several approaches to study the taxonomic and genetic diversity of microbial communities have been developed. 16S rRNA gene hypervariable regions amplicon sequencing allows one to directly assign taxonomy to sequenced DNA fragments and estimate alpha- and beta-diversity, i.e., the diversity of taxa within a sample or between different samples [307–309]. However, amplicon sequencing introduces biases [310, 311], does not distinguish close strains [311] and, more importantly, does not provide information about genes other than 16S rRNA genes. In contrast, by sequencing all DNA

fragments found within a sample, short-read shotgun metagenomic sequencing (on Illumina or BGI platforms) allows one, with sufficient quality of input DNA and depth of sequencing, to reconstruct larger genomic fragments (contigs) or even complete bacterial genomes and identify functional gene clusters [312–314]. In combination with long-read sequencing (using Oxford Nanopore or Pacific Biosciences technologies), one can perform hybrid assembly, which significantly increases the length of contigs [315, 316] but requires higher quality (reduced fragmentation) and quantity of input DNA.

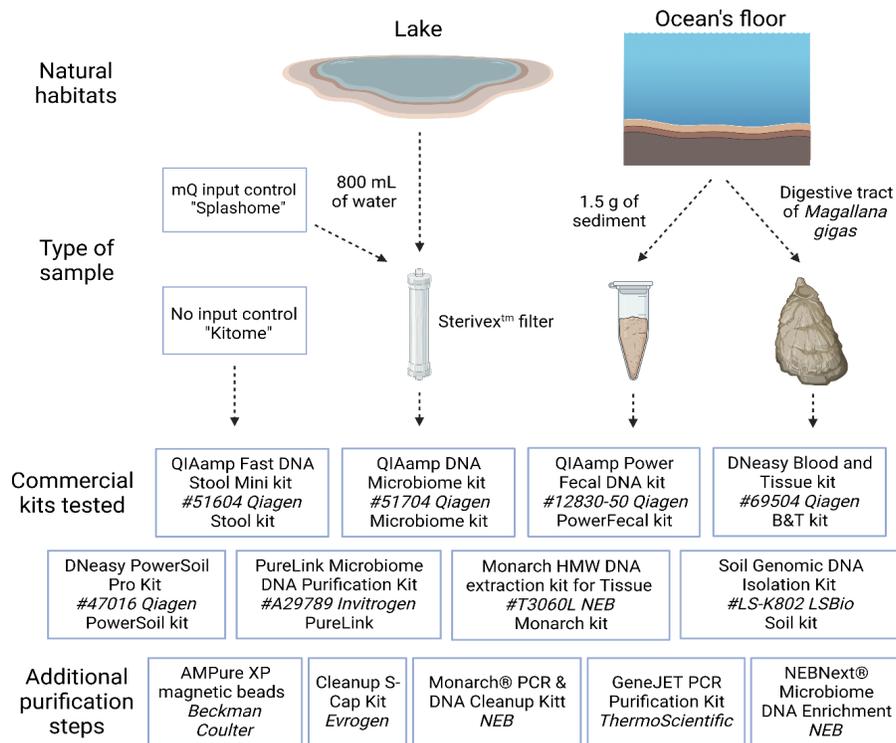
Purification of high-quality metagenomic DNA from natural sources faces a variety of challenges. Sample handling, the choice of DNA purification method, and even homogenization techniques used significantly affect the resulting community composition [317–320]. In this work, we processed three types of samples: fresh water, ocean sediments, and digestive system of a model marine invertebrate – Pacific oyster (*Magallana gigas* also known as *Crassostrea gigas*). To select the best strategy for microbial DNA isolation from each type of sample we systematically evaluated available commercial DNA purification kits, and estimated the quality of resulting DNA, its applicability for the 16S rRNA amplicon, short-read and long-read sequencing, and compared resulting communities' composition (**Fig. 3.1**). Our goal was to select DNA purification procedures that optimize each of the following parameters:

- DNA quantity, the most evident, yet the most important characteristics, especially for Oxford Nanopore platform that, according to manufacturer recommendations, requires at least 1 µg of DNA (although in some reports this quantity was lowered to just 1 ng [321]). Obtaining sufficient quantities of DNA is also particularly important for Arctic Ocean samples given comparatively low numbers of prokaryotic cells in cold waters [322, 323].
- DNA purity and the presence of contaminants that can significantly impact the efficiency of PCR and the quality of sequencing libraries [324]. A well-known example of such contaminant is humic acid that often co-purifies with DNA from soil samples and inhibits PCR in concentrations starting from 10 ng/µL [325, 326]. PCR is also sensitive to the presence of chelating agents that sequester Mg²⁺ and to a plethora of organic compounds (36). Such contaminations can be roughly estimated by UV adsorption at 230 nm, while co-purifying proteins can be detected by UV adsorption at 280 nm.

- DNA fragmentation. Extensive DNA fragmentation can be a consequence of poor sample storage, mechanical shearing during purification, or enzymatic degradation by nucleases released at the stage of cell homogenization [327]. The parameter is especially important for long-read sequencing as it determines the lengths of resulting reads.
- Admixture of eukaryotic DNA. This problem is relevant for shotgun sequencing of DNA from symbiotic microbial communities. Unless specific measures are taken, the proportion of microbial reads can be below a few percent [328–330]. Although the presence of non-microbial DNA seemingly should not affect 16S rRNA amplicon sequencing, incorrect estimation of the amount of microbial DNA in the sample can lead to PCR amplification biases caused by target DNA underrepresentation.
- Reagent and laboratory contamination with microbial DNA (“kitome” and “splashome”, respectively). This is one of the major problems that precludes correct analysis of microbial communities in sparsely-populated environments, since the signal from contaminating DNA can be higher than that from DNA of interest. Primarily, avoidance of contaminating DNA is achieved by complying with aseptic practices at all stages of sample collection and DNA purification, as well as by sterilization and decontamination of equipment and laboratory space [331–334]. Although bioinformatic decontamination procedures can be applied at the stage of data analysis, complete removal of non-sample specific sequences is impossible, as the *a priori* community composition is not known [334, 335]. To estimate the signal from contaminating DNA, negative control samples are processed and sequenced in parallel with experimental samples. An unavoidable and often neglected source of DNA contamination is the DNA purification kit itself, as even minuscule quantities of microbial DNA in the kit solutions can be detected during sequencing, and each commercial kit can be described in the terms of the “kitome”, e.g., a set of contaminating taxa it contributes [336–338].
- Method-specific biases in the composition of microbial taxa. Another major source of biases in metagenomic analysis is the varied efficiency of DNA purification for different microbial groups. Given the same starting material, different DNA isolation methods will result in vastly different quantities and proportions of group-specific DNA molecules [339–342]. The major source of this bias is the efficiency of cell lysis. Many gram-positive bacteria or specific groups that form

endospores or dormant cells (such as *Firmicutes*, *Actinobacteria*, *Cyanobacteria*, *Myxococcales* and *Azotobacteraceae*) resist homogenization and lysis by convenient methods [343, 344]. At the same time, biases in water samples can be associated with filter retention efficiency that could vary for different cells [345]. The effects of the homogenization and cell lysis approach on the resulting community composition are described in many reports [340, 346–348].

In this work, we applied eight commercially available and widely used microbial DNA purification kits to three types of samples: fresh water, deep sea sediments, and digestive system of a model marine invertebrate – Pacific oyster *M. gigas* (**Fig. 3.1**). For each sample-kit combination, we estimated: 1) DNA quantity; 2) DNA purity; 3) DNA fragmentation; 4) presence of PCR inhibitors; 5) admixture of eukaryotic DNA; 6) contamination by “kitome” and “splashome; 7) reproducibility of the 16S rRNA amplicon sequencing in 3 biological replicates; 8) alpha-diversity of the sample. The results allowed us to rank each kit and build a comprehensive description matrix that should aid in the selection of the best DNA isolation method for a specific sample type and purpose.



Parameters of purified DNA

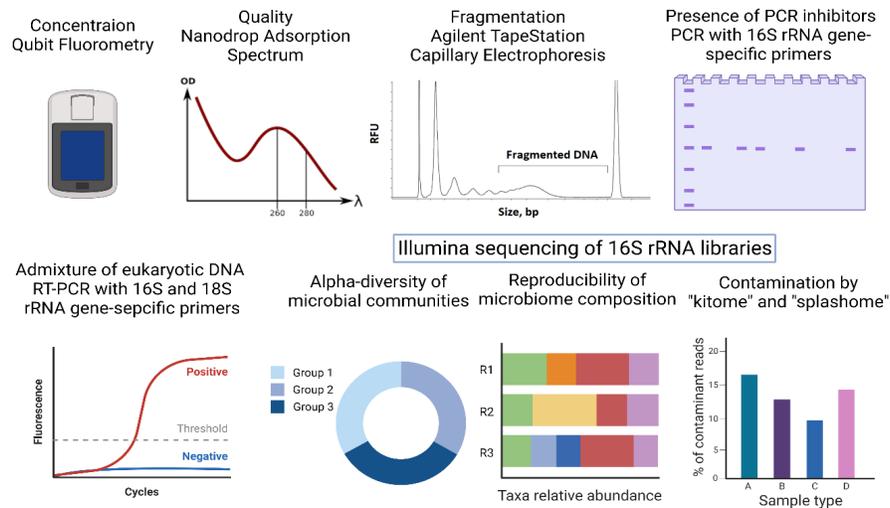


Figure 3.1. A scheme representing the methodology of the study. Three types of samples were collected. Each sample, along with two types of negative controls (No-input NC; Milli-Q NC), was processed in triplicates with eight commercial microbial DNA purification kits. Additional purification steps were applied to some of the resulting DNA samples to follow the removal of contaminants and reduction of eukaryotic DNA load. All samples were evaluated for an indicated set of parameters to select the best DNA purification strategy.

3.2 Methods

3.2.1 DNA purification kits (*conducted by other authors*)

All DNA purification steps were performed according to manufacturer's recommendations. When applied, we specifically mention some minor changes introduced into the protocols. General description of each kit and its short identification name used throughout the paper is provided below. When bead beating step was necessary, we used TissueLyser LT (Qiagen) and treated samples at conditions of 50 Hz for 10 minutes, unless other time was specified in the protocol.

QIAamp Fast DNA Stool Mini kit (#51604 Qiagen) = **Stool** kit. Purification protocol begins with absorption of the compounds that can degrade DNA and inhibit downstream enzymatic reactions. Next, microbial cells are lysed, and proteins are degraded in the presence of Proteinase K. Then, the sample is loaded onto the QIAamp spin column. DNA bound to the silica membrane is washed two times and concentrated DNA is then eluted.

QIAamp DNA Microbiome kit (#51704 Qiagen) = **Microbiome** kit. The first step is lysis of eukaryotic cells and degradation of host nucleic acids with benzonaze. The second step is disruption of bacterial cells through bead beating. Then, the sample is loaded onto the QIAamp UCP mini column. DNA bound to the silica membrane is washed two times and concentrated DNA is then eluted.

QIAamp PowerFecal DNA kit (#12830-50 Qiagen) = **PowerFecal** kit. The first step of the protocol is mechano-chemical cell disruption through bead beating in lysis buffer. The Inhibitor Removal Technology[®] is then used to remove common substances that interfere with downstream PCR applications. DNA from supernatant is captured on MB Spin Column, washed two times and eluted from the silica membrane.

DNeasy Blood and Tissue kit (#69504 Qiagen) = **B&T** kit. After cell lysis and protein degradation in the presence of Proteinase K, the sample is loaded onto the DNeasy Mini spin column. DNA is then washed two times and eluted from the silica membrane.

DNeasy PowerSoil Pro Kit (#47016 Qiagen) = **PowerSoil** kit. The protocol starts with mechano-chemical cell disruption through beat beating in lysis solution. In this kit, Inhibitor Removal Technology[®] is also applied, which should help in removal of organic and inorganic materials such as humic acids, cell debris, and proteins that could interfere with downstream PCR applications. After

that, the sample is loaded onto MB Spin Column with silica membrane. DNA is then washed two times and eluted from the silica membrane.

PureLink Microbiome DNA Purification Kit (#A29789 Invitrogen) = **PureLink** kit. Multiple protocols are proposed for this kit, we applied soil procedure for all types of samples. In this protocol, the microorganisms are lysed by a combination of heat, chemical, and mechanical disruption, e.g., bead beating is applied. After that, the sample is treated with a cleanup buffer to eliminate PCR inhibitors. The sample is then applied to a PureLink™ spin column. DNA is washed one time and then eluted from the silica membrane.

Monarch HMW DNA extraction kit for Tissue (#T3060L NEB) = **Monarch** kit. We skipped the homogenization step, as *M. gigas* samples were uniformly homogenized at preliminary step, common for all kits, while water and soil samples did not require homogenization with pestle. Then, we followed extraction protocol for bacterial samples. All samples were incubated in a STET solution supplied with lysozyme (see below) and heat-treated as recommended by the manufacturer. Next, lysis master mix solution was added to the samples, followed by proteins and RNA removal. In order to preserve the integrity of DNA, it was purified using glass beads, in contrast to silica membrane utilized in all other kits. DNA was gently washed through rotation and eluted from the beads.

Soil Genomic DNA Isolation Kit (#LS-K802 LSBio) = **Soil** kit. The sample is homogenized by bead beating and then treated in a special buffer that contains detergent, which serves the purpose of PCR inhibitors removal, such as humic acids, proteins, polysaccharides, and other contaminants. DNA is further bound to a silica spin-column. After one-time washing, DNA is eluted from the silica membrane.

3.2.2 Sample collection (*conducted by other authors*)

Ocean sediment samples were hand-collected with a scoop from a bottom sediment horizon of 0–5 cm at the littoral in Kola Bay (Minkino, 69.00203 N, 33.0201 E) in December 2021 (**Fig. S3.1A**). Samples were sterile packed, transferred and stored at -20°C until further processing.

Fresh water samples were collected from Skolkovo pond (Moscow, 55.69491 N, 37.35393 E) in June 2022 (**Fig. S3.1A**). For each sample, 800 mL of water from the same batch was filtered through the 0.22 µm Sterivex unit (Merck-Millipore) with a peristaltic pump. 800 mL of Milli-Q water was

filtered as a negative control in parallel to capture and investigate “splashome”. Filters were subsequently stored at -20°C until processing.

8 individuals of *M. gigas* were collected in the Melkovodnaya Bay (Vladivostok, 43.018832 N, 131.885805 E) in July 2022 (**Fig. S3.1A, B**). Digestive tracts were isolated and stored at -20°C until processing.

3.2.3 DNA extraction (conducted by other authors)

At all stages where the use of water was required, we used sterile nuclease-free water (B1500L, NEB).

Sediment samples. For DNA extraction, 1.5 g of thawed soil was used per replicate. Studied sediment material was stored in one tube and it was rigorously mixed beforehand to avoid heterogeneity. DNA was isolated in accordance with protocols recommended by manufacturers. For Monarch kit, homogenization step was skipped.

Water samples. Sterivex filter membrane was removed from the cover as described [349]. The membrane was transferred into a sterile and single-use Petri dish with a cell-coated surface facing up and cut into small pieces. Membrane fragments were transferred into the sterile 2-mL microcentrifuge tube. Before further processing, filter fragments were incubated in 200 µL of the lysis buffer STET (50 mM TrisHCl, pH=8.0; 50 mM EDTA; 5% Triton-X100; NaCl – 200 mM; freshly supplied with 10mg/ml lysozyme) at 37°C for 1 h in a heating block with shaking at 600 rpm. After incubation, all liquid was collected for downstream processing in accordance with the manufacturer’s protocols. For samples, which were processed by Monarch kit, the homogenization step was skipped. For samples, which were processed by Microbiome kit, the homogenization step was done before addition of the ATL Buffer (provided in the Microbiome kit).

Gut flora samples. Digestive tracts from eight *M. gigas* individuals were thawed, pooled, and then homogenized for 15 min at 50Hz using TissueLyser LT in the Tissue Disruption Tube (QIAamp Fast DNA Tissue Kit (Qiagen)), containing a single stainless-steel bead. Any other tissue homogenization method can be applied as this step. Resultant homogenate was split into 24 aliquots (three replicates for each of eight kits used). DNA was isolated in accordance with the manufacturer’s protocols. For the Monarch kit, the internal homogenization step was skipped.

“Kitomes” and “splashomes”. For each kit, two types of negative controls were prepared. For No-input negative control (for Soil and Gut flora samples) no starting material was added, and DNA was extracted solely from the buffers contained in the kit. For Milli-Q negative control (for Water samples), we took laboratory Milli-Q water purified consequently in two steps using the Barnstead Pacific TII (Thermo Scientific) and Simplicity systems (Millipore). 800 mL of Milli-Q was filtered through Sterivex filter units, and DNA was extracted from filter membranes as described for Water samples processing.

3.2.4 Additional purification procedures (*conducted by other authors*)

All additional purification was conducted according to the manufacturer’s instructions. An equal volume of samples was taken for each procedure.

AMPure XP Reagent (A63880 Beckman Coulter). The Agencourt AMPure XP purification system utilizes solid-phase reversible immobilization (SPRI) paramagnetic bead technology and optimized buffer for high-throughput purification of DNA fragments. Salts, inhibitors, nucleotides, and enzymes are removed using a washing procedure with magnetic separation, resulting in a purified DNA product.

Cleanup S-Cap (BC041L Evrogen). DNA fragments bind to the column membrane in the presence of highly concentrated chaotropic salts and optimal pH of binding buffer. Subsequent wash steps allow to get rid of nucleotides, short fragments of nucleic acids, salts, proteins, inhibitors of enzymatic reactions, and other impurities of organic compounds. DNA elution occurs under slightly alkaline conditions in a low salt buffer.

GeneJET PCR Purification Kit (#K0702 Thermo Scientific). DNA combined with the binding buffer is added to a purification column. A chaotropic agent in the binding buffer denatures proteins and together with optimal pH promotes DNA binding to the silica membrane in the column. Impurities are removed with a wash step and purified DNA is then eluted from the column with the elution buffer.

Monarch PCR & DNA Cleanup Kit (#T1030L NEB). This method also employs a bind-wash-elute workflow. Binding Buffer is used to dilute the samples and ensure they are compatible for loading onto the proprietary silica column under high salt conditions. The Wash Buffer ensures

enzymes, detergents and other low-molecular weight reaction components are removed, thereby providing high-purity DNA after elution.

NEBNext Microbiome DNA Enrichment Kit (E2612L NEB). Methylated host DNA in a DNA mixture is selectively bound to the mCpG binding domain of human MBD2-Fc protein. After capture, the microbial DNA which is not CpG methylated, or is minimally CpG methylated, remains in the supernatant with minimal sample loss.

RNaseA (#T3018-2 NEB). Co-purification of RNA during DNA extraction is a common problem that leads to the overestimation of DNA yield and complication in NGS library preparation. Most commercial kits utilize low alcohol binding conditions, that should result in low RNA co-purification. However, an additional RNaseA treatment might be required for samples with obvious saturation in low molecular weight fragments. 10 μ L of sample was incubated with 0.4 μ L of RNaseA per at room temperature for 10 minutes.

3.2.5 DNA quantification and quality assessment (*conducted by other authors*)

DNA concentration was measured using the Qubit dsDNA High Sensitivity Assay Kit or Qubit dsDNA Broad Range Assay Kit on the Qubit 3.0 Fluorometer (Invitrogen). All DNA samples were eluted with 60 μ L of elution buffer provided in the corresponding kit so the total amount can be directly compared within one sample type. Detection limit of DNA concentration in the sample for Qubit dsDNA High Sensitivity Assay Kit is 0.1 ng/ μ L. Thus, samples that were not measurable by this assay should contain less than 6 ng of total DNA. We consider samples with DNA yield above 1500 ng as “passable” for downstream short-read and long-read DNA sequencing.

DNA purity was assessed by measuring the ratios of absorbance at 260 nm to 230 nm (260/230) and 260 nm to 280 nm (260/280) using the NanoDrop1000 spectrophotometer (Thermo Scientific). Samples with a 260/230 ratio between \sim 2.0-2.2 and 260/280 \sim 1.7-2.0 are assumed as “pure”.

The integrity of genomic DNA was assessed using Agilent TypeStation 4150 (Agilent Technologies) with Genomic DNA ScreenTape System according to the manufacturer’s instructions. 1 μ L of sample was used. Samples with DIN above 7.0 were assumed as “high-quality” and acceptable for long-read sequencing.

3.2.6 PCR Amplification (*conducted by other authors*)

16S rRNA V3-V4 region was amplified using universal Illumina V3/V4 PCR primers [350]. Each PCR reaction contained 1 μ L of DNA, 0.25 μ L of forward and reverse primer (final concentration of 0.5 mM), 5 μ L of Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB) and 3 μ L of nuclease-free water (NEB) for final reaction volume of 10 μ L. PCR conditions were 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 55°C for 20 s, and 72°C for 30 s, with a final extension time of 5 min at 72°C. Amplicons were visualized on a 1% agarose gel containing 0.25 μ g/ μ L ethidium bromide running in 1xTris-EDTA buffer at 100V with a target product size in a range of 400-500bp (62). Products of successful amplification were clearly visible on gel, otherwise the initial DNA sample was diluted 10 or 100 times for PCR re-examination.

3.2.7 Quantitative real-time PCR (*conducted by other authors*)

To estimate a ratio between bacterial DNA and co-extracted eukaryotic DNA, quantitative real-time PCR (qPCR) targeting different variable regions of 16S rRNA and 18S rRNA genes was performed using universal primers (**Table 2.1**). Since many DNA samples contained PCR inhibitors, they required additional dilution in order to achieve a robust amplification, yet the same DNA batch was always used for 16S and 18S amplification. qPCR reactions were performed in technical triplicates in optical 96 well plates using the QuantStudio 3 Real-Time PCR System (Applied Biosystems). Each reaction was performed in a final volume of 10 μ L, containing 1 μ L of template DNA, 5 μ L of 2x iTaq Universal SYBR Green Supermix (Bio-Rad), 0.5 μ L of forward and reverse primers (0.5 mM final concentration). qPCR protocol consisted of initial denaturation at 95°C for 5 min, followed by 40 cycles of 95°C (30 s), 55°C (20 s), and 72°C (30 s), and the final generation of the dissociation melting curves to verify amplification specificity. The load of eukaryotic DNA was estimated as the difference in the threshold cycle (Ct value) obtained for each sample with 16S- and 18S-specific primers (Ct(18S)–Ct(16S)) [351].

To test the amplification efficiency of the selected 16S and 18S-specific primer pairs, qPCR reactions were performed with five ten-fold dilution series (ranging from 10 to 0.0001 ng/ μ l) of,

correspondingly, *E. coli* and *Homo sapiens* genomic DNA. Primer pairs efficiencies were higher than 98% and were calculated as described in [351].

Table 2.1. Primers used in the study.

16S	V3-V4	F	CCTACGGGNGGCWGCAG	[350]
		R	GACTACHVGGGTATCTAATCC	
18S	V4	F	CCAGCAGCCGCGGTAATTCC	[352]
		R	ACTTTCGTTCTTGATTAA	

3.2.8 16S rRNA libraries sequencing (conducted by other authors)

Amplification of the V3-V4 region of 16S rRNA and library preparations were performed according to the Illumina manual [350]. The amplicon libraries were barcoded, pooled in a single batch and sequenced in Evrogen using NovaSeq 6000, 2 x 250 bp paired-end protocol.

3.2.9 16S data analysis (conducted by me)

I contributed to updating 16S_snakemake.smk pipeline in the GitHub repository of my colleague D. Sutormin and conducted the computational analysis of 16S rRNA data described in this section. All data related to the 16S analysis is also presented in public at https://github.com/mavic9/16S_rRNA_analysis.

For 16S data analysis, a snakemake pipeline was developed (https://github.com/sutormin94/16S_analysis/). Briefly, quality of sequencing data was checked using FastQC and then reads were trimmed and filtered using Trimmomatic v. 0.39 (SE -phred 33 HEADCROP 17 ILLUMINACLIP:2:30:10 MINLEN:150). Only forward reads were used for downstream analysis. Reads that passed quality control were processed with DADA2 pipeline v. 3.6.2. Resultant amplicon sequence variants (ASVs) were clustered using MMseqs2 v. 10-6d92c (coverage > 0.95, identity > 0.98) and representative sequences were further treated as operative taxonomic units (OTUs). OTUs were returned to DADA2, and taxonomy was assigned to OTUs using the SILVA SSU database v.138 [353]. Contamination was removed using R package decontam v. 1.14.0 in the “either”

mode with a threshold 0.5 [335]. PCoA (Principal coordinates analysis), alpha-diversity, and taxonomic analyses were performed with R packages phyloseq v. 1.30.0 [354], ggplot2 v. 3.3.6 (<https://github.com/tidyverse/ggplot2>), dplyr v. 1.0.10 (<https://github.com/tidyverse/dplyr>), vegan v. 2.6.2 (<https://github.com/vegandevs/vegan>). Shannon index, calculated by phyloseq, was used as a metric that incorporates richness and dominance of OTUs in a community.

3.2.10 Shotgun sequencing (*conducted by other authors*)

Libraries for shotgun sequencing were prepared from 300–500 ng of starting DNA using MGIEasy Universal DNA Library Prep Set (MGI Tech), following the manufacturer's instructions. DNA was sonicated using a Covaris S-220 followed by selection of 200-250 bp-long fragments on magnetic beads (MagBio). The concentration of the prepared libraries was measured using Qubit Flex (Life Technologies) with the dsDNA HS Assay Kit. The quality of the prepared libraries was assessed using Bioanalyzer 2100 with the High Sensitivity DNA kit (Agilent). DNA libraries were further circularized and sequenced by a paired end sequencing using DNBSEQ-G400 with the High-throughput Sequencing Set PE100 following the manufacturer's instructions (MGI Tech) with an average coverage of 100x. FastQ files were generated using the zebacallV2 software (MGI Tech).

3.3 Results

3.3.1 Sample collection and processing

The results of section 3.3.1 were conducted by other authors.

The overall scheme of our study is presented in **Fig. 3.1**. We processed three types of samples: fresh water, sea sediments, and digestive system of a marine invertebrate (“gut flora”). Given the low titer of prokaryotic cells in the Arctic Ocean water masses, we decided to validate DNA purification from water using a richer sample from a freshwater lake with an intermediate level of algal bloom (**Fig. S3.1A**). 800 mL water aliquots collected from a single batch were filtered through Sterivex filters using a peristaltic pump. 800 mL aliquots of Milli-Q water were loaded on separate series of filters as negative controls (“Milli-Q negative controls”) and processed before experimental water samples. All

filters were frozen at -20°C and then processed together in a single day. STET lysis solution supplemented with lysozyme was added to fragmented filter membranes and collected liquid served as input for downstream processing (see Methods). The sea sediment sample was collected at the littoral zone of the White Sea (Arctic Ocean water area) (**Fig. S3.1A**). The sample was thoroughly mixed to avoid granulometric inhomogeneities, and equivalent of 1.5 grams of dry weight was taken per replicate for processing with each kit. To study the efficiency of microbial DNA purification from marine invertebrates, we selected eight individuals of giant Pacific oyster *M. gigas* (also known as *C. gigas*, NCBI:txid29159) collected at the shores of the Sea of Japan (Pacific Ocean water area) (**Fig. S3.1B,C**). The digestive tracts were extracted, homogenized (see Methods) and pooled together for downstream processing (300 mg of homogenate was taken per replicate). All experimental samples were treated in triplicates. Additional negative controls with no input material (“No-input negative controls”), which allowed us to estimate the “kitome” composition, were processed in parallel.

Each sample was purified with eight commercially available kits. Five kits included a beat beating stage for homogenization of the sample (Qiagen’s Microbiome, PowerSoil, PowerFecal, LSBio’s Soil, and Invitrogen’s PureLink) while the other three did not (Qiagen’s B&T and Stool, NEB’s Monarch HMW Tissue). Sample processing was carried out according to manufacturers’ instructions with minor alterations. Purified DNA was eluted in 60 µL of kit elution buffers and stored at -20°C. Samples were evaluated for various DNA parameters and used for 16S rRNA sequencing.

Since sea sediment samples produced good quantities of DNA, yet contained PCR inhibitors (see below), we re-purified these samples using several different protocols (**Fig. 3.1**) and repeated our analyses. To assess if the presence of PCR inhibitors affected efficiency of NGS library preparation, we subjected one initial and one re-purified DNA sea sediment sample for shot-gun BGI sequencing. In addition, we evaluated the efficiency of host DNA removal by the NEBNext Microbiome DNA enrichment kit for *M. gigas* samples.

3.3.2 DNA yield

I conducted the visualization of Fig. 3.2.

The concentration of DNA isolated by each kit was determined with Qubit dsDNA High Sensitivity or Broad Range Assay on the Qubit 3.0 Fluorometer. With every sample type, we observed remarkable differences in DNA extraction efficiency between different kits (**Fig. 3.2, Table 3.2**). For

sediment samples, PowerFecal and PowerSoil kits produced the highest yields, while Monarch and PureLink treated samples did not contain detectable quantities of DNA. For water samples, PowerFecal and PowerSoil were also the most efficient and on par with the B&T kit, which is often used for DNA purification from water samples [355–357]. Notably, in the case of this type of sample, all tested extraction kits isolated measurable quantities of DNA, though Monarch and PureLink continued to be the least efficient. For *M. gigas* samples, only Monarch kit failed to extract DNA. PowerSoil and PowerFecal resulted in high yields, suggesting their potential applicability for different types of samples. The highest amount was isolated with the Stool kit. In this case the amount of isolated DNA exceeded that obtained with other kits more than 10-fold. Since a low molecular weight smear was evident after gel electrophoreses of Stool kit purified DNA, we checked for RNA contamination by treating the purified DNA with DNase or RNase. The result confirmed that RNA was abundant in the sample (Fig. S3.2).

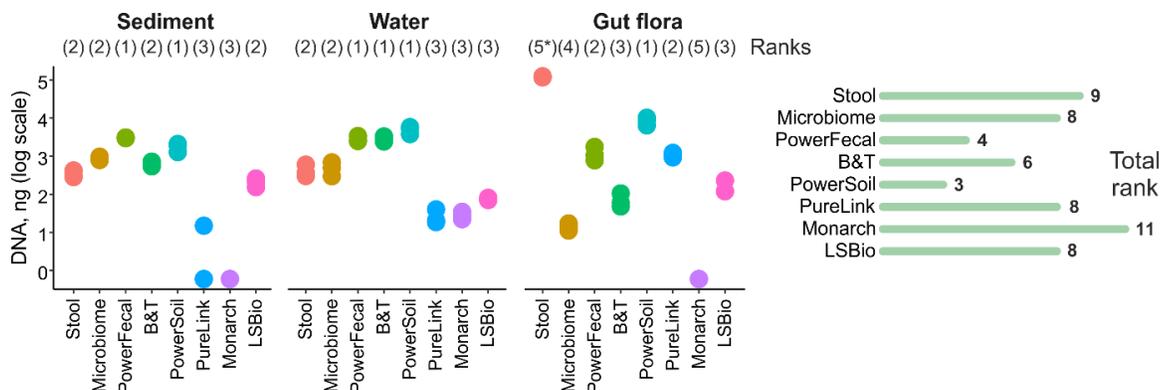


Figure 3.2. Total DNA amounts purified by different kits from three types of samples. Data for three technical replicates are shown. Right - kit ranking (a sum of ranks for different sample types). Lower ranks indicate higher DNA yeild. (*) DNA extracted from gut flora samples with a Stool kit was of low quality and was heavily contaminated with RNA.

3.3.3 DNA fragmentation

I conducted the visualization of Fig. 3.3.

The extent of DNA fragmentation was estimated by capillary electrophoresis on the TapeStation 4150 System (Agilent). This analysis allows one to calculate the DNA integrity number (DIN). DINs above 7 (roughly corresponding to samples with median fragment lengths above 15 kbp) are usually considered applicable for long-read (ONT or PacBio) sequencing [358, 359]. For complex environmental samples, such quality of DNA is rarely achieved, and storage of frozen samples can

result in increased fragmentation. Extensive DNA fragmentation (DIN below 3) can affect all types of downstream applications: long-read, short-read, and even 16S rRNA sequencing [359–361]. We measured DIN after purified DNA samples were kept frozen at -20°C for 3 months (**Fig. 3.3, S3.3 and Table 2**). For sediment samples, all kits that extracted measurable quantities of DNA, except B&T kit, demonstrated acceptable DINs of ~6-7. Stool and PowerSoil kits resulted in DINs of ~6-7 for water samples, while the commonly used B&T kit produced DNA with a DIN of ~4. *M. gigas* samples demonstrated intense fragmentation, reflecting possible contamination with host-derived nucleases. The highest DIN for these samples (~5.5) was achieved by PowerFecal and PowerSoil kits. Unexpectedly, the Monarch HMW kit, designed for purification of high molecular weight DNA from tissue samples, performed poorly resulting in DNA quantities that could not be measured on the TapeStation (**Fig. 3.2, 3.3**).

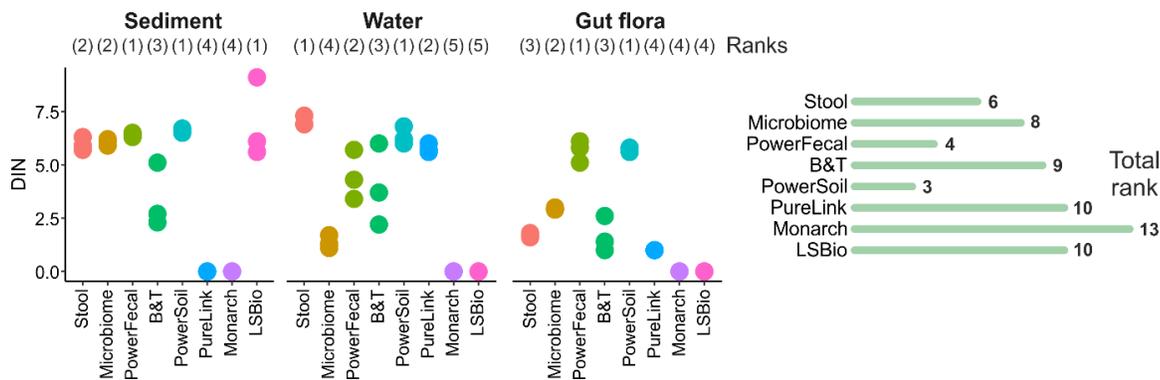


Figure 3.3. DNA integrity numbers (DINs) obtained with three types of samples subjected to purification by different kits. Data for three technical replicates are shown. Right - total kit ranking (a sum of ranks for different sample types). Lower ranks indicate higher DNA integrity.

3.3.4 DNA purity

I conducted the visualization of Fig. 3.4.

The purity of DNA samples was measured on a NanoDrop spectrophotometer. The 260/280 absorption ratio of 1.8-1.9 is usually considered as evidence of absence of protein contamination, while the 260/230 absorption ratio in a range of 1.8-2.5 demonstrates the lack of organic impurities [362]. In general, all purified DNA samples lacked protein contamination but contained excess compounds absorbing at 230 nm (**Fig. 3.4 A, B**). Stool, PowerFecal and PowerSoil kits performed best for sediment

samples; Stool, PowerFecal and B&T– for water samples; while PureLink, PowerFecal, and PowerSoil produced best results with *M. gigas* samples.

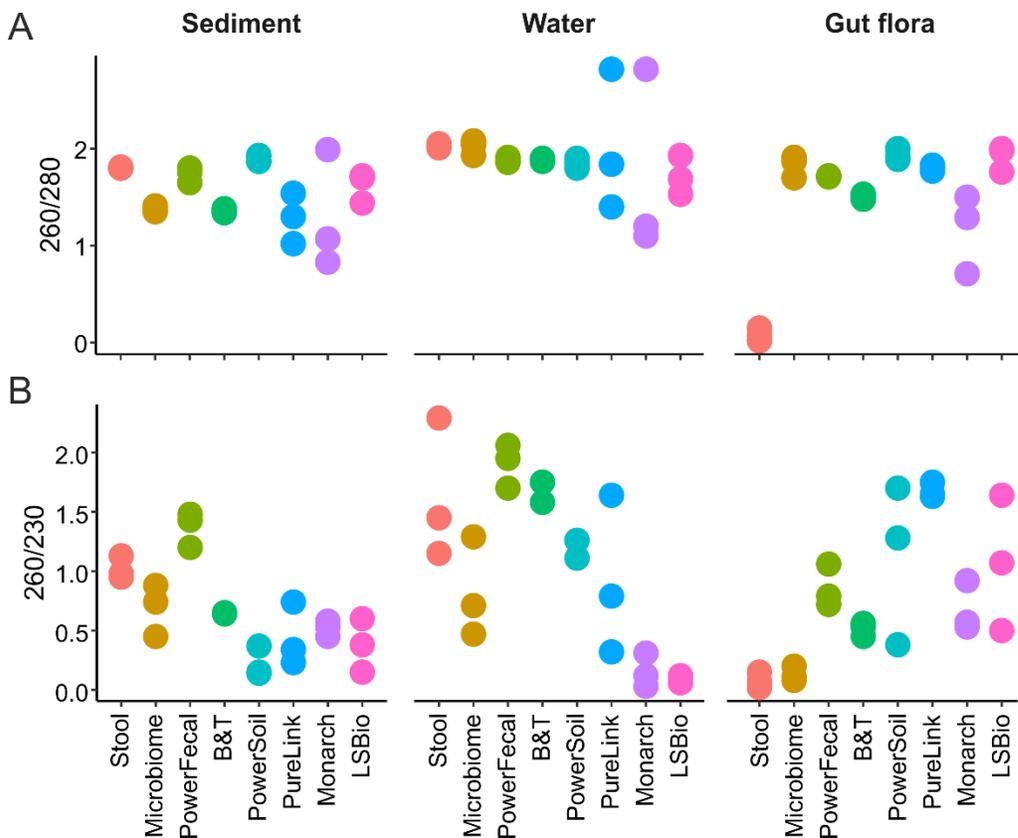


Figure 3.4. DNA purity assessed by 260/280 (A) and 260/230 (B) absorption ratios. Data for three technical replicates are shown.

3.3.5 Presence of PCR inhibitors (conducted by other authors)

As inhibition of downstream enzymatic reactions is the major adverse effect of DNA contaminants, another way to assess DNA purity is to estimate the efficiency of PCR amplification with purified DNA samples. To do so, we performed PCR using universal 16S rRNA gene-specific primers and 25 amplification cycles with Phusion DNA polymerase (NEB) according to manufacturer’s recommendations. For all sediment and *M. gigas* samples, except for those purified with PowerSoil, we failed to identify the expected amplicon, indicating the presence of PCR-inhibiting impurities (Fig. 3.5A, S3.4, Table 3.2). Although water samples were expected to contain less PCR inhibitors, samples purified with Stool, Microbiome and B&T kits also failed to produce a PCR

product. A common way to overcome PCR inhibition issue for DNA obtained from natural sources is dilution of samples, which lowers concentration of inhibitors below a threshold while still providing sufficient DNA for amplification [363–365]. Upon 100x dilution, all samples (except the negative “kitome” controls) produced 16S rRNA gene amplicons. Thus, we used a dilution factor as a rough estimate for the presence of PCR-inhibiting impurities (**Fig. 3.5A, S3.4, Table 3.2**). PowerSoil, PowerFecal and Stool kits include a specific inhibitors removal step, and indeed PowerSoil performed best in this comparison.

Since contamination with PCR inhibitors represented a serious issue for the sea sediment samples, we investigated if additional purification steps could help to cope with this problem. We selected DNA samples purified with Stool and Microbiome kits, which required 10x or 100x dilution for efficient 16S PCR amplification, respectively. Samples were divided in equal aliquots and re-purified using protocols described in the Methods section. For each re-purification, 50 ng (Stool) or 150 ng (Microbiome) of DNA was taken and retention efficiency was close to 100 % (**Fig. S3.4B**). To assess efficiency of PCR inhibitors removal, we performed 16S PCR with non-diluted and 10x diluted non-purified and re-purified samples. Both input samples did not produce the expected PCR product (**Fig. S3.4B**). In the case of Microbiome samples, only columns re-purification with Evrogen kit led to the appearance of 16S gene amplicons, while for Stool kit-purified samples all additional purifications protocols used were efficient (**Fig. S3.4C**).

