Chromosomal Architecture and its Influence on Gene Expression in Native and Engineered Bacteria

Dissertation

zur Erlangung des Grades eines
Doktor der Naturwissenschaften
(Dr. rer. nat.)
des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von
Marc Teufel

Aus Aachen

Marburg 2022

Die Untersuchungen zur vorliegenden Arbeit wurden von Juni 2018 bis Mai 2022 am Zentrum für Synthetische Mikrobiologie (SYNMIKRO) der Phillips-Universität Marburg unter der Leitung von Dr. Patrick Sobetzko durchgeführt.

Vom Fachbereich Biologie
der Philipps-Universität Marburg als Dissertation
angenommen am:
07.11.2022

Erstgutachter: Dr. Patrick Sobetzko

Zweitgutachterin: Prof. Dr. Anke Becker

Tag der Disputation:
25.11.2022

The results presented in this dissertation are included in the following original or pre-printed publications. The pre-print version is additionally submitted for publication in a scientific journal.

Teufel, M., Klein, C. A., Mager, M., Sobetzko, P. (2022). A multifunctional system for genome editing and large-scale interspecies gene transfer. Nature Communications, 13(1). https://doi.org/10.1038/s41467-022-30843-1

Teufel, M., Henkel, W., Sobetzko, P. (Submitted). The role of replication-induced chromosomal copy numbers in spatio-temporal gene regulation and evolutionary chromosome plasticity

Pre-print:

Teufel, M., Henkel, W., Sobetzko, P. (2022). The role of replication-induced chromosomal copy numbers in spatio-temporal gene regulation and evolutionary chromosome plasticity. In bioRxiv. https://doi.org/10.1101/2022.03.30.486354

Contents

1 Introduction

1.1 The bacterial chromosome

Most bacterial chromosomes are circular DNA molecules of about 2-8 megabases (Mb) in size, highly compacted into a structure called nucleoid, and occupy a relatively small fraction of the cell during growth [de Vries, 2010, Reyes-Lamothe et al., 2008]. This compaction is necessary due to the size of the chromosome. If the chromosome of *Escherichia coli*, for instance, were spread out, it would be 850 times larger than the cell length [Krogh et al., 2018]. The required compaction is imposed by molecular crowding, DNA polymer dynamics, DNA supercoiling and nucleoid-associated proteins (NAPs) [Dorman, 2013]. Besides the necessity of chromosome compaction due to its size, another important aspect of nucleoid formation is the functional organization of the DNA into a three-dimensional form. It needs to be compatible with DNA transaction processes such as replication, recombination, segregation and transcription [Verma et al., 2019]. Therefore, bacterial chromosomes are tightly balanced between efficient compaction within the cell and the accessibility for such DNA transaction processes [Glinkowska et al., 2021].

Research over the past decades has shown that chromosomal compaction takes place at multiple scales [Verma et al., 2019]. At the smallest scale (1kb or less) NAPs participate in the local compaction of DNA by bending, looping, bridging or wrapping the DNA (see Fig. 1.1A). At a larger scale (10kb or larger) transcription, replication and the action of topoisomerases generate supercoiling in the DNA helix resulting in the formation of plectonemic loops [Lagomarsino et al., 2015].

DNA supercoiling refers to an additional spatial twisting of the helical DNA that leads to a higher-order DNA structure. The DNA molecule consists of two interconnected strands that wind around each other with approximately 10.5bp per turn in its relaxed

6

state [Deweese et al., 2008]. However, the relaxed DNA structure does not exist in nature. In most living organisms, the chromosome is slightly underwound, which means the DNA exhibits on average more than 10.5bp per helical turn. In addition, the DNA can also be locally overwound. Both states are primarily catalyzed by an enzyme class called topoisomerases, but can also be induced by transcription and replication via unwinding DNA and the subsequent progression of the polymerases [Dorman, 2013]. The bacterial chromosome, as a double-stranded DNA molecule with a helical structure, is topologically constrained into a covalently closed circular shape, eliminating rotation of the free ends [Verma et al., 2019]. Therefore, changes in helical turns result in torsional and axial stress in the DNA molecule, which, in turn, causes the DNA to writhe about itself forming supercoils similar to twisted telephone cords [Deweese et al., 2008].

These plectonemic supercoils vary greatly in size and are highly dynamic [Sinden and Pettijohn, 1981, Verma et al., 2019]. They are organized in several topological microdomains defined by supercoiling diffusion barriers. The barriers act as fixation points for their topological domains and insulate them by inhibiting supercoiling diffusion. In other words, if a supercoiled DNA molecule, which is divided into several topological domains, is cut in one of those domains, only the corresponding domain is relaxed but not the others (see Fig. 1.1B) [Verma et al., 2019]. Possible mechanisms responsible for the formation of such supercoiling diffusion barriers can be NAP binding to two distinct sites, bacterial interspersed mosaic elements (BIMEs) or transcription [Verma et al., 2019]. The composition of NAPs and the global superhelical density of DNA change with the growth phase [Le Berre et al., 2022]. This suggests that bacterial chromosomes are not static and constrained into a single structure, but rather change their structure depending on the metabolic state of the cell. This further indicates a regulatory role of the organization of bacterial chromosomes in terms of gene expression. It could be shown that genes within the same microdomain tend to be co-regulated in *Mycoplasma pneumoniae*, suggesting that chromosome organization influences transcriptional regulation [Trussart et al., 2017]. Furthermore, inhibition of DNA supercoiling led to a decrease in the microdomain border strengths, indicating that supercoiling might play a role in the regulation of these domains.

The microdomains in the form of plectonemic loops can be joined together at the megabase scale and thereby forming distinct structures called macrodomains, which are characterised by enhanced DNA-DNA interactions within the same macrodomain
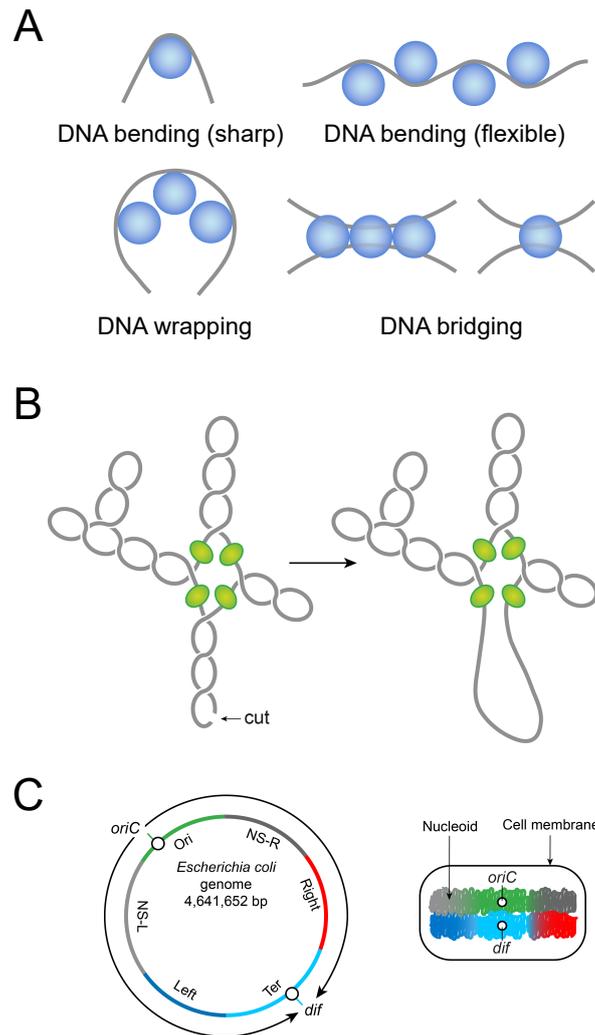
Figure 1.1: Nucleoid formation. **A** Shown are different modes of DNA (grey lines) organization by NAPs (blue circle) responsible for chromosome compaction. **B** Organization of plectonemic supercoils into topological microdomains by supercoiling diffusion barriers. A double-strand break in one of those topological domains would relax the DNA molecule and disband the supercoils. Due to supercoiling diffusion barriers (green), the DNA relaxation does not diffuse into neighbouring topological domains insulating each domain from adjacent supercoiling. **C** Illustration of an open conformation of the circular chromosome of *E. coli*. Replication is bi-directional (arrows) starting at the replication origin (oriC) and terminating at the *dif* site in the terminus region. The *E. coli* chromosome has four structured macrodomains (ori, ter, left, right) and two unstructured spatial domains (NS-L, NS-R). After compaction, those domains are specifically localized within the cell. In a newly born *E. coli* cell, for instance, oriC and dif are localized in the mid-cell, while the left and right macrodomains tend to be localized towards the cell ends. Adapted from [Verma et al., 2019].

and with less or no interactions with other macrodomains [Valens et al., 2004]. *E. coli* have four different macrodomains: the ter macrodomain, which spans the region of the replication terminus, the left and the right macrodomains flanking the terminus macrodomain, and the ori macrodomain covering the region around the origin of replication (see Fig. 1.1C). The ori macrodomain is additionally flanked by flexible and non-structured regions (NS-L and NS-R), which interact with macrodomains on both sites. In addition, these macrodomains are spatially organized during the cell cycle [Verma et al., 2019]. Furthermore, evolutionary conservation of systems associated with these macrodomains suggests the functional importance of macrodomains for bacteria. The MatP/*matS* site-specific system, for instance, is responsible for the organization of the ter region into a macrodomain by binding of MatP to *matS* sites within the ter domain [Mercier et al., 2008]. Inactivation of MatP leads to a less compacted ter macrodomain and also to impaired chromosome segregation. For the ori macrodomain, there is a similar system called parABS, which is conserved in many bacterial species [Jalal and Le, 2020].

1.2 Transcriptional gene regulation in bacteria

The central dogma of molecular biology describes the flow of genetic information: DNA, which stores the information of life, is transcribed into RNA by enzymes called RNA polymerases (RNAP). The information of the RNA is then translated into polypeptides or proteins by macromolecular complexes called ribosomes. However, transcription and translation do not occur in the same ratio, instead, the flow of information can be modulated in its strength. In particular, the information of one DNA molecule can result in the formation of several thousand protein molecules or none. This modulation or regulation allows the cell to adapt to external and internal conditions and is executed by different mechanisms at different levels and stages. It can occur at the transcription level (DNA to RNA), where it is mainly determined by the ability of the RNA polymerase to bind DNA or at the translation level (RNA to protein), where it is determined either by the binding of ribosomes to the RNA (initiation) or by ribosome processivity (elongation, termination) [Browning and Busby,

2016, Tollerson and Ibba, 2020].

In bacteria, functionally related genes are often clustered into so-called operons. Those functioning units of DNA consist of several structural genes, which are transcribed into a single messenger RNA (mRNA) molecule, a RNAP binding site called promoter upstream of the genes, the 5'- untranslated region (5'-UTR), which determines translation initiation by ribosomes, as well as regulatory binding sites (activator-binding site, operator). The promoter is already a regulatory site by itself as it determines the RNAP affinity to the DNA and thereby the transcription rate of the corresponding gene. In general, the promoter begins upstream of the first transcribed base (+1) of a gene and consists of defined regions in the following order (upstream): the discriminator, -10-region, the spacer, -35-region and the UP-element. Variation in the sequence of each of these promoter regions can affect the binding of the RNAP to the DNA [Klein et al., 2021, Yan and Fong, 2017, Forquet et al., 2021].

However, as the promoter sequence is fixed information in the DNA, it can not transiently change to adapt to alternating conditions. For that, regulatory sites, which are generally located close to or even overlapping the RNAP binding site can be bound by trans-acting proteins called transcription factors (TFs), which in turn can transiently regulate the transcription rate. A classical example is the repression of the lac operon, which is responsible for metabolizing lactose in *Escherichia coli* (see Fig. 1.2). Simply put, the repressor protein LacI binds at different operator sites within the lac operon in the absence of lactose [Matthews and Nichols, 1998]. This leads to the formation of a DNA loop, which inhibits the transcription of the lac operon by preventing the RNAP from binding to the promoter. By this simple form of regulation, the cell ensures that no unnecessary proteins for lactose metabolism are produced when lactose is not present. However, when lactose is encountered by the cell, a lactose metabolite, allolactose, binds LacI, which results in an allosteric shift of the protein and consequently in the inability of the protein to bind the DNA, allowing the RNAP to bind to the promoter. Additionally, the lac operon is also regulated by direct activation. As glucose is the preferred carbon source for *E. coli*, the lac operon is further regulated to ensure that lactose is only metabolized when no glucose is present. The RNAP binding affinity to the lac promoter is weak and needs the activator cAMP receptor protein (CRP) bound to the signal molecule cAMP (CRP-cAMP) to enhance transcription [Malan et al., 1984]. The concentration of cAMP is inversely proportional to that of glucose in the cell, resulting in increased binding of cAMP to CRP when glucose is absent, which

in turn activates the transcription of the lac operon [Notley-McRobb et al., 1997]. While TFs like LacI are specific for the regulation of one operon or gene, others like CRP regulate several hundreds of genes and constitute a global mechanism of gene regulation in bacteria. For instance, CRP-cAMP is also involved in the regulation of other energy-related metabolic pathways, like galactose metabolism, arabinose metabolism, citrate metabolism and the phosphoenolpyruvate group translocation system in *E. coli* [Taniguchi et al., 1979, Ogden et al., 1980, Saier, 1998]. Similar to the regulation of the lac operon, CRP-cAMP ensures the activation of secondary carbon source metabolisms only when a preferred carbon source like glucose is not present (catabolite repression). Two further important global TFs are, for instance, ArcA, which represses a variety of genes involved in aerobic metabolism under anaerobic conditions, and OmpR, which plays a central regulatory role in acid and osmotic stress responses [Park et al., 2013, Chakraborty and Kenney, 2018].

Another kind of global transcriptional regulation is the RNAP-centred regulation, where factors interact directly with the RNAP to influence its activity at different promoters [Browning and Busby, 2016]. The RNAP core enzymes consist of the large β- and β'-subunit, two α-subunits and the small ω-subunit. However, the RNAP core enzyme is not able to bind specific promoters on its own. For this, RNAPs need to bind specific proteins known as sigma factors, which are associated with individual promoters and thereby guide the RNAP to these promoters. All bacteria have at least one essential sigma factor, which is responsible for genes related to housekeeping functions, but many bacteria also have alternative sigma factors for different promoters related to different functions [Gruber and Gross, 2003]. Available results suggest that sigma factors compete for a limited amount of RNAP core enzyme [Ishihama, 2000, Ramnaresh Gupta and Chatterji, 2016, Farewell et al., 1998]. The housekeeping sigma factor σ^{70} (RpoD) from *E. coli* directs transcription of genes essential for growth and is the dominant sigma factor during exponential phase [Sharma and Chatterji, 2010]. While the alternative sigma factor σ^{38} (RpoS), which is responsible for stress- and stationary phase-related gene transcription, is low concentrated during the exponential phase, its concentration increases during the transition to the stationary phase [Schellhorn, 2020]. Even though σ^{38} has a lower affinity to the RNAP core enzyme than σ^{70}, it competes with σ^{70} for available RNAP core enzymes during stationary phase and stress-related conditions with additional support from different factors [Maeda et al., 2000]. The TF Crl binds directly to σ^{38}, which favors its association with

the RNAP core enzyme and thereby increases its activity [Cavaliere and Norel, 2016]. Furthermore, the stringent response alarmone guanosine tetraphosphate (ppGpp), whose concentration is increased during the stationary phase, reduces the ability of σ^{70} to compete with σ^{38} for core binding [Jishage et al., 2002].

This results in the recruiting of RNAP to more σ^{38} specific promoters and therefore in the expression of stationary phase and stress-related genes.
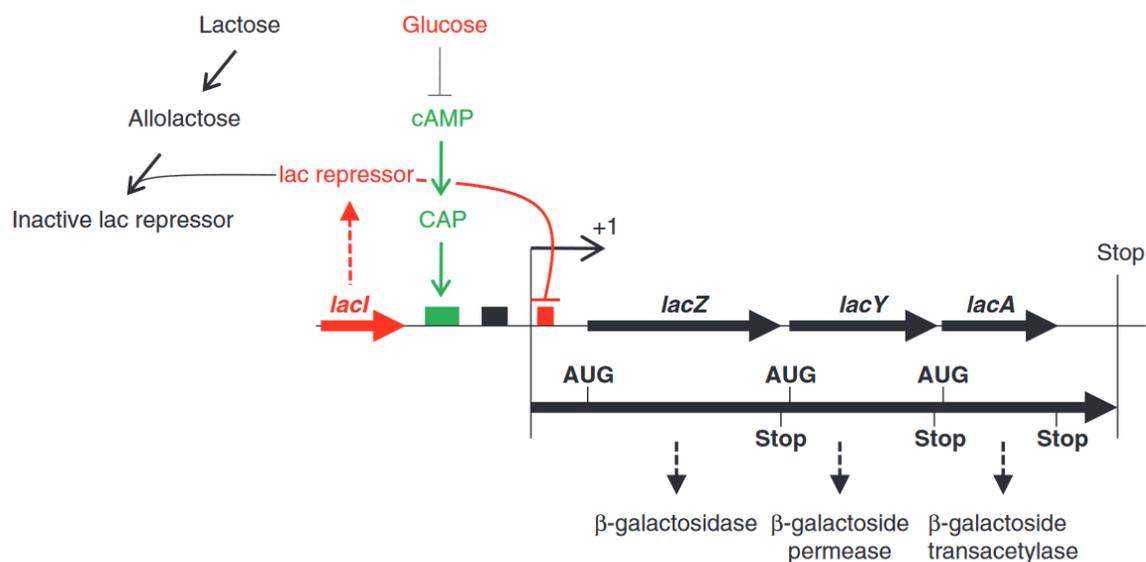
Figure 1.2: Regulation of the lac opeorn. Shown is the lac operon of *E. coli* consisting of an activator-binding site (green), a promoter (black), an operator (red) and the three structural genes (*lacZ, lacY, lacA*). Lactose abundance leads to the inhibition of the repressor LacI. The absence of glucose correlates with an increased cAMP concentration in the cell, which, in turn, binds together with the cAMP receptor protein (CRP or CAP) at the activator-binding site. Both mechanisms then facilitate the transcription of the *lac* genes. Taken from [Georis, 2013].

1.2.1 Global gene regulation by DNA supercoiling and nucleoid associated Proteins (NAPs)

While DNA supercoiling and NAPs are two of the major players in nucleoid formation, they can also have a direct influence on gene expression. For the model organism *E. coli*, the main actors in DNA supercoiling homeostasis are the antagonists DNA

Gyrase (Type II topoisomerase) and Topoisomerase I [Dorman, 2019], which introduce negative and positive supercoils into the DNA, respectively. These enzymes modulate tension in the DNA molecule and directly affect transcription initiation and termination [Hatfield and Benham, 2002]. Activity and abundance of the antagonists are tightly regulated and change upon transition between growth phases [Balke and Gralla, 1987]. The two NAPs, Fis and H-NS, are not only responsible for modulating the nucleoid structure by looping and bridging DNA, respectively, but also for modulating DNA supercoiling [Skoko et al., 2006, Lim et al., 2014]. Fis forms a concentration gradient of being higher concentrated during exponential phase compared to stationary phase [Ball et al., 1992]. As Fis also reduces the activity of DNA gyrase, it thereby counteracts the predominant negative DNA supercoiling during the early exponential phase [Muskhelishvili and Travers, 2003]. H-NS can modulate DNA supercoiling by its two distinct binding modes, stiffening and bridging [Lim et al., 2014]. The stiffening mode is caused by nucleoprotein filament formation and can suppress DNA plectoneme formation. The bridging mode instead can promote DNA plectoneme formation.

Topological changes by DNA supercoiling can affect the accessibility of the RNAP to the DNA and thus alter gene expression. About 50% of all *E. coli* genes, for example, are sensitive to DNA supercoiling [Blot et al., 2006]. Therefore, a global regulatory mechanism for genes is realised through the modulation and regulation of DNA supercoiling. Moreover, DNA topology can be altered locally by transcription activity in the neighbourhood of promoters following the Liu Wang Model [Liu and Wang, 1987, Riebet and Raibaud, 1991, Chen and Lilley, 1999]. Through unwinding and subsequent RNAP progression during transcription, positive supercoils are generated ahead of the polymerase and negative supercoils behind it. Consequently, orientation and activity of neighbouring genes and sensitivity of the affected promoter form another layer of locally organised regulation [Sobetzko, 2016, El Houdaigui et al., 2019].

NAPs also constitute a global regulatory mechanism as most of them have a direct influence as transcription factors on the expression of many genes by protein-DNA interactions. H-NS for example regulates gene expression of almost 200 genes in *E. coli* and acts as a repressor for horizontally acquired pathogenicity islands in enteric bacteria by silencing genes through restricting RNA-Polymerase from binding to DNA [Santos-Zavaleta et al., 2019, Lucchini et al., 2006]. As mentioned before, Fis concentration is higher during the exponential phase compared to the stationary phase in

E. coli. Fis regulates over 200 genes and it is assumed that it serves as an early signal of a nutritional upshift [Santos-Zavaleta et al., 2019, Ball et al., 1992]. For the integration host factor (IHF), which regulates almost 500 genes in *E. coli*, it has been suggested that its regulatory mechanism is facilitated by its ability to bend DNA and thereby bringing regulatory proteins and RNA-Polymerase into proximity [Santos-Zavaleta et al., 2019, Santero et al., 1992].

In summary, DNA supercoiling has a major impact on bacterial gene expression and its interaction with NAPs represents a dynamic regulatory system in bacteria.

1.2.2 Chromosome replication as a global regulator

In contrast to promoter regulation or transcriptional regulation, gene expression can also be increased by adding more copies of a gene. If cellular growth is restricted due to insufficient gene expression, another copy of the gene can provide an evolutionary advantage. The seven ribosomal RNA (rRNA) copies in the genome of *E. coli*, for instance, are considered to have evolved to maintain the high demand of ribosomes during fast growth [Jinks-Robertson and Nomura, 1987]. Even though not all seven rRNA operons are essential for fast growth, they are necessary for optimal adaptation to changing physiological conditions and thereby provide evolutionary advantage [Condon et al., 1995].

In contrast to stable gene duplication, bacterial cells are also able to transiently change gene copies during exponential growth caused by chromosomal replication. Bacterial chromosomes are usually replicated bidirectionally, starting at the origin of replication (oriC) and ending in the terminus region opposite the oriC. During the progression of the replication forks, genes exist either in one copy in front of the replication apparatus or in two copies behind it. Consequently, genes in oriC proximity are copied earlier than genes close to the terminus and therefore have a higher copy number for a longer time period.

A direct link between replication-induced copy numbers and gene expression was first shown by Masters and Pardee, who discovered that the synthesis of enzymes

under inducible conditions doubles when the number of copies of the corresponding gene also doubles due to replication [Masters and Pardee, 1965]. Chandler and Pritchard then showed that the output of genes, which are not promoter regulated, is proportional to the average gene copy number per unit mass [Chandler and Pritchard, 1975]. Furthermore, expression of the *lacZ* gene inserted at positions in oriC proximity revealed increased accumulation of β-galactosidase *in-vivo* compared to ter-proximal locations due to the increase in copy numbers [Sousa et al., 1997].
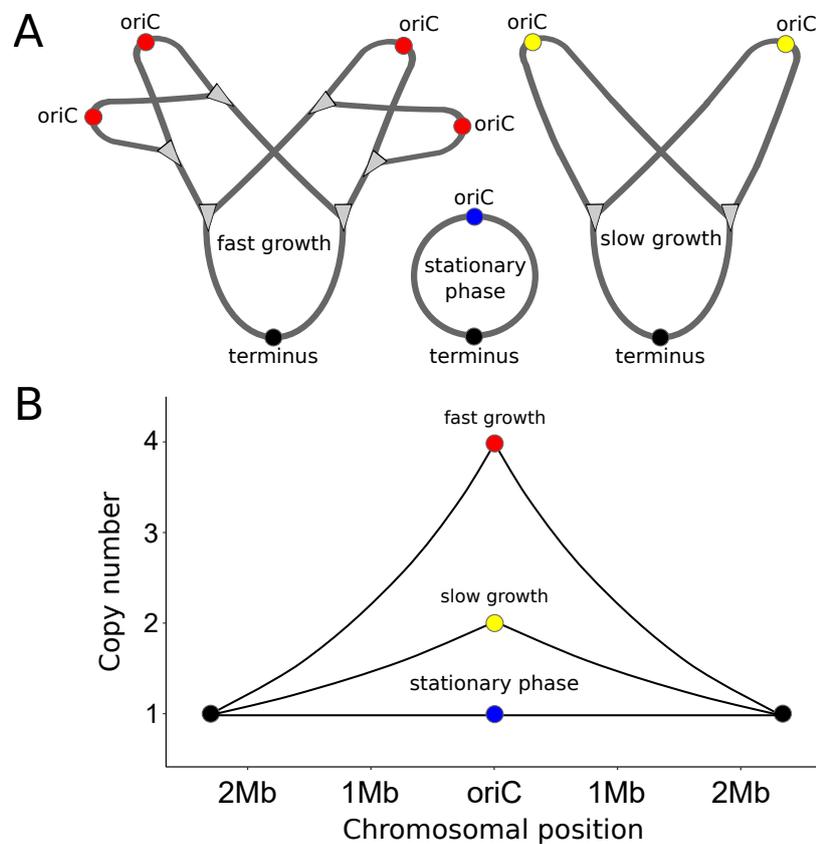
Figure 1.3: Replication-induced increase in chromosomal copy numbers. **A** Scheme of the bacterial chromosome and its replication during fast-growth, slow-growth and the stationary phase. Grey triangles represent the replication forks. Colored dots indicate the oriC, while black dots indicate the terminus of replication. Adapted from chapter 3.1. **B** Shown are hypothetical averaged copy numbers of the different growth stages of a bacterial population with respect to chromosomal position.

Considering the maximal velocity of the DNA polymerase of about 1000 base pairs (bp) per second during fast growth, the expected elongation phase in which the chromosome is fully replicated extends the doubling time of some bacteria [Huang and Ito, 1999, Egan et al., 2004, Fijalkowska et al., 2012]. *E. coli* cells, for instance, can divide every 20 minutes in rich medium, the expected time for a full replication of its chromosome, however, is around 40 minutes. This discrepancy is solved by multiple replication initiations during the cell cycle occurring at 2, 4, or 8 origins, depending on the growth rate (see Fig. 1.3A) [Skarstad and Katayama, 2013]. Therefore, overlapping replication can lead to multiple gene copies, each increasing the transcriptional output of those genes. This also leads to a gene copy gradient along the oriC-ter axis during the cell cycle where genes closer to the oriC have a higher copy number than genes closer to the terminus [Gowrishankar, 2015].

Considering an exponentially growing bacterial population, the cells are in different stages of the cell cycle. Consequently, their replication fork progression also varies. However, since oriC DNA has a higher copy number than terminus DNA over a longer period of time, the average amount of DNA will always be the highest for oriC DNA and will gradually decrease towards the terminus region (see Fig. 1.3B) [Maduike et al., 2014, Sueoka and Yoshikawa, 1965].

1.3 Genetic Engineering in Bacteria

Genetic engineering or genome editing is the process of altering the DNA of an organism to enhance or modify its characteristics. These modifications range from single-point mutation to large-scale DNA insertions, deletions and rearrangements. Genetic modifications facilitate the production of improved and novel organisms, the investigation of genes and their regulation, and are also the foundation of synthetic biology.

The basic steps of bacterial genome editing are the introduction of exogenous template DNA in the form of plasmids, linear single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA) and the subsequent integration into the host chromosome via homologous recombination. Homologous recombination plays a major role in bacterial DNA repair and is one of the main mechanisms for incorporating foreign

DNA into the genome. This process is believed to be a driving factor in bacterial evolution by generating additional diversity and is harnessed by scientists to specifically change the genome of interest [Vos, 2009]. While native recombination systems tend to be inefficient in promoting recombination with exogenous template DNA, phage recombination systems like the RecET-system from Rac prophage or the λRED system from λ phage were shown to be highly efficient in modifying bacterial chromosomes [Zhang et al., 1998, Murphy, 2016]. The λRED system, for instance, comprises the proteins Gam, Exo and Beta (see Fig. 1.4A) [Sawitzke et al., 2007]. Gam prevents endogenous RecBCD and SbcCD exonucleases in *E. coli* from binding and digesting incoming linear DNA. Exo is a 5'-3' dsDNA-dependent exonuclease and is responsible for the formation of ssDNA used for the recombination mechanism. Beta then protects ssDNA from digestion and promotes its annealing to a complementary ssDNA target within the cell. Even though it is supposed that there is not only one distinct molecular mechanism of λRED recombination, the dominant mechanism, however, seems to occur at the replication fork, preferentially through the interaction with the lagging strand template and involves a ssDNA intermediate (see Fig. 1.4B) [Fels et al., 2020, Mosberg et al., 2010]. It is presumed, that the linear, double-stranded template DNA is first made single-stranded via Exo and then incorporated into the replication fork in an Okazaki fragment-like manner. However, the efficiency of recombination with a ssDNA intermediate dramatically decreases when the size of the insertion fragment exceeds 1kb [Maresca et al., 2010]. Therefore, for larger insertions, other mechanisms, which e.g. includes only partially resected dsDNA at the 5'-end and strand-invasion via ssDNA overlaps must be predominant (see Fig. 1.4C) [Court et al., 2002, Murphy, 2016]. It has to be mentioned that only for a handful of bacterial species the functionality of the λRED system was demonstrated. For other species, scientists have utilized other bacteria-specific phage homologous recombination systems. The genes *gp60* and *gp61* from Che9c mycobacteriophage are *E. coli* Rac prophage RecE and RecT homologues, respectively, and catalyze the deletion of different genes in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* [van Kessel and Hatfull, 2007, Fels et al., 2020]. However, there are still plenty of bacterial species for which genetic engineering is highly inefficient or not feasible.

Even though heterologous recombination systems like λRED can greatly increase the recombination and editing efficiency compared to native recombination systems, the overall efficiency to find correct clones is still low. This mainly originates from
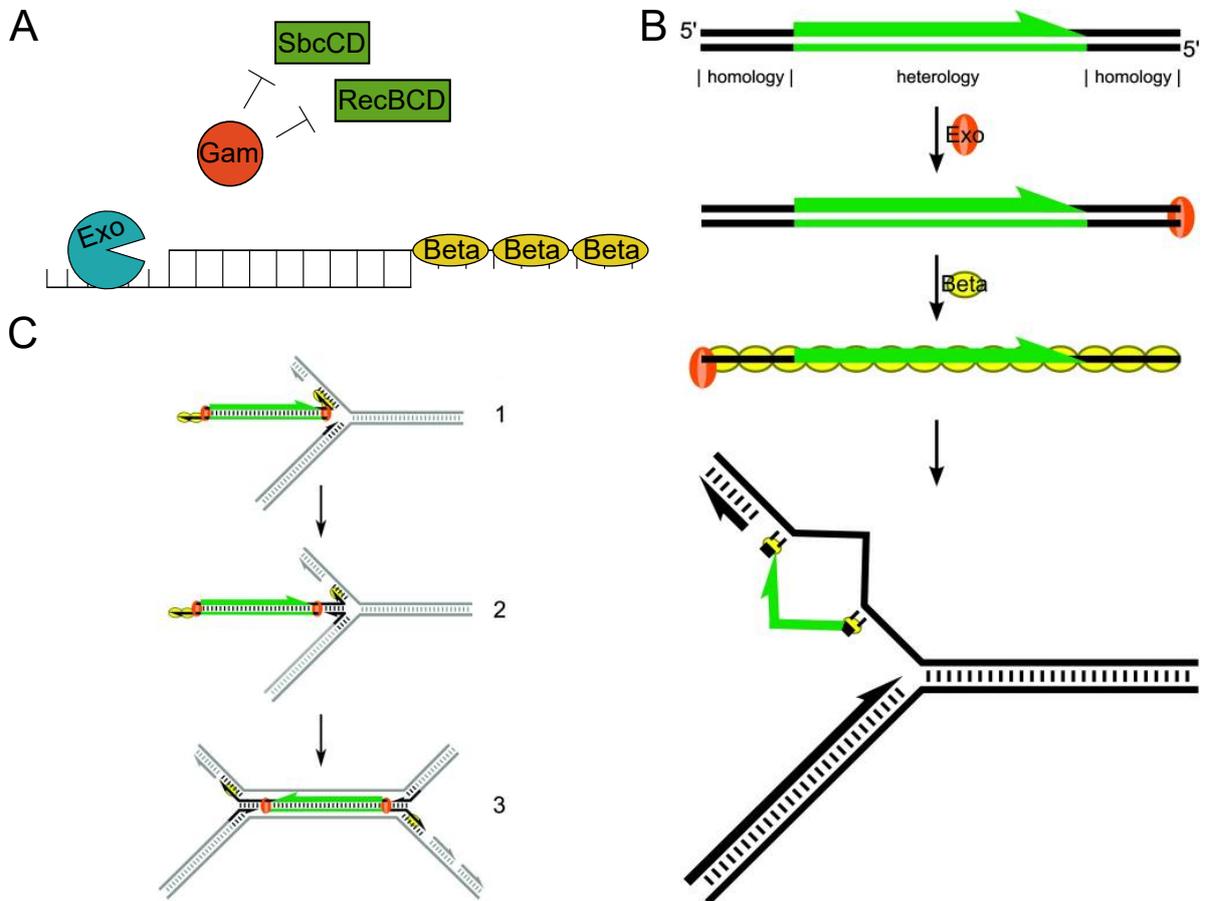
Figure 1.4: λ**RED**. **A** Functionality of λRED. Exo forms ssDNA by 5'-3' dsDNA-dependent exonuclease activity. Beta protects the ssDNA. Gam inhibits the endogenous RecBCD and SbcCD exonuclease activity. Adapted from https://blog.addgene.org. **B** Proposed mechanism for λRED recombination. A linear dsDNA template is made single-stranded and incorporated into the replication fork in an Okazaki fragment-like manner. Taken from [Mosberg et al., 2010]. **C** Previous proposed model for λRED recombination by Court et al. [Court et al., 2002]. This model does not represent the λRED recombination mechanism for small linear dsDNA templates. However, for larger templates, this or a similar model still could be the prevailing mechanism. Taken from [Mosberg et al., 2010].

different factors e.g. transformation efficiency of the template DNA, the efficiency of the recombination system and chromosome segregation, as recombination takes place at the replication fork and efficiency is therefore diluted by replication [Fels et al., 2020, Pines et al., 2015]. To overcome the low efficiency and the accompanying screening effort, several strategies have been developed e.g. the integration of selectable markers like antibiotic resistance genes or metabolic markers [Ambler and Scott, 1978, Reiss et al., 1984, STACEY and SIMSON, 1965, Fels et al., 2020]. Together with the insert, a resistance cassette or a metabolic marker gene is integrated into the chromosome. This confers an advantage over non-edited cells when cells are grown with the corresponding substrate. However, the number of possible edits in one cell is limited by the number of available markers as only one edit per marker is possible. To circumvent this problem, site-specific recombinases such as Cre- and FLP recombinase can be used [Hoess and Abremski, 1985, Cox, 1983]. By flanking the selection marker with such recombination sites, it is possible to remove the marker after integration and upon expression of the corresponding recombinase. However, this adds an extra step to the editing process and leaves active recombination sites in the genome, which can recombine with other recombination sites after another edit. Even with the use of alternative recombination sites, this limits the application of this approach [Turan et al., 2010]. In the case of generating point mutations or small insertion and deletions, oligo-mediated allelic replacement (OMAR) can be utilized. Instead of using dsDNA harbouring a selection marker, short single-stranded oligonucleotides are incorporated into the genome by an Okazaki-like allelic-replacement event at the replication fork [Ellis et al., 2001]. Due to the high efficiency of up to 25% (one-quarter of the screened clones) when the methyl-directed mismatch repair system (MMR) is inactivated, selection markers are not necessary as the screening effort is manageable [Costantino and Court, 2003]. Furthermore, automation and cyclizing of this method allow genome modifications of several genes at the same time [Wang et al., 2009]. However, OMAR is limited by the size of ssDNA oligos used and therefore insertions of larger size are not possible.

To avoid the utilization of genetic markers or heterologous DNA and obtain scarless mutations, approaches using homing endonucleases, e.g. I-SceI from *Saccharomyces cerevisiae*, can be used. Such meganucleases recognize specific DNA sequences of 12-40bp in length and introduce DNA double-strand breaks (DSBs) [Pósfai et al., 1999, Seligman et al., 1997]. When the cell is provided with a template harbouring

homologies flanking the recognition site, DSB-induced recombination can occur resulting in a scarless insertion or deletion while non-recombinants cannot survive due to the chromosomal DSB. Using this approach large inserts of about 6.5kb were achieved with an efficiency of 75-90% [Tas et al., 2015]. However, an additional recombination step is required as the recognition site needs to be first integrated into the chromosomal target region.

With the discovery of CRISPR/Cas9, the initial integration of a specific restriction site was no longer necessary and due to its simplicity new tools for different kinds of genetic engineering were created in recent years.

1.3.1 CRISPR/Cas9 as a genome editing tool and beyond

Between the early 1980s and the late 2000s, a new bacterial immune system was discovered called clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated (Cas) protein-system. Its mechanism is based on the recognition of invading DNA from previous infections and subsequent elimination of the foreign genetic material by Cas proteins, which are guided by specific RNA molecules. The ability of CRISPR/Cas to recognize and cut specific DNA sequences has made it possible to develop a genome editing tool that surpasses previous tools in its simplicity and efficiency. The type II system CRISPR/Cas9 from *Streptococcus pyogenes* can be programmed with a single chimera RNA molecule to cleave specific DNA sites in the genome of an organism [Jinek et al., 2012]. Due to its simplicity, CRISPR/Cas9 became the most widely used genome editing tool in eukaryotes and bacteria. It requires only two components, the Cas9 protein and a single-guide RNA (sgRNA), which are usually introduced into the cell as expression cassettes on a vector. The Cas9 protein consists of two nuclease domains, HNH and RuvC, which cleave the target and non-target DNA strand, respectively and thus producing a DNA double-strand break (see Fig. 1.5) [Anders et al., 2014]. The Cas9 is guided by the sgRNA, which consists of the guide sequence at the 5'-end and a backbone. The guide sequence is generally 20bp long and complementary to the target strand. It is the only variable part of the CRISPR/Cas9 system and should be carefully designed by the experimenter. It has been shown that different guide sequences can have different activity efficiencies, thus

several prediction tools for highly efficient sgRNAs have been developed [Xiang et al., 2021, Guo et al., 2018]. Additionally, it has been shown that mismatches in the guide sequence can be potentially tolerated near the 5'-end, while mismatches closer to the 3'-end (7-12bp) are less tolerated [Lu et al., 2019, Cong et al., 2013]. This can increase the risk of CRISPR/Cas9-induced DSBs at unwanted sites and it is recommended that such off-target sites in the genome are considered during the design of the sgRNA. Therefore, the prediction tools often output rated sgRNAs based on possible binding sites and corresponding off-target sites, GC-content and self-complementarity [Ding et al., 2020]. The sgRNA backbone forms a specific secondary structure and binds the Cas9 protein, which leads to conformational changes in the protein activating the nuclease domains [Xu et al., 2017, Nishimasu et al., 2014]. For the recognition of the target site, a PAM sequence (NGG) directly downstream of the target site is necessary. When bound, the Cas9 nuclease domains cleave the DNA 3bp upstream of the PAM and produce blunt ends [Jinek et al., 2012]. Depending on the organism, the resulting DSB can either be repaired by the host via non-homologous end joining (NHEJ) or by homologous recombination if a homologous repair template is available. As NHEJ does not rely on sequence homology, it is an error-prone repair mechanism and can lead to small insertion and deletions at the Cas9 cut site. Therefore, this repair mechanism can be harnessed to produce e.g. gene knockouts by unspecific mutations. While in eukaryotes, NHEJ plays a major role in the repair of DSBs [Lieber, 2008, Gorbunova and Levy, 1997], it is absent in most bacteria and homologous recombination is the dominant repair mechanism [Hiom, 2009, Ayora et al., 2011]. If a DNA template is present, homologous recombination facilitates the insertion of the template into the chromosome and consequently enables a precise edit of the target site. As no non-homologous repair mechanisms are usually present in bacteria, CRISPR/Cas9-induced DSBs lead to cell death of unedited cells. Therefore, it serves as an efficient tool for counter-selection without the need to insert antibiotic resistance markers into the genome.

In the classical bacterial CRISPR/Cas9 gene editing approach, a phage-derived recombination system, e.g. λRED or RecET, is first induced. Then, the cell is transformed with a linear DNA template and a CRISPR/Cas9 plasmid carrying the Cas9 gene and the sgRNA for counter-selection of non-edited cells. This approach has been successfully tested in various bacteria such as *E. coli*, *Pseudomonas putida*, *Enterobacter aerogenes*, *Klebsiella pneumoniae* and *Lactobacillus reuteri* [Arroyo-Olarte et al., 2021, Jiang

et al., 2013]. Based on that, Reisch and Prather modified the system for better control and to avoid putative inefficiencies due to co-transformation of plasmid and template DNA [Reisch and Prather, 2015]. By placing the Cas9 expression under the control of the inducible tetracycline promoter (P_{tet}) and adding a *ssrA* degradation tag at the Cas9 C-Terminus, they were able to stable maintain the CRISPR/Cas9 plasmids after transformation. This opened up the possibility of iterative genome editing and accelerated the genome editing process to as little as three days. Furthermore, tight control of the CRISPR/Cas9 system is also important for the application in other bacterial species as overexpression of the Cas9 protein can have cytotoxic effects [Jiang et al., 2017, Cho et al., 2018].

Based on the CRISPR/Cas9 principle, many CRISPR/Cas9 technologies have been developed over the years. Cas9 nickases (Cas9n), for instance, can reduce potential off-target effects. By using two mutated copies of the *cas9* gene, each harbouring either a RuvC or HNH nuclease domain knockout mutation, it was possible to introduce two adjacent DNA nicks with two different sgRNAs resulting in a DSB and a greatly improved specificity in gene targeting [Ran et al., 2013a].

Knock-out mutations of both nuclease domains within one copy of the *cas9* gene result in a defective Cas9 protein with an inactivated nuclease activity (dCas9). This dCas9 can be used to inhibit gene transcription supported by sgRNA, which is called CRISPR interference technology (CRISPRi) [Larson et al., 2013]. The inactivated dCas9 is guided by the sgRNA to the target site and acts as a "roadblock" for RNA-polymerase by restricting it from either binding to the promoter or by preventing it from sliding on the DNA to inhibit transcriptional extension [Zhao et al., 2020]. In contrast to transcriptional repression, CRISPR activation (CRISPRa) uses a transcriptional activation domain fused to the dCas9, which then can activate the transcription of specific genes [Cheng et al., 2013]. Additionally, a fluorescence protein can also be fused to the dCas9 protein and enables fluorescent localization of specific sites in the genome [Chen et al., 2013]. One of the recent advances in CRISPR/Cas9 genome editing are base editors. Fusing cytidine-deaminases or adenine-deaminase to dCas9 or Cas9n facilitates nucleotide exchanges (C to T; A to G) without the need for recombination machinery or introducing DSB [Arroyo-Olarte et al., 2021].

