How transformation affects bacterial genome dynamics and opens up alternative evolutionary trajectories

INAUGURAL - DISSERTATION
ZUR ERLANGUNG DES DOKTORGRADES (DR. RER. NAT.)
DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER UNIVERSITÄT ZU KÖLN

VORGELEGT VON

MONA FÖRSTER

AUS WIPPERFÜRTH

First examiner
Prof. Dr. Berenike MAIER

Second examiner
Prof. Dr. Isabel GORDO

Disputation:
16TH JULY 2024

Abstract

For non-sexually reproducing organisms such as bacteria, natural transformation, a common mode of horizontal gene transfer (HGT), plays an important role in increasing genetic diversity, spreading beneficial mutations and curing deleterious ones. However, little is known about the genome-wide barriers, the effects of transformation on fitness and the role of the genetic background of the strains.

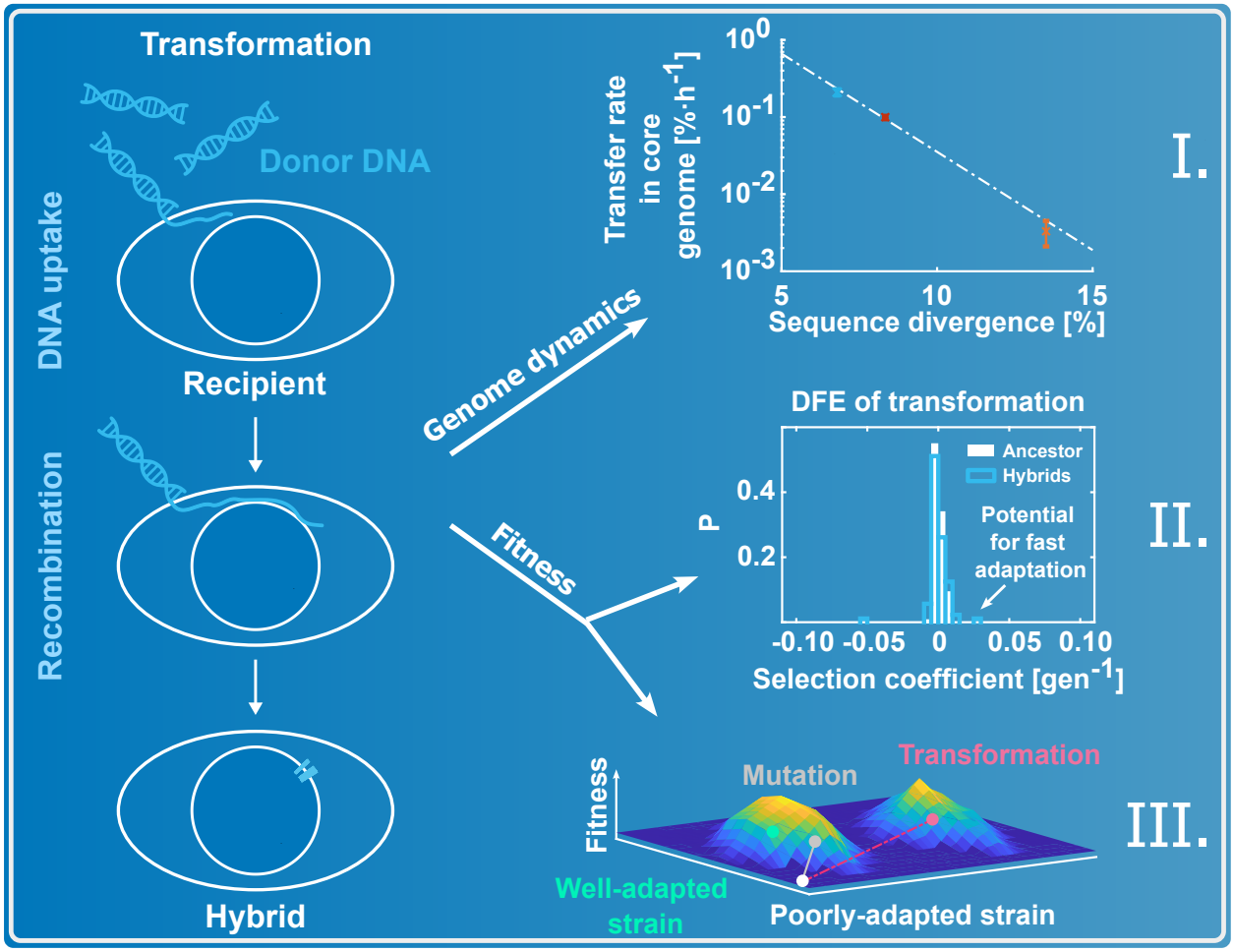
In the first project, we elucidate the genome-wide barriers of transformation and their dependence on the genetic distance between donor and recipient. We conduct replacement accumulation experiments with *Bacillus subtilis* as recipient and different *Bacillus* species as well as a *Geobacillus* species as donors. For 20 cycles, we allow the recipient to transform for two hours in presence of donor DNA, ensure minimal selection by plating the hybrids directly after transformation, and apply a single cell bottleneck by selecting a monoclonal hybrid for the start of the next cycle. Following cycle 10 and 20, the DNA of the hybrids is sent for whole genome sequencing and the fitness is measured by competition experiments. By pooling all hybrids, we find 96 % of core genes affected by replacements when supplying *B. subtilis* with *Bacillus spizizenii* DNA, indicating an almost unrestricted exchange between closely related species. In the less-related species *Bacillus atrophaeus*, we find replacement primarily in regions with high sequence identity, with an overrepresentation of ribosomal genes. A comparison of all donor-recipient pairs reveals an exponential relationship between core genome identity and transfer rate. Core genome identity also affects the length of replaced segments. At a low rate, the replacements entail insertions and deletions of accessory regions. We conclude that transformation between closely related species is strongly dependent on sequence identity and is dominated by replacements.