We also investigated whether the presence of PCR inhibitors affects the quality of DNA libraries for shotgun sequencing on the BGI platform. A DNA sample purified from sediments with the PowerFecal kit (which required a 10x dilution for efficient PCR) and a corresponding sample re-purified with DNA Purification kit (Evrogen) (no dilutions required for efficient PCR) were used. Both samples were prepared according to a standard procedure that involved AMPure XP beads purification (see Methods). Samples passed through library preparation steps with similar efficiency and produced assemblies of a comparable quality (**Fig. S3.5**), suggesting that samples with moderate contamination with PCR inhibitors can be subjected to shotgun sequencing with a standard library preparation method without the additional purification step.

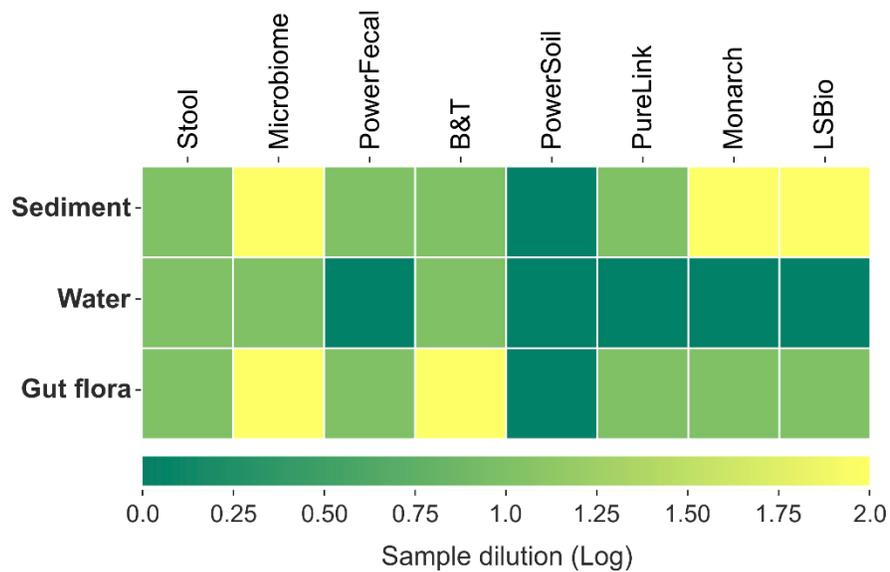


Figure 3.5. The presence of DNA inhibitors estimated as a dilution factor required to achieve visible production of a 16S rRNA amplicon.

3.3.6 Admixture of eukaryotic DNA

I conducted the visualization of Fig. 3.6.

The quantities of eukaryotic DNA in samples were assessed using qPCR with primers specific for 16S rRNA and 18S rRNA genes. The universal 18S rRNA primers used annealed to the 18S rRNA gene of *M. gigas* genome [366]. The difference in 18S to 16S amplicons abundance was estimated by subtracting threshold cycles (the Ct(18S)–Ct(16S) value). Values above 0 indicate the predominance of 16S rRNA-containing DNA molecules in the sample and 3 Ct cycles are expected to roughly correspond to a 10-fold difference of template molecules. As can be seen from **Fig. 3.6** and **Table 2**, sediment and water samples produced a ~100-1000x higher 16S rRNA signal, although, given the substantially larger sizes of eukaryotic genomes, this does not directly indicate a similar excess of prokaryotic DNA. For the *M. gigas* samples, many kits demonstrated Ct(18S)–Ct(16S) values above 0, and the best result was achieved with the Microbiome kit which is specifically designed to deplete host DNA. At the same time, Stool, PowerSoil, and PowerFecal kits demonstrated the highest levels of eukaryotic DNA admixture. It should be noted that though the PowerFecal and PowerSoil kits performed best in other tests and produced the largest amount of DNA from *M. gigas* samples.

However, the high DNA yields may be due to eukaryotic DNA admixture, suggesting that the weakness of these kits lies in the inability to discriminate between microbial and host DNA. In addition, PureLink-purified *M. gigas* sample showed a Ct(18S)–Ct(16S) value close to 0, and the amount of DNA purified with this kit was significantly higher compared to other kits with intermediate level of eukaryotic DNA admixture. Thus, considering the amount of total DNA and abundance of 16S rRNA in samples, B&T, PureLink, and Microbiome kits performed best for *M. gigas* samples.

To evaluate whether eukaryotic DNA admixture can be lowered by an additional treatment, we employed the NEBNext Microbiome DNA Enrichment kit, which selectively captures CpG methylated eukaryotic DNA [367], and re-purified Stool, PowerFecal and PowerSoil samples that contained the highest amounts of eukaryotic DNA. Even though mollusk’s DNA is considered hypomethylated [368, 369], the treatment of selected samples improved the Ct(18S)–Ct(16S) value by 3 Ct units on average, though microbial DNA still remained significantly underrepresented (**Fig. S3.6**).

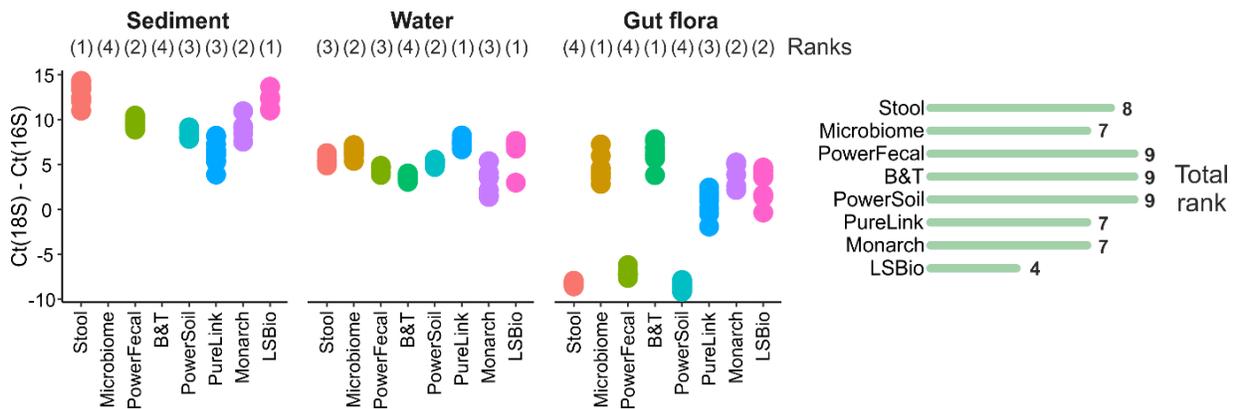


Figure 3.6. Eukaryotic DNA admixture estimated by the qPCR Ct values obtained with 18S and 16S rRNA gene-specific primers. Data for three technical qPCR replicates for each of the three kit purification replicates are shown. Below - total kit rankings (a sum of ranks for different sample types). Lower ranks indicate higher proportion of microbial DNA.

Table 3.2. Quality and quantity of DNA purified from three types of samples by various DNA extraction kits.

Sample	Kit	Stool	Microbiome	PowerFecal	Blood& Tissue	Power Soil	PureLink	Monarch	Soil
Sediment	DNA amount, ng	336.4 ± 75.3	858 ± 99	3036 ± 67	628 ± 90	1764 ± 430	<6	<6	211.2 ± 53.9
	Quality (260/280)	1.81 ± 0.01	1.38 ± 0.03	1.74 ± 0.08	1.36 ± 0.02	1.89 ± 0.03	1.29 ± 0.26	1.3 ± 0.6	1.62 ± 0.16
	Quality (260/230)	1.02 ± 0.10	0.69 ± 0.22	1.26 ± 0.34	0.65 ± 0.01	0.22 ± 0.13	0.44 ± 0.27	0.52 ± 0.07	0.38 ± 0.32
	Fragmentation (DIN)	5.97 ± 0.31	6.07 ± 0.15	6.4 ± 0.1	3.37 ± 1.51	6.6 ± 0.1	NA	NA	6.93 ± 1.89
	18S _{Ct} – 16S _{Ct} **	12.98 ± 1.04	NA	9.65 ± 0.49	NA	8.45 ± 0.47	6.33 ± 1.32	7.77 ± 1.02	11.94 ± 0.96
Water	DNA amount, ng	423.2 ± 156.8	494.8 ± 196.2	2828 ± 485	2896 ± 442	4960 ± 1042	26.4 ± 12.0	28 ± 6	75.6 ± 5.1
	Quality (260/280)	2.03 ± 0.02	2.02 ± 0.08	2.05 ± 0.33	1.88 ± 0.02	1.85 ± 0.01	2.02 ± 0.73	1.7 ± 0.96	1.72 ± 0.20
	Quality (260/230)	1.63 ± 0.59	0.82 ± 0.41	1.95 ± 0.22	1.69 ± 0.10	1.21 ± 0.09	0.92 ± 0.67	0.08 ± 0.06	0.09 ± 0.03
	Fragmentation (DIN)	7.03 ± 0.23	1.37 ± 0.31	4.46 ± 1.16	3.97 ± 1.91	6.33 ± 0.42	5.85 ± 0.21	NA	NA
	18S _{Ct} – 16S _{Ct} *	5.43 ± 0.42	6.37 ± 0.5	4.47 ± 0.34	3.53 ± 0.29	5.2 ± 0.29	7.37 ± 0.46	3.58 ± 1.51	6.77 ± 1.37
<i>M.gigas</i>	DNA amount, ng	119760 ± 438	14.3 ± 3.1	1196 ± 501	71.2 ± 30.0	1038 ± 105	1096 ± 151	<6	156.8 ± 63.7
	Quality (260/280)	0.05 ± 0.04	1.83 ± 0.11	1.71 ± 0.01	1.50 ± 0.03	1.94 ± 0.06	1.80 ± 0.03	1.17 ± 0.41	1.91 ± 0.13
	Quality (260/230)	0.05 ± 0.04	0.13 ± 0.06	0.86 ± 0.18	0.51 ± 0.06	1.12 ± 0.67	1.68 ± 0.06	0.67 ± 0.21	1.07 ± 0.81
	Fragmentation (DIN)	1.7 ± 0.1	2.93 ± 0.06	5.67 ± 0.51	1.67 ± 0.83	5.73 ± 0.12	1.0 ± 0.0	NA	1.0 ± 0.0
	18S _{Ct} – 16S _{Ct} **	-8.2 ± 0.16	4.56 ± 1.31	-7.14 ± 0.38	6.31 ± 1.12	-8.47 ± 0.41	0.83 ± 1.3	3.76 ± 1.23	2.88 ± 1.67

* - 10x dilution was taken for qPCR reaction due to the presence of PCR inhibitors.

** - 100x dilution was taken for qPCR reaction due to the presence of PCR inhibitors.

NA – No result was obtained due to low DNA quality or the presence of PCR inhibitors.

3.3.7 Contamination levels and determination of “kitomes”

I conducted the statistics analysis, subsampling OTUs, visualization and the analysis of contamination.

To further explore technical biases associated with the benchmarked DNA extraction kits, V3-V4 regions of 16S rRNA gene were PCR amplified from DNA extracted from environmental microbial communities and from mock samples prepared without input material (No-input negative controls representing “kitome”) or with sterile Milli-Q water (Milli-Q negative controls representing “kitome” + “splashome”). Obtained 16S amplicons were subjected to HTS using Illumina NovaSeq 6000 and the microbial diversity was studied at the operational taxonomic unit (OTU) level.

For environmental samples, on average, 93054 reads (standard deviation (SD) of 34626 reads) were obtained. For mock samples, as expected, the average number of reads was lower - 77564 (SD of 63798 reads). For further analysis, samples with a large number of reads were subsampled to 100,000 reads. Inspection of rarefaction curves at the level of OTUs demonstrated that for all samples, the sequencing depth reached the saturation level (**Fig. S3.7**).

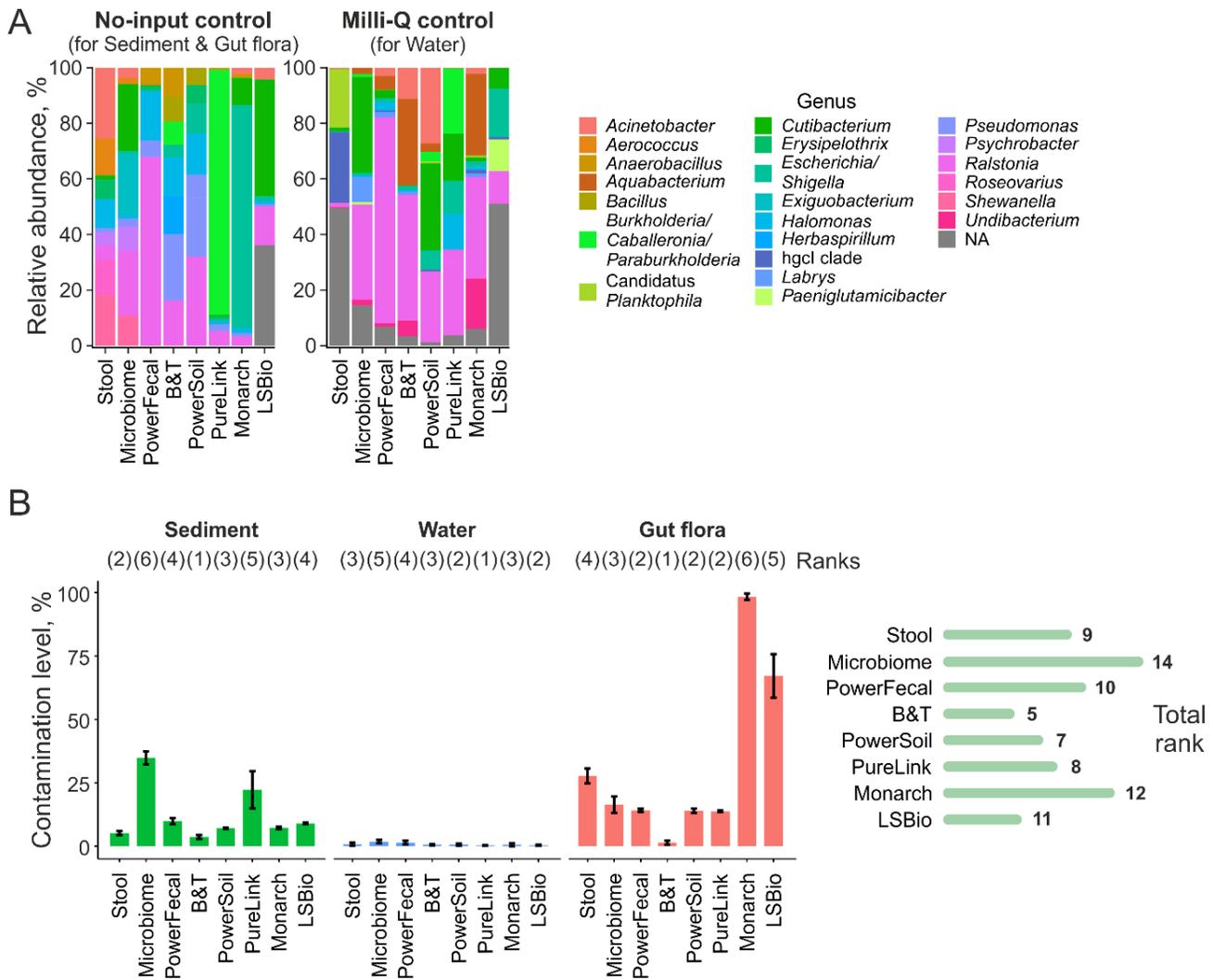


Figure 3.7. Determination of kitomes of commercial DNA extraction kits. (A) Diversity of kitomes on a genus level. Data is shown for genera with relative abundances >5%. (B) Left - contamination levels of natural samples processed with commercial kits. Means of three technical replicates +/- SD are shown. Kit rankings for each sample type are shown above the bar plot. Right - total kit ranking (a sum of ranks for different sample types). Lower ranks indicate lower contamination levels.

In the mock samples, the most abundant taxa were represented by Beta-proteobacteria (*Burkholderia*, *Ralstonia*), Gamma-proteobacteria (*Acinetobacter*, *Escherichia/Shigella*), Actinomycetia (*Cutibacterium*), and other bacterial genera known to be kit and laboratory contaminants (Fig. 3.7A) [370–373]. A full list of OTUs found in the mock samples can be found in Table S3.1.

Using the data obtained for mock samples, contamination was removed from natural samples with the Decontam package (47). Decontamination was performed in the “either” mode which utilizes both DNA concentration and the OTU frequency data. In order to effectively remove contamination, the probability threshold was set to an aggressive 0.5 value (compared to 0.1 default value). With this threshold, sequences that are more prevalent in mock rather than in natural samples, are considered as contaminants. The fraction of OTUs classified as contamination (the sum of contaminant OTU frequencies divided by the total frequency in the sample) was calculated for environmental samples and is further referred to as “contamination level”. Contamination levels varied significantly for different sample types with water samples associated with lower average level of contamination (0.8% vs 12% for sediment and 32% for gut flora samples), indicating that some sample types may be more prone for contamination than others. The increased contamination levels in samples from digestive tract of *M. gigas* are likely due to low amounts of bacterial DNA in these samples (**Table 3.2**). Contamination levels reached 20-80% in sediment samples processed with Microbiome and PureLink kits and in gut flora samples processed with Stool (also reflecting poor quality of this sample, see **Fig. S3.2**), Monarch, and LSBio kits (**Fig. 3.7B, S3.8**). After decontamination, the composition of these bacterial communities changed drastically (**Fig. S3.9**). The B&T kit demonstrated the lowest level of contamination for sediments and gut flora, while PureLink showed the lowest contamination level for water samples. DNA isolation with B&T and PowerSoil kits was associated with the overall lower contamination levels, while extraction with Microbiome and Monarch kits led to higher contamination levels (**Fig. 3.7B**).

3.3.8 Technical reproducibility of DNA extraction kits

The results of this section were obtained in collaboration with D. Sutormin.

To inspect the effect of DNA extraction kits on the technical variability of bacterial communities' composition, lists of OTUs were compared within sets of three technical replicates after the decontamination step. Reproducibility level was measured as a ratio of OTUs present in all three technical replicates to the total number of unique OTUs found in at least one replicate. Reproducibility levels varied significantly between sample types with the lowest and the highest levels associated with,

correspondingly, gut flora and water samples (**Fig. 3.8A, S3.10**). Lower reproducibility of gut flora and sediment samples may be caused by the low amounts of extracted microbial DNA subjected for sequencing (*M. gigas* samples are enriched with host DNA, **Table 3.2**) and/or by innate heterogeneity and high diversity of microbial communities (sediment samples, see below). Within sample types, kits demonstrated high variability. For sediment samples, the LSBio and PowerSoil kits designed to extract DNA from soil were ranked the first and the second best, while the PureLink kit performed the worst. For water samples, several kits (Stool, PowerSoil, and Monarch) performed equally well, while PowerFecal and LSBio kits demonstrated the lowest reproducibility levels. For gut flora samples, the PowerFecal and PowerSoil kits shared the first place, while the Monarch kit was the least reproducible, likely reflecting the poor quality of DNA (**Table 3.2**). Overall, in terms of reproducibility, PowerSoil outperformed other kits, ranked the first for two out of three sample types, while Microbiome and PureLink kits performed poorly (**Fig. 3.8A**).

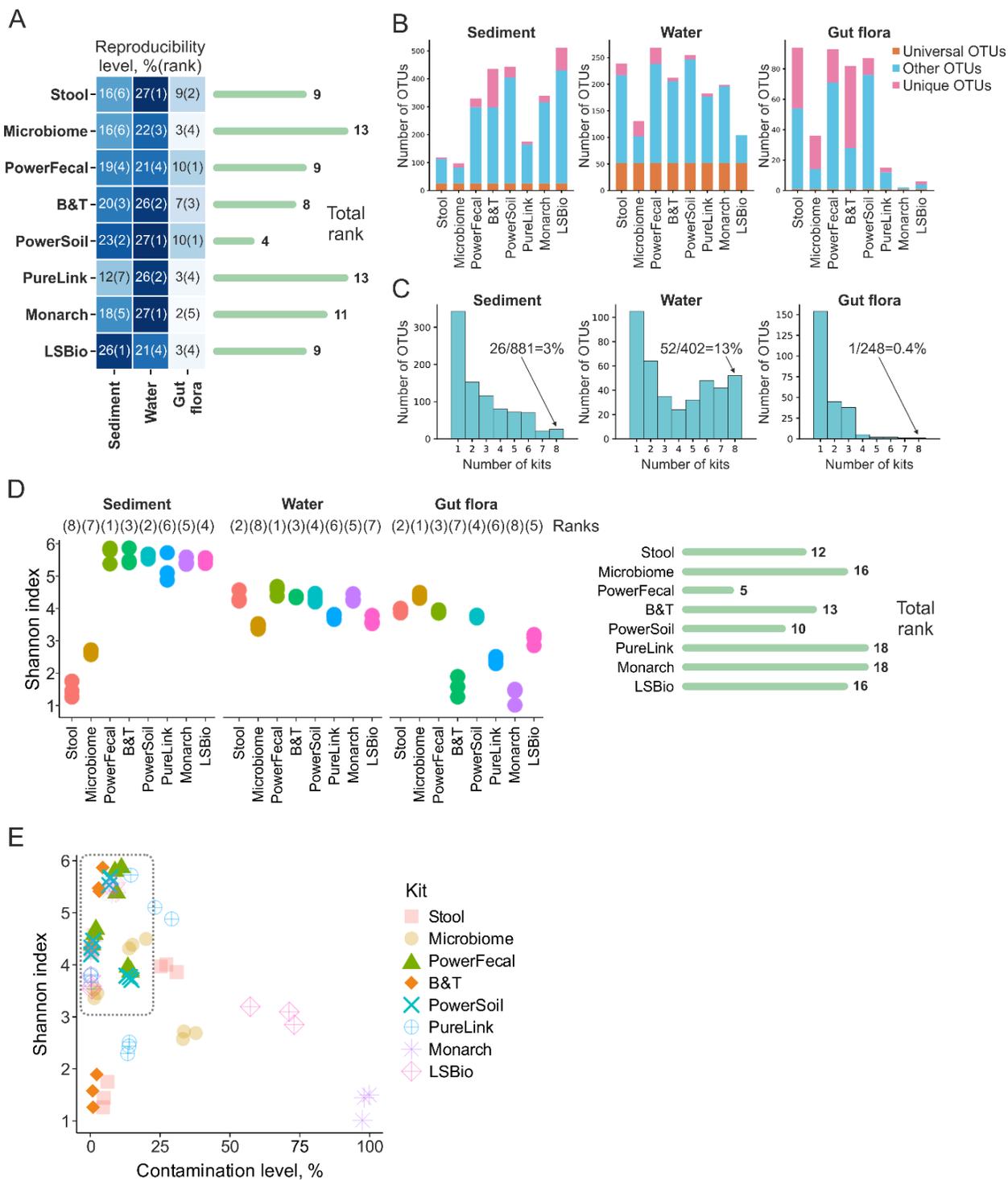


Fig 3.8. Reproducibility of DNA extraction kits and the effect of kits on alpha-diversity of bacterial communities. (A) Reproducibility levels of commercial DNA extraction kits for different sample types. Right - total kit rankings (a sum of ranks for different sample types). Lower rankings indicate higher reproducibility

levels. **(B)** Bar plots representing fractions of OTUs shared between all DNA extraction kits benchmarked for a particular sample type (Universal OTUs, orange), OTUs found by just one kit (Unique OTUs, pink), and other OTUs (found in DNA prepared by two to seven kits, light blue). **(C)** Fractions of OTUs shared between different numbers of DNA extraction kits. Numbers of OTUs found by all 8 studied DNA extraction kits are shown. **(D)** Microbial alpha diversity (the Shannon index) of DNA samples prepared using different kits. Data for three technical replicates are shown. Right - total kit rankings (a sum of ranks for different sample types). Lower rankings indicate higher Shannon index values. **(E)** Contamination level and alpha diversity (Shannon index) of samples processed with different DNA extraction kits. A group of samples with high diversity and low contamination levels is marked with a dashed rectangle.

3.3.9 Alpha diversity of microbial communities

I conducted the analysis of alpha diversity of different samples.

To assess the effects of DNA extraction kits on the diversity of microbial communities, alpha diversity (the Shannon index) was calculated for purified DNA samples after the contamination removal step (**Fig. 3.8D**). Higher Shannon indexes were observed for sediment samples and lower Shannon indexes were observed for gut flora samples, indicating high and low complexities of the two communities, correspondingly. The Shannon index was significantly decreased for samples processed with Microbiome (sediment and water samples), PureLink and LSBio (water and gut flora samples), Stool (sediment samples), and B&T and Monarch (gut flora samples) kits. The Microbiome kit was associated with the highest Shannon index for gut flora samples and the PowerFecal kit was associated with the highest Shannon indexes for soil and water samples. Considering contamination levels and alpha diversities associated with different DNA extraction kits, the PowerSoil and PowerFecal kits resulted in high alpha diversity and low contamination levels for all types of samples analyzed (**Fig. 3.8E**).

3.3.10 Effects of DNA extraction kits on the composition of bacterial communities

I conducted the beta-diversity analysis, PERMANOVA and the relative abundance visualisation. D. Sutormin assisted with reproducibility analysis.

To investigate the effects of DNA extraction kits on the composition of microbial communities, beta diversity of natural microbiomes after the decontamination step was assessed using the Bray-Curtis dissimilarity and NMDS for visualization. Visual inspection of relative abundance and NMDS

plots (**Fig. 3.9**) revealed outlier kits with distorted microbial community composition. All sample types processed with the Microbiome kit, sediment samples processed with the Stool kit, and water samples processed with the LSBio kit had dramatically distorted composition. Moreover, gut flora samples processed with LSBio and Monarch kits had distorted and highly heterogenous compositions, which corresponds to low level of technical reproducibility and poor quality of DNA purified with these kits (**Fig. 3.8A**). Using the PERMANOVA method, kit selection was found to be a factor significantly affecting the composition of microbial communities for all sample types (p -values=0.001).

We further estimated how the communities' composition overlaps between samples purified by different kits. For this analysis, only OTUs found in 3 replicates of DNA prepared with each kit were considered. As DNA from water samples was efficiently purified by most kits, more than 50 OTUs were identified in all samples, while the level of kit-specific OTUs was low (**Fig. 3.8B, C**). For sediment samples, the number of universal OTUs was lower, and the B&T kit purified more unique OTUs than other kits (**Fig. 3.8B, C**). Because of the overall lower quality of DNA from *M. gigas* and the failure of some kits to generate detectable amounts of DNA, these samples were the least consistent (**Fig. 3.8B, C**). With PureLink, Monarch, LSBio, and Microbiome kits, less OTUs were identified, while increased amount of unique OTUs was observed in Stool and B&T purified DNA.

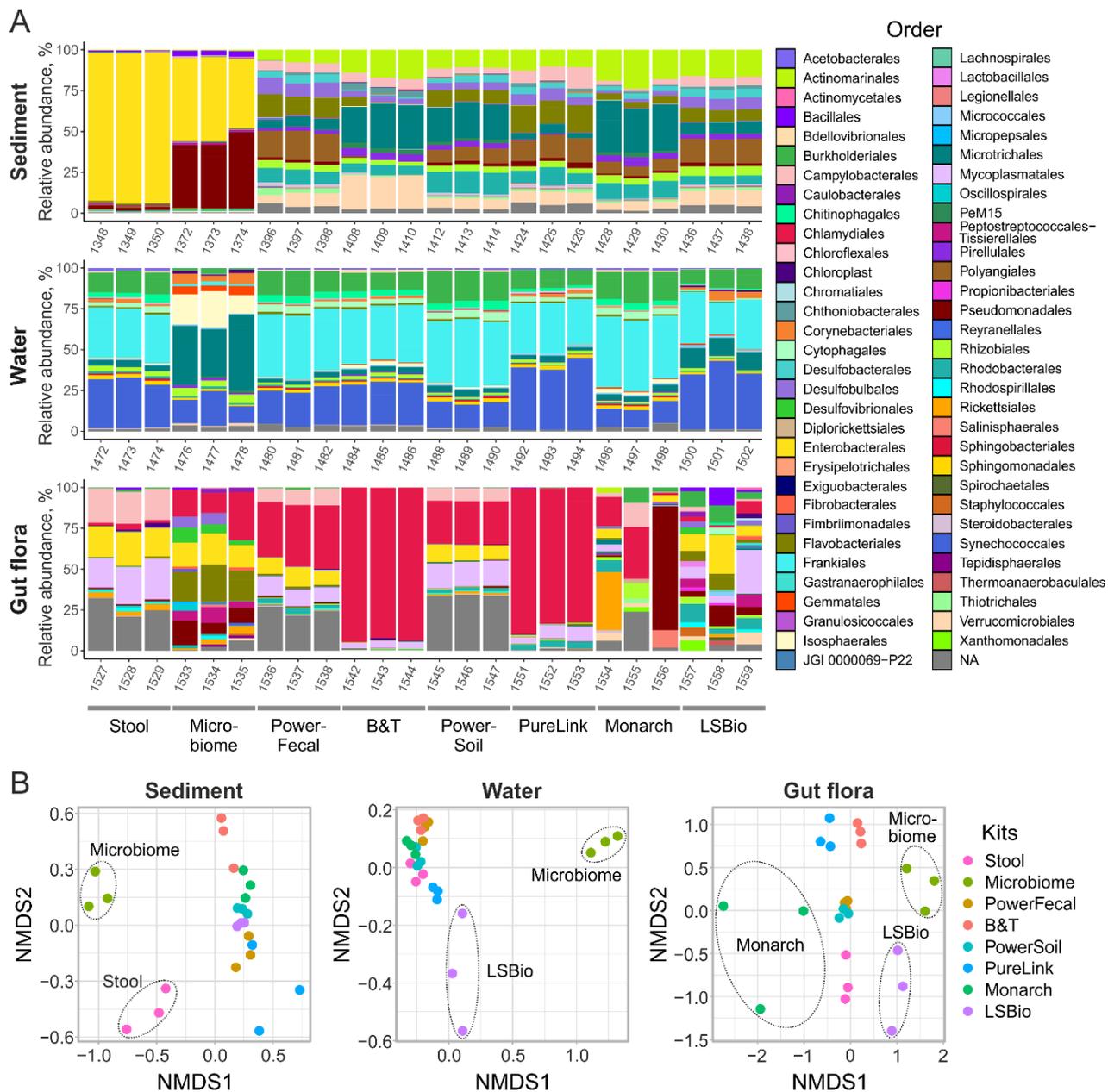


Figure 3.9. Effects of DNA extraction kits on the composition of bacterial communities. (A) Microbial communities' composition of environmental samples at the order level after the decontamination step. Data is shown for all technical replicates independently. Orders with relative abundances >1% are shown. (B) NMDS plots with points representing microbial communities. Dashed ellipses indicate sample groups distant from the majority of samples.

3.4 Discussion

All authors contributed in this section. I was responsible for the visualization of Fig. 3.10.

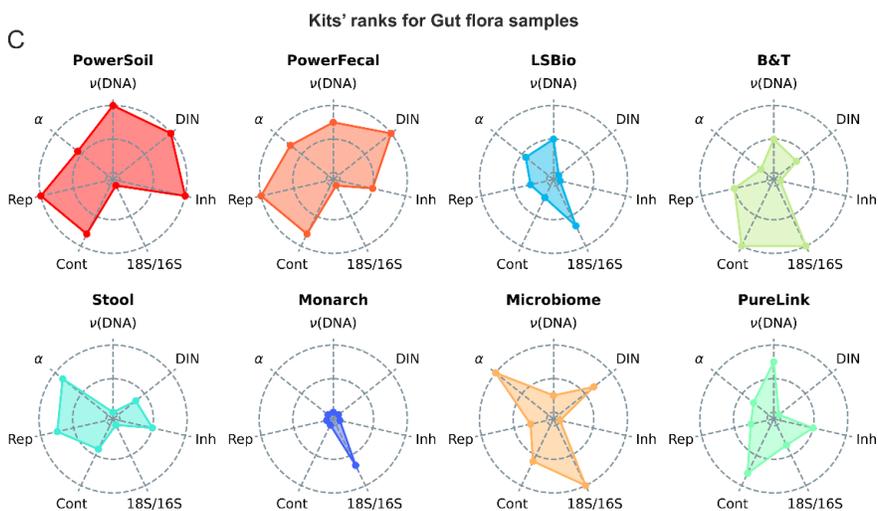
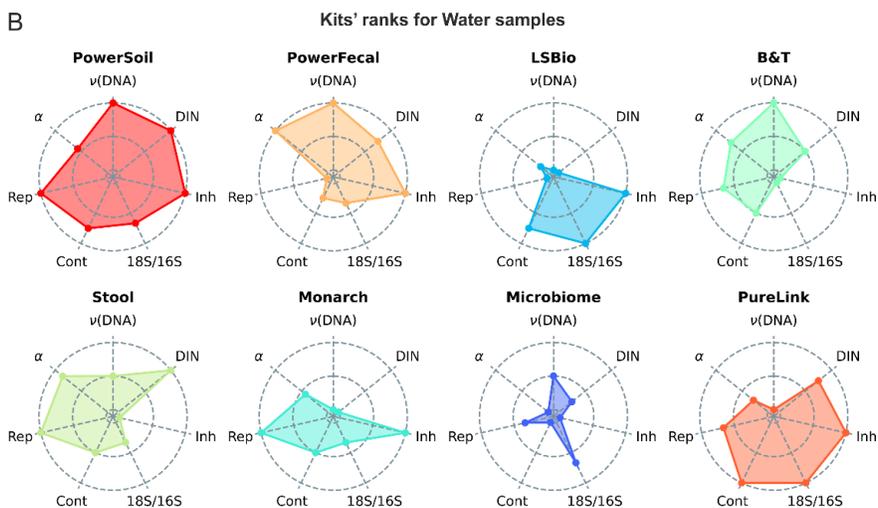
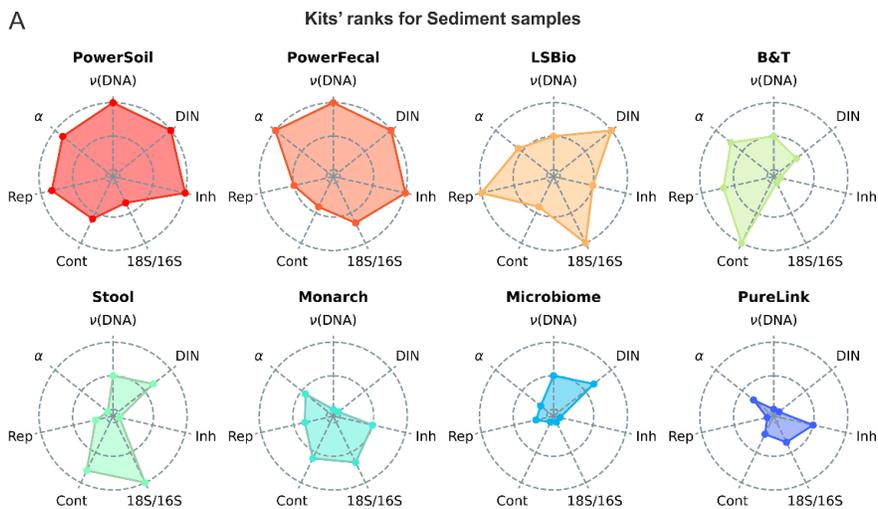
The results of any metagenomic analysis depend on the quality of input DNA. Given that DNA purification steps introduce multiple biases [374–378], it is highly advantageous to know the strong sides and limitations of available DNA isolation tools. Comprehensive study of marine microbial communities requires processing of different sample types, including water, sea floor sediments, and symbionts of multicellular organisms. DNA isolation from various types of sample may face challenges, such as low titers of microbial cells, the presence of PCR inhibitors, or eukaryotic DNA. Despite multiple studies discussing the choice of DNA isolation methods for human microbiota [319, 378–380] and some studies focused on specific marine communities [381, 382], a holistic investigation of DNA purification kits performance with marine samples is lacking.

To cover this gap, we selected eight commercially available DNA purification kits, and benchmarked them using three sample types (sea sediment, water, and gut flora of *M. gigas*). For each kit-sample combination, we measured the quality of purified DNA and its applicability for 16S metagenomics and described the resulting community composition. We estimated DNA concentration, DNA integrity (DIN), DNA purity (260/230 and 260/280 absorbance ratios), presence of PCR inhibitors, admixture of eukaryotic DNA (the 18S/16S ratio obtained using the Ct(18S)-Ct(16S) value), and three parameters describing the composition of microbial communities revealed by 16S metagenomics: contamination level, alpha diversity of microbial communities, and reproducibility of results. Using these parameters, the kits were ranked according to their performance. Higher DNA yield, DIN, reproducibility level, alpha diversity, and lower presence of inhibitors, 18S/16S ratio, and contamination level were considered to be linked to better performance.

Based on average rankings obtained for three sample types, the PowerSoil and PowerFecal kits, and the Microbiome and Monarch kits had the highest and the lowest performance levels, respectively (**Fig. 3.10, S3.11**). The PowerSoil and PowerFecal kits had the worst 18S/16S ratios, indicating that they may have a bias toward eukaryotic DNA extraction, which, however, did not have apparent negative effects on the profiling of microbial communities.

PowerSoil was ranked the best for all sample types, followed by PowerFecal for soil and gut flora samples and PureLink for water samples (**Fig. 3.10A, B**). The LSBio kit designed for DNA extraction from soil was ranked the third for this type of sample (**Fig. 3.10A**). The Microbiome kit designed to extract microbial DNA from samples with high eukaryotic load demonstrated the best 18S/16S ratio and high alpha diversity. However, the high level of PCR inhibitors and low reproducibility in our opinion limits its applicability, although by the sum of parameters the kit was ranked third for the gut flora samples (**Fig. 3.10C**). Overall, the PowerSoil and PowerFecal kits can be considered as most versatile and useful for a wide range of environmental samples. The LSBio and PureLink kits can be recommended for DNA extraction from, respectively, soil and water samples.

An inevitable source of biases associated with any DNA purification procedure is the presence of microbial DNA contaminants in kit solutions and columns, which may result in identification of OTUs that are actually absent from the studied sample. The set of such contaminants is referred to as the “kitome”. Using samples with no input material, we describe OTUs specifically associated with each tested kit. The list of detected OTUs can be used for decontamination at the stage of 16S data analysis. In addition, we studied the contribution of contamination from the laboratory Milli-Q water – another important source of non-specific 16S signal. Using Milli-Q water as an input, we described the combined composition of the “kitome” (kit-specific) and “splashome” (laboratory-specific) contaminations. The results showed that samples of poor quality in general demonstrated much higher levels of contamination.



1 2 3 4 5 6 7 8 Plot area ranking

Figure 3.10. Radar-plots demonstrating the performance of DNA extraction kits with Sediment (A), Water (B), and *M. gigas* gut flora samples (C). $v(\text{DNA})$ – DNA yield, DIN – DNA integrity, Inh – presence of PCR inhibitors (higher rank indicates the lower level of inhibitors), 18S/16S – 18S/16S ratio (higher rank indicates the lower ratio), Cont – contamination level (higher rank indicates the lower level of contamination), Rep – reproducibility level, α – alpha-diversity. Kits were ordered by the sum of rankings.

Our study has obvious limitations, as a multitude of additional parameters, such as details of sampling, storage and homogenization procedures can affect the quality of purified DNA and communities' composition [317, 346, 372, 380, 383]. Also, the usage of environmental samples for kit benchmarking does not allow to identify kit biases for specific bacterial taxa (“taxa-specific biases”), as the “ground truth” composition of investigated communities' was not known. To overcome this limitation, mock communities with an *a priori* known composition are typically used [384–386]. Mock communities, however, do not reflect biases associated with natural samples (“sample-specific biases”), e.g., heterogeneity, presence of PCR inhibitors, presence of eukaryotic cells, high complexity, etc. To account for both taxa-specific and sample-specific biases a combined approach is needed, based on benchmarking on mock and environmental communities. Nevertheless, purification of DNA from initially identical replicates of environmental samples with a set of eight DNA purification procedures allowed us to characterize sample-specific biases and highlight kits with the highest performance. While more elaborate DNA purification protocols can be further developed based on standard recommendations for these kits our results should aid in selection of the DNA isolation techniques most appropriate for different types of marine samples and downstream applications (amplicon and shotgun short-read or long-read NGS sequencing).