These new CRISPR applications are complemented by the discovery and implementation of other CRISPR systems. For instance, Cas12a (Cpf1) from *Francisella novicida* has been characterized as a natural double-nicking CRISPR nuclease, which is smaller

than the Cas9 from *S. pyogenes* and recognizes a T-rich PAM [Zetsche et al., 2015]. Furthermore, a Cas9 variant has been generated (xCas9), which can recognize a new set of PAMs (NG, GAA, and GAT) and therefore extend the range of possible target sites [Hu et al., 2018].
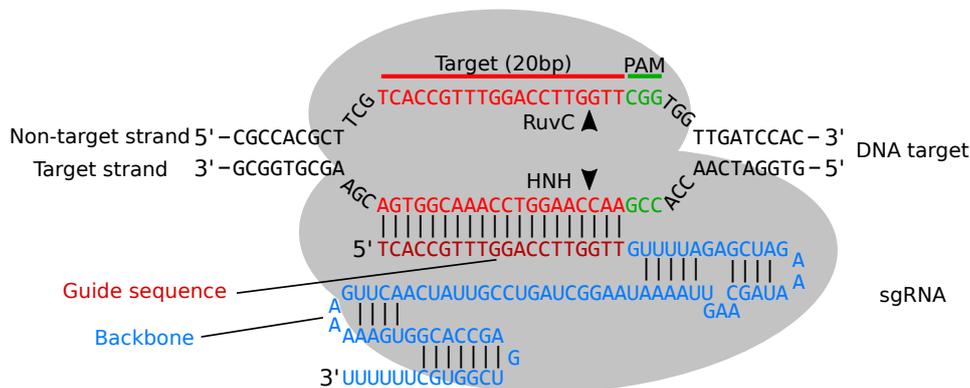

Figure 1.5: DNA binding of the Cas9-sgRNA complex. Shown is the binding of the Cas9-sgRNA complex to the target DNA (red). The sgRNA consists of a backbone (blue) responsible for Cas9 binding and a guide sequence (dark red) essential for DNA binding. The guide sequence binds the complementary target strand directly upstream of the complementary PAM sequence (NGG) (green). The Cas9 protein then produces DNA blunt ends 3bp upstream of the PAM sequence by cleaving the target and non-target strands via the nuclease domains HNH and RuvC, respectively. Adapted from [Ran et al., 2013b].

1.3.2 Modular cloning

Modular cloning (MoClo) is a relatively new method for DNA assembly and cloning [Weber et al., 2011]. It is based on Golden-Gate cloning, which uses type IIs restriction enzymes [Engler et al., 2008]. While for the classical type II restriction enzymes (e.g. EcoRI, BamHI) the cut-site lies within the recognition site, type IIs restriction enzymes cut outside of their recognition site (see Fig. 1.6A). This gives Golden-Gate cloning advantages over classical cloning strategies in terms of the maximum amount of fragments and the combination of cloning steps. With the proper design of cleavage sites and the resulting overhangs, two DNA fragments can be cloned into a new product lacking the original restriction sites. This allows the combination of

the digestion and ligation step as well as the use of several fragments at the same time in a one-pot, one-step reaction. In contrast to classical cloning, the ligation products are enriched during the reaction, as no restriction occurs after ligation. This greatly increases the cloning efficiency and thereby the possible amount of fragments utilized in one cloning step. With Golden-Gate cloning, 5-10 fragments can usually be assembled simultaneously, however, also up to 35 fragments are possible when a design tool for the assembly reaction is used [Pryor et al., 2020]. Based on Golden-Gate cloning, MoClo utilizes standardized parts to allow fast and reliable assembly of complex devices, such as functional genetic circuits comprising several genes [Weber et al., 2011]. The standardized parts are cloning vectors of different levels (0,1,2). Beginning on level 0, basic parts such as promoters, UTRs/RBS, coding sequences and terminators can be amplified via PCR and cloned into a level 0 destination vector using the type IIs restriction enzyme BbsI. The restriction sites can be introduced by primers flanking the corresponding DNA. Besides BbsI sites, which are removed upon level 0 cloning, the destination vectors additionally harbour a second type IIs restriction cut site (BsaI). During another restriction-ligation step, all level 0 parts can be combined into a functional level 1 transcription unit in the background of a level 1 destination vector (see Fig. 1.6B). This is feasible due to the standardized overhangs of each destination vector upon restriction with BsaI. Therefore, it is possible to create libraries of different level 0 parts and combine them individually to, in turn, create different level 1 transcription units (e.g. testing different promoter and ribosomal binding site combinations for the gene of interest). If necessary, up to six level 1 modules can be combined and cloned into a level 2 destination vector using again BbsI. The standardization and modular design of the MoClo system allow the recycling of previously validated genetic elements in different applications as well as the free exchange of parts between different users. It can help simplify the planning of cloning strategies and minimize the number of cloning steps required to obtain the desired construct. Such Moclo systems have been constructed for different organisms like plants, bacteria and yeast, providing a new set of standardized parts and partly increasing the flexibility of the original MoClo system [Engler et al., 2014, Klein et al., 2019, Stukenberg et al., 2021, Moore et al., 2016, Pérez-González et al., 2017].
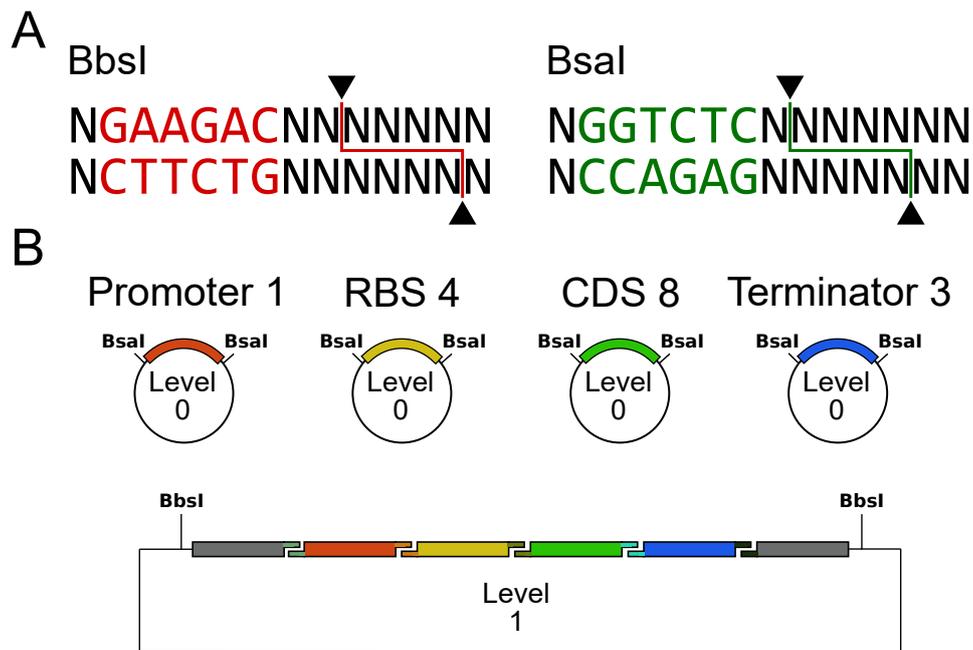
Figure 1.6: Modular cloning. A Shown are the recognition sites (red/green characters) and cut sites (triangle) of the two type IIs restriction enzymes BsaI and BbsI. **B** Principal of modular cloning. A library of basic modules like promoters, UTR/RBS, CDS and terminators can be built by cloning each part into a level 0 entry vector. From this library, different basic modules can be combined into a transcription unit using BsaI and DNA ligase in a one-pot, one-step reaction. This is facilitated by fixed overhangs upon restriction with the enzyme. The newly combined transcription unit within the level 1 vector can be further combined with other transcription units using BbsI.

2 Aim of the study

While transcriptional gene regulation in bacteria has been extensively studied in recent decades, the chromosomal organization of genes and its influence on gene expression is often not considered.

The present work aims to contribute to the understanding of chromosomal organization and its influence on bacterial gene regulation, especially in terms of the replication-induced copy number effect. This is not only important for the understanding of bacterial chromosome evolution, but also for modifying the genomic landscape via genetic engineering.

The first objective of this study is the investigation of the *Escherichia coli* gene expression gradient along the genome when comparing the exponential with the stationary phase and how it is influenced by the replication-induced copy number effect. Besides the copy number effect, another factor, which could influence this pattern, is the spatial organization of transcription factor binding sites along the chromosome. Since both factors act in the same direction, the individual influence of the two factors on the gene expression pattern cannot be readily determined. This requires changes in one of these factors, preferably the copy number effect, as this one is easier to change than the genetic regulatory network.

Therefore, the second objective is to develop a tool, which would facilitate the required chromosomal changes. This genome editing tool should be based on CRISPR/Cas9 and, beyond its original purpose, enable major chromosomal rearrangements and gene exchange between bacterial species to meet the desired need for such tools in research today.

The third objective is to examine the influence of the replication-induced copy number effect on individual genes in *Escherichia coli*. For this purpose, the copy number effect should first be extracted from the transcription data of two different growth phases

and these data should then be compared with each other.

The final objective is to investigate the influence of the replication-induced copy number effect on the evolutionary development of bacterial chromosomes. For that, the conservation of predominantly copy number regulated and predominantly promoter regulated genes as well as exponential genes of *Escherichia coli* should be determined in two other bacterial species. In addition, the conservation of genes between closely related, but in their growth rate different bacterial species should be analysed. This analysis should be extended to the entire bacterial kingdom.

3 Results

The results of this doctoral thesis are divided into two chapters. Each chapter contains an individual scientific study, which is either published (Chapter 3.2) or submitted (Chapter 3.1) for publication in a scientific journal. The scientific studies are introduced with a summary and a clarification about the contribution of each author.

3.1 The Role of Replication-induced Chromosomal Copy Numbers in Spatio-temporal Gene Regulation and Evolutionary Chromosome Plasticity

Bacteria are constantly exposed to changes in their environment, to which they respond with changes in their gene expression. For most bacteria, promoter regulation (promoter recognition and transcription initiation/elongation) is the key regulator step that modulates gene expression and has been extensively studied over the last decades. While transcription factors can modulate the expression of single operons, they can also orchestrate the regulation of hundreds of genes and thus constitute a global mechanism of gene regulation. Another global regulatory mechanism is the chromosomal copy number effect, which describes the transient increase in gene expression due to a replication-induced increase of gene copies during cell growth. Especially in fast-growing bacteria, replication constitutes a spatio-temporal gradient in gene copies along the oriC-ter axis.

In this study, the gene expression profile of the exponential and stationary phase of *E. coli* is investigated. It was previously shown that during the exponential phase, oriC-proximal genes are higher expressed than genes closer to the terminus compared to the stationary phase. This expression gradient may either emerged from promoter regulation by strategic positioning of regulator binding sites that correspond to the expression gradient or by the replication-induced copy number effects, which also exhibit a similar oriC-ter gradient. To decouple both effects, an *E. coli* strain (INV) was constructed, which has reduced chromosomal copy numbers during the exponential phase compared to the wild type (WT). By comparing the WT with the INV strain, it was demonstrated that the observed expression pattern is caused by the replication-induced copy number effect and not by promoter regulation.

Furthermore, the regulatory impact of the copy number effect was investigated compared to the promoter regulation. It was shown that about 40% of the *E. coli* genes are predominantly copy number regulated when comparing the exponential with the stationary phase. It was found that genes, which are part of the coenzyme metabolism are strongly overrepresented in the fraction of predominantly copy number regu-

lated genes. This fits with previous studies, where a correlation between coenzyme metabolism and growth was shown, but the regulation was unknown. Furthermore, the 'amino acid metabolism' and 'Trafficking/secretion' categories are also overrepresented, indicating a regulation of the corresponding pathways via the copy number effect.

In two other bacterial species (*V. natriegens* and *D. dadantii*), orthologous genes of predominantly copy number regulated genes of *E. coli* show less variation in oriC distance than promoter regulated genes. Additionally, exponential genes from *E. coli* were analysed in *V. natriegens*, a fast-growing bacterium with a steep copy number gradient and *D. dadantii*, a slow-growing bacterium with a flatter gradient. It was shown, that these exponential genes are located closer to the replication origin in *V. natriegens* than in *D. dadantii*. An additional extension of the analysis to various clades in the bacterial kingdom revealed a connection between the growth rate of a bacterium and the positional conservation along the oriC-ter axis. These findings indicate an evolutionary importance of the copy number effect as a global regulatory system.

Patrick Sobetzko, Marc Teufel and Werner Henkel conceived the experimental design. Marc Teufel conducted the experiments. Next-Generation Sequencing data analysis of *E. coli* experiments was performed by Marc Teufel in consultation with Patrick Sobetzko. Analyses of different bacterial species were conducted by Patrick Sobetzko and Marc Teufel. The manuscript was written by Marc Teufel and Patrick Sobetzko.

The Role of Replication-induced Chromosomal Copy Numbers in Spatio-temporal Gene Regulation and Evolutionary Chromosome Plasticity

Marc Teufel[1], Werner Henkel[2], and Patrick Sobetzko[1,*]

[1]Philipps Universität Marburg, Synthetic Microbiology Center Marburg (SYNMIKRO), Marburg, D-35043, Germany
[2]Jacobs University Bremen, Transmission Systems Group, Campus Ring 1, Bremen, D-28759, Germany
[*]patrick.sobetzko@synmikro.uni-marburg.de

ABSTRACT

For a coherent response to environmental changes, bacterial evolution has formed a complex transcriptional regulatory system comprising classical DNA binding proteins sigma factors and modulation of DNA topology. In this study, we investigate replication-induced gene copy numbers - a regulatory concept that is unlike the others not based on modulation of promoter activity but on replication dynamics. We show that a large fraction of genes are predominantly affected by transient copy numbers and identify cellular functions and central pathways governed by this mechanism in *Escherichia coli*. Furthermore, we show quantitatively that the previously observed spatio-temporal expression pattern between different growth phases mainly emerges from transient chromosomal copy numbers. We extend the analysis to the plant pathogen *Dickeya dadantii* and the biotechnologically relevant organism *Vibrio natriegens*. The analysis reveals a connection between growth phase dependent gene expression and evolutionary gene migration in these species. A further extension to the bacterial kingdom shows that chromosome evolution is governed by growth rate related transient copy numbers.

Introduction

Bacteria interact dynamically with the environment and adapt to external and internal conditions. The first level of adaption is the regulation of gene expression to integrate various signals in a concerted manner. Global regulators of gene expression are DNA binding proteins comprising abundant nucleoid associated proteins with hundreds of target genes and a plethora of regulators with few targets[1]. The actions of these regulators form the transcriptional regulatory network (TRN) of a bacterial cell. It is capable of transmitting information between different parts of the nucleoid, especially between compact macrodomains[2,3]. Hence, this regulatory concept connects spatially distant or isolated chromosomal regions.

DNA topology is another major regulator. Here, the 3D structure of the DNA and tension within the molecule is converted in more or less favourable conditions for RNAP and regulator binding. The main actors are the antagonists DNA Gyrase and Topoisomerase I[4]. These enzymes remove or add helical turns to the DNA and thereby modulate tension in the DNA molecule. Activity and abundance of the antagonists are tightly regulated and change upon transition between growth phases[5]. Through modulation of activity and abundance of DNA Gyrase and Topoisomerase I, DNA supercoiling levels are controlled realising a global regulation. Moreover, DNA topology can be altered locally by transcription activity in the neighborhood of promoters following the Liu Wang Model[6–8]. Consequently, orientation and activity of neighboring genes and sensitivity of the affected promoter form another layer of locally organised regulation[9,10]. In summary, the regulatory mechanisms act alone or in combination on promoter activity and are, therefore, subsumed under promoter regulation in this manuscript.

In contrast to this strategy, gene expression can be increased by adding more copies of a gene. This evolutionary strategy can be observed for highly transcribed genes like stable RNA operons, where promoter regulatory optimisation is exhausted[11,12] or fast adaptation to new environments is required[13]. Gene duplication does not alter individual promoter regulation unless titration of regulators to the increased number of binding sites is involved. As the majority of regulatory sites are covered by abundant proteins that bind hundreds of sites, a few additional sites usually have no relevant effect on the binding site to regulator ratio. The increase in expression by adding copies of a gene takes place at an evolutionary time scale.

However, there is also a mechanism for transient changes in copy numbers within the life cycle of a bacterial cell. During DNA replication, genes are either present in one copy in front of the replication apparatus or in two copies after replication of its locus. Hence, the closer a gene is to the origin of replication (oriC) the earlier it is copied. Consequently, it produces double the amount of RNA for a longer time period within the cell cycle than a gene located close to the terminus. Even with a maximum velocity of about 1000 bp/s for fast replicating bacteria[14,15], the expected C period may extend beyond the doubling time (40

min vs 20 min for *E. coli*) for fast growing bacteria. To overcome this limitation, fast growing bacteria turn to overlapping replication rounds, where new rounds of replication are initiated before the template DNA is fully replicated. This can increase gene copies up to 8 copies in *E. coli* in the oriC proximal region in comparison to the terminus region[16, 17]. Furthermore, this copy number effect is linked to specific growth conditions of the cell. Under rich nutrient conditions, the copy number effect is maximal, whereas under conditions of starvation or stress no replication is initiated and locus copies are uniform along the chromosome[18]. Earlier studies identified a link between gene expression of individual genes and their copy number[19–22]. A systematic analysis of the impact of copy numbers on gene expression, functional regulation and its impact on chromosome plasticity has not been performed yet. In 2013, we identified a gradient of activated genes following the oriC-ter axis[23]. This gradient covers the full chromosome and potentially comprises a plethora of genes. In this study, we analyse and quantify the impact of copy numbers on forming a spatio-temporal expression pattern. We also quantify its impact on gene regulation of individual genes, functional groups and pathways. Furthermore, we show how copy numbers drive the arrangement of genes during evolution depending on species growth rates.

Results

The Spatial-temporal Gene Expression Pattern between exponential and stationary phase might be explained by two different regulatory concepts

Comparing the *E. coli* expression profiles of the exponential and stationary phase revealed a higher expression of oriC-proximal genes during the exponential phase, whereas genes closer to the terminus region showed a lower expression compared to stationary phase[23] (see Fig.1A). This spatio-temporal gene expression pattern may reflect a cellular program to adapt to changing conditions. The pattern may emerge due to the strategic positioning of genes regulated by global transcription factors such as σ^{70}, σ^{38} or abundant regulatory proteins like Fis, H-NS, IHF, the cAMP receptor protein (CRP) or the leucine-responsive regulatory protein (Lrp). These factors regulate hundreds of genes and therefore may impact a global pattern. Analysis of target gene positioning of global regulators revealed a gradient of regulated genes in the σ^{70}, σ^{38}, CRP and Lrp regulons along the oriC-ter axis (see Fig. 1B,C,D and Fig. S6). Activity of these factors depend on the cellular state. While σ^{70} is the dominant transcription factor during exponential phase, σ^{38} competes with σ^{70} for RNA polymerase (RNAP) during stationary phase. Curated regulatory data of the regulonDB database revealed that genes specifically regulated by σ^{70} are more abundant in proximity to oriC (see Fig. 1B), which may contribute to the observed increase of oriC-proximal genes during exponential phase. Genes regulated specifically by σ^{38}, however, are more abundant at the terminus region (see Fig 1C), which would lead to an increase of oriC-distal gene expression in stationary phase compared to exponential phase. Furthermore, σ^{70} regulated genes reduce activity in stationary phase due to a reduced fraction of RNAPσ^{70} triggered by σ^{38} competition for RNAP. CRP and Lrp, both important regulators during starvation and stationary phase[24, 25], negatively regulate genes with a characteristic distribution gradient along the oriC-ter axis (see Fig.1D). In this case, repression of more oriC-proximal genes during stationary phase would also contribute to the observed increase in gene expression of oriC-proximal genes during exponential phase. In general, a combination of different regulatory proteins with non-uniform distribution of target genes along the oriC-ter axis might be the source of the spatial gene expression pattern observed when comparing exponential and stationary phase. These patterns can be supported by DNA supercoiling sensitivity of promoters mediated by DNA structure and regulatory proteins. All mentioned factors act on promoter activity and are subsumed under promoter regulation.

Besides promoter regulation, which differs in its activity regarding growth phases, replication activity is another potential regulatory factor. During exponential growth, many bacteria perform multifork replication to ensure chromosome replication when the doubling time is shorter than the duration of replication (see Fig.1E). Consequently, another round of replication begins before the previous round terminates. Depending on the organism, several replication initiations can occur during the cell cycle, resulting in multiple transient gene copies in the oriC-proximal region in contrast to a single copy in the terminus region. This copy number effect leads to a higher expression of oriC-proximal genes in exponential phase compared to stationary phase, in which the copy number of each gene is one along the oriC-ter axis as no rounds of replication are initiated. Towards the terminus region, this effect is gradually reduced. Marker-Frequency-Analysis (MFA) allows to visualise and quantify the copy number effect when using whole-genome DNA sequencing of exponential growing cells (see Fig.1F). For *E. coli*, we observed a gradual decrease of reads along the oriC-ter axis representing the average copy number of the sequenced culture. However, both regulatory factors, promoter regulation and copy number effect, act in the same direction regarding increasing and decreasing oriC-proximal/distal genes during exponential and stationary phase. Therefore, it is only possible to determine each of the factors influence on gene expression by isolating a single factor.

A strain to dissect the impact of promoter regulation and copy number effects

To determine the influence of promoter regulation and the copy number effect on the gene expression pattern of exponential phase compared to stationary phase, altering only one of those factors is necessary. As changes in the genetic regulatory network would be difficult due to the diversity of regulatory proteins and DNA topology[9, 26, 27], we decided to significantly alter

the copy number effect. By relocation of the oriC into the terminus region, we would get an opposite copy number profile compared to wild type. In such a strain, during exponential phase, genes of the (wild type) terminus region would have a higher copy number than (wild type) oriC-proximal genes. This inversion of gene copy number would either result in an unaltered, a disturbed or an inverted expression profile when compared to stationary phase depending on the impact of each of the regulatory factors. However, relocating the oriC into the terminus can cause massive biological problems.

Replication in *E. coli* is bidirectional. Both replication forks move along the left and right replichore of the circular chromosome and meet in the terminus region opposite of oriC. In this region, the replication forks are trapped at specific DNA sites called ter sites, which are bound by the terminus utilization substance protein (Tus)[28]. This protein-DNA complex unidirectionally arrests DNA replication, allowing replication forks to pass ter sites only in the origin-to-terminus direction. An oriC in the terminus region would therefore lead to replication fork stalling shortly after initiation and prevent replication of the remaining chromosome. To circumvent this problem, we generated a *E. coli* MG1655 Δtus strain to abolish replication stalling at ter sites. This would then allow the replication forks to pass freely from the former wild type terminus to the oriC region.

Another problem would be head-on replication-transcription conflicts of the highly transcribed ribosomal RNA (rrn) operons and the replication forks, as the *rrn* operons are transcribed in oriC-ter direction. These head-on collisions seem to significantly delay fork progression and especially the *rrnCABE* cluster and the *rrnH* operon causes substantial problems to replication progression[29]. We therefore needed to alter the transcription direction of the *rrn* operons. The Cre-lox and FLP/FRT systems, which are based on site-specific recombinases, allow excision and inversions of chromosomal DNA flanked by two identical target sites depending on its relative orientation. By flanking *rrn* operons with facing FRT or loxP site pairs, it would be possible to invert the transcription direction and circumvent head-on replication-transcription machinery collisions, when relocating the oriC to the terminus. To minimize crosstalk between FRT/loxP sites of different inversion sites, different FRT/loxP variants were used for each pair[30][31]. As the *rrnCABE* cluster consists of four closely located ribosomal operons in oriC-proximity, we only used one pair of FRT sites to invert the whole region instead of inverting every single operon on its own. All insertions and deletions were carried out using the CRISPR SWAPnDROP system, which allows consecutive chromosomal changes based on CRISPR/Cas9 counter-selection[32].

For the relocation of the oriC into the terminus, we first replaced the native oriC with the F-plasmid origin of replication oriS flanked by a pair of tandem FRT sites to allow excision of the oriS. Furthermore, this would allow a parallel inversion of the ribosomal RNA operons together with the oriS deletion to avoid head-on collisions in intermediate strains for a sequential approach. After the replacement of oriC with oriS, we inserted the native oriC into the terminus region of the chromosome. The strain was then transformed with a plasmid harbouring the Cre recombinase and Flippase under the control of the pBAD promoter. Additionally, we used CRISPR/Cas9 to actively remove oriS DNA looped-out during excision by FLP/FRT-recombination and prevent reintegration. In summary, the final strain was able to freely invert its *rrn* operons and remove the placeholder oriS upon induction of the CRISPR/Cas9 and recombination systems to generate a strain with an inverted copy number.

After induction, streaked colonies appeared in different sizes ranging from very small to wild type-like size. OriS elimination could only be found in small and middle-size colonies. Surprisingly, *rrn* operon inversions occurred rarely and could not be found in combination with the oriS elimination. Colonies of different sizes and with oriS knockout were re-streaked for further investigation. Re-streak of the small-size colonies resulted in a mix of small and middle-sized colonies indicating instability of the strain due to frequent suppressor mutation. For stability reasons, we used one of the middle-sized colonies that originated from a re-streaked small-size colony for further investigations. MFA-analysis of the clone during exponential phase revealed an inversion spanning half of the chromosome, mainly the left replichore (see Fig.2 A-D). This inversion resulted in a relocation of the oriC from the terminus back into the wild type oriC region with a distance of about 381kb from the native oriC site. Thereby, for most right replichore genes, a wild type oriC distance configuration was restored whereas most genes of the left replichore remained inverted with respect to oriC distance. Furthermore, for the *rrnCABE* cluster and the *rrnH* operon the wild type leading strand arrangement was restored and therefore head-on collision with replication was prevented. This might have improved fitness and explains the increased colony size[29]. Additionally, this strain revealed a significantly reduced maximal copy number during exponential phase compared to wild type (see Fig.2D and Fig.1F). Consistently, the doubling time of the inversion strain (INV) with around 62 minutes is three times greater compared to the wild type (see Fig. S1A).

With its stability and the strongly reduced copy-number in exponential phase, it allows further investigation of copy number impact on the spatio-temporal expression pattern. If copy numbers have a strong impact, analysis of the exponential and stationary phase should show a reduced or abolished expression gradient along the oriC-ter axis. For the reference strain, we removed the oriC in the terminus region of the INV precursor strain, before CRISPR/Cas9 and recombinase induction. Subsequently, the oriS in the native oriC site was replaced by the native oriC resulting in a wild type-like strain (see Fig. S1A,B), referred to as wild type (WT) for the rest of the manuscript. MFA-analysis shows a replication profile similar to *E. coli* MG1655 regarding spatial copy number distribution (see Fig.2E and 1F). In analogy to the RNA-seq data analysis of *E. coli* CSH50 wild type[23], a sliding window approach was used to determine spatial biases of up/down regulation (see Material and

Methods). RNA-seq analysis of the exponential and stationary phase revealed the same spatio-temporal gene expression pattern seen in the reference study (see Fig.2F and Fig.1A). Additional comparison to *E. coli* MG1655, the direct precursor wild type strain showed a similarity in gene expression that equals the similarity of replicates of a single strain (see Fig. S7C,D). Hence, from the perspective of mRNA levels, WT cannot be distinguished from its wild type precursor *E. coli* MG1655.

Copy number is the dominant effect of the spatio-temporal expression pattern

For comparison of WT and INV transcription patterns, RNA-seq of the INV and WT strains was carried out in triplicates. In analogy to WT, we analysed the expression profile of the INV strain in the exponential (EXP) and stationary (STAT) phase. For better comparison, the expression profile was mapped against the WT chromosome. Mapping against the INV chromosome would alter the coordinate system of the chromosome with respect to WT comprising altered replichores and inverted regions. Comparison and interpretation of the data would therefore be difficult. As a consequence of the sliding window approach, windows spanning the inversion break points are not present in both data sets and are therefore excluded from the analysis (see Material and Methods and Fig.3A et seqq.). As seen in Fig.3A, the spatial expression gradient of the INV strain along the oriC-ter axis is mostly attenuated. Nevertheless, local characteristic peaks of the spatial pattern are still consistent with the wild type pattern suggesting a state of the promoter regulatory system similar to WT (see Fig. S2). This indicates that instead of promoter regulation, copy number effects may play the major role in the formation of the gradual expression pattern. More compellingly, for the left replichore, genes closer to the terminus show a higher expression in the exponential phase compared to the stationary phase. Due to the inversion, these genes are situated close to oriC$_{INV}$ in the INV strain. Hence, this expression profile reflects the still abundant influence of the reduced copy number effect in this strain.

To verify the effects of copy number and study it isolated from other regulatory effects, exponential phase of WT and INV strains were compared. Both strains differ in copy number, whereas, in the same growth phase, differences in promoter regulation are expected to be minimal. The comparison revealed a very strong expression bias gradient along the oriC-ter axis (see Fig.3B). The vast majority of oriC-proximal genes show a higher expression in the WT strain, while genes closer to the terminus are predominantly higher expressed in the INV strain. Interestingly, in the putative absence of promoter regulation, the gradient is more pronounced than between the exponential and stationary phase indicating the dominance of copy number effects in forming a gradual spatial expression pattern. Additionally, the characteristic local peaks seen in the analysis of WT EXP/STAT and INV EXP/STAT cannot be observed. This underpins the promoter regulatory origin of the local peaks between the exponential and stationary phase. Furthermore, the expression biases of both WT EXP/STAT (see Fig.2F) as well as the WT/INV EXP (see Fig.3B) follow their corresponding copy number differences (see Fig.3C,D). The expression bias even reflects the small steps in copy number differences at the inversion break points(see Fig.3B,D).

Even though the previous data suggest a major role of the copy number effect on the expression profile, the exact impact is not yet quantified. Whether other regulatory factors systematically contribute positively or negatively to the spatio-temporal expression pattern is still an open question. Multiple copies of a gene cause an increase in gene expression proportional to the number of copies. Consequently, the average expression fold change should match the corresponding copy number differences, if the copy number effect is the dominant factor. If other systemic regulatory factors influence the spatio-temporal pattern, the average expression fold change should deviate significantly from the copy number difference. As seen in Fig.3E, the average expression fold change of WT EXP/STAT corresponds well to the copy number difference (see Fig.3C). For the case of WT/INV EXP (see Fig.3D,F), where copy numbers were systematically reduced in the INV strain, the fold changes also matched the copy number differences supporting the role of copy numbers in forming spatial expression patterns. In this case, the characteristic local peaks (see Fig. S2) observed between exponential and stationary phase is flattened out, indicating a promoter regulatory source between these phases.

In certain cases, it could be important to remove copy number effects from expression data. Such cases could be mutant studies in which regulatory effects of the mutant are investigated. A growth defect often observed in regulator mutants would introduce a bias caused by copy number differences between wild type and mutant (see Fig. 5)[33]. Consequently, gene expression data and deduced regulatory interactions might be biased. We therefore tested this scenario by subtracting the copy number difference between WT and INV in exponential phase from the WT exponential phase expression data. We then compared the corrected WT exponential phase expression data with its stationary phase expression data resulting in a flat spatial expression pattern. A comparison with the INV EXP/STAT expression pattern revealed a remarkable similarity (see Fig. S4). This shows that copy number data can be used to compensate for copy number differences between samples and underpins the impact of copy numbers on forming spatial expression patterns.

Copy numbers regulate distinct cellular functions

We have shown that the copy number effect plays a major role in forming a spatio-temporal gene expression pattern between exponential and stationary phase. This may also indicate a central role in the regulation of individual genes and pathways. However, a spatial bias induced by copy numbers does not necessarily imply a major role in single gene regulation. Promoter regulation may alter gene expression several hundred-fold[34]. Regarding total fold change, the fold change of copy number can

be neglected in such cases. To estimate the impact of copy number relative to other regulatory factors, we analysed the single gene expression fold change data of WT EXP/STAT. The expression fold change of a gene is determined by its difference in copy number and promoter regulation. We have shown that on a large scale, the expression fold change follows the copy number between exponential and stationary phase. Therefore, we can remove the copy number effect (f_{copy}) of single genes from its expression fold change (f_{total}) to determine the influence of the remaining promoter regulation (f_{reg}).

$$f_{reg} = \frac{f_{total}}{f_{copy}} \qquad (1)$$

We compared the copy number influence factor and the promoter influence factor of each individual gene and divided them into two categories: predominantly copy number regulated and predominantly promoter regulated, depending on the proportion of each factor on the total fold change of gene expression. About 40% of the genes are predominantly regulated by copy number when comparing the exponential with the stationary phase (see Fig.4A). For more than three quarter (78%) of genes, its copy number still covers more than one fourth of its total fold change. This underlines the broad relevance of the copy number effect in gene expression. However, there are also genes, which have a significantly higher fraction of promoter regulation (up to 1256-fold). For those genes, the copy number effect presumably plays an inferior role in regulation.

When comparing the influence factors of WT/INV EXP, the majority of the genes (67%) are predominantly regulated by the copy number effect with a reduced influence of other regulators (see Fig.4B). In this case, about 92% of the genes have a copy number influence factor, which is greater or equal to one fourth of the influence factor of other regulators. This reflects the mild influence of promoter regulation and the copy number dominance in this experimental design. The remaining fraction of altered promoter regulation could originate from the altered expression of regulators located in the inversion region of the INV strain and its secondary effects.

The large set of genes dominantly regulated by copy numbers may indicate a concerted regulatory mechanism. In the case of WT/INV EXP, the experimental setup was rather artificial and did not follow a process in the life cycle of *E. coli*. Therefore, we focused on the WT EXP/STAT experiment to see if there is a link between regulation by copy number and specific cell functions. For genes predominantly regulated by the copy number, frequencies of functional categories were investigated. We found a significant overrepresentation of genes in the 'Coenzyme metabolism', 'Amino acid metabolism' and 'Trafficking/secretion' categories, while in 'Energy production and conversion' genes regulated by copy number were underrepresented. For coenzyme metabolism, it was shown that in *E. coli* coenzyme synthesis is directly correlated to growth[35]. *E. coli* is capable to effectively adjusting de novo coenzyme synthesis to counteract varying dilution rates during growth, but the regulation is still unknown. Hence, a direct linking to replication rounds and therefore to copy number appears plausible. Also, amino acid metabolism is involved in biomass formation and is, therefore, a plausible candidate for coupling to copy numbers. Analysis of individual metabolic pathways further supports a coherent regulation by copy numbers. Here, specific pathways were strongly enriched in copy number dominated genes e.g. the aspartate pathway (see Fig. S3). Interestingly, in this pathway, the copy number regulation of intermediate pathway steps is complemented by promoter regulation at neuralgic steps at the entry and exit points. Lists with all metabolic pathways from *E. coli* and the corresponding copy number influence, as well as raw data for fold change analysis, can be found in the supplementary data.

Regulation via transient copy numbers determines chromosomal architecture in the course of evolution

As the coenzyme and amino acid metabolism are essential metabolisms in bacteria, their genes are evolutionary conserved. With respect to their regulation by copy number, those genes may also exhibit conservation regarding the location on the chromosome. Genes, which are coupled to growth and copy number, should be located close to the oriC or at least maintain their relative position to oriC. In order to investigate the evolutionary conservation of genes predominantly regulated by copy number, we first divided all genes into different sets depending on the extent of copy number regulation (see Fig.4D). Three opposing sets were generated with increased stringency for either copy number or promoter regulation dominance. We then estimated the variation of those genes in two species (*Dickeya dadantii* and *Vibrio natriegens*) with respect to their oriC distance in *E. coli*. *D. dadantii* is part of the Enterobacterales and is the causative agent of bacterial stem and root rot affecting potatoes and other crops, while *V. natriegens* is part of the Vibrionales and of increasing biotechnological relevance. The stronger the gene regulation is dominated by the copy number, the less variation in oriC distance is observed in these two species compared to *E. coli*. In contrast, the stronger their regulation is dominated by promoter regulation, the more variation is detected. This indicates an oriC distance conservation of genes regulated by copy number and high spatial flexibility of genes governed by promoter regulation.

The two selected species flank *E. coli* with respect to doubling time during exponential growth. *D. dadantii* (approx. 100 min) has a longer doubling time than *E.coli* (approx. 20 min) whereas *V. natriegens* (approx. 10 min) exhibits a far shorter doubling time. As DNA polymerase speed is a limiting factor for fast growing bacteria, a reduced doubling time is reflected in intensified overlapping replication increasing copy numbers (see Fig. 4E). Consequently, in these species, three different levels of copy

numbers are realised allowing us to investigate the impact of copy numbers on chromosome evolution with respect to gene location. For comparison of the three species, genes predominantly expressed during exponential phase in *E. coli* (p-value < 0.05) were selected. Due to the difference in expression between the exponential and stationary phase, copy number can potentially positively regulate these genes also in other organisms. Selective pressure for copy number regulation could depend on the extent of the available copy number effect. Consequently, for faster growing species, a higher portion of genes active during exponential growth could exploit copy number effects for regulatory purposes and migrate towards oriC. The distribution of orthologs in the three species revealed a link between growth rate and stringency of gene positioning (see Fig. 4F). Orthologs in the slow growing bacterium *D. dadantii* were less focused on the oriC proximal region than orthologs in *E. coli*. Orthologs in *V. natriegens*, the fastest growing bacterium, were even more focused in the oriC proximal region than in *E. coli*.

We further investigated this observation using a larger set of species. For most species, the oriC position is not determined[36] or hidden in countless publications. Therefore, we devised a method that is based on the oriC-ter symmetry found across the bacterial kingdom[3]. Hence, the chromosomes of two species match best with respect to oriC distance if oriCs of both species are superimposed (see Fig.5A). To identify the oriC-ter axis required to determine the oriC distance conservation of a species pair, all chromosomal constellations were tested for optimal mapping by relative rotation of both chromosomes (see Fig. 5A,B). The approach requires a minimal evolutionary distance in which several gene relocations took place between species. Whether the requirement is fulfilled can be determined by constellation analysis itself. If the evolutionary distance is too close and genomes are actually identical concerning gene positions, diagonals instead of circles will form. A weak upward diagonal connecting the circles can still be seen for the comparison of *E. coli* and *D. dadantii* belonging to the same phylogenetic order (see Fig. 5B). Analysis of different phylogenetic categories identified the family category to be the lower limit for a proper analysis. Using this approach, oriC distance correlation was determined for species pairs of various phylogenetic classes including gram-positive and gram-negative bacteria. The species within these classes were selected to be different in their family membership to ensure a standardised evolutionary distance. To approximate growth rates, we used the correlation of growth rate and the number of ribosomal operons of a species. This correlation was first verified using growth rates of Couturier and Rocha 2006[20] and NCBI 16S rRNA annotations (see Fig. 5 C). Species in each class were split into slow growing (16S rRNA $\leq$ 3) and fast growing (16S rRNA $\geq$ 6). For all species pairs in these sets, oriC distance correlation was determined. Consistent with the initial analysis in Figure 4F, fast growing species showed a stronger correlation of oriC distance between its orthologs than slow growing species in all investigated classes. This indicates that copy number regulation is also involved in the evolutionary shaping of bacterial chromosomes proportional to its regulatory potential.

Discussion

In this study, we gave new insights into spatio-temporal regulation in bacteria caused by replication-induced chromosomal copy number effects. We could show that the gene expression pattern observed when comparing the exponential and stationary phase is induced by copy number differences between the two growth phases instead of spatio-temporal promoter regulation.

The initial approach was the construction of a strain, which harbours inverse copy numbers due to the relocation of the oriC into the terminus. However, moving the oriC would cause conflicts with several cellular systems coupled to replication. Conflicts involved the directional tus/ter replication termination system that would block replication from the terminus towards the native oriC location. Furthermore, the strain was made flexible for rRNA operon inversions by flanking loxP and FRT sites to make the strain ready to avoid head-on collisions of transcription and replication machinery. Other non-essential systems were not altered to reduce further invasive chromosome modifications. Systems like nucleoid occlusion (SlmA)[37], oriC macrodomain formation (maoP/maoS)[38] and ter domain formation (matS/matP)[39] depend on strategic positioning of binding sites. Transplantation to the new oriC or ter sites would have caused massive genome modifications potentially disrupting the local sequence context with unpredictable effects on chromosome integrity or transcription proximal to the deletion and insertion sites. Although not essential[40], we attempted to transfer the dif site to the native oriC locus but were not able to get positive clones. However, the resulting strain was viable, but showed frequent mutants with higher fitness. Genome analysis of a mutant revealed a large inversion covering the left replichore. This suggests a connection to the rRNA operons in the right replichore. The inversion improved copy number of these operons significantly by moving oriC in close proximity and at the same time abolished head-to-head collisions of RNAP and DNAP. A connection to macrodomains or overall chromosome structure is unlikely as the inversion both disrupts the ori and ter macrodomains. Other chromosomal organisation systems such as nucleoid occlusion (SlmA) or terminus segregation FtsK/KOPS usually symmetric to the oriC-ter axis are also disrupted by the inversion of a single replichore. This might be an indicator that these systems would have shown a minor contribution to improving the design of the initial approach with fully inverted copy numbers. However, the strain harbouring the inversion (INV) shows an increased doubling time (61.5 min) compared to WT (19.6 min). This can be due to a combination of the altered transcription due to the inversion and the disruption of the above mentioned systems. Nevertheless, the strain was stable and showed a reduced copy number gradient compared to WT and met the requirements of the study. In particular, the difference in copy number compared to WT allowed us to isolate the copy number effect from other regulatory factors and investigate the global

expression pattern. Expression analysis showed that, except for intended differences due to modified copy numbers, the overall mRNA levels were mainly consistent with WT expression in exponential phase (see Fig. S2 and S4). This indicates a mainly retained regulatory state without global perturbations.

When comparing the exponential and stationary phase of the INV strain, the gradual expression pattern is reduced giving the first indication of the major role of the copy number effect as a global regulator. By comparing the exponential phases of WT and INV, we were able to mainly eliminate promoter regulation as the same growth phases were compared. The analysis revealed a strong gradually expression pattern primary resulting from the differences in copy number between the strains. Furthermore, computational normalisation of the wild type expression with respect to individual gene copy numbers generated a pattern resembling the INV strain expression pattern, where gene copy numbers were strongly reduced by design (see Fig. S4). Quantification of the average fold changes between exponential and stationary phase tightly followed the measured copy numbers. This proved the general ability of the copy number effect to produce the distinctive expression pattern we observed in the WT. Promoter regulation also influences the spatial pattern more locally. Overlapping the copy number pattern, characteristic local peaks were present in the wild type and INV strain suggesting a connection to promoter regulation. This assumption was supported by the comparison of WT and INV both in exponential phase in which the promoter regulation differences are expected to be minimal. Consequently, the characteristic local peaks were not present. Moreover, the expression fold change was almost identical to the copy number differences between the two strains, indicating an even more pronounced impact of copy number on the spatial expression pattern.

Although fold change averages strictly follow the copy number, single genes can still strongly deviate positively or negatively from the average, but cancel each other out during averaging. Therefore, promoter regulation may not play a crucial role in global spatial pattern formation but can still dominate the regulation of single genes. To investigate regulation on the single gene level, we decomposed the fold change of each gene into a copy number and a promoter regulation component. We tested this analysis by comparing the exponential phase of the WT and INV strain, which revealed that most genes are predominantly copy number regulated as expected when promoter regulation differences are minimal (see Fig.4B). The remaining promoter regulation derived from differences between the two strains e.g. the inversion of the left replichore and its secondary effects. For the native growth phase transition from exponential to stationary phase, about 40% of the genes still showed a dominant copy number regulation and even 75% of the genes were at least controlled to 25% by copy numbers. These numbers indicate, that the copy number effect acts as a regulatory principle that can be compared with other major regulators like the transcriptional regulatory network[41], major sigma factors and DNA supercoiling[42]. The latter control specific cellular functions and thereby contribute to a coherent organisation of the cell. To test for a putative specific regulation of the copy number effect, predominantly copy number regulated genes were investigated with respect to their abundance in various functional groups. A significantly high frequency of these genes was involved in coenzyme metabolism and amino acid metabolism known to be related to growth[35,43]. Processes like these, directly coupled to cell growth and division, by their nature require fold changes in gene expression in the range of division rates. Transient copy numbers are linked to the division rate by DNA replication frequency. Consequently, such processes are prone to be regulated by transient copy numbers. On the other hand, genes in the group of energy metabolism were underrepresented. Energy metabolism depends on the presence of various molecular sources of energy[43,44]. Consistently, promoter regulation is more pronounced for this set of genes and copy numbers play a minor role. Hence, processes requiring complex regulation or higher fold changes are predominantly regulated by other regulatory mechanisms like transcription factors. Interestingly, a combination of both regulatory concepts can be observed in the aspartate pathway (see Fig. 3). Here, the basal level of the pathway is controlled by copy numbers and the internal balance of alternate fluxes is modulated by promoter regulation. A list of all *E. coli* metabolic pathways together with the information of their corresponding genes regarding the individual influence of copy number between the exponential and stationary phase is provided in the supplemental data.