In the second project, we characterize the fitness effects of transformation. To this end, we create a hybrid library with *B. subtilis* as recipient and *B. vallismortis* as donor. The distribution of fitness effects (DFE) of the hybrid library is measured via competition experiments under different conditions. These include experiments conducted at varying temperature, with and without lag phase, and in more or less complex media. Two additional libraries, one with random replacement of whole genes and one with *B. spizizenii* as donor are generated, measured under one condition and compared to the first library. In contrast to the DFE previously reported for mutations, we do not observe a shift towards negative fitness values in most conditions. Instead, we find a predominantly neutral DFE with some beneficial and deleterious outliers in all but one characterized environments. Some of these outliers show antagonistic pleiotropic effects in the different conditions, supporting the idea that a benefit is derived through a shared gene pool between closely

related species. Furthermore, we verify the predictive value of the DFEs by performing an evolution experiment in two of the measured conditions. We conclude that DFEs are a suitable tool for predicting evolution and that transformation has the potential to accelerate adaptation.

In the third project, we investigate the role of preceding adaptation on the fitness effects of transformation. For that purpose, we create two hybrid populations with different histories of adaptation, one whose recipient is well adapted to growth in liquid environments and another one whose recipient is well adapted to growth in structured environments. Using these hybrid populations, we address two questions. First, we assess whether transformation with genomic DNA of *B. vallismortis* increases the fitness of *B. subtilis* that were already well adapted to a specific growth condition. By means of laboratory evolution, we find additional benefit of transformation during exponential growth in liquid. Second, we investigate whether transformation speeds up adaption to new growth environments. We find that during adaptation, transformation leads to an increase in fitness in both poorly adapted populations. Interestingly, in the transformed cells different genes are affected by replacements or mutation than in the non-transformed control populations. We conclude that transformation opens up new trajectories of adaptive evolution.

To summarize, we find extensive exchange between closely related species, mostly neutrality for fitness of gene replacements and the opportunity for new pathways of adaptive evolution. This indicates that HGT between closely related species is frequent and has a major impact on how bacteria adapt to changing environments.


Transformation
Donor DNA
DNA uptake
Recipient
Recombination
Hybrid
Genome dynamics
Fitness
Transfer rate in core genome [%·h-1]
10^0
10^-1
10^-2
10^-3
5
10
15
Sequence divergence [%]
I.
DFE of transformation
P
0.4
0.2
Ancestor
Hybrids
Potential for fast adaptation
-0.10 -0.05 0 0.05 0.10
Selection coefficient [gen-1]
II.
Fitness
Mutation
Transformation
Well-adapted strain
Poorly-adapted strain
III.

Zusammenfassung

Bakterien pflanzen sich nicht geschlechtlich fort. Daher spielt die natürliche Transformation, eine gängige Form des horizontalen Gentransfers (HGT), eine wichtige Rolle bei der Steigerung der genetischen Vielfalt, der Verbreitung nützlicher und der Korrektur schädlicher Mutationen. Über die genomweiten Barrieren, die Auswirkungen der Transformation auf die Fitness und die Rolle des genetischen Hintergrunds der Stämme ist jedoch nur wenig bekannt.

Im ersten Projekt werden die genomweiten Barrieren der Transformation und ihre Abhängigkeit von der genetischen Distanz zwischen Donor und Rezipienten untersucht. Dazu werden Gentransfer-Akkumulationsexperimente mit *Bacillus subtilis* als Rezipienten und verschiedenen *Bacillus*-Spezies sowie einer *Geobacillus*-Spezies als Donor durchgeführt. Für 20 Zyklen wird der Empfänger zwei Stunden lang mit der Donor-DNA transformiert. Für minimale Selektion wird gesorgt, indem die Hybride direkt nach der Transformation ausplattiert werden. Ein Einzelzellen-Engpass wird angewendet, indem ein monoklonaler Hybrid für den Beginn des nächsten Zyklus ausgewählt wird. Nach 10 bzw. 20 Zyklen wird die DNA des gesamten Genoms sequenziert und die Fitness der Hybride durch Kompetitionsexperimente gemessen. Unter Berücksichtigung aller Ersetzungen in 42 Hybriden wird gezeigt, dass 96 % der Kerngene von Ersetzungen betroffen sind, wenn *B. subtilis* mit *Bacillus spizizenii* DNA transformiert wird. Das deutet auf einen nahezu uneingeschränkten Austausch zwischen eng verwandten Arten hin. Bei der weniger verwandten Spezies *Bacillus atrophaeus* werden hauptsächlich Regionen mit hoher Sequenzidentität ausgetauscht, wobei Regionen mit ribosomalen Genen überrepräsentiert sind. Ein Vergleich aller Donor-Rezipienten-Paare zeigt eine exponentielle Beziehung zwischen Kerngenomidentität und Transferrate. Die Kerngenomidentität wirkt sich auch auf die Länge der ersetzten Segmente aus. Die Ersetzungen führen manchmal zu Insertionen und Deletionen von akzessorischen Regionen. Wir kommen zu dem Schluss