Chapter 4: Bacteriocin-Producing *Escherichia coli* Q5 and C41 with Potential Probiotic Properties: *In Silico*, *In Vitro*, and *In Vivo* Studies

The results of Chapter 4 are published in:

Mihailovskaya, V., Sutormin, D., Karipova, M., Trofimova, A., Mamontov, V., Severinov, K., Kuznetsova, M. Bacteriocin-Producing *Escherichia coli* Q5 and C41 with Potential Probiotic Properties: *In Silico*, *In Vitro*, and *In Vivo* Studies. *International Journal of Molecular Sciences*. 2023; 24(16):12636.

I conducted genome assemblies and annotations of assemblies for In Silico analysis of E. coli strains. In Vitro and In Vivo studies presented in this chapter were conducted by other authors. The text of the publication is presented only for the completeness of the storytelling.

4.1 Introduction

Infectious diseases cause significant economic damage to agricultural enterprises by reducing productivity and overall animal welfare [387]. Treatment of infectious diseases in farm animals is complicated by the spread of multidrug-resistance (MDR) among most pathogenic bacteria, including *E. coli* [388, 389]. With the growing need for the prevention and treatment of diseases caused by such strains, probiotics have gained new attention [390]. Probiotics are living microorganisms that benefit the host by improving microbial balance, regulating mucosal and systemic immunity, and antagonizing pathogenic and opportunistic microorganisms [391, 392]. Probiotics are particularly advantageous as they allow for a decrease in the use of antibiotics in livestock.

Commensal *E. coli* isolates producing antimicrobial peptides (bacteriocins) can help control the spread of pathogenic enterobacteria [393]. *E. coli* is known to produce two types of bacteriocins, classified by their molecular weight into colicins (>10 kDa) and microcins (<10 kDa), with diverse mechanisms of action: pore formation (colicins A, E1, K, N, U, S4, B, Ia, Ib, microcins V and L), nuclease activity (colicins E2, E3, E4, E5, E6, E7, E8, E9, D), inhibition of peptidoglycan biosynthesis (colicin M), inhibition of DNA gyrase (microcin B17), RNA polymerase (microcin J25), an aminoacyl tRNA synthetase (microcin C7), or ATP synthase (microcin H47) [394]. The widely used probiotic “Mutaflor” contains the *E. coli* Nissle 1917 strain producing microcins M and H47 [395, 396], while “Symbioflor 2” contains *E. coli* G3/10 producing microcin C7 [390]. The probiotics “Colibakterin”

and “Bifikol” used in Russia are developed on the basis of *E. coli* M-17, a producer of bacteriocins B, M, and microcin V [397].

The ability of bacteriocin-producing *E. coli* to inhibit pathogens *in vitro* is well described in many studies [398, 399]. However, only in a few studies has a direct correlation been demonstrated between the effectiveness of bacteriocin production *in vitro* and protection against pathogenic bacteria *in vivo* [400, 401]. The latter is determined, at least in part, by the colonization ability caused by the adhesive activity of a microorganism. High adhesive activity allows bacteria to stay in the intestine for longer periods of time and, consequently, increases the duration of their positive effect in the gastrointestinal tract on the host microbiota and immune system [402]. An additional requirement for a probiotic strain is the absence of virulence factors and antibiotic resistance genes (ARGs) [403].

Earlier, we characterized a collection of commensal bacteriocin-producing *E. coli* strains isolated from healthy farm animals and identified two strains that hold promise for probiotic development [404]. In this study, we present a comprehensive characterization of these strains, including whole-genome sequencing and analysis, and demonstrate their efficacy against pathogens *in vitro* and *in vivo*.

4.2 Results

4.2.1 *In silico* Analysis of *E. coli* Q5 and C41 Genomes

The results presented in sections 4.2.1.1-4.2.1.4 were conducted in collaboration with D. Sutormin. I was responsible for genome assemblies and annotations and D. Sutormin assisted with feature analysis of annotated genomes of E. coli strains.

4.2.1.1 General Genome Features

Complete genomes of *E. coli* Q5 and C41 were obtained by long-read sequencing using the Oxford Nanopore Technology (MinION) (Table S4.1). The genome of *E. coli* Q5 comprises a 4,948,409 bp chromosome with a GC content of 51.0% and five plasmids, named pQ501 (137,557 bp, GC content of 48.0%), pQ502 (98,317 bp, GC content of 48.0%), pQ503 (58,014 bp, GC content of 42.0%), pQ504 (17,107 bp, GC content of 48.0%), and pQ505 (7791 bp, GC content of 51.0%) (Fig. 4.1). The genome of *E. coli* C41 comprises a 5,037,330 bp chromosome with a GC content of 51.3% and two plasmids, named pC4101 (114,534 bp, GC content of 48.0%) and pC4102 (13,295 bp, GC content of 49.0%).

All plasmids share a high level of identity (>99%) with known sequences from the nt database (released on 2023-05-04). A total of 4945 protein-coding sequences (CDSs) were identified in the Q5 genome and 4878 CDSs—in the C41 genome. Both strains encoded 22 rRNAs, 89 tRNAs, and 1 tmRNA (**Table S4.1**). The Q5 genome shared 96.2% sequence average nucleotide identity with that of C41 (PMID: **26585518**). According to PHASTEST, the Q5 genome contains seven intact prophages, whereas the C41 genome contains 10 intact and one incomplete prophages (**Table S4.2**).

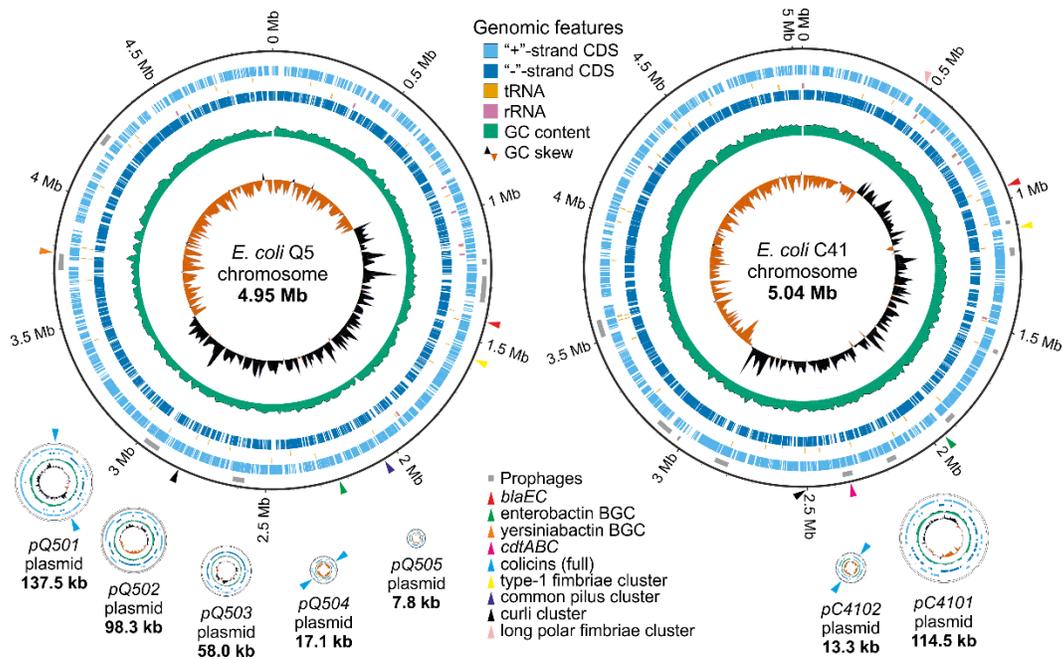


Figure 4.1. Genomes of *E. coli* Q5 and C41 strains were visualized using GenoVi [405]. Plasmids and chromosomes are not shown to scale. Grey rectangles indicate prophages; colored triangles indicate genomic features related to probiotic properties.

Plasmids pQ501, pQ502, pQ503, and pQ504 have low (~1–2) copy-numbers, whereas pQ505 has a high copy-number (~90) (**Table S4.3**). The pQ501 and pQ503 plasmids carry the F and P-type conjugation systems, respectively. The pQ504 and pQ505 plasmids encode mobilization systems. The pQ502 plasmid was recognized as a Punavirus (Uroviricota) by BLAST (89% coverage and 99% identity with the Punavirus P1 sequence MH422554.1). With the exception of pQ505, which was not assigned to any known incompatibility group, all Q5 plasmids belong to different incompatibility groups, providing supporting evidence that they are not the result of a misassembly. Both plasmids of the C41 strain are single-copy and belong to different incompatibility groups. The pC4101 plasmid

contains an F-type conjugation system, and the pC4102 plasmid has a mobilization system (**Table S4.3**).

4.2.1.2 Antimicrobial Resistance and Virulence-Associated Genes

Mobile ARGs and virulence-associated genes (VAGs) are undesirable in probiotic strains. Prediction of ARGs with Abricate and VRprofile2 in the *E. coli* Q5 and C41 genomes revealed, respectively, chromosomal *blaEC-5* and *blaEC-18* genes encoding beta-lactamases (marked with red triangles in **Fig. 4.1**). No ARGs were found in the plasmids or inside mobile genetic elements, and no mutations in the *gyrA/parC/parE* loci conferring resistance to fluoroquinolone antibiotics were detected. No enterotoxins, cytotoxins, or hemolysin genes were found in the Q5 genome. However, the C41 strain carried a chromosomal *cdtABC* gene cluster encoding the cytolethal distending toxin (CDT). Several biosynthetic gene clusters (BGCs) of iron-chelating compounds (siderophores) were detected in both genomes. Enterobactin and yersiniabactin BGCs were found in the Q5 chromosome (marked with green and orange triangles in **Fig. 4.1**). The enterobactin BGC was also detected in the C41 chromosome. The limited repertoire of ARGs, the absence of toxin-encoding genes, and the presence of siderophore biosynthesis genes make *E. coli* Q5 a promising probiotic candidate.

4.2.1.3 Adhesion-Related Genes

Adhesion is the first step in the colonization of the intestine by microorganisms and is thus a required property for a probiotic strain. Adhesion-related genes were found in both sequenced genomes. The *E. coli* Q5 chromosome carries genes encoding the type 1 fimbriae (*fimB*, *fimE*, and *fimAICDFGH*), the *E. coli* common pilus (*ecpABCDE*), curli (*csgDEFG* and *csgBAC*), and the FdeC adhesin (*fdeC*). The *E. coli* C41 chromosome contains gene clusters encoding the type 1 fimbriae (*fimB*, *fimE*, and *fimAICDFGH*), long polar fimbriae (*lpfABCD*), and curli (*csgDEFG* and *csgBAC*). The presence of detected adhesins should allow *E. coli* Q5 and C41 to attach to surfaces and colonize the intestine (see also below). Neither genome contains genes of fimbria associated with pyelonephritis (*pap*), fimbriae S and F1C (*sfa* and *foc*), afimbrial adhesins (*afa/dra*), or the *eae* gene, which codes for a protein required for the formation of attaching and effacing lesions.

4.2.1.4 Bacteriocin Gene Clusters

The production of bacteriocins is believed to help probiotic strains compete with pathogenic strains for an ecological niche [391]. In order to evaluate the antimicrobial potential of the two strains, their genomes were screened with antiSMASH, PRISM4, and BAGEL4. *E. coli* Q5 genome contained three sets of genes required for production and export of (and self-immunity to) colicins Ia and Ib (both on the pQ501 plasmid) and colicin Y (on the pQ504 plasmid) (marked with blue triangles in **Fig. 4.1**). An incomplete set of genes for microcin V production was found on the chromosome and in the pQ501 plasmid. Additionally, the *cbrA* gene responsible for resistance to colicin M was found on the Q5 chromosome (**Table 4.1**). The pC4102 plasmid contained a full set of functional genes needed for the production and export of pore-forming colicin E1 (*cea*, *cei*, *cel*). Additionally, the C41 chromosome contained the *cvpA* gene encoding microcin V production protein and the *cbrA* and *cbrC* genes conferring resistance to, respectively, colicins M and E2 (**Table 4.1**). The presence of complete gene sets for the production of different colicins supports the potential of the two strains (especially *E. coli* Q5) for development into probiotics.

Table 4.1. Description of bacteriocin-related genes found in the *E. coli* Q5 and C41 genomes.

Strain	Colicin	Genes Found
<i>E. coli</i> Q5	colicin Ia *	<i>cia</i> (QQ972_24345, pQ501), <i>iia</i> (QQ972_24350, pQ501)
	colicin Ib *	<i>cib</i> (QQ972_24045, pQ501), <i>iib</i> (QQ972_24050, pQ501)
	colicin Y *	<i>crl</i> , <i>cui</i> , <i>cya</i> (pQ504)
	microcin V	<i>cvaC</i> (frameshifted, pQ501), <i>cvi</i> (<i>cvi</i> and QQ972_24735, pQ501), <i>cvpA</i> (chr)
	colicin M	<i>cbrA</i> (chr)
<i>E. coli</i> C41	colicin E1 *	<i>cei</i> (QQ971_24605), <i>cea</i> , <i>cel</i> (QQ971_24650) (all pC4102)
	microcin V	<i>cvpA</i> (chr)
	colicin M	<i>cbrA</i> (chr)
	colicin E2	<i>cbrC</i> (chr)

Note. «*»—indicates a full set of functional genes required for bacteriocin production, processing, and export.

4.2.2 *In vitro* Analysis of *E. coli* Q5 and C41 Potential as Possible Probiotics

The results in sections 4.2.2.1-4.2.2.3 were conducted by other authors.

4.2.2.1 Antimicrobial Activity in Spent Media

Spent media (cell-free supernatants) from *E. coli* Q5 and C41 cultures grown for 22 h were tested for their ability to inhibit the growth of various test strains. Such inhibition can be expected if supernatants contain bacteriocin produced during cultivation. Spent medium from *E. coli* M-17, a component of the commercial probiotic “Colibakterin,” was used as a control. *E. coli* Q5 and C41 supernatants inhibited, to various extent, the growth of avian pathogenic *E. coli* (BR4, BR35, BR37), diarrheagenic *E. coli* (CA29, CA43, CA46), *Klebsiella pneumoniae*, and *Staphylococcus aureus* (Table 4.2, Fig. S4.1). *E. coli* O157 was modestly inhibited by the M-17 supernatant alone. *E. coli* Q5, but not other supernatants, inhibited the growth of *S. flexneri*. *E. coli* Q5 and C41 supernatants inhibited the growth of *E. coli* BR35 and CA46 more effectively than the M-17 control ($p < 0.05$). The C41 supernatant was also a better inhibitor of *Klebsiella pneumoniae* ($p < 0.05$). Neither supernatant affected on the growth of *Salmonella* Typhimurium, *Proteus mirabilis*, or *Pseudomonas aeruginosa*.

Table 4.2. Antagonistic activity of cell-free supernatants of *E. coli* M-17, Q5, and C41 against test-strains, M ± m.

Test-Strain	Growth Inhibition Index of Test-Strains after 22 h of Cultivation, %		
	<i>E. coli</i> M-17	<i>E. coli</i> Q5	<i>E. coli</i> C41
<i>E. coli</i> BR4	42.1 ± 9.5	38.3 ± 7.9	46.9 ± 13.3
<i>E. coli</i> BR35	31.9 ± 8.2	46.6 ± 5.4 *	55.9 ± 5.7 *
<i>E. coli</i> BR37	25.7 ± 11.3	41.3 ± 5.6	43.2 ± 2.6
<i>E. coli</i> CA29	38.4 ± 5.3	21.0 ± 3.5	22.4 ± 4.6
<i>E. coli</i> CA43	32.4 ± 5.2	24.6 ± 2.6	33.9 ± 4.5
<i>E. coli</i> CA46	10.2 ± 1.5	31.3 ± 2.8 *	37.0 ± 4.3 *
<i>E. coli</i> O157	12.6 ± 9.7	0	0
<i>K. pneumoniae</i>	35.6 ± 4.0	25.8 ± 0.9	55.1 ± 2.5 *
<i>S. aureus</i>	18.8 ± 7.3	15.2 ± 2.8	14.2 ± 3.8
<i>S. flexneri</i>	0	25.6 ± 11.4 *	0
<i>S. typhimurium</i>	0	0	0
<i>P. mirabilis</i>	0	0	0
<i>P. aeruginosa</i>	0	0	0

Note. «*»—significantly more than *E. coli* M-17 (t -test, $p < 0.05$).

4.2.2.2 Adhesion Ability

The level of nonspecific adhesion of *E. coli* Q5 to a hydrophilic surface was comparable, and in the case of adhesion to a hydrophobic surface, significantly lower than that of the control probiotic strain *E. coli* M-17 ($p < 0.01$). *E. coli* Q5 was low-adhesive to human red blood cells (RBC): the average adhesion index (AAI) was 0.83 ± 0.12 , the adhesion coefficient (AC) was 0.47 ± 0.07 , and the index adhesiveness of microorganisms (IAM) was 1.78 ± 0.09 (Fig. 4.2). However, this strain was medium-adhesive to bovine RBC (AAI = 2.08 ± 0.18 , AC = 0.76 ± 0.02 , and IAM = 2.73 ± 0.16). *E. coli* C41 had a level of adhesion to a hydrophobic surface comparable to that of *E. coli* M-17. Adhesion to a hydrophilic surface was $22.8 \pm 1.2\%$, which is significantly higher than that of the *E. coli* M-17 control ($p < 0.01$). *E. coli* C41 was classified as medium-adhesive and adhered well to both human RBC (AAI = 1.21 ± 0.04 , AC = 0.48 ± 0.01 , and IAM = 2.53 ± 0.09) and bovine RBC (AAI = 1.65 ± 0.08 , AC = 0.65 ± 0.08 , and IAM = 2.56 ± 0.05). It is worth noting that both strains, in contrast to *E. coli* M-17, had a greater affinity for bovine RBC, which could help to effectively colonize the intestines of animals.

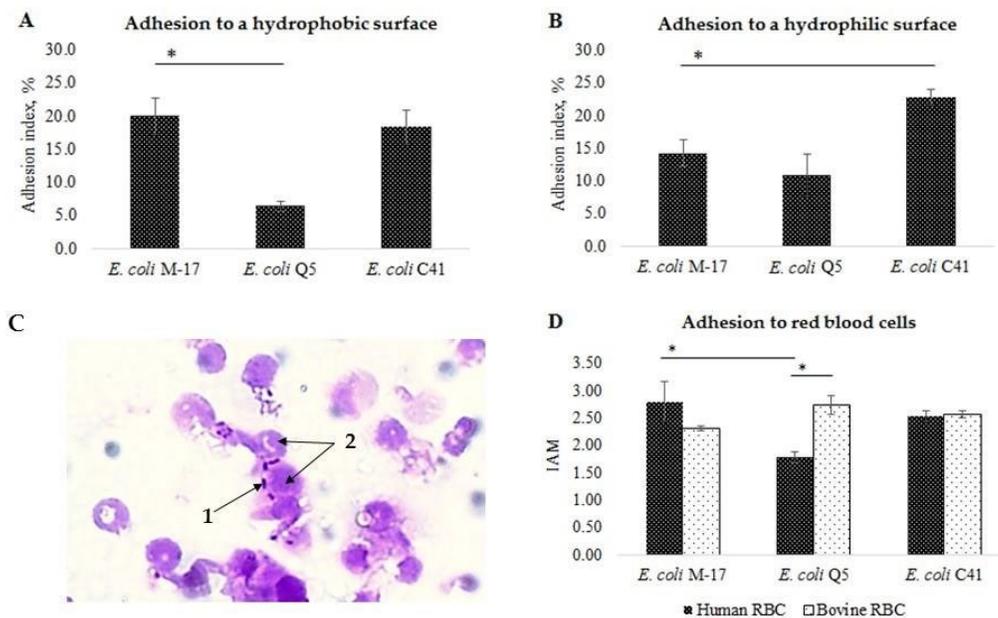


Figure 4.2. Adhesive properties of *E. coli* Q5, C41, and the control commercial probiotic strain M17. (A) Nonspecific adhesion to a hydrophobic surface (polystyrene). (B) Nonspecific adhesion to a hydrophilic surface (glass). (C) An example of specific adhesion of *E. coli* Q5, 30 min, staining with gentian violet, 1000 \times : 1—bacterial cells; 2—human red blood cells; (D) Level of specific adhesion. IAM—index adhesiveness of

microorganisms. Columns—means; bars—mean deviations; «*» indicates a significant difference between levels of adhesion between strains (*t*-test, *p* < 0.05).

4.2.2.3 Antimicrobial Susceptibility and Lysogeny

The antimicrobial susceptibility of *E. coli* Q5 and C41 was analyzed by the disc diffusion method for a number of antibiotics, including ampicillin, cefoperazone, ceftriaxone, cefepime, aztreonam, meropenem, gentamicin, amikacin, norfloxacin, ciprofloxacin, levofloxacin, tetracycline, and chloramphenicol. Both strains were sensitive to every antibiotic tested.

Lysogeny is a potentially high-risk factor for a probiotic strain [406]. Ultraviolet (UV) irradiation of both strains for 70 s and 150 s, a condition that mobilizes prophages, did not lead to lysis, suggesting that *E. coli* Q5 and C41 strains are not lysogenic despite the presence of multiple integrated prophages (**Table S4.2**), which might be inactivated by mutations and/or are not mobilized in our experimental conditions.

4.2.3 In vivo Analysis of Probiotic Properties of *E. coli* Q5 and C41

The results in sections 4.2.3.1-4.2.3.4 were conducted by other authors.

4.2.3.1 The Effect of *E. coli* Q5 and C41 on the Physiological Parameters of Rats

Upon five-day oral administration of *E. coli* Q5 and C41 at daily doses of 5×10^8 or 5×10^{10} colony-forming units (CFU)/per rat, the survival rate was 100%. There were no symptoms of disease or behavioral abnormalities; the animals were active. The average weight of rats fed with *E. coli* Q5 and C41 at a dose of 5×10^8 CFU/rat·day exceeded the control by 2.7% and 0.5%, respectively (**Table 4.3**). At the higher dose of *E. coli* Q5 and C41, the growth-stimulating effect was lost, and the average body weight (BW) became lower than in the control group.

Table 4.3. Measured physiological parameters of the rats, $M \pm m$.

Parameter	Control	<i>E. coli</i> Q5, CFU/Rat·Day		<i>E. coli</i> C41, CFU/Rat·Day	
		5×10^8	5×10^{10}	5×10^8	5×10^{10}
Survival rate, %	100	100	100	100	100
BW before administration, g	372 ± 35	358 ± 14	372 ± 26	349 ± 28	358 ± 14
BW after administration, g	392 ± 39	387 ± 22	371 ± 28	369 ± 27	372 ± 16
BW gain, g	20 ± 5	29 ± 13	-1 ± 5	20 ± 8	14 ± 6
BW gain, %	5.3 ± 1.1	8.0 ± 4.6	-0.3 ± 1.4 *	5.8 ± 2.4	4.0 ± 4.3

Note. BW—body weight. «*»—significant difference from the control (intact rats) (t -test, $p < 0.05$).

4.2.3.2 Composition of Rat Intestinal Microbiota after Administration of *E. coli* Q5 and C41 and upon Experimental Infection with Toxigenic *E. coli* C55 after Preliminary Administration of *E. coli* Q5 and C41

The basic content of microorganisms in the intestinal microbiota of rats before the experiment was: 8.0 ± 0.0 lg CFU/g *Bifidobacterium* and *Lactobacillus*; 6.1 ± 0.2 lg CFU/g *Enterococcus*; 5.4 ± 0.2 lg CFU/g *E. coli*; 7.8 ± 0.1 lg CFU/g *Staphylococcus*; and 4.3 ± 0.3 lg CFU/g *Candida albicans*. When comparing the content of the intestinal microbiota of animals after the administration of *E. coli* Q5 and *E. coli* C41, there were no significant differences in the content of *Bifidobacterium*, *Lactobacillus*, and *Enterococcus*, whose content varied in the range of 8.0–8.7 lg CFU/g, 7.3–8.7 lg CFU/g, and 5.3–6.1 lg CFU/g, respectively (**Fig. 4.3**). The content of *E. coli* significantly increased, from 5.4 ± 0.2 to 6.3 ± 0.1 lg CFU/g after administration of *E. coli* Q5 ($p = 0.003$) and from 5.4 ± 0.2 to 6.8 ± 0.4 lg CFU/g after the use of *E. coli* C41 ($p = 0.002$). Compared to the control, after simultaneous administration of both strains, there was a dramatic (average of two orders of magnitude) decrease in the number of *Staphylococcus* in the feces of rats.

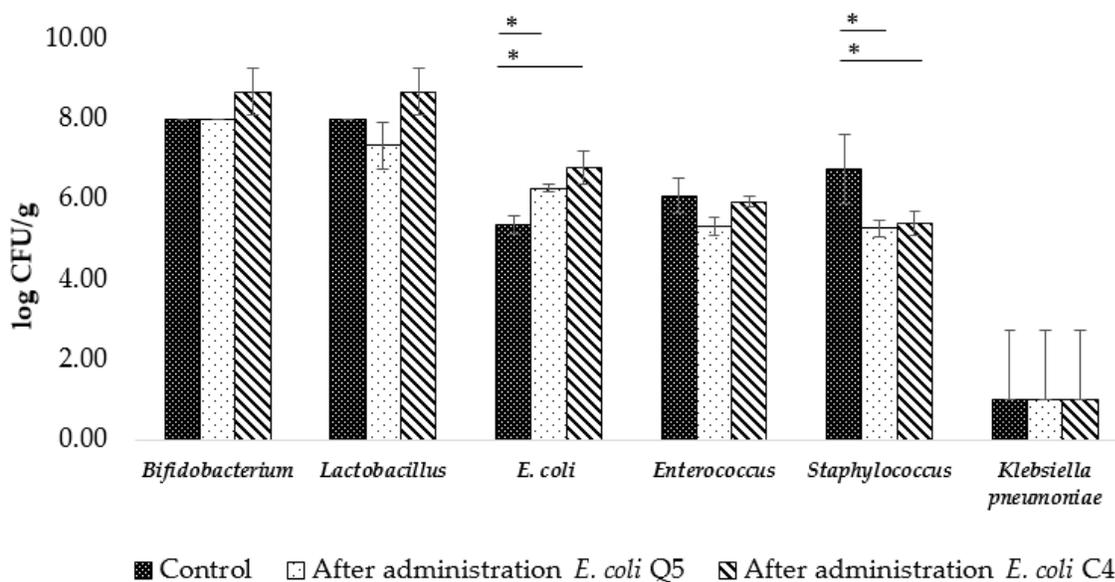


Figure 4.3. Changes in abundance of representative members of rat intestinal microbiota after 5-day administration of *E. coli* Q5 or C41 (5×10^8 CFU/rat-day). Columns—means; bars—mean deviations; «*» indicates a significant difference from control (intact rats) (*t*-test, $p < 0.05$).

As a result of toxigenic *E. coli* C55 infection, hemolytic *E. coli* (*E. coli* hem+) appeared in the intestinal microbiota of rats in the amount of 4.6 ± 1.2 lg CFU/g (**Fig. 4.4**). The amount of *E. coli* hem+ was significantly lower in animals to which *E. coli* Q5 was administered ($p = 0.04$). After administration of *E. coli* C41, *E. coli* hem+ was not detected. It is important to emphasize that in the infection control group, the number of *C. albicans* significantly increased ($p = 0.03$) and the content of *K. pneumoniae* was on average 2.2 lg CFU/g higher than in the group of animals that received the probiotic strains prior to the introduction of toxigenic *E. coli* C55.

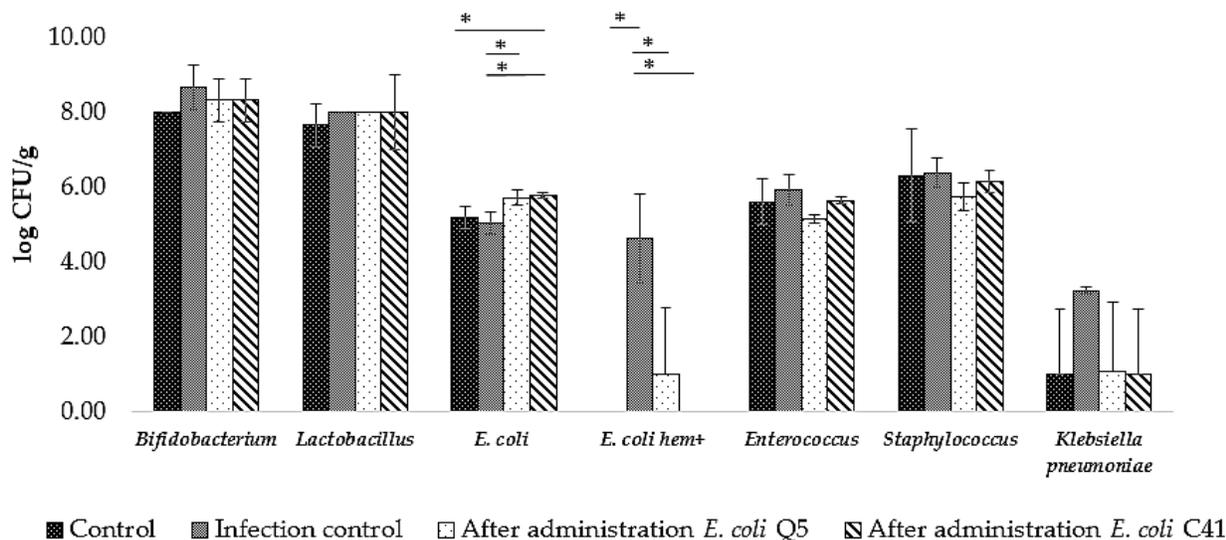


Figure 4.4. Changes in abundance of representative members of rat intestinal microbiota after experimental infection with toxigenic strain *E. coli* C55 with or without preliminary 5-day administration of *E. coli* Q5 or C41 (5×10^8 CFU/rat·day). Columns—means; bars—mean deviations; «*» indicates a significant difference from intact rats control or infection control (animals infected with *E. coli* C55) (*t*-test, $p < 0.05$).

4.2.3.3 Hematological and Biochemical Parameters of Rats after Administration of *E. coli* Q5 and C41 and during Experimental Infection with Toxigenic *E. coli* C55 after Preliminary Administration of *E. coli* Q5 and C41

The hematological and biochemical indices of rats are presented in **Table 4.4**. Red blood cell count (RBC), hemoglobin concentration ([Hb]), hematocrit (Ht), platelet count (PLT), white blood cell count (WBC), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), and glucose did not significantly differ from controls ($p > 0.05$) and were within the norm according to Wikivet [407]. After 5-day administration of *E. coli* Q5 or C41 at a dose of 5×10^8 CFU/rat·day, the proportion of monocytes significantly increased ($p < 0.05$) but remained within the normal range of 0–5% [407]. Total protein and urea were normal (59–78 g/L and 3.07–7.28 $\mu\text{mol/L}$ [408]) and did not show a significant difference ($p > 0.05$). The alanine aminotransferase (ALT) level decreased significantly after the introduction of probiotic bacteria but remained within the normal range of 35–80 U/L [409].

In the blood of rats in the control group infected with toxigenic *E. coli* C55, the concentration of alkaline phosphatase (ALP) increased 1.7 times but remained at the control level in animals that received *E. coli* Q5 or C41 prior to infection. In addition, the concentration of urea in the infection control group was slightly above the norm (3.07–7.28 $\mu\text{mol/L}$ [408])

Table 4.4. Hematological and biochemical indices of the rats, $M \pm m$.

Parameter	Probiotic Administration (5 Days)			Experimental Infection <i>E. coli</i> C55 (8 Days)			
	Control	<i>E. coli</i> Q5	<i>E. coli</i> C41	Control	Infection Control	After Administration <i>E. coli</i> Q5	After Administration <i>E. coli</i> C41
Hematological parameters							
RBC, $\times 10^{12}/\text{L}$	8.8 \pm 0.4	8.7 \pm 0.2	9.3 \pm 0.1	8.7 \pm 0.3	8.9 \pm 0.2	8.6 \pm 0.4	8.9 \pm 0.2
Hb, g/L	153 \pm 4	151 \pm 7	161 \pm 4	152 \pm 6	158 \pm 4	149 \pm 6	158 \pm 3
Ht, %	45 \pm 1	45 \pm 2	48 \pm 1 *	45 \pm 2	47 \pm 1	44 \pm 2	47 \pm 1
PLT, $\times 10^9/\text{L}$	760 \pm 94	653 \pm 34	673 \pm 80	839 \pm 112	620 \pm 94	739 \pm 39	770 \pm 75
WBC, $\times 10^9/\text{L}$	13.0 \pm 1.0	13.2 \pm 1.6	10.3 \pm 2.4	12.5 \pm 3.2	11.1 \pm 2.0	12.5 \pm 1.8	12.1 \pm 1.5
Neutrophils, %	26 \pm 7	38 \pm 7	29 \pm 6	37 \pm 11	38 \pm 8	29 \pm 9	36 \pm 7
Lymphocytes, %	71 \pm 8	55 \pm 9	67 \pm 6	65 \pm 7	58 \pm 7	54 \pm 19	60 \pm 7
Eosinophils, %	2 \pm 2	2 \pm 2	1 \pm 1	2 \pm 2	1 \pm 1	2 \pm 1	1 \pm 1
Basophils, %	0	0	0	0	0	0	0
Monocytes, %	1 \pm 0	5 \pm 2 *	3 \pm 1 *	2 \pm 1	3 \pm 2	4 \pm 2	2 \pm 1
MCV, fL	51.5 \pm 1.3	51.8 \pm 0.6	41.5 \pm 16.6	51.4 \pm 1.4	52.9 \pm 2.1	51.4 \pm 0.5	52.6 \pm 1.1
MCH, pg	17.6 \pm 0.4	17.4 \pm 0.3	17.4 \pm 0.3	17.5 \pm 0.3	17.8 \pm 0.5	17.4 \pm 0.1	17.7 \pm 0.2
MCHC, %	34.1 \pm 0.1	33.7 \pm 0.3	33.8 \pm 0.2	34.0 \pm 0.3	33.7 \pm 0.5	33.8 \pm 0.3	33.7 \pm 0.4
Biochemical parameters							
Glucose, mmol/L	6.8 \pm 0.3	6.5 \pm 0.5	7.1 \pm 0.4	6.5 \pm 0.3	6.6 \pm 0.4	6.7 \pm 0.4	7.1 \pm 0.5
Total protein, g/L	67.6 \pm 1.7	67.1 \pm 1.9	66.3 \pm 2.1	73.1 \pm 2.9	71.2 \pm 2.8	68.9 \pm 2.1	69.0 \pm 2.2
Creatinine, $\mu\text{mol/L}$	64.2 \pm 1.0	59.2 \pm 3.4 *	55.8 \pm 3.8 *	61.8 \pm 6.8	54.4 \pm 4.2	45.9 \pm 4.7 #	47.2 \pm 4.2 #

Urea, $\mu\text{mol/L}$	7.1 \pm 0.6	7.2 \pm 0.3	7.1 \pm 0.8	6.8 \pm 0.7	7.6 \pm 0.6	7.2 \pm 0.5	6.8 \pm 0.5
ALP, U/L	475 \pm 135	539 \pm 117	604 \pm 101	521 \pm 69	921 \pm 79 #	569 \pm 73 α	602 \pm 43 α
ALT, U/L	85.2 \pm 9.0	69.3 \pm 7.8	60.8 \pm 3.9 *	90.6 \pm 7.9	93.0 \pm 7.6	92.3 \pm 8.4	75.9 \pm 4.8 # α

Note. RBC—red blood cell, Hb—hemoglobin, Ht—hematocrit, PLT—platelet count, WBC—total leukocyte count, PI—phagocytic index, ALP—alkaline phosphatase, ALT—alanine aminotransferase, MCV—mean corpuscular volume, MCH—mean corpuscular hemoglobin, MCHC—mean corpuscular Hb concentration. «*» or «#»—significant difference from the control (intact rats, 5 days) and control (intact rats, 8 days), respectively (*t*-test, $p < 0.05$), α —significant difference from the infection control group in which animals were infected with *E. coli* C55 (*t*-test, $p < 0.05$).

4.2.3.4 Histological Analysis of Small Intestine, Peyer’s Patches, Spleen, and Liver Morphology of Rats in Experimental Infection with Toxigenic *E. coli* C55 after Preliminary Administration of the *E. coli* Q5 and C41 Strains

Administration of toxigenic *E. coli* C55 did not cause lethal effects but led to the appearance of distinct histopathological changes in the organs of rats in the infection control group compared to uninfected animals (Fig. 4.5, 4.6). Lymphocytic cell infiltrates were found in the liver lobules. Hepatocytes showed degenerative changes: vascularization and dystrophic inflammation of liver cells were observed. Scattered areas of hemorrhage were recorded in the hepatic parenchyma of infected animals (Fig. 4.6a). There was swelling of the stroma and the subepithelial part of the villi in the small intestine, and congestion of blood and lymphatic vessels was recorded in the mucosal and submucosal layers. There was an abundance of lymphocytes and granulocytes in the stroma of villi and crypts. Desquamation of the epithelium was observed on the surface of the mucous membrane of the small intestine (Fig. 4.6b). In addition, activation of lymphoid tissue in Peyer’s patches, especially in the B-dependent zone, was detected compared with the control group (Fig. 4.6c). In the colon, there was an increase in focal lymphocytic infiltration of the intestinal wall compared to the control—accumulations of lymphocytes are determined in the mucosa and submucosa, as well as in the muscular and serous layers (data not shown). In the spleen, swollen stromal cells were determined in the red and white pulp. Lymphoid nodules of the white pulp were predominantly medium and small; most of them did not contain germinal centers (Fig. 4.6d). There was also a decrease in the number and size of secondary follicles compared with the control group.

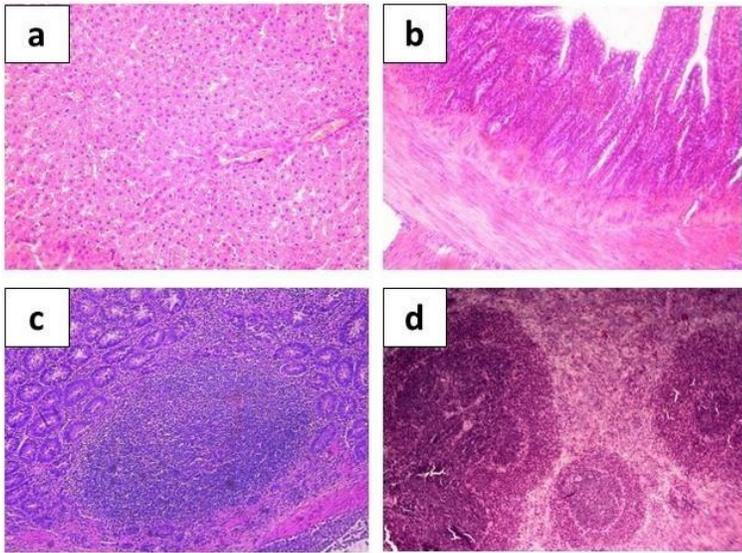


Figure 4.5. Hematoxylin-eosin-stained organ sections taken from control group animals show normal histological structures of rat hepatic tissue (**a**, $\times 200$), intestine (**b**, $\times 200$), Peyer's patches (**c**, $\times 400$), and spleen (**d**, $\times 200$).