Another indicator of coherent regulation by copy numbers was the conservation of position in two other species. The set of preferentially copy number regulated genes showed a reduced deviation of oriC distance in the course of evolution than promoter regulated genes. Hence, copy number regulation forces genes to keep their oriC distance. For related species with higher copy numbers, these genes are automatically expressed at higher levels during exponential phase. This could be a simple mechanism to shift up metabolism output for fast growth during adaptation to new environments[45] and is consistent with the spontaneous emergence of fast growing bacteria in various branches of the bacterial kingdom[20]. More compellingly, depending on the impact of copy numbers in these species, genes differentially expressed between the exponential and stationary phase were more or less sorted along the oriC-ter axis. For the slower growing plant pathogen *D.dadantii*, genes relevant during exponential phase were less driven towards oriC than in *E. coli*, whereas in the fast growing *V. natriegens* those genes were significantly shifted towards oriC compared to *E. coli*. An extension to various clades in the bacterial kingdom revealed a connection between the growth rate of a bacterium and the positional conservation along the oriC-ter axis. The higher the growth rate of an organism, the more pronounced the sorting along the oriC-ter axis. Hence, the copy number effect is more exploited in fast growing bacteria. The observation in both gram-positive and gram-negative bacteria indicates a fundamental

evolutionary concept of gene regulation and chromosome architecture coupled to replication dynamics and growth rates. Our findings may have various consequences. Copy numbers impact expression patterns of exponentially growing cells. The analysis of regulatory relationships using mutants that often exhibit growth defects may be biased. A reduced copy number effect of the slow growing mutant compared to the wild type may systematically alter the set of differentially expressed genes and thereby indicate false regulatory interactions. Also, other copy number modifications including overinitiation of replication or replication stalling can be detected by copy number analysis and can be corrected. We have shown that copy number effects can be computationally removed with the help of copy number analysis by DNA-sequencing (see Fig. S4) to be consistent with biological reality. This can be coupled with the verification of mutants by DNA sequencing and would therefore not lead to extra expenses. From the point of evolution, copy numbers are an interesting regulatory concept. Gene expression can be changed smoothly by shifting the gene along the oriC-ter axis with a range of several folds without the expense of regulatory proteins. As copy number regulation is fundamentally different to promoter regulation both can be applied independently with minimal crosstalk, which allows for fast evolutionary optimisation processes. For the emerging field of synthetic biology, the copy number effect can be exploited by strategic positioning of metabolic pathways minimising regulatory complexity. Especially for biotechnological applications with fast growing organisms, the continuous copy number gradient between oriC and ter is an ideal tuning vehicle for pathway integration.

Material and Methods

Strain cultivation and sequencing

The INV and WT strains were cultivated in LB medium (10g/L tryptone, 5g/L yeast extract, 10g/L NaCl) at 37°C under aerobic conditions in flasks with shaking at 200rpm. For the analyses of the exponential phase, both strains were harvested at an OD_{600} of 0.3 pelleted and immediately suspended in RNA*later* (Thermo Fisher) according to the manufacturer's instructions. The stationary phases of the WT and INV strains were harvested when no significant changes in OD_{600} could be observed for about 20 minutes. Subsequently, cells were also pelleted and suspended in RNA*later*. Samples were then split for DNA- and RNA-sequencing. Isolation of bacterial genomic DNA was performed according to Bruhn et al.[46]. For RNA-sequencing, lysis of cells and subsequent isolation of total RNA were carried out using the lysing matrix B/FastPrep® sample preparation system (MP Biomedicals) and the miRNeasy Mini Kit (Qiagen), respectively. Ribosomal RNA depletion (RNA) and library preparation (RNA/DNA) was conducted by Eurofins Genomics using the Illumina Technology (strand-specific; paired-end; 2x150bp read length). All samples (INV/WT exponential and stationary phase) were carried out in biological triplicates. For the comparative genomics analysis of three species (*Dickeya dadantii* 3937; *Escherichia coli* MG1655; *Vibrio natriegens* ATCC 14048), all species were grown in rich medium. For *D. dadantii* and *E. coli* LB medium was used. For the halophilic *V. natriegens* LBV2 (LB + 204mM NaCl, 4.2mM KCl, 20.14mM $MgCl_2$) was used. All three species were grown under aerobic conditions in baffled flasks with orbital shaking at 200 rpm. For optimal growth, *E. coli* and *V. natriegens* were grown at 37°C and *D. dadantii* at 30°C. Cells were harvested at OD_{600}=0.3 in mid-exponential phase. DNA was also extracted according to Bruhn et al.[46] and Illumina-sequenced by Eurofins Genomics yielding 5M 150bp paired-end reads.

Regulatory and functional data

Data concerning sigma factor and transcription factor regulation was obtained from regulonDB (v10.9). For chromosomal sigma factor distribution, genes solely regulated by the respective sigma factor were selected. Data concerning functional identity were derived from the NCBI COG database (https://www.ncbi.nlm.nih.gov/research/cog-project/).

Construction of the inversion (INV) and reference strain (WT)

For the construction of the inversion strain (INV) to dissect the impact of promoter regulation and copy number effect as well as for reference strain (WT) we used CRISPR SWAPnDROP to make all relevant chromosomal changes in *E. coli* MG1655[32]. For each deletion and insertion, a different (pSwap) plasmid was constructed harbouring homology regions, sgRNAs and inserts. All primers used for the amplification of the homology regions, for the sgRNA construction as well as for each insert are available in the supplementary data. An overview of each chromosomal edit done in each of the strains is given in Table1.

Analysis of copy numbers and marker frequency

DNA read mapping was done with the R QuasR package. Marker frequency analysis (MFA) was performed to measure copy number[29,47,48]. Genome coverage of exponential phase samples was first averaged over 5kb sliding windows relative to the corresponding stationary phase to get robust estimates of local copy numbers (see data points in MFA plots). A $\log_2$ linear regression of local copy numbers was performed for each replichore separately. The intersection ordinate of the two replichore regression curves was used as oriC and terminus (ter) copy number estimates. The data were normalised to a terminus copy number of 1 to simplify illustrations. For copy number difference, the fold change between the regression curves of the

Table 1. Chromosomal edits of the *E. coli* INV and WT strain

Edit	Strain	Purpose
Δtus	INV/WT	ter-site inactivation
oriC::oriSFRT	INV/WT	new oriS with flanking FRT sites for removal
insPQ::oriC	INV/WT	relocation of oriC
FRT insertion downstream of rrnE opeorn	INV/WT	Inversion of the rrnCABE operons
FRTm insertion upstream of rrnH operon	INV/WT	Inversion of the rrnH operon
FRTm insertion downstream of rrnH operon	INV/WT	Inversion of the rrnH operon
loxP511 insertion upstream of rrnG operon	INV/WT	Inversion of the rrnG operon
loxP511 insertion downstream of rrnG operon	INV/WT	Inversion of the rrnG operon
loxP insertion upstream of rrnD operon	INV/WT	Inversion of the rrnD operon
loxP insertion downstream of rrnD operon	INV/WT	Inversion of the rrnD operon
ΔoriC(insPQ)	WT	reconstitution of the wild type
oriSFRT::oriC	WT	reconstitution of the wild type

investigated growth phases and strains was calculated at the corresponding locus. For copy numbers of individual genes, the ordinate of the respective replichore regression curve at the gene locus was used.

Expression analyses

RNA-sequencing reads of each gene were first normalized for gene length and the total number of reads in each sample. All samples of this study were quantile-normalized in one batch to harmonize differences in overall gene expression distributions caused by technical variation. For the expression bias analyses, the differences in the number of up- and down-regulated genes between the growth phases were determined for sliding windows of 300 genes. The chromosomal location of each window was set to the average location of all genes in the window. For the fold change analyses, the average gene expression fold changes between the growth phases were determined for sliding windows of 300 genes. As the spatial expression pattern represents a systematic bias in expression data, the average fold changes were further corrected for relative frequency biases. This systemic bias originates from the imbalance of relative frequency when one component is enriched leading to a depletion of all other components. In the concrete case, copy number causes an increase of oriC-proximal gene expression levels, which in turn reduces expression levels of the terminus-proximal genes. This results in negative average fold change values in the terminus region. However, we assume the average fold change to be 1 at the chromosomal location where no copy number difference is present between samples. Therefore, we corrected the fold changes accordingly. The location of no copy number difference was first extracted from the copy number difference curve. The average fold change bias at that location was determined by taking the ordinate value of a regression curve of the fold change data on both replichores. All fold changes were corrected for that ordinate value. In the spatial analyses of WT and INV, reference chromosome coordinates were set to wild type. The inversion in the INV strain causes new neighborhoods of genes at the break points of the inversion. Therefore, windows spanning these breakpoints contain gene sets that are not present in WT (e.g. WT oriC-proximal genes paired with terminus-proximal genes). These windows were omitted in the analysis as no counterpart was present in WT.

Analysis of gene migration in *Dickeya dadantii*, *Escherichia coli* and *Vibrio natriegens*

Orthologs of genes in the three species were determined using proteinortho v6[49]. Only orthologs with a single copy (no paralogs) in all three species were considered. Gene positions were transformed to relative oriC proximities and normalised by half of the chromosome size. Consequently, values range between 1 (oriC) and 0 (ter).

For the analysis of variability of gene position, promoter and copy number regulated genes were split into 3 subsets with increasing stringency of the regulatory type (promoter or copy number). For each gene in a set the difference of oriC proximity was determined for *E. coli* vs *D. dadantii* and *E. coli* vs *V. natriegens*. The two resulting differences were averaged to reduce species-specific biases. Then, the average and standard errors of these averaged differences were determined for the different sets.

For the comparison of ortholog positions in the three species, genes significantly up-regulated (p-value < 0.05) in exponential phase relative to stationary phase of *E. coli* WT expression data were chosen.

Significance of functional groups

The significance of functional groups was determined by generating 1000 random sets of genes of the same size as the set of predominantly copy number regulated genes. For these sets mean frequencies **m** and the corresponding standard deviations **s**

for functional groups were determined to compute a z-score **z**

$$z(x) = \frac{x - m}{s} \tag{2}$$

for each functional group, where x is the number of genes in the respective function group of the experimental data.

Comparative genomics analysis of bacterial chromosome arrangement

The full NCBI set of completely assembled genomes was first screened for NCBI taxonomy information to cluster species according to phylogenetic categories. The remaining set was split into distinct phylogenetic classes that were analysed separately. To cover the diversity of a class and avoid a representation bias, one species was selected out of every family of the class. Each species was a randomly selected representative species of its family, according to the NCBI database. The family category was chosen to select species with a defined range of evolutionary distance for later comparison. Furthermore, categories below a family with closer evolutionary distance yielded little chromosomal diversity for analysis. Species within a class were split into the set of fast or slow growing species. This was done by the number of 16S rRNAs that correlate with growth speed. The number of 16S rRNA was extracted from the species annotation files (GFF3) provided by NCBI. Data about the doubling times were taken from Couturier and Rocha 2006[20]. For all species pairs within a set, orthologs were determined using proteinortho v6[49]. Only orthologs with a single copy (no paralogs) in the two species of a pair were considered to determine reliable chromosome positions of orthologs. The oriC-ter axis was determined by finding the putative oriC position in both species that give rise to the best Pearson correlation coefficient of distances to oriC (see constellation analysis). These correlation coefficients were used as an indicator of the strength of oriC-ter axis symmetry.

Constellation analysis

For most species, oriC position is not determined. However, the oriC position causes a chromosomal symmetry due to positional conservation of genes relative to oriC[3]. This can be used to determine the oriC-ter axis. To determine the axis, the best overlay of chromosomes of two related species is determined. First, an arbitrary oriC position is assigned individually to both chromosomes (e.g. position 100000 for species 1 and position 500000 for species 2). The relative distances to all genes on both chromosomes to the assigned oriCs are determined (see Figure 5A top row). For all orthologous gene pairs of the two species, the distance to the respective oriC is compared yielding a correlation coefficient (Pearson correlation coefficient). This approach is repeated with other arbitrary oriC positions until all combinations were tested (see Figure 5B). The positions yielding the highest correlation are taken to be oriC. In (see Figure 5B) several optima are present due to intrinsic symmetries. For example, taking the correct oriC position of *E. coli* and *D. datantii* yields the same maximal correlation as taking the correct terminus position of *E. coli* and the terminus position of *D. datantii* as this directly infers that oriC is also superimposed. Which of the two optima are chosen has no impact on further analysis in this study as the level of the correlation is central not the related positions that yielded the maximal correlation. To distinguish which optimum is associated with oriC and ter, additional information such as *dnaA* or ribosomal operon location can be used. However, this is not relevant for the study. The strength of the correlation is an indicator of the conservation of gene position relative to oriC. For the analysis, it is important to consider the total evolutionary distance, as closely related species, in general, show a higher degree of conservation. This can be accomplished by choosing species within the same phylogenetic ranges (e.g. class, family, phylum etc.). Furthermore, there is a lower limit for this method, when little chromosomal rearrangements occurred between species. In this case, constellation analysis indicates this by a diagonal line of similarly high correlations. This is caused by the fact that correlation remains stable when rotating putative oriC positions on both chromosomes simultaneously (e.g. moving in 5000 bp steps in a clockwise direction). If gene positions are similar on both chromosomes, also distances to the oriCs change accordingly during synchronous rotation and retain the same correlation coefficient. The more genes have moved to other locations e.g. the opposite replichore, the more correlation differs during synchronous rotation. In Figure 5B a faint diagonal line can still be seen, indicating a small percentage of orthologous genes with a similar chromosome position in both species. As long as the two optima can still be clearly determined, the method is applicable.

References

1. de Matos Simoes, R., Dehmer, M. & Emmert-Streib, F. Interfacing cellular networks of S. cerevisiae and E. coli: connecting dynamic and genetic information. *BMC Genomics* **14**, 324 (2013).

2. Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodomain organization of the Escherichia coli chromosome. *EMBO J* **23**, 4330–4341 (2004).

3. Sobetzko, P., Glinkowska, M., Travers, A. & Muskhelishvili, G. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol Biosyst* **9**, 1643–1651 (2013).

4. Menzel, R. & Gellert, M. Regulation of the genes for e. coli dna gyrase: homeostatic control of dna supercoiling. *Cell* **34**, 105–113 (1983).

5. Balke, V. L. & Gralla, J. Changes in the linking number of supercoiled dna accompany growth transitions in escherichia coli. *Journal bacteriology* **169**, 4499–4506 (1987).

6. Liu, L. F. & Wang, J. C. Supercoiling of the dna template during transcription. *Proceedings National Academy Sciences* **84**, 7024–7027 (1987).

7. Riebet, E. & Raibaud, O. Supercoiling is essential for the formation and stability of the initiation complex at the divergent malep and malkp promoters. *Journal molecular biology* **218**, 529–542 (1991).

8. Chen, D. & Lilley, D. M. Transcription-induced hypersupercoiling of plasmid dna. *Journal molecular biology* **285**, 443–448 (1999).

9. Sobetzko, P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. *Nucleic Acids Res* **44**, 1514–1524 (2016).

10. El Houdaigui, B. *et al.* Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res* **47**, 5648–5657 (2019).

11. Jinks-Robertson, S. & Nomura, M. Ribosomes and trna. escherichia coli and salmonella typhimurium molecular and cellular biology. (1987).

12. Condon, C., Liveris, D., Squires, C., Schwartz, I. & Squires, C. L. rrna operon multiplicity in escherichia coli and the physiological implications of rrn inactivation. *Journal bacteriology* **177**, 4152–4156 (1995).

13. Slack, A., Thornton, P. C., Magner, D. B., Rosenberg, S. M. & Hastings, P. J. On the mechanism of gene amplification induced under stress in Escherichia coli. *PLoS Genet* **2**, e48 (2006).

14. Huang, Y.-P. & Ito, J. Dna polymerase c of the thermophilic bacterium thermus aquaticus: classification and phylogenetic analysis of the family c dna polymerases. *Journal molecular evolution* **48**, 756–769 (1999).

15. Egan, E. S., Løbner-Olesen, A. & Waldor, M. K. Synchronous replication initiation of the two vibrio cholerae chromosomes. *Current Biology* **14**, R501–R502 (2004).

16. Donachie, W. D. Relationship between cell size and time of initiation of dna replication. *Nature* **219**, 1077–1079 (1968).

17. Bipatnath, M., Dennis, P. P. & Bremer, H. Initiation and velocity of chromosome replication in escherichia coli b/r and k-12. *Journal bacteriology* **180**, 265–273 (1998).

18. Ferullo, D. J., Cooper, D. L., Moore, H. R. & Lovett, S. T. Cell cycle synchronization of Escherichia coli using the stringent response, with fluorescence labeling assays for DNA content and replication. *Methods* **48**, 8–13 (2009).

19. Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology (Reading)* **143 (Pt 6)**, 2071–2078 (1997).

20. Couturier, E. & Rocha, E. P. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol Microbiol* **59**, 1506–1518 (2006).

21. Schmid, M. B. & Roth, J. R. Gene location affects expression level in Salmonella typhimurium. *J Bacteriol* **169**, 2872–2875 (1987).

22. Chandler, M. G. & Pritchard, R. H. The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. *Mol Gen Genet* **138**, 127–141 (1975).

23. Sobetzko, P., Glinkowska, M., Travers, A. & Muskhelishvili, G. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol Biosyst* **9**, 1643–1651 (2013).

24. Schultz, J. E., Latter, G. I. & Matin, A. Differential regulation by cyclic AMP of starvation protein synthesis in Escherichia coli. *J Bacteriol* **170**, 3903–3909 (1988).

25. Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O. & Matthews, R. G. Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proc Natl Acad Sci U S A* **99**, 13471–13476 (2002).

26. Martis B, S., Forquet, R., Reverchon, S., Nasser, W. & Meyer, S. DNA Supercoiling: an Ancestral Regulator of Gene Expression in Pathogenic Bacteria? *Comput Struct Biotechnol J* **17**, 1047–1055 (2019).

27. Klein, C. A., Teufel, M., Weile, C. J. & Sobetzko, P. The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. *Sci Rep* **11**, 24399 (2021).

28. Hill, T. M. Arrest of bacterial DNA replication. *Annu Rev Microbiol* **46**, 603–633 (1992).

29. Ivanova, D. *et al.* Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res* **43**, 7865–7877 (2015).

30. Missirlis, P. I., Smailus, D. E. & Holt, R. A. A high-throughput screen identifying sequence and promiscuity characteristics of the loxP spacer region in Cre-mediated recombination. *BMC Genomics* **7**, 73 (2006).

31. Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J Mol Biol* **402**, 52–69 (2010).

32. Teufel, M., Klein, C. A., Mager, M. & Sobetzko, P. A multifunctional system for genome editing and large-scale interspecies gene transfer. *Nat. Commun.* **13** (2022).

33. Beber, S. P. M. G. H. M., M.E. Interplay of digital and analog control in time-resolved gene expression profiles. *EPJ Nonlinear Biomed Phys* **4** (2016).

34. Lu, P. The lac operon: A short history of a genetic paradigm. *Science* **275**, 938–939 (1997).

35. Hartl, J., Kiefer, P., Meyer, F. & Vorholt, J. A. Longevity of major coenzymes allows minimal de novo synthesis in microorganisms. *Nat Microbiol* **2**, 17073 (2017).

36. Gao, F. & Zhang, C.-T. DoriC: a database of oriC regions in bacterial genomes. *Bioinformatics* **23**, 1866–1867, DOI: 10.1093/bioinformatics/btm255 (2007). https://academic.oup.com/bioinformatics/article-pdf/23/14/1866/469973/btm255.pdf.

37. el Hajj, H. H., Zhang, H. & Weiss, B. Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *J Bacteriol* **170**, 1069–1075 (1988).

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

39. Mercier, R. *et al.* The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell* **135**, 475–485 (2008).

40. Blakely, G., Colloms, S., May, G., Burke, M. & Sherratt, D. Escherichia coli XerC recombinase is required for chromosomal segregation at cell division. *New Biol* **3**, 789–798 (1991).

41. Marr, C., Geertz, M., Hütt, M. T. & Muskhelishvili, G. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol* **2**, 18 (2008).

42. Blot, N., Mavathur, R., Geertz, M., Travers, A. & Muskhelishvili, G. Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep* **7**, 710–715 (2006).

43. Zampieri, M., Hörl, M., Hotz, F., Müller, N. F. & Sauer, U. Regulatory mechanisms underlying coordination of amino acid and glucose catabolism in Escherichia coli. *Nat Commun* **10**, 3354 (2019).

44. Stülke, J. & Hillen, W. Carbon catabolite repression in bacteria. *Current opinion microbiology* **2**, 195–201 (1999).

45. Bremer, H., Dennis, P. P., Neidhardt, F. *et al.* Escherichia coli and salmonella: cellular and molecular biology. *Washington (DC): American Society for Microbiology. Chapter, Modulation chemical composition other parameters cell by growth rate* 1553–1569 (1996).

46. Bruhn, M. *et al.* .

47. Maduike, N. Z., Tehranchi, A. K., Wang, J. D. & Kreuzer, K. N. Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. *Mol Microbiol* **91**, 39–56 (2014).

48. Kemter, F. S. *et al.* Synchronous termination of replication of the two chromosomes is an evolutionary selected feature in Vibrionaceae. *PLoS Genet* **14**, e1007251 (2018).

49. Lechner, M. *et al.* Proteinortho: detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
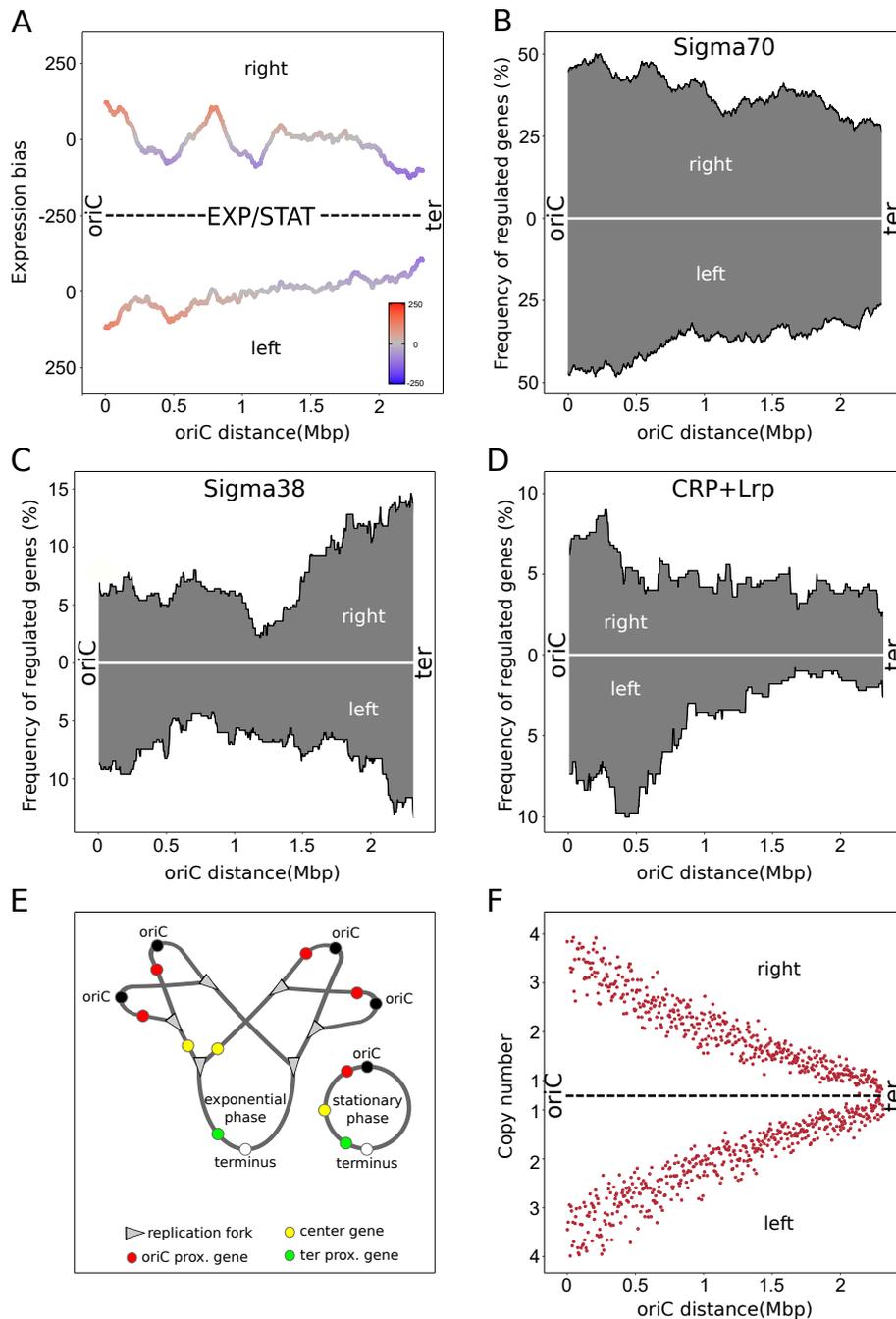
Acknowledgements

Figure 1. Overview of the *E. coli* wild type Distributions and biases were calculated using a sliding window of 300 genes. Distributions were normalized over the total gene number of each window. The replichores (right/left) are organized from the left to the right representing the oriC and the terminus, respectively. **(A)** Spatial bias of up- and down-regulated genes between exponential and stationary phase in the *E. coli* CSH50 wild type. Transcriptomic data originated from Sobetzko et al.[23] **(B)** Spatial frequency of Sigma70 regulated genes. **(C)** Spatial frequency of Sigma38 regulated genes. **(D)** Combined spatial frequency of CRP or LRP repressed genes. **(E)** Scheme of overlapping replication in exponential phase and its consequences on transient gene copies compared to stationary phase. **(F)** Marker frequency analysis of the *E. coli* MG1655 wild type strain for exponential phase.
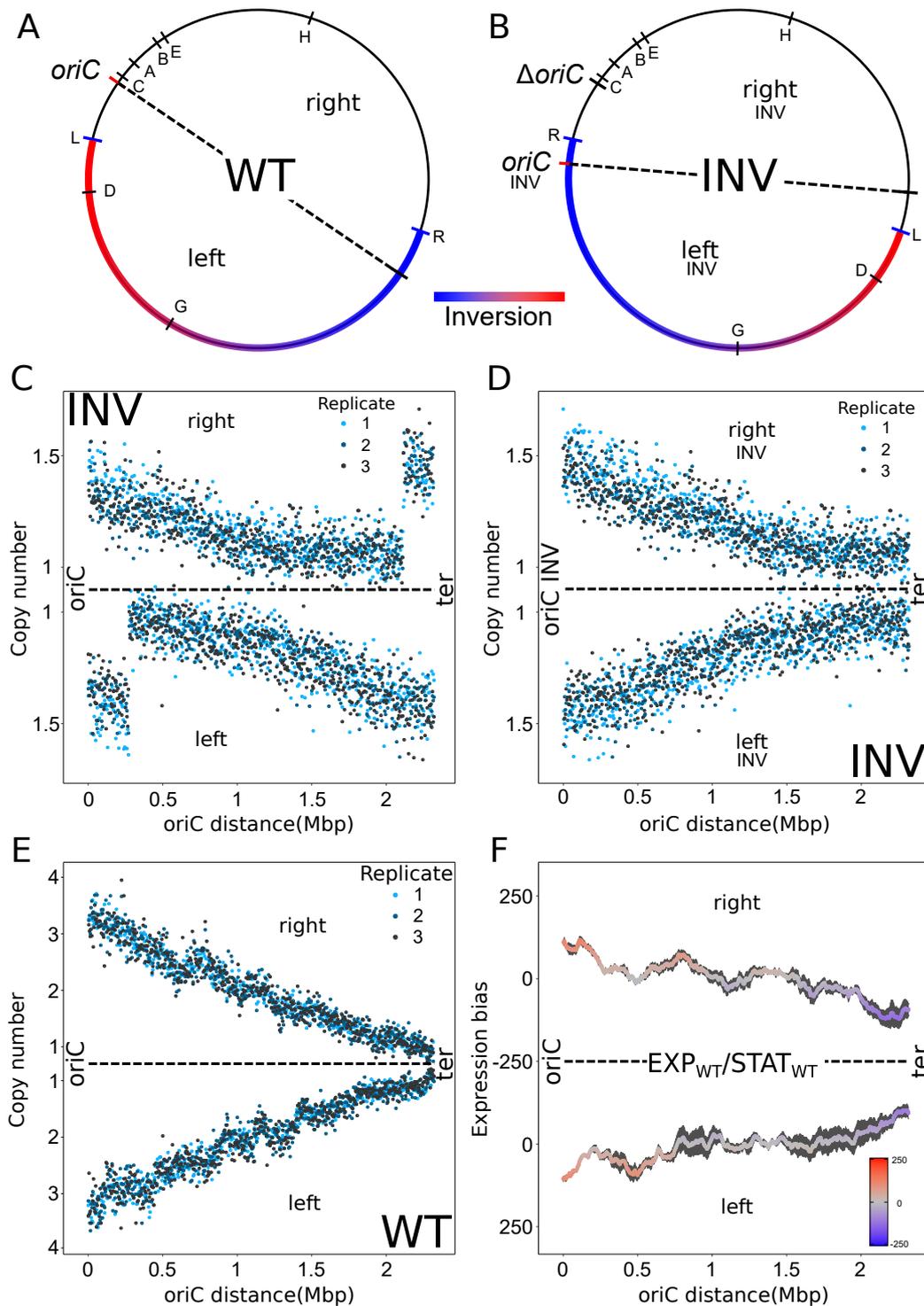
Figure 2. Characterisation of the INV and WT strain. (A) Chromosomal map of the WT strain. The inversion region for the derived INV strain is indicated in a red to blue gradient. Ribosomal RNA operon positions are indicated by capital letters within the circle. The letters L and R outside the circle indicate inversion break points. The dashed line indicates the chromosomal symmetry axis. **(B)** Chromosomal map of the INV strain with the same indications as in A. **(C)** Marker frequency analysis of the INV strain for exponential phase mapped against the WT genome. **(D)** Marker frequency analysis of the INV strain for exponential phase mapped against the INV genome. OriC INV, right$_{INV}$, left$_{INV}$ indicate the new chromosomal organization of the INV strain. **(E)** Marker frequency analysis of the WT strain for exponential phase. **(F)** Spatial bias of up- and down-regulated genes of three replicates between exponential and stationary phase in the WT strain using a sliding window of 300 genes. Standard error is indicated by a black area around the mean.
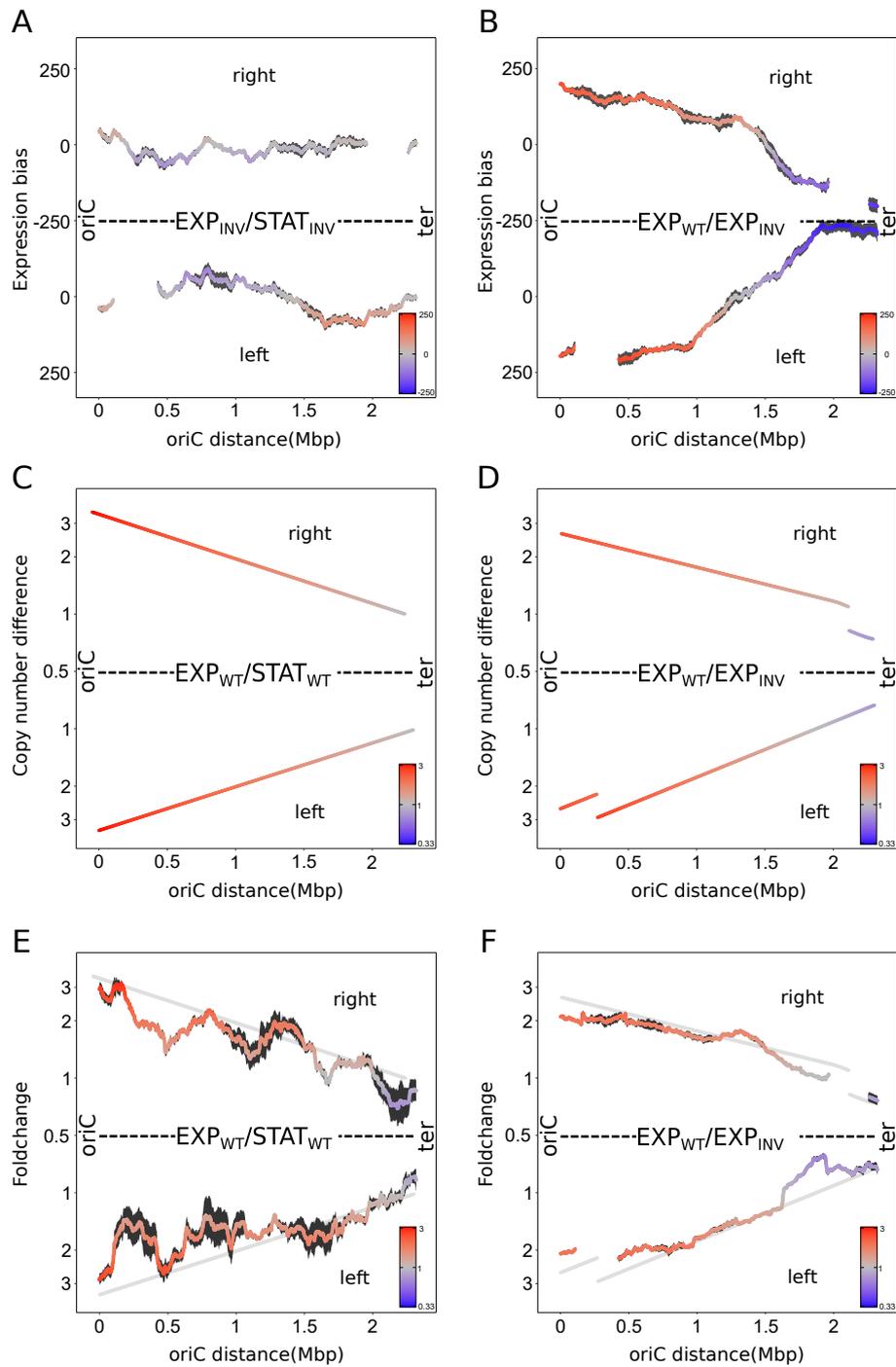
Figure 3. Impact of copy number on chromosomal spatial gradients. (A) Spatial bias of up- and down-regulated genes (sliding window of 300 genes) of three replicates between exponential and stationary phase in the INV strain mapped against the WT genome. Standard error is indicated by a black area around the mean. (B) Same as in A, but for the comparison of WT and INV strain during exponential phase. (C) Average local difference in copy number derived from MFA analysis of three replicates between exponential and stationary phase in the WT strain. (D) Same as in C, but for the comparison of WT and INV strain during exponential phase. (E) Average local expression fold change (sliding window of 300 genes) of three replicates between exponential and stationary phase in the WT strain. The data was normalized for relative frequency biases (see Material and Methods). The grey line indicates the local copy number differences of C. (F) Same as in E, but for the comparison of WT and INV strain during exponential phase. The grey line indicates the local copy number differences of D.
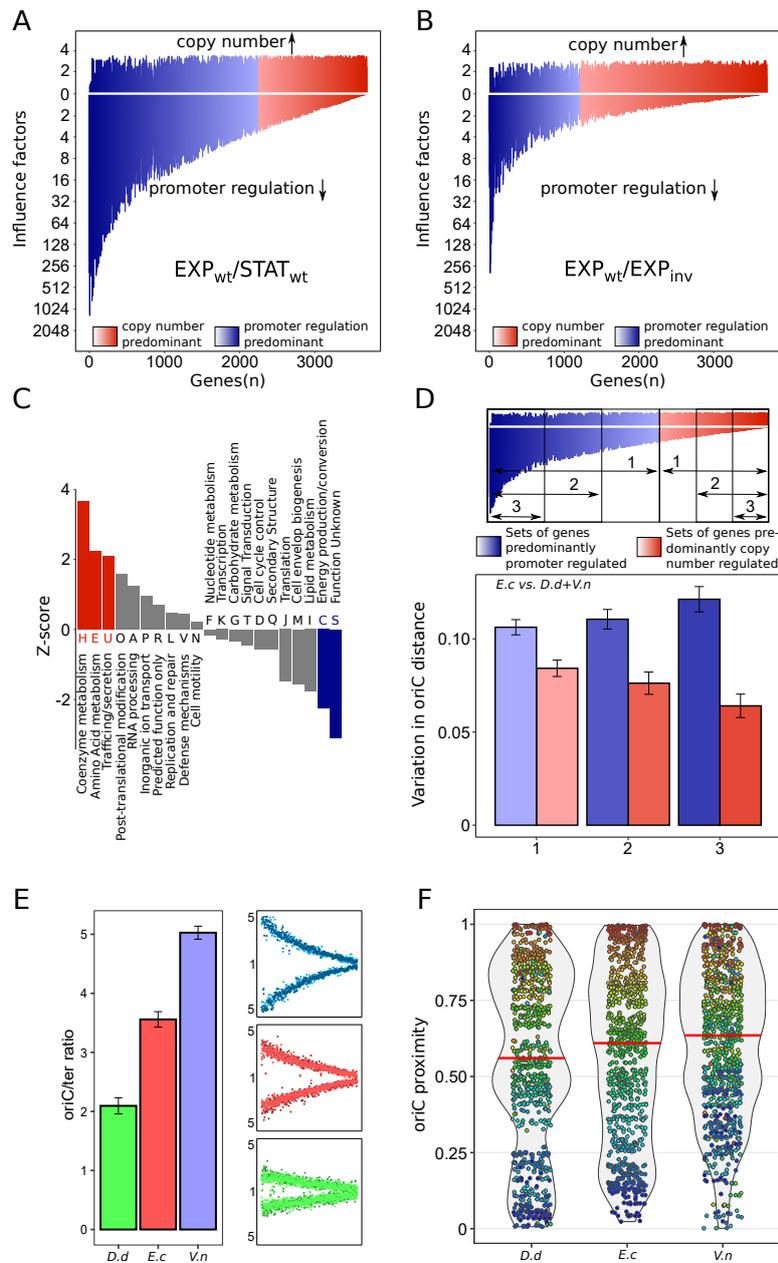
Figure 4. Impact of copy number on gene regulation and evolutionary gene migration. **(A)** Copy number factor and promoter regulation factor of all genes sorted by its ratio in the comparison of wild type between exponential and stationary phase. Rightmost genes show the highest impact of copy number effect on its total regulation. Blue colors indicate a higher impact of promoter regulation whereas red colors indicate a higher impact of copy number regulation. **(B)** Same as in A, but for the comparison of WT and INV strain during exponential phase. **(C)** Significance (z-score) of over- and underrepresented functional groups of WT genes predominantly regulated by copy number for the comparison of exponential and stationary phase. **(D)** Conservation of oriC distance of *E. coli* orthologous gene present in *D. dadantii* and *V.natriegens*. Variation is the fraction of the full oriC-ter distance. Red and blue colors indicate the sets of predominantly copy number and promoter regulated gene sets with different stringency, respectively. **(E)** Copy number of *D. dadantii*, *E. coli* and *V.natriegens* and the corresponding marker frequency plots for exponential growth. **(F)** oriC distance violin plots with orthologs of *D. dadantii*, *E. coli* and *V.natriegens*. Orthologous genes are active during exponential growth in *E. coli*. Median values are indicated by horizontal red lines. Individual genes and its orthologs are indicated as dots and are color coded according to the oriC-ter order in *E. coli*.
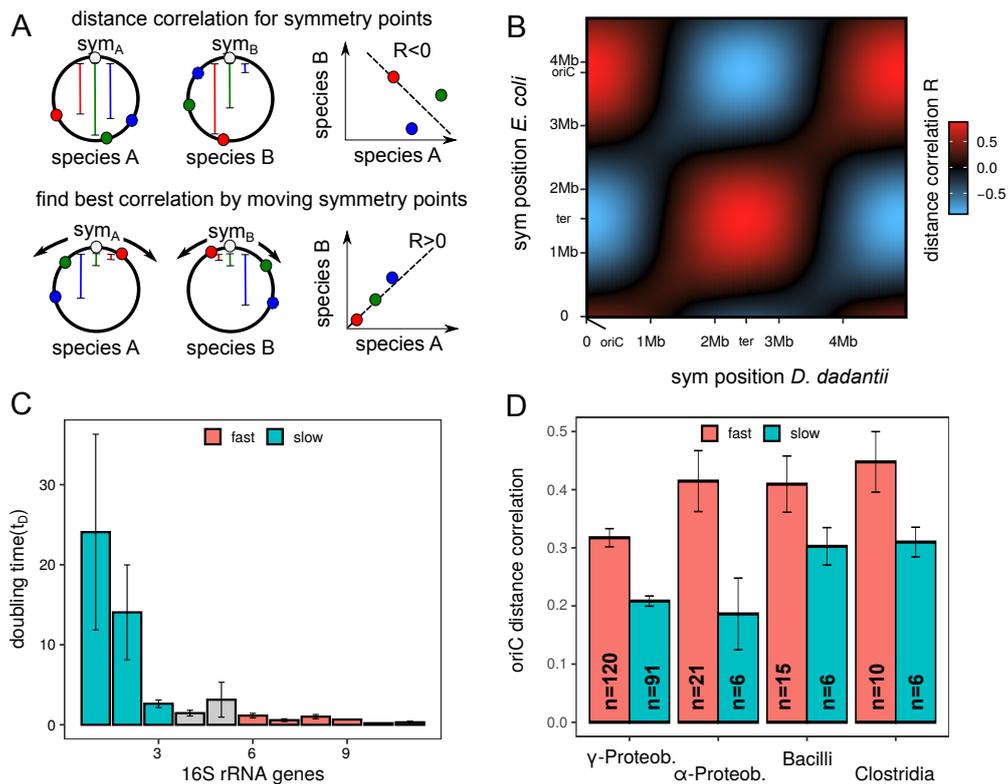
Figure 5. **Analysis of oriC distance conservation in slow and fast growing bacteria.** (**A**) Scheme of screening for the oriC-ter axis in species without known oriC. The axes in both species move through the putative oriC positions that yields the best oriC distance correlation of orthologs. (**B**) oriC-ter axis analysis for *E. coli* and *D. dadantii*. The correct oriC positions are indicated matching the maximum correlation (red). (**C**) Interdependence of doubling time and number of 16S rRNA genes. Selected groups for fast and slow growing species are indicted by red and blue. (**D**) Average oriC distance correlation of species from different classes. The number of pairs used for averaging are indicated in the bars. Red and blue colors indicate the groups of fast and slow growing species.
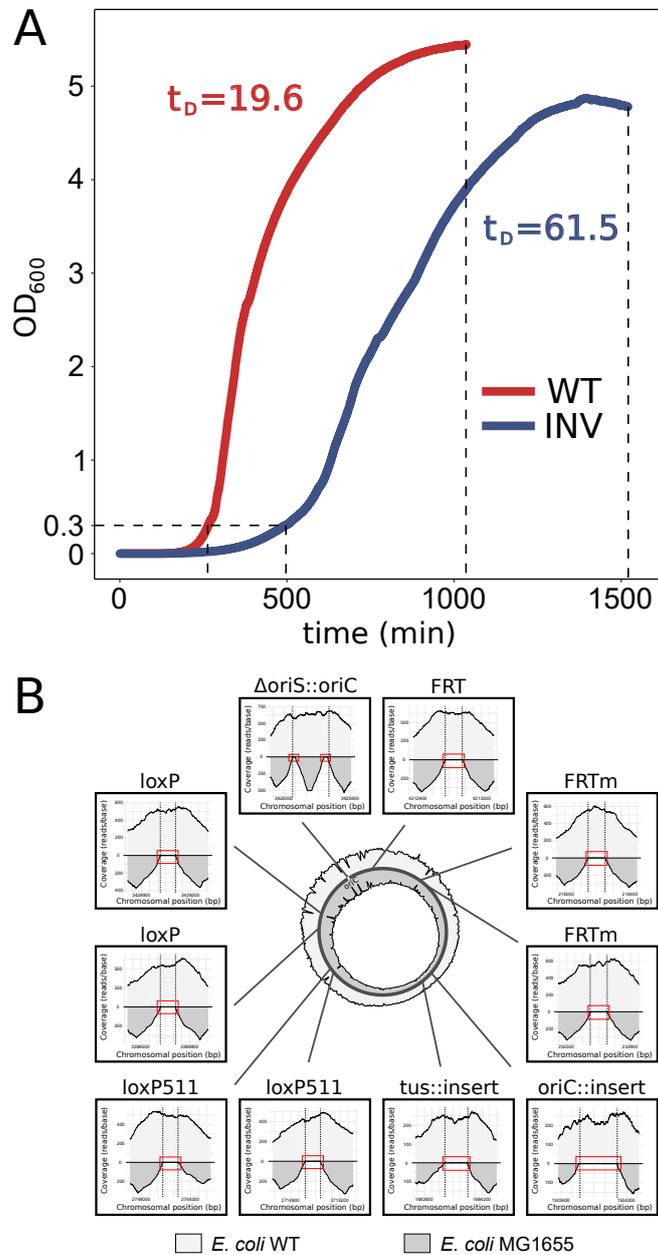
Figure S1. Growth curve and chromosomal edits in the WT and INV strains. (A) Growth curve of the WT and the INV strain in LB medium at 37°C and aeration. Doubling times (t_D) are indicated. Harvesting of exponential phase and stationary phase samples is indicated by dashed lines. **(B)** Next-Generation Sequencing of the *E. coli* WT strain. Shown is the next-generation sequencing (NGS) coverage of the *E. coli* WT strain after 12 consecutive edits (light grey) compared to its precursor *E. coli* MG1655 (dark grey). For the wild type strain, several recombination sites (FRT/loxP) as well as random DNA and origins of replication were inserted into the chromosome using iterative CRISPR SWAPnDROP genome editing. *E. coli* WT strain and MG1655 NGS reads were aligned against the WT strain reference genome and the sectors of each edited site as well as the complete genome coverage (circle) are shown. Reads at all insertion locations (dashed lines) are present for the wild type strain, while no reads are present for MG1655 (red rectangle).
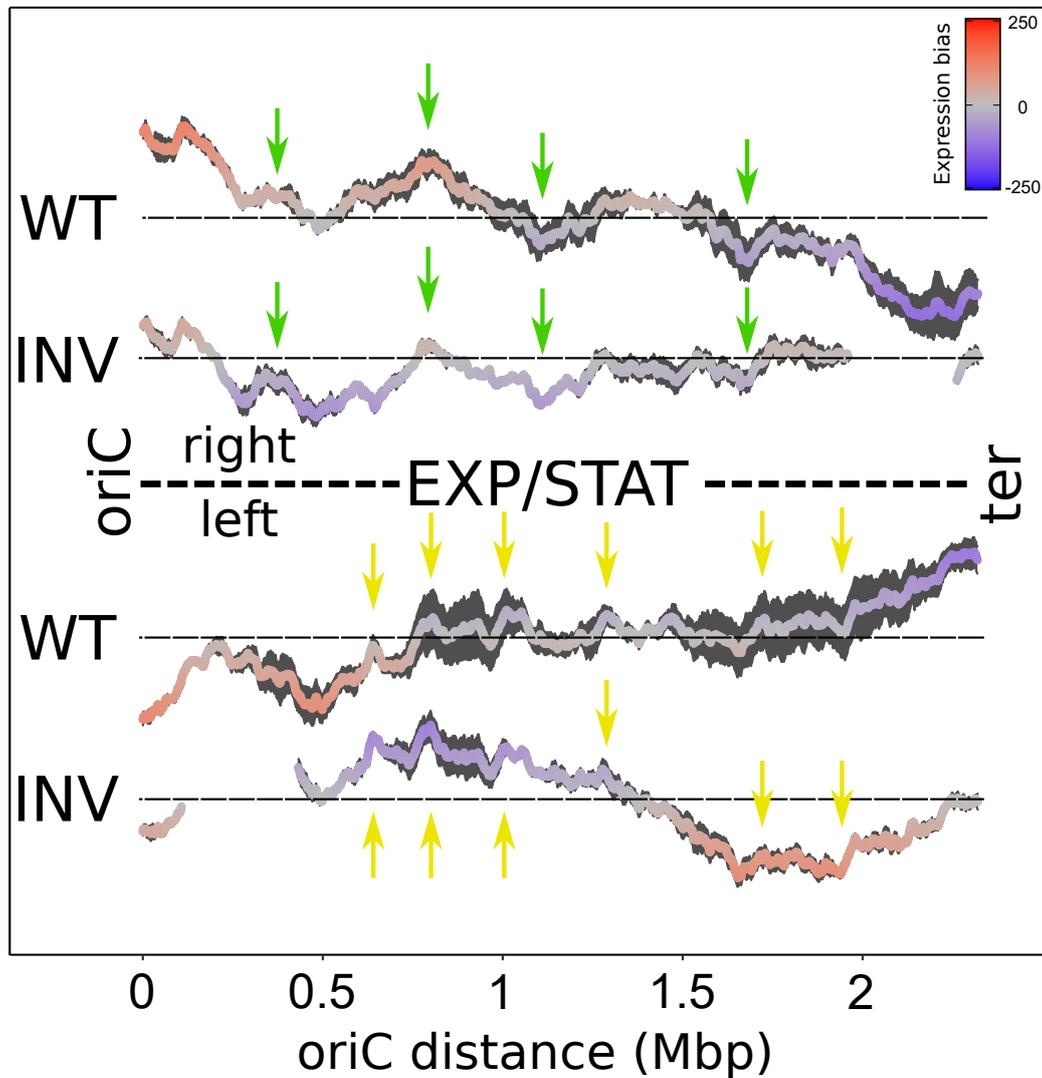
Figure S2. Comparison of the local spatial expression pattern of WT and INV strains between exponential and stationary phase. Replichores of both strains were aligned by position but shifted vertically to avoid overlapping. Expression biases of both strains are mapped against the WT chromosomal location. The zero level of each curve is indicated by a black horizontal line. Green and yellow arrows indicate characteristic local peaks on the left and the right replichore, respectively. Gaps for the INV strain are due to the absence of corresponding wild type windows comprising the inversion break points.
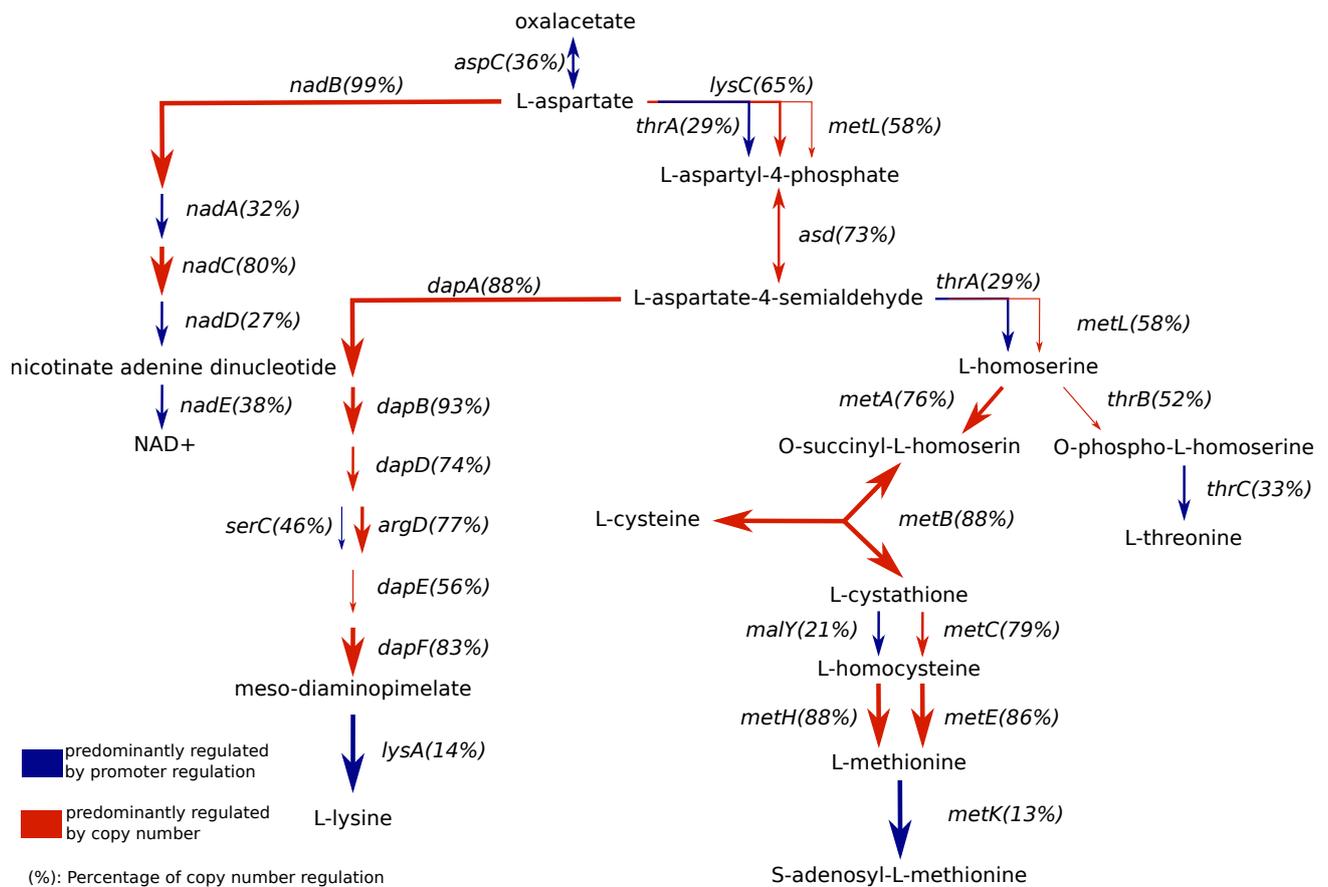
Figure S3. Impact of copy number on the aspartate pathway regulation. Shown is the aspartate pathway of *E. coli*. Red and blue arrows indicate dominant copy number and promoter regulation, respectively. The thickness of the arrow indicate the degree of the dominance also indicated in percent next to the gene coding for the enzyme involved in the enzymatic step.
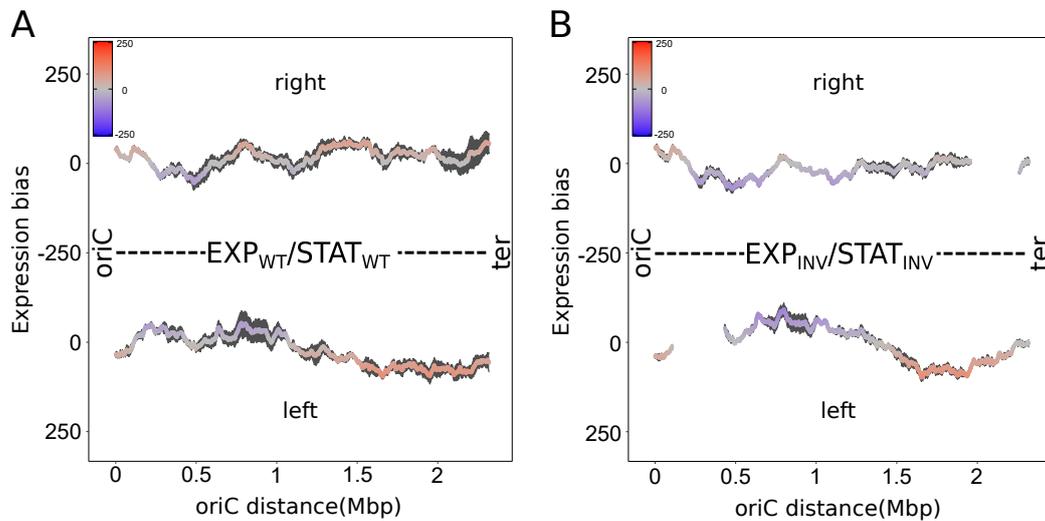
Figure S4. Computational copy number normalisation vs. biological copy number reduction. Comparison of the spatial expression pattern of WT with copy number normalisation (A) and the spatial pattern of the INV strain with a strongly reduced copy number between exponential and stationary phase (B). Colors indicate the extent of the spatial expression bias. Gaps for the INV strain are due to the absence of corresponding WT windows comprising the inversion break points.
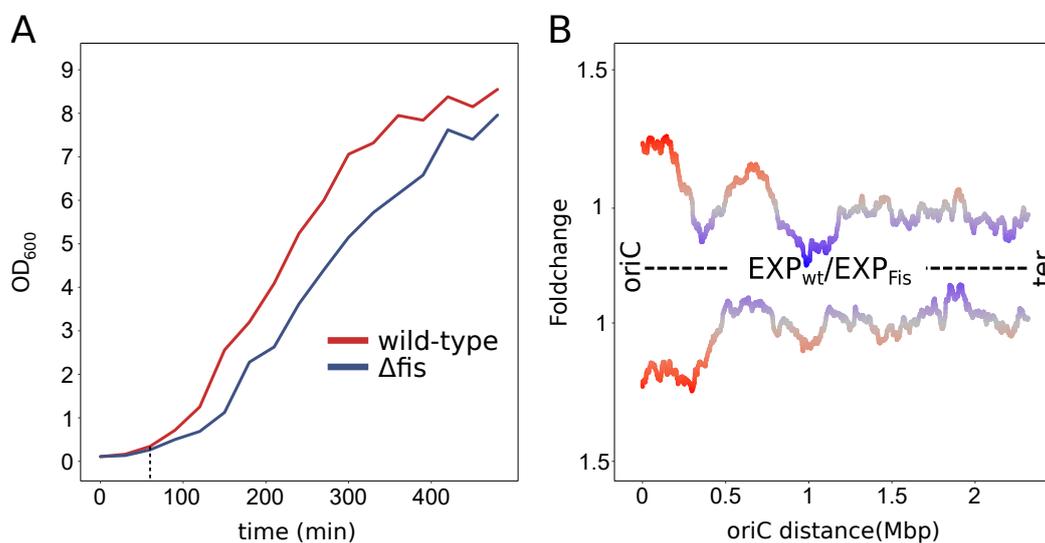
Figure S5. Putative impact of growth defects on spatial expression in mutant analysis. Expression data and growth curves were taken from Beber et al. 2016[33]. **(A)** Growth curves of *E. coli* wild type and its Fis deletion mutant. Harvesting of exponential phase samples is indicated by dashed lines. **(B)** Spatial expression pattern of *E. coli* wild type (CSH50) compared to its Fis deletion mutant. Average fold changes (wt/Δfis) of gene expression within a sliding window of 300 genes is depicted.
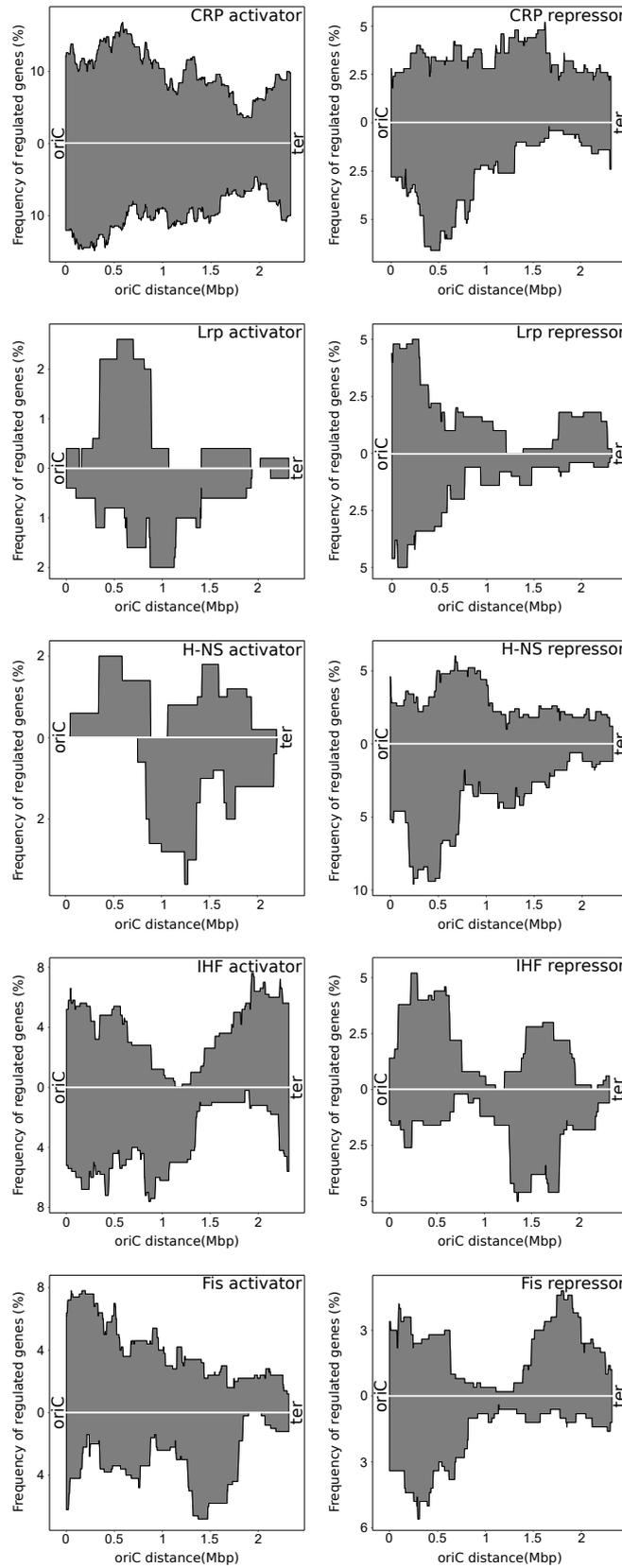
Figure S6. Spatial frequency of genes regulated by global regulators. Shown is the spatial frequency of genes regulated by the global regulators CRP, Lrp, H-NS, IHF and Fis subdivided according to their activation or repression activity.