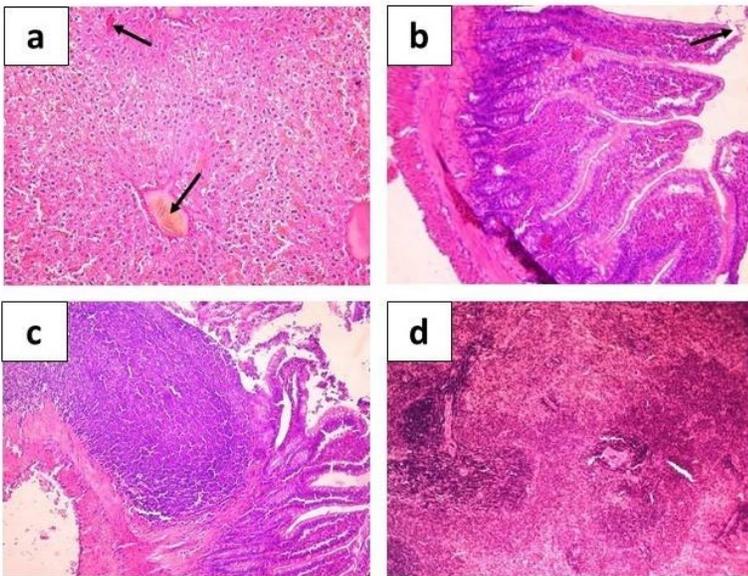


Figure 4.6. Hematoxylin-eosin stained organ sections of rats infected with *E. coli* C55 show histological structures of liver parenchyma with hemorrhage sites and congestion of blood vessels (**a**, arrows, $\times 200$), intestine with desquamation of the epithelium (**b**, arrow, $\times 200$) swelling of the stroma and the subepithelial part of the villi, Peyer's patches (**c**, $\times 400$), and spleen (**d**, $\times 200$).

Compared to infection control, a noticeable improvement in the state of organs was recorded in the group of rats infected after a preliminary 5-day administration of *E. coli* Q5 and C41 at a dose of $5 \times$

10^8 CFU/rat·day. There were no infection-associated changes in the liver (**Fig. 4.7a, 4.8a**), and in the small intestine, epithelial cells formed an even monolayer without epithelial desquamation foci (**Fig. 4.7b, 4.8b**). Compared with the control group (intact rats), an increase in the number of active goblet cells and an increase in the mitotic activity of cells in the crypts were visually noted, and moderate diffuse lymphocyte infiltration of the mucosa was diagnosed. These data indicate the ability of probiotics to positively influence epithelial cell tight junction stability and intestinal goblet cell mucus production. Peyer's patches were represented by clusters of large lymphoid nodules located in the mucosa and submucosa of the intestine (**Fig. 4.7c, 4.8c**). The nodules contained large germinal centers, occupying most of the follicle. In addition, there were many secondary lymphoid nodules containing germinal centers in the spleen (**Fig. 4.7d, 4.8d**). These data indicate antigenic stimulation of the host by probiotic bacteria, which in turn should stimulate intestinal immune cells, which contribute to the induction of mucosal immunity.

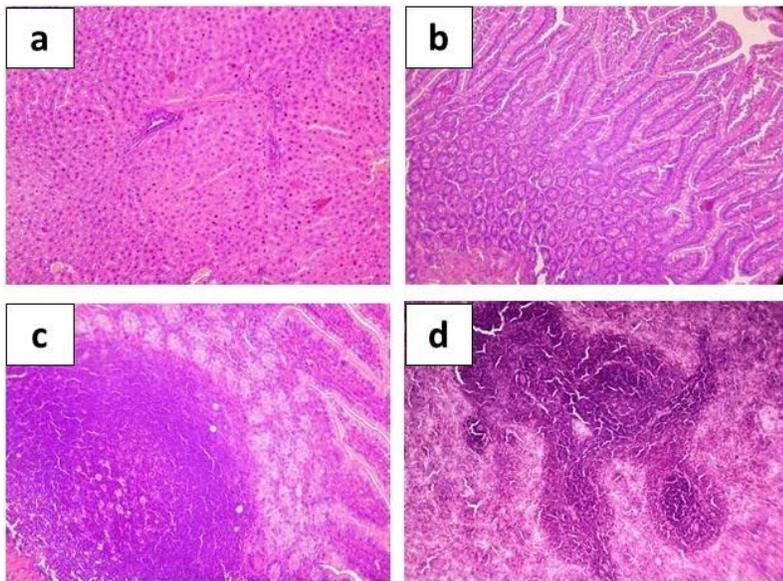


Figure 4.7. Hematoxylin-eosin stained organ sections taken from rats after experimental infection with toxigenic strain *E. coli* C55 after 5-day preliminary administration of *E. coli* Q5 (5×10^8 CFU/rat·day) showing histological structure of hepatic tissue (**a**, $\times 200$), intestine (**b**, $\times 200$), Peyer's patches (**c**, $\times 400$), and spleen (**d**, $\times 200$).

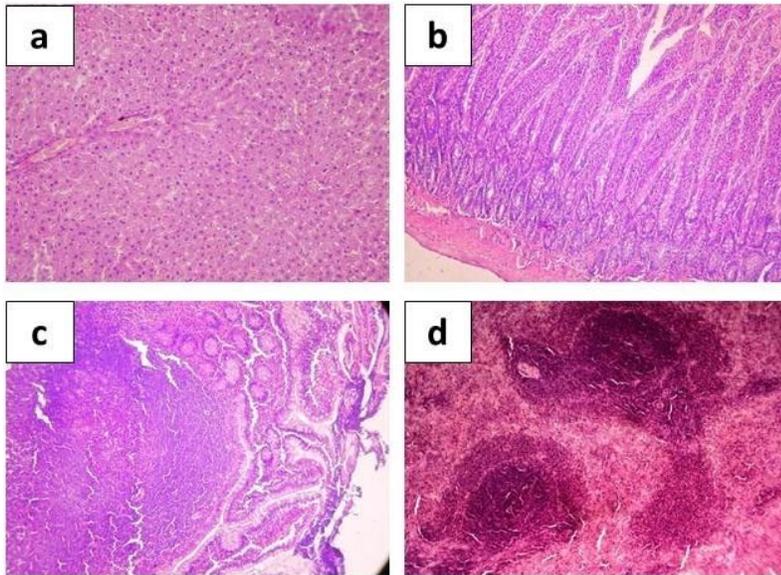


Figure 4.8. Hematoxylin-eosin stained organ sections taken from rats after experimental infection with toxigenic strain *E. coli* C55 after 5-day preliminary administration of *E. coli* C41 (5×10^8 CFU/rat·day) showing histological structure of hepatic tissue (**a**, $\times 200$), intestine (**b**, $\times 200$), Peyer's patches (**c**, $\times 400$), and spleen (**d**, $\times 200$).

4.3 Discussion

Probiotics are living microorganisms that play an important role in maintaining overall health, strengthening the immune system, and preventing severe intestinal diseases in farm animals [392, 410]. Significant progress has been made in the field of probiotics in recent decades; however, their mechanisms of action are still not fully understood. In this work, we describe and present two bacteriocin-producing *E. coli* strains that, based on the results of our analysis, hold promise for development as probiotics.

The production of bacteriocins is a key mechanism that allows probiotic *E. coli* to compete with pathogenic microorganisms in the intestine by inhibiting their growth [411]. Numerous studies have shown that *E. coli* bacteriocins are effective against diarrheagenic *E. coli* [398, 399, 401] and related enteropathogenic bacteria such as *Klebsiella*, *Salmonella*, and *Shigella* [412, 413]. The antagonistic effect of commercial probiotic strain *E. coli* M-17 is due to the production of pore-forming colicin B and microcin V, which inhibit the synthesis of peptidoglycan by hydrolyzing lipid II [397]. The antagonistic properties of *E. coli* Nissle 1917 are due to the siderophores microcin M and microcin

H47, which inhibit the ATP synthase [395, 396]. For the *E. coli* Q5 strain studied in this work, antagonistic *in vitro* and *in vivo* activity is probably associated with the production of pore-forming colicins Ia, Ib, and Y since corresponding complete biosynthetic gene clusters have been found in its genome. In the case of *E. coli* C41, the production of pore-forming colicin E1 is the likely reason for antagonistic activity. Pore-forming bacteriocins bind to receptors of the Toll or Ton systems and become embedded in the lipid bilayer, leading to the formation of channels and leakage of cellular contents [394]. Both strains contain genes responsible for resistance to colicin M (and additionally to colicin E2 for *E. coli* C41), which should prevent their displacement by resident or pathogenic bacteriocin producers.

E. coli Q5 and C41 demonstrated *in vitro* antagonistic activity against most enteropathogens tested, including *E. coli* causing colibacillosis in farm animals. Oral administration of *E. coli* Q5 and C41 *in vivo* eliminated *S. aureus* and decreased *K. pneumoniae* titers during experimental infection with the enterotoxigenic *E. coli* C55. This is a very promising result given that *S. aureus* is considered the main causative agent of “contagious” mastitis in bovines [414], and *K. pneumoniae* is often associated with pneumonia and septicemia in foals [415].

Another anti-pathogenic mechanism of probiotic action is the binding and blocking of receptors in intestinal epithelial cells. The effectiveness of the interaction of microorganisms with surfaces depends on the expression of an extensive repertoire of genes encoding fimbrial and afimbrial adhesins [416, 417]. The type 1 fimbriae encoded by *E. coli* Q5 and C41 attach in a mannose-dependent manner to eukaryotic cell receptors [417]. The *E. coli* Nissle 1917 probiotic strain also has type 1 fimbriae and is curly [396]. Enteropathogenic *E. coli* uses type 1 fimbriae to attach to intestinal epithelial cells. Thus, probiotic strains exclude the binding of pathogens by attaching to the same receptors. Adhesive amyloids (curly), encoded by both strains, are involved in adhesion to surfaces [418]. The FdeC adhesin encoded by *E. coli* Q5 has a high affinity for epithelial cells and provides protection against urinary tract infections [419].

Strains with good adhesive ability colonize the intestine better [420]. Yet, high-adhesive strains are not considered promising for probiotic development, as they displace not only pathogenic but also autochthonous microorganisms [421]. Therefore, most of the probiotic strains used are low- or medium-adhesive [396]. *In vitro* studies have shown that both strains were medium-adhesive and had

a greater affinity for bovine RBC than *E. coli* M-17, which may allow *E. coli* Q5 and C41 to effectively colonize the intestines of animals.

A crucial property for the practical application of probiotic strains is biosafety. Functional annotation of the *E. coli* Q5 genome allowed us to confirm the absence of enterotoxin genes, hemolysins, virulence-associated fimbriae (such as *pap*, *sfa*, *afa/dra* operons), and mobile ARGs. The *E. coli* C41 chromosome encodes the Cdt toxin, a pathogenicity factor. However, there was no toxic effect when *E. coli* C41 was administered at a dose of 5×10^{10} CFU/rat·day. The presence of *cdtABC* is a risk factor that; however, can be eliminated by removing the gene. Another risk factor is the presence of several prophage elements in the genomes of *E. coli* Q5 and C41. However, our *in vitro* studies indicated the apparent absence of lysogenic activity in both strains.

The two strains had no negative effect on the physiological (the body weight gain was within the norm) or hematological parameters of the rats. The total proteins, ALP, and ALT were normal, which confirms the absence of hepatotoxicity in the strains [410, 422]. In addition, the administration of *E. coli* Q5 and C41 countered the increase in ALT levels during experimental infection with a toxigenic *E. coli*. According to Shahverdi et al., this effect is characteristic of probiotic strains and indicates their hepatoprotective role [422]. Maintaining the levels of urea and creatinine in the normal range indicates the absence of any kidney disorders in rats. Finally, administration of both strains did not lead to significant changes in the composition of the native microbiota; the level of beneficial representatives such as *Bifidobacterium*, *Lactobacillus*, and *Enterococcus* remained unchanged.

Stabilization and maintenance of the integrity of the intestinal barrier are mechanisms of probiotic action that provide protection against pathogens and the toxins they produce. The most severe animal diseases are caused by enterohemorrhagic (producing shiga-toxin Stx1 and/or Stx2 that stop protein synthesis in endothelial target cells) and enterotoxigenic (producing enterotoxin EAST1 and/or enterohemolysin EhxA) *E. coli* strains [401, 423–425]. These toxins, produced by beta-lactam-resistant *E. coli* C55, lead to pathological changes in the intestines of infected rats (inflammation, epithelial desquamation, focal lymphocytic infiltration), and increased ALP levels. We show that preliminary administration of bacteriocin-producing *E. coli* for 5 days at a dose of 5×10^8 CFU/rat·day protected rats from colonization and pathogenic effects of *E. coli* C55. If the proportion of *E. coli* hem⁺ in the infection control group was more than half of all *Escherichia*, then after the preliminary administration

of *E. coli* Q5, *E. coli* hem⁺ were detected in only one animal, and after the administration of *E. coli* C41, *E. coli* hem⁺ were not detected at all. Pre-emptive oral administration of our strains prevented the destruction of the intestinal barrier (there was no epithelial desquamation or inflammation), presumably by blocking the access of *E. coli* C55 and its metabolites to subepithelial cells. In a mouse model, the introduction of *E. coli* Nissle 1917 protected the intestinal barrier from dysfunction due to a more pronounced expression of the tight junction molecules regulating intestinal permeability [426]. The mechanism(s) of intestinal barrier protection operational in the cases of *E. coli* C41 and Q5 remain to be determined.

In summary, positive effects of bacteriocin-producing *E. coli* are associated with inhibition of enteropathogens through bacteriocin production, competition for adhesion sites, improving the balance of the natural intestinal microbiota, and maintaining the integrity of the epithelial barrier by stimulating the secretion of mucin glycoproteins, antimicrobial proteins, tight junction molecules, modulation of metabolic and immune processes, and likely other mechanisms. Thus, our work demonstrated that short-term oral administration of *E. coli* Q5 and C41 to rats contributed to the preservation of intestinal homeostasis and provided protection from external influences, including infection with an enterotoxigenic beta-lactam-resistant *E. coli* strain. Given all the evidence, these two strains are promising candidates for development as probiotics for farm animals.

4.4 Materials and Methods

4.4.1 Bacterial Strains (conducted by other authors)

Earlier, we studied 97 *E. coli* isolates obtained from fecal samples of healthy farm animals from industrial and private farms in Russia [404]. As a result of the study, two bacteriocin-producing strains were selected: *E. coli* Q5 was obtained from a healthy quail and *E. coli* C41 from a healthy cow. These strains presumably had high probiotic potential. *E. coli* Q5 and *E. coli* C41 strains were deposited in the All-Russian Collection of Microorganisms (VKM) under the numbers B-3706D and B-3707D, respectively. A toxigenic strain of *E. coli* C55 was isolated from a calf with diarrhea. *E. coli* C55 produced intestinal toxins (Stx1, East1, and EhxA) and was resistant to beta-lactam antibiotics (ampicillin, ceftriaxone, cefepime, and cefoperazone). This strain was used in the current work to

simulate experimental toxicoinfection. The characteristics of all strains used in this work are presented in **Table 4.5**.

Table 4.5. Bacterial strains used in this work.

Strain	Collection/Source	Collection Number
Studied bacteriocin-producing strains		
<i>Escherichia coli</i> Q5	VKM/Feces of healthy quail from industrial farms, Perm, Russia	B-3706D
<i>Escherichia coli</i> C41	VKM/Feces of healthy cattle from industrial farms, Perm, Russia	B-3707D
Test-strains used for the antagonistic activity experiment		
<i>Escherichia coli</i> BR4	“Ex culture collection“, University of Ljubljana, Slovenia	L-5838
<i>Escherichia coli</i> BR35		L-5865
<i>Escherichia coli</i> BR37		L-5868
<i>Escherichia coli</i> CA29	Feces of cattle from industrial farms, Perm, Russia	-
<i>Escherichia coli</i> CA43		-
<i>Escherichia coli</i> CA46		-
<i>Escherichia coli</i> O157	State collection of pathogenic microorganisms and cell cultures (SCPM), Obolensk, Russia	240329
<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> ATCC 700603		B-7474
<i>Proteus mirabilis</i> №H-237		160120
<i>Pseudomonas aeruginosa</i> ATCC27853		41501
<i>Staphylococcus aureus</i> ATCC 6538 (FDA 209P)		201108
<i>Shigella flexneri</i> №170		232151
<i>Salmonella enterica</i> serovar Typhimurium №1135	Feces of a patient with acute enteritis in the medical facility, Perm, Russia	-
Control strain		
<i>Escherichia coli</i> M-17	“Colibakterin”	-
Strain used to simulate experimental infection		
<i>Escherichia coli</i> C55	Feces of cattle from industrial farms, Perm, Russia	-

4.4.2 Genome Sequencing and Assembly

A. Trofimova conducted genome sequencing. Genome assemblies were conducted by me.

E. coli Q5 and *E. coli* C41 genomic DNA was extracted from the overnight cultures grown at 37 °C using the GeneJET Genomic DNA purification kit (Thermo Scientific, Waltham, Massachusetts, USA). DNA was sequenced using Oxford Nanopore Technologies (ONT). Sequencing libraries were prepared from the non-sheared DNA using the Native Barcoding kit (SQK-NBD114-24; ONT, Oxford, UK) with enrichment of long fragments using the Long Fragment Buffer (LFB) according to the manufacturer’s protocol. Sequencing was performed on MinION using the R10.4.1 flow cell (FLO-

MIN114; ONT, Oxford, UK) with a translocation rate of 400 bps. Basecalling was performed using Guppy 6.0.1 [427] in the “hac” mode. Default parameters were used for all software unless otherwise specified. Draft genomes were assembled with Flye (v 2.9.1) [428]. The assembly was subsequently polished with medaka (v 1.7.2) using ONT reads, and assembly graphs were manually inspected in Bandage (v 0.8.1) [429].

4.4.3 Genome Annotation and Analysis

Genome annotation was conducted by me. D. Sutromin assisted with analysis of annotated genomes.

Polished genome assemblies containing circular replicons were further annotated using PGAP (v 6.1) [430]. Virulence-associated genes (VAGs) were detected using VirulenceFinder (v 2.0) [431] and VRprofile2 [432]. Antibiotic-resistance genes (ARGs) were predicted with Abricate [433] using the NCBI AMRFinderPlus database [434]. Mutations conferring antibiotic resistance were searched using ResFinder (v 4.1) [431]. Biosynthetic gene clusters (BGCs) and bacteriocins were predicted using antiSMASH (v 7.0) in a “loose” mode [435, 436], PRISM4 [437], and BAGEL4 [438]. Prophages were predicted with PHASTEST [439]. Plasmid incompatibility groups were predicted with PlasmidFinder-2.0 [440].

4.4.4 Data Deposition

I prepared data for the deposition. D. Sutromin conducted the data deposition.

Raw reads for *E. coli* C41 and *E. coli* Q5 whole-genome sequencing were deposited in the Sequence Read Archive (SRA) under SRR24834172 and SRR24834173 accessions, respectively. Annotated genome assemblies obtained in this study were deposited in the NCBI BioProject PRJNA980458, GenBank accession numbers CP127252-CP127254 (*E. coli* C41) and CP127255-CP127260 (*E. coli* Q5).

4.4.5 Antimicrobial Activity of Cell-Free Supernatants of *E. coli* Strains (conducted by other authors)

The *in vitro* antagonistic effect of probiotic *E. coli* was assessed by evaluating the bacterial growth of test-strains (Table 4.5) in the presence of cell-free supernatants of the studied *E. coli* strains in the culture medium. *E. coli* M-17 was obtained from the probiotic “Colibakterin” (MICROGEN NPO JSC, Nizhniy Novgorod, Russia) and used as a control strain. *E. coli* Q5, *E. coli* C41, and *E. coli* M-17 strains were overnight cultured in liquid Luria-Bertani medium (LB medium, “Difco,” Le Pont de Claix, France) at 37 °C without aeration. The grown bacterial cultures were transferred into Eppendorf tubes and centrifuged for 10 min at 13,000 rpm. The supernatants were sterilized using Millex®-GS membrane filters (“Merck Milli-pore Ltd.,” Carrigtwohill, Ireland) with a pore diameter of 0.22 µm. Supernatants were stored at –20 °C. Suspensions of 24 h cultures of the test-strains diluted to a concentration of 10⁶ CFU/mL and cell-free supernatants of probiotic strains were introduced into the wells of the 96-well microtiter plates in a ratio of 1:1 and incubated at 37 °C for 24 h without shaking. Subsequently, the optical density OD₆₀₀ of cultures was measured using the plate reader INFINITE M1000 (Tecan Austria GmbH, Grödig, Austria), and the percentage of growth inhibition after 22 h of co-cultivation was calculated, taking as 100% the optical density of the culture grown in the control wells.

4.4.6 Nonspecific Adhesion of *E. coli* Strains (conducted by other authors)

The study of bacterial nonspecific adhesion was carried out in glass penicillin vials (hydrophilic surface) and in polystyrene 96-well plates (Medpolimer, Saint Petersburg, Russia) (hydrophobic surface), according to Nikolaev Yu.A. [441]. Bacterial cells were deposited at 8000 rpm, washed twice in a phosphate buffer, standardized to 0.150–0.200 OD₅₄₀ units, and 3.0 mL were injected into vials and 200 mL into the wells of the microplate. Vials and plates were placed for 1 h in a thermostat at 37 °C with stirring at 150 rpm. The adhesion index was understood as the number of cells adhering to the walls of the vial/plates, expressed in % of their initial number, and was calculated as follows:

$$\text{Adhesion index} = \left(1 - \frac{OD_{final}}{OD_{initial}}\right) \cdot 100\%,$$

where $OD_{initial}$ and OD_{final} are the optical densities at the initial moment of time and after 1 h, respectively.

4.4.7 Specific Adhesion of *E. coli* Strains (conducted by other authors)

The study of bacterial specific adhesion to red blood cells was carried out according to the Brillis method in Eppendorf tubes [442]. To account for the adhesive properties of bacteria, human red blood cells O (I) of the Rh (+) blood group were used (“Biomed,” a branch of FSUE “Microgen,” Perm, Russia). Erythrocytes contain glycophorin on their surface, which is identical to the glycocalyx of epithelial cells [443]. Erythrocytes were washed in saline phosphate buffer (PBS), then diluted to 10^8 cells/mL. The bacteria were grown overnight, washed with phosphate buffer, and a suspension was prepared at a concentration of 10^8 cells/mL. Then a bacterial suspension was mixed with erythrocyte mass in a ratio of 1:1 and incubated at 37 °C with stirring at 120 rpm for 30 min. Blood smears were prepared and stained with a 0.5% solution of gentian violet [444]. During optical microscopy of the preparations, the following indicators were taken into account: average adhesion index (AAI), which is the average number of microorganisms attached to the surface of a single red blood cell; and adhesion coefficient (AC), the percentage of red blood cells having bacteria on the surface. The index adhesiveness of microorganisms (IAM) was calculated as follows:

$$IAM = \frac{AAI}{AC}$$

Counting was carried out on 100 cells, looking through the entire glass slide. Depending on the IAM values, microorganisms were considered non-adhesive (IAM < 1.75), low-adhesive (IAM = 1.76–2.49), medium-adhesive (IAM = 2.50–3.99), and highly adhesive (IAM > 4.0).

4.4.8 Antimicrobial Susceptibility (conducted by other authors)

The strains were tested by the disk-diffusion method using Muller-Hinton agar (“FBIS SRCAMB”, Obolensk, Russia) and disks (“NICF”, St. Petersburg, Russia) for sensitivity to ampicillin (10 µg), cefoperazone (75 µg), ceftriaxone (30 µg), cefepime (30 µg), meropenem (10 µg), aztreonam (30 µg), amikacin (30 µg), gentamicin (10 µg), ciprofloxacin (5 µg), levofloxacin (5 µg); norfloxacin (10 µg), tetracycline (30 µg), chloramphenicol (30 µg). The determination of the sensitivity of *E. coli*

strains to antibiotics was carried out in accordance with the clinical guidelines “Determination of the sensitivity of microorganisms to antimicrobial drugs” of the Interregional Association for Clinical Microbiology and Antimicrobial Chemotherapy (IACMAC, Version-2018-03).

4.4.9 Bacteriophage Induction (*conducted by other authors*)

Bacterial overnight cultures were diluted in PBS in order to obtain a concentration of 1×10^5 to 1×10^6 bacteria per ml, and 20 mL of such diluted overnight cultures were transferred into standard Petri dishes for exposure to the continuous UV-light treatment (260 nm) for 70 s or 150 s. After the UV exposure, the cultures were incubated for 1 h at 37 °C and then mixed with a culture of the sensitive strain *E. coli* DH5a and added to melted 0.6% agar (46 °C), mixed, and poured onto LB agar plates. After a 24 h incubation at 37 °C the presence of lysis zones in the sensitive strain was screened for.

4.4.10 Probiotic and Pathogenic Inocula Preparation (*conducted by other authors*)

To prepare a probiotic suspension, *E. coli* Q5 and C41 were grown in LB broth for 24 h at 37 °C without aeration. Then the suspensions of microorganisms were centrifuged at 5 000 rpm for 10 min, the supernatant was removed, and the sediment was resuspended in saline. The OD was measured and brought to the final concentration of 5×10^8 or 5×10^{10} CFU/mL. The pathogenic inoculum of *E. coli* C55 at a concentration of 5×10^8 CFU/mL was prepared in a similar way. The suspensions were stored in vials at a temperature of 4 °C and used for administration to rats.

4.4.11 Experimental Design *In vivo* (*conducted by other authors*)

Forty-eight 180 day-old white male rats of the Wistar line were used for *in vivo* experiments. Experiments on rats were conducted following guidelines set by the Ethics Committee. General animal care was carried out in accordance with State Standard No. 33215-2014, “Guidelines for accommodation and care of animals. Environment, housing and management” [445]. The rats were caged in the animal house, where the temperature ranged from 23 °C to 26 °C. The animals received free access to feed (standard pellets) and drinking water (*ad libitum*) during all experiments. Three rats were used to analyze background hematological and biochemical parameters and the composition of the intestinal microbiota before the experiment. The remaining rats were randomly divided into six groups. The design of the *in vivo* experiment is presented in **Table 4.6**.

The first group (control) included intact animals ($n = 10$) that received 1 mL of saline throughout the experiment.

The second group (infection control) included rats ($n = 5$) that received 1 mL of saline for 5 days, then per animal once orally infected with the toxigenic *E. coli* C55 (5×10^8 CFU suspended in 1 mL of saline).

The third ($n = 10$) and fourth ($n = 10$) groups included rats that received *E. coli* Q5 or *E. coli* C41, respectively, orally (5×10^8 CFU suspended in 1 mL of saline), daily for 5 days, with drinking water. Then, after administration of probiotic bacteria, five rats from each group were removed for analysis of hematological and biochemical parameters. The remaining animals were infected with *E. coli* C55 (5×10^8 CFU suspended in 1 mL of saline) orally with water per animal. After 3 days after infection, all rats were euthanized, and the blood and organs of the rats were taken for analysis.

The fifth ($n = 5$) and sixth ($n = 5$) groups included rats that received *E. coli* Q5 or *E. coli* C41, respectively, orally (5×10^8 CFU suspended in 1 mL of saline) daily for 5 days with drinking water. The body weight (BW) of rats in the first, second, third, and fourth groups was measured before the experiment, after administration of probiotic bacteria, and after infection with *E. coli* C55. The BW of animals in the fifth and sixth groups was measured before the experiment and after taking probiotic microorganisms. Throughout the study, the behavior and appearance of animals, water consumption, and food consumption were monitored to determine whether there were any deviations from normal behavior.

Table 4.6. *In vivo* experiment design.

Time Period, Days	Number of Euthanized Rats						Action
	Control (10 in Total)	Infection Control (5 in Total)	<i>E. coli</i> Q5, 5×10^8 CFU (10 in Total)	<i>E. coli</i> C41, 5×10^8 CFU (10 in Total)	<i>E. coli</i> Q5, 5×10^{10} CFU (5 in Total)	<i>E. coli</i> C41, 5×10^{10} CFU (5 in Total)	
0							Body mass measurement
1-4							Administration probiotic bacteria at a dose 5×10^8 or 5×10^{10} CFU
5	-5		-5	-5	-5	-5	Body mass measurement, analysis of microbiota composition, hematological and biochemical parameters, infection with a toxigenic <i>E. coli</i> C55
6-7							Monitoring the condition of rats
8	-5	-5	-5	-5			Body mass measurement, analysis of microbiota composition, hematological and biochemical parameters, organ removal *

Note. «*»—organs for histologic analysis were removed from two randomly selected rats.

4.4.12 Analysis of the Composition of the Intestinal Microbiota (conducted by other authors)

The feces of randomly chosen rats from each group were used as material for bacteriological analysis. The bacteriological analysis of the microbial intestinal community was performed by direct plating (colony-forming unit count, CFU) on selective solid media: Pseudomonas CN Agar (Laboratorios Conda S.A., Madrid, Spain), Endo Agar for *E. coli*, Ploskireva Agar for Proteus, Egg-salt Agar for Staphylococcus, Blaurocca medium for Bifidobacteria, MRS Agar for Lactobacillus, Iron Sulfite Modified Agar №3 for Clostridium, and Sabouraud Agar №2 for Candida (“FBIS SRCAMB”, Obolensk, Russia). After infection, feces were inoculated on blood agar with ampicillin. Ampicillin-resistant colonies with hemolysis representing an experimental *E. coli* C55 infection were counted. The obtained CFU were recalculated to 1 g of the chyme content.

4.4.13 Hematological and Biochemical Blood Analysis (*conducted by other authors*)

Blood samples were taken directly from the heart using a syringe. Analysis of red blood cell count (RBC), hemoglobin concentration ([Hb]), hematocrit (Ht), platelet count (PLT), and white blood cell count (WBC) was performed using the automated Hematological Analyzer (MINDRAY BS-3600, Shenzhen, China). Using RBC, Ht, and [Hb], the average corpuscular volume (MCV), average corpuscular hemoglobin (MCH), and average concentration of corpuscular hemoglobin (MCHC) were calculated according to standard formulas [446].

Blood samples were centrifuged at 1000 RPM for 10 min (Eppendorf 5415R, Germany) and analyzed for the following serum biochemical parameters: glucose, total protein, creatinine, urea, and levels of enzymes phosphatase (ALP) and alanine aminotransferase (ALT) using a Biochemical Analyzer (MINDRAY BS-200, Shenzhen, China).

4.4.14 Histologic Analysis (*conducted by other authors*)

Samples of intestine, Peyer's patches, spleen, and liver from rats were fixed in 10% neutral formalin in phosphate buffer (pH 7.2) and poured into "Histomix" paraffin (BioVitrum, Saint Petersburg, Russia). The paraffin sections were stained with hematoxylin (BioVitrum, Russia) and eosin (BioVitrum, Saint Petersburg, Russia) to evaluate tissue morphology under a light microscope (Olympus, Tokyo, Japan).

4.4.15 Statistical Analysis (*conducted by other authors*)

The data were presented as the arithmetic mean and its mean deviation ($M \pm m$). Statistical analysis was performed using the Student's *t*-test in STATISTICA 10.0. A *p*-value of less than 0.05 was considered significant.

4.5 Conclusions

This study presents a comprehensive assessment of the probiotic characteristics of two bacteriocin-producing strains (*E. coli* Q5 and C41) using *in silico*, *in vitro*, and *in vivo* approaches. The results demonstrate that oral administration of *E. coli* Q5 and C41 to rats did not cause side effects or signs of

clinical disease but contributed to the preservation of intestinal homeostasis and had a preventive effect by protecting against the pathogenic effects of a toxigenic *E. coli* strain. Given that maintaining effective symbiosis between the host organism and the intestinal microbiota is currently considered a necessary component of the veterinary strategy to ensure animal health, our results form the basis for research and development of a probiotic based on the studied strains to be used for the treatment and prevention of infectious diseases in farm animals.

Chapter 5: Conclusions

My doctoral thesis is based on the study of escape mechanisms of MGEs against adaptive immune CRISPR-Cas system in *E. coli* cells and the methodology for studying natural bacterial strains and 16S rRNA metagenomic data. Both these topics are related to the studies of diversity in microbial communities.

The results of Chapter 2 illustrate that plasmids targeted by the CRISPR spacer have a miserable yet actual likelihood to overcome the highly induced CRISPR-Cas system in *E. coli*. The mathematical model outlined the escape mechanism with the molecular dynamics of plasmid replication and CRISPR interference. The outcome of molecular dynamics in a single cell cannot be determined in advance but can emerge at the population level. In the results of microbiological experiments, I demonstrated that MGEs can reshape the phenotypic composition of bacterial populations, even in mono-strain communities. The transformed *E. coli* populations remained stable during the cycles of reseeded on new selective media. Using confocal microscopy and labeling the plasmid by fluorescent gene GFP, I demonstrated that CRISPR ON *E. coli* colonies consist of two subpopulations: plasmid-bearing and plasmid-less cells. Fluorescence-based cytometry further confirmed these results. During the experiments, we also detected that a small fraction of CRISPR ON colonies after transformation were fluorescent (**Fig. 2.3 a**). The results of sequencing fluorescent CRISPR ON colonies revealed inactive CRISPR-Cas systems (**Table S2.1**) but no escape mutations were detected in the g8 protospacer in the plasmids (**Fig S2.1 a**). The possibility that the minor subpopulation of plasmid-bearing cells in CRISPR ON colony represents cells either bearing inactive CRISPR-Cas system or escape mutant plasmids was ruled out with cross-transformation by compatible plasmids to test the activity of CRISPR-Cas system (**Fig. 2.1 d**) and the transformation by the plasmids extracted from these CRISPR ON colonies to test a presence of escape mutations (**Fig. 2.1 c**). Moreover, the several rounds of reseeded on selective plates demonstrated the persistence of plasmids in a small fraction of CRISPR ON cells (**Fig. 2.2**). Alternatively, if the reseeded CRISPR ON colony contains a fraction of cells with an inactive CRISPR-Cas system or mutated plasmids, we might expect that the CFU ratio on the antibiotic-supplemented plate and LB-agar plate would align with each other by the second or third round of reseeded. This alignment would occur due to antibiotic selective pressure, as in this case without CRISPR interference

due to mutations the majority of the cells remained plasmid-bearing. In addition, the fluorescence of CRISPR ON colonies, as the indicative mark of genetic alterations, generally was not observed during the rounds of reseeded under CRISPR interference conditions. Thus, the reseeded experiments demonstrated the plasmid persistence under CRISPR interference conditions without genetic alterations (**Fig. 2.2**). In the subsequent step, I performed the long-term cultivation experiments in antibiotic-free conditions (**Fig. 5.1**). The cultivation in antibiotic-free conditions provides a challenge for selecting plasmid-bearing cells without genetic alterations. If antibiotic pressure offers favorable conditions for selecting CRISPR ON plasmid-bearing cells, then during antibiotic-free cultivation, plasmids affect additional metabolic costs for *E. coli* cells. We demonstrated this fitness cost of plasmids, measuring the growth rate. The growth rate of plasmid-bearing cells is noticeably lower compared to plasmid-free cells (**Fig. S2.3**). Nevertheless, after 72 h of antibiotic-free cultivation, we successfully detected CRISPR ON *E. coli* cells bearing active CRISPR-Cas system and pG8 plasmid with wild-type g8 protospacer (**Fig. S2.3 b, c**), but the fraction of these CRISPR ON cells was significantly low among the total population (**Fig. S2.3 a**). Thus, the results of Chapter 2 validate our hypothesis about the persistence of plasmids with the wild-type protospacer in *E. coli* cells with the active type I-E CRISPR-Cas system and are consistent with the mathematical model.

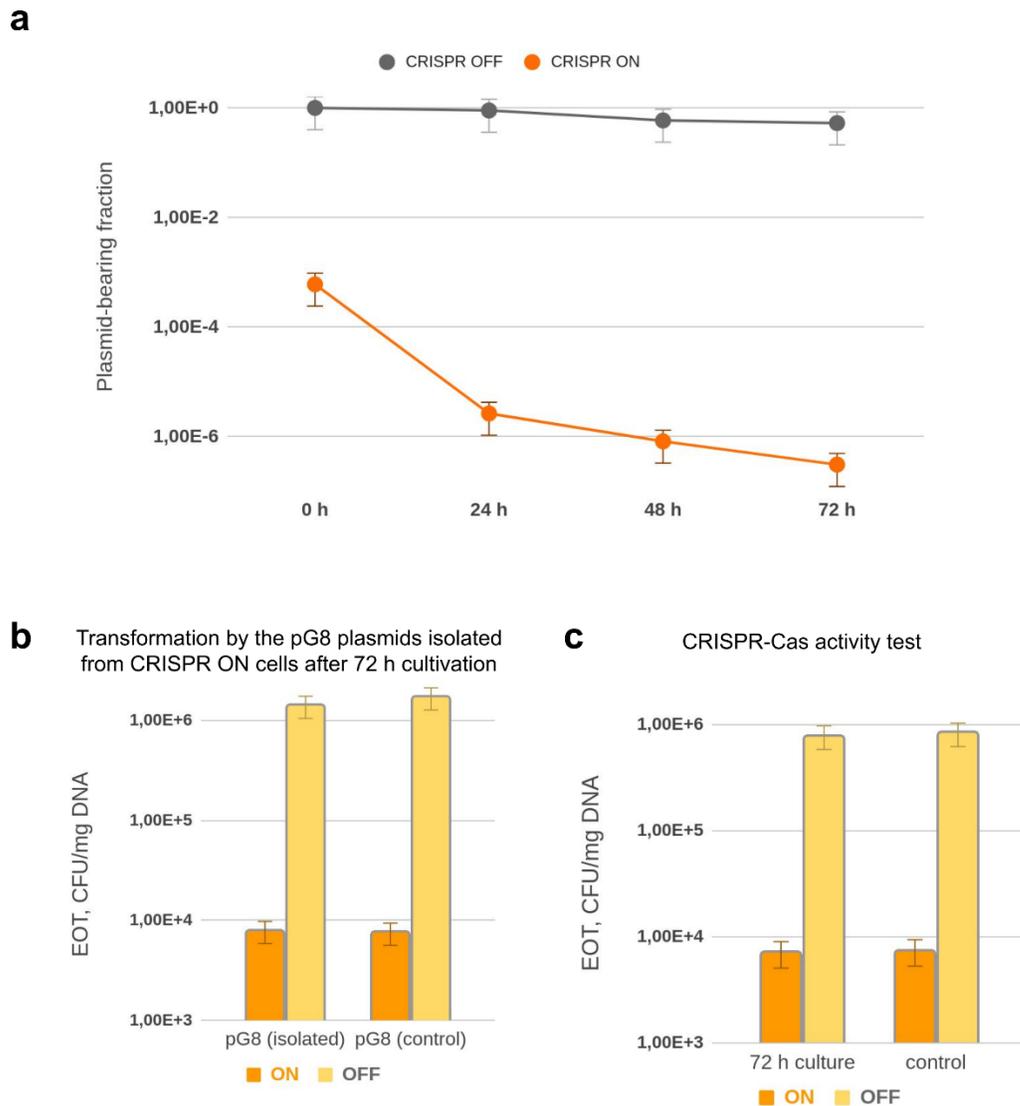


Figure 5.1. Reseeding experiments of CRISPR ON and CRISPR OFF cells bearing the plasmid pG8 in antibiotic-free LB media. **a**, The plot shows plasmid-bearing fractions in CRISPR ON and CRISPR OFF cultures depending on the cultivation time. CRISPR ON cells were cultivated in LB media supplemented with *cas* genes inducers, CRISPR OFF cells were cultivated in LB media only. The cultures were transferred into appropriate fresh LB media every 12 h. **b**, The results of transformation of KD263 competent cells with plasmids extracted from CRISPR ON colonies formed on Ab/Ind plates after 72 h cultivation. The original pG8 plasmid was used as a control. **c**, The results of transformation of competent cells prepared from CRISPR ON colonies used in **b** with the pRSFG8 plasmid, freshly prepared competent KD263 cells were used as a control.

The transformation experiments with the dsDNA replicative form of phage M13 complemented the study of the interaction between plasmids and the CRISPR-Cas system. I demonstrated that escape mutants of phage M13 can emerge during infection of CRISPR ON *E. coli* cells. These findings are backed by at least two experimental results. Firstly, a single CRISPR ON *E. coli* colony after transformation can contain various types of mutant phages M13. The results of sequencing indicated that different types of point mutations in M13 genomes correspond to distinct sequence reads. Since the formation of a colony after transformation likely originates from one cell, and the chance of transforming a single cell with two or more DNA molecules is significantly low, the sequencing results suggest that the escape mutations arise independently. Secondly, the sequencing results of the initial population of phage M13 showed no mutations in the protospacer, or if present, these mutations were undetectable by sequencing. If the initial population of phage M13 had contained escape mutations, then the fraction of such mutants was significantly lower than the demonstrated efficiency of CRISPR ON transformation. Similar to the CRISPR interference assay with plasmids (**Figure 2.1 b**), the efficiency of transformation of CRISPR ON cells by the phage M13 replicative form is at least two orders of magnitude lower compared to the CRISPR OFF control transformation (**Figure 2.7 a**). Thus, the transformation efficiency by the phage M13 replicative form is still detectable, except for the potential fraction of escape mutants in the initial population of the phage M13 replicative form. However, in contrast to the plasmids, in the experiments with phage M13, we revealed CRISPR adaptation (**Figure 2.7 c**) and phage M13 escape mutants in CRISPR ON colonies (**Figure 2.8**) after the transformation. Regarding the plasmids, we can suggest that detecting their mutant forms in CRISPR ON colonies is still challenging, as plasmids propagate vertically from a parent cell to the offspring so there is no fast propagation among the CRISPR ON colony. Therefore, I assume that the transformation predominantly occurred with the wild-type dsDNA replicative form of phage M13, and phage M13 escape mutants appear during rounds of infections and rapidly propagate in CRISPR ON *E. coli* colonies.