Figure S7. Comparison of the _E. coli_ WT and MG1655 strains. Comparison of the functionality group analysis of the WT (a) and MG1655 (b) strain. (c) Expression data (log$_{10}$) from MG1655 replicate 1 was plotted against MG1655 replicate 2. (d) Averaged expression data (log$_{10}$) from MG1655 was plotted against averaged expression data from WT strain.

3.2 A multifunctional system for genome editing and large-scale interspecies gene transfer

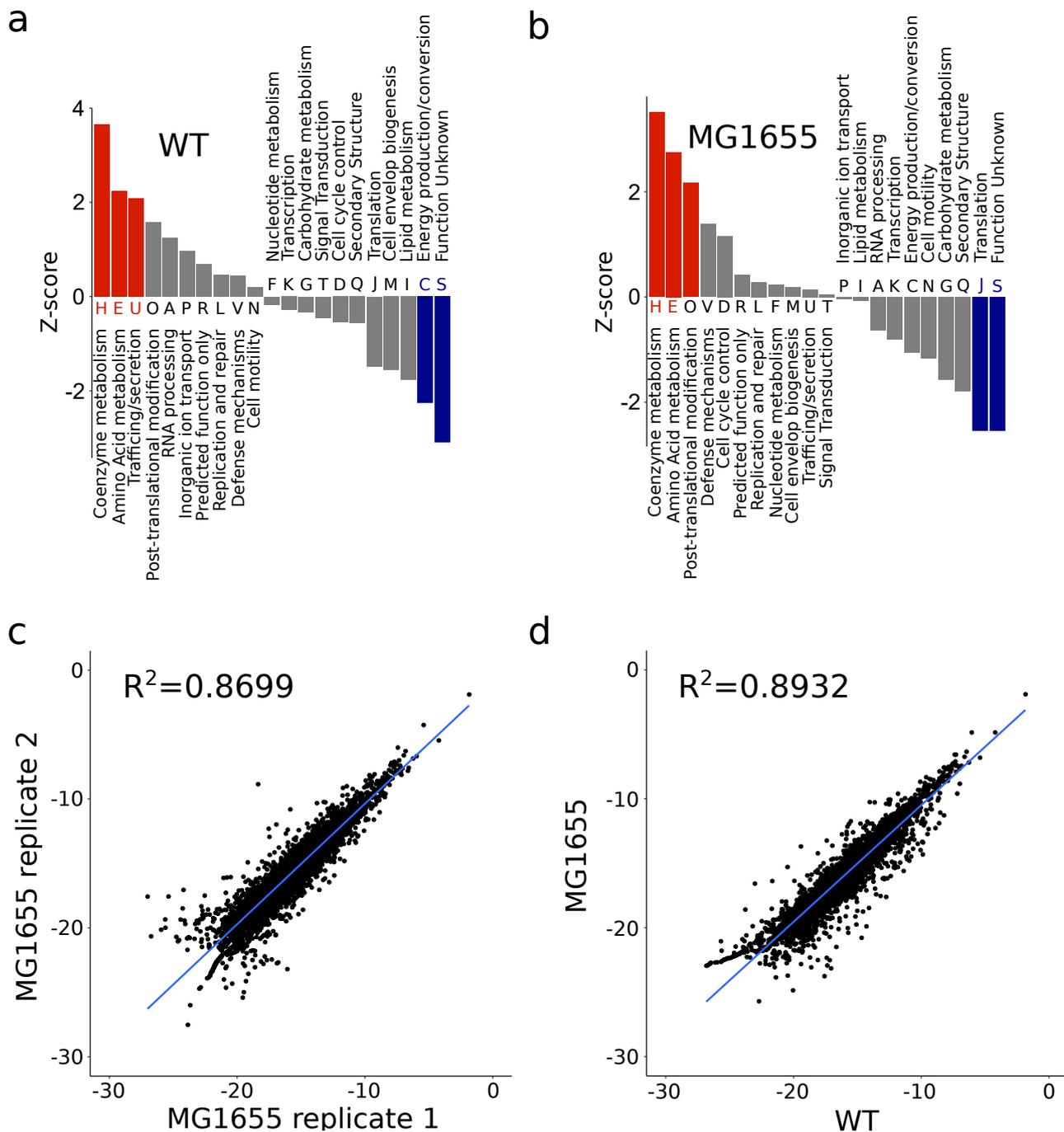CRISPR/Cas9 became an important tool for the genetic manipulation of organisms over recent years. Its simplicity and high efficiency in producing all kinds of chromosomal changes made it the tool of choice for big as well as smaller laboratories. In bacteria, it is used as an efficient counter-selection system against non-recombinant clones in combination with homologous recombination and therefore allows scarless genome editing without the use of metabolic or antibiotic resistance markers.

In this study, the CRISPR/Cas9 tool "CRISPR SWAPnDROP" for bacterial species is described. In addition to standard genetic modifications comprising scarless, marker-free, iterative and parallel insertions and deletions, it also facilitates the transfer of large chromosomal regions between bacterial species. Its versatile range of applications makes it a useful tool for metabolic engineering, synthetic biology and chromosome studies.

Editing efficiencies of CRISPR SWAPnDROP were shown to be up to 100% in *lacZ* and *araB* scarless reconstitution experiments for *E. coli*. Multiplex genome editing was demonstrated with the parallel editing of up to four genes with an average efficiency of 98% and 83% for dual-edits and quadruple-edits, respectively. By the excision of a 151kb chromosomal region and subsequent transfer from one E. coli strain to another, the principle of large chromosomal DNA transfer was proven. The concept of CRISPR SWAPnDROP includes a modular design for the construction of the systems' core plasmid harbouring the core components of the genome editing tool (Cas9 and a recombination system). This simplifies the adaption of the CRISPR SWAPnDROP system for new bacterial species. By exchanging and testing different replication origins, promoters, antibiotic resistance cassettes and recombination systems using this modular approach, CRISPR SWAPnDROP could be established for the model organism *E. coli*, the biotechnology relevant *Vibrio natriegens* and the plant pathogen *Dickeya dadantii*. Furthermore, the exchange of small and large chromosomal regions between those species was demonstrated. The lac operons of *E. coli* MG1655 and DH5*α* were transferred and integrated into the *V. natriegens* chromosome. As a result, *V. natriegens*

was able to metabolise lactose as the only carbon source and it was possible to perform a blue/white screen in this organism. Additionally, the RP4 conjugation system was transferred from *E. coli* and integrated into the chromosomes of *V. natriegens* and *D. dadantii*, generating strains of these organisms capable of conjugational transfer for the first time. To test the interspecies chromosomal transfer to *E. coli*, the ORF of the β-galactosidase GanB from *D. dadantii* was successfully transferred into the ORF of *lacZω* from *E. coli* DH5α.

The carrier plasmid of CRISPR SWAPnDROP harbouring the variable components (e.g. sgRNA, insert DNA, homology arms) was also designed in a modular approach and is compatible with existing modular cloning libraries. This enables the user to exchange the variable parts and simply combine them in a new manner. Furthermore, the multi-color scarless co-selection system significantly improves the editing efficiency and provides visual quality controls throughout the assembly and editing process.

Patrick Sobetzko, Marc Teufel and Carlo Klein conceived and conducted the construction of the plasmids. Experiments for the establishment of CRISPR SWAPnDROP in *V. natriegens* were conceived by Marc Teufel and conducted by Maurice Mager. Editing efficiency and multiplex experiments for *E. coli* were conceived by Marc Teufel and conducted by Maurice Mager in consultation with Patrick Sobetzko. Further experiments for *E. coli* and *V. natriegens* were conceived and conducted by Marc Teufel. Next-generation sequencing analysis was conducted by Marc Teufel in consultation with Patrick Sobetzko. Experiments in *D. dadantii* were conceived and conducted by Marc Teufel and Patrick Sobetzko. The manuscript was written by Marc Teufel and Patrick Sobetzko.

ARTICLE

Check for updates

A multifunctional system for genome editing and large-scale interspecies gene transfer

Marc Teufel[1], Carlo A. Klein [1], Maurice Mager [1] & Patrick Sobetzko [1✉]

CRISPR SWAPnDROP extends the limits of genome editing to large-scale in-vivo DNA transfer between bacterial species. Its modular platform approach facilitates species specific adaptation to confer genome editing in various species. In this study, we show the implementation of the CRISPR SWAPnDROP concept for the model organism *Escherichia coli*, the fast growing *Vibrio natriegens* and the plant pathogen *Dickeya dadantii*. We demonstrate the excision, transfer and integration of large chromosomal regions between *E. coli*, *V. natriegens* and *D. dadantii* without size-limiting intermediate DNA extraction. CRISPR SWAPnDROP also provides common genome editing approaches comprising scarless, marker-free, iterative and parallel insertions and deletions. The modular character facilitates DNA library applications, and recycling of standardized parts. Its multi-color scarless co-selection system significantly improves editing efficiency and provides visual quality controls throughout the assembly and editing process.

[1] Philipps Universität Marburg, Synthetic Microbiology Center Marburg (SYNMIKRO), Marburg 35043, Germany. ✉email: patrick.sobetzko@synmikro.uni-marburg.de

I n recent years, the relevance of genome editing in bacteria rapidly increased in basic research, biotechnology and synthetic biology. Novel technologies like CRISPR/Cas9 and large scale DNA synthesis brought genome editing within reach of a broad scientific community. However, large-size modifications, incompatibility of individual tools as well as high-throughput systems are still challenging.

The basic steps of genome editing in bacteria is the introduction of exogenous template DNA in the form of plasmids, ssDNA or dsDNA and the subsequent integration into the chromosome via homologous recombination. Homologous recombination is a very inefficient process, even when improved with recombination systems like λRED. Different strategies have been developed to overcome time-consuming screening for edited cells, e.g. the integration of antibiotic resistance markers or metabolic markers[1–4]. However, the number of possible edits in one cell is limited by the number of available markers as only one edit per marker is possible. In addition, the introduction of additional marker genes into the chromosome can cause undesired interference with adjacent transcription units[5,6]. To circumvent the problem of marker limitation, site-specific recombinases such as Cre- and FLP recombinase are introduced to remove the selection marker after chromosomal integration[7,8]. This adds an additional step to the editing process and leaves active recombination sites in the genome, which eventually limits the application of this approach even with the use of alternative recombination sites[9]. Another strategy to avoid selection markers and other scars is oligo-mediated allelic replacement (OMAR). Short single-stranded (ss) DNA are incorporated into the genome by an Okazaki-like allelic-replacement event at the replication fork and facilitate point mutations, small deletions and insertions[10]. Automation and cyclizing of this method allow genome modifications of several genes[11]. However, OMAR is limited by the size of ssDNA oligos used and therefore larger edits are not possible.

For larger scarless genome editing, counter-selection methods can be applied. Such methods rely on efficient counter-selectors like the meganucleases I-SceI and I-CreI, which introduce double-strand breaks at a specific site of 18 and 22 bp in length[12,13]. The double-strand break leads to cell death of non-edited cells. This strongly enriches the viable population for successfully edited cells. However, the approach requires an additional classical editing step, in which the meganucleases' target site is first integrated into the chromosome at the site of interest.

The discovery of CRISPR/Cas9 abolished the initial integration of a specific restriction site. Similar to a meganuclease, Cas9 is an endonuclease. In contrast to a meganuclease, Cas9 has no DNA sequence specificity. Specificity is mediated by a guide RNA (gRNA) complementary to the target sequence. Possible target sites are only limited by the protospacer adjacent motive (PAM) sequence (5′-NGG-3′), which is required to be located upstream of the target sequence. Due to its simplicity, Cas9 counter-selection is widely used. Plasmid systems harbouring an inducible Cas9 and a recombination system were designed to select for gene deletions, point mutations and short insertions[14,15]. These systems depend on the transformation of synthetic oligonucleotides or linear double-stranded template DNA for homologous recombination and are therefore not suitable for large-size insertion. To overcome size limitations, REXER provides the template DNA on an episomal replicon[16]. Inserts of up to 100 kb are excised in vivo by CRISPR/Cas9 to facilitate λRED recombination. Selection occurs via positive and negative selection markers, which leads to scars at the integration site and requires a previous integration of the first selection-marker set.

For modularization and increased flexibility of CRISPR/Cas9-based methods, modular cloning techniques like Golden Gate Cloning have been employed for sgRNA arrays[17,18]. Furthermore, joining of insert and homology arms as well as the preparation of the locus-specific protospacer require an additional cloning step favouring the concept of modular cloning. Independent of the CRISPR/Cas9 methods, general approaches for modular cloning were developed to assemble larger constructs from small standardized parts, e.g. MoClo, MoCloFlex, Golden Mutagenesis and Marburg Collection[19–22]. However, modular cloning for CRISPR/Cas9 editing has not yet been systematically implemented. The flexibility of modular systems comes at a price. The sequence integrity of the parts from verified plasmid-born DNA libraries is highly stable and needs no additional sequencing after assembly.

Here we present CRISPR SWAPnDROP, a versatile genome-editing system for bacteria. Following a modular concept, CRISPR SWAPnDROP includes a scarless and marker-free system for large-scale insertions, deletions and in vivo chromosomal DNA transfer between strains and even species. For compatibility with automated genome editing, it is capable of multiplexing and iterative genome modifications. Its multi-colour selection system efficiently avoids errors in system assembly and provides a scarless co-selection system for increased editing efficiency.

Results

Concept. DNA synthesis technologies and modular cloning approaches allow for the assembly of large DNA fragments at a scale of bacterial chromosomes. Such large fragments play a role in biotechnology for complex pathway assembly or tailored organism design. Moreover, in synthetic biology and basic chromosome research, rearranged or even completely synthetic chromosomes receive increasing attention[16,23,24]. Such approaches, however, require reliable and versatile handling of large DNA fragments and DNA libraries. We have developed CRISPR SWAPnDROP to meet these requirements. CRISPR SWAPnDROP is based on homologous recombination, CRISPR/Cas9 counter-selection and a scarless multi-colour co-selection system. CRISPR SWAPnDROP provides a framework for the assembly of large genomic sequences, the rearrangement of chromosomal parts as well as the sequence transfer between strains and even to other organisms (see Fig. 1). Furthermore, it comes with a full set of editing tools for convenient iterative and parallel scarless chromosomal deletions and insertions. Supplementary Table 5 provides an overview of the whole workflow regarding days, tasks and expenditure of time.

CRISPR SWAPnDROP is based on homologous recombination of linear double-stranded DNA. It allows for insertion and deletions as well as transfer of large DNA fragments between strains or species (Fig. 1). For insertions and deletions (Indel) two flanking homologies (HA, HB) are used for chromosomal integration of the insert (INS). The fragment HA-INS-HB is released from a plasmid inside the cell by restriction with Cas9. Furthermore, Cas9 is used as a counter-selector using locus-specific sgRNAs expressed from the same plasmid. For the transfer of DNA between cells, chromosomal DNA is loaded onto the plasmid by another set of homologies (H1, H2) matching the flanks of the chromosomal fragment (Swap). Fragment excision and plasmid opening is mediated by Cas9 and sgRNAs specific for the flanks of the chromosomal fragment and the plasmid. After loading the fragment onto the plasmid, it is transferred to another cell via RP4 conjugation or plasmid transformation. In the new host, in analogy to the insertion process, the fragment is released by flanking Cas9 restriction and another pair of homologies (Hα, Hβ) located at the edges of the fragment confer recombination (Drop). The editing process is supported by a colour system based on the deoxyviolacein pathway and the *mScarlet* gene that provides positive feedback of the experimental
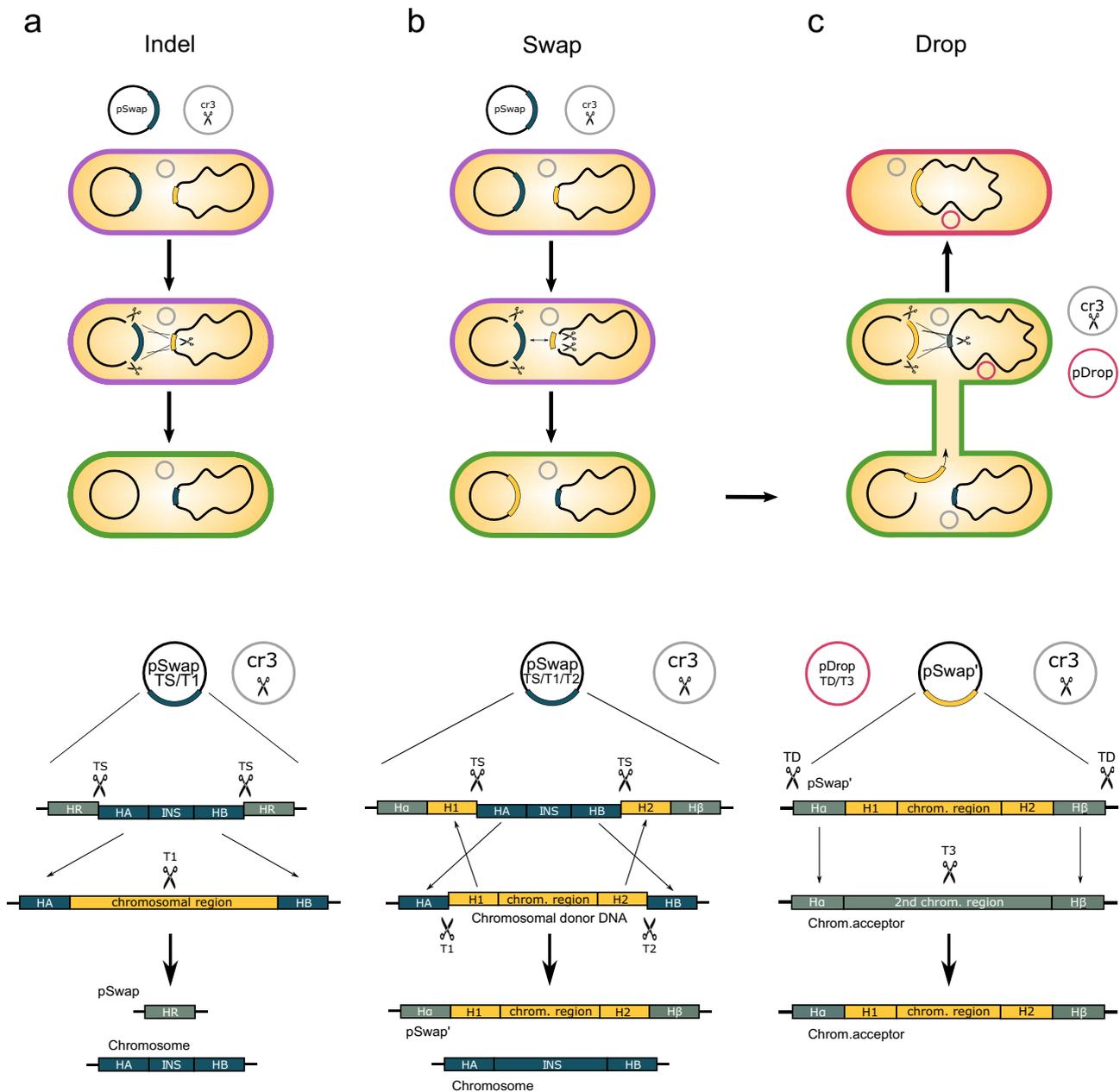
Fig. 1 Overview and mechanism of CRISPR SWAPnDROP. CRISPR SWAPnDROP is a genome editing system based on CRISPR/Cas9 counter-selection, homologous recombination and a multi-colour scarless co-selection. It facilitates scarless insertions and deletions (Indel) (**a**) and is capable of transferring chromosomal DNA between organisms (**b**, **c**). The pSwap plasmid harbours the template for homologous recombination including the homologous regions (HA, HB, HR, H1α, H2β) and the insert DNA (INS), the sgRNA (TS) expression cassettes for the excision of the double-stranded donor template and for the chromosomal counter-selection (T1, T2). The excision is catalysed by CRISPR/Cas9 (Scissors), expressed from the helper plasmid cr3 together with the genes for efficient homologous recombination. For the transfer of chromosomal regions, chromosomal donor DNA (yellow) is additionally excised (T1, T2). The two regions are swapped resulting in a pSwap plasmid loaded with a chromosomal region of interest (pSwap'). The loaded pSwap' plasmid can then be conjugated to a cell harbouring cr3 and the second helper plasmid pDrop, which catalyses the excision of the donor DNA on pSwap' (TD) and the subsequent integration into the acceptor chromosome via counter-selection (T3). Cell border colour represents the multi-colour co-selection system. Cells change their colour from purple to green and from green to red after the Indel/Swap and Drop events, respectively.

state in each step of editing by changing colours (Fig. 1a). Cells harbouring the correct assembled pSwap plasmid produce a purple colour. Cells, which successfully performed Cas9 restriction and homologous recombination during the Indel/Swap and the Drop steps produce a green and red colour, respectively. For CRISPR SWAPnDROP to function properly it is central that recombination and Cas9 restriction works efficiently. Although the CRISPR SWAPnDROP concept described above is generally

applicable, it is apparently not possible to provide a static system that will be functional in many species. Given the diversity of species, central components like origins of replication, resistance cassettes, recombination systems or individual promoters and RBS need to be adapted to the individual species. Therefore, we have designed CRISPR SWAPnDROP in a highly modular fashion to allow for efficient combinatorial screenings for the implementation in new species. In this study, we show the

implementation process for the model organism *Escherichia coli*, the fast-growing marine bacterium *Vibrio natriegens* and the plant pathogen *Dickeya dadantii*. Details about the features, mechanisms, helper plasmids and modularity are provided in the following paragraphs.

cr3—the CRISPR SWAPnDROP workhorse. The cr3 helper plasmid harbours the enzymes required for genome editing. These enzymes comprise the Cas9 endonuclease and a species-specific recombination system. The plasmid itself is assembled via the MoCloFlex system[20]. With the MoCloFlex system, the assembly of up to 5 components in any orientation, order and composition is possible in one step. This allows the user to adapt the cr3 plasmid to different species if necessary by exchanging parts, e.g. origin of replication, antibiotic resistance or the recombination system. In this study, we show the composition of cr3 for *E. coli* (cr3Ec), *V. natriegens* (cr3Vn) and *D. dadantii* (cr3Dd).

pSwap—the modular carrier plasmid. The pSwap plasmid represents the variable component of the CRISPR SWAPnDROP system. It harbours all edit-specific parts including insertion fragments, sgRNAs and homologous regions for the integration and transfer of DNA. The assembly is based on modular cloning and confers edit-specific customization of the pSwap plasmid.

pDrop—the chromosomal transfer helper plasmid. The pDrop plasmid supports the integration of the transferred chromosomal DNA into the new chromosomal locus. It consists of a fixed and a flexible sgRNA, which are responsible for the in vivo excision of the transferred DNA for homologous recombination and the counter-selection for a successful integration, respectively.

Cas9-based selection and recombination. CRISPR SWAPn-DROP can be used for chromosomal modifications such as insertions and deletions (Indels) as well as for the rearrangement and transfer of chromosomal regions between bacterial strains and species (see Fig. 1). For the creation of Indels, the template for homologous recombination is located on the pSwap plasmid consisting of homologous regions HA and HB flanking the desired insert (INS). Induction of the CRISPR/Cas9 and the recombination system leads to the excision of the template and its integration into the chromosome. Recombination of the homologous regions HR, which flank the excised fragment (HA-INS-HB) and have the same nucleotide sequence, recircularize the pSwap plasmid after excision of the template DNA. This is required to maintain plasmid integrity when no loading of the pSwap for DNA transfer is required (see Fig. 1a). Expression of the T1 sgRNA/Cas9 targeting the chromosomal region allows selection for a successful recombination event. Only cells which lost the target site upon integration of the excised fragment survive. For scarless deletions, it is possible to omit the INS fragment leaving only the homologous regions HA and HB. For the transfer of chromosomal regions, the pSwap can also be loaded with chromosomal DNA. In this case, the regions H1 and H2 are homologous to the flanking regions of the chromosomal region to be transferred (see Fig. 1b). Expression of T1 and T2 sgRNA/Cas9 as well as TS sgRNA/Cas9 cause excision of the chromosomal region and the HA-INS-HB fragment, respectively. The recombination system then catalyses the swap of both DNA fragments leading to a loaded pSwap (pSwap') with chromosomal DNA and a chromosome with a deletion or a substitute fragment. Here, the sgRNAs act as excision and selection tools both on the chromosome and the pSwap plasmid. For dropping the loaded

sequence at the desired location, the swap approach can be further extended by an additional set of homologies Hα and Hβ that flank the region of integration. This is supported by the helper plasmid pDrop expressing sgRNAs TD and T3, which are used for the in vivo excision of the loaded DNA and the selection at the insertion locus, respectively (see Fig. 1c).

Modular assembly of the pSwap plasmid. The pSwap plasmid consists of seven locus-specific parts necessary for genome editing, e.g. homologous regions (HA, HB, H1, H2, Hα, Hβ), sgRNA expression cassettes for counter-selection and excision (T1, T2) or insert DNA (INS). Each part has its own vector that can be joined into a pSwap plasmid with the desired parts, allowing for the recycling of parts, e.g. the recycling of homologies and sgRNAs to insert different sequence at the same location, or the recycling of the insert at different locations (see Fig. 2a and b). This approach reduces cloning efforts and supports the storage-efficient implementation of a parts library. Moreover, for the pSwap plasmid, assembled from sequenced parts, no additional sequencing is required. The pSwap assembly follows a simple hierarchical topology similar to other cloning systems[19,20] and is also based on Golden-Gate cloning. Each of the seven modules has its specific level 1 entry vector for cloning the desired parts (see Fig. 2a). In addition, the INS entry vector is level 1 MoClo[19], MoCloFlex[20] and Marburg Collection[22] compatible. Hence, level 1 transcription units build with the MoClo-System or Marburg Collection as well as larger assemblies build with MoCloFlex can be cloned into the INS entry vector. Therefore, already present libraries for these systems can be accessed by CRISPR SWAPnDROP and used for chromosomal integration. Except for the H2, each entry vector contains a high-copy pUC origin of replication, a kanamycin resistance cassette and all contain the *ccdB* and *lacZα* genes for selective cloning[25]. The selective cassette is flanked by two BsaI restriction sites for Golden Gate cloning of the desired DNA fragments. As BsaI is a type-IIS restriction enzyme, cleavage occurs outside of its recognition sequence and therefore corresponding sites are lost upon restriction. This allows cloning in a one-pot, one-step reaction, in which the selective cassette is removed and replaced with the desired DNA fragment. Each of the resulting level 1 vectors contain two additional type-IIS BbsI restriction sites flanking the cloned DNA fragment for the assembly of a level 2 pSwap plasmid. All necessary level 1 modules are cleaved with BbsI and the DNA fragments are assembled via DNA ligase in a one-pot, one-step reaction due to fixed overhangs (see Fig. 2b). In addition to the other level 1 entry vectors, the different versions of the H2 entry vector (Cm^R, Gm^R) contain a single-copy F-origin of replication and a chloramphenicol or gentamycin cassette for the final pSwap. The single-copy H2 plasmid also contains a RP4 origin of transfer for conjugation of large genomic regions. For the scarless integration into the chromosome, HAIB plasmid can be used replacing the modules HA, INS and HB. A standard assembly of pSwap comprises 7 parts. With a growing number of parts in a Golden Gate reaction, the number of correctly assembled plasmids decrease[20]. The selection for a correctly assembled pSwap is guided by co-expression of the deoxyviolacein pathway, whose individual genes are distributed among the parts T1, T2, H1α and HA. A successful assembly results in the formation of purple colonies after transformation (see Fig. 2c, e and Supplementary Fig. 1a). This allows the direct transformation of non-cloning strains that naturally exhibit higher frequencies of incorrect clones. Within this study, all selected purple colonies lead to successful edits indicating correct pSwap assemblies. A detailed plasmid map of pSwap is depicted in Fig. 3.
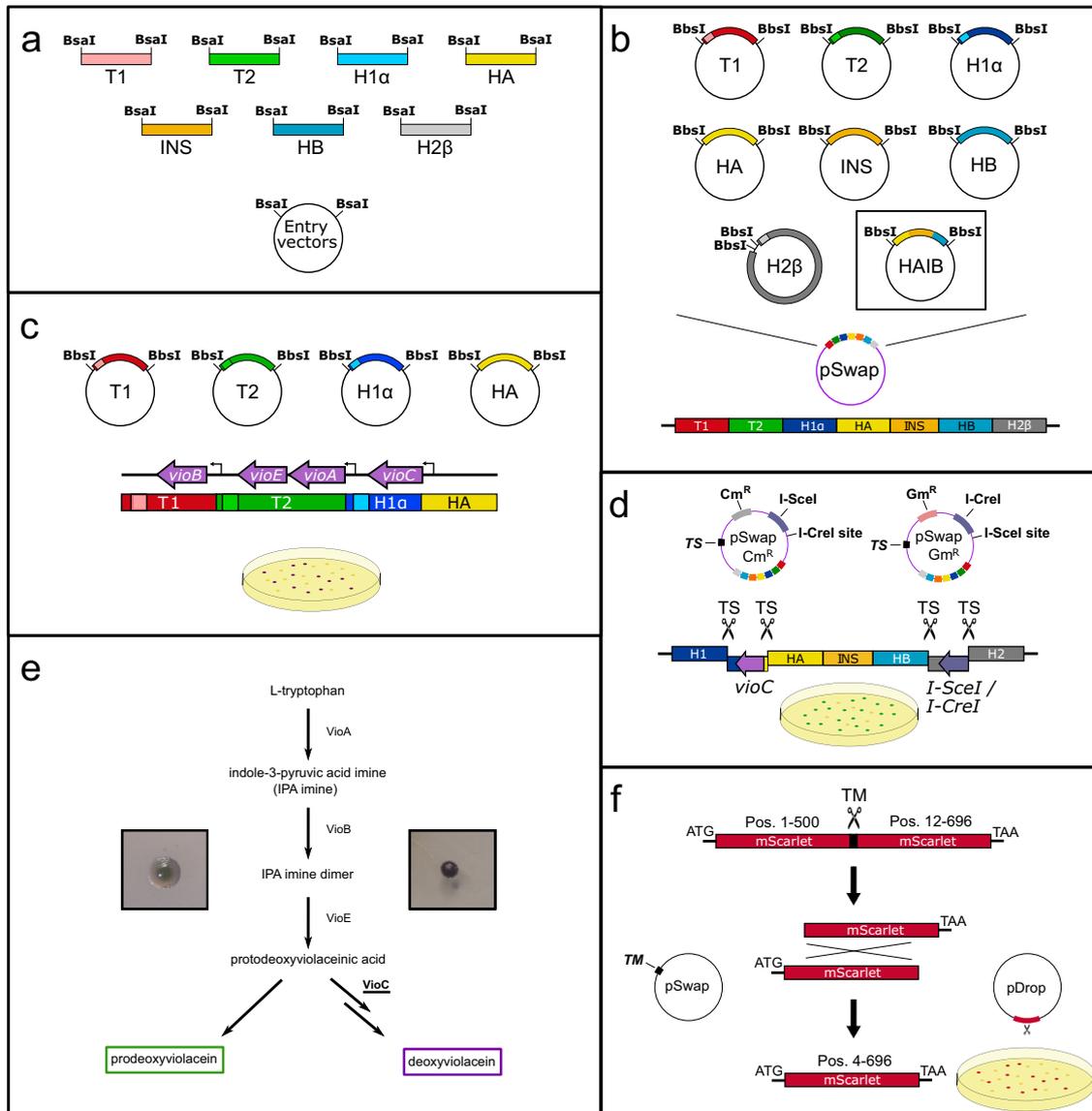
Fig. 2 pSwap assembly and recombination overview. CRISPR SWAPnDROP uses a modular assembly system for the pSwap construction. It consists of seven modules (five for scarless integration) harbouring sgRNA expression cassettes for target sites (T1,T2), homologous regions (Hα, Hβ, H1, H2, HA, HB) and insert DNA (INS). **a** Each of those parts can be cloned into a different entry vector using the type-IIS restriction enzyme BsaI. **b** Each of those seven (five for scarless integration) modules are used to assemble the pSwap plasmid. The assembly takes place in a one-pot, one-step reaction using the type-IIS restriction enzyme BbsI. Therefore, it is possible to combine and recycle different modules for the desired approach. **c** The correct assembly of the pSwap plasmid is ensured by the expression of the biosynthetic pathway of deoxyviolacein. Each of the fragments (T1, T2, H1α, HA) harbour parts of the deoxyviolacein expression cassettes (vioBEAC). Only if assembled correctly, production of deoxyviolacein is possible. Colonies appear purple after transformation. **d** During linearization of the pSwap plasmid by sgRNA TS and Cas9, the vioC gene is removed resulting in the formation of prodeoxyviolacein after successful recombination. Colonies then appear green. In addition, the tool box comprises two pSwap plasmids (Cm[R]/Gm[R]) harbouring either the I-SceI or I-CreI meganuclease genes and the opposite recognition site. The removal of the meganuclease genes during recombination allows iterative use of the pSwap plasmids and therefore consecutive genomic edits. **e** Shown is the biosynthetic pathway of L-tryptophan to deoxyviolacein. If the gene (vioC) for the conversion of protodeoxyviolacein to deoxyviolacein, which is a purple pigment, is missing, a metabolic shift takes place towards prodeoxyviolacein, which presents a green colour. **f** The pDrop plasmid harbours a partly duplicated, non-functional mScarlet gene. During Drop recombination, the pDrop plasmid is cut by the sgRNA TM, which is expressed on the pSwap plasmid and subsequently recombined resulting in a functional mScarlet gene. Colonies appear red after successful recombination.

Design and construction of cr3Ec, cr3Vn and cr3Dd plasmids. The construction of the cr3Ec, cr3Vn and cr3Dd plasmids was carried out using the modular cloning systems MoCloFlex[20] and the Marburg Collection[22]. The Marburg collection was used to assemble individual transcription units from promoter, RBS, CDS, tag and terminator parts in the library. In addition, the library was extended with newly designed parts. From the Marburg collection, the origin of replication as well as resistance marker, cas9 and recombination transcription units were transferred into position vectors AB, CD, EF and IJ of the MoCloFlex (MCF) system, respectively. The position vectors together with MCF linkers BC, DE, FI and JA allowed us to assemble a set of cr3 variants, each harbouring different combinations of origins of replication, resistance cassettes and induction systems for
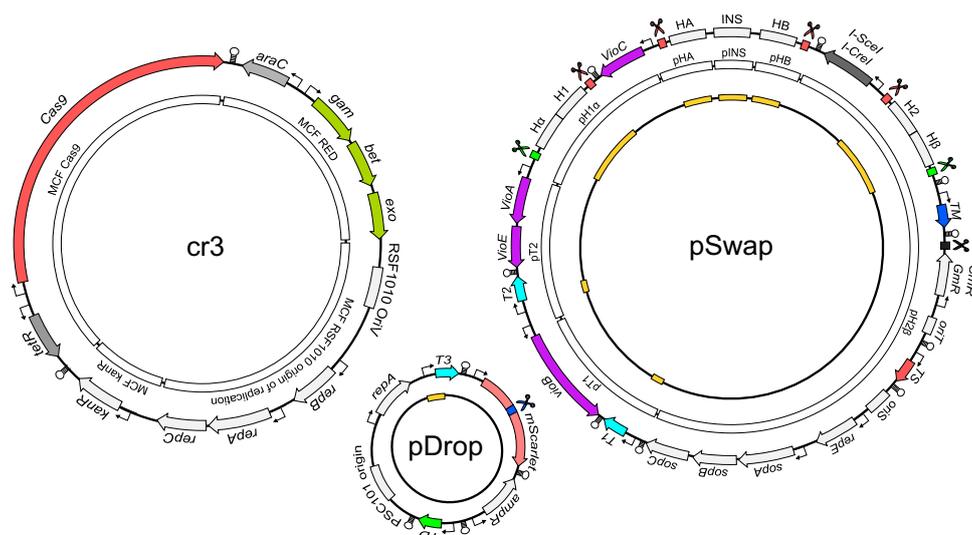
Fig. 3 Detailed map of the plasmids cr3Ec, pSwap and pDrop. The outer layer contains relevant features of each plasmid. The white inner layer (if applicable) represents the modular parts prior to assembly using MoCloFlex (MCF) or the CRISPR SWAPnDROP assembly system. The black inner layer (if applicable) shows the location of variable elements (orange) of the plasmids specific for each edit. Cas9 and meganuclease target sites are indicated by scissors. The sgRNA loci and the respective target sites share the same colour.

CRISPR/Cas9 and homologous recombination. The variants were then screened for compatibility with the targeted organism and the other helper plasmids pSwap and pDrop. For the construction of the cr3Ec, the broad host range origin of replication RSF1010 (MCF RSF1010 origin of replication), a kanamycin-resistant cassette (MCF KanR), Cas9 transcription unit (MCF Cas9) and a λRED transcription unit (MCF RED) (see Fig. 3) was assembled. In accordance with Reisch and Prather[14], the Cas9 transcription unit was assembled using an inducible tetracycline promoter (P_{tet}), a weak ribosomal binding site, which was integrated as a level 0 part into the Marburg Collection library, the Cas9 coding sequence, the M0051 ssrA degradation tag and the B0015 terminator. For the Cas9 coding sequence, Esp3I, BbsI and BsaI recognition sites were removed by introducing silent mutations to confer compatibility with the Marburg Collection, MoCloFlex and CRISPR SWAPnDROP Golden Gate systems. For efficient homologous recombination in *E. coli* and *V. natriegens*, the λRED system was used. Its transcription unit was assembled using the promoter part harbouring the P_{BAD} and *araC* of *E. coli*, the Gam, Beta, Exo coding sequence and the B0015 terminator[22]. For the construction of the cr3Vn, the broad host range RSF1010 as well as a codon-optimized kanamycin resistance gene were combined, which were known to ensure plasmid stability in *V. natriegens*[22]. In earlier studies, phage-derived or native recombination systems were used to confer efficient homologous recombination in *V. natriegens*[26,27]. In this study, we modified the *E. coli* λRED operon by setting it under the control of the native *V. natriegens* P_{BAD} promoter to support homologous recombination. The repression was conferred by the native *araC* gene located on the chromosome. For the construction of the cr3Dd, we screened for a species-specific recombination system as no functional recombination system was yet known for *D. dadantii*. This was done by a sequence homology search for known recombination systems in bacteria. In the *D. dadantii* strain Yana2-2[28], we identified a prophage containing an operon consisting of five genes where two genes showed high amino acid homology to a recombination system present in *Salmonella enterica*. The genes comprised *bet* (89% identity), *exo* (92%) identity. These two genes also showed homology to *bet* and *exo* of λRED (bet: 84% and exo: 86%). The third gene in the operon was identified to be a DNA methyl transferase. Genes four and five were short predicted genes with

unknown function. Before transfer of the recombination system to CRISPR SWAPnDROP, we screened for resistance cassettes, origins of replications and induction systems in *D. dadantii* using the modular MoCloFlex approach already applied for *E. coli* and *V. natriegens*. The RSF1010 origin of replication together with the kanamycin resistance cassette of cr3Vn turned out to be also stable in *D. dadantii*. For the induction systems, we chose the cr3Ec arabinose and tetracycline inducible systems. Functionality was verified by mCherry and mVenus fluorescence reporters instead of Cas9 and the recombination system on CR3Dd. Induction and background fluorescence levels were similar to *E. coli*. However, transformation of cr3Dd with Cas9 and the recombination system was not feasible. Replacement of either Cas9 or the recombination system by its fluorescence reporter parts indicated a problem with the recombination system. Therefore, we designed shortened versions of the recombination operon: bet-exo, bet-exo with the methyl transferase and the full operon as control. Again the full operon was not transferable. However, both shorter versions could be transferred. In the next step, we tested induction to see whether toxic levels of induction could occur. For the bet-exo version, no toxic effects were detected. For the bet-exo with methyl transferase, a strong reduction of viable cells was detected for full induction and also with lower levels. During efficiency determination, it turned out that only the tetp-bet-exo version was functional. Hence, the final cr3Dd harbours this recombination cassette.

Enhanced editing efficiency by multi-colour co-selection. CRISPR SWAPnDROP enhances editing efficiency by a multi-colour scarless co-selection strategy. This co-selection strategy selects for functional CRISPR/Cas9 and recombination systems, thus decreasing suppressor mutations (see Supplementary Table 6). Consequently, editing efficiency is increased without scars or the trade-off of additional chromosomal markers[29]. The CRISPR SWAPnDROP co-selection is based on the elimination of the *vioC* gene from the pSwap plasmid by CRISPR/Cas9-mediated excision and subsequent recombination upon induction (see Fig. 2d). The lack of *vioC* leads to a metabolic shift towards prodeoxyviolacein and, in turn, to the formation of green colonies (see Figs. 2e and 4a). To assess the gene-editing efficiency of CRISPR SWAPnDROP two *E. coli* genes, *lacZ* and *araB*, were
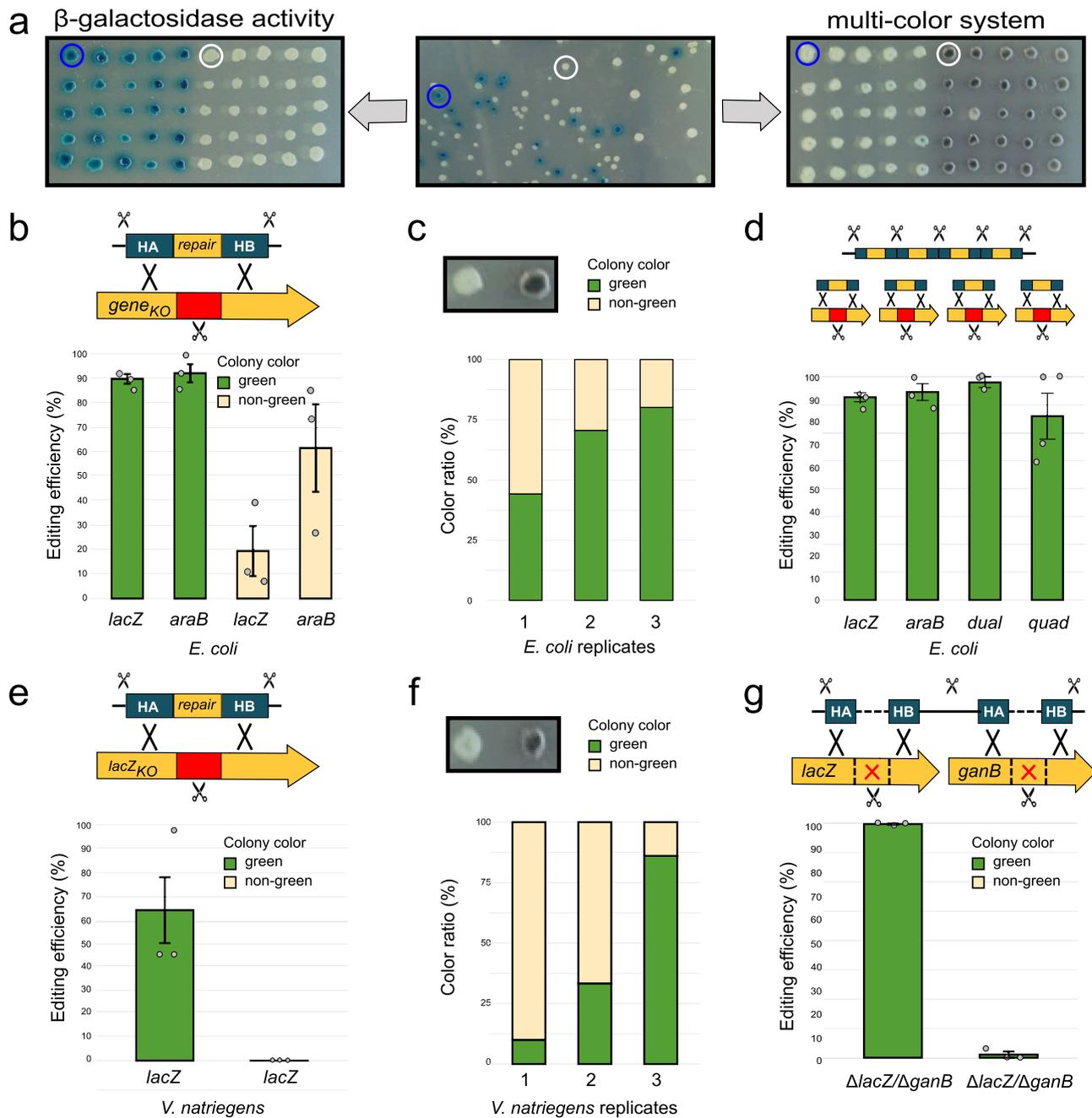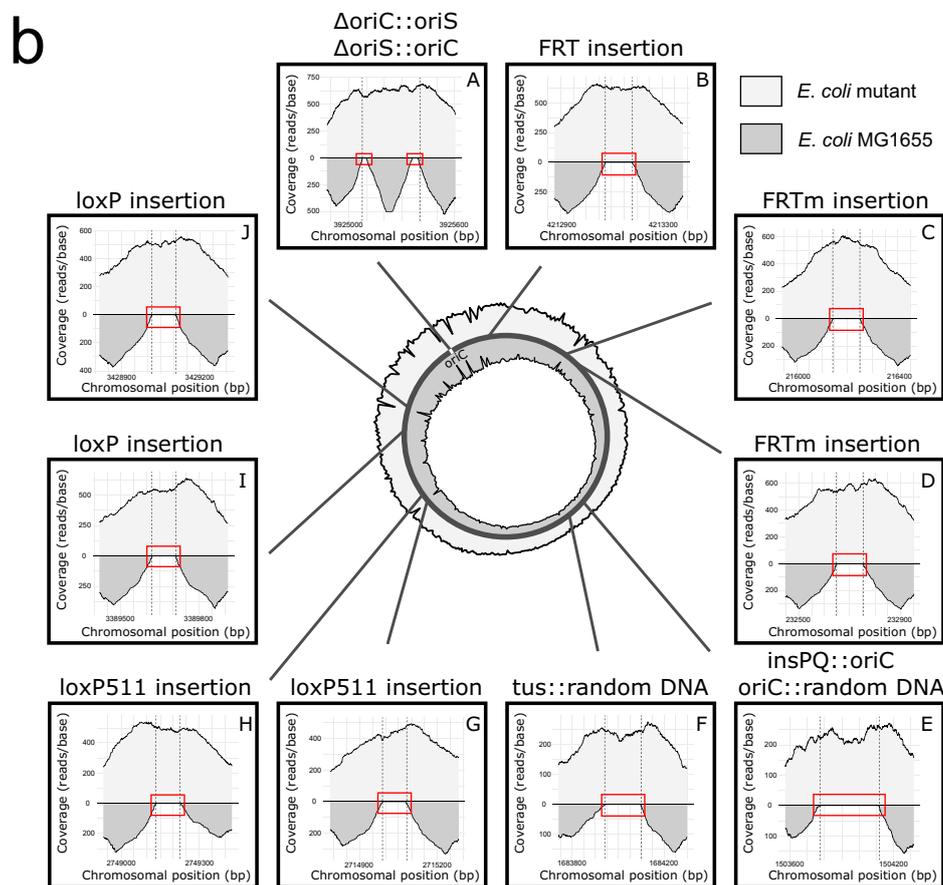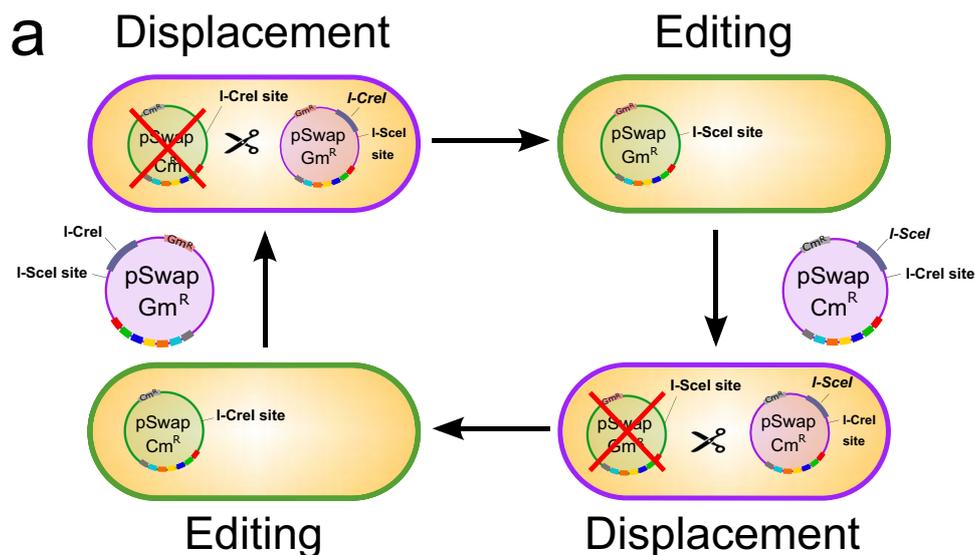
Fig. 4 Editing efficiency of CRISPR SWAPnDROP in three species. a Test of the multi-colour system by a *β*-galactosidase knockout in *Dickeya dadantii*. In the left panel, *β*-galactosidase positive and negative colonies are picked after editing (centre panel) and grown at 37 °C on LB agar supplemented with IPTG and X-Gal. The same colonies were also transferred to a second plate and grown on LB at 30 °C to develop the green colour (right panel). Circles indicate identical colonies on three plates. White colonies in the left and centre panels indicate successful editing. **b** Shown is the editing efficiency of green and non-green colonies for the reconstituted *lacZ* and *araB* genes in *E. coli*. Experiments were carried out in triplicates. 45 green and 45 non-green colonies were tested in total for both *lacZ* and *araB* reconstitution. Data are presented as mean values ± SEM. **c** Shown is the colour distribution of colonies in three different *lacZ* reconstitution experiments. **d** Shown is the editing efficiency of green colonies of single and multiplex editing. Error bars represent the standard error. Experiments were carried out in triplicates, except for the quadruple edit, which was carried out in quadruplicates. For dual editing (*lacZ* and *araB*), 45 colonies and for the quadruple edit (*lacZ*, *araB*, *xylA* and *dapA*), 40 colonies were tested in total. Data are presented as mean values ± SEM. **e** Shown is the editing efficiency of green and non-green colonies for the reconstituted *lacZ* gene in *V. natriegens*. Experiments were carried out in triplicates. In total, 43 green and 32 non-green colonies were tested. Data are presented as mean values ± SEM. **f** Shown is the colour distribution of colonies in three different *lacZ* reconstitution experiments in *V. natriegens*. Error bars represent the standard error. **g** Shown is the editing efficiency of green and non-green colonies for the double knock-out of *lacZ* and *ganB* genes in *D. dadantii*. Experiments were carried out in triplicates. 75 green and 75 non-green colonies were tested in total. Data are presented as mean values ± SEM. Source data are provided as a Source data file.

reconstituted and phenotypically tested. For that, two *E. coli* knockout strains were first generated each lacking either a functional *lacZ* or *araB* gene by deleting parts of the corresponding coding sequence and introducing Cas9 target sites. Deleting only parts of the coding sequence renders the genes non-functional. At the same time it ensures that the reconstituting fragment on its own does not restore gene function in the subsequent reconstitution step. In this way, neither the pSwap plasmid containing the reconstituting fragment nor an off-target insertion into another locus causes false-positive results. The genes *lacZ* and *araB* encode for the β-galactosidase and ribulokinase, both essential for the lactose and arabinose metabolism, respectively. Therefore, both strains were not able to grow on M9 minimal medium supplemented with lactose or arabinose, respectively. Two pSwap plasmids (pSwap lacZ/pSwap araB) harbouring either the missing part of *lacZ* or *araB* (INS) flanked by corresponding homologous regions HA and HB as well as sgRNAs (T1), specific for the deletion region, were then assembled and transformed into the *lacZ* and *araB* knock-out strains. After induction of the CRISPR/Cas9 and λRED systems, green and non-green colonies were spotted on M9 plates supplemented with either lactose or arabinose. Clones, which scarlessly integrated the missing coding sequence into their genome, were again able to grow on the corresponding saccharide. Clones were verified exemplarily by PCR and Sanger sequencing. For *lacZ* and *araB* reconstitution, green colonies attained an editing efficiency of up to 100% with low variation, while non-green colonies resulted in an editing efficiency of about 20% and 60% with a higher variation, respectively (see Fig. 4b). In addition, the colony colour distribution in different *lacZ* reconstitution experiments revealed a high variation of the green to non-green ratio (see Fig. 4c). In some experiments, the number of green colonies dropped to a few counts on the whole plate whereas editing efficiency of green colony remained high. The frequency of green colonies depends on the timing of suppressor mutant emergence and subsequent outgrowth and is, therefore, random and not controllable. The earlier the mutation during growth, the higher the percentage of suppressors on the plate. Consequently, for some experiments, the total efficiency would drop to a low percentage, due to accumulation of suppressor mutants. However, within the pool of green colonies, the editing efficiency is significantly higher, very stable between replicates and not correlated to the frequency of green colonies. Hence, the multi-colour system of SWAPnDROP increases and stabilizes the expected editing efficiency of CRISPR/Cas9 gene editing. Analogous to the in vivo pSwap cleavage of the green-colour selection for the Indel and Swap recombination, a red-colour selection was implemented for the Drop recombination. The pDrop plasmid harbours a partly duplicated, non-functional *mScarlet* gene, which is cleaved by CRISPR/Cas9 during the Drop. The cleavage is mediated by the sgRNA TM (target mScarlet) located on the pSwap plasmid (see Fig. 2f). Homologous recombination of the overlapping sequences results in a functional *mScarlet* gene and therefore in the formation of red colonies (see Fig. 2f and Supplementary Fig. 1c, d). For a detailed plasmid map of pDrop see Fig. 3.

Fast iterative genome editing using alternating plasmids. The CRISPR SWAPnDROP system comes with two complementary versions of the pSwap plasmid: pSwap Cm^R and Gm^R. The first harbours a constitutively expressed *I-SceI*, a *I-CreI* recognition site and a Cm^R cassette. The second harbours a constitutively expressed *I-CreI*, a *I-SceI* recognition site and a Gm^R cassette (see Fig. 5a). During recombination events in the editing process, the meganuclease genes are removed from the pSwap plasmid by CRISPR/Cas9 excision together with the *vioC* gene (see Fig. 2d).

The resulting green edited clones carry a pSwap' plasmid only harbouring the recognition site of the other meganuclease (see Fig. 2d). Transformation of a fresh pSwap plasmid cures the clone from the old pSwap by expression of the meganuclease located on the fresh pSwap and allows for another round of editing (see Fig. 5a). Functionality was tested in *E. coli* MG1655 wild type for 12 consecutive genomic inserts at 10 distinct loci using alternating pSwap Cm^R and pSwap Gm^R plasmids. During these 12 edits no problem in plasmid curing or evident functional loss of the editing system was observed. During the functionality test, different recombination sites (FRT/loxP) as well as random DNA and origins of replication (F-plasmid oriS/native *E. coli* oriC) were inserted and replaced throughout the genome of *E. coli* MG1655 (see Fig. 5b). The use of alternating pSwap plasmids facilitates one round of editing every 3 days. This is particularly useful in case of the absence of a suitable PAM site. A new site can be introduced by a first edit in close proximity and the desired edit can be done efficiently in a consecutive edit reverting the first edit and introducing the desired modification. Moreover, the approach not only facilitates and speeds up consecutive rounds of edits, but also makes CRISPR SWAPnDROP compatible with automated editing approaches.

Multiplex genome editing. To assess the efficiency of CRISPR SWAPnDROP multiplex genome editing, the two *E. coli* genes, *lacZ* and *araB*, were reconstituted simultaneously. In a first step, an *E. coli* strain was generated by two consecutive gene knock-outs of *lacZ* and *araB* via CRISPR SWAPnDROP. Consequently, the strain was unable to grow on M9 plates supplemented with lactose or arabinose. In a second step, a pSwap for parallel repair was assembled. The pSwap plasmid contained the two repair templates (HA*lacZ*-INS*lacZ*-HB*lacZ*-TS-HA*araB*-INS*araB*-HB*araB*) separated by an additional Cas9 excision site (TS) as well as the sgRNAs (T1, T2) targeting the non-functional *lacZ* and *araB* deletion sites (see Supplementary Fig. 2). The repair templates were cloned via Golden Gate cloning in pHAIB using 6 PCR fragments (HA*lacZ*, INS*lacZ*, HB*lacZ*-TS, HA*araB*, INS*araB*, HB*araB*). The additional Cas9 target site (TS) separates the templates for each locus of separate recombination via λRED. After induction of CRISPR/Cas9 and λRED, green colonies were transferred to M9 plates containing either lactose or arabinose. Consequently, only colonies with successfully reconstituted *lacZ* and *araB* genes were able to grow on both plates. The editing efficiency of the dual-repair approach was 98%, which means almost all clones restored their ability to grow on both saccharides. Even though, CRISPR SWAPnDROP is currently not able to generate more than two simultaneous edits due to its limitation of two sgRNAs, we wanted to assess the multiplex editing potential of larger numbers of parallel edits for future extensions of the system by sgRNA arrays[18]. Therefore, two additional genes, *xylA* and *dapA*, were disrupted in the *lacZ* and *araB* double mutant. *xylA* encodes for a xylose isomerase and is essential for the xylose catabolism, while *dapA* encodes for a dihydrodipicolinate synthase, which is essential for cell wall synthesis. Disrupting those genes leads to a strain, which is DAP auxotroph and not able to use xylose as a carbon source. The resulting strain contained the four non-functional genes *lacZ*, *araB*, *xylA* and *dapA*. The lack of two sgRNAs was compensated by introducing the same Cas9 target sites for later repair in *xylA* and *dapA* knockouts as used for the *lacZ* and *araB* knockouts respectively. This allowed us to simultaneously modify four different loci, while only using two sgRNAs (T1, T2). In analogy to the double repair, the pSwap plasmid used for the reconstitution contained the four repair templates, each separated by an additional Cas9 target site (TS). Green colonies were tested on their ability to grow solely on lactose, arabinose and xylose each supplemented with DAP as

well as on glucose without DAP. On average, 83% of the tested clones were able to grow on all saccharides as well as without DAP (see Fig. 4d). This shows the high editing efficiency of CRISPR SWAPnDROP even for up to four simultaneous edits and supports future extensions of the system by sgRNA arrays.

Transfer of large chromosomal regions. CRISPR/Cas9-mediated excision of double-stranded DNA and homologous recombination generally permit the handling and alteration of large DNA fragments in vivo[16]. CRISPR SWAPnDROP extends the handling of large DNA fragments to a systematic excise and insert.

Coupled with the modular assembled pSwap and pDrop plasmids, CRISPR SWAPnDROP provides a tool to rearrange the chromosome and transfer large chromosomal DNA between bacterial strains. As a proof of concept, we transferred and integrated a 151 kb chromosomal fragment from one E. coli strain into another (see Fig. 6a). For the sequence to be transferred we chose a chromosomal region (del4), known to cause little impact on strain fitness upon deletion[30] (see Supplementary Fig. 3). Using CRISPR SWAPnDROP, we first generated a E. coli wild-type knockout strain of the del4 region as acceptor for the region. The del4 knock-out avoids possible toxic effects and homology

Fig. 5 Iterative genome editing with CRISPR SWAPnDROP. a The use of the two pSwap plasmids (Cm^R/Gm^R) allows iterative genome editing. Each harbouring either the *I-SceI* or *I-CreI* meganuclease genes and the opposite recognition site I-CreI site or I-SceI site, respectively. During the editing, the meganuclease genes are removed leaving the pSwap plasmid only with the opposite recognition site. Subsequent transformation of the other pSwap plasmid leads to the elimination of the previous plasmid and allows for another round of editing. Cell colours represent the corresponding editing step. **b** Next-generation sequencing of *E. coli* mutant after 12 consecutive edits. Shown is the next-generation sequencing (NGS) coverage of a *E. coli* mutant strain after 12 consecutive edits (light grey) compared to the wild-type *E. coli* MG1655 (dark grey). For the mutant strain, several recombination sites (FRT/loxP) as well as random DNA and origins of replication were inserted into the chromosome using iterative CRISPR SWAPnDROP genome editing. Mutant and wild-type *E. coli* NGS reads were aligned against the mutant reference genome and the sectors of each edited site (I–X) as well as the complete genome coverage (circle) are shown. Reads at all insertion locations (dashed lines) are present for the mutant strain, while no reads are present for the wild-type strain (red rectangle). (I) and (V) show the coverage of a chromosomal location after 2 different edits at the same position. For (I), the native origin of replication *oriC* of *E. coli* was first replaced with the F-plasmid derived *oriS* origin of replication. *oriS* was then replaced again with *oriC* while introducing scar sites flanking the *oriC*. Scar sites were introduced due to using HA, INS and HB plasmids instead of HAIB plasmid. The absence of those flanking scars in the wild-type is highlighted with red rectangles in (I). For (V), *oriC* was first inserted into the *insPQ* locus and then replaced with a random sequence. The absence of the random sequence in the wild-type strain is highlighted with a red square.

issues due to an additional copy of the del4 region within the *E. coli* chromosome during and after transfer and facilitates transfer validation. For that, *E. coli* MG1655 was first transformed with cr3Ec. A pSwap plasmid was assembled harbouring homologies HA and HB flanking the del4 region as well as a short random insert sequence, which eventually replaced the deleted region. After deletion, the *E. coli* MG1655 Δdel4 was transformed with a pDrop plasmid (pDrop del4) harbouring the sgRNA for selection upon reintegration. A pSwap (pSwap del4) was then assembled and transformed into the *E. coli* RP4 conjugation strain MFD*pir* to load the del4 region onto the plasmid. The del4 region of *E. coli* MFD*pir* is about 151 kb in size and a bit smaller than the deleted *E. coli* MG1655 del4 region due to a ~16 kb deletion within the region. The loaded pSwap' plasmid was conjugated to *E. coli* MG1655 Δdel4 and the 151 kb region was integrated into the Δdel4 chromosomal region (*E. coli* MG1655 Δdel4::MFDdel4). Colony pcr revealed an integration efficiency of up to 60% and on average around 40% (45 clones were tested). Then, pcr and sanger sequencing was used for verification of each strain. As seen in Fig. 6b, we were able to amplify DNA from the del4 region in *E. coli* MG1655 as well as in *E. coli* MG1655 Δdel4::MFDdel4, while in *E. coli* MG1655 Δdel4 those bands did not appear on the electrophoresis gel. The 3 kb PCR band spanning the del4 region appeared only in *E. coli* MG1655 Δdel4, as for the strains still harbouring the del4 region the PCR product would have been above 167 kb. To distinguish between *E. coli* MG1655 and *E. coli* MG1655 Δdel4::MFDdel4, homologies Hα and Hβ were chosen to delete small regions flanking the del4 region after integration. DNA bands of lower size from PCRs spanning this small deleted region indicate the integration of the MFDdel4 region into the Δdel4 mutant chromosome (see Fig. 6b). To rule out that specific bands originate from the pSwap' plasmid, the strain was tested phenotypically and via PCR for the loss of the plasmid. The strain was unable to grow on the corresponding antibiotic and no plasmid specific amplicons were detected. *E. coli* MG1655 Δdel4::MFDdel4 is a chimera of the MG1655 and the *E. coli* MFD*pir* at the del4 locus. To verify the chimeric nature of the strain, we performed next-generation sequencing (NGS) of the *E. coli* MG1655 Δdel4::MFDdel4 (see Fig. 6c). NGS reads of *E. coli* MG1655 Δdel4::MFDdel4 and *E. coli* MG1655 were mapped against the *E. coli* MG1655 reference genome. While for *E. coli* MG1655, the del4 region is 167 kb in size, about 16 kb are missing in the del4 region of *E. coli* MFD*pir*. Consistent with a successful edit, reads of the 16 kb deletion as well as the deleted del4 flanking regions were missing, while reads of the complete MFDdel4 region were present in *E. coli* MG1655 Δdel4::MFDdel4.

Interspecies gene transfer and genome editing in *V. natriegens*. The fast-growing gammaproteobacterium *V. natriegens* has the

potential to rise as an important organism in biotechnology and molecular biology. In recent years, this organism was made genetically accessible, but still lacks a working CRISPR/Cas9-based gene editing system[31,32]. Using MoCloFlex, we adapted CRISPR SWAPnDROP to *V. natriegens*. In contrast to the pSwap plasmid, the cr3Ec plasmid was not compatible with *V. natriegens*. Hence, we generated a new cr3 version called cr3Vn (see the 'Design and construction of cr3Ec, cr3Vn and cr3Dd plasmids' section). To test the system, a pSwap plasmid (pSwap recJ) was assembled containing the selective sgRNA (recJ) and homology arms HA and HB of ~1.2 kb and ~1.4 kb in length, which flank the exonuclease gene *recJ*. Inactivation of *recJ* was shown to enhance natural transformation in other *Vibrio* species[33]. Therefore, the *recJ* locus was chosen with regard to possible future cloning applications. *V. natriegens* was transformed with both cr3Vn and pSwap recJ and purple colonies were used for induction of the Cas9 and λRED genes. Subsequent colony PCR and sanger sequencing revealed *recJ* knock-out clones in 23% of the tested clones (3 out of 13). After showing editing activity of CRISPR SWAPnDROP in *V. natriegens*, we systematically determined the editing efficiency and optimised the induction protocol (see the 'Methods' section). In analogy to the quantification of the editing efficiency in *E. coli*, we intended to use a *lacZ* knockout repair approach. For that, we needed a knockout *lacZ* variant of a functional lac operon. As *V. natriegens* has no functional lac operon, a native *lacZ* knockout was no option[34]. Consequently, we planed to transfer the *E. coli* Δ*lacZ* variant used for *E. coli* to *V. natriegens*. First, we transferred the native *E. coli* lac operon to *V. natriegens* to test its functionality. For the transfer, we assembled a pSwap plasmid with homology arms (H1α, H2β) and sgRNAs (T1, T2) flanking the *E. coli* lac operon (pSwap lacVn). The lac operon was loaded onto the pSwap', transferred to *V. natriegens* via conjugation and subsequently dropped into the *recJ* locus (see Fig. 8a). Without additional adaptations to the lac operon, *V. natriegens* was able to process X-gal to form blue colonies upon induction with IPTG. In the next step, we transferred the lac operon of the *E. coli* Δ*lacZ* knockout strain to *V. natriegens* harbouring cr3Vn and pDropVn recJ and dropped it into the *recJ* chromosomal locus. For the selective sgRNA of pDropVn recJ, the spacer sequence of the initial *recJ* knockout test was used. For the transfer, we reused pSwap lacVn initially designed to allow for a drop into the *recJ* of *V. natriegens*. Integration was verified by PCR and Sanger sequencing of the transitional region of the insert fragment and the flanking chromosomal region. pSwap plasmid elimination was verified phenotypically by the inability to grow on corresponding antibiotics. With the resulting *V. natriegens* lac$_{Ec}$ Δ*lacZ* strain, repair was performed using the *lacZ* repair pSwap (pSwap lacZ) already used in *E. coli* genome editing efficiency
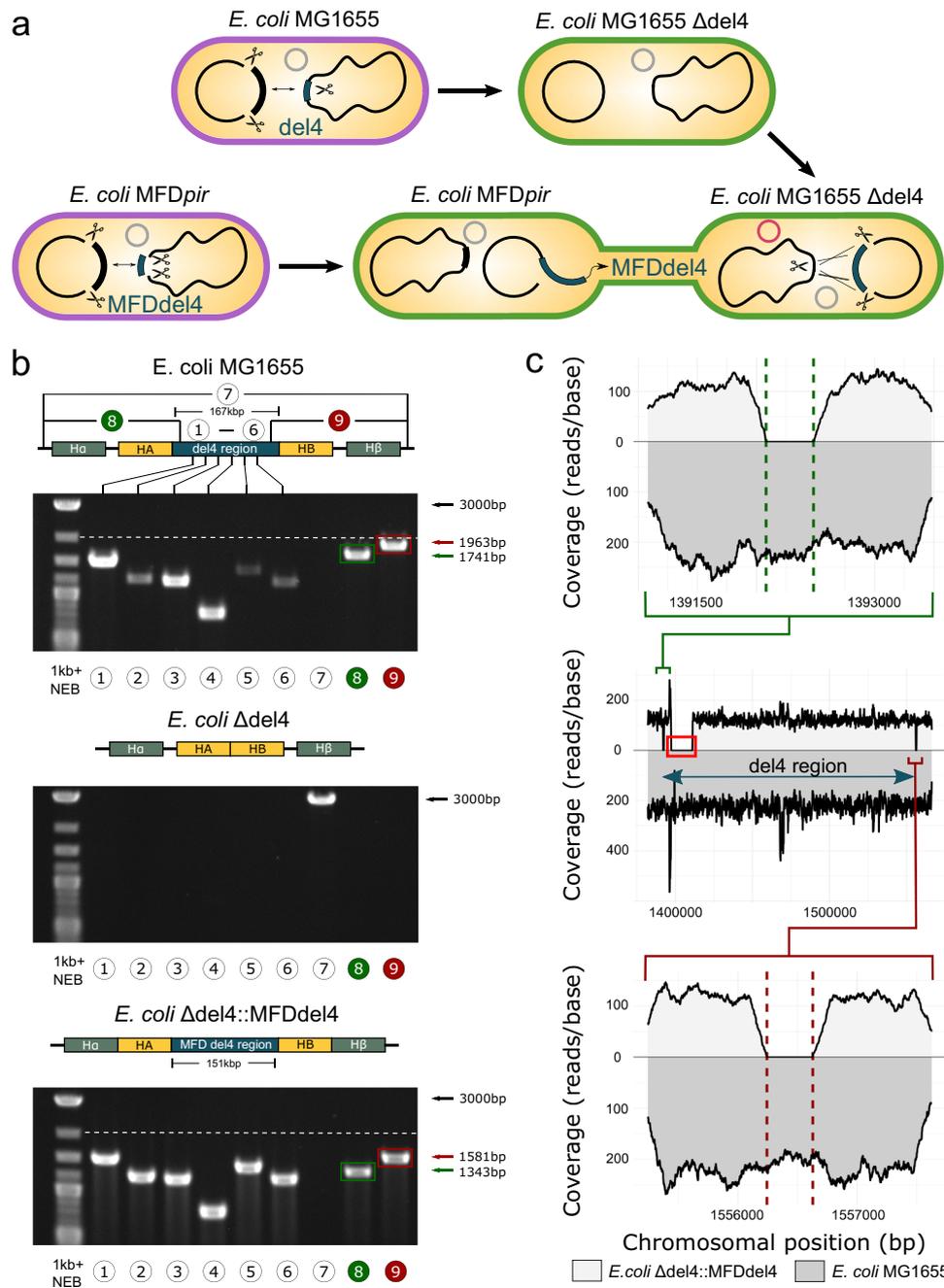
Fig. 6 Transfer and integration of a 151 kb chromosomal region via CRISPR SWAPnDROP. a Scheme of the deletion of the 167 kb del4 chromosomal region in *E. coli* MG1655 (top). Swap and conjugational transfer of the 151 kb MFDdel4 chromosomal region from *E. coli* MFD*pir* to *E. coli* MG1655 Δdel4 and subsequent integration (bottom). Cell colours represent the corresponding editing step. **b** Shown are agarose gels of PCRs from the integration of a 151 kb chromosomal region in *E. coli*. PCRs were done for *E. coli* MG1655 (acceptor precursor), the acceptor (*E. coli* MG1655 Δdel4) and the edited strain (*E. coli* MG1655 Δdel4::MFDdel4). PCRs "1–6" result in the amplification of fragments of the del4 region evenly distributed over the 167 kb (*E. coli* MG1655) and 151 kb (*E. coli* MFD*pir*), respectively. PCR "7" result in the amplification of a 3000 bp fragment, if the del4 region is missing. PCR "8" and "9" amplify the transition of the del4 region to the adjacent chromosomal regions. Upon integration of the 151 kb MFDdel4 region into the *E. coli* MG1655 Δdel4 chromosome, small regions flanking the 151 kb region were deleted resulting in DNA bands of lower size compared to the wild-type chromosome. This is additionally highlighted with the colours green and red. Dashed line indicates the band size of a 2000 bp DNA fragment for a better comparison. PCR was repeated once with identical results. Chromosomal overviews including homologous regions (HA, HB, Hα, Hβ of each PCR are shown above each agarose gel. **c** Shown is the next-generation sequencing (NGS) coverage (reads/base) of *E. coli* MG1655 Δdel4::MFDdel4 (light grey) and *E. coli* MG1655 (dark grey) mapped against the *E. coli* MG1655 reference genome. The middle plot shows the coverage of the del4 region. For *E. coli* MG1655 Δdel4::MFDdel4, reads for the 16 kb deletion, only present in the integrated MFDdel4 region, are missing (red rectangle). The upper and lower plots showing the detailed coverage of the del4 flanking regions. Upon integration of the MFDdel4 region, small flanking regions were deleted, shown as missing reads between the dashed lines.

determination. Green and white colonies were examined for their ability to form blue colour upon growth with IPTG and X-gal, which implies a successful editing event. The editing efficiency of the *lacZ* repair in *V. natriegens* was up to 92% and on average around 65% (see Fig. 4e, f). Interestingly, no non-green colonies appeared blue upon growth with IPTG and X-gal indicating an enhanced editing efficiency of green colonies also in *V. natriegens*. In addition, we tested alpha-complementation in *V. natriegens*. In analogy to the previous lac operon transfer, we transferred the lac operon of DH5α harbouring only the *lacZω* fragment, capable of alpha-complementation. After transfer and integration into the *recJ* locus by CRISPR SWAPnDROP, the strain was transformed with PICH41308[19], constitutively expressing the *lacZα* fragment. Blue colonies on plates containing IPTG and X-gal indicated a successful α-complementation (see Fig. 8b). We then tested the transfer between species and subsequent integration of a large DNA fragment. For this purpose, we transferred the 151 kb *E. coli* MFDdel4 region to *V. natriegens* and integrated it into the *recJ* chromosomal locus. Analogous to the swap and drop of this region between two *E. coli* strains, we first assembled the same pSwap as for the *E. coli* swap, but exchanged the Hα and Hβ regions (pSwap del4Vn) to target the *recJ* locus. Afterwards, the MFDdel4 region was loaded onto the pSwap plasmid in *E. coli* and subsequently conjugated to the *V. natriegens* strain harbouring cr3Vn and pDropVn recJ. After dropping the 151 kb region into the *recJ* locus of *V. natriegens*, colony PCR revealed an efficiency of 90%. Plasmid specific PCR and the inability to grow on corresponding antibiotics indicated the elimination of the pSwap' plasmid. As seen in Fig. 7a, pcr from *V. natriegensrecJ*::MFDdel4 showed DNA bands of the del4 region (see Fig. 7a "1–6") as well as of the transition between the del4 and the adjacent chromosomal regions (see Fig. 7a "8–9"). In addition, amplification of the *recJ* locus was no longer possible (see Fig. 7a "7"). In order to verify the integrity of the region, we performed next-generation sequencing (NGS). NGS reads of *V. natriegensrecj*::MFDdel4 and *V. natriegens* wild-type (ATCC 14048) were mapped against the *V. natriegensrecj*::MFDdel4 reference genome (see Fig. 7c). Data analysis revealed that indeed reads for the complete 151 kb MFDdel4 region were present in the *recJ* locus of *V. natriegensrecj*::MFDdel4, while they were absent in *V. natriegens* wild-type (see Fig. 7b). After successful transfer of a large region to *V. natriegens*, we transferred the RP4 conjugation system from *E. coli* MFDpir to *Vibrio natriegens* into the recJ locus to allow future transfer of large DNA regions from *V. natriegens* to other species (see Supplementary Fig. 4a). In *E. coli* MFDpir, RP4 is flanked by two resistance cassettes. These were excluded from the transfer to avoid undesired antibiotic resistances in the final *V. natriegensrecJ*::RP4 strain. After transfer and integration analogous to the MFDdel4 region, chromosomal integration was verified by whole-genome sequencing (see Supplementary Fig. 4b). Conjugation proficiency was tested by transformation of *V. natriegensrecJ*::RP4 with a pSwap plasmid and consecutive conjugational transfer of the plasmid to *E. coli* DH5α (see Supplementary Fig. 4c).