To sum up the outcomes of Chapter 2, I suggest the possibility of plasmids without genetic alterations to persist in the bacterial population under permanent CRISPR interference pressure. This imposes additional requirements for CRISPR-based tools for eliminating plasmids, especially for AMR

plasmids. Given the presence of additional anti-defense mechanisms provided by AMR plasmids [209], the task becomes more complex. In contrast to the plasmids, the experiments with phages demonstrated the quick propagation of escape mutants among the bacterial population. While not absolute, an optimized CRISPR-Cas defense affords bacteria a window of opportunities to acquire beneficial MGEs that can provide a competitive edge under adverse conditions.

Chapter 3 demonstrates the importance of selecting appropriate methods for the extraction of metagenomes from various environments such as soil, water and gut microbiomes. Given the complexity of interactions within natural microbial communities, it is critical to accurately determine microbial compositions for a deeper understanding of intra-community interactions and the formulation of biological hypotheses. Thus, the appropriate choice of methodology, described in Chapter 3, for metagenomic studies is pivotal. In this study, my contribution involved the analysis of 16S rRNA data and the development of a computational pipeline to evaluate the accuracy of different kits for the extraction of 16S rRNA. This pipeline included the analysis of relative abundance in control samples to evaluate the taxa bias provided by the kits. By employing the decontam tool, I identified and assessed the extent of contamination in the samples (see Chapter 3, Methods). After removing the contaminating OTUs from samples, alpha diversity (Shannon index) was defined for each sample. The best kits for the extraction of 16S rRNA should demonstrate a low level of contamination and a relatively high Shannon index indicating that there is no bias provided by kit. Using visualization of relative abundance at the order level and ordination of the relative abundance matrix, I revealed that even after removing the contaminating OTUs the samples remained significantly different for particular kits. To objectively rank the kits in their ability to extract the actual 16S rRNA profile of natural microbial community, all analyzed features were normalized and scaled from 0 to 1. We determined that the PowerSoil, PowerFecal, and LSBio kits received the highest ranks respectively and can be considered universal choices for studying the diversity of natural microbial communities.

Chapter 4 highlights the potential of long-read sequencing technology for analyzing bacterial isolates. It also demonstrates how the annotation of biosynthetic clusters can facilitate the experimental validation. In Chapter 4, I contributed to *in silico* analysis by providing long-read genome assemblies

of isolated strains using flye tool, annotations of the assemblies, and searching for biosynthetic clusters (see Chapter 4, Methods). Out of the 10 *E. coli* strains examined, *E. coli* Q5 and C41 strains emerged as particularly noteworthy and were thus chosen for in-depth study. The detailed analysis has demonstrated that *E. coli* Q5 is a potential probiotic candidate. The assemblies of *E. coli* isolates were marked by possessing several different plasmids per cell: 5 plasmids in *E. coli* Q5 and 2 plasmids in *E. coli* C41. Bioinformatic analysis of these plasmids uncovered gene clusters responsible for producing various microcins and colicins, bacteriocins known for their antimicrobial properties. The subsequent microbiological experiments conducted by colleagues validated the antimicrobial activity of these predicted bacteriocins inhibiting the growth of pathogenic strains. Thus, the plasmids provide significant benefits to *E. coli* Q5 and C41 isolates in competing with other strains in the natural microbial community. Our results illustrate the diversity of *E. coli* strains in natural microbial communities and the efficiency of bioinformatic approaches to search for valuable biosynthetic gene clusters based on long-read sequencing technology.

My thesis combines the results of three studies and demonstrates the diversity of interactions between MGEs and bacterial strains. The mechanism of plasmid persistence in the *E. coli* population based on replication dynamics was described. The computational approach for analyzing the methods of 16S rRNA extraction was developed. The potential of long-read sequencing and *in silico* analysis for the study of new bacteria isolates was demonstrated. The methodology described in my thesis is applicable to other microbiology and system biology studies.

Bibliography

1. Cury, J., Oliveira, P. H., de la Cruz, F., & Rocha, E. P. C. (2018). Host range and genetic plasticity explain the coexistence of integrative and extrachromosomal mobile genetic elements. *Molecular Biology and Evolution*, *35*(9), 2230–2239.
2. Rocha, E. P. C., & Bikard, D. (2022). Microbial defenses against mobile genetic elements and viruses: Who defends whom from what? *PLoS Biology*, *20*(1), e3001514.
3. van Dijk, B., Hogeweg, P., Doekes, H. M., & Takeuchi, N. (2020). Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *ELife*, *9*.
4. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences*, *96*(5), 2192–2197.
5. Mushegian, A. R. (2020). Are there 10^{31} virus particles on earth, or more, or fewer? *Journal of Bacteriology*, *202*(9).
6. Batinovic, S., Wassef, F., Knowler, S. A., Rice, D., Stanton, C. R., Rose, J., Tucci, J., Nittami, T., Vinh, A., Drummond, G., Sobey, C., Chan, H., Seviour, R., Petrovski, S., Franks, A.E. (2019). Bacteriophages in natural and artificial environments. *Pathogens*, *8*(3), 100.
7. Keen, E. C. (2015). A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays*, *37*(1), 6–9.
8. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J.*, *11*(7), 1511–1520.
9. Zhang, M., Zhang, T., Yu, M., Chen, Y.-L., & Jin, M. (2022). The life cycle transitions of temperate phages: regulating factors and potential ecological implications. *Viruses*, *14*(9), 1904.
10. Young, R., Wang, I.-N., & Roof, W. D. (2000). Phages will out: strategies of host cell lysis. *Trends in Microbiology*, *8*(3), 120–128.
11. Davies, E. v., James, C. E., Williams, D., O'Brien, S., Fothergill, J. L., Haldenby, S., Paterson, S., Winstanley, C., & Brockhurst, M. A. (2016). Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences*, *113*(29), 8266–8271.
12. Harrison, E., & Brockhurst, M. A. (2017). Ecological and evolutionary benefits of temperate phage: what does or doesn't kill you makes you stronger. *BioEssays*, *39*(12).
13. Young, R., Wang, I.-N., & Roof, W. D. (2000). Phages will out: strategies of host cell lysis. *Trends in Microbiology*, *8*(3), 120–128.
14. Makarova, K. S., Wolf, Y. I., Snir, S., & Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of Bacteriology*, *193*(21), 6039–6056.
15. Doron, S., Melamed, S., Ofir, G., Leavitt, A., Lopatina, A., Keren, M., Amitai, G., & Sorek, R. (2018). Systematic discovery of antiphage defense systems in the microbial pangenome. *Science*, *359*(6379).
16. Hochhauser, D., Millman, A., & Sorek, R. (2023). The defense island repertoire of the *Escherichia coli* pan-genome. *PLoS Genetics*, *19*(4), e1010694.
17. Bertozzi Silva, J., Storms, Z., & Sauvageau, D. (2016). Host receptors for bacteriophage adsorption. *FEMS Microbiology Letters*, *363*(4), fnw002.
18. Dowah, A. S. A., & Clokie, M. R. J. (2018). Review of the nature, diversity and structure of bacteriophage receptor binding proteins that target Gram-positive bacteria. *Biophysical Reviews*, *10*(2), 535–542.
19. Hao, G., Chen, A. I., Liu, M., Zhou, H., Egan, M., Yang, X., Kan, B., Wang, H., Goulian, M., & Zhu, J. (2019). Colistin Resistance-Mediated Bacterial Surface Modification Sensitizes Phage Infection. *Antimicrobial Agents and Chemotherapy*, *63*(12).
20. Meyer, J. R., Dobias, D. T., Weitz, J. S., Barrick, J. E., Quick, R. T., & Lenski, R. E. (2012). Repeatability and Contingency in the Evolution of a Key Innovation in Phage Lambda. *Science*, *335*(6067), 428–432.
21. Luria, S. E., & Human, M. L. (1952). A nonhereditary, host-induced variation of bacterial viruses. *Journal of Bacteriology*, *64*(4), 557–569.
22. Luria, S. E. (1953). Host-induced modifications of viruses. *Cold Spring Harbor Symposia on Quantitative Biology*, *18*(0), 237–244.

23. Kobayashi, I. (2001). Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Research*, 29(18), 3742–3756.
24. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G., & Murray, N. E. (2014). Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research*, 42(1), 3–19.
25. Sumbly, P., & Smith, M. C. M. (2002). Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 44(2), 489–500.
26. Sumbly, P., & Smith, M. C. M. (2003). Phase variation in the phage growth limitation system of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 185(15), 4558–4563.
27. Laity, C., Chater, K. F., Lewis, C. G., & Buttner, M. J. (1993). Genetic analysis of the ϕ C31 - specific phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 7(2), 329–336.
28. Bedford, D. J., Laity, C., & Buttner, M. J. (1995). Two genes involved in the phase-variable phi C31 resistance mechanism of *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*, 177(16), 4681–4689.
29. Isaev, A., Drobiazko, A., Sierro, N., Gordeeva, J., Yosef, I., Qimron, U., Ivanov, N., & Severinov, K. (2020). Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Research*, 48(10), 5397–5406.
30. Pu, T., Mei, Z., Zhang, W., Liang, W., Zhou, X., Liang, J., Deng, Z., & Wang, Z. (2020). An *in vitro* DNA phosphorothioate modification reaction. *Molecular Microbiology*, 113(2), 452–463.
31. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak-Amikam, Y., Afik, S., Ofir, G., & Sorek, R. (2015). BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO Journal*, 34(2), 169–183.
32. Gordeeva, J., Morozova, N., Sierro, N., Isaev, A., Sinkunas, T., Tsvetkova, K., Matlashov, M., Truncaitè, L., Morgan, R. D., Ivanov, N. v, Siksnys, V., Zeng, L., & Severinov, K. (2019). BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Research*, 47(1), 253–265.
33. Chopin, M.-C., Chopin, A., & Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Current Opinion in Microbiology*, 8(4), 473–479.
34. Dy, R. L., Richter, C., Salmond, G. P. C., & Fineran, P. C. (2014). Remarkable mechanisms in microbes to resist phage infections. *Annual Review of Virology*, 1(1), 307–331.
35. Lopatina, A., Tal, N., & Sorek, R. (2020). Abortive infection: bacterial suicide as an antiviral immune strategy. *Annual Review of Virology*, 7(1), 371–384.
36. Rostøl, J. T., & Marraffini, L. (2019). (Ph)ighting phages: how bacteria resist their parasites. *Cell Host & Microbe*, 25(2), 184–194.
37. Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G., & Sorek, R. (2019). Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, 574(7780), 691–695.
38. Snyder, L., & McWilliams, K. (1989). The rex genes of bacteriophage lambda can inhibit cell function without phage superinfection. *Gene*, 81(1), 17–24.
39. Toothman, P., & Herskowitz, I. (1980). Rex-dependent exclusion of lambdoid phages II. Determinants of sensitivity to exclusion. *Virology*, 102(1), 147–160.
40. Chopin, M.-C., Chopin, A., & Bidnenko, E. (2005). Phage abortive infection in lactococci: variations on a theme. *Current Opinion in Microbiology*, 8(4), 473–479.
41. Domingues, S., Chopin, A., Ehrlich, S. D., & Chopin, M. C. (2004). The Lactococcal abortive phage infection system AbiP prevents both phage DNA replication and temporal transcription switch. *Journal of Bacteriology*, 186(3), 713–721.
42. Parreira, R., Ehrlich, S. D., & Chopin, M.-C. (1996). Dramatic decay of phage transcripts in lactococcal cells carrying the abortive infection determinant AbiB. *Molecular Microbiology*, 19(2), 221–230.
43. Samson, J. E., Spinelli, S., Cambillau, C., & Moineau, S. (2013). Structure and activity of AbiQ, a lactococcal endoribonuclease belonging to the type III toxin-antitoxin system. *Molecular Microbiology*, 87(4), 756–768.
44. Durmaz, E., & Klaenhammer, T. R. (2007). Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. *Journal of Bacteriology*, 189(4), 1417–1425.
45. Parma, D. H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E., & Gold, L. (1992). The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes & Development*, 6(3), 497–510.

46. Schmitt, C. K., Kemp, P., & Molineux, I. J. (1991). Genes 1.2 and 10 of bacteriophages T3 and T7 determine the permeability lesions observed in infected cells of *Escherichia coli* expressing the F plasmid gene pifA. *Journal of Bacteriology*, *173*(20), 6507–6514.
47. Ngiam, L., Weynberg, K. D., & Guo, J. (2022). The presence of plasmids in bacterial hosts alters phage isolation and infectivity. *ISME Communications*, *2*(1), 75.
48. Yamaguchi, Y., Park, J.-H., & Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annual Review of Genetics*, *45*(1), 61–79.
49. Page, R., & Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology*, *12*(4), 208–214.
50. Qi, Q., Kamruzzaman, M., & Iredell, J. R. (2021). The *higBA*-Type toxin-antitoxin system in *incC* plasmids is a mobilizable ciprofloxacin-inducible system. *MSphere*, *6*(3).
51. Gerdes, K., Helin, K., Christensen, O. W., & Løbner-Olesen, A. (1988). Translational control and differential RNA decay are key elements regulating postsegregational expression of the killer protein encoded by the *parB* locus of plasmid R1. *Journal of Molecular Biology*, *203*(1), 119–129.
52. Melderer, L., Bernard, P., & Couturier, M. (1994). Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. *Molecular Microbiology*, *11*(6), 1151–1157.
53. Tsuchimoto, S., Nishimura, Y., & Ohtsubo, E. (1992). The stable maintenance system *pem* of plasmid R100: degradation of PemI protein may allow PemK protein to inhibit cell growth. *Journal of Bacteriology*, *174*(13), 4205–4211.
54. Choi, J. S., Kim, W., Suk, S., Park, H., Bak, G., Yoon, J., & Lee, Y. (2018). The small RNA, SdsR, acts as a novel type of toxin in *Escherichia coli*. *RNA Biology*, *15*(10), 1319–1335.
55. Gerdes, K., Bech, F. W., Jørgensen, S. T., Løbner-Olesen, A., Rasmussen, P. B., Atlung, T., Boe, L., Karlstrom, O., Molin, S., & von Meyenburg, K. (1986). Mechanism of postsegregational killing by the *hok* gene product of the *parB* system of plasmid R1 and its homology with the *relF* gene product of the *E. coli* *relB* operon. *The EMBO Journal*, *5*(8), 2023–2029.
56. Ogura, T., & Hiraga, S. (1983). Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences*, *80*(15), 4784–4788.
57. Bahassi, E. M., O’Dea, M. H., Allali, N., Messens, J., Gellert, M., & Couturier, M. (1999). Interactions of CcdB with DNA Gyrase. *Journal of Biological Chemistry*, *274*(16), 10936–10944.
58. Fineran, P. C., Blower, T. R., Foulds, I. J., Humphreys, D. P., Lilley, K. S., & Salmond, G. P. C. (2009). The phage abortive infection system, ToxIN, functions as a protein–RNA toxin–antitoxin pair. *Proceedings of the National Academy of Sciences*, *106*(3), 894–899.
59. Masuda, H., Tan, Q., Awano, N., Wu, K.-P., & Inouye, M. (2012). YeeU enhances the bundling of cytoskeletal polymers of MreB and FtsZ, antagonizing the CbtA (YeeV) toxicity in *Escherichia coli*. *Molecular Microbiology*, *84*(5), 979–989.
60. Wang, X., Lord, D. M., Cheng, H.-Y., Osbourne, D. O., Hong, S. H., Sanchez-Torres, V., Quiroga, C., Zheng, K., Herrmann, T., Peti, W., Benedik, M. J., Page, R., & Wood, T. K. (2012). A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nature Chemical Biology*, *8*(10), 855–861.
61. Aakre, C. D., Phung, T. N., Huang, D., & Laub, M. T. (2013). A bacterial toxin inhibits DNA replication elongation through a direct interaction with the β sliding clamp. *Molecular Cell*, *52*(5), 617–628.
62. Marimon, O., Teixeira, J. M. C., Cordeiro, T. N., Soo, V. W. C., Wood, T. L., Mayzel, M., Amata, I., Garcia, J., Morera, A., Gay, M., Vilaseca, M., Orekhov, V. Y., Wood, T. K., & Pons, M. (2016). An oxygen-sensitive toxin–antitoxin system. *Nature Communications*, *7*(1), 13634.
63. Hazan, R., & Engelberg-Kulka, H. (2004). *Escherichia coli* mazEF-mediated cell death as a defense mechanism that inhibits the spread of phage P1. *Molecular Genetics and Genomics*, *272*(2), 227–234.
64. Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C., & Fineran, P. C. (2014). A widespread bacteriophage abortive infection system functions through a Type IV toxin–antitoxin mechanism. *Nucleic Acids Research*, *42*(7), 4590–4605.
65. Helinski, D. R. (2022). A Brief History of Plasmids. *EcoSal Plus*, *10*(1).
66. Grohmann, E., Muth, G., & Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiology and Molecular Biology Reviews*, *67*(2), 277–301.
67. Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *Journal of Experimental Medicine*, *79*(2), 137–158.

68. Muschiol, S., Balaban, M., Normark, S., & Henriques-Normark, B. (2015). Uptake of extracellular DNA: Competence induced pili in natural transformation of *Streptococcus pneumoniae*. *BioEssays*, 37(4), 426–435.
69. Dionisio, F., Conceição, I. C., Marques, A. C. R., Fernandes, L., & Gordo, I. (2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biology Letters*, 1(2), 250–252.
70. Coluzzi, C., Garcillán-Barcia, M. P., de la Cruz, F., & Rocha, E. P. C. (2022). Evolution of plasmid mobility: origin and fate of conjugative and nonconjugative plasmids. *Molecular Biology and Evolution*, 39(6).
71. Sheahan, M. L., Coyne, M. J., Flores, K., Garcia-Bayona, L., Chatzidaki-Livanis, M., Sundararajan, A., Holst, A. Q., Barquera, B., & Comstock, L. E. (2023). A ubiquitous mobile genetic element disarms a bacterial antagonist of the gut microbiota. *BioRxiv*, 2023.08.25.553775.
72. Rozwandowicz, M., Brouwer, M. S. M., Fischer, J., Wagenaar, J. A., Gonzalez-Zorn, B., Guerra, B., Mevius, D. J., & Hordijk, J. (2018). Plasmids carrying antimicrobial resistance genes in *Enterobacteriaceae*. *Journal of Antimicrobial Chemotherapy*, 73(5), 1121–1137.
73. Sher, A.A.; VanAllen, M.E.; Ahmed, H.; Whitehead-Tillery, C.; Rafique, S.; Bell, J.A.; Zhang, L.; Mansfield, L.S. (2023). Conjugative RP4 plasmid-mediated transfer of antibiotic resistance genes to commensal and multidrug-resistant enteric bacteria in vitro. *Microorganisms*, 11, 193.
74. McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6), 344–355.
75. Ravindran, S. (2012). Barbara McClintock and the discovery of jumping genes. *Proceedings of the National Academy of Sciences*, 109(50), 20198–20199.
76. Fedoroff, N., Wessler, S., & Shure, M. (1983). Isolation of the transposable maize controlling elements Ac and Ds. *Cell*, 35(1), 235–242.
77. Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8(12), 973–982.
78. Sota, M., Yano, H., Nagata, Y., Ohtsubo, Y., Genka, H., Anbutsu, H., Kawasaki, H., & Tsuda, M. (2006). Functional analysis of unique class II insertion sequence IS1071. *Applied and Environmental Microbiology*, 72(1), 291–297.
79. Torres, O. R., Korman, R. Z., Zahler, S. A., & Dunny, G. M. (1991). The conjugative transposon Tn925: enhancement of conjugal transfer by tetracycline in *Enterococcus faecalis* and mobilization of chromosomal genes in *Bacillus subtilis* and *E. faecalis*. *Molecular and General Genetics MGG*, 225(3), 395–400.
80. Rice, L. B. (1998). Tn916 family conjugative transposons and dissemination of antimicrobial resistance determinants. *Antimicrobial Agents and Chemotherapy*, 42(8), 1871–1877.
81. Pyatkov, K. I., Arkhipova, I. R., Malkova, N. v., Finnegan, D. J., & Evgen'ev, M. B. (2004). Reverse transcriptase and endonuclease activities encoded by *Penelope*-like retroelements. *Proceedings of the National Academy of Sciences*, 101(41), 12437–12442.
82. Hughes, S. H. (2015). Reverse transcription of retroviruses and LTR retrotransposons. *Microbiology Spectrum*, 3(2).
83. Wicker, T., Schulman, A. H., Tanskanen, J., Spannagl, M., Twardziok, S., Mascher, M., Springer, N. M., Li, Q., Waugh, R., Li, C., Zhang, G., Stein, N., Mayer, K. F. X., & Gundlach, H. (2017). The repetitive landscape of the 5100 Mbp barley genome. *Mobile DNA*, 8(1), 22.
84. Wu, L., Gingery, M., Abebe, M., Arambula, D., Czornyj, E., Handa, S., Khan, H., Liu, M., Pohlschroder, M., Shaw, K. L., Du, A., Guo, H., Ghosh, P., Miller, J. F., & Zimmerly, S. (2018). Diversity-generating retroelements: natural variation, classification and evolution inferred from a large-scale genomic survey. *Nucleic Acids Research*, 46(1), 11–24.
85. Roux, S., Paul, B. G., Bagby, S. C., Nayfach, S., Allen, M. A., Attwood, G., Cavicchioli, R., Chistoserdova, L., Gruninger, R. J., Hallam, S. J., Hernandez, M. E., Hess, M., Liu, W.-T., McAllister, T. A., O'Malley, M. A., Peng, X., Rich, V. I., Saleska, S. R., & Elie-Fadrosh, E. A. (2021). Ecology and molecular targets of hypermutation in the global microbiome. *Nature Communications*, 12(1), 3076.
86. Dai, W., Hodes, A., Hui, W. H., Gingery, M., Miller, J. F., & Zhou, Z. H. (2010). Three-dimensional structure of tropism-switching *Bordetella* bacteriophage. *Proceedings of the National Academy of Sciences*, 107(9), 4347–4352.
87. Carattoli, A., Filetici, E., Villa, L., Dionisi, A. M., Ricci, A., & Luzzi, I. (2002). Antibiotic resistance genes and *Salmonella* genomic island 1 in *Salmonella enterica* serovar typhimurium isolated in Italy. *Antimicrobial Agents and Chemotherapy*, 46(9), 2821–2828.

88. Vázquez, X., García-Fierro, R., Fernández, J., Bances, M., Herrero-Fresno, A., Olsen, J. E., Rodicio, R., Ladero, V., García, V., & Rodicio, M. R. (2023). Incidence and genomic background of antibiotic resistance in food-borne and clinical isolates of *Salmonella enterica* serovar derby from Spain. *Antibiotics*, *12*(7), 1204.
89. Hickman, A. B., & Dyda, F. (2015). Mechanisms of DNA transposition. *Microbiology Spectrum*, *3*(2).
90. Siguier, P., Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews*, *38*(5), 865–891.
91. Consuegra, J., Gaffé, J., Lenski, R. E., Hindré, T., Barrick, J. E., Tenaillon, O., & Schneider, D. (2021). Insertion-sequence-mediated mutations both promote and constrain evolvability during a long-term experiment with bacteria. *Nature Communications*, *12*(1), 980.
92. Mormann, S., Lömker, A., Rückert, C., Gaigalat, L., Tauch, A., Pühler, A., & Kalinowski, J. (2006). Random mutagenesis in *Corynebacterium glutamicum* ATCC 13032 using an IS6100-based transposon vector identified the last unknown gene in the histidine biosynthesis pathway. *BMC Genomics*, *7*(1), 205.
93. Ruiz-Martínez, L., López-Jiménez, L., d'Ostuni, V., Fusté, E., Vinuesa, T., & Vinas, M. (2011). A mechanism of carbapenem resistance due to a new insertion element (ISPa133) in *Pseudomonas aeruginosa*. *International Microbiology: The Official Journal of the Spanish Society for Microbiology*, *14*, 51–58.
94. Inui, M., Tsuge, Y., Suzuki, N., Vertès, A. A., & Yukawa, H. (2005). Isolation and characterization of a native composite transposon, Tn14751, carrying 17.4 kilobases of *Corynebacterium glutamicum* chromosomal DNA. *Applied and Environmental Microbiology*, *71*(1), 407–416.
95. Kim, E., & Aoki, T. (1994). The transposon-like structure of IS26-tetracycline, and kanamycin resistance determinant derived from transferable R plasmid of fish pathogen, *Pasteurella piscicida*. *Microbiology and Immunology*, *38*(1), 31–38.
96. Zhou, K., Yu, W., Shen, P., Lu, H., Wang, B., Rossen, J. W. A., & Xiao, Y. (2017). A novel Tn1696-like composite transposon (Tn6404) harboring bla IMP-4 in a *Klebsiella pneumoniae* isolate carrying a rare ESBL gene bla SFO-1. *Scientific Reports*, *7*(1), 17321.
97. Sun, Y., Liu, Y., Wu, H., Wang, L., Liu, J., Yuan, L., Pan, Y., He, D., & Hu, G. (2019). IS26-flanked composite transposon Tn6539 carrying the tet(M) Gene in IncHI2-type conjugative plasmids from *Escherichia coli* isolated from ducks in China. *Frontiers in Microbiology*, *9*.
98. Iyer, A., Barbour, E., Azhar, E., Salabi, A. A. el, Hassan, H. M. A., Qadri, I., Chaudhary, A., Abuzenadah, A., Kumosani, T., Damanhoury, G., Alawi, M., Na'was, T., Nour, A. M. A., & Harakeh, S. (2013). Transposable elements in *Escherichia coli* antimicrobial resistance. *Advances in Bioscience and Biotechnology*, *04*(03), 415–423.
99. Nicolas, E., Lambin, M., Dandoy, D., Galloy, C., Nguyen, N., Oger, C. A., & Hallet, B. (2015). The Tn3-family of Replicative Transposons. *Microbiology Spectrum*, *3*(4).
100. Nordmann, P., Dortet, L., & Poirel, L. (2012). Carbapenem resistance in *Enterobacteriaceae*: here is the storm! *Trends in Molecular Medicine*, *18*(5), 263–272.
101. Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). In Chapter 25 - Mobile DNA, *Molecular Biology (Third Edition)* (pp. 793–829). *Academic Cell*.
102. Chaconas, G., & Harshey, R. M. (2007). Transposition of phage Mu DNA. In *Mobile DNA II* (pp. 384–402). Wiley.
103. Shapiro, J. A. (1979). Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. *Proceedings of the National Academy of Sciences*, *76*(4), 1933–1937.
104. Vergnaud, G., Midoux, C., Blouin, Y., Bourkaltseva, M., Krylov, V., & Pourcel, C. (2018). Transposition Behavior Revealed by High-Resolution Description of *Pseudomonas aeruginosa* Saltovirus Integration Sites. *Viruses*, *10*(5), 245.
105. Harshey, R. M. (2014). Transposable Phage Mu. *Microbiology Spectrum*, *2*(5).
106. Maxwell, A., Craigie, R., & Mizuuchi, K. (1987). B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proceedings of the National Academy of Sciences*, *84*(3), 699–703.
107. Han, Y.-W., & Mizuuchi, K. (2010). Phage Mu transposition immunity: protein pattern formation along DNA by a diffusion-ratchet mechanism. *Molecular Cell*, *39*(1), 48–58.
108. Braid, M. D., Silhavy, J. L., Kitts, C. L., Cano, R. J., & Howe, M. M. (2004). Complete Genomic Sequence of Bacteriophage B3, a Mu-Like Phage of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, *186*(19), 6560–6574.
109. Fogg, P. C. M., Hynes, A. P., Digby, E., Lang, A. S., & Beatty, J. T. (2011). Characterization of a newly discovered Mu-like bacteriophage, RcapMu, in *Rhodobacter capsulatus* strain SB1003. *Virology*, *421*(2), 211–221.

110. Zehr, E. S., Tabatabai, L. B., & Bayles, D. O. (2012). Genomic and proteomic characterization of SuMu, a Mu-like bacteriophage infecting *Haemophilus parasuis*. *BMC Genomics*, *13*(1), 331.
111. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, *315*(5819), 1709–1712.
112. Marraffini, L. A., & Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in *Staphylococci* by targeting DNA. *Science*, *322*(5909), 1843–1845.
113. Makarova, K., Wolf, Y., Alkhnbashi, O. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat Rev Microbiol*, *13*, 722–736.
114. Makarova, K.S., Wolf, Y.I., Iranzo, J. *et al.* (2020) Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*, *18*, 67–83.
115. Jansen, Ruud., Embden, Jan. D. A. van, Gaastra, Wim., & Schouls, Leo. M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, *43*(6), 1565–1575.
116. Jackson, S. A., McKenzie, R. E., Fagerlund, R. D., Kieper, S. N., Fineran, P. C., & Brouns, S. J. J. (2017). CRISPR-Cas: Adapting to change. *Science*, *356*(6333).
117. Koonin, E. V., & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772), 20180087.
118. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433.
119. Mojica, F. J. M., Ferrer, C., Juez, G., & Rodríguez-Valera, F. (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea *Haloferax mediterranei* and *Haloferax volcanii* and could be involved in replicon partitioning. *Molecular Microbiology*, *17*(1), 85–93.
120. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174–182.
121. Horvath, P., Romero, D. A., Coûté-Monvoisin, A.-C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., & Barrangou, R. (2008). Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *Journal of Bacteriology*, *190*(4), 1401–1412.
122. Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, *190*(4), 1390–1400.
123. Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* *155*(3):733-740.
124. Semenova, E., Jore, M. M., Datsenko, K. A., Semenova, A., Westra, E. R., Wanner, B., van der Oost, J., Brouns, S. J. J., & Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, *108*(25), 10098–10103.
125. Datsenko, K., Pougach, K., Tikhonov, A. *et al.* (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications* *3*, 945.
126. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821.
127. Jiang, W., Bikard, D., Cox, D., Zhang, F., & Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, *31*(3), 233–239.
128. Musharova, O., Sitnik, V., Vlot, M., Savitskaya, E., Datsenko, K. A., Krivoy, A., Fedorov, I., Semenova, E., Brouns, S. J. J., & Severinov, K. (2019). Systematic analysis of type I-E *Escherichia coli* CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Molecular Microbiology*, *111*(6), 1558–1570.
129. Yosef, I., Goren, M. G., & Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Research*, *40*(12), 5569–5576.
130. Mulepati, S., & Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated Protein 3 (Cas3). *Journal of Biological Chemistry*, *286*(36), 31896–31903.

131. Jackson, R. N., Lavin, M., Carter, J., & Wiedenheft, B. (2014). Fitting CRISPR-associated Cas3 into the Helicase Family Tree. *Current Opinion in Structural Biology*, 24, 106–114.
132. Huo, Y., Nam, K. H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M. D., Zhou, S., Rajashankar, K., Kurinov, I., Zhang, R., & Ke, A. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature Structural & Molecular Biology*, 21(9), 771–777.
133. Nimkar, S., & Anand, B. (2020). Cas3/I-C mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. *Nucleic Acids Research*, 48(5), 2486–2501.
134. Mulepati, S., & Bailey, S. (2011). Structural and biochemical analysis of nuclease domain of clustered regularly interspaced short palindromic repeat (CRISPR)-associated protein 3 (Cas3). *Journal of Biological Chemistry*, 286(36), 31896–31903.
135. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K. A., Djordjevic, M., Wanner, B. L., & Severinov, K. (2010). Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Molecular Microbiology*, 77(6), 1367–1379.
136. Touchon, M., Charpentier, S., Clermont, O., Rocha, E. P. C., Denamur, E., & Branger, C. (2011). CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. *Journal of Bacteriology*, 193(10), 2460–2467.
137. Zhou, Y., Tang, Y., Fu, P., Tian, D., Yu, L., Huang, Y., Li, G., Li, M., Wang, Y., Yang, Z., Xu, X., Yin, Z., Zhou, D., Poirel, L., & Jiang, X. (2020). The type I-E CRISPR-Cas system influences the acquisition of *bla*_{KPC}-IncF plasmid in *Klebsiella pneumoniae*. *Emerging Microbes & Infections*, 9(1), 1011–1022.
138. Swarts, D. C., Mosterd, C., van Passel, M. W. J. & Brouns, S. J. J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS One* 7, e35888.
139. Musharova, O., Medvedeva, S., Klimuk, E., Guzman, N. M., Titova, D., Zgoda, V., Shiriaeva, A., Semenova, E., Severinov, K., & Savitskaya, E. (2021). Pre-spacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase. *Proceedings of the National Academy of Sciences*, 118(22).
140. Lin, J., Feng, M., Zhang, H., & She, Q. (2020). Characterization of a novel type III CRISPR-Cas effector provides new insights into the allosteric activation and suppression of the Cas10 DNase. *Cell Discovery*, 6(1), 29.
141. Goldberg, G., Jiang, W., Bikard, D. *et al.* (2014) Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature* 514, 633–637.
142. Sofos, N., Feng, M., Stella, S., Pape, T., Fuglsang, A., Lin, J., Huang, Q., Li, Y., She, Q., & Montoya, G. (2020). Structures of the Cmr-β complex reveal the regulation of the immunity mechanism of type III-B CRISPR-Cas. *Molecular Cell*, 79(5), 741-757.e7.
143. Samai, P., Pyenson, N., Jiang, W., Goldberg, G. W., Hatoum-Aslan, A., & Marraffini, L. A. (2015). Co-transcriptional DNA and RNA cleavage during type III CRISPR-Cas immunity. *Cell*, 161(5), 1164–1174.
144. Deng, L., Garrett, R. A., Shah, S. A., Peng, X., & She, Q. (2013). A novel interference mechanism by a type IIIB CRISPR-Cmr module in *Sulfolobus*. *Molecular Microbiology*, 87(5), 1088–1099.
145. Niewoehner, O., Garcia-Doval, C., Rostøl, J. T., Berk, C., Schwede, F., Bigler, L., Hall, J., Marraffini, L. A., & Jinek, M. (2017). Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, 548(7669), 543–548.
146. Kazlauskienė, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G., & Siksnys, V. (2017). A cyclic oligonucleotide signaling pathway in type III CRISPR–Cas systems. *Science*, 357(6351), 605–609.
147. Garcia-Doval, C., Schwede, F., Berk, C., Rostøl, J. T., Niewoehner, O., Tejero, O., Hall, J., Marraffini, L. A., & Jinek, M. (2020). Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature Communications*, 11(1), 1596.
148. Shmakov, S. A., Makarova, K. S., Wolf, Y. I., Severinov, K., & Koonin, E. V. (2018). Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proceedings of the National Academy of Sciences*, 115(23).
149. Garcia-Doval, C., Schwede, F., Berk, C., Rostøl, J. T., Niewoehner, O., Tejero, O., Hall, J., Marraffini, L. A., & Jinek, M. (2020). Activation and self-inactivation mechanisms of the cyclic oligoadenylate-dependent CRISPR ribonuclease Csm6. *Nature Communications*, 11(1), 1596.

150. Domgaard, H., Cahoon, C., Armbrust, M., Redman, O., Jolley, A., Thomas, A., Jackson, R. (2023). CasDinG is a 5'-3' dsDNA and RNA/DNA helicase with three accessory domains essential for type IV CRISPR immunity. *Nucleic Acids Research*, 51, 15, 8115–8132.
151. Özcan, A., Pausch, P., Linden, A. *et al.* (2019). Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*. *Nat Microbiol* 4, 89–96.
152. Pinilla-Redondo, R., Mayo-Muñoz, D., Russel, J., Garrett, R. A., Randau, L., Sørensen, S. J., & Shah, S. A. (2020). Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Research*, 48(4), 2000–2012.
153. Kamruzzaman, M., & Iredell, J. R. (2020). CRISPR-Cas system in antibiotic resistance plasmids in *Klebsiella pneumoniae*. *Frontiers in Microbiology*, 10:2934.
154. Crowley, V. M., Catching, A., Taylor, H. N., Borges, A. L., Metcalf, J., Bondy-Denomy, J., & Jackson, R. N. (2019). A type IV-A CRISPR-Cas system in *Pseudomonas aeruginosa* mediates RNA-guided plasmid interference *in vivo*. *The CRISPR Journal*, 2(6), 434–440.
155. Deltcheva, E., Chylinski, K., Sharma, C. *et al.* (2011). CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* 471, 602–607.
156. Collias, D., Beisel, C.L. (2021). CRISPR technologies and the search for the PAM-free nuclease. *Nat Commun* 12, 555.
157. Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., & Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 468(7320), 67–71.
158. Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature Communications*, 9(1), 1911.
159. Ledford, H., & Callaway, E. (2020). Pioneers of revolutionary CRISPR gene editing win chemistry Nobel. *Nature*, 586(7829), 346–347.
160. Wang, SW., Gao, C., Zheng, YM. *et al.* (2022). Current applications and future perspective of CRISPR/Cas9 gene editing in cancer. *Mol Cancer* 21, 57.
161. Nelles, D. A., Fang, M. Y., O'Connell, M. R., Xu, J. L., Markmiller, S. J., Doudna, J. A., & Yeo, G. W. (2016). Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell*, 165(2), 488–496.
162. Spanjaard, B., Hu, B., Mitic, N. *et al.* (2018). Simultaneous lineage tracing and cell-type identification using CRISPR–Cas9-induced genetic scars. *Nat Biotechnol* 36, 469–473.
163. Fernandes, L.G.V., Guaman, L.P., Vasconcellos, S.A. *et al.* (2019). Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in *Leptospira*. *Sci Rep* 9, 1839.
164. Riedmayr, L.M., Hinrichsmeyer, K.S., Karguth, N. *et al.* (2022). dCas9-VPR-mediated transcriptional activation of functionally equivalent genes for gene therapy. *Nat Protoc* 17, 781–818.
165. Shmakov, S., Smargon, A., Scott, D. *et al.* (2017). Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology* 15, 169–182.
166. Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K. S., Koonin, E. V., & Zhang, F. (2019). Engineering of CRISPR-Cas12b for human genome editing. *Nature Communications*, 10(1), 212.
167. Gao, P., Yang, H., Rajashankar, K. *et al.* (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res* 26, 901–913.
168. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I. M., Li, Y., Fedorova, I., Nakane, T., Makarova, K. S., Koonin, E. V., Ishitani, R., Zhang, F., & Nureki, O. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell*, 165(4), 949–962.
169. Harrington, L. B., Ma, E., Chen, J. S., Witte, I. P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N. C., Burstein, D., Banfield, J. F., & Doudna, J. A. (2020). A scoutRNA Is Required for Some Type V CRISPR-Cas Systems. *Molecular Cell*, 79(3), 416-424.e5.
170. Takeda, S. N., Nakagawa, R., Okazaki, S., Hirano, H., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Nishimasu, H., & Nureki, O. (2021). Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Molecular Cell*, 81(3), 558-570.e3. \
171. Bandyopadhyay, A., Kancharla, N., Javalkote, V. S., Dasgupta, S., & Brutnell, T. P. (2020). CRISPR-Cas12a (Cpf1): a versatile tool in the plant genome editing tool box for agricultural advancement. *Frontiers in Plant Science*, 11:584151.

172. Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., & Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*, 353(6299).
173. East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*, 538(7624), 270–273.
174. Liu, L., Li, X., Wang, J., Wang, M., Chen, P., Yin, M., Li, J., Sheng, G., & Wang, Y. (2017). Two distant catalytic sites are responsible for C2c2 RNase activities. *Cell*, 168(1–2), 121-134.e12.
175. Slaymaker, I. M., Mesa, P., Kellner, M. J., Kannan, S., Brignole, E., Koob, J., Feliciano, P. R., Stella, S., Abudayyeh, O. O., Gootenberg, J. S., Strecker, J., Montoya, G., & Zhang, F. (2019). High-resolution structure of Cas13b and biochemical characterization of RNA targeting and cleavage. *Cell Reports*, 26(13), 3741-3751.e5.
176. Aman, R., Mahas, A., Butt, H., Aljedaani, F., & Mahfouz, M. (2018). Engineering RNA virus interference via the CRISPR/Cas13 machinery in *Arabidopsis*. *Viruses*, 10(12), 732. v10120732
177. Kavuri, N. R., Ramasamy, M., Qi, Y., & Mandadi, K. (2022). Applications of CRISPR/Cas13-Based RNA editing in plants. *Cells*, 11(17), 2665.
178. Vink, J. N. A., Martens, K. J. A., Vlot, M., McKenzie, R. E., Almendros, C., Estrada Bonilla, B., Brocken, D. J. W., Hohlbein, J., & Brouns, S. J. J. (2020). Direct visualization of native CRISPR target search in live bacteria reveals Cascade DNA surveillance mechanism. *Molecular Cell*, 77(1), 39-50.e10.
179. McKenzie, R. E., Keizer, E. M., Vink, J. N. A., van Lopik, J., Büke, F., Kalkman, V., Fleck, C., Tans, S. J., & Brouns, S. J. J. (2022). Single cell variability of CRISPR-Cas interference and adaptation. *Molecular Systems Biology*, 18(4).
180. Liu, Z.-L., Hu, E.-Z., & Niu, D.-K. (2023). Investigating the relationship between CRISPR-Cas content and growth rate in bacteria. *Microbiology Spectrum*, 11(3).
181. Nobrega, F. L., Walinga, H., Dutilh, B. E., & Brouns, S. J. J. (2020). Prophages are associated with extensive CRISPR-Cas auto-immunity. *Nucleic Acids Research*, 48(21), 12074–12084.
182. Zaayman, M., & Wheatley, R. M. (2022). Fitness costs of CRISPR-Cas systems in bacteria. *Microbiology*, 168(7).
183. Szczelkun, M. D., Tikhomirova, M. S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., & Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proceedings of the National Academy of Sciences*, 111(27), 9798–9803.
184. Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M. S., Siksnys, V., & Seidel, R. (2015). Directional R-loop formation by the CRISPR-Cas surveillance complex Cascade provides efficient off-target site rejection. *Cell Reports*, 10(9), 1534–1543.
185. Redding, S., Sternberg, S. H., Marshall, M., Gibb, B., Bhat, P., Guegler, C. K., Wiedenheft, B., Doudna, J. A., & Greene, E. C. (2015). Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell*, 163(4), 854–865.
186. Yoshimi, K., Takeshita, K., Kodera, N., Shibumura, S., Yamauchi, Y., Omatsu, M., Umeda, K., Kunihiro, Y., Yamamoto, M., & Mashimo, T. (2022). Dynamic mechanisms of CRISPR interference by *Escherichia coli* CRISPR-Cas3. *Nature Communications*, 13(1), 4917.
187. Dillard, K. E., Brown, M. W., Johnson, N. v., Xiao, Y., Dolan, A., Hernandez, E., Dahlhauser, S. D., Kim, Y., Myler, L. R., Anslyn, E. v., Ke, A., & Finkelstein, I. J. (2018). Assembly and translocation of a CRISPR-Cas primed acquisition complex. *Cell*, 175(4), 934-946.e15.
188. Hayes, R., Xiao, Y., Ding, F. *et al.* (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530, 499–503.
189. Xiao, Y., Luo, M., Hayes, R. P., Kim, J., Ng, S., Ding, F., Liao, M., & Ke, A. (2017). Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR-Cas system. *Cell*, 170(1), 48-60.e11.
190. Lee, H., Popodi, E., Tang, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proceedings of the National Academy of Sciences*, 109(41).
191. Stern, A., Keren, L., Wurtzel, O., Amitai, G., & Sorek, R. (2010). Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends in Genetics*, 26(8), 335–340.
192. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L.A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* 12: 177–186 (2012).

193. Jiang, W., *et al.* (2013) Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. *PLoS Genet.* 9(9):e1003844.
194. Stout, E.A., *et al.* (2018). Deletion-based escape of CRISPR-Cas9 targeting in *Lactobacillus gasseri*. *Microbiology* 164(9):1098-1111.
195. Sheng, Y., Wang, H., Ou, Y. *et al.* (2023). Insertion sequence transposition inactivates CRISPR-Cas immunity. *Nature Communications* 14, 4366.
196. Mahendra, C., Christie, K. A., Osuna, B. A., Pinilla-Redondo, R., Kleinstiver, B. P., & Bondy-Denomy, J. (2020). Broad-spectrum anti-CRISPR proteins facilitate horizontal gene transfer. *Nat Microbiol.* 5(4), 620–629.
197. Bondy-Denomy, J., Garcia, B., Strum, S. *et al.* (2015). Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* 526, 136–139.
198. Lu, W.-T., Trost, C. N., Müller-Esparza, H., Randau, L., & Davidson, A. R. (2021). Anti-CRISPR AcrIF9 functions by inducing the CRISPR–Cas complex to bind DNA non-specifically. *Nucleic Acids Research*, 49(6), 3381–3393.
199. Pawluk, A., Shah, M., Mejdani, M., Calmettes, C., Moraes, T. F., Davidson, A. R., & Maxwell, K. L. (2017). Disabling a type I-E CRISPR-Cas nuclease with a bacteriophage-encoded anti-CRISPR protein. *MBio*, 8(6):e01751-17.
200. Huang, L., *et al.* (2021). AcrDB: a database of anti-CRISPR operons in prokaryotes and viruses. *Nucleic Acids Research*, 49(D1), D622–D629.
201. Bondy-Denomy, J., Davidson, A. R., Doudna, J. A., Fineran, P. C., Maxwell, K. L., Moineau, S., Peng, X., Sontheimer, E. J., & Wiedenheft, B. (2018). A unified resource for tracking anti-CRISPR names. *The CRISPR Journal*, 1(5), 304–305.
202. Trasanidou, D., Gerós, A. S., Mohanraju, P., Nieuwenweg, A. C., Nobrega, F. L., & Staals, R. H. J. (2019). Keeping CRISPR in check: diverse mechanisms of phage-encoded anti-CRISPR. *FEMS Microbiology Letters*, 366(9).
203. Pinilla-Redondo, R., Shehreen, S., Marino, N.D. *et al.* (2020). Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nature Communications* 11, 5652.
204. Davidson, A. R., Lu, W.-T., Stanley, S. Y., Wang, J., Mejdani, M., Trost, C. N., Hicks, B. T., Lee, J., & Sontheimer, E. J. (2020). Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems. *Annual Review of Biochemistry*, 89(1), 309–332.
205. León, L. M., Park, A. E., Borges, A. L., Zhang, J. Y., & Bondy-Denomy, J. (2021). Mobile element warfare via CRISPR and anti-CRISPR in *Pseudomonas aeruginosa*. *Nucleic Acids Research*, 49(4), 2114–2125.
206. Marino, N. D., Pinilla-Redondo, R., Csörgő, B., & Bondy-Denomy, J. (2020). Anti-CRISPR protein applications: natural brakes for CRISPR-Cas technologies. *Nature Methods*, 17(5), 471–479.
207. Gussow, A.B., Park, A.E., Borges, A.L. *et al.* (2020). Machine-learning approach expands the repertoire of anti-CRISPR protein families. *Nature Communications* 11, 3784.
208. Wandera, K. G., Alkhnabshi, O. S., Bassett, H. v. I., Mitrofanov, A., Hauns, S., Migur, A., Backofen, R., & Beisel, C. L. (2022). Anti-CRISPR prediction using deep learning reveals an inhibitor of Cas13b nucleases. *Molecular Cell*, 82(14), 2714–2726.e4.
209. Wang, C., Sun, Z., Hu, Y., Li, D., Guo, Q., & Wang, M. (2023). A novel anti-CRISPR AcrIE9.2 is associated with dissemination of *bla*_{KPC} plasmids in *Klebsiella pneumoniae* sequence type 15. *Antimicrobial Agents and Chemotherapy*, 67(4).
210. Kraus, C., & Sontheimer, E. J. (2023). Applications of anti-CRISPR proteins in genome editing and biotechnology. *Journal of Molecular Biology*, 435(13), 168120.
211. Aljeldah, M. M. (2022). Antimicrobial resistance and its spread is a global threat. *Antibiotics*, 11(8), 1082.
212. Davies, J., & Davies, D. (2010). Origins and evolution of antibiotic resistance. *Microbiology and Molecular Biology Reviews*, 74(3), 417–433.
213. Murray, C. J. L., Ikuta, K. S., Sharara, F., Swetschinski, L., Robles Aguilar, *et al.* (2022). Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325), 629–655.
214. Coculescu B.I. (2009). Antimicrobial resistance induced by genetic changes. *J Med Life*. Apr-Jun;2(2):114-23.
215. Michaelis, C., & Grohmann, E. (2023). Horizontal gene transfer of antibiotic resistance genes in biofilms. *Antibiotics*, 12(2), 328.
216. Humphrey, S., San Millán, Á., Toll-Riera, M. *et al.* (2021). Staphylococcal phages and pathogenicity islands drive plasmid evolution. *Nature Communications* 12, 5845.

217. Benz, F., & Hall, A. R. (2023). Host-specific plasmid evolution explains the variable spread of clinical antibiotic-resistance plasmids. *Proceedings of the National Academy of Sciences*, 120(15).
218. Lewis, K. (2008). Multidrug tolerance of biofilms and persister cells. In: Romeo, T. (eds) *Bacterial Biofilms. Current Topics in Microbiology and Immunology*, vol 322. Springer, Berlin, Heidelberg.
219. Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). mobile genetic elements associated with antimicrobial resistance. *Clinical Microbiology Reviews*, 31(4).
220. Li, X.Z., Hauer, B. & Rosche, B. Single-species microbial biofilm screening for industrial applications. *Appl Microbiol Biotechnol* 76, 1255–1262 (2007).
221. Andersson, S., Kuttuva Rajarao, G., Land, C. J., & Dalhammar, G. (2008). Biofilm formation and interactions of bacterial strains found in wastewater treatment systems. *FEMS Microbiology Letters*, 283(1), 83–90.
222. Mah, T.-F. (2012). Biofilm-specific antibiotic resistance. *Future Microbiology*, 7(9), 1061–1072.
223. Hatfull, G. F., Dedrick, R. M., & Schooley, R. T. (2022). Phage therapy for antibiotic-resistant bacterial infections. *Annual Review of Medicine*, 73(1), 197–211.
224. Srinivasan, R., Kannappan, A., Shi, C., & Lin, X. (2021). Marine bacterial secondary metabolites: a treasure house for structurally unique and effective antimicrobial compounds. *Marine Drugs*, 19(10), 530.
225. Bar-On, Y. M., Phillips, R., & Milo, R. (2018). The biomass distribution on Earth. *Proceedings of the National Academy of Sciences*, 115(25), 6506–6511.
226. Louca, S., Mazel, F., Doebeli, M., & Parfrey, L. W. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLoS Biology*, 17(2), e3000106.
227. Parks, D.H., Chuvochina, M., Chaumeil, P.A. *et al.* (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* 38, 1079–1086.
228. Steen, A.D., Crits-Christoph, A., Carini, P. *et al.* (2019). High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J* 13, 3126–3130.
229. Pérez-Cobas, A. E., Gomez-Valero, L., & Buchrieser, C. (2020). Metagenomic approaches in microbial ecology: an update on whole-genome and marker gene sequencing analyses. *Microbial Genomics*, 6(8):mgen000409.
230. Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), 2761–2764.
231. Xie, H., Yang, C., Sun, Y., Igarashi, Y., Jin, T., & Luo, F. (2020). PacBio long reads improve metagenomic assemblies, gene catalogs, and genome binning. *Frontiers in Genetics*, 11:516269.
232. Chen, L., Zhao, N., Cao, J. *et al.* (2022). Short- and long-read metagenomics expand individualized structural variations in gut microbiomes. *Nature Communications* 13, 3175.
233. Santana-Pereira, A. L. R., Sandoval-Powers, M., Monsma, S., Zhou, J., Santos, S. R., Mead, D. A., & Liles, M. R. (2020). Discovery of novel biosynthetic gene cluster diversity from a soil metagenomic library. *Frontiers in Microbiology*, 11.
234. Trivedi, P., Leach, J.E., Tringe, S.G. *et al.* (2020). Plant–microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 18, 607–621.
235. de Jonge, N., Carlsen, B., Christensen, M.H, Pertoldi, C. & Nielsen, J.L. (2022). The gut microbiome of 54 mammalian species. *Front. Microbiol.* 13:886252.
236. Dekaboruah, E., Suryavanshi, M., Chettri, D. *et al.* (2020). Human microbiome: an academic update on human body site specific surveillance and its possible role. *Arch Microbiol* 202, 2147–2167.
237. Samaddar, S., Karp, D. S., Schmidt, R., Devarajan, N., McGarvey, J. A., Pires, A. F. A., & Scow, K. (2021). Role of soil in the regulation of human and plant pathogens: soils' contributions to people. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376(1834), 20200179.
238. Stothart, M. R., Greuel, R. J., Gavriiliuc, S., Henry, A., Wilson, A. J., McLoughlin, P. D., & Poissant, J. (2021). Bacterial dispersal and drift drive microbiome diversity patterns within a population of feral hindgut fermenters. *Molecular Ecology*, 30(2), 555–571.
239. Penesyan, A., Paulsen, I.T., Kjelleberg, S. *et al.* (2021). Three faces of biofilms: a microbial lifestyle, a nascent multicellular organism, and an incubator for diversity. *npj Biofilms Microbiomes* 7, 80.
240. Gordon, V., Bakhtiari, L., & Kovach, K. (2019). From molecules to multispecies ecosystems: the roles of structure in bacterial biofilms. *Physical Biology*, 16(4), 041001.

241. Bai, Y., Müller, D., Srinivas, G. *et al.* (2015). Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* 528, 364–369.
242. Seymour, J., Amin, S., Raina, J.B. *et al.* (2017). Zooming in on the phycosphere: the ecological interface for phytoplankton–bacteria relationships. *Nat Microbiol* 2, 17065.
243. Tara Ocean Foundation., Tara Oceans., European Molecular Biology Laboratory (EMBL). *et al.* (2022). Priorities for ocean microbiome research. *Nat Microbiol* 7, 937–947.
244. Hoshino, T., Doi, H., Uramoto, G.-I., Wörmer, L., Adhikari, R. R., Xiao, N., Morono, Y., D’Hondt, S., Hinrichs, K.-U., & Inagaki, F. (2020). Global diversity of microbial communities in marine sediment. *Proceedings of the National Academy of Sciences*, 117(44), 27587–27597.
245. Paoli, L., Ruscheweyh, H.J., Forneris, C.C. *et al.* (2022). Biosynthetic potential of the global ocean microbiome. *Nature* 607, 111–118.
246. Travin, D. Y., Bikmetov, D., & Severinov, K. (2020). Translation-targeting RiPPs and where to find them. *Frontiers in Genetics*, 11:226.
247. Suttle, C. (2007). Marine viruses – major players in the global ecosystem. *Nat Rev Microbiol* 5, 801–812.
248. Tamulaitis, G., *et al.* (2014). Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Molecular Cell* 56(4):506-17.
249. Jiang, W., Samai, P. & Marraffini, L.A. (2016). Degradation of phage transcripts by CRISPR-associated rnases enables type III CRISPR-Cas immunity. *Cell* 164(4): 710–721.
250. van Houte, S. *et al.* The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* 532: 385–388 (2016).
251. Deveau, H., *et al.* (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol.* 190(4):1390–1400.
252. Severinov, K., Ispolatov, I. & Semenova, E. (2016). The influence of copy-number of targeted extrachromosomal genetic elements on the outcome of CRISPR-Cas defense. *Front. Mol. Biosci.* 3:45.
253. Shmakov, S., *et al.* (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Research*. 42(9):5907– 5916.
254. Semenova, E., *et al.* (2016). Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl Acad. Sci. USA* 113(27):7626–7631.
255. Alseth, E.O., *et al.* (2019). Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature* 574:549–552.
256. Lee, C. L., Ow, D.S.W. & Oh, S.K.W. (2006). Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *J Microbiol Methods* 65(2), 258–267.
257. Hall, B.G., Acar, H., Nandipati, A. & Barlow, M. (2014). Growth rates made easy. *Mol Biol Evol* 31, 232–238.
258. McDonald, J.C., *et al.* (2000). Fabrication of microfluidic systems in poly(dimethylsiloxane). *Electrophoresis* 21(1):27-40.
259. Kingsland, S. (1982). The refractory model: the logistic curve and the history of population ecology. *The Quarterly Review of Biology* 57(1):29–52.
260. Van Kampen, N.G. (2007). Stochastic processes in physics and chemistry, *vol 1 Elsevier*.
261. Robinson, J., Thorvaldsdóttir, H., Winckler, W. *et al.* (2011). Integrative genomics viewer. *Nat Biotechnol* 29, 24–26.
262. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. (2016). Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. *Microb Cell Fact* 15: 211.
263. Held, D., Yaeger, K. & Novy, R. (2003). New coexpression vectors for expanded compatibilities in *E. coli*. *inNovations* 18:4–6.
264. Ciofu, O., Beveridge, T.J., Kadurugamuwa, J., Walther-rasmussen, J. & Høiby, N. (2000). Chromosomal beta-lactamase is packaged into membrane vesicles and secreted from *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* 45:9–13.
265. Amanatidou, E., *et al.* (2019). Biofilms facilitate cheating and social exploitation of β -lactam resistance in *Escherichia coli*. *npj Biofilms Microbiomes* 5, 36.
266. Medaney, F., Dimitriu, T., Ellis, R. & Raymond, B. (2015). Live to cheat another day: bacterial dormancy facilitates the social exploitation of β -lactamases. *ISME J.* 10: 778–787.
267. Nicoloff, H. & Andersson, D.I. (2016). Indirect resistance to several classes of antibiotics in cocultures with resistant bacteria expressing antibiotic-modifying or -degrading enzymes. *J Antimicrob Chemother* 71(1):100-10.

268. Hoffman L.R., *et al.* (2005). Aminoglycoside antibiotics induce bacterial biofilm formation. *Nature* 436:1171–1175.
269. Westra E.R. *et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular Cell* 46(5):595–605 (2012).
270. Strotskaya, A. *et al.* The action of *Escherichia coli* CRISPR–Cas system on lytic bacteriophages with different lifestyles and development strategies. *Nucleic Acids Res* 45(4):1946–1957 (2017).
271. Gophna, U. *et al.* (2015). No evidence of inhibition of horizontal gene transfer by CRISPR–Cas on evolutionary timescales. *ISME J.* 9(9): 2021–2027.
272. Reisner, A., Haagenzen, J.A., Schembri, M.A., Zechner, E.L. & Molin, S. (2003). Development and maturation of *Escherichia coli* K-12 biofilms. *Molecular Microbiology* 48(4):933–46.
273. Stewart, P. & Franklin, M. (2008). Physiological heterogeneity in biofilms. *Nat Rev Microbiol* 6:199–210.
274. Nadell, C.D., Foster, K.R. & Xavier, J.B. (2010). Emergence of spatial structure in cell groups and the evolution of cooperation. *PLoS Comp. Biol.* 6(3): e1000716.
275. Nadell, C.D., Drescher, K. & Foster, K.R. (2016). Spatial structure, cooperation and competition in biofilms. *Nat Rev Microbiol* 14(9):589–600.
276. Frost, I., *et al.* (2018). Cooperation, competition and antibiotic resistance in bacterial colonies. *ISME J.* 12:1582–1593.
277. Haerter, J.O., Trusina, A. & Sneppen, K. (2011). Targeted bacterial immunity buffers phage diversity. *Journal of Virology* 85(20):10554–10560.
278. Haerter, J.O. & Sneppen, K. (2012). Spatial structure and lamarckian adaptation explain extreme genetic diversity at CRISPR locus. *MBio* 3(4):e00126–12–e00126–12.
279. Alseth, E.O., *et al.* (2019). Bacterial biodiversity drives the evolution of CRISPR-based phage resistance. *Nature* 574:549–552.
280. Kuczynska-Wisnik, D., Kędzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipinska, B., & Laskowska, E. (2002). The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. *Microbiology*, 148(6), 1757–1765.
281. Jair, K. W., Martin, R. G., Rosner, J. L., Fujita, N., Ishihama, A., & Wolf, R. E. (1995). Purification and regulatory properties of MarA protein, a transcriptional activator of *Escherichia coli* multiple antibiotic and superoxide resistance promoters. *Journal of Bacteriology*, 177(24), 7100–7104.
282. Vinué, L., McMurry, L. M., & Levy, S. B. (2013). The 216-bp *marB* gene of the *marRAB* operon in *Escherichia coli* encodes a periplasmic protein which reduces the transcription rate of *marA*. *FEMS Microbiology Letters*, 345(1), 49–55.
283. Sun, C., Guo, Y., Tang, K., Wen, Z., Li, B., Zeng, Z., & Wang, X. (2017). MqsR/MqsA Toxin/Antitoxin System Regulates Persistence and Biofilm Formation in *Pseudomonas putida* KT2440. *Frontiers in Microbiology*, 8:840.
284. Jovanovic, G., Engl, C., Mayhew, A. J., Burrows, P. C., & Buck, M. (2010). Properties of the phage-shock-protein (Psp) regulatory complex that govern signal transduction and induction of the Psp response in *Escherichia coli*. *Microbiology*, 156(10), 2920–2932.
285. Dworkin, J., Jovanovic, G., & Model, P. (2000). The PspA Protein of *Escherichia coli* is a negative regulator of ζ^{54} -dependent transcription. *Journal of Bacteriology*, 182(2), 311–319.
286. Kashiwagi, K., Miyamoto, S., Nukui, E., Kobayashi, H., & Igarashi, K. (1993). Functions of *potA* and *potD* proteins in spermidine-preferential uptake system in *Escherichia coli*. *Journal of Biological Chemistry*, 268(26), 19358–19363.
287. Gul, N., & Poolman, B. (2013). Functional reconstitution and osmoregulatory properties of the ProU ABC transporter from *Escherichia coli*. *Molecular Membrane Biology*, 30(2), 138–148.
288. Pyenson, N. C., & Marraffini, L. A. (2020). Co-evolution within structured bacterial communities results in multiple expansion of CRISPR loci and enhanced immunity. *ELife*, 9:e53078.
289. Costello, M.J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, H., Miloslavich, P. (2010). A census of marine biodiversity knowledge, resources, and future challenges. *PLoS One* 5:e12110.
290. Costello, M.J., Chaudhary, C. (2017). Marine biodiversity, biogeography, deep-sea gradients, and conservation. *Current Biology* 27:R511–R527.
291. Costello, M.J., Tsai, P., Wong, P.S., Cheung, A.K.L., Basher, Z., Chaudhary, C. (2017). Marine biogeographic realms and species endemism. *Nature Communications* 8:1057.

292. Lovejoy, C., Massana, R., Pedrós-Alió, C. (2006). Diversity and distribution of marine microbial eukaryotes in the Arctic Ocean and adjacent seas. *Appl Environ Microbiol* 72:3085–3095.
293. Costello, M.J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D., Hoeksema, B.W., Poore, G.C.B., van Soest, R.W.M., Stöhr, S., Walter, T.C. (2013). Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS One* 8:e51629.
294. Eilers, H., Pernthaler, J., Glöckner, F.O., Amann, R. (2000). Culturability and in situ abundance of pelagic bacteria from the North Sea. *Appl Environ Microbiol* 66:3044–3051.
295. Zhang, X-H., Ahmad, W., Zhu X-Y., Chen, J., Austin, B. (2021). Viable but nonculturable bacteria and their resuscitation: implications for cultivating uncultured marine microorganisms. *Mar Life Sci Technol* 3:189–203.
296. Tringe, S.G., Von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C. (2005). Comparative metagenomics of microbial communities. *Science* 308:554–557.
297. Sogin, M., Morrison, H., Huber, J., Welch, D., Huse, S., Neal, P., Arrieta, J., Herndl, G. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences* 103:12115–12120.
298. Taş, N., de Jong, A.E.E., Li, Y., Trubl, G., Xue, Y., Dove, N.C. (2021). Metagenomic tools in microbial ecology research. *Curr Opin Biotechnol* 67:184–191.
299. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A. (2015). Structure and function of the global ocean microbiome. *Science* 348:1261359.
300. Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doucier, G., Acinas, S.G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J.M. (2015). Patterns and ecological drivers of ocean viral communities. *Science* 348:1261498.
301. Royo-Llonch, M., Sánchez, P., Ruiz-González, C., Salazar, G., Pedrós-Alió, C., Sebastián, M., Labadie, K., Paoli, L., Ibarbalz, F., Zinger, L. (2021). Compendium of 530 metagenome- assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol* 6:1561–1574.
302. Gilbert, J.A., Jansson, J.K., Knight, R. (2018). Earth microbiome project and global systems biology. *mSystems* 3:3.
303. Gilbert, J.A., Jansson, J.K., Knight, R. (2014). The Earth Microbiome project: successes and aspirations. *BMC Biol* 12:1–4.
304. Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J., Prill, R.J., Tripathi, A., Gibbons, S.M., Ackermann, G. (2017). A communal catalogue reveals Earth’s multiscale microbial diversity. *Nature* 551:457–463.
305. Mavrich, T. N., Hatfull, G.F. (2017). Bacteriophage evolution differs by host, lifestyle and genome. *Nat Microbiol* 2:1–9.
306. Jordan, T.C., Burnett, S.H., Carson, S., Caruso, S.M., Clase, K., DeJong, R.J., Dennehy, J.J., Denver, D.R., Dunbar, D., Elgin, S. (2014). A broadly implementable research course in phage discovery and genomics for first-year undergraduate students. *mBio* 5:e01051-13.
307. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., Glöckner, F.O. (2014). The SILVA and “all-species living tree project (LTP)” taxonomic frameworks. *Nucleic Acids Research* 42:D643–D648.
308. Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.-H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* 12:635–645.
309. Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J. (2017). Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613.
310. DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069–5072.
311. Janda, J.M., Abbott, S.L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J Clin Microbiol* 45:2761–2764.
312. Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., Schulz, F., Jarett, J., Rivers, A.R., Eloe-Fadrosh, E.A. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731.
313. Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B.J., Evans, P. N., Hugenholtz P., Tyson, G.W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542.

314. Tully, B.J., Graham, E.D., Heidelberg, J.F. (2018). The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data* 5:1–8.
315. Antipov, D., Korobeynikov, A., McLean, J.S., Pevzner, P.A. (2016). hybridSPAdes: an algorithm for hybrid assembly of short and long reads. *Bioinformatics* 32:1009–1015.
316. Chen, Z., Erickson, D.L., Meng, J. (2020). Benchmarking hybrid assembly approaches for genomic analyses of bacterial pathogens using Illumina and Oxford Nanopore sequencing. *BMC Genomics* 21:1–21.
317. Videnska, P., Smerkova, K., Zwinsova, B., Popovici, V., Micenkova, L., Sedlar, K., Budinska, E. (2019). Stool sampling and DNA isolation kits affect DNA quality and bacterial composition following 16S rRNA gene sequencing using MiSeq Illumina platform. *Sci Rep* 9:1–14.
318. Abellan-Schneyder, I., Matchado, M.S., Reitmeier, S., Sommer, A., Sewald, Z., Baumbach, J., List, M., Neuhaus, K. (2021). Primer, pipelines, parameters: issues in 16S rRNA gene sequencing. *mSphere* 6:e01202-20.
319. Teng, F., Darveekaran Nair, S.S., Zhu, P., Li, S., Huang, S., Li, X., Xu, J., Yang, F. (2018). Impact of DNA extraction method and targeted 16S-rRNA hypervariable region on oral microbiota profiling. *Sci Rep* 8:16321.
320. Flint, A., Laidlaw, A., Li, L., Raitt, C., Rao, M., Cooper, A., Weedmark, K., Carrillo, C., Tamber, S. (2022). Choice of DNA extraction method affects detection of bacterial taxa from retail chicken breast. *BMC Microbiol* 22:230.
321. Simon, S.A., Schmidt, K., Griesdorn, L., Soares, A.R., Bornemann, T.L.V., Probst, A.J. (2023). Dancing the Nanopore limbo—Nanopore metagenomics from small DNA quantities for bacterial genome reconstruction. *BMC Genomics* 24, 727.
322. Nikrad, M.P., Cottrell, M.T., Kirchman, D.L. (2012). Abundance and single-cell activity of heterotrophic bacterial groups in the western Arctic Ocean in summer and winter. *Appl Environ Microbiol* 78:2402–2409.
323. Cardozo-Mino, M.G., Fadeev, E., Salman-Carvalho, V., Boetius, A. (2021). Spatial distribution of Arctic bacterioplankton abundance is linked to distinct water masses and summertime phytoplankton bloom dynamics (Fram Strait, 79 N). *Front Microbiol* 12:658803.
324. Schrader, C., Schielke, A., Ellerbroek, L., Johne, R. (2012). PCR inhibitors—occurrence, properties and removal. *J Appl Microbiol* 113:1014–1026.
325. Wang, Y., Fujii, T. (2011). Evaluation of methods of determining humic acids in nucleic acid samples for molecular biological analysis. *Biosci Biotechnol Biochem* 75:355–357.
326. Matheson, C.D., Gurney, C., Esau, N., Lehto, R. (2010). Assessing PCR inhibition from humic substances. *Open Enzym Inhib J* 3:38–45.
327. Bertrand, H., Poly, F., Lombard, N., Nalin, R., Vogel, T.M., Simonet, P. (2005). High molecular weight DNA recovery from soils prerequisite for biotechnological metagenomic library construction. *J Microbiol Methods* 62:1–11.
328. Lazarevic, V., Gaña, N., Girard, M., Mauffrey, F., Ruppé, E., Schrenzel, J. (2022). Effect of bacterial DNA enrichment on detection and quantification of bacteria in an infected tissue model by metagenomic next-generation sequencing. *ISME Communications* 2:122.
329. Nelson, M.T., Pope, C.E., Marsh, R.L., Wolter, D.J., Weiss, E.J., Hager, K.R., Vo, A.T., Brittnacher, M.J., Radey, M.C., Hayden, H.S. (2019). Human and extracellular DNA depletion for metagenomic analysis of complex clinical infection samples yields optimized viable microbiome profiles. *Cell Rep* 26:2227–2240.
330. Marchukov, D., Li, J., Juillerat, P., Misselwitz, B., Yilmaz, B. (2023). Benchmarking microbial DNA enrichment protocols from human intestinal biopsies. *Front Genet* 14.
331. Martí, J.M. (2019). Recentrifuge: Robust comparative analysis and contamination removal for metagenomics. *PLoS Comput Biol* 15:e1006967.
332. Borst, A., Box, A., Fluit, A.C. (2004). False-positive results and contamination in nucleic acid amplification assays: suggestions for a prevent and destroy strategy. *European journal of clinical microbiology and infectious diseases* 23:289–299.
333. Davidi, D., Fitzgerald, S., Glaspell, H.L., Jalbert, S., Klapperich, C.M., Landaverde, L., Maheras, S., Mattoon, S.E., Britto, V.M., Nguyen, G.T. (2021). Amplicon residues in research laboratories masquerade as COVID-19 in surveillance tests. *Cell reports methods* 1:100005.
334. Jurasz, H., Pawłowski, T., Perlejewski, K. (2021). Contamination issue in viral metagenomics: problems, solutions, and clinical perspectives. *Front Microbiol* 12:745076.

335. Davis, N.M., Proctor, D.M., Holmes, S.P., Relman, D.A., Callahan, B.J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* 6:1–14.
336. Thoendel, M., Jeraldo, P., Greenwood-Quaintance, K.E., Yao, J., Chia, N., Hanssen, A.D., Abdel, M.P., Patel, R. (2017). Impact of contaminating DNA in whole-genome amplification kits used for metagenomic shotgun sequencing for infection diagnosis. *J Clin Microbiol* 55:1789–1801.
337. Paniagua Voirol, L.R., Valsamakis, G., Yu, M., Johnston, P.R., Hilker, M. (2021). How the ‘kitome’ influences the characterization of bacterial communities in lepidopteran samples with low bacterial biomass. *J Appl Microbiol* 130:1780–1793.
338. Olomu, I.N., Pena-Cortes, L.C., Long, R.A., Vyas, A., Krichevskiy, O., Luellwitz, R., Singh, P., Mulks, M.H. (2020). Elimination of “kitome” and “splashome” contamination results in lack of detection of a unique placental microbiome. *BMC Microbiol* 20:1–19.
339. Roopnarain, A., Mukhuba, M., Adeleke, R., Moeletsi, M. (2017). Biases during DNA extraction affect bacterial and archaeal community profile of anaerobic digestion samples. *3 Biotech* 7:1–12.
340. Nearing, J.T., Comeau, A.M., Langille, M.G.I. (2021). Identifying biases and their potential solutions in human microbiome studies. *Microbiome* 9:1–22.
341. Vishnivetskaya, T.A., Layton, A.C., Lau, M.C.Y., Chauhan, A., Cheng, K.R., Meyers, A.J., Murphy, J.R., Rogers, A.W., Saaranya, G.S., Williams, D.E. (2014). Commercial DNA extraction kits impact observed microbial community composition in permafrost samples. *FEMS Microbiol Ecol* 87:217–230.
342. Carrigg, C., Rice, O., Kavanagh, S., Collins, G., O’Flaherty, V. (2007). DNA extraction method affects microbial community profiles from soils and sediment. *Appl Microbiol Biotechnol* 77:955–964.
343. Kearney, S.M., Gibbons, S.M., Poyet, M., Gurry, T., Bullock, K., Allegretti, J.R., Clish, C.B., Alm, E.J. (2018). Endospores and other lysis-resistant bacteria comprise a widely shared core community within the human microbiota. *ISME J* 12:2403–2416.
344. Junier, T., Paul, C., Corona Ramírez, A., Filippidou, S., Fatton, M., Lehmann, A., Chain, P.S., Ariztegui, D., Vennemann, T., Viollier, P.H. (2022). High diversity of lysis-resistant cells upon the application of targeted physical and chemical lysis to environmental samples originating from three different water bodies. *Environmental DNA* 4:1092–1105.
345. Padilla, C.C., Ganesh, S., Gantt, S., Huhman, A., Parris, D.J., Sarode, N., Stewart, F.J. (2015). Standard filtration practices may significantly distort planktonic microbial diversity estimates. *Front Microbiol* 6:547.
346. Zhang, B., Brock, M., Arana, C., Dende, C., van Oers, N.S., Hooper, L.V., Raj, P. (2021). Impact of bead-beating intensity on the genus- and species-level characterization of the gut microbiome using amplicon and complete 16S rRNA gene sequencing. *Front Cell Infect Microbiol* 11:678522.
347. Ketchum, R.N., Smith, E.G., Vaughan, G.O., Phippen, B.L., McParland, D., Al-Mansoori, N., Carrier, T.J., Burt, J.A., Reitzel, A.M. (2018). DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Front Mar Sci* 5:255.
348. de Liphthay, J.R., Enzinger, C., Johnsen, K., Aamand, J., Sørensen, S.J. (2004). Impact of DNA extraction method on bacterial community composition measured by denaturing gradient gel electrophoresis. *Soil Biol Biochem* 36:1607–1614.
349. Cruaud, P., Vigneron, A., Fradette, M., Charette, S.J., Rodriguez, M.J., Dorea, C.C., Culley, A.I. (2017). Open the Sterivex™ casing: An easy and effective way to improve DNA extraction yields. *Limnol Oceanogr Methods* 15:1015–1020.
350. Amplicon PCR, Clean-Up PCR, Index PCR. (2013). 16S Metagenomic sequencing library preparation. Illumina: San Diego, CA, USA.
351. Rutledge, R.G., Cote, C. (2003). Mathematics of quantitative kinetic PCR and the application of standard curves. *Nucleic Acids Research* 31:e93–e93.
352. Monier, A., Worden, A.Z., Richards, T.A. (2016). Phylogenetic diversity and biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the oceans. *Environ Microbiol Rep* 8:461–469.
353. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41:D590–D596.
354. McMurdie, P.J., Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217.

355. Miya, M., Minamoto, T., Yamanaka, H., Oka, S., Sato, K., Yamamoto, S., Sado, T., Doi, H. (2016). Use of a filter cartridge for filtration of water samples and extraction of environmental DNA. *Journal of Visualized Experiments* e54741.
356. Tsuji, S., Takahara, T., Doi, H., Shibata, N., Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis—A review of methods for collection, extraction, and detection. *Environmental DNA* 1:99–108.
357. Wong, M., Nakao, M., Hyodo, S. (2020). Field application of an improved protocol for environmental DNA extraction, purification, and measurement using Sterivex filter. *Sci Rep* 10:1–13.
358. Lu, H., Giordano, F., Ning, Z. (2016). Oxford Nanopore MinION sequencing and genome assembly. *Genomics Proteomics Bioinformatics* 14:265–279.
359. Heavens, D., Chooneea, D., Giolai, M., Cuber, P., Aanstad, P., Martin, S., Alston, M., Misra, R., Clark, M.D., Leggett, R.M. (2021). How low can you go? Driving down the DNA input requirements for nanopore sequencing. *bioRxiv* 2010–2021.
360. Manzari, C., Oranger, A., Fosso, B., Piancone, E., Pesole, G., D’Erchia, A.M. (2020). Accurate quantification of bacterial abundance in metagenomic DNAs accounting for variable DNA integrity levels. *Microb Genom* 6.
361. Hiramatsu, K., Matsuda, C., Masago, K., Toriyama, K., Sasaki, E., Fujita, Y., Haneda, M., Ebi, H., Shibata, N., Hosoda, W. (2023). Diagnostic utility of DNA integrity number as an indicator of sufficient DNA quality in next-generation sequencing-based genomic profiling. *Am J Clin Pathol* aqad046.
362. Olson, N.D., Morrow, J.B. (2012). DNA extract characterization process for microbial detection methods development and validation. *BMC Res Notes* 5:1–14.
363. Wang, H., Qi, J., Xiao, D., Wang, Z., Tian, K. (2017). A re-evaluation of dilution for eliminating PCR inhibition in soil DNA samples. *Soil Biol Biochem* 106:109–118.
364. McKee, A.M., Spear, S.F., Pierson, T.W. (2015). The effect of dilution and the use of a post-extraction nucleic acid purification column on the accuracy, precision, and inhibition of environmental DNA samples. *Biol Conserv* 183:70–76.
365. Wiedbrauk, D.L., Werner, J.C., Drevon, A.M. (1995). Inhibition of PCR by aqueous and vitreous fluids. *J Clin Microbiol* 33:2643–2646.
366. Penaloza, C., Gutierrez, A.P., Eöry, L., Wang, S., Guo, X., Archibald, A.L., Bean, T.P., Houston, R.D. (2021). A chromosome-level genome assembly for the Pacific oyster *Crassostrea gigas*. *Gigascience* 10: giab020.
367. Thoendel, M., Jeraldo, P.R., Greenwood-Quaintance, K.E., Yao, J.Z., Chia, N., Hanssen, A.D., Abdel, M.P., Patel, R. (2016). Comparison of microbial DNA enrichment tools for metagenomic whole genome sequencing. *J Microbiol Methods* 127:141–145.
368. Zemach, A., McDaniel, I.E., Silva, P., Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328:916–919.
369. Li, Q., Chen, Y., Zhang, S., Lyu, Y., Zou, Y., Li, J. (2022). DNA Enrichment Methods for Microbial Symbionts in Marine Bivalves. *Microorganisms* 10:393.
370. Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol* 12:1–12.
371. Knight, R., Vrbanac, A., Taylor, B.C., Aksenov, A., Callewaert, C., Debelius, J., Gonzalez, A., Kosciolek, T., McCall, L.I., McDonald, D. (2018). Best practices for analysing microbiomes. *Nat Rev Microbiol* 16:410–422.
372. Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N. (2017). Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35:833–844.
373. Blauwkamp, T.A., Thair, S., Rosen, M.J., Blair, L., Lindner, M.S., Vilfan, I.D., Kawli, T., Christians, F.C., Venkatasubrahmanyam, S., Wall, G.D. (2019). Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. *Nat Microbiol* 4:663–674.
374. Greathouse, K.L., Sinha, R., Vogtmann, E. (2019). DNA extraction for human microbiome studies: the issue of standardization. *Genome Biol* 20:1–4.
375. Ganda, E., Beck, K.L., Haiminen, N., Silverman, J.D., Kawas, B., Cronk, B.D., Anderson, R.R., Goodman, L.B., Wiedmann, M. (2021). DNA extraction and host depletion methods significantly impact and potentially bias bacterial detection in a biological fluid. *mSystems* 6:e00619-21.
376. Guo, F., Zhang, T. (2013). Biases during DNA extraction of activated sludge samples revealed by high throughput sequencing. *Appl Microbiol Biotechnol* 97:4607–4616.