Establishing a lactose/galactose pathway in *V. natriegens*. The *V. natriegens* lac_Ec strain created during editing efficiency determination was able to process X-gal. Therefore, we further analysed the lactose metabolism. Growth experiments indicated that *V. natriegens* lac_Ec was able to grow on M9 minimal medium with lactose as its sole carbon source (see Fig. 8a, c), even with a slightly higher growth rate ($\mu = 0.67 \, h^{-1}$) compared to growth on glucose ($\mu = 0.50 \, h^{-1}$). A similar phenomenon was observed for the disaccharide sucrose compared to glucose[34]. Growth rates on

rich medium and glucose were slightly reduced compared to wild type. This was also observed in pure Δ*recJ* strains of the initial knockout experiments. Utilization of the second downstream product of lactose degradation, galactose, was strongly improved ($\mu = 0.59 \, h^{-1}$) compared to wild type ($\mu = 0.08 \, h^{-1}$). In order to exclude secondary mutations in *V. natriegenslac_Ec* as a source of these changes, we removed the lac operon using CRISPR SWAPnDROP. The resulting Δ*lac_Ec* strain showed the loss of lactose metabolisation and the strongly reduced wild type level of galactose utilization[34]. It retained the slightly reduced growth rates on glucose and rich medium of a *recJ* mutant. Hence, the improved galactose utilization is directly connected to the presence of the *E. coli* lac operon. To link the observation to a single gene in the *lac* operon, we performed single knockouts for each gene (see Fig. 8d). The knockouts of *lacA* and *lacI* showed no deviating phenotype. The *lacZ* knockout resulted in the inability to grow on lactose but had no impact on growth rates on galactose. Finally, the knockout of the lactose transporter *lacY* abolished growth on lactose and restored the low growth rates of *V. natriegens* on galactose. In *E. coli*, *lacY* is known to transport lactose but also has a limited affinity for galactose[35,36]. Hence, the *E. coli lacY* may compensate the suboptimal transport of galactose in *V. natriegens*.

Interspecies gene transfer and genome editing in *D. dadantii*. With the successful implementation of CRISPR SWAPnDROP for *E. coli* and *V. natriegens* we investigated the possibility to establish CRISPR SWAPnDROP for the plant pathogen *Dickeya dadantii* formerly known as *Erwinia chrysanthemi*. It is the causative agent of bacterial stem and root rot affecting potatoes and other crops. For this organism no effective genome editing was feasible due to the lack of a functional recombination system. We identified a phage-derived recombination systems in the *Dickeya dadantii* Yana2-2[28] strain using a homology-based approach (see 'Design and construction of cr3Ec, cr3Vn and cr3Dd plasmids'). Furthermore, we tested resistance cassettes, origins of replication and induction systems for functionality in *D. dadantii* to assemble cr3Dd. In a first test, using the newly assembled cr3Dd and a pSwap targeting the *lacZ* locus of *D. dadantii*, we were able to generate a *lacZ* knock-out with high efficiency. As *D. dadantii* harbours two β-galactosidase genes, *lacZ* and *ganB*, phenotypic verification analogous to gene editing in *E. coli* and *V. natriegens* was not possible at this stage. We turned to a multiplex approach to knock-out *lacZ* and *ganB* in parallel. All three replicates showed a 100% editing efficiency for green colonies and only 2% for white colonies (see Fig. 4a, g). Hence, editing efficiency was high and the multi-colour system was functional in *D. dadantii*. For the test of gene transfer, we decided to transfer the RP4 conjugation system from *E. coli* MFDpir to the *D. dadantii lacZ* locus to generate a strain for future transfer of *D. dadantii* genes to other organisms. In analogy to the other two species, homologies for RP4 and the lacZ locus as well as target sites for Cas9 were designed and pSwap/pDrop plasmids were assembled. As for the transfer and insertion of the RP4 conjugation system in *V. natriegens*, the flanking resistance cassettes of the *E. coli* MFDpir RP4 region were excluded. In the first step, RP4 was loaded onto the pSwap plasmid using the *E. coli* induction protocol. Green colonies were verified to harbour RP4 by flanking PCRs and conjugated to *D. dadantii* already harbouring cr3Dd and pDrop lacZ. Induction of the Drop system in *D. dadantii* followed the editing protocol for *D. dadantii*. Red colonies were screened for the edit and a positive clone was used for whole-genome sequencing (see Fig. 9a). Activity of the RP4 conjugation system in the *D. dadantii*
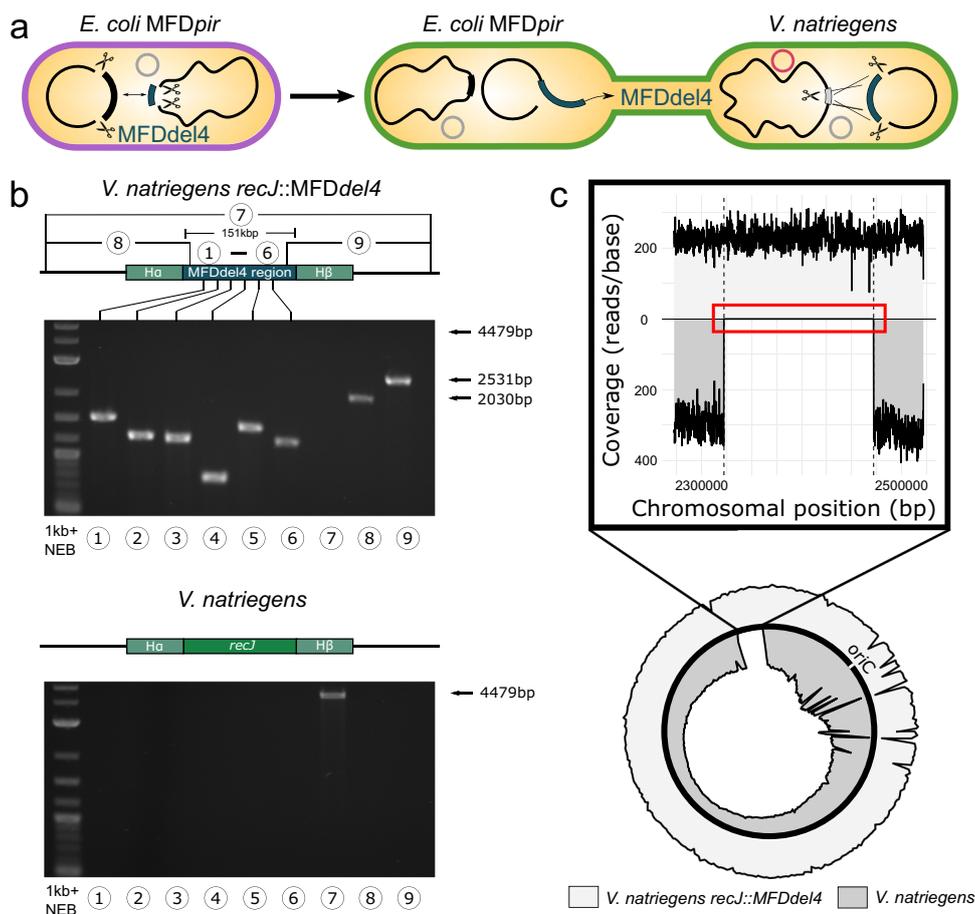
Fig. 7 Integration of a 151 kb _E. coli_ chromosomal region into the _V. natriegens_ chromosome. a Scheme of the swap and conjugational transfer of the 151 kb MFDdel4 chromosomal region from _E. coli_ to _V. natriegens_ and subsequent integration. Cell colours represent the corresponding editing step. **b** Shown are agarose gels of PCRs from the transfer and integration of a 151 kb chromosomal region from _E. coli_ into the _recJ_ locus of _V. natriegens_. PCRs were done for the acceptor strain (_V. natriegens_) and for the edited strain (_V. natriegensrecJ_::MFDdel4). PCRs "1–6" result in the amplification of evenly distributed fragments of the _E. coli_ del4 region within the _V. natriegens_ chromosome. PCR "7" results in the amplification of a 4479 bp fragment, if the _recJ_ locus is unchanged. PCR "8" and "9" amplify the transition of the del4 region to the adjacent chromosomal regions. PCR was repeated once with identical results. Chromosomal overviews including homologous regions (Hα, Hβ of each PCR are shown above each agarose gel. **c** Shown is the next-generation sequencing (NGS) coverage of the edited strain _V. natriegensrecJ_::MFDdel4 (light grey) compared to the reference _V. natriegens_ strain (dark grey). Both NGS reads were aligned against the edited reference genome (_recJ_::MFDdel4). The sector of the del4 insertion as well as the complete genome coverage (circle) are shown. Reads of the complete del4 insertion (between dashed lines) are present for the edited strain, while no reads are present for the wild-type strain (red rectangle).

lacZ::RP4 strain was tested by transfer of a pSwap plasmid from _D. dadantii_ to _E. coli_ DH5α (see Fig. 9b). The established conjugation system allows for direct transfer of DNA from _D. dadantii_ to other species (see Fig. 9c). To test interspecies chromosomal DNA transfer to _E. coli_, the ORF of the second β-galactosidase GanB was selected for scarless transfer into the ORF of _lacZω_ of the DH5α strain. All parts were designed and assembled in analogy to previous transfers. Swap transfer and drop was performed according to _D. dadantii_ and _E. coli_ protocols. Transfer was verified phenotypically as well as by PCR and Sanger sequencing (see Fig. 9d–f). As DH5α lacks the α subunit of LacZ, it is not able to process X-gal to form blue colonies. The replacement of _lacZω_ ORF by the _ganB_ ORF of _D. dadantii_ restored the blue colony phenotype upon IPTG induction with supplemented X-Gal. Compared to the native _lacZ_, X-Gal processing of GanB is lower and takes more time to develop the full-colour intensity. As the delay was not observed in the _D. dadantii_ _lacZ_ knockout strain only harbouring the GanB β-galactosidase, this may be linked to a suboptimal codon usage or activity of the GanB enzyme in _E. coli_.

Discussion

With CRISPR SWAPnDROP we present a versatile scarless and marker-free genome editing system able to perform indels consecutively or in parallel and transfer chromosomal regions between species independent of size and with high editing efficiencies. In this study, we implement the CRISPR SWAPnDROP concept and test various features for the three species _E. coli_, _V. natriegens_ and _D. dadantii_. With its multi-colour co-selection system we introduce an approach to co-selection not relying on chromosomal insertions or other persistent scars and at the same time increasing editing efficiencies and monitoring the assembly process of the modular CRISPR SWAPnDROP system. With its high editing efficiency of above 90% for _E. coli_, CRISPR SWAPnDROP matches the current CRISPR/Cas9-based editing systems[14,15,37] and integrates desirable features of several systems with high flexibility (see Supplementary Table 7).

CRISPR SWAPnDROP is also capable of multiplexed genome editing. We could show, that even with four parallel edits, the editing efficiency remains well above 80%. This suggests the
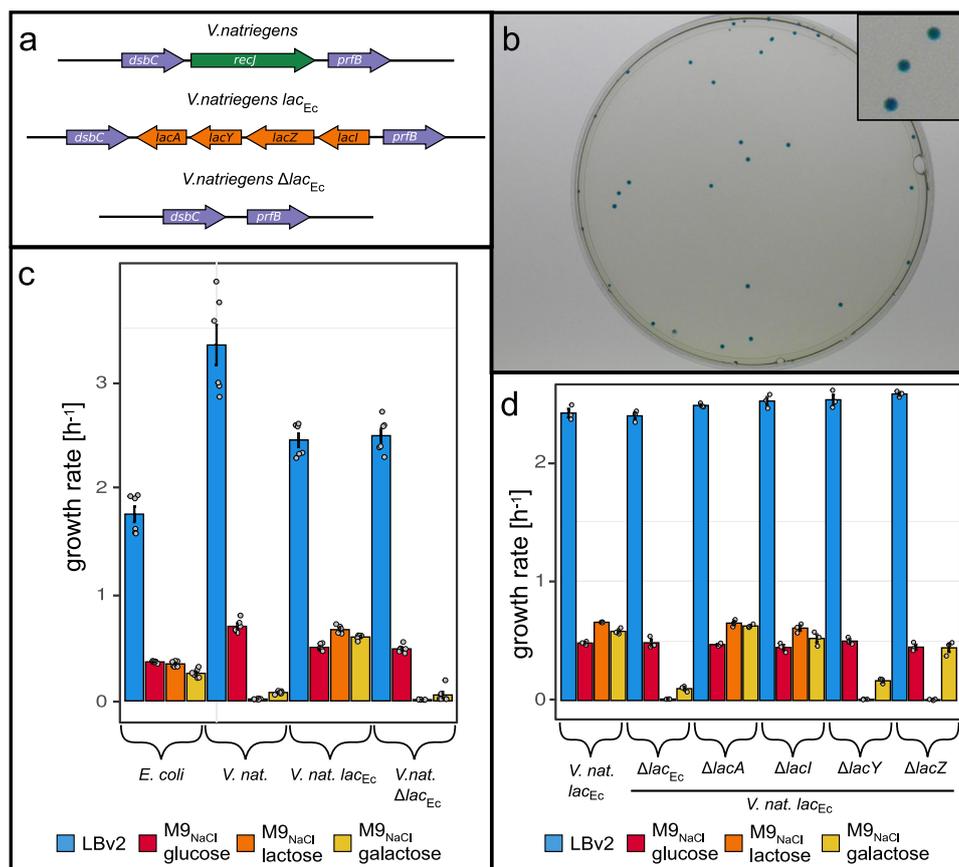
Fig. 8 Functionality of the *E. coli* lactose operons in *V. natriegens*. a Shown is the gene map of the *V. natriegens* strains tested for lactose metabolism. The native *recJ* gene is replaced by the *E. coli* lac operon in *V. natriegens* lac_Ec. This strain was further edited by the elimination of the lac_Ec operon resulting in *V. natriegens* Δlac_Ec, which is also a Δ *recJ* knock out. **b** Shown is a picture of *V. natriegens* colonies, harbouring the *E. coli* DH5α lac operon, transformed with a plasmid containing the *lacZα* fragment. Cells were plated on LBv2 agar supplemented with IPTG and X-gal. **c** Shown are the growth rates of *V. natriegens* strains and *E. coli* grown in different media and carbon sources. For each sample, six independent replicates were performed. Cells were grown in M9 minimal medium supplemented with either 0.4% glucose (red), lactose (orange) and galactose (yellow) as well as in LB (*E. coli*)/LBv2 medium (*V. natriegens*) (blue) using a microplate reader. Data are presented as mean values ± SEM. **d** Shown are the growth rates of *V. natriegenslac_EC* strain and knockouts of the lac operon and of each gene in the lac operon grown in different media and carbon sources. For each sample, three independent replicates were performed. Cells were grown in M9 minimal medium (+2% NaCl) supplemented with either 0.4% glucose (red), lactose (orange) and galactose (yellow) as well as in LB (*E. coli*)/LBv2 medium (*V. natriegens*) (blue) using a microplate reader. Data are presented as mean values ± SEM. Source data are provided as a Source data file.

integration of sgRNAs arrays that have shown to perform up to 10 parallel edits[17,18] in future extensions.

Its modular design renders integration of these arrays possible without alteration of essential concepts, e.g. by replacing the T1 and T2 modular plasmids by a sgRNA array plasmid. Another aspect of the modular design is the application of libraries for chromosomal integration. In the insert plasmid INS, sequence libraries can be cloned resulting in a broad variety of inserts for a specific location. Furthermore, CRISPR SWAPnDROP is compatible with current cloning systems such as MoClo, MoCloFlex and the Marburg Collection[19,20,22]. This grants access to already established large DNA libraries and facilitates integration of CRISPR SWAPnDROP to present library systems. DNA libraries can be applied in protein modification or pathway optimization and can be directly tested at native chromosomal loci without intermediate plasmid constructs. The modular assembly of the key plasmids cr3 and pSwap simplifies adaptation to other species and facilitates updates with new Cas or recombination systems[38,39]. Using MoCloFlex[20] and Marburg collection[22] systems, CRISPR SWAPnDROP was implemented for *E. coli*, *V. natriegens* and *D. dadantii*.

Although the earliest common ancestor of *E. coli* and *V. natriegens* is situated at the phylogenetic level of class (Gammaproteobacteria), only a few parts carrying the resistance cassettes and induction systems needed to be modified to adapt CRISPR SWAPnDROP from *E. coli* to *V. natriegens*. For even more distantly related species, more adaptations may be necessary. In general, for implementing CRISPR SWAPnDROP in a new species, species-specific parts (e.g. promoters, origin of replication and resistance cassettes) need to be chosen. Concerning Cas9 activity, it has been shown that Cas9 is active in all kingdoms of life including animals, fungi, plants, bacteria and archea[40–43]. Therefore, we can assume Cas9 activity in a wide range of species. However, recombination systems are more species-specific. Therefore, for each species, a functional recombination system has to be present. Such recombination systems were found in all kingdoms of life[44–46] and could be incorporated into CRISPR SWAPnDROP to adapt it to new species. Recombination systems are usually found in species-specific phages/viruses or in the organism itself. In this study, this approach is shown for *Dickeya dadantii* for which no functional heterologous recombination system was known. We identified a phage recombination system
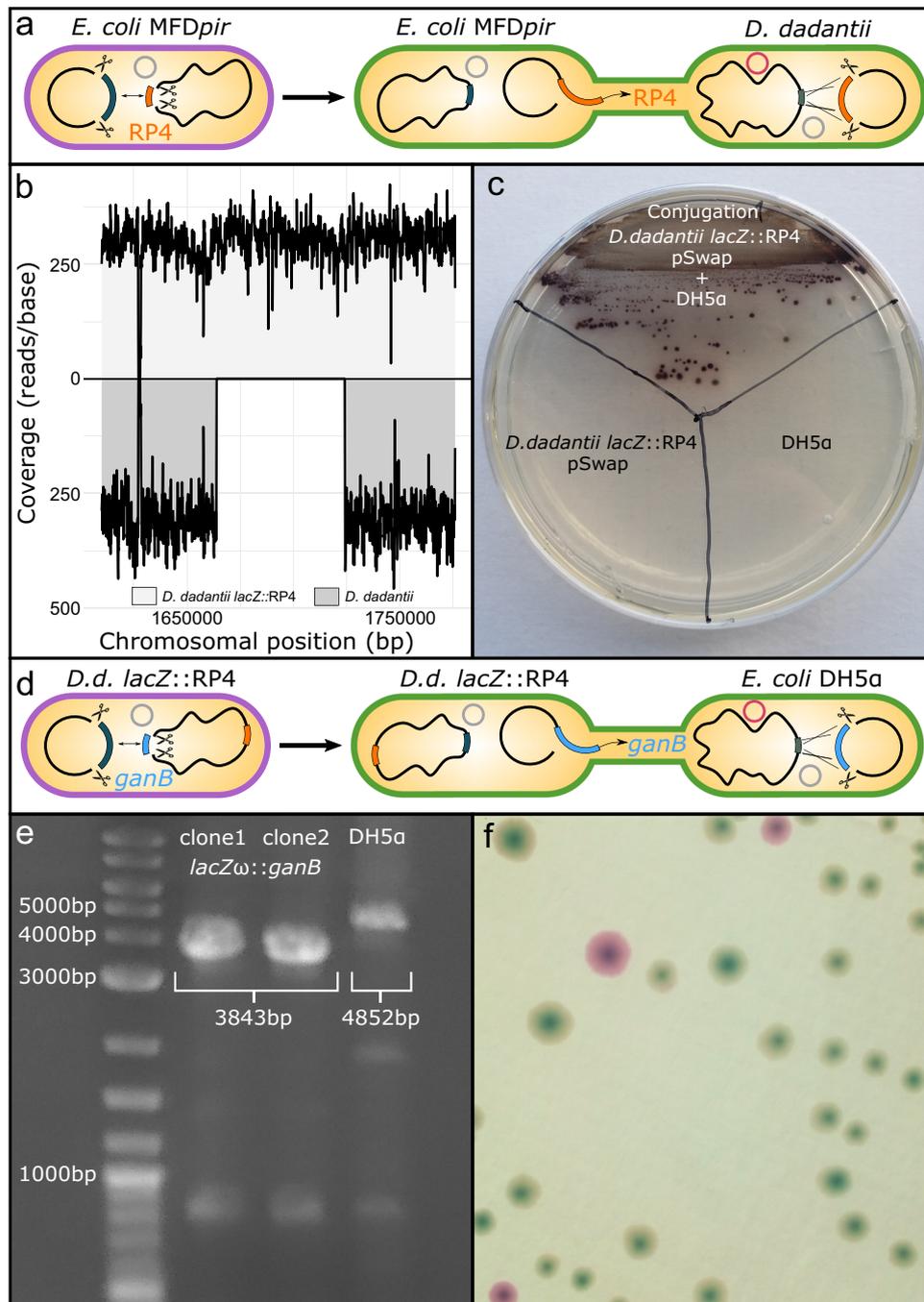
Fig. 9 Transfer of the RP4 conjugation system to *D. dadantii* and transfer of the *β*-galactosidase GanB ORF to *E. coli* DH5*α*. a Scheme of the swap and conjugational transfer of the RP4 conjugation system from *E. coli* MFDpir to *D. dadantii* and subsequent integration. Cell colours represent the corresponding editing step. **b** Shown is the next-generation sequencing (NGS) coverage (reads/base) of *D. dadantii lacZ*::RP4 (light grey) and *D. dadantii* (dark grey) at the RP4 insert locus mapped against the *D. dadantii lacZ*::RP4 reference genome. **c** Conjugation test of pSwap from *D. dadantii lacZ*::RP4 to *E. coli* DH5*α*. Controls for donor and acceptor strains are indicated. **d** Scheme of the swap and conjugational transfer of the *ganB* ORF from *D. dadantii* (D.d.) *lacZ*::RP4 to *E. coli* DH5*α* and subsequent integration. Cell colours represent the corresponding editing step. **e** PCR verification of the Drop of *D. dadantii ganB* ORF in the *E. coli* DH5*α lacZω* ORF. Two clones after Drop compared to DH5*α*. PCR was repeated once with identical results. **f** Blue colony phenotype of DH5*α lacZω*::*ganB* upon induction with IPTG and X-gal. The pDrop plasmid is still present in some colonies of the plated non-selective overnight culture resulting in red colonies.

in a related strain via sequence analysis and tested different variants of this system using the modular design of cr3. With the resulting cr3Dd and a functional recombination system, we were able to facilitate gene transfer and gene editing in a new species. Furthermore, present recombination systems can be rendered functional in other species by additional expression of missing or incompatible components like single-stranded DNA-binding proteins (SSB)[47]. Using one of the approaches mentioned above, suitable recombination systems should be available for a wide range of species.

The modular design of CRISPR SWAPnDROP is based on the type IIS restriction enzymes BbsI and BsaI. Therefore, homology

arms and inserts should be free of these sites to be applied in the system. The recognition sites of the applied restriction enzymes are 6 bases long and therefore occur on average every 4096 bp. For optimal recombination efficiency, homology arms should be about 500 bp in *E. coli*[48], about 1000 to 3000 bp for *V. natriegens*[27] and about 1000 bp for *D. dadantii*. However, it is possible to reduce the length to 50 bp for *E. coli* and 200 bp for *V. natriegens* and *D. dadantii*. Therefore, the frequency of a restriction site in a homologous arm is statistically about 1/10 for *V. natriegens* and *D. dadantii* and 1/20 for *E. coli*. Shortening of homology arm, however, may cause a reduction in editing efficiency. Alternatively, restriction sites can be preserved, especially within inserts, by adding a final ligation step to the Golden Gate reaction protocol to religate the preserved sites (see 'Methods'). It is apparent, that the additional restriction sites reduce assembly efficiency. For cloning of homologies and insert parts, usually one to three fragments are assembled. Within this range, efficiency of Golden Gate cloning is very high, especially with ccdB/lacZα counter-selection used in CRISPR SWAPnDROP. Therefore, a reduced efficiency is no practical burden. For the seven parts of the pSwap assembly, the violacein colour system provides efficient selection for proper clones. Cloning efficiency can be further improved by transformation of a cloning strain instead of the strain of interest.

A general problem of Cas9-based counter-selection is the dependence on PAM sites (NGG) at the locus of interest. Although PAM sites occur on average every 16 bp, in some cases the PAM site might not be situated at the perfect location for a scarless edit. In this case, an additional edit needs to be performed to introduce a PAM site at the desired location. With this new PAM site, the actual scarless edit is possible in a second step. With its ability for stable iterative rounds of genome editing, CRISPR SWAPnDROP facilitates such consecutive edits. In this study, we have shown the stability of the system in 12 consecutive edits. Therefore, the system can be potentially applied in automated cloning approaches[29].

The application of the meganucleases I-SceI and I-CreI to cure plasmids from the previous iteration, restricts this method to organisms without naturally occurring sites for these meganucleases. However, the length of 18 and 22 bp for I-SceI and I-CreI practically almost excludes occurrences of such sites. The sites occur on average every 68 Gb for I-SceI and every 17 Tb for I-CreI. To illustrate this: Even in the 1000-fold larger human genome no such site is present. However, for the rare case of an occurrence in the genome, the meganucleases and its site on the H2 plasmids can be modified to avoid undesired restriction[49,50].

The multi-colour selection system introduced in this study improves editing efficiency by selecting out the major part of suppressor mutants that are in previous concepts part of the clones screened for successful edits. Its mechanism provides quality control for the recombination and restriction systems involved in the edit. In general, it relies on parallel plasmid-born edit with a clear phenotype. This concept is realised by Cas9 mediated *vioC* removal and subsequent recombination to reconstitute the pSwap plasmid. Only if Cas9 restriction and recombination are functional green colonies appear on the plate. Therefore, green colonies have a higher chance to yield successful edits and results between replicates are more stable. Like any other selector, the multi-colour system cannot filter all suppressor mutants (e.g. mutations on the target site or sgRNA mutations), therefore editing efficiency is below 100% and residual variance between replicates is detected.

In comparison to *E. coli* and *D. dadantii*, editing efficiency in *V. natriegens* is slightly lower. The decrease in editing efficiency may lie in the efficiency of *E. coli* λRED applied in *V. natriegens*. The recombination system is evolutionary optimized for *E. coli*

and apart from the induction system, we made no attempts to further optimize codon usage or RBS strength for λRED in *V. natriegens*. An optimized expression of λRED may improve efficiency even more. Still, screening of two colonies yields a positive clone on average. Therefore CRISPR SWAPnDROP is also highly effective in *V. natriegens*.

CRISPR SWAPnDROP is capable of transferring large chromosomal regions between strains and species. We tested the system in Hfr+ strains. By adding an RP4 plasmid derivative or by triparental mating, the approach can in principle be extended to any strain of interest. However, CRISPR SWAPnDROP can also be used to construct permanent conjugation strains of new species as we have shown for *V. natriegens* and *D. dadantii*. RP4 is a broad-host-range conjugation system, functional in Gram-negative bacteria[51] and various Gram-positive bacteria[52]. Large scale DNA transfer of CRISPR SWAPnDROP can be applied in synthetic chromosome construction for biotechnology and basic research[23,24,53]. The technical and human effort for the in vitro assembly of synthetic chromosomes strongly increases with its size[19]. In addition, with size, the chance for type IIS restriction sites within the DNA of interest rises. In the 6 kb lac operon, already 4 natural BbsI and 3 Esp3I sites usually used in modular cloning systems are present. In the 151 kb region 37 BbsI, 8 BsaI and 37 Esp3I are present. Hence, at a certain size of the construct, either site removal[25] or homology-based methods are required. Similar to modular cloning, assembly approaches based on yeast recombination suffer from laborious and time-consuming transfer back and forth between yeast and the organism of interest via in vitro methods[54]. With an iterative approach using modular assembly at the small scale to design customized parts and CRISPR SWAPnDROP for the assembly of the smaller parts to large chromosomes technical size limitations can be overcome and costs could be reduced drastically. With the transfer of 151 kb of DNA at an efficiency above 40%, we have shown that CRISPR SWAPnDROP is not limited in size within the limits of species-specific chromosome plasticity. In addition, the presented transfer of the *E. coli* lac operon into *V. natriegens* as well as the transfer of the GanB ORF from *D. dadantii* to *E. coli* is an example for a successful gain of function genome edit. Hence, it is possible to introduce desired properties from different organisms or a DNA library to construct tailor-made organisms.

Methods

Strains, media and reagents. *Escherichia coli* MG1655 was used for all genome editing experiments in *E. coli*, while *E. coli* DH5α was used for cloning purposes. For conjugation experiments, a precursor strain of MFD*pir* harbouring a stable RP4 conjugation machinery and additionally modified with *dapA* and *endA* knockouts was utilized[55]. All *E. coli* strains were cultivated in LB (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) at 37 °C under aerobic conditions in flasks with shaking at 200 rpm. If needed, LB was supplemented with anhydrotetracycline (100 ng/ml), arabinose (1%) or 300 μM DAP. For experiments requiring minimal medium, M9 medium (33.7 mM Na_2HPO_4; 22 mM KH_2PO_4; 8.6 mM NaCl; 9.4 mM NH_4Cl; 1 mM $MgSO_4$; 0.3 mM $CaCl_2$) supplemented with different carbon sources (lactose, arabinose, xylose, glucose: 0.5%) was used. For cloning, Q5 HF Polymerase (NEB), BsaI-HFv2 (NEB) and BbsI (NEB) were used. In this study, experiments with *V. natriegens* were carried out in *V. natriegens* ATCC 14048 ΔVNP12[56]. *V. natriegens* was cultivated in LB supplemented with V2 salts (204 mM NaCl, 4.2 mM KCl, 20.14 mM $MgCl_2$) (LBv2) at 37 °C under aerobic conditions in flasks with shaking at 200 rpm. For growth of *V. natriegens* in minimal medium, M9 medium supplemented with different carbon sources (0.4%) and additional 2% NaCl was used. Experiments in *D. dadantii* were carried out in *D. dadantii* 3937. Cultivation was done in LB medium at 30 °C under aerobic conditions in flasks with shaking at 200 rpm. Preparation of chemically competent cells and transformation was carried out according to Green and Rogers[57] and Stukenberg et al.[22] for *E. coli* and *V. natriegens*, respectively. For *D. dadantii*, cr3Dd and pDrop were transferred via electroporation[58]. The pSwap plasmid was transferred via conjugation[59].

Sequencing of bacterial genomes and data analysis. Isolation of bacterial genomic DNA was performed via phenol/chloroform extraction according to

Bruhn et al.[60]. Next-generation sequencing (NGS) was done by Eurofins Genomics Germany GmbH (Ebersberg, Germany). Reads were mapped and coverage was determined by the R package QuasR using a custom R script. The mutant reference genomes used for mapping were made using SnapGene on the basis of the *E. coli* MG1655 genome.

pSwap assembly. The pSwap plasmid consists of seven modules each harbouring a part for the Swap and Drop recombination system: T1, T2 (sgRNA expression cassettes); H1α, H2β, HA, HB (homologous regions) and INS (insert fragment). For the T1 and T2 construction, 1 μl of complementary oligonucleotides (10 μM) consisting of the 20 bp target sequence and corresponding overhangs (see Supplementary Table 1), 10 μl of 10X T4-Ligase buffer (NEB) and 88 μl MilliQ water were incubated at 95 °C for 10 min and then slowly cooled down to room temperature. 1 μl of annealed oligonucleotides was used for cloning. For the construction of the other parts, PCR products using primer with corresponding overhangs were used (see Supplementary Table 1). Approximately 40 fmol (~25 ng/1000 bp) of PCR products were utilized for cloning. Annealed oligonucleotides or PCR products were mixed with 40 fmol of the corresponding entry vectors, 1X T4-Ligase buffer (NEB), 1 μl (20 units) BsaI-HFv2 (NEB), 1 μl (400 units) T4-Ligase (NEB) and MilliQ water to a final volume of 20 μl. The mix was incubated for 2–5 h at 37 °C for in vitro assembly. In case of internal BsaI sites in the inserts, a final 30 min ligation step at 16 °C prior to heat inactivation is added for optimal Ligase activity and reduced BsaI activity. The mix was transformed into chemically competent DH5α cells and subsequently plated on LB Agar supplemented with 1mM IPTG and 250 μM X-gal. White colonies were tested for a successful integration into the entry vectors via sanger sequencing. Antibiotic concentrations used for cultivation of each part is seen in Table Supplementary 1. If needed, Hα- and Hβ homologies for H1α and H2β were added as an additional PCR fragment without fixed overhangs upstream of the 5′-end of the H1 homology or downstream of the 3′-end of the H2 homology. For the pSwap assembly, 40 fmol of each part (T1, T2, H1α, H2β, HA, HB, INS), 1X T4-Ligase buffer, 1 μl (10 units) BbsI (NEB), 1 μl (400 units) T4-Ligase and MilliQ water were incubated in a final volume of 20 μl for 5–7 h at 37 °C. In case of internal BbsI sites in the inserts, a final 30 min ligation step at 16 °C prior to heat inactivation can be added for optimal Ligase activity and reduced BbsI activity. The restriction-ligation could then be transformed into a cloning (DH5α) or directly into the desired wild-type strain. Successfully assembled pSwap plasmids were verified by purple colonies after transformation. Different parts of the deoxyviolacein pathway are distributed on different modules. However, only when the modules are correctly assembled into the pSwap, a functional deoxyviolacein pathway result in purple colonies. Antibiotics were used depending on the corresponding H2β (see Supplementary Table 1). For insertion or deletion experiments, T2 was not essential, but still mandatory for the pSwap assembly. Therefore, a random DNA sequence, which was not present in the strains' chromosomal sequence, was cloned into the entry vector and used for the assembly. Homologies (H1, H2) for the loading of the pSwap were also not needed, but a recombination between those random sequences was important to maintain the integrity of the pSwap plasmid during genome modifications. Therefore, identical 175 bp random DNA sequences were cloned into H1α and H2β (HR, see Fig. 1a). In general homology arms of 500[48] and 1500 bp[27] were used for *E. coli* and *V. natriegens*, respectively. As all parts were digested with BsaI and BbsI for cloning, the presence of those restriction sites within the parts (e.g. the homologies or insert fragment) were avoided. Therefore, in some cases the size for homology arms were chosen between 50 and 500 bp for *E. coli* and between 1000 and 1500 bp for *V. natriegens* to avoid present restriction sites. A list with Primers, which were used for the construction of each individual pSwap plasmid can be found in the supplementary data.

CRISPR SWAPnDROP genome editing protocol (Indel/Swap). *E. coli* MG1655 was first transformed with the helper plasmid cr3Ec. For all CRISPR/Cas9 experiments, cr3Ec harbouring the P$_{tet}$ repressor (TetR) needed to be transformed first to establish P$_{tet}$ repression prior to the transformation with pSwap. Double transformation with pSwap and cr3Ec lead to dysfunctional components in the Cas9 system[14]. After transformation with cr3Ec, *E. coli* MG1655 cr3Ec was transformed either directly with the pSwap restriction-ligation mix or the prepared pSwap plasmid. The transformants were grown on LB Agar with appropriate antibiotics (see Supplementary Table 2) at 37 °C overnight. The plates were then incubated for 2–5 h at room temperature allowing the colonies to form a purple colour. Subsequently, purple colonies were inoculated in LB with 100 ng/ml anhydrotetracyclin (aTet), 1% arabinose and appropriate antibiotics (see Supplementary Table 2). Cultures were grown at 37 °C shaking (200 rpm) overnight and plated on LB Agar with the same antibiotics at 37 °C. After ~16 h, the plates were incubated at room temperature for 2–5 h now allowing the colonies to form a green colour. Green colonies were then tested for a successful editing event via PCR and Sanger sequencing.

Vibrio natriegens genome editing protocol (Indel/Swap). The *V. natriegens* strain ATCC 14048 ΔVNP12 was first transformed with cr3Vn. Subsequently, this strain was transformed with pSwap Cm^R and incubated on LBv2 Agar with

corresponding antibiotics at 37 °C overnight (see Supplementary Table 3). The plates were incubated at room temperature for several hours to develop its purple colour. Purple colonies were inoculated and grown until OD$_{600}$ reached 4 or overnight in LBv2 with appropriate antibiotics at 37 °C and 200 rpm shaking. Induction of the λRED genes was carried out by adding 0.4% arabinose for 1–2 h followed by the induction of the Cas9 gene and the sgRNA with aTet (80 ng/ml) for further 2–4 h at 37 °C. Subsequently, the cultures were plated on LBv2 with 80 ng/ml aTet, 0.4% arabinose, antibiotics (see Supplementary Table 3) and grown at 37 °C overnight. The following day, colonies were incubated at room temperature for several hours or until the next day. Green colonies were then tested via PCR and Sanger sequencing.

Dickeya dadantii genome editing protocol (Indel/Swap). *D. dadantii* 3937 was first transformed with cr3Dd via electrotransformation. After transformation/conjugation of the pSwap, cells were incubated on LB Agar with corresponding antibiotics at 30 °C overnight (see Supplementary Table 4). Purple colonies were inoculated in LB medium with appropriate antibiotics and grown overnight at 30 °C. Induction of the λRED genes was carried out by adding 0.4% arabinose for 1–2 h followed by the induction of the Cas9 gene and the sgRNA with aTet (80 ng/ml) for further 2–4 h at 30 °C. The culture was then plated on LB Agar supplemented with 100 ng/ml aTet, 0.4% arabinose and antibiotics (see Supplementary Table 4).

Conjugation and drop protocol. For the transfer and integration of chromosomal DNA, the loading of the pSwap plasmid with a chromosomal region was carried out in the donor strain *E. coli* MFDpir Δ dapA Δ endA, harbouring the RP4 conjugation system[55]. Simultaneously, the acceptor strain was transformed with the helper plasmids cr3Ec and pDrop. Similar to the sgRNA level 1 parts (T1, T2) of the pSwap assembly, the spacer sequence (T3) for the counter-selection in the acceptor strain needed to be cloned into the pDrop entry vector and subsequently by sanger sequencing. Both, the donor strain harbouring the loaded pSwap plasmid (pSwap') and the RP4 conjugation system as well as the acceptor strain harbouring the helper plasmids were cultivated in LB with correct antibiotics at 37 °C overnight (see Supplementary Table 2). The culture of the *E. coli* donor strain was additionally supplemented with 300 μM diaminopimelic acid (DAP). The donor and acceptor strains were diluted to an OD$_{600}$ of 3 with fresh LB medium after washing and then mixed with a 1:1 ratio. Subsequently, the mix was spotted on LB Agar plates only supplemented with 300 μM DAP and incubated at 37 °C overnight. The spot was scrapped off the plate and washed several times with LB medium to remove residual DAP. Selection for successful conjugation was performed on LB Agar with appropriate antibiotics (see Supplementary Table 2). After the conjugation of the pSwap' plasmid, the acceptor strain was inoculated and directly induced in LB with the corresponding antibiotics for the selection of cr3Ec and pDrop (see Supplementary Table 2). Selection for the pSwap was omitted, as during the excision of the recombination template, the pSwap plasmid was destroyed. Correct function of the CRISPR/Cas9- and λRED-systems was ensured by the cut and reconstitution of the pDrop non-functional mScarlet gene yielding red colonies. Induction was again carried out with 1% arabinose and 100 ng/ml aTet. The induced acceptor strain was cultivated overnight at 37 °C and then plated on LB with the same antibiotics. Red colonies were tested for the integration into chromosome via PCR and sanger sequencing. Conjugation and subsequent chromosome integration from *E. coli* to *V. natriegens* were carried out similarly. For *V. natriegens* pDropVn containing a gentamycin resistance cassette, as selection with ampicillin did not provide satisfying results. For the conjugation, the stationary phase cultures diluted to an OD$_{600}$ of 3 were mixed with a 1:9 ratio (*V. natriegens*:*E. coli*) and then spotted on LBv2 agar plates, also supplemented with DAP. *V. natriegens* now harbouring the pSwap' plasmid, cr3Vn and pDropVn were inoculated and grown overnight at 37 °C with corresponding antibiotics (see Supplementary Table 3). The next day, induction was performed in the stationary phase culture by inducing first the λRED recombination with 0.4% arabinose for 1–2 h and then the CRISPR/Cas9 with 80 ng/ml aTet for another 2–4 h. The induced culture was then plated on LBv2 agar with antibiotics as well as 0.4% arabinose and 80 ng/ml aTet. Red colonies were tested via PCR and Sanger sequencing. Conjugational transfer of pSwap from *V. natriegens*recJ::RP4 to *E. coli* was performed in analogy to the reverse direction. For conjugation spots, LB instead of LBv2 agar plates were used. Conjugation and subsequent chromosome integration from *E. coli* to *D. dadantii* were carried out in analogy to the *V. natriegens* protocol, but using LB with *D. dadantii* specific antibiotic concentrations (see Supplementary Table 4) and 30 °C for optimal growth. Due to the lower growth rate of *D. dadantii* compared to *V. natriegens*, a 1:1 mix ratio was used during conjugation.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The DNA sequencing data for edit verification generated in this study have been deposited in the sequence read archive (SRA) database under accession code PRJNA824474. The plot raw data generated in this study are provided in the

Supplementary Information/Source data file. Plasmid maps and cloning templates for pSwap construction can be found in the supplementary data. Source data are provided with this paper.

References

1. Ambler, R. P. & Scott, G. K. Partial amino acid sequence of penicillinase coded by *Escherichia coli* plasmid R6K. *Proc. Natl Acad. Sci. USA* **75**, 3732–3736 (1978).
2. Reiss, B., Sprengel, R. & Schaller, H. Protein fusions with the kanamycin resistance gene from transposon Tn5. *EMBO J.* **3**, 3317–3322 (1984).
3. Stacey, K. A. & Simson, E. Improved method for the isolation of thymine-requiring mutants of *Escherichia coli*. *J. Bacteriol.* **90**, 554–555 (1965).
4. Fels, U., Gevaert, K. & Van Damme, P. Bacterial genetic engineering by means of recombineering for reverse genetics. *Front. Microbiol.* **11**, 548410 (2020).
5. Sobetzko, P. Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. *Nucleic Acids Res.* **44**, 1514–1524 (2016).
6. El Houdaigui, B. et al. Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Res.* **47**, 5648–5657 (2019).
7. Hoess, R. H. & Abremski, K. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J. Mol. Biol.* **181**, 351–362 (1985).
8. Cox, M. M. The FLP protein of the yeast 2-microns plasmid: expression of a eukaryotic genetic recombination system in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **80**, 4223–4227 (1983).
9. Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J. Mol. Biol.* **402**, 52–69 (2010).
10. Ellis, H. M., Yu, D., DiTizio, T. & Court, D. L. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc. Natl Acad. Sci. USA* **98**, 6742–6746 (2001).
11. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
12. Pósfai, G., Kolisnychenko, V., Bereczki, Z. & Blattner, F. R. Markerless gene replacement in *Escherichia coli* stimulated by a double-strand break in the chromosome. *Nucleic Acids Res.* **27**, 4409–4415 (1999).
13. Seligman, L. M., Stephens, K. M., Savage, J. H. & Monnat, R. J. Genetic analysis of the *Chlamydomonas reinhardtii* I-CreI mobile intron homing system in *Escherichia coli*. *Genetics* **147**, 1653–1664 (1997).
14. Reisch, C. R. & Prather, K. L. The no-SCAR (scarless Cas9 assisted recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5**, 15096 (2015).
15. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
16. Wang, K. et al. Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016).
17. Cress, B. F. et al. CRISPathBrick: modular combinatorial assembly of type II-A CRISPR arrays for dCas9-mediated multiplex transcriptional repression in *E. coli*. *ACS Synth. Biol.* **4**, 987–1000 (2015).
18. Vad-Nielsen, J., Lin, L., Bolund, L., Nielsen, A. L. & Luo, Y. Golden Gate Assembly of CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell Mol. Life Sci.* **73**, 4315–4325 (2016).
19. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765 (2011).
20. Klein, C. A., Emde, L., Kuijpers, A. & Sobetzko, P. MoCloFlex: a modular yet flexible cloning system. *Front. Bioeng. Biotechnol.* **7**, 271 (2019).
21. Püllmann, P. et al. Golden mutagenesis: an efficient multi-site-saturation mutagenesis approach by Golden Gate Cloning with automated primer design. *Sci. Rep.* **9**, 10932 (2019).
22. Stukenberg, D. et al. The Marburg Collection: a Golden Gate DNA assembly framework for synthetic biology applications in *Vibrio natriegens*. *ACS Synth. Biol.* **10**, 3236 (2021).
23. Messerschmidt, S. J. et al. Optimization and characterization of the synthetic secondary chromosome synVicII in *Escherichia coli*. *Front Bioeng. Biotechnol.* **4**, 96 (2016).
24. Yoneji, T., Fujita, H., Mukai, T. & Su'etsugu, M. Grand scale genome manipulation via chromosome swapping in *Escherichia coli* programmed by three one megabase chromosomes. *Nucleic Acids Res.* **49**, 8407–8418 (2021).
25. Schindler, D., Milbredt, S., Sperlea, T. & Waldminghaus, T. Design and assembly of DNA sequence libraries for chromosomal insertion in bacteria based on a set of modified MoClo vectors. *ACS Synth. Biol.* **5**, 1362–1368 (2016).
26. Lee, H. H., Ostrov, N., Gold, M. A. & Church, G. M. Recombineering in *Vibrio natriegens*. Preprint at *bioRxiv* https://doi.org/10.1101/130088 (2017).
27. Dalia, T. N. et al. Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in *Vibrio natriegens*. *ACS Synth. Biol.* **6**, 1650–1655 (2017).
28. Fujikawa, T., Hatomi, H. & Ota, N. Draft genome sequences of seven strains of *Dickeya dadantii*, a quick decline-causing pathogen in fruit trees, isolated in Japan. *Microbiol. Resour. Announc.* **9**, e00609-20 (2020).
29. Lajoie, M. J., Gregg, C. J., Mosberg, J. A., Washington, G. C. & Church, G. M. Manipulating replisome dynamics to enhance lambda Red-mediated multiplex genome engineering. *Nucleic Acids Res.* **40**, e170 (2012).
30. Hirokawa, Y. et al. Genetic manipulations restored the growth fitness of reduced-genome *Escherichia coli*. *J. Biosci. Bioeng.* **116**, 52–58 (2013).
31. Lee, H. H. et al. Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi. *Nat. Microbiol.* **4**, 1105–1113 (2019).
32. Hoff, J. et al. *Vibrio Natriegens*: an ultrafast-growing marine bacterium as emerging synthetic biology chassis. *Environ. Microbiol.* **22**, 4394–4408 (2020).
33. Dalia, T. N. et al. Enhancing multiplex genome editing by natural transformation (MuGENT) via inactivation of ssDNA exonucleases. *Nucleic Acids Res.* **45**, 7527–7537 (2017).
34. Hoffart, E. et al. High substrate uptake rates empower *Vibrio natriegens* as production host for industrial biotechnology. *Appl. Environ. Microbiol.* **83**, e01614-17 (2017).
35. Venkatesan, P. & Kaback, H. R. The substrate-binding site in the lactose permease of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **95**, 9802–9807 (1998).
36. Guan, L., Sahin-Toth, M. & Kaback, H. R. Changing the lactose permease of *Escherichia coli* into a galactose-specific symporter. *Proc. Natl Acad. Sci. USA* **99**, 6613–6618 (2002).
37. Li, Y. et al. Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 meditated genome editing. *Metab. Eng.* **31**, 13–21 (2015).
38. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60**, 385–397 (2015).
39. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
40. Li, J. F. et al. Multiplex and homologous recombination-mediated genome editing in Arabidopsis and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* **31**, 688–691 (2013).
41. Schuster, M. & Kahmann, R. CRISPR-Cas9 genome editing approaches in filamentous fungi and oomycetes. *Fungal Genet. Biol.* **130**, 43–53 (2019).
42. Nayak, D. D. & Metcalf, W. W. Cas9-mediated genome editing in the methanogenic archaeon *Methanosarcina acetivorans*. *Proc. Natl Acad. Sci. USA* **114**, 2976–2981 (2017).
43. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
44. De Souza Silva, O. & Blokesch, M. Genetic manipulation of *Vibrio cholerae* by combining natural transformation with FLP recombination. *Plasmid* **64**, 186–195 (2010).
45. Nayak, D. D. & Metcalf, W. W. Genetic techniques for studies of methyl-coenzyme M reductase from *Methanosarcina acetivorans* C2A. *Methods Enzymol.* **613**, 325–347 (2018).
46. Horwitz, A. A. et al. Efficient multiplexed integration of synergistic alleles and metabolic pathways in yeasts via CRISPR-Cas. *Cell Syst.* **1**, 88–96 (2015).
47. Filsinger, G. T. et al. Characterizing the portability of phage-encoded homologous recombination proteins. *Nat. Chem. Biol.* **17**, 394–402 (2021).
48. Ao, X. et al. Based on CRISPR-Cas12a. *Front. Microbiol.* **9**, 2307 (2018).
49. Seligman, L. M. et al. Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.* **30**, 3870–3879 (2002).
50. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).
51. Thomas, C. M. & Smith, C. A. Incompatibility group p plasmids: genetics, evolution, and use in genetic manipulation. *Annu. Rev. Microbiol.* **41**, 77–101 (1987).
52. Trieu-Cuot, P., Arthur, M. & Courvalin, P. Origin, evolution and dissemination of antibiotic resistance genes. *Microbiol. Sci.* **4**, 263–266 (1987).
53. Zumkeller, C., Schindler, D. & Waldminghaus, T. Modular assembly of synthetic secondary chromosomes. *Methods Mol. Biol.* **1837**, 71–94 (2018).
54. Adam, A. C., González-Blasco, G., Rubio-Texeira, M. & Polaina, J. Transformation of *Escherichia coli* with DNA from Saccharomyces cerevisiae cell lysates. *Appl. Environ. Microbiol.* **65**, 5303–5306 (1999).
55. Ferriéres, L. et al. Silent mischief: bacteriophage Mu insertions contaminate products of *Escherichia coli* random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J. Bacteriol.* **192**, 6418–6427 (2010).

56. Pfeifer, E. et al. Generation of a prophage-free variant of the fast-growing bacterium *Vibrio natriegens*. *Appl. Environ. Microbiol.* **85**, e00853-19 (2019).
57. Green, R. & Rogers, E. J. Transformation of chemically competent *E. coli*. *Methods Enzymol.* **529**, 329–336 (2013).
58. Sambrook, J. in *Synthetic Oligonucleotides* (1989).
59. Miller, V. L. & Mekalanos, J. J. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires toxR. *J. Bacteriol.* **170**, 2575–2583 (1988).
60. Bruhn, M. et al. Functionality of Two Origins of Replication in Vibrio cholerae Strains With a Single Chromosome. *Front Microbiol.* **9**, 2932 (2018).

Acknowledgements

Author contributions

M.T., C.K. and P.S. conceived the experiments. M.T., C.K., M.M. and P.S. conducted the experiments. M.T., C.K., M.M. and P.S. analysed the results. M.T. and P.S. wrote the manuscript. All authors reviewed the manuscript.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-30843-1.

Correspondence and requests for materials should be addressed to Patrick Sobetzko.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

4 Discussion

4.1 Impact of chromosomal organization on bacterial gene regulation

In recent decades investigations of bacterial gene regulation were highly focused on transcription factors and their local regulation of gene expression. However, this is only one factor of many that in concert constitute a diverse regulation machinery of bacterial cells. Another factor whose importance for gene expression is often underestimated, but which contributes greatly to the regulatory landscape of bacterial cells, is the organization of the chromosome.

4.1.1 One-dimensional (1D) organization of the chromosome

When investigating the local genetic context of bacterial genomes, the chromosome is a linear entity beginning and ending at the origin of replication. Genes are distributed along this chromosome and are controlled by different promoters and regulatory proteins. However, this distribution is not random, but the genes are ordered relative to the oriC resulting in a symmetrical axis from oriC to terminus (oriC-ter axis) and thereby forming a one-dimensional (1D) organization of the chromosome. It has been shown, that gene expression is increased for genes closer to the oriC compared to genes closer to the terminus and that genes are less conserved with increasing distance to the oriC [Lato and Golding, 2020, Couturier and Rocha, 2006]. This is further supported by the fact that orthologous genes are mainly rearranged symmetrically around the

75

origin of replication in various bacterial species [Sobetzko et al., 2012, Eisen et al., 2000, Darling et al., 2008]. This means, that genes are more likely to 'jump' between both replichores as long as their oriC distance is preserved. Furthermore, binding sites of specific transcription factors also tend to be organized relative to the oriC-ter axis [Sobetzko et al., 2012]. The sigma factor RpoD (sigma 70), which is the dominant sigma factor during the exponential phase, regulates more genes in oriC proximity than in the terminus region. Conversely, the sigma factor RpoS (sigma 38), which is the primary regulator of stationary phase genes, controls more genes closer to the terminus region. Similar organized are the binding sites of several NAPS. Fis and Lrp for instance are important global regulators during exponential and stationary phase, respectively. While Fis activates many genes in oriC proximity, Lrp represses oriC proximal genes instead [Sobetzko et al., 2012]. The bias in gene location relative to the oriC and the temporal factor of different growth phases implies that the one-dimensional chromosome organization is strongly related to DNA replication. During cell division, chromosome replication starts at the oriC and moves bidirectionally until the replication forks meet at the terminus. During this process, gene copies are increased, leading to an increased expression. Since this is a mandatory characteristic of life, one can assume that evolution has led to the use of this as a regulatory system. Indeed it has been demonstrated that for fast-growing bacteria, highly expressed genes associated with transcription and translation are preferentially located close to the oriC and that such positioning is under strong selection [Couturier and Rocha, 2006]. Sobetzko et al. could observe a spatio-temporal expression pattern in *E. coli*, which resembles the replication-induced copy numbers [Sobetzko et al., 2013]. It could be demonstrated in chapter 3.1 that this expression pattern is caused by the copy number effect during the exponential phase. Furthermore, it was also shown that for the fast-growing bacterium *V. natriegens*, genes, which are expressed during the exponential phase, tend to be closer located to the oriC than in the slower-growing bacterium *D. dadantii*. Additionally, an extension to different clades in the bacterial kingdom showed a relationship between the growth rate of a bacterium and position maintenance along the oriC-ter axis. These findings suggest that the replication-induced copy number effect is a driving factor in the evolution of genome organization for fast-growing bacteria and therefore genes closer to the oriC are subject to its regulation. Further transcriptome analysis revealed that around 40% of the *E. coli* genes are mainly regulated by the copy number effect between exponential and stationary phase. Especially most

genes that are assigned to the coenzyme metabolism are regulated by copy numbers. This implies a regulatory mechanism to counteract coenzyme dilution during cell growth [Hartl et al., 2017]. Ribosomal RNA operons, which code for the most abundant form of RNA in bacterial cells, are often in close proximity to the oriC, suggesting a correlation of their expression with the environmental condition [Soler-Bistué et al., 2015, Couturier and Rocha, 2006]. When growth conditions improve by e.g. the abundance of a nutrient-rich medium, multifork replication and the resulting increase in copy numbers provide the cell with the required amount of ribosomal RNA for functional homeostasis. Soler-Bistué et al. showed an oriC distance-dependent growth rate reduction when relocating most of the rrn operons in *V. cholerae* [Soler-Bistué et al., 2015]. Therefore, it was speculated in chapter 3.1 that the main reason for the inversion was the increased distance of oriC to the *rrnCABE* operon cluster when relocating the oriC into the terminus.

It has to be mentioned that this one-dimensional chromosome organization is probably less important for slow-growing bacteria, which was shown by a reduced conservation of orthologous genes along the oriC-ter axis for slow-growing bacteria in chapter 3.1. In these bacteria, promoter regulation probably takes a larger share of the total regulation than the copy number effect, as growth efficiency or efficient carbon utilisation is more pronounced compared to fast-growing bacteria [Roller et al., 2016].

Although the maximum impact factor of the copy number effect on gene expression may be much smaller compared to promoter regulation, the fact that for most genes in *E. coli* the copy number effect accounts for a significant part of their regulation suggests its role as one of the most important regulators.

Another important factor for gene expression is the gene orientation relative to the oriC. As DNA replication and transcription occur on the same DNA molecule, collision of these machineries is a regular event and can cause substantial problems to replication progression and transcription [Ivanova et al., 2015, French, 1992]. This effect is especially strong for most essential genes, but can also cause significant transcription defects of highly expressed genes and large operons [Rocha and Danchin, 2003, Omont and Képès, 2004, Price et al., 2005]. Therefore, transcription of most of the essential and highly expressed genes in *E. coli* and *Bacillus subtilis*, for instance, are co-oriented with the replication direction. However, due to the inversion of the left replichore in chapter 3.1, two of the seven highly expressed rRNA operons in the INV strain are convergently oriented to the replication direction, resulting in a head-on collision of

the transcription and replication machinery. One could argue that this leads to the increased doubling time of the strain, but it is rather one factor of many. The chromosome segregation system FtsK/KOPS, for instance, which is part of the divisome in *E. coli*, is most likely impaired in the INV strain. In the wild-type strain, the KOPS sites distributed along the chromosome are all oriented towards the terminus region to help the FtsK translocase reach the *dif* site where the final chromosome unlinking occurs [Bigot et al., 2007]. In the INV strain, the KOPS sites on the left replichore are oriented in the direction of the oriC due to the inversion. This will most likely cause problems in chromosome segregation and therefore could influence the doubling time of the strain. Additionally, the ter macrodomain is affected as the inversion occurred within this domain. This could also lead to impaired chromosome segregation. Furthermore, the altered gene regulation due to the inversion will presumably also affect the growth rate of the INV strain. Considering that only five rRNA operons are necessary to support optimal growth on complex media [Condon et al., 1995], one could speculate that the inverted orientation of the two rRNA operons in the INV strain has less detrimental effects than assumed.

4.1.2 Three-dimensional (3D) organization of the chromosome

The replication-induced copy number effect most likely formed the genome organization on the one-dimensional level (oriC-ter axis) and constitute a global regulatory system for gene expression in bacterial cells. It is therefore conceivable that the three-dimensional structure of bacterial chromosomes also influences gene expression. Furthermore, gene order could have been adapted accordingly in the course of evolution to exploit this for efficient gene regulation. Since the chromosome compaction is highly dynamic one could further assume, that the chromosome structure changes according to changes in the prevailing conditions to facilitate the expression of genes required for the new situation. During the exponential phase, the chromosome is highly condensed mainly due to NAPS and DNA supercoiling, while during the transition into the stationary phase the nucleoid relaxes and expands into the entire cytoplasm [Krogh et al., 2018]. This compaction can bring locally distant genes in spatially proximity indicating a regulatory effect [Le et al., 2013, Marbouty et al., 2015].

In fast-growing cells, RNA-polymerases are localized to distinct transcriptional foci suggesting active clustering of highly transcribed genes [Krogh et al., 2018]. Interestingly, several studies have found a periodic distribution of conserved gene pairs in the *E. coli* chromosome, which are assumed to be in spatially proximity after chromosome folding [Wright et al., 2007, Mathelier and Carbone, 2010, Képès, 2004]. Krogh et al. proposed that the observed patterns can result in transcriptional spilling [Krogh et al., 2020]. The transcriptional activation of one gene and the associated recruitment and increase in the local concentration of RNAP might spill onto spatially close genes increasing the chance of successful gene transcription initiation. This could be a regulatory mechanism of chromosome organization to allow the co-expression of functional-related genes. The genes for pectinolysis in the bacterial genera *Dickeya* and *Pectobacterium*, which are responsible for the soft rot disease in several plants, are scattered in multiple pathogenicity islands in the genome [Le Berre et al., 2022]. A study found, that these pathogenicity islands are distributed periodically along the genome and form a spatially proximal "archipelago" [Bouyioukos et al., 2016]. It is proposed that this co-localization and co-transcription would favour the efficient funnelling of pectinases at convergent points within the cell, which constitutes a crucial strategy for the successful degradation of the plant cell wall by the full blend of pectinases.

4.2 Challenges for genetic engineering with respect to chromosomal architecture

Genetic modifications are an integral part of today's research, whether to improve the biological understanding of organisms or to modify them for our benefit. When planning such chromosomal changes, the main focus is often only on the modification itself rather than on the chromosomal context in which it is introduced. However, the chromosomal context can have a decisive influence on the outcome of the experiment depending on the modification. As discussed in the previous section, gene expression can be modulated by the chromosomal gene order either on the one-dimensional level (oriC-ter axis) or the three-dimensional level (chromosome topology). Therefore, the integration of synthetic and exogenous transcription units into the chromosome or

modifications affecting the transcription of a gene are prone to be influenced by the chromosomal architecture and vice versa.

Studies have found that expression of a reporter gene can vary up to 300-fold in *E. coli* depending on the chromosomal location it was integrated [Bryant et al., 2014, Scholz et al., 2019]. Especially high expression peaks could be observed in proximity to the rrn operons [Scholz et al., 2019]. In general, a high transcriptional propensity was positively correlated with Fis binding sites and negatively correlated with H-NS binding sites. These differences in gene expression were independent of replication-induced copy number effects and indicate a structural influence of the chromosome topology as previously discussed. It has to be mentioned that for *E. coli*, 99% of the chromosomal locations show only small differences in transcriptional activity, which may be acceptable when integrating new expression cassettes into the chromosome [Scholz et al., 2019]. However, transcriptionally silent extended protein occupancy domains (tsEPODs), which are genomic regions of high protein occupancy appear to correlate with low transcription levels [Vos, 2009]. It is proposed, that these tsEPODs are forming chromosomal organizing centers by high occupancy of NAPs like Fis, IHF and H-NS. Integration into such regions would therefore lead to little or no transcriptional activity and expression of the heterologous genes.

The copy number effect as a spatio-temporal mechanism, which can transiently increase the expression of genes is also rarely addressed when planning strategies to integrate new expression cassettes into the genome. This is particularly important for synthetic metabolic pathways or regulatory networks where fine-tuning of the individual components is required [Li et al., 2020]. It could be shown that the expression of reporter genes correlates positively with the position relative to the oriC and that generally the largest differences in gene expression are due to the replication-induced copy number effect [Block et al., 2012, Sauer et al., 2016]. This fits with the moderate topology-related changes in gene expression in the majority of the *E. coli* chromosome mentioned before [Scholz et al., 2019]. It was shown in chapter 3.1, that gene expression normalized to the copy number resembles biological reality. Therefore, relative copy numbers, e.g. obtained by marker frequency analysis, could be used to calculate the expected position-dependent differences in gene expression of new transcription units. In cases where the expression of synthetic or heterologous genes needs to be constant throughout the growth phases, integration sites at terminus-proximal positions could be chosen where the copy number effect is less pronounced. However, horizontally

acquired genes preferentially accumulate near the (AT-rich) terminus region and are usually silenced by H-NS [Lawrence and Ochman, 1998, Touchon et al., 2014, Singh et al., 2016]. Therefore one can assume that the integration of expression cassettes into this region has a higher probability to be negatively influenced by silencing effects [Scholz et al., 2019]. Another option could be negative autoregulation of the integrated expression cassette, which would compensate for fluctuations in gene dosage due to the copy number effect [Klumpp et al., 2009].

Although head-on collision of transcription and replication in bacteria can have detrimental effects [Rocha and Danchin, 2003, Omont and Képès, 2004, Price et al., 2005], none of the above studies found any influence of gene orientation on reporter gene expression in this regard [Block et al., 2012, Sauer et al., 2016, Scholz et al., 2019]. However, gene orientation concerning the local chromosomal context still could have an effect. It is known that promoters are sensitive to DNA supercoiling and that during transcription, DNA ahead of the RNAP becomes overwound (positively supercoiled) while DNA behind it becomes underwound (negative supercoiled) [Dorman, 2019]. It could be shown that the expression of reporter genes is influenced by the expression of neighbouring genes and vice versa caused by transcription-induced supercoiling [Bryant et al., 2014, Yeung et al., 2017]. Therefore, the integration of an expression cassette into a particular chromosomal location could influence and be influenced by the local genetic context and the expression can vary depending on the orientation. Bryant et al. observed a 4-fold reduction of the reporter gene expression, when it was downstream of the inducible native *melAB* operon, irrespective of the orientation of the reporter gene [Bryant et al., 2014]. Furthermore, transcription of the reporter gene did not affect the expression of *melA* when it was directed away from the operon. However, there was a 50% reduction in *melA* expression when the transcription of the reporter cassette was directed towards the operon. This indicates a negative impact of the introduction of positive supercoils on both of these promoters. Therefore, the effects of the local genetic context could be of great importance if the integration site is close to essential genes. Transcription of the heterologous or synthetic expression cassette could then influence the usually highly regulated transcription of the essential gene, which could lead to deleterious effects in the cell. On the other hand, transcription-induced supercoiling of chromosomal genes could influence the expression of the integrated expression cassette, which is relevant to synthetic gene network design. It was demonstrated, that transcription-induced supercoiling of neighbouring genes

can alter key features of synthetic gene induction, such as dynamic range, maximum expression, and the activation threshold [Yeung et al., 2017]. This can be problematic, when using synthetic gene circuits, which comprise programmed genetic elements functioning as logic gates (AND, OR, NAND etc.). These genetic circuits are realized using repressors and inducible promoters [Yokobayashi et al., 2002, Guet et al., 2002]. Disturbance in gene induction of those circuits by neighbouring gene transcription could lead to impaired or absent function.

4.3 Tool requirements for genetic engineering

Genetic engineering comprises chromosomal modifications ranging from single nucleotide insertions, deletions and substitutions over integration and deletion of whole genes and operons to the recoding and construction of entire chromosomes. Each of these modifications places different demands on the tools used.

4.3.1 Small-scale genetic engineering

For smaller insertions (up to 30bp) and deletions (up to 45kb) or point mutations, only small templates in the form of oligonucleotides are necessary [Fels et al., 2020]. The high efficiency of oligo-mediated allelic replacements (OMAR) eliminates the need for selection markers, when the methyl-directed mismatch repair system is inactivated. Instead of eliminating the mismatch-repair system, CRISPR/Cas9 can also be used to increase the efficiency by eliminating non-edited cells [Jiang et al., 2013]. Furthermore, the use of a single-stranded DNA template only requires the *bet* gene of λRED for successful recombination, which delivers the oligonucleotides to the replication fork where they are incorporated.

For larger inserts ($\sim$5-10kb) or deletions ($\sim$ >50kb), double-stranded DNA templates are needed, which can be introduced into the cell as a linear molecule (PCR fragment) or in the form of a plasmid (PCR fragment cloned into a vector). Recombination with a linear template always requires reduced or non-functional intracellular exonucleases

[Madyagol et al., 2011]. Otherwise, incoming linear DNA is degraded during transformation. Therefore, this is either possible with natural competent species, which harbours a native system to protect foreign DNA [Blokesch, 2016], species with specific genetic background (e.g. exonuclease knockout) or by using phage-based recombination systems (e.g. λRED), which inhibit exonuclease activity.

However, the recombination efficiency is still quite low and further selection methods are needed to obtain a reasonable amount of positive clones. This can be selection markers or, if a scarless integration is required, CRISPR/Cas9. Selection markers usually have to be additionally removed from the genome, e.g. by site-specific recombination using Cre/*loxP* or FLP/*FRT*, respectively.

For some bacterial species, transformation with a linear DNA fragment is not possible. Therefore an integrative plasmid harbouring the insert is used instead. This requires the use of appropriate vectors, which are unable to replicate under certain conditions [Hamilton et al., 1989]. This can be for instance a temperature-sensitive replicon (pSC101), which facilitate the integration of the plasmid into the chromosome upon temperature increase. In addition, the plasmid harbours a homologous region comprising the insert for chromosome integration and a counter-selection marker (e.g. antibiotic resistance cassette). Upon temperature increase and subsequent loss of the replicon, only cells with the integrated plasmid and counter-selection marker survive. In a second step, the cointegrate can be resolved by recombination between the two homologies (mutant/wildtype) resulted from the single crossover event. Thereby, the region between those homologous regions are removed and either the wild type or the mutant allele remains. This approach relies on the functionality of the native recombination protein RecA and allows seamless recombineering. The second recombination event usually occurs at low frequency, therefore another selection method is often required [Madyagol et al., 2011]. This can be, for instance, the integration of a I-SceI restriction site into the vector, which facilitates the cointegrate resolution upon enzyme restriction. Another approach uses a vector harbouring the *sacB* gene , which converts sucrose to levan, a toxic compound for gram-negative bacteria. By adding sucrose, cells, which did not resolve the conitegrate and still harbouring the *sacB* gene, are eliminated.

To circumvent the two-step homologous recombination events and the reduced efficiency due to the possible remain of the wildtype allele, CRISPR/Cas9 can be helpful. As described in chapter 3.2 and also by Wang et al., CRISPR/Cas9 can be utilized to

cut the vector and supply the cell with a linear DNA template [Wang et al., 2016]. This facilitates the use of a high efficient phage-based recombination system like λRED due to the linear DNA fragment and provides a simple method for the integration of very large DNA fragments into the chromosome.

4.3.2 Large-scale genetic engineering

Large-scale genetic engineering can comprise the construction of minimal genomes, the introduction of large exogenous genomic regions with specific functionality, e.g. conjugation systems (see chapter 3.2), as well as recoding or constructing entire chromosomes via synthesized DNA. While the integration of pre-existing heterologous DNA requires DNA transfer systems such as transposon mutagenesis or CRISPR SWAPnDROP, chromosome recoding and construction first require methods for the assembly of the large synthetic DNA. In recent years, the cost of DNA synthesis has steadily decreased, while its fidelity has increased. DNA synthesis methods for synthesizing DNA on a microarray surface have advanced to allow the synthesis of at least 55,000 unique 200-mer oligonucleotides on a single microarray at a cost of about 1$/oligonucleotide (Agilent) [Lynch and Gill, 2012, Song et al., 2021]. Despite the great advances in DNA synthesis in terms of fidelity, length and yield, the technology is still far from synthesizing large DNA fragments or whole chromosomes. Therefore, DNA assembly methods are required to construct such large DNA fragments. Previous approaches for recoding or constructing bacterial chromosomes uses either *in-vitro* methods, *in-vivo* methods or a combination of both [Gibson et al., 2009, Fredens et al., 2019, Gibson et al., 2008]. *In-vitro* methods such as Gibson assembly, which uses exonucleases for suitable overhangs, or Golden-Gate cloning can be used to assemble DNA fragments of up to 583kb and 100kb, respectively [Gibson et al., 2009, Schindler, 2016]. However, Golden-Gate cloning requires that the recognition sites of the restriction enzymes used are absent within the DNA fragments. Furthermore, Gibson assembly requires homologous regions flanking the synthetic DNA, which reduces the size of each fragment. In addition, the handling of such large DNA fragments *in-vitro* can be impractical mainly due to the shear forces that can fragment the DNA [Schindler, 2016]. Therefore, a heterologous host, such as the yeast *Saccharomyces cerevisiae*, is often used above a certain size to assemble DNA fragments *in-vivo* by transformation-associated

recombination cloning [Kouprina and Larionov, 2016]. However, the efficient recombination system of *S. cerevisiae* can lead to unstable DNA constructs due to the small homology size required [Resnick and Nilsson-Tillgren, 1990, Schindler, 2016]. Furthermore, this method is also more time-consuming than *in-vitro* methods as it requires additional isolation and transformation steps.

Once assembled, these large DNA molecules need to be transferred to the corresponding organism. The lack of suitable transformation methods for large DNA molecules can be a bottleneck. The upper limit of bacterial artificial chromosomes (BAC), which can be transformed into *E. coli*, for instance, seems to be about 300kb [Sheng et al., 1995, Gibson et al., 2009]. However, just recently the *E. coli* chromosome could be split into three 1Mb replicons of which one could successfully electroporate into another *E. coli* strain [Yoneji et al., 2021]. The authors assumed, that this was possible due to the physical compactness of the supercoiled form of the split chromosome and that supercoiling is a key feature for the successful electroporation of very large DNA molecules.

In the case of the genome transplantation of the entire *Mycoplasma mycoides* genome into *Mycoplasma capricolum* cells, a time-consuming and complex chemical transformation method has been used [Gibson et al., 2010, Lartigue et al., 2007]. However, this was probably only possible due to the specific physiology of the *Mycoplasma* species. This bacterial species have a small genome of about 1Mb and a total lack of a bacterial cell wall. This helped the scientists to establish existing protocols for eukaryotic genome transplantation due to the resemblance of the cell surface to the plasma membrane of eukaryotic cells. It should be mentioned that the unmethylated synthetic chromosome of *M. mycoides* assembled in *S. cerevisiae* was unprotected against the restriction system of the recipient cell, which was not the case for the native isolated chromosome [Gibson et al., 2010]. Therefore, additional methylation of the synthetic chromosome or the disruption of the recipient cell's restriction system was necessary. This needs to be considered when synthetic DNA is assembled by *in-vitro* assembly methods or *in-vivo* recombination, in which DNA methylation does not occur and the DNA sequence is prone to the host restriction system [Meselson et al., 1972, Wilson, 1991].

Another method for transferring large DNA molecules into a bacterial cell is conjugation. In chapter 3.2, a 151kb chromosomal region was transferred from one cell to another and then integrated into the recipient chromosome with the help of CRISPR/-

Cas9. Another study showed, that it was possible to recode the entire genome of *E. coli* by successive conjugation of *in-vivo* assembled synthetic DNA and subsequent integration into the chromosome [Fredens et al., 2019]. With the ability of CRISPR SWAPnDROP to additionally transfer large DNA fragments from existing replicons, it is possible to assemble or modify bacterial chromosomes on a large scale *in-vivo*. The assembly step in *S. cerevisiae* and the associated elaborate isolation and transformation into the bacterial cell could be omitted. Bacterial cells would first be transformed in parallel with different DNA fragments assembled *in-vitro* using Golden-Gate or modular cloning. CRISPR SWAPnDROP could then be used to transfer each synthetic fragment from the donor strains to the recipient strain and successively assemble or recode the synthetic or native chromosome, respectively. Its feature of stable iterative genome editing would further accelerate and simplify this process.

4.4 How to adapt a CRISPR/Cas9 genome editing tool for other bacterial species

Even though CRISPR/Cas9 technology has been adapted for a variety of bacterial species, there are still species of industrial and scientific interest that cannot, or at least with difficulty, be genetically modified. This is due to the different evolution of living organisms and thus to variations in their metabolism, transcription, DNA repair mechanisms, genomic nucleotide content, codon usage or host defence mechanisms against foreign DNA. These differences often lead to failed attempts in adapting genome editing tools to new organisms, since their biological interaction with the heterologous systems is not known. While in *E. coli* the CRISPR/Cas9 system generally works efficiently for genome editing, it is toxic in *Corynebacterium glutamicum* [Jiang et al., 2017]. It could only be speculated that the cytotoxic effects may result from a CRISPRi-like effect of unspecific Cas9 binding, however, it is not known how this is a problem for *C. glutamicum* but not for *E. coli*. Another study compared the cytotoxic effects of Cas9 and dCas9 expression in five different bacterial species, which showed differential levels of Cas9 toxicity ranging from low cytotoxic effects in *E. coli* and *S. typhimurium* to be highly toxic in the GC rich microbes, *M. smegmatis*, *X. campestris* and *D. radiodurans* [Misra et al., 2019]. Due to the missing information of those interactions,

trial and error approaches are often carried out to adapt genome editing systems to new species. However, there are a few basic steps that can be used as a guideline. The general genetic accessibility of an organism is a prerequisite for successful genome editing approaches. That means genetic transformation methods (electroporation, chemical transformation, natural competence) and functional genetic parts such as origins of replication, promoters, RBS, terminators, and antibiotic resistance markers are available. It is beneficial to have a variety of genetic parts at your disposal as the prospect of success of trial and error approaches increases with the number of possible solutions. A modular cloning approach in combination with a library of genetic parts, as used in chapter 3.2, is of great help in this respect. It facilitates the fast and simple exchange of parts to test them in a variety of combinations on different levels as it is often not known whether the used protein itself is non-functional or toxic, or whether the expression strength is the problem. For the adaption of CRISPR SWAPnDROP to *V. natriegens*, a broad host range origin of replication was combined with a codon-optimized kanamycin resistance gene from the library, which was known to ensure plasmid stability in *V. natriegens* [Stukenberg et al., 2021]. As those parts were already part of the modular library, the first constructs could be tested in *V. natriegens* within three days. Preliminary tests have shown, that the Cas9 transcription unit, which also worked in *E. coli*, was functional in eliminating bacterial cells, however, no recombination events using the λRED expression cassette from *E. coli* could be observed. Therefore, the P_{BAD} promoter from the λRED system was exchanged with the native P_{BAD} promoter of *V. natriegens* to ensure proper transcription of the recombination system, which eventually lead to edited clones. The functionality of the λRED system in *V. natriegens* was not yet reported and underlines again that testing and fine-tuning existing genes via modular libraries can help adapt genome editing systems for new species.

It has to be mentioned that the use of inducible promoters like P_{tet} or P_{BAD} for the expression of the Cas9 protein, sgRNA and the recombination system are recommended if these expression cassettes are to be kept stable on a plasmid in the cell [Reisch and Prather, 2015]. In classical approaches, the plasmid harbouring the *cas9* gene and the sgRNA is co-transformed with the DNA template to avoid premature DSBs [Jiang et al., 2013]. The editing event happens during the transformation, and the control of Cas9 and sgRNA expression is not necessary. However, in the case of iterative genome edits or when large DNA templates decrease the co-transformation efficiency, it is necessary

to stably maintain plasmids with *cas9* and sgRNA in the cell. Therefore, tight control of Cas9 and sgRNA expression to avoid premature DSBs is important. Furthermore, overexpression of Cas9 can have cytotoxic effects and a constitutive expression of phage-derived recombination systems like λRED can lead to an increased mutation rate [Murphy and Campellone, 2003].

Choosing a suitable recombination system is another important factor and often the bottleneck for adapting genome editing systems for new bacterial species. While λRED or RecET work efficiently as recombination systems for a variety of bacteria, they are non-functional or even toxic for many non-model organisms [Corts et al., 2019, Joseph et al., 2018, Guo et al., 2019]. This is not surprising, since both systems derived from *E. coli* phages, which evolved to be functional in their host. In those cases, identifying genes analogous to λRED and the recombination protein RecT in phylogenetically related organisms can help to find functional and efficient recombination systems for the new species [Corts et al., 2019, Joseph et al., 2018, Guo et al., 2019]. Testing other recombination systems like the argonaute protein from *Natronobacterium gregoryi*, which enhances gene insertions or deletions in prokaryotes at high efficiencies, could also be helpful [Fu et al., 2019]. Furthermore, single-stranded DNA-binding proteins (SSBs), which are species specific and play a important role in phage recombination pathways, can be heterologously expressed in the organism of interest [Filsinger et al., 2021]. This can facilitate the interaction of phage-encoded single-stranded DNA annealing proteins (e.g. λRED-β) with a single-stranded DNA template, which may not be the case when using the native SSB protein. With this approach, the genome editing efficiency can be significantly improved using a heterologous recombination system, which would otherwise be highly inefficient or nonfunctional in the host.

In chapter 3.2, CRISPR SWAPnDROP was adapted for the plant pathogen *D. dadantii*, whose native recombination system in combination with classical marker exchange approaches has been used for chromosomal changes so far [Leonard et al., 2021, Jiang et al., 2022]. Initial tests failed to detect efficient recombination with λRED in *D. dadantii*, therefore species-specific recombination systems were sought in related species and strains. A prophage operon of a *D. dadantii* strain was identified, consisting of five genes, two of which show high amino acid homology to λRED (*bet, exo*). One of the other genes coded for a methyl transferase, while the remaining two genes were of unknown function. To investigate the functionality of the operon in *D. dadantii*, the operon was split into different versions harbouring either only the *bet* and *exo*

homologous genes, *bet* and *exo* in combination with the methyl transferase gene or the complete operon. All versions were placed under the control of the *E. coli* P_{BAD} promoter with the help of the modular MoCloFlex cloning approach and each version was tested for its genome editing capability [Klein et al., 2019]. The functionality of the *E. coli* P_{BAD} promoter was previously tested using a fluorescence reporter system in *D. dadantii*. Transformation of *D. dadantii* cells with the cr3Dd plasmid harbouring the complete operon was not possible indicating high toxicity of the complete operon and also the *bet+exo*+methyl transferase version showed a strong reduction of viable cells upon induction. Only the use of the *bet+exo* version showed no toxic effects and resulted in a high editing efficiency of up to 100%.

4.5 Conclusion

Bacterial gene regulation occurs on many levels ranging from modulations by transcription factors to modulations by chromosome organization. To better understand how those mechanisms constitute the entire regulatory system, systemic approaches, which cover the complete genetic landscape of an organism, can be helpful.

The results presented in this work contributed to the understanding of the replication-induced copy number effect in bacteria and its impact on gene regulation and chromosome evolution. The systemic analysis provided new insights into how and to what extent bacteria utilize this effect.

Furthermore, the genome editing tool CRISPR SWAPnDROP, which was developed in the course of this work, facilitates a simple and efficient way to modify bacterial genomes. Its set of features outcompetes other tools in terms of applicability and versatility. The modular assembly helps users to establish a genome editing tool for new bacterial species, which is often complicated and time-consuming. Additionally, the unique mechanism of CRISPR SWAPnDROP allows the transfer and rearrangement of large chromosomal regions, which is of great benefit for chromosome research and synthetic biology.

5 Summary

Research in recent years has yielded new insights into the influence of chromosomal architecture at different levels on bacterial gene regulation and expression. On the topological level, chromosome compaction can bring distant genes or regions in spatially proximity, suggesting a regulatory concept of co-expression of distant genes. DNA supercoiling, which is highly dynamic and one of the major factors in nucleoid formation, can have a significant influence on gene expression by modulatory effects on transcription. Furthermore, the replication-induced copy number effect increases the expression of genes by transiently increasing the number of gene copies during replication. However, how this impacts the organism on a systemic level (global gene expression) has not been shown yet.

In this work, the influence of the replication-induced copy number effect on gene expression in *Escherichia coli* has been investigated. It was previously shown, that genes closer to the replication origin (oriC) are higher expressed during the exponential phase compared to the stationary phase. This effect decreases with increasing distance to the oriC. In the course of this work, it was demonstrated that this expression pattern is due to the copy number effect instead of the strategic positioning of genes regulated by global transcription factors. Furthermore, the regulatory impact of the replication-induced copy number effect was determined for individual genes. It could be shown that around 40% of the genes are predominantly copy number regulated, suggesting an important role of the copy number effect for gene regulation and expression in *E. coli*. In addition, the influence of the copy number effect on the chromosome organization was investigated. The conservation of the position of genes relative to oriC indicates a strong influence of the copy number effect on bacterial chromosome evolution, especially in fast-growing bacteria.