377. Feinstein, L.M., Sul, W.J., Blackwood, C.B. (2009). Assessment of bias associated with incomplete extraction of microbial DNA from soil. *Appl Environ Microbiol* 75:5428– 5433.
378. Lim, M.Y., Song, E.J., Kim, S.H., Lee, J., Nam, Y.D. (2018). Comparison of DNA extraction methods for human gut microbial community profiling. *Syst Appl Microbiol* 41:151– 157.
379. Karstens, L., Siddiqui, N.Y., Zaza, T., Barstad, A., Amundsen, C.L., Sysoeva, T.A. (2021). Benchmarking DNA isolation kits used in analyses of the urinary microbiome. *Sci Rep* 11:1–9.
380. Kazantseva, J., Malv, E., Kaleda, A., Kallastu, A., Meikas, A. (2021). Optimisation of sample storage and DNA extraction for human gut microbiota studies. *BMC Microbiol* 21:1– 13.
381. Corcoll, N., Österlund, T., Sinclair, L., Eiler, A., Kristiansson, E., Backhaus, T., Eriksson, K.M. (2017). Comparison of four DNA extraction methods for comprehensive assessment of 16S rRNA bacterial diversity in marine biofilms using high-throughput sequencing. *FEMS Microbiol Lett* 364.
382. Ketchum, R.N., Smith, E.G., Vaughan, G.O., Phippen, B.L., McParland, D., Al-Mansoori, N., Carrier, T.J., Burt, J.A., Reitzel, A.M. (2018). DNA extraction method plays a significant role when defining bacterial community composition in the marine invertebrate *Echinometra mathaei*. *Front Mar Sci* 5:255.
383. Dahn, H.A., Mountcastle, J., Balacco, J., Winkler, S., Bista, I., Schmitt, A.D., Pettersson, O.V., Formenti, G., Oliver, K., Smith, M. (2022). Benchmarking ultra-high molecular weight DNA preservation methods for long-read and long-range sequencing. *Gigascience* 11.
384. Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J., Cotter, P.D. (2016). 16S rRNA gene sequencing of mock microbial populations-impact of DNA extraction method, primer choice and sequencing platform. *BMC Microbiol* 16:1–13.
385. Turlousse, D.M., Narita, K., Miura, T., Ohashi, A., Matsuda, M., Ohyama, Y., Shimamura, M., Furukawa, M., Kasahara, K., Kameyama, K. (2022). Characterization and Demonstration of Mock Communities as Control Reagents for Accurate Human Microbiome Community Measurements. *Microbiol Spectr* 10:e01915-21.
386. Mancabelli, L., Ciociola, T., Lugli, G.A., Tarracchini, C., Fontana, F., Viappiani, A., Turroni, F., Ticinesi, A., Meschi, T., Conti, S. (2022). Guideline for the analysis of the microbial communities of the human upper airways. *J Oral Microbiol* 14:2103282.
387. Tomley, F.M.; Shirley, M.W. (2009) Livestock infectious diseases and zoonoses. *Philos Trans R Soc Lond B Biol Sci*, 364, 2637–2642.
388. European Food Safety Authority (EFSA); European Centre for Disease prevention and Control (ECDC). The European Union Summary Report on Antimicrobial Resistance in zoonotic and indicator bacteria from humans, animals and food in 2018/2019. *EFSA J* 2021, 19, e06490.
389. Okello, E., Williams, D.R., ElAshmawy, W.R., Adams, J., Pereira, R.V., Lehenbauer, T.W., Aly, S.S. (2021). Survey on antimicrobial drug use practices in California preweaned dairy calves. *Front Vet Sci*, 8, 636670.
390. Parker, J.K., Davies, B.W. (2022). Microcins reveal natural mechanisms of bacterial manipulation to inform therapeutic development. *Microbiology*, 168, 001175.
391. Das, T.K., Pradhan, S., Chakrabarti, S., Mondal, K.C., Ghosh, K. (2022). Current status of probiotic and related health benefits. *App Food Res*, 2, 100185.
392. Hossain, M.I., Sadekuzzaman, M., Ha, S.D. (2017). Probiotics as potential alternative biocontrol agents in the agriculture and food industries: A review. *Food Res. Int.*, 100, 63–73.
393. Mazurek-Popczyk, J., Pisarska, J., Bok, E., Baldy-Chudzik, K. (2020). Antibacterial activity of bacteriocinogenic commensal *Escherichia coli* against zoonotic strains resistant and sensitive to antibiotics. *Antibiotics*, 9, 411.
394. Rebuffat, S. (2011). Bacteriocins from gram-negative bacteria: a classification? In *Prokaryotic Antimicrobial Peptides*; Drider, D., Rebuffat, S., Eds.; Springer: New York, NY, USA, pp. 55–72.
395. Grozdanov, L., Raasch, C., Schulze, J., Sonnenborn, U., Gottschalk, G., Hacker, J., Dobrindt, U. (2004). Analysis of the genome structure of the nonpathogenic probiotic *Escherichia coli* strain Nissle 1917. *J Bacteriol*, 186, 5432–5441.
396. Sonnenborn, U., Schulze, J. (2009) The non-pathogenic *Escherichia coli* strain Nissle 1917—features of a versatile probiotic. *Microb Ecol Health Dis*, 21, 122–158.
397. Belova, I.V., Tochilina, A.G., Solovieva, I.V., Gorlova, I.S., Efimov, E.I., Zhirnov, V.A., Ivanova, T.P. (2017) Phenotypic and genotypic characteristics of the probiotic strain *E. coli* M-17. *Mod Probl Sci Educ*, 3, 26533 (In Russian). Available online: <https://science-education.ru/ru/article/view?id=26533> (accessed on 18 May 2023).

398. Schamberger, G.P., Diez-Gonzalez, F. (2004). Characterization of colicinogenic *Escherichia coli* strains inhibitory to enterohemorrhagic *Escherichia coli*. *J Food Prot*, 67, 486–492.
399. Cameron, A., Zaheer, R., Adator, E.H., Barbieri, R., Reuter, T., McAllister, T.A. (2019) Bacteriocin occurrence and activity in *Escherichia coli* isolated from bovines and wastewater. *Toxins*, 11, 475.
400. Kozlovsky, Y.E., Ovcharova, A.N., Petrova, V.A., Plugina, I.V., Pustovalov, S.A., Petnikov, A.Y., Khomyakova, T.I., Magomedova, A.D., Chertovich, N.F. (2012). Comparative effectiveness estimation of some probiotic strains of *Escherichia coli* in experimental disbiosis and toxicoinfection. Achievements of science and technology of the agro-industrial complex. *NTP Anim Husb Feed Prod*, 4, 64–66 (In Russian).
401. Hrala, M., Bosák, J., Mícenková, L., Křenová, J., Lexa, M., Pirková, V., Tomáštková, Z., Koláčková, I., Šmajš, D. (2021). *Escherichia coli* strains producing selected bacteriocins inhibit porcine enterotoxigenic *Escherichia coli* (ETEC) under both in Vitro and in Vivo conditions. *Appl Environ Microbiol*, 87, e0312120.
402. van Zyl, W.F., Deane, S.M., Dicks, L.M.T. (2020). Molecular insights into probiotic mechanisms of action employed against intestinal pathogenic bacteria. *Gut Microbes*, 9, 1831339.
403. Eiseul, K., Seung-Min, Y., Dayoung, K., Hae-Yeong, K. (2022). Complete genome sequencing and comparative genomics of three potential probiotic strains, *Lacticaseibacillus casei* FBL6, *Lacticaseibacillus chiayiensis* FBL7, and *Lacticaseibacillus zae* FBL8. *Front Microbiol*, 12, 794315.
404. Kuznetsova, M.V., Mihailovskaya, V.S., Remezovskaya, N.B., Starčić Erjavec, M. (2022). Bacteriocin-producing *Escherichia coli* isolated from the gastrointestinal tract of farm animals: Prevalence, molecular characterization and potential for application. *Microorganisms*, 10, 1558.
405. Cumsille, A., Durán, R.E., Rodríguez-Delherbe, A., Saona-Urmeneta, V., Cámara, B., Seeger, M., Araya, M., Jara, N., Buil-Aranda, C. (2023). GenoVi, an open-source automated circular genome visualizer for bacteria and archaea. *PLOS Comput Biol*, 19, e1010998.
406. Jarocki, P., Komoń-Janczara, E., Młodzińska, A., Sadurski, J., Kołodzińska, K., Łaczmanski, Ł., Panek, J., Frać, M. (2023). Occurrence and genetic diversity of prophage sequences identified in the genomes of *L. casei* group bacteria. *Sci Rep*, 13, 8603.
407. Wikivet (2012). Rat Haematology. Available online: https://en.wikivet.net/Rat_Haematology.
408. Voitenko, N.G., Makarova, M.N., Zueva, A.A. (2020). Variability of blood biochemical parameters and establishment of reference intervals in preclinical studies. Message 1: Rats. *Lab Anim Sci Res*, 1, 3.
409. Abrashova, T.V., Gushchin, Y.A., Kovaleva, M.A., Rybakova, A.V., Selezneva, A.I., Sokolova, A.P., Khodko, S.V. (2013). Handbook. Physiological, Biochemical and Biometric Indicators of the Norm of Experimental Animals, Publishing house “LEM”, St. Petersburg, Russia, p. 116.
410. Radwan, M., Rashed RHamoda, A.F., Amin, A., Sakaya, R.B. (2021). Experimental Infection with *E. coli* O157 in Rats and Its Toxic Effect, Biochemical and Histopathological Changes with Referee to Modern Therapy. *Ann Microbiol Immunol*, 4, 1024.
411. Upatissa, S., Mitchell, R.J. (2023). The “cins” of our fathers: Rejuvenated interest in colicins to combat drug resistance. *J Microbiol*, 61, 145–158.
412. Cursino, L., Smajš, D., Smarda, J., Nardi, R.M., Nicoli, J.R., Chartone-Souza, E., Nascimento, A.M. (2006). Exoproducts of the *Escherichia coli* strain H22 inhibiting some enteric pathogens both in vitro and in vivo. *J Appl Microbiol*, 100, 821–829.
413. Sassone-Corsi, M., Nuccio, S.P., Liu, H., Hernandez, D., Vu, C.T., Takahashi, A.A., Edwards, R.A., Raffatellu, M. (2016). Microcins mediate competition among *Enterobacteriaceae* in the inflamed gut. *Nature*, 540, 280–283.
414. Unnerstad, H.E., Lindberg, A., Waller, K.P., Ekman, T., Artursson, K., Nilsson-Öst, M., Bengtsson, B. (2009). Microbial aetiology of acute clinical mastitis and agent-specific risk factors. *Vet Microbiol*, 137, 90–97.
415. Saishu, N., Ozaki, H., Murase, T. (2014). CTX-M-type extended-spectrum beta-lactamase-producing *Klebsiella pneumoniae* isolated from cases of bovine mastitis in Japan. *J Vet Med Sci*, 76, 1153–1156.
416. Kline, K.A., Fälker, S., Dahlberg, S., Normark, S., Henriques-Normark, B. (2009). Bacterial adhesins in host-microbe interactions. *Cell Host Microbe*, 5, 580–592.
417. Yoshida, M., Thiriet-Rupert, S., Mayer, L., Beloin, C., Ghigo, J.-M. (2022). Selection for nonspecific adhesion is a driver of FimH evolution increasing *Escherichia coli* biofilm capacity. *microLife*, 3, uqac001.

418. Bhoite, S., van Gerven, N., Chapman, M.R., Remaut, H. (2019). Curli biogenesis: bacterial amyloid assembly by the type VIII secretion pathway. *EcoSal Plus*, 8, 163–171.
419. Nesta, B., Spraggon, G., Alteri, C., Moriel, D.G., Rosini, R., Veggi, D., Smith, S., Bertoldi, I., Pastorello, I., Ferlenghi, I., *et al.* (2012). FdeC, a novel broadly conserved *Escherichia coli* adhesin eliciting protection against urinary tract infections. *mBio*, 10, e00010–e00012.
420. Monteagudo-Mera, A., Rastall, R.A., Gibson, G.R., Charalampopoulos, D., Chatzifragkou, A. (2019). Adhesion mechanisms mediated by probiotics and prebiotics and their potential impact on human health. *Appl Microbiol Biotechnol*, 103, 6463–6472.
421. Papadimitriou, K., Zoumpopoulou, G., Foligné, B., Alexandraki, V., Kazou, M., Pot, B., Tsakalidou, E. (2015). Discovering probiotic microorganisms: In vitro, in vivo, genetic and omics approaches. *Front Microbiol*, 6, 6–58.
422. Shahverdi, S., Barzegari, A.A., Bakhshayesh, R.V., Nami, Y. (2023). *In-vitro* and *in-vivo* antibacterial activity of potential probiotic *Lactobacillus paracasei* against *Staphylococcus aureus* and *Escherichia coli*. *Heliyon*, 9, e14641.
423. Melton-Celsa, A.R. (2014). Shiga toxin (Stx) classification, structure, and function. *Microbiol Spectr*, 2, EHEC-0024-2013.
424. Veilleux, S., Dubreuil, J.D. (2006). Presence of *Escherichia coli* carrying the EAST1 toxin gene in farm animals. *Vet Res*, 37, 3–13.
425. Lorenz, S.C., Son, I., Maounounen-Laasri, A., Lin, A., Fischer, M., Kase, J.A. (2013). Prevalence of hemolysin genes and comparison of ehxA subtype patterns in Shiga toxin-producing *Escherichia coli* (STEC) and non-STEC strains from clinical, food, and animal sources. *Appl Environ Microbiol*, 79, 6301–6311.
426. Ukena, S. N., Singh, A., Dringenberg, U., Engelhardt, R., Seidler, U., Hansen, W., Bleich, A., Bruder, D., Franzke, A., Rogler, G., Suerbaum, S., Buer, J., Gunzer, F., & Westendorf, A. M. (2007). Probiotic *Escherichia coli* Nissle 1917 inhibits leaky gut by enhancing mucosal integrity. *PLoS One*, 2(12), e1308.
427. Wick, R.R., Judd, L.M., Holt, K.E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol*, 20, 129.
428. Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*, 37, 540–546.
429. Wick, R.R., Schultz, M.B., Zobel, J., Holt, K.E. (2015). Bandage: Interactive visualisation of de novo genome assemblies. *Bioinformatics*, 31, 3350–3352.
430. Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E.P., Zaslavsky, L., Lomsadze, A., Pruitt, K.D., Borodovsky, M., Ostell, J. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44, 6614–6624.
431. Kleinheinz, K.A., Joensen, K.G., Larsen, M.V. (2014). Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. *Bacteriophage*, 4, e27943.
432. Wang, M., Goh, Y.X., Tai, C., Wang, H., Deng, Z., Ou, H.Y. (2022). VRprofile2: Detection of antibiotic resistance-associated mobilome in bacterial pathogens. *Nucleic Acids Research*, 50, W768–W773.
433. Seemann, T. (2023). Abricate, Github. Available online: <https://github.com/tseemann/abricate>.
434. Feldgarden, M., Brover, V., Haft, D.H., Prasad, A.B., Slotta, D.J., Tolstoy, I., Tyson, G.H., Zhao, S., Hsu, C.H., McDermott, P.F., *et al.* (2019). Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother*, 63, e00483-19.
435. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., Weber, T. (2021). antiSMASH 6.0: Improving cluster detection and comparison capabilities. *Nucleic Acids Research*, 49, W29–W35.
436. Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., Fetter, A., Terlouw, B.R., Metcalf, W.W., Helfrich, E.J.N., *et al.* (2023). antiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research*, 51, W46–W50.
437. Skinnider, M.A., Johnston, C.W., Gunabalasingam, M., Merwin, N.J., Kieliszek, A.M., MacLellan, R.J., Li, H., Ranieri, M.R.M., Webster, A.L.H., Cao, M.P.T., *et al.* (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nature Communications*, 11, 6058.
438. van Heel, A.J., de Jong, A., Song, C., Viel, J.H., Kok, J., Kuipers, O.P. (2018). BAGEL4: A user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research*, 46, W278–W281.

439. Wishart, D.S., Han, S., Saha, S., Oler, E., Peters, H., Grant, J.R., Stothard, P., Gautam, V. (2023). PHASTEST: Faster than PHASTER, better than PHAST. *Nucleic Acids Research*, 51, W443-W450.
440. Carattoli, A., Zankari, E., García-Fernández, A., Larsen, M.V., Lund, O., Villa, L., Møller, F. (2014). *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob Agents Chemother*, 58, 3895–3903.
441. Nikolaev, Y.A. (2000). Regulation of adhesion in *Pseudomonas fluorescens* bacteria under the influence of distant intercellular interactions. *Microbiology*, 69, 356–361 (In Russian).
442. Brilis, V.I., Brilene, T.A., Lentsner, K.P., Lentsner, A.A. (1986). Metodika izucheniya adgezivnogo processa mikroorganizmov. *Lab Delo*, 4, 210–215.
443. Eskova, A.I., Andryukov, B.G., Yakovlev, A.A., Kim, A.V., Ponomareva, A.L., Obuhova, V.S. (2022). Horizontal transfer of virulence factors by pathogenic *Enterobacteria* to marine saprotrophic bacteria during co-cultivation in biofilm. *BioTech*, 11, 17.
444. Lenchenko, E., Blumenkrants, D., Sachivkina, N., Shadrova, N., Ibragimova, A. (2020). Morphological and adhesive properties of *Klebsiella pneumoniae* biofilms. *Vet World*, 13, 197–200.
445. GOST 33215-2014, Guidelines for accommodation and care of animals. En-vironment, housing and management, Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals, and European Convention for the Protection of Vertebrate Animals Used for Experimental and other Scientific Purposes, ETS 123, Strasbourg, 1986.
446. Sarma, P.R. (1990). Red Cell Indices. *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., Walker, H.K., Hall, W.D., Hurst, J.W. Eds., Butterworths: Boston, MA, USA, Chapter 152.

Supplementary Information

For Chapter 2

a



b



Figure S2.1. Sequences of PAM and g8 protospacer derived from a pool of pG8 plasmids. a, Sequence logos show no escape mutations in PAM and protospacer of the pG8 plasmid purified from CRISPR ON cells similar to the plasmids purified from CRISPR OFF cells. **b,** PCR assay of CRISPR arrays in CRISPR OFF (OFF) and CRISPR ON (ON) colonies.

Figure S2.2 is put in Chapter 5 as Figure 5.1

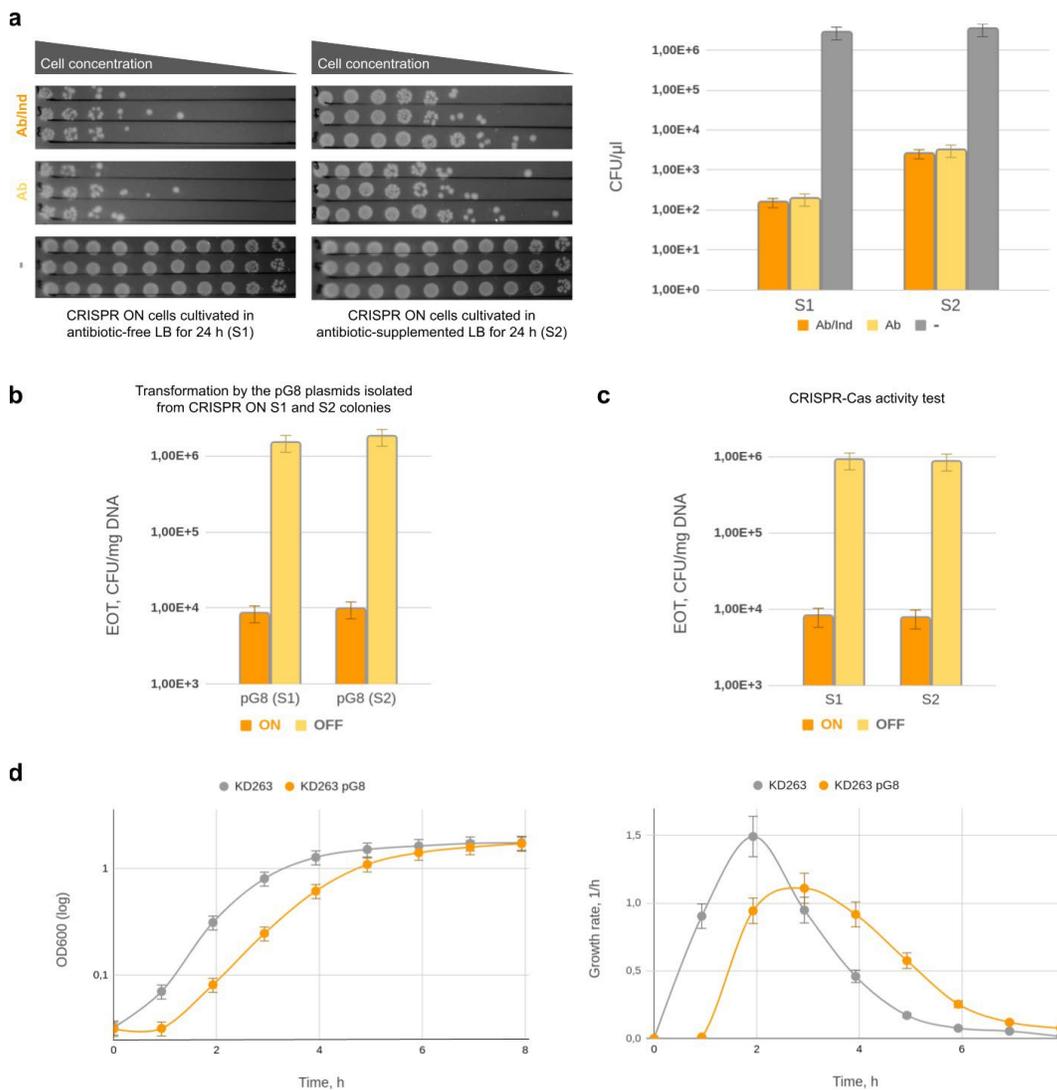


Figure S2.3. Cultivation of CRISPR ON cells bearing the pG8 plasmid in antibiotic-free LB media supplemented with *cas* genes expression inducers. **a**, The results of reseeds of cells from three randomly chosen CRISPR ON colonies after 24 h cultivation in LB supplemented with the *cas* genes expression inducers only (S1) or with both the inducers and 100 μ g/ml ampicillin (S2). 5 μ l aliquots of 4-fold serial dilutions of each culture were dropped on the surface LB agar plates without any additions (“-”, grey), supplemented with ampicillin only (“Ab”, yellow), or with both *cas* gene expression inducers and ampicillin (“Ab/Ind”, orange). The bars show CFU per μ l of culture calculated from results shown on the left. The error bars illustrate standard deviations. **b**, The results of transformation of KD263 competent cells with plasmids extracted from S1 and S2 CRISPR ON colonies formed on Ab/Ind plates in **a**. **c**, The results of transformation of competent cells prepared from CRISPR ON S1 and S2 colonies used in **b** with the pRSFG8 plasmid. **d**, Growth curves and calculated growth rates of plasmid-free KD263 cells and cells bearing the pG8 plasmid in LB media at 37 °C.

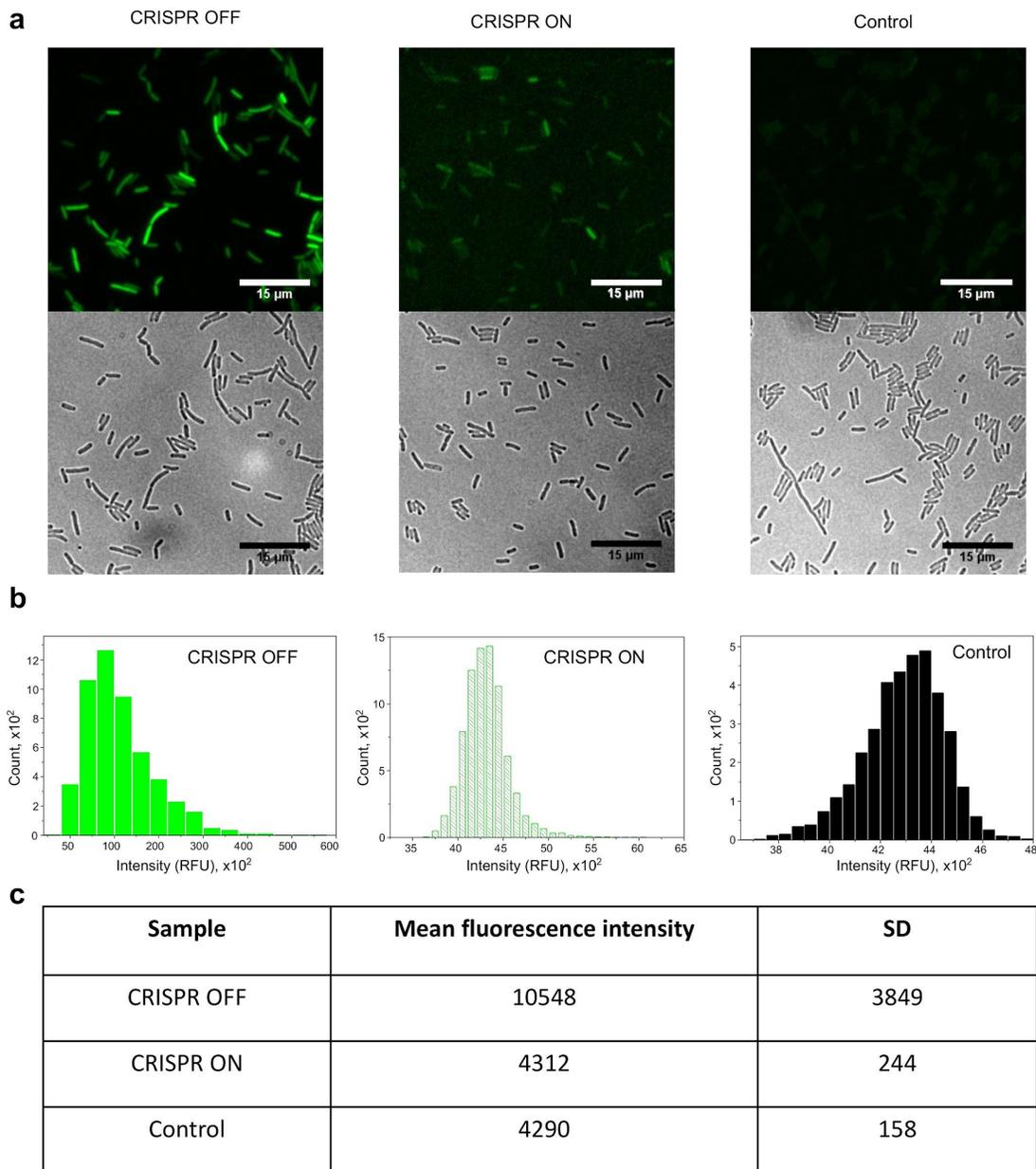


Figure S2.4. Fluorescence intensity of CRISPR ON and CRISPR OFF transformant cells. **a**, Images of cells derived from CRISPR ON and CRISPR OFF colonies. KD263 cells were used as a negative control. **b**, Histograms show the distribution of fluorescence of cells in relative fluorescence units (RFU). **c**, Statistics of fluorescence intensity.

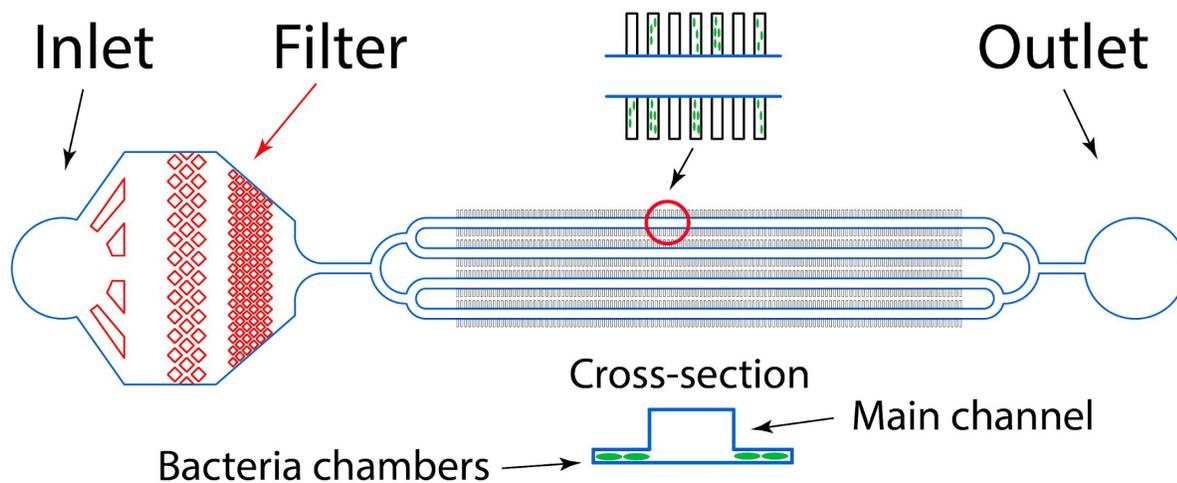


Figure S2.5. A scheme of the microfluidic device for live cell microscopy. The device contains 1000 growth chambers $80 \times 20 \times 1 \mu\text{m}$ located on both sides of the four major channels of $100 \mu\text{m}$ width and $40 \mu\text{m}$ depth each that are used for introducing bacteria cells into the growth chambers and for the fresh LB medium circulation. The inlet of the device contains a $25 \mu\text{m}$ filter to prevent clogging.

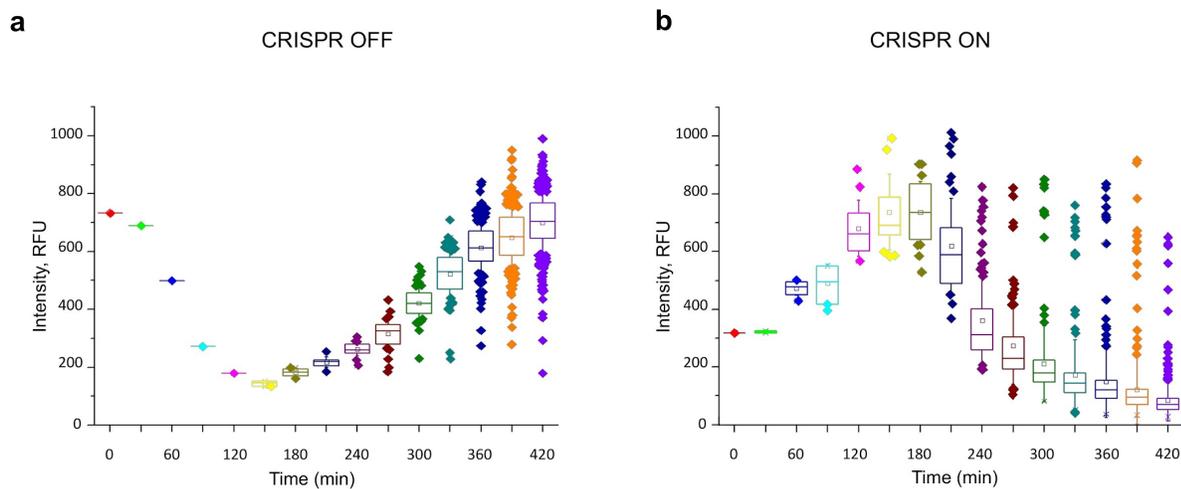


Figure S2.6. Fluorescence intensity change of CRISPR ON and CRISPR OFF cells in microfluidic channels over time.

a, CRISPR OFF fluorescent cells show normal fluorescence intensity in relative fluorescence units (RFU) as a positive control. **b**, The fluorescence intensity of cells in CRISPR ON experiment decreases over time after activation of the CRISPR-Cas system at 2-3 hour points, but a small amount of cells ($\approx 6-7\%$) remain fluorescent. At the same time, the intensity of fluorescent CRISPR ON cells becomes less than the mean level (≈ 700 RFU) inherent to CRISPR OFF cells.

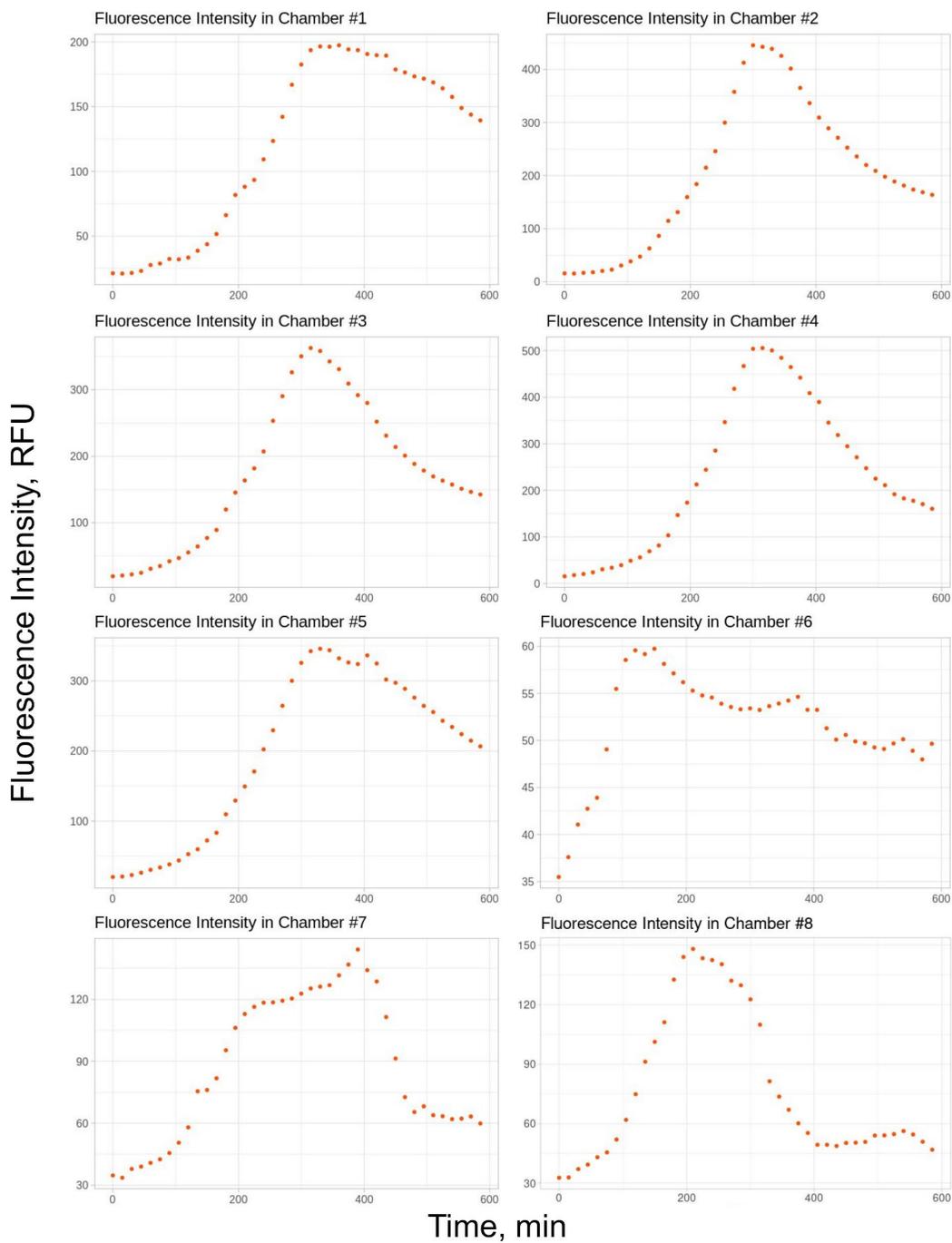


Figure S2.7. Mean fluorescent intensity per microfluidics channel area over time. Cultivation of KD263 cells bearing plasmid pG8-GFP was observed in LB media supplemented with *cas* genes expression inducers. Mean fluorescence intensity in relative fluorescence units (RFU) from all KD263 cells for each growth chamber was normalized on background level and calculated. All plots demonstrate an increase of mean fluorescence intensity at the initial stages of observation, but starting from 300 min after induction the fluorescence intensity decreases.

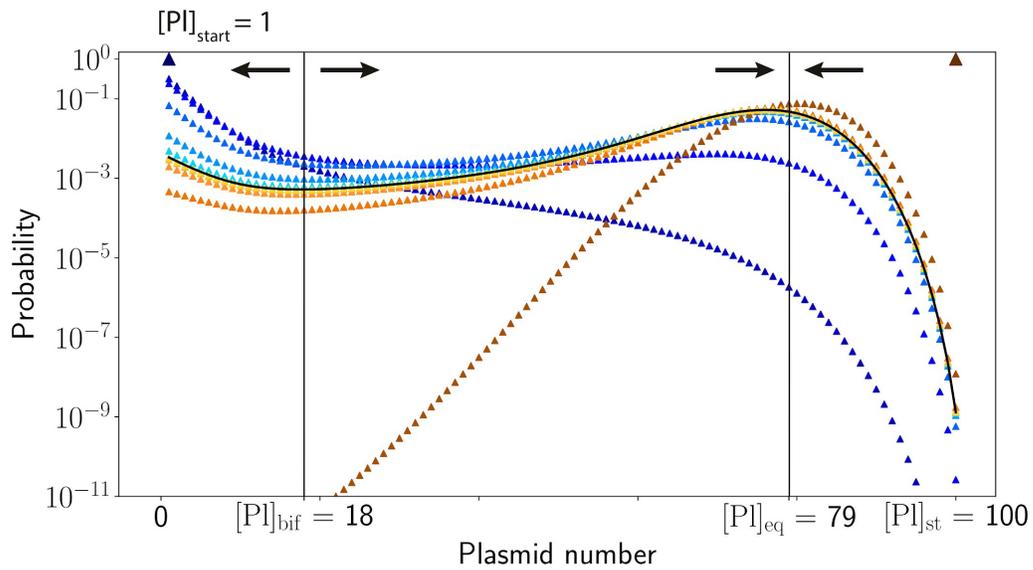


Figure S2.8. Convergence of PCN probability distribution to the universal scaling form. The blue triangles show the evolution of $P_k(t)$ when cells initially contain a single plasmid, while the orange/brown squares show the evolution of $P_k(t)$ when cells initially contain $[PI]_{st}$ plasmids. Shades of blue and red correspond to the different generations of cells from generation 0 (dark triangle and square) to generation 5 (the lightest blue and orange). Both families of curves converge to the universal asymptotic curve shown by a black line.