Moreover, a genome editing tool based on CRISPR/Cas9 (CRISPR SWAPnDROP) was established for the chromosomal modifications required for these investigations.

Beyond its initial purpose, this tool was designed to facilitate large chromosomal rearrangements and the transfer of chromosomal regions between bacterial species. As a proof of principle, a 151kb chromosomal region was transferred from one *E. coli* strain to another as well as to the biotechnology relevant *Vibrio natriegens*. In addition, the RP4 conjugation system of *E. coli* was transferred to both *V. natriegens* and the plant pathogen *Dickeya dadantii* and its functionality was demonstrated in these organisms. Furthermore, the transfer of the *E. coli* lac operon to *V. natriegens* and the transfer of the GanB ORF from *D. dadantii* to *E. coli* demonstrate successful gain of function genome edits using CRISPR SWAPnDROP.

6 Zusammenfassung

Die Forschung der letzten Jahre hat neue Erkenntnisse über den Einfluss der chromosomalen Architektur auf die bakterielle Genregulation und -expression erbracht. Auf der topologischen Ebene kann die Chromosomenkompaktierung weit voneinander entfernte Gene oder Regionen in räumliche Nähe bringen, was auf ein Regulierungskonzept der Koexpression entfernter Gene hindeutet. DNA-Supercoiling, das hochdynamisch und einer der Hauptfaktoren bei der Nukleoidbildung ist, kann die Genexpression durch modulierende Effekte auf die Transkription erheblich beeinflussen. Außerdem erhöht der durch die Replikation hervorgerufene Kopienzahl-Effekt die Expression von Genen, indem er die Anzahl der Genkopien während der Replikation vorübergehend erhöht. Wie sich dies auf den Organismus auf einer systemischen Ebene (globale Genexpression) auswirkt, wurde jedoch noch nicht untersucht.

In dieser Arbeit wurde der Einfluss des Kopienzahl-Effekts auf die Genexpression in *Escherichia coli* untersucht. Frühere Publikationen haben gezeigt, dass Gene, die näher am Replikationsursprung (oriC) liegen, in der exponentiellen Phase stärker exprimiert werden als in der stationären Phase. Dieser Effekt nimmt mit zunehmender Entfernung zum Replikationsursprung ab. Im Zuge dieser Arbeit konnte nachgewiesen werden, dass dieses Expressionsmuster auf den Kopienzahl-Effekt zurückzuführen ist und nicht auf die strategische Positionierung von Genen, die durch globale Transkriptionsfaktoren reguliert werden. Darüber hinaus wurde der regulatorische Einfluss des Kopienzahl-Effekts für einzelne Gene bestimmt. Es konnte gezeigt werden, dass etwa 40% der Gene überwiegend über die Kopienzahl reguliert sind, was auf eine wichtige Rolle des Kopienzahl-Effekts für die Genregulation und -expression in *E. coli* hinweist. Darüber hinaus wurde der Einfluss des Kopienzahl-Effekts auf die Chromosomenorganisation untersucht. Die Konservierung der Position von Genen relativ zum Replikationsursprung deutet auf einen starken Einfluss des Kopienzahl-Effekts auf die bakterielle Chromosomenevolution hin, insbesondere bei schnell wachsenden

Bakterien.

Darüber hinaus wurde ein auf CRISPR/Cas9 basierendes *Genome Editing Tool* (CRISPR SWAPnDROP) etabliert, das für die erforderlichen chromosomalen Veränderungen dieser Untersuchungen benötigt wurde. Dieses *Tool* wurde außerdem entwickelt, um große chromosomale Veränderungen und den Transfer von chromosomalen Regionen zwischen Bakterienarten zu ermöglichen. Als *proof of principle* wurde eine 151kb große chromosomale Region sowohl zwischen zwei *E. coli*-Stämmen als auch von *E. coli* auf das biotechnologisch relevante Bakterium *Vibrio natriegens* übertragen. Zusätzlich wurde das RP4-Konjugationssystem von *E. coli* sowohl auf *V. natriegens* als auch auf das Pflanzenpathogen *Dickeya dadantii* übertragen und seine Funktionalität in diesen Organismen nachgewiesen. Abschließend konnten mit der Übertragung des lac-Operons von *E. coli* auf *V. natriegens* und der Übertragung des GanB ORFs von *D. dadantii* auf *E. coli* erfolgreiche *gain of fucntion* Genomeditierungen gezeigt werden.

Bibliography

[Ambler and Scott, 1978] Ambler, R. P. and Scott, G. K. (1978). Partial amino acid sequence of penicillinase coded by Escherichia coli plasmid R6K. Proc Natl Acad Sci U S A, 75(8):3732–3736.

[Anders et al., 2014] Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature, 513(7519):569–573.

[Arroyo-Olarte et al., 2021] Arroyo-Olarte, R. D., Bravo Rodríguez, R., and Morales-Ríos, E. (2021). Genome Editing in Bacteria: CRISPR-Cas and Beyond. Microorganisms, 9(4).

[Ayora et al., 2011] Ayora, S., Carrasco, B., Cárdenas, P. P., César, C. E., Cañas, C., Yadav, T., Marchisone, C., and Alonso, J. C. (2011). Double-strand break repair in bacteria: a view from Bacillus subtilis. FEMS Microbiol Rev, 35(6):1055–1081.

[Balke and Gralla, 1987] Balke, V. L. and Gralla, J. (1987). Changes in the linking number of supercoiled dna accompany growth transitions in escherichia coli. Journal of bacteriology, 169(10):4499–4506.

[Ball et al., 1992] Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C. (1992). Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. J Bacteriol, 174(24):8043–8056.

[Bigot et al., 2007] Bigot, S., Sivanathan, V., Possoz, C., Barre, F. X., and Cornet, F. (2007). FtsK, a literate chromosome segregation machine. Mol Microbiol, 64(6):1434–1441.

[Block et al., 2012] Block, D. H., Hussein, R., Liang, L. W., and Lim, H. N. (2012). Regulatory consequences of gene translocation in bacteria. Nucleic Acids Res, 40(18):8979–8992.

[Blokesch, 2016] Blokesch, M. (2016). Natural competence for transformation. Curr. Biol., 26(21):R1126–R1130.

[Blot et al., 2006] Blot, N., Mavathur, R., Geertz, M., Travers, A., and Muskhelishvili, G. (2006). Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. EMBO Rep, 7(7):710–715.

[Bouyioukos et al., 2016] Bouyioukos, C., Reverchon, S., and Képès, F. (2016). From multiple pathogenicity islands to a unique organized pathogenicity archipelago. Sci Rep, 6:27978.

[Browning and Busby, 2016] Browning, D. F. and Busby, S. J. (2016). Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol, 14(10):638–650.

[Bryant et al., 2014] Bryant, J. A., Sellars, L. E., Busby, S. J., and Lee, D. J. (2014). Chromosome position effects on gene expression in Escherichia coli K-12. Nucleic Acids Res, 42(18):11383–11392.

[Cavaliere and Norel, 2016] Cavaliere, P. and Norel, F. (2016). Recent advances in the characterization of Crl, the unconventional activator of the stress sigma factor σS/RpoS. Biomol Concepts, 7(3):197–204.

[Chakraborty and Kenney, 2018] Chakraborty, S. and Kenney, L. J. (2018). A new role of OmpR in acid and osmotic stress in salmonella and e. coli. Front. Microbiol., 9:2656.

[Chandler and Pritchard, 1975] Chandler, M. G. and Pritchard, R. H. (1975). The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. Mol Gen Genet, 138(2):127–141.

[Chen et al., 2013] Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell, 155(7):1479–1491.

[Chen and Lilley, 1999] Chen, D. and Lilley, D. M. (1999). Transcription-induced hypersupercoiling of plasmid dna. Journal of molecular biology, 285(2):443–448.

[Cheng et al., 2013] Cheng, A. W., Wang, H., Yang, H., Shi, L., Katz, Y., Theunissen, T. W., Rangarajan, S., Shivalila, C. S., Dadon, D. B., and Jaenisch, R. (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Res, 23(10):1163–1171.

[Cho et al., 2018] Cho, S., Choe, D., Lee, E., Kim, S. C., Palsson, B., and Cho, B. K. (2018). High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli. ACS Synth Biol, 7(4):1085–1094.

[Condon et al., 1995] Condon, C., Liveris, D., Squires, C., Schwartz, I., and Squires, C. L. (1995). rrna operon multiplicity in escherichia coli and the physiological implications of rrn inactivation. Journal of bacteriology, 177(14):4152–4156.

[Cong et al., 2013] Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science, 339(6121):819–823.

[Corts et al., 2019] Corts, A. D., Thomason, L. C., Gill, R. T., and Gralnick, J. A. (2019). A new recombineering system for precise genome-editing in Shewanella oneidensis strain MR-1 using single-stranded oligonucleotides. Sci Rep, 9(1):39.

[Costantino and Court, 2003] Costantino, N. and Court, D. L. (2003). Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci U S A, 100(26):15748–15753.

[Court et al., 2002] Court, D. L., Sawitzke, J. A., and Thomason, L. C. (2002). Genetic engineering using homologous recombination. Annu Rev Genet, 36:361–388.

[Couturier and Rocha, 2006] Couturier, E. and Rocha, E. P. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. Mol Microbiol, 59(5):1506–1518.

[Cox, 1983] Cox, M. M. (1983). The FLP protein of the yeast 2-microns plasmid: expression of a eukaryotic genetic recombination system in Escherichia coli. Proc Natl Acad Sci U S A, 80(14):4223–4227.

[Darling et al., 2008] Darling, A. E., Miklós, I., and Ragan, M. A. (2008). Dynamics of genome rearrangement in bacterial populations. PLoS Genet, 4(7):e1000128.

[de Vries, 2010] de Vries, R. (2010). DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. Biochimie, 92(12):1715–1721.

[Deweese et al., 2008] Deweese, J. E., Osheroff, M. A., and Osheroff, N. (2008). DNA Topology and Topoisomerases: Teaching a "Knotty" Subject. Biochem Mol Biol Educ, 37(1):2–10.

[Ding et al., 2020] Ding, W., Zhang, Y., and Shi, S. (2020). Development and Application of CRISPR/Cas in Microbial Biotechnology. Front Bioeng Biotechnol, 8:711.

[Dorman, 2013] Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: making progress towards a unified model? Nat Rev Microbiol, 11(5):349–355.

[Dorman, 2019] Dorman, C. J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. BMC Mol Cell Biol, 20(1):26.

[Egan et al., 2004] Egan, E. S., Løbner-Olesen, A., and Waldor, M. K. (2004). Synchronous replication initiation of the two vibrio cholerae chromosomes. Current Biology, 14(13):R501–R502.

[Eisen et al., 2000] Eisen, J. A., Heidelberg, J. F., White, O., and Salzberg, S. L. (2000). Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol, 1(6):RESEARCH0011.

[El Houdaigui et al., 2019] El Houdaigui, B., Forquet, R., Hindré, T., Schneider, D., Nasser, W., Reverchon, S., and Meyer, S. (2019). Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. Nucleic Acids Res, 47(11):5648–5657.

[Ellis et al., 2001] Ellis, H. M., Yu, D., DiTizio, T., and Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A, 98(12):6742–6746.

[Engler et al., 2008] Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. PLoS One, 3(11):e3647.

[Engler et al., 2014] Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014). A golden gate modular cloning toolbox for plants. ACS Synth Biol, 3(11):839–843.

[Farewell et al., 1998] Farewell, A., Kvint, K., and Nyström, T. (1998). Negative regulation by RpoS: a case of sigma factor competition. Mol. Microbiol., 29(4):1039–1051.

[Fels et al., 2020] Fels, U., Gevaert, K., and Van Damme, P. (2020). Bacterial Genetic Engineering by Means of Recombineering for Reverse Genetics. Front Microbiol, 11:548410.

[Fijalkowska et al., 2012] Fijalkowska, I. J., Schaaper, R. M., and Jonczyk, P. (2012). DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. FEMS Microbiol Rev, 36(6):1105–1121.

[Filsinger et al., 2021] Filsinger, G. T., Wannier, T. M., Pedersen, F. B., Lutz, I. D., Zhang, J., Stork, D. A., Debnath, A., Gozzi, K., Kuchwara, H., Volf, V., Wang, S., Rios, X., Gregg, C. J., Lajoie, M. J., Shipman, S. L., Aach, J., Laub, M. T., and Church, G. M. (2021). Characterizing the portability of phage-encoded homologous recombination proteins. Nat Chem Biol, 17(4):394–402.

[Forquet et al., 2021] Forquet, R., Pineau, M., Nasser, W., Reverchon, S., and Meyer, S. (2021). Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters. mSystems, 6(4):e0097821.

[Fredens et al., 2019] Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., Chia, T., Schmied, W. H., Dunkelmann, D. L., Beránek, V., Uttamapinant, C., Llamazares, A. G., Elliott, T. S., and Chin, J. W. (2019). Total synthesis of Escherichia coli with a recoded genome. Nature, 569(7757):514–518.

[French, 1992] French, S. (1992). Consequences of replication fork movement through transcription units in vivo. Science, 258(5086):1362–1365.

[Fu et al., 2019] Fu, L., Xie, C., Jin, Z., Tu, Z., Han, L., Jin, M., Xiang, Y., and Zhang, A. (2019). The prokaryotic Argonaute proteins enhance homology sequence-directed recombination in bacteria. Nucleic Acids Res, 47(7):3568–3579.

[Georis, 2013] Georis, I. (2013). Catabolite gene activator protein. In Brenner's Encyclopedia of Genetics, pages 442–446. Elsevier.

[Gibson et al., 2008] Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science, 319(5867):1215–1220.

[Gibson et al., 2010] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O., and Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. Science, 329(5987):52–56.

[Gibson et al., 2009] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods, 6(5):343–345.

[Glinkowska et al., 2021] Glinkowska, M., Waldminghaus, T., and Riber, L. (2021). Editorial: Bacterial Chromosomes Under Changing Environmental Conditions. Front Microbiol, 12:633466.

[Gorbunova and Levy, 1997] Gorbunova, V. and Levy, A. A. (1997). Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res, 25(22):4650–4657.

[Gowrishankar, 2015] Gowrishankar, J. (2015). End of the beginning: elongation and termination features of alternative modes of chromosomal replication initiation in bacteria. PLoS Genet, 11(1):e1004909.

[Gruber and Gross, 2003] Gruber, T. M. and Gross, C. A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol.

[Guet et al., 2002] Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002). Combinatorial synthesis of genetic networks. Science, 296(5572):1466–1470.

[Guo et al., 2018] Guo, J., Wang, T., Guan, C., Liu, B., Luo, C., Xie, Z., Zhang, C., and Xing, X. H. (2018). Improved sgRNA design in bacteria via genome-wide activity profiling. Nucleic Acids Res, 46(14):7052–7069.

[Guo et al., 2019] Guo, T., Xin, Y., Zhang, Y., Gu, X., and Kong, J. (2019). A rapid and versatile tool for genomic engineering in Lactococcus lactis. Microb Cell Fact, 18(1):22.

[Hamilton et al., 1989] Hamilton, C. M., Aldea, M., Washburn, B. K., Babitzke, P., and Kushner, S. R. (1989). New method for generating deletions and gene replacements in Escherichia coli. J Bacteriol, 171(9):4617–4622.

[Hartl et al., 2017] Hartl, J., Kiefer, P., Meyer, F., and Vorholt, J. A. (2017). Longevity of major coenzymes allows minimal de novo synthesis in microorganisms. Nat Microbiol, 2:17073.

[Hatfield and Benham, 2002] Hatfield, G. W. and Benham, C. J. (2002). DNA topology-mediated control of global gene expression in Escherichia coli. Annu Rev Genet, 36:175–203.

[Hiom, 2009] Hiom, K. (2009). DNA repair: common approaches to fixing double-strand breaks. Curr Biol, 19(13):R523–525.

[Hoess and Abremski, 1985] Hoess, R. H. and Abremski, K. (1985). Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. J Mol Biol, 181(3):351–362.

[Hu et al., 2018] Hu, J. H., Miller, S. M., Geurts, M. H., Tang, W., Chen, L., Sun, N., Zeina, C. M., Gao, X., Rees, H. A., Lin, Z., and Liu, D. R. (2018). Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature, 556(7699):57–63.

[Huang and Ito, 1999] Huang, Y.-P. and Ito, J. (1999). Dna polymerase c of the thermophilic bacterium thermus aquaticus: classification and phylogenetic analysis of the family c dna polymerases. Journal of molecular evolution, 48(6):756–769.

[Ishihama, 2000] Ishihama, A. (2000). Functional modulation of escherichia coli RNA polymerase. Annu. Rev. Microbiol., 54(1):499–518.

[Ivanova et al., 2015] Ivanova, D., Taylor, T., Smith, S. L., Dimude, J. U., Upton, A. L., Mehrjouy, M. M., Skovgaard, O., Sherratt, D. J., Retkute, R., and Rudolph, C. J. (2015). Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. Nucleic Acids Res, 43(16):7865–7877.

[Jalal and Le, 2020] Jalal, A. S. B. and Le, T. B. K. (2020). Bacterial chromosome segregation by the ParABS system. Open Biol, 10(6):200097.

[Jiang et al., 2022] Jiang, D., Zeng, Q., Banerjee, B., Lin, H., Srok, J., Yu, M., and Yang, C. H. (2022). The phytopathogen Dickeya dadantii 3937 cpxR locus gene participates in the regulation of virulence and the global c-di-GMP network. Mol Plant Pathol.

[Jiang et al., 2013] Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol, 31(3):233–239.

[Jiang et al., 2017] Jiang, Y., Qian, F., Yang, J., Liu, Y., Dong, F., Xu, C., Sun, B., Chen, B., Xu, X., Li, Y., Wang, R., and Yang, S. (2017). CRISPR-Cpf1 assisted genome editing of Corynebacterium glutamicum. Nat Commun, 8:15179.

[Jinek et al., 2012] Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 337(6096):816–821.

[Jinks-Robertson and Nomura, 1987] Jinks-Robertson, S. and Nomura, M. (1987). Ribosomes and trna. escherichia coli and salmonella typhimurium molecular and cellular biology.

[Jishage et al., 2002] Jishage, M., Kvint, K., Shingler, V., and Nyström, T. (2002). Regulation of sigma factor competition by the alarmone ppGpp. Genes Dev, 16(10):1260–1270.

[Joseph et al., 2018] Joseph, R. C., Kim, N. M., and Sandoval, N. R. (2018). Recent developments of the synthetic biology toolkit for clostridium. Front. Microbiol., 9:154.

[Klein et al., 2019] Klein, C. A., Emde, L., Kuijpers, A., and Sobetzko, P. (2019). Mo-CloFlex: A Modular Yet Flexible Cloning System. Front Bioeng Biotechnol, 7:271.

[Klein et al., 2021] Klein, C. A., Teufel, M., Weile, C. J., and Sobetzko, P. (2021). The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. Sci Rep, 11(1):24399.

[Klumpp et al., 2009] Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. Cell, 139(7):1366–1375.

[Kouprina and Larionov, 2016] Kouprina, N. and Larionov, V. (2016). Transformation-associated recombination (TAR) cloning for genomics studies and synthetic biology. Chromosoma, 125(4):621–632.

[Krogh et al., 2020] Krogh, T. J., Franke, A., Møller-Jensen, J., and Kaleta, C. (2020). Elucidating the Influence of Chromosomal Architecture on Transcriptional Regulation in Prokaryotes - Observing Strong Local Effects of Nucleoid Structure on Gene Regulation. Front Microbiol, 11:2002.

[Krogh et al., 2018] Krogh, T. J., Møller-Jensen, J., and Kaleta, C. (2018). Impact of Chromosomal Architecture on the Function and Evolution of Bacterial Genomes. Front Microbiol, 9:2019.

[Képès, 2004] Képès, F. (2004). Periodic transcriptional organization of the E.coli genome. J Mol Biol, 340(5):957–964.

[Lagomarsino et al., 2015] Lagomarsino, M. C., Espéli, O., and Junier, I. (2015). From structure to function of bacterial chromosomes: Evolutionary perspectives and ideas for new experiments. FEBS Lett, 589(20 Pt A):2996–3004.

[Larson et al., 2013] Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc, 8(11):2180–2196.

[Lartigue et al., 2007] Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., Smith, H. O., and Venter, J. C. (2007). Genome transplantation in bacteria: changing one species to another. Science, 317(5838):632–638.

[Lato and Golding, 2020] Lato, D. F. and Golding, G. B. (2020). Spatial Patterns of Gene Expression in Bacterial Genomes. J Mol Evol, 88(6):510–520.

[Lawrence and Ochman, 1998] Lawrence, J. G. and Ochman, H. (1998). Molecular archaeology of the Escherichia coli genome. Proc Natl Acad Sci U S A, 95(16):9413–9417.

[Le et al., 2013] Le, T. B., Imakaev, M. V., Mirny, L. A., and Laub, M. T. (2013). High-resolution mapping of the spatial organization of a bacterial chromosome. Science, 342(6159):731–734.

[Le Berre et al., 2022] Le Berre, D., Reverchon, S., Muskhelishvili, G., and Nasser, W. (2022). Relationship between the Chromosome Structural Dynamics and Gene Expression-A Chicken and Egg Dilemma? Microorganisms, 10(5).

[Leonard et al., 2021] Leonard, S., Villard, C., Nasser, W., Reverchon, S., and Hommais, F. (2021). RNA chaperones hfq and ProQ play a key role in the virulence of the plant pathogenic bacterium dickeya dadantii. Front. Microbiol., 12:687484.

[Li et al., 2020] Li, N., Zeng, W., Xu, S., and Zhou, J. (2020). Toward fine-tuned metabolic networks in industrial microorganisms. Synth Syst Biotechnol, 5(2):81–91.

[Lieber, 2008] Lieber, M. R. (2008). The mechanism of human nonhomologous DNA end joining. J Biol Chem, 283(1):1–5.

[Lim et al., 2014] Lim, C. J., Kenney, L. J., and Yan, J. (2014). Single-molecule studies on the mechanical interplay between DNA supercoiling and H-NS DNA architectural properties. Nucleic Acids Res, 42(13):8369–8378.

[Liu and Wang, 1987] Liu, L. F. and Wang, J. C. (1987). Supercoiling of the dna template during transcription. Proceedings of the National Academy of Sciences, 84(20):7024–7027.

[Lu et al., 2019] Lu, Z., Yang, S., Yuan, X., Shi, Y., Ouyang, L., Jiang, S., Yi, L., and Zhang, G. (2019). CRISPR-assisted multi-dimensional regulation for fine-tuning gene expression in Bacillus subtilis. Nucleic Acids Res, 47(7):e40.

[Lucchini et al., 2006] Lucchini, S., Rowley, G., Goldberg, M. D., Hurd, D., Harrison, M., and Hinton, J. C. (2006). H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog, 2(8):e81.

[Lynch and Gill, 2012] Lynch, S. A. and Gill, R. T. (2012). Synthetic biology: new strategies for directing design. Metab Eng, 14(3):205–211.

[Maduike et al., 2014] Maduike, N. Z., Tehranchi, A. K., Wang, J. D., and Kreuzer, K. N. (2014). Replication of the Escherichia coli chromosome in RNase HI-deficient cells: multiple initiation regions and fork dynamics. Mol Microbiol, 91(1):39–56.

[Madyagol et al., 2011] Madyagol, M., Al-Alami, H., Levarski, Z., Drahovská, H., Turňa, J., and Stuchlík, S. (2011). Gene replacement techniques for Escherichia coli genome modification. Folia Microbiol (Praha), 56(3):253–263.

[Maeda et al., 2000] Maeda, H., Fujita, N., and Ishihama, A. (2000). Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res, 28(18):3497–3503.

[Malan et al., 1984] Malan, T. P., Kolb, A., Buc, H., and McClure, W. R. (1984). Mechanism of CRP-cAMP activation of lac operon transcription initiation activation of the P1 promoter. J Mol Biol, 180(4):881–909.

[Marbouty et al., 2015] Marbouty, M., Le Gall, A., Cattoni, D. I., Cournac, A., Koh, A., Fiche, J. B., Mozziconacci, J., Murray, H., Koszul, R., and Nollmann, M. (2015). Condensin- and Replication-Mediated Bacterial Chromosome Folding and Origin Condensation Revealed by Hi-C and Super-resolution Imaging. Mol Cell, 59(4):588–602.

[Maresca et al., 2010] Maresca, M., Erler, A., Fu, J., Friedrich, A., Zhang, Y., and Stewart, A. F. (2010). Single-stranded heteroduplex intermediates in lambda Red homologous recombination. BMC Mol Biol, 11:54.

[Masters and Pardee, 1965] Masters, M. and Pardee, A. B. (1965). Sequence of enzyme synthesis and gene replication during the cell cycle of Bacillus subtilis. Proc Natl Acad Sci U S A, 54(1):64–70.

[Mathelier and Carbone, 2010] Mathelier, A. and Carbone, A. (2010). Chromosomal periodicity and positional networks of genes in Escherichia coli. Mol Syst Biol, 6:366.

[Matthews and Nichols, 1998] Matthews, K. S. and Nichols, J. C. (1998). Lactose repressor protein: functional properties and structure. Prog Nucleic Acid Res Mol Biol, 58:127–164.

[Mercier et al., 2008] Mercier, R., Petit, M. A., Schbath, S., Robin, S., El Karoui, M., Boccard, F., and Espéli, O. (2008). The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. Cell, 135(3):475–485.

[Meselson et al., 1972] Meselson, M., Yuan, R., and Heywood, J. (1972). Restriction and modification of DNA. Annu Rev Biochem, 41:447–466.

[Misra et al., 2019] Misra, C. S., Bindal, G., Sodani, M., Wadhawan, S., Kulkarni, S., Gautam, S., Mukhopadhyaya, R., and Rath, D. (2019). Determination of Cas9/dCas9 associated toxicity in microbes.

[Moore et al., 2016] Moore, S. J., Lai, H. E., Kelwick, R. J., Chee, S. M., Bell, D. J., Polizzi, K. M., and Freemont, P. S. (2016). EcoFlex: A Multifunctional MoClo Kit for E. coli Synthetic Biology. ACS Synth Biol, 5(10):1059–1069.

[Mosberg et al., 2010] Mosberg, J. A., Lajoie, M. J., and Church, G. M. (2010). Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics, 186(3):791–799.

[Murphy, 2016] Murphy, K. C. (2016). λ Recombination and Recombineering. EcoSal Plus, 7(1).

[Murphy and Campellone, 2003] Murphy, K. C. and Campellone, K. G. (2003). Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. BMC Mol Biol, 4:11.

[Muskhelishvili and Travers, 2003] Muskhelishvili, G. and Travers, A. (2003). Transcription factor as a topological homeostat. Front Biosci, 8:d279–285.

[Nishimasu et al., 2014] Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of cas9 in complex with guide RNA and target DNA. Cell, 156(5):935–949.

[Notley-McRobb et al., 1997] Notley-McRobb, L., Death, A., and Ferenci, T. (1997). The relationship between external glucose concentration and cAMP levels inside Escherichia coli: implications for models of phosphotransferase-mediated regulation of adenylate cyclase. Microbiology (Reading), 143 (Pt 6):1909–1918.

[Ogden et al., 1980] Ogden, S., Haggerty, D., Stoner, C. M., Kolodrubetz, D., and Schleif, R. (1980). The Escherichia coli L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. Proc Natl Acad Sci U S A, 77(6):3346–3350.

[Omont and Képès, 2004] Omont, N. and Képès, F. (2004). Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. Bioinformatics, 20(16):2719–2725.

[Park et al., 2013] Park, D. M., Akhtar, M. S., Ansari, A. Z., Landick, R., and Kiley, P. J. (2013). The bacterial response regulator ArcA uses a diverse binding site architecture to regulate carbon oxidation globally. PLoS Genet, 9(10):e1003839.

[Pines et al., 2015] Pines, G., Freed, E. F., Winkler, J. D., and Gill, R. T. (2015). Bacterial Recombineering: Genome Engineering via Phage-Based Homologous Recombination. ACS Synth Biol, 4(11):1176–1185.

[Price et al., 2005] Price, M. N., Alm, E. J., and Arkin, A. P. (2005). Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. Nucleic Acids Res, 33(10):3224–3234.

[Pryor et al., 2020] Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., and Lohman, G. J. S. (2020). Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. PLoS One, 15(9):e0238592.

[Pérez-González et al., 2017] Pérez-González, A., Kniewel, R., Veldhuizen, M., Verma, H. K., Navarro-Rodríguez, M., Rubio, L. M., and Caro, E. (2017). Adaptation of the GoldenBraid modular cloning system and creation of a toolkit for the expression of heterologous proteins in yeast mitochondria. BMC Biotechnol, 17(1):80.

[Pósfai et al., 1999] Pósfai, G., Kolisnychenko, V., Bereczki, Z., and Blattner, F. R. (1999). Markerless gene replacement in Escherichia coli stimulated by a double-strand break in the chromosome. Nucleic Acids Res, 27(22):4409–4415.

[Ramnaresh Gupta and Chatterji, 2016] Ramnaresh Gupta, K. and Chatterji, D. (2016). Sigma factor competition inescherichia coli: Kinetic and thermodynamic perspectives. In Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria, pages 196–202. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[Ran et al., 2013a] Ran, F. A., Hsu, P. D., Lin, C. Y., Gootenberg, J. S., Konermann, S., Trevino, A. E., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013a). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 154(6):1380–1389.

[Ran et al., 2013b] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013b). Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 8(11):2281–2308.

[Reisch and Prather, 2015] Reisch, C. R. and Prather, K. L. (2015). The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in Escherichia coli. Sci Rep, 5:15096.

[Reiss et al., 1984] Reiss, B., Sprengel, R., and Schaller, H. (1984). Protein fusions with the kanamycin resistance gene from transposon Tn5. EMBO J, 3(13):3317–3322.

[Resnick and Nilsson-Tillgren, 1990] Resnick, M. A. and Nilsson-Tillgren, T. (1990). DNA homology and chromosome stability: a sensitive yeast genetic system for identifying double-stranded DNA damage. Prog Clin Biol Res, 340B:363–369.

[Reyes-Lamothe et al., 2008] Reyes-Lamothe, R., Wang, X., and Sherratt, D. (2008). Escherichia coli and its chromosome. Trends Microbiol, 16(5):238–245.

[Riebet and Raibaud, 1991] Riebet, E. and Raibaud, O. (1991). Supercoiling is essential for the formation and stability of the initiation complex at the divergent malep and malkp promoters. Journal of molecular biology, 218(3):529–542.

[Rocha and Danchin, 2003] Rocha, E. P. C. and Danchin, A. (2003). Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet., 34(4):377–378.

[Roller et al., 2016] Roller, B. R., Stoddard, S. F., and Schmidt, T. M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nat Microbiol, 1(11):16160.

[Saier, 1998] Saier, M. H. (1998). Multiple mechanisms controlling carbon metabolism in bacteria. Biotechnol Bioeng, 58(2-3):170–174.

[Santero et al., 1992] Santero, E., Hoover, T. R., North, A. K., Berger, D. K., Porter, S. C., and Kustu, S. (1992). Role of integration host factor in stimulating transcription from the sigma 54-dependent nifH promoter. J Mol Biol, 227(3):602–620.

[Santos-Zavaleta et al., 2019] Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., García-Sotelo, J. S., Alquicira-Hernández, K., Muñiz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D. A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C. F., Galagan, J., and Collado-Vides, J. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. Nucleic Acids Res, 47(D1):D212–D220.

[Sauer et al., 2016] Sauer, C., Syvertsson, S., Bohorquez, L. C., Cruz, R., Harwood, C. R., van Rij, T., and Hamoen, L. W. (2016). Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis. ACS Synth Biol, 5(9):942–947.

[Sawitzke et al., 2007] Sawitzke, J. A., Thomason, L. C., Costantino, N., Bubunenko, M., Datta, S., and Court, D. L. (2007). Recombineering: in vivo genetic engineering in E. coli, S. enterica, and beyond. Methods Enzymol, 421:171–199.

[Schellhorn, 2020] Schellhorn, H. E. (2020). Function, evolution, and composition of the RpoS regulon in escherichia coli. Front. Microbiol., 11:560099.

[Schindler, 2016] Schindler, D. (2016). Konstruktion synthetischer sekundärer chromosomen zur charakterisierung von dna-reparatur und segregation in escherichia coli. [PhD thesis, Philipps-Universität Marburg], Retrieved from http://archiv.ub.uni-marburg.de/diss/z2017/0043/pdf/dds.pdf.

[Scholz et al., 2019] Scholz, S. A., Diao, R., Wolfe, M. B., Fivenson, E. M., Lin, X. N., and Freddolino, P. L. (2019). High-Resolution Mapping of the Escherichia coli Chromosome Reveals Positions of High and Low Transcription. Cell Syst, 8(3):212–225.

[Seligman et al., 1997] Seligman, L. M., Stephens, K. M., Savage, J. H., and Monnat, R. J. (1997). Genetic analysis of the Chlamydomonas reinhardtii I-CreI mobile intron homing system in Escherichia coli. Genetics, 147(4):1653–1664.

[Sharma and Chatterji, 2010] Sharma, U. K. and Chatterji, D. (2010). Transcriptional switching in Escherichia coli during stress and starvation by modulation of sigma activity. FEMS Microbiol Rev, 34(5):646–657.

[Sheng et al., 1995] Sheng, Y., Mancino, V., and Birren, B. (1995). Transformation of Escherichia coli with large DNA molecules by electroporation. Nucleic Acids Res, 23(11):1990–1996.

[Sinden and Pettijohn, 1981] Sinden, R. R. and Pettijohn, D. E. (1981). Chromosomes in living Escherichia coli cells are segregated into domains of supercoiling. Proc Natl Acad Sci U S A, 78(1):224–228.

[Singh et al., 2016] Singh, K., Milstein, J. N., and Navarre, W. W. (2016). Xenogeneic Silencing and Its Impact on Bacterial Genomes. Annu Rev Microbiol, 70:199–213.

[Skarstad and Katayama, 2013] Skarstad, K. and Katayama, T. (2013). Regulating DNA replication in bacteria. Cold Spring Harb Perspect Biol, 5(4):a012922.

[Skoko et al., 2006] Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S. M., Marko, J. F., and Johnson, R. C. (2006). Mechanism of chromosome compaction and looping by the Escherichia coli nucleoid protein Fis. J Mol Biol, 364(4):777–798.

[Sobetzko, 2016] Sobetzko, P. (2016). Transcription-coupled DNA supercoiling dictates the chromosomal arrangement of bacterial genes. Nucleic Acids Res, 44(4):1514–1524.

[Sobetzko et al., 2013] Sobetzko, P., Glinkowska, M., Travers, A., and Muskhelishvili, G. (2013). DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. Mol Biosyst, 9(7):1643–1651.

[Sobetzko et al., 2012] Sobetzko, P., Travers, A., and Muskhelishvili, G. (2012). Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. Proc Natl Acad Sci U S A, 109(2):42–50.

[Soler-Bistué et al., 2015] Soler-Bistué, A., Mondotte, J. A., Bland, M. J., Val, M. E., Saleh, M. C., and Mazel, D. (2015). Genomic location of the major ribosomal protein gene locus determines Vibrio cholerae global growth and infectivity. PLoS Genet, 11(4):e1005156.

[Song et al., 2021] Song, L. F., Deng, Z. H., Gong, Z. Y., Li, L. L., and Li, B. Z. (2021). Oligonucleotide Synthesis for Whole-Genome Synthesis and Data Storage: Challenges and Opportunities. Front Bioeng Biotechnol, 9:689797.

[Sousa et al., 1997] Sousa, C., de Lorenzo, V., and Cebolla, A. (1997). Modulation of gene expression through chromosomal positioning in Escherichia coli. Microbiology (Reading), 143 (Pt 6):2071–2078.

[STACEY and SIMSON, 1965] STACEY, K. A. and SIMSON, E. (1965). IMPROVED METHOD FOR THE ISOLATION OF THYMINE-REQUIRING MUTANTS OF ESCHERICHIA COLI. J Bacteriol, 90:554–555.

[Stukenberg et al., 2021] Stukenberg, D., Hensel, T., Hoff, J., Daniel, B., Inckemann, R., Tedeschi, J. N., Nousch, F., and Fritz, G. (2021). The marburg collection: A golden gate DNA assembly framework for synthetic biology applications in vibrio natriegens. 10(8):1904–1919.

[Sueoka and Yoshikawa, 1965] Sueoka, N. and Yoshikawa, H. (1965). The chromosome of Bacillus subtilis. I. Theory of marker frequency analysis. Genetics, 52(4):747–757.

[Taniguchi et al., 1979] Taniguchi, T., O'Neill, M., and de Crombrugghe, B. (1979). Interaction site of Escherichia coli cyclic AMP receptor protein on DNA of galactose operon promoters. Proc Natl Acad Sci U S A, 76(10):5090–5094.

[Tas et al., 2015] Tas, H., Nguyen, C. T., Patel, R., Kim, N. H., and Kuhlman, T. E. (2015). An Integrated System for Precise Genome Modification in Escherichia coli. PLoS One, 10(9):e0136963.

[Tollerson and Ibba, 2020] Tollerson, R. and Ibba, M. (2020). Translational regulation of environmental adaptation in bacteria. J Biol Chem, 295(30):10434–10445.

[Touchon et al., 2014] Touchon, M., Bobay, L. M., and Rocha, E. P. (2014). The chromosomal accommodation and domestication of mobile genetic elements. Curr Opin Microbiol, 22:22–29.

[Trussart et al., 2017] Trussart, M., Yus, E., Martinez, S., Baù, D., Tahara, Y. O., Pengo, T., Widjaja, M., Kretschmer, S., Swoger, J., Djordjevic, S., Turnbull, L., Whitchurch, C., Miyata, M., Marti-Renom, M. A., Lluch-Senar, M., and Serrano, L. (2017). Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. Nat Commun, 8:14665.

[Turan et al., 2010] Turan, S., Kuehle, J., Schambach, A., Baum, C., and Bode, J. (2010). Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. J Mol Biol, 402(1):52–69.

[Valens et al., 2004] Valens, M., Penaud, S., Rossignol, M., Cornet, F., and Boccard, F. (2004). Macrodomain organization of the Escherichia coli chromosome. EMBO J, 23(21):4330–4341.

[van Kessel and Hatfull, 2007] van Kessel, J. C. and Hatfull, G. F. (2007). Recombineering in Mycobacterium tuberculosis. Nat Methods, 4(2):147–152.

[Verma et al., 2019] Verma, S. C., Qian, Z., and Adhya, S. L. (2019). Architecture of the Escherichia coli nucleoid. PLoS Genet, 15(12):e1008456.

[Vos, 2009] Vos, M. (2009). Why do bacteria engage in homologous recombination? Trends Microbiol, 17(6):226–232.

[Wang et al., 2009] Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature, 460(7257):894–898.

[Wang et al., 2016] Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., and Chin, J. W. (2016). Defining synonymous codon compression schemes by genome recoding. Nature, 539(7627):59–64.

[Weber et al., 2011] Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011). A modular cloning system for standardized assembly of multigene constructs. PLoS One, 6(2):e16765.

[Wilson, 1991] Wilson, G. G. (1991). Organization of restriction-modification systems. Nucleic Acids Res, 19(10):2539–2566.

[Wright et al., 2007] Wright, M. A., Kharchenko, P., Church, G. M., and Segrè, D. (2007). Chromosomal periodicity of evolutionarily conserved gene pairs. Proc Natl Acad Sci U S A, 104(25):10559–10564.

[Xiang et al., 2021] Xiang, X., Corsi, G. I., Anthon, C., Qu, K., Pan, X., Liang, X., Han, P., Dong, Z., Liu, L., Zhong, J., Ma, T., Wang, J., Zhang, X., Jiang, H., Xu, F., Liu, X., Xu, X., Wang, J., Yang, H., Bolund, L., Church, G. M., Lin, L., Gorodkin, J., and Luo, Y. (2021). Enhancing CRISPR-Cas9 gRNA efficiency prediction by data integration and deep learning. Nat Commun, 12(1):3238.

[Xu et al., 2017] Xu, J., Lian, W., Jia, Y., Li, L., and Huang, Z. (2017). Optimized guide RNA structure for genome editing via Cas9. Oncotarget, 8(55):94166–94171.

[Yan and Fong, 2017] Yan, Q. and Fong, S. S. (2017). Study of in vitro transcriptional binding effects and noise using constitutive promoters combined with UP element sequences in escherichia coli. J. Biol. Eng., 11(1).

[Yeung et al., 2017] Yeung, E., Dy, A. J., Martin, K. B., Ng, A. H., Del Vecchio, D., Beck, J. L., Collins, J. J., and Murray, R. M. (2017). Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. Cell Syst, 5(1):11–24.

[Yokobayashi et al., 2002] Yokobayashi, Y., Weiss, R., and Arnold, F. H. (2002). Directed evolution of a genetic circuit. Proc Natl Acad Sci U S A, 99(26):16587–16591.

[Yoneji et al., 2021] Yoneji, T., Fujita, H., Mukai, T., and Su'etsugu, M. (2021). Grand scale genome manipulation via chromosome swapping in Escherichia coli programmed by three one megabase chromosomes. Nucleic Acids Res, 49(15):8407–8418.

[Zetsche et al., 2015] Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., van der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell, 163(3):759–771.

[Zhang et al., 1998] Zhang, Y., Buchholz, F., Muyrers, J. P., and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat Genet, 20(2):123–128.

[Zhao et al., 2020] Zhao, J., Fang, H., and Zhang, D. (2020). Expanding application of CRISPR-Cas9 system in microorganisms. Synth Syst Biotechnol, 5(4):269–276.

Danksagung

Ich möchte mich an dieser Stelle bei den Menschen bedanken, die mich auf meinem bisherigen Weg begleitet und die mir diese Arbeit ermöglicht haben.

Zuallererst möchte ich mich bei meinen Eltern Ursula und Albert Teufel bedanken, die mich immer in all meinen Entscheidungen unterstützt haben und ohne die ich nicht dort wäre, wo ich jetzt bin. Auch meinen Freunden gilt ein besonderer Dank, da sie immer ein offenes Ohr für mich haben.

Bei meinem Doktorvater Patrick Sobetzko, bedanke ich mich ganz herzlich für die Möglichkeit zu promovieren und auch für die große Unterstützung während meiner Zeit als Doktorand. Ich werde unsere täglichen Gespräche und Diskussionen vermissen!

Prof. Dr. Torsten Waldmignhaus und Prof. Dr. Peter Graumann möchte ich dafür danken, dass Sie mich in Ihre Labore aufgenommen haben. Ich bedanke mich ebenfalls bei allen Mitgliedern der Arbeitsgruppen Waldminghaus und Graumann, da sie mir stets das Gefühl gegeben haben, Teil ihrer Arbeitsgruppe zu sein. Zusätzlich möchte ich mich bei den Arbeitsgruppen Becker und Fritz für den regen wissenschaftlichen Austausch bedanken. Ein besonderer Dank gilt Sabrina Steidl, Mona Bastian, Dr. Jan-Wolfhard Kellmann, Bettina Happel und Karin Sievers für einen reibungslosen Labor- und Arbeitsablauf. Bei Prof. Dr. Anke Becker und Prof. Dr. Lennart Randau bedanke ich mich für ihre Prüfertätigkeit.

Ein großes Dankeschön gilt auch meinen Kollegen Carlo A. Klein, Franziska Kemter und David Velázquez. Danke für die humorvolle und freundschaftliche Arbeitsatmosphäre.

Ein Dankeschön auch an meinen ehemaligen Studenten Maurice Mager. Die Zusammenarbeit mit dir hat mir großen Spaß gemacht.

Erklärung der selbständigen Erarbeitung der Dissertation

Hiermit erkläre ich, dass ich die vorliegende Dissertation

„Chromosomal Architecture and its Influence on Gene Expression in Native and Engineered Bacteria"

selbstständig und ohne unerlaubte Hilfsmittel angefertigt habe. Ich habe mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient. Die Dissertation wurde in der jetzigen oder in einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinem sonstigen Prüfungszweck gedient.

Marburg, den 13.09.22

_____

Marc Teufel