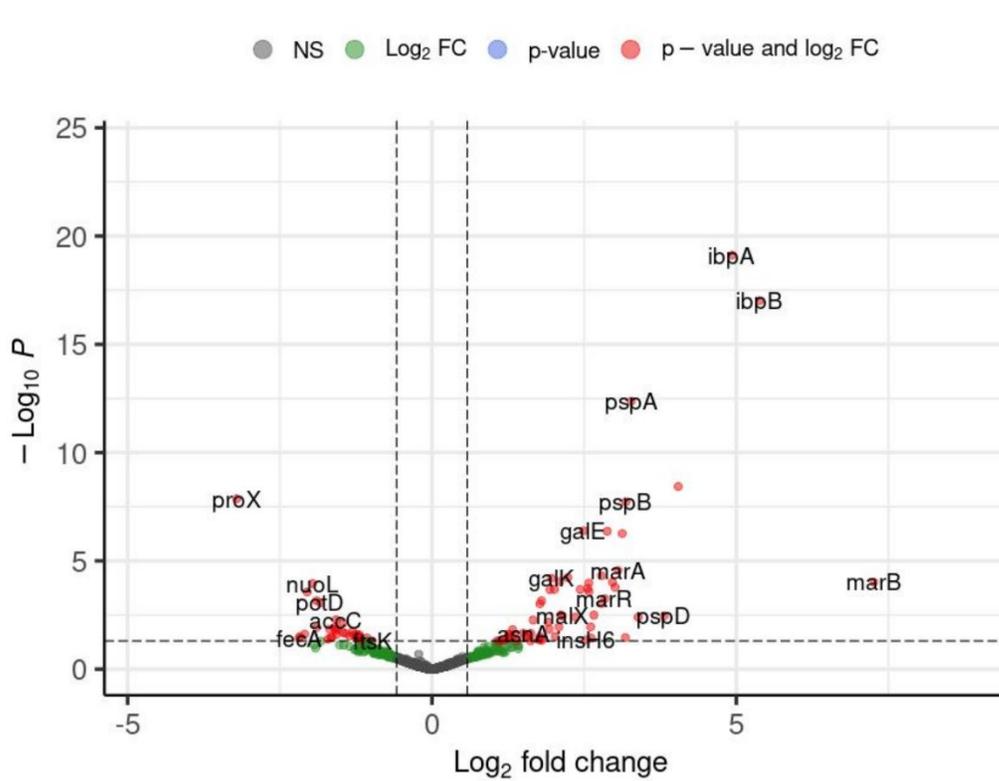


Figure S2.9. Volcano plots of differentially expressed genes of CRISPR ON cells. The key genes of CRISPR ON cells demonstrating the significant difference in expression level in relation to CRISPR OFF control are indicated with red color. The analysis is based on three biological repeats.

Table S2.1. The list of mutations disrupting CRISPR-Cas activity in cells from bright CRISPR ON colonies.

Mutation	Position	Portion of reads, %	Gene	Translation
C>T	1,344,539	18	<i>cse1</i>	L93I
A>C	1,344,542	33	<i>cse1</i>	N94H
insertion of G	1,334,547	34	<i>cse1</i>	frameshift
T>C	1,334,560	21	<i>cse1</i>	P100L
insertion of TGCAA	1,343,360	79	Non-coding region near promoter araBp	---
T>C	1,343,476	100	Non-coding region near promoter araBp	---
T>C	1,343,696	100	Non-coding region near promoter araBp	---
T>G	1,343,711	100	Non-coding region near promoter araBp	---

Table S2.2. Nucleotide sequences of primers used for cloning.

Primer name	Primer sequence	Purpose
pG8_dir	GACTCTAGAGGATCTACTAGTCATATGG	Amplification of the pG8 plasmid
pG8_rev	GACCTGCAGGCATGCAAGC	
TagGFP2_dir	AAAGCTTGCATGCCTGCAGGTCATGAGCGGGGCGAGGAG	Molecular cloning of <i>tagGFP2</i> gene into the pG8 plasmid (Gibson Assembly)
TagGFP2_rev	CTAGTAGATCCTCTAGAGTCTTACCTGTACAGCTCGTCCATGC	
GyrA_dir	CGGTCAACATTGAGGAAGAGC	qPCR assay of genome DNA
GyrA_rev	TACGTCACCAACGACACGG	
Bla_dir	TGAGTATTCAACATTTCCGTGTCCG	qPCR assay of the pG8 plasmid
Bla_rev	CGAAAACCTCTCAAGGATCTTACCG	
pRSF_ori_dir	GTCCGCTCTCCTGTTCCG	qPCR assay of the pRSFG8 plasmid
pRSF_ori_rev	AGCGTGAGCTATGAGAAAGCG	
Ec_LDR_F	AAGGTTGGTGGGTTGTTTTTATGG	PCR assay of CRISPR adaptation
M13_g8	GGATCGTCACCCTCAGCAGCG	

For Chapter 3

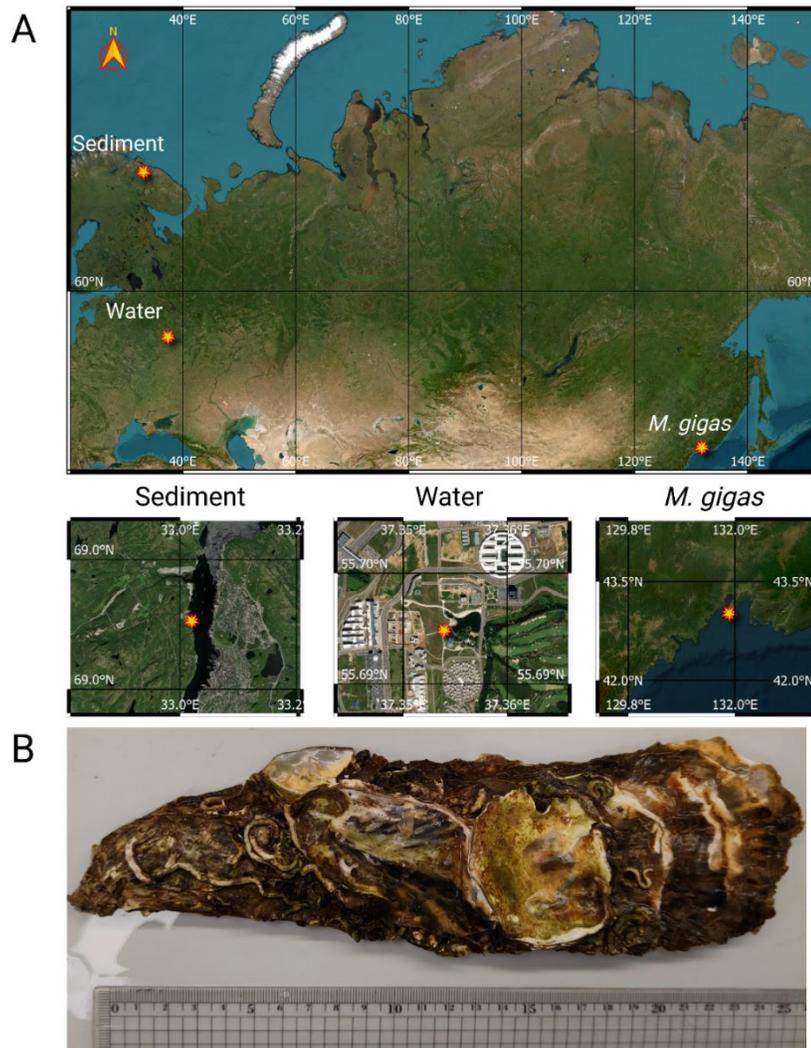


Fig. S3.1. A) Geographical locations for water, sea sediment, and Pacific oyster (*M. gigas*) collection spots. Map was prepared using Open Source Geospatial Foundation Project (<http://qgis.org>) **B)** Selected individual of Pacific oyster (*M. gigas*) aligned with a centimeter ruler.

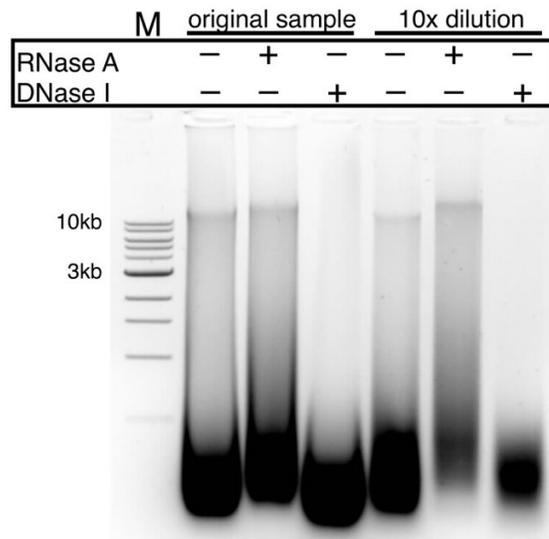


Fig. S3.2. DNase and RNase treatment of *M.gigas* DNA sample isolated with Stool kit.

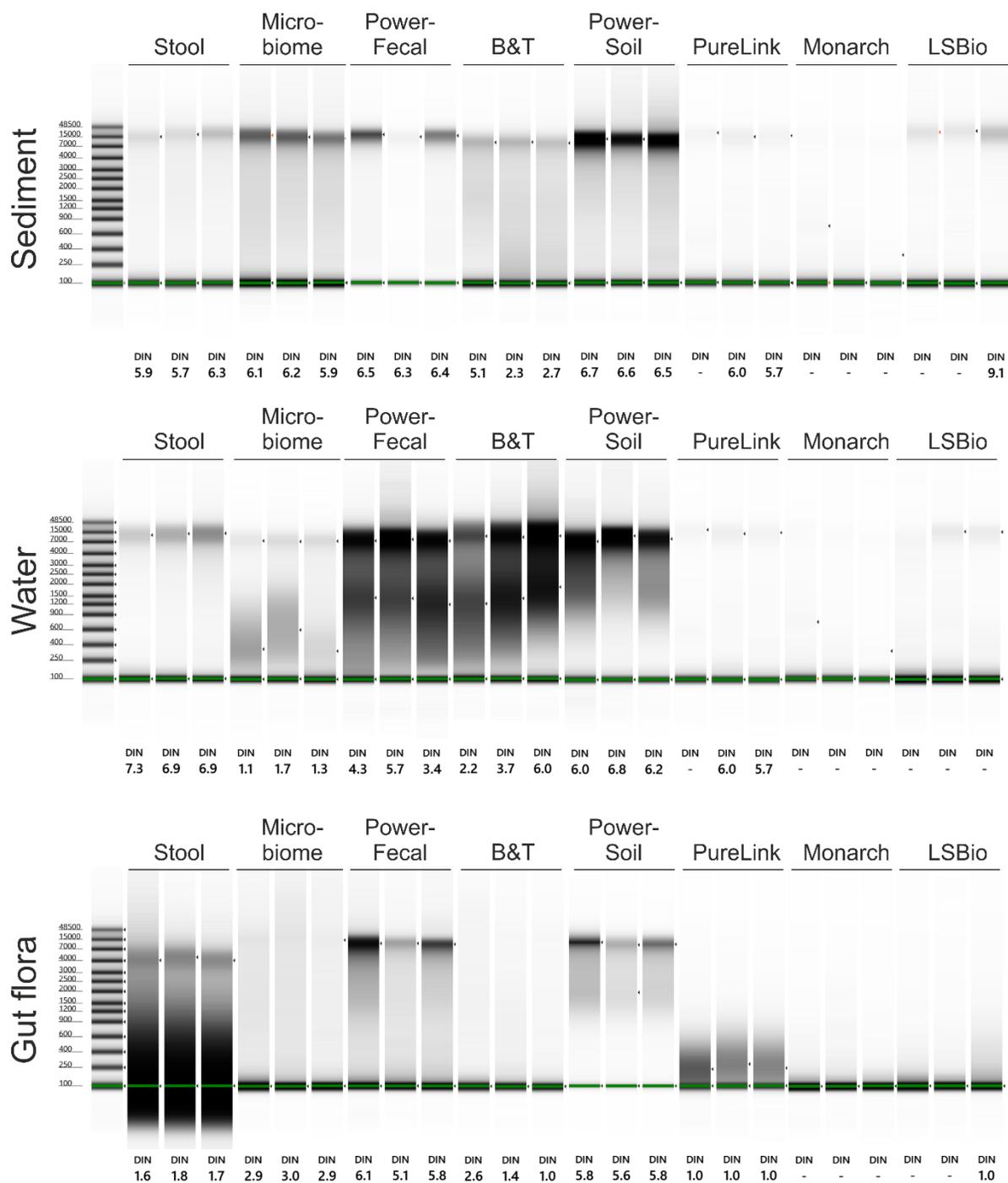


Fig. S3.3. Capillary electrophoresis estimation of the DIN values performed on a TapeStation 4150 (Agilent) with Genomic DNA ScreenTape System.

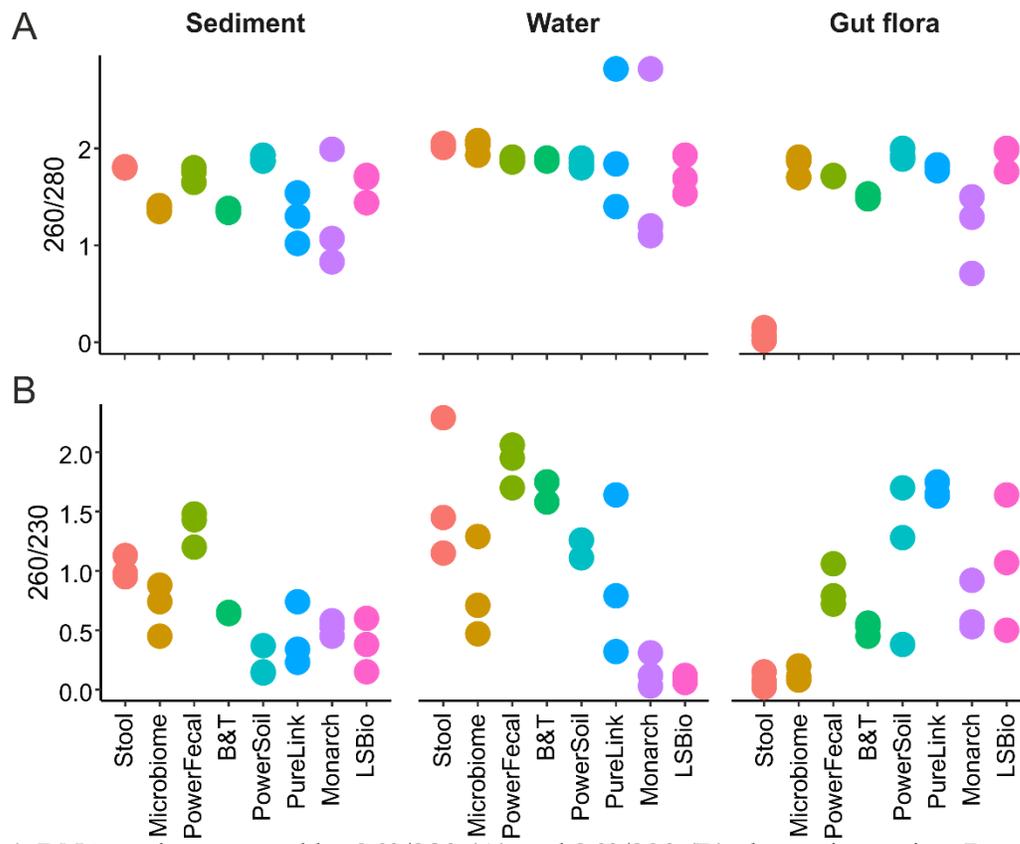


Fig. S4.4. DNA purity assessed by 260/280 (**A**) and 260/230 (**B**) absorption ratios. Data for three technical replicates are shown.

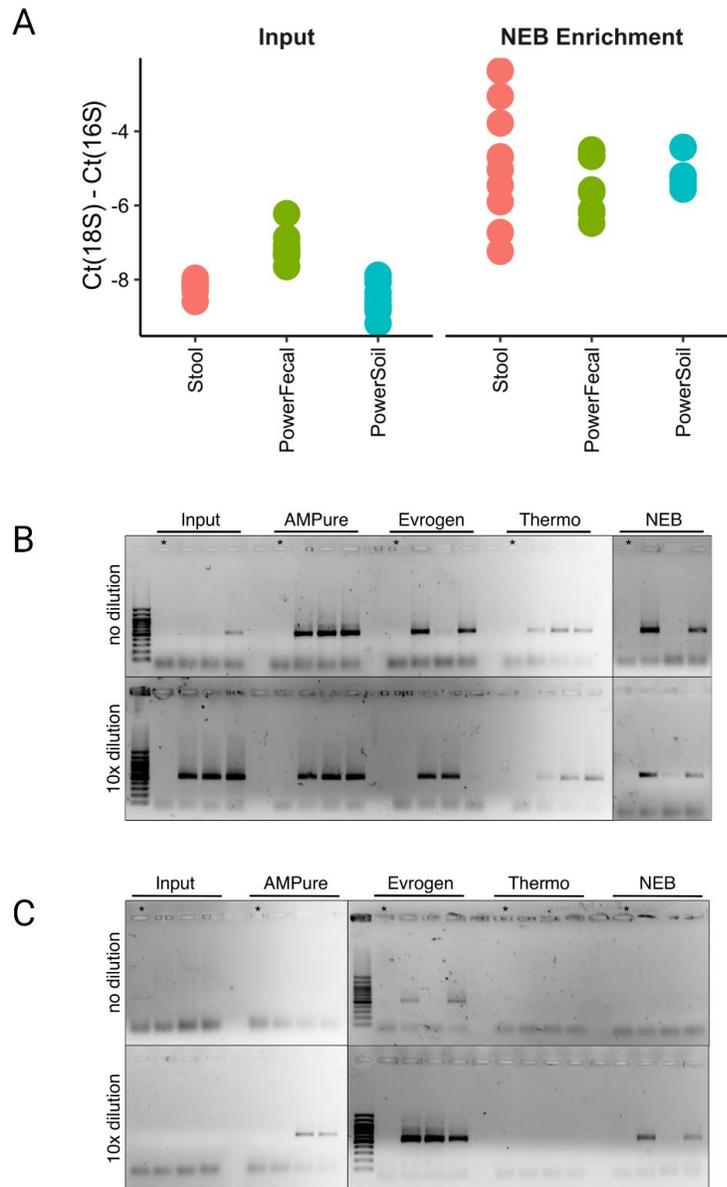


Fig. S4.5. (A) Retention of DNA relative to the input amount (50 ng for Stool or 150 ng for Microbiome kits) after additional re-purification procedures. (B-C) 16S rRNA gene PCR with sea sediment samples purified with Stool (B) and Microbiome (C) kits. PCR was performed with non-diluted and 10-fold diluted input DNA or DNA additionally re-purified with indicated kits (see Methods). *- no-input control that was re-purified in parallel with experimental samples to estimate potential contamination of the re-purification kit solutions with microbial DNA.

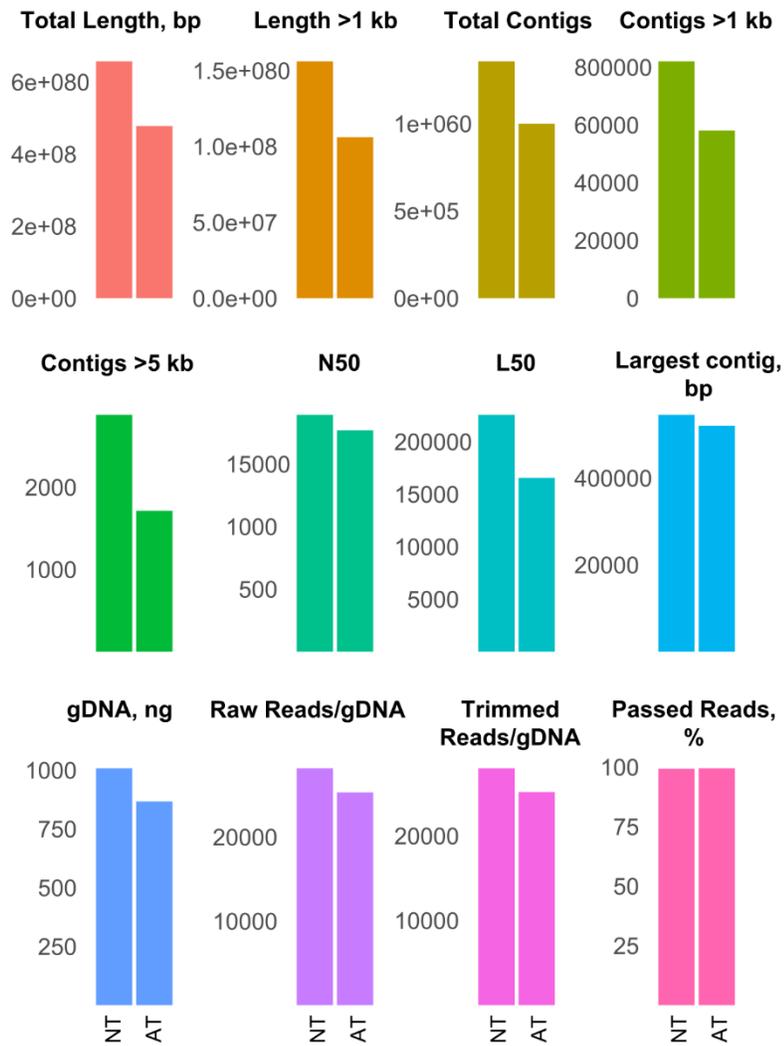


Fig. S4.6. Parameters of shotgun BGI libraries obtained from sediment sample purified with PowerFecal kit (NT – no treatment) and additionally re-purified with Evrogen column kit (AT - additional treatment).

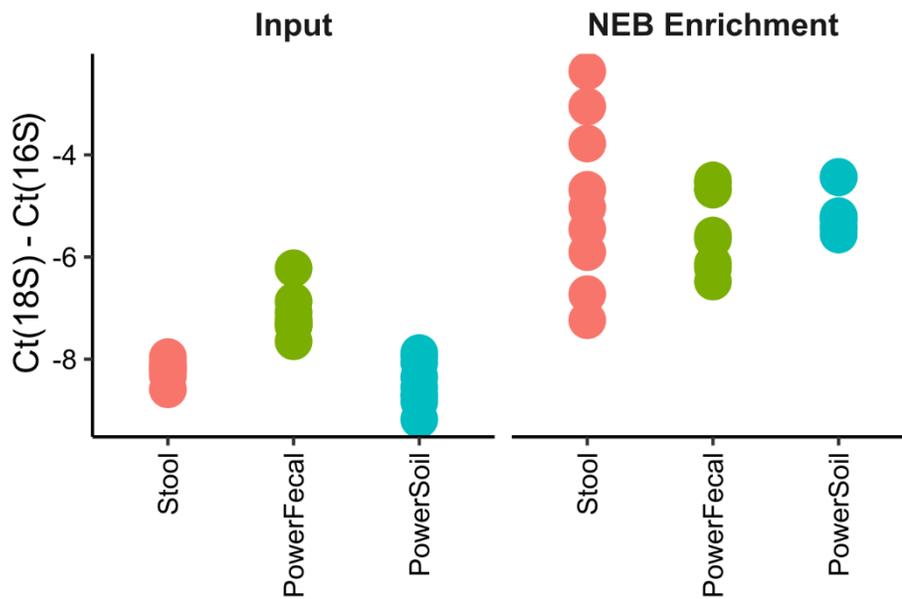


Fig. S4.7. qPCR Ct values obtained with 18S and 16S rRNA gene-specific primers with the input and NEB Enrichment re-purified DNA samples. Data for three technical qPCR replicates for each of the three kit purification replicates are shown.

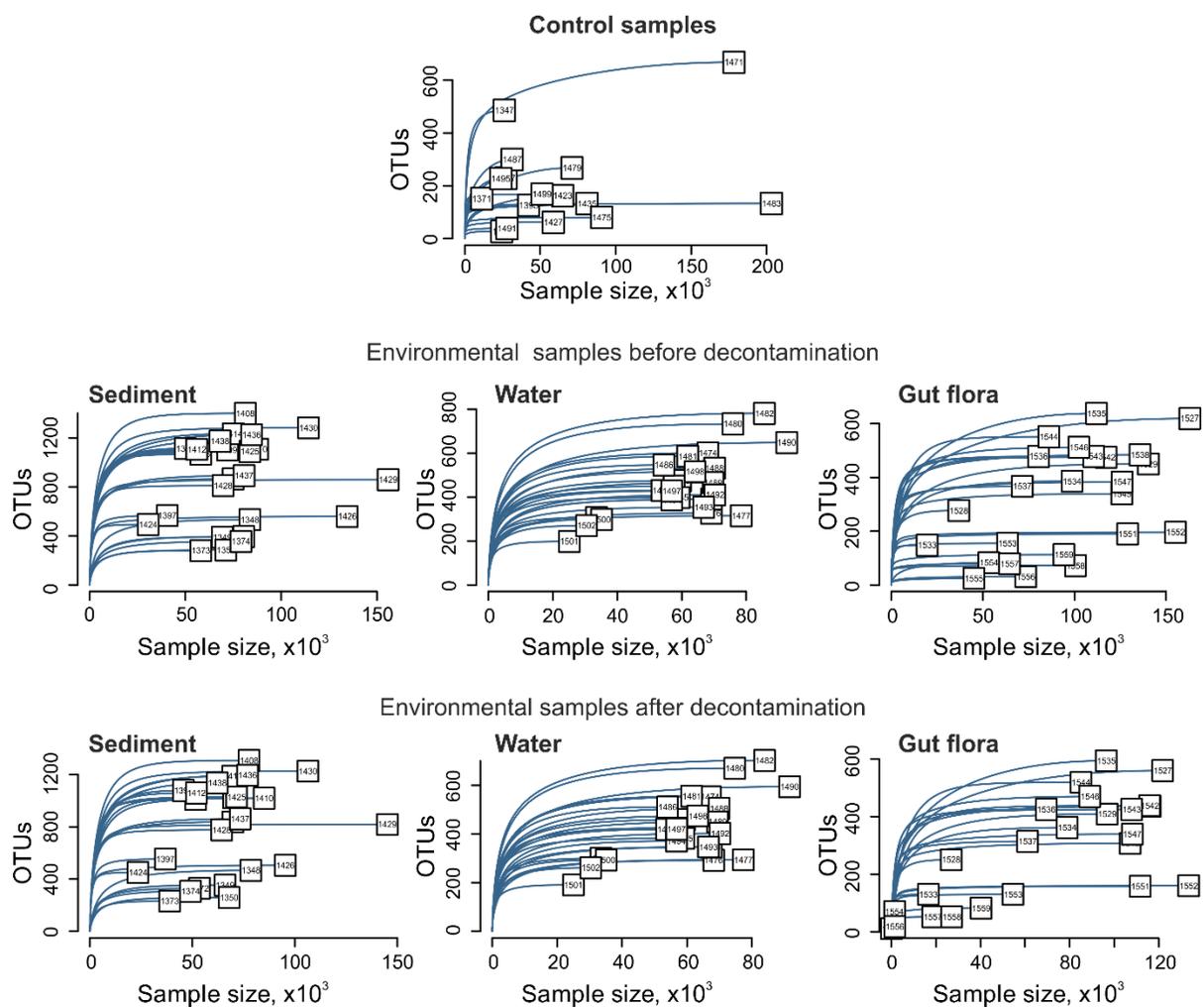


Fig. S4.8. Rarefaction curves for control samples (upper row), natural samples before decontamination (middle), and natural samples after decontamination (bottom row).

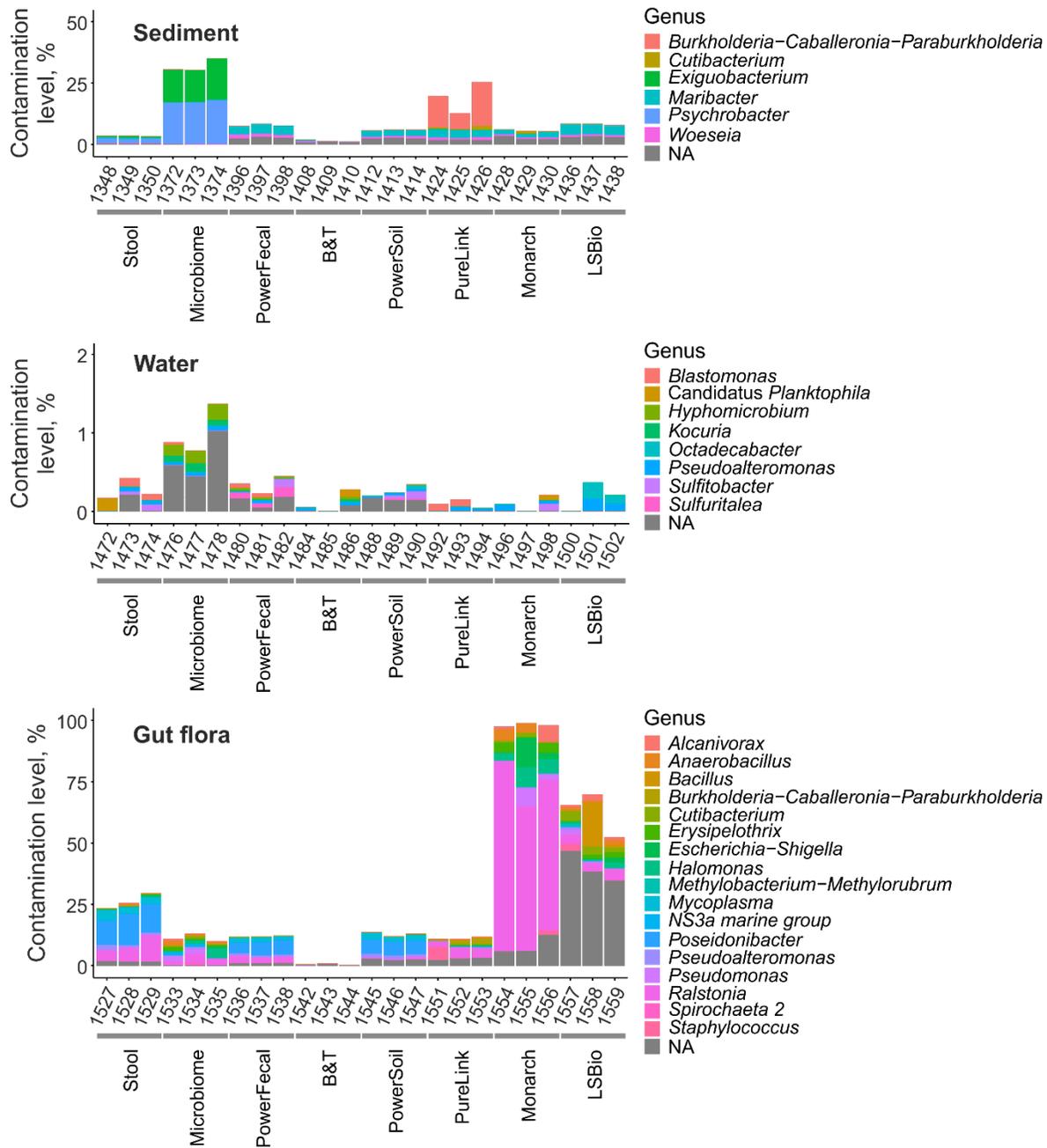


Fig. S4.9. Contamination levels of natural samples. Data is shown for all technical replicates independently. Genera with relative abundances >1% are shown.

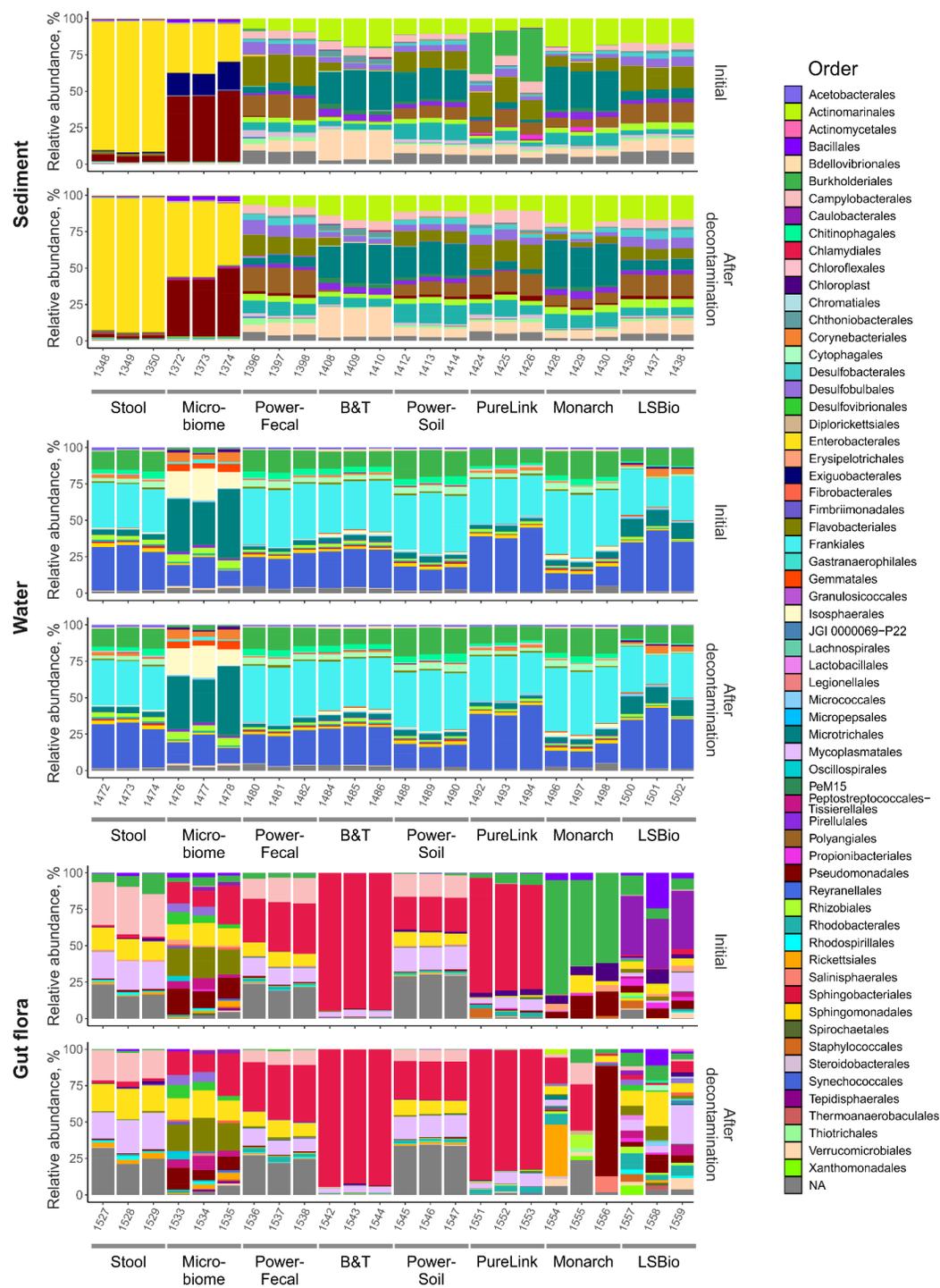


Fig. S4.10. Microbial communities' composition of natural samples before decontamination (Initial) and after the decontamination on an order level. Data is shown for all technical replicates independently. Orders with relative abundances >1% are shown.

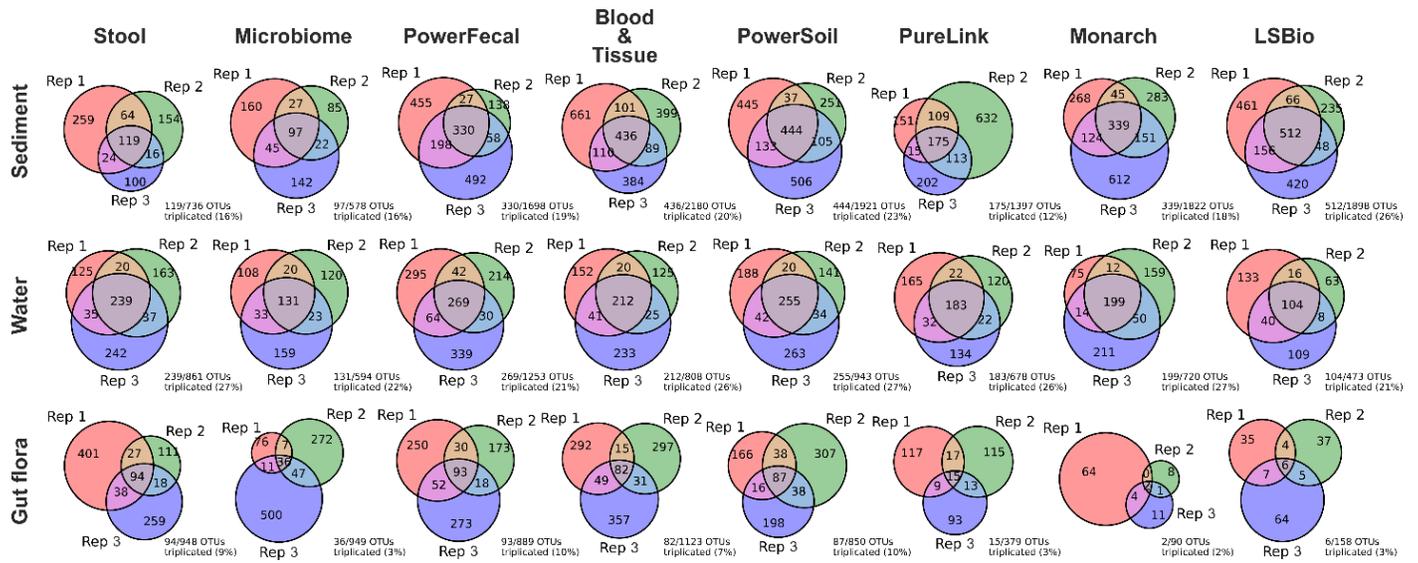


Fig. S4.11. Reproducibility of DNA-extraction kits. Venn diagrams representing the intersections of lists of non-zero OTUs (OTUs with a non-zero abundance) for three technical replicates obtained with specified DNA-extraction kits. Below the diagram, reproducibility level is shown (%) as a fraction of non-zero OTUs found in all three replicates from the total number of unique non-zero OTUs found in at least one replicate.

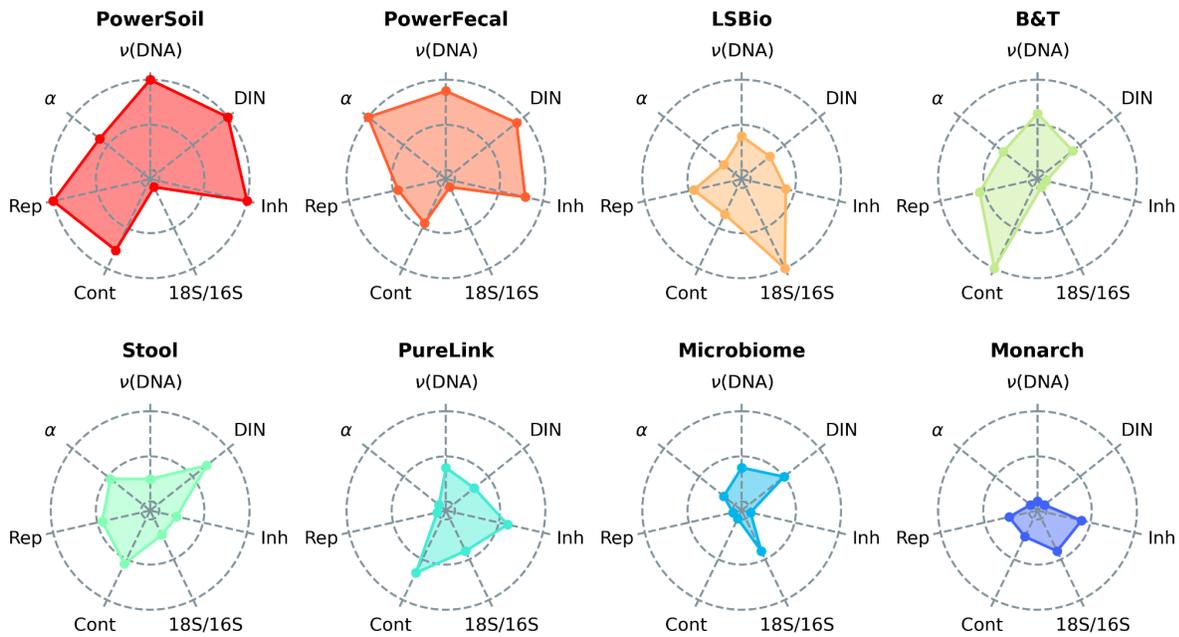


Fig. S4.12. Radar-plots demonstrating the average performance of DNA-extraction kits. $\nu(\text{DNA})$ - DNA yield, DIN - DNA integrity, Inh - presence of PCR inhibitors (higher rank indicates the lower level of inhibitors), 18S/16S - 18S/16S ratio (higher rank indicates the lower ratio), Cont - contamination level (higher rank indicates the lower level of contamination), Rep - reproducibility level, α - alpha-diversity. Kits were ordered by the sum of ranks.

For Chapter 4

The supporting information can be downloaded at:
<https://www.mdpi.com/article/10.3390/ijms241612636/s1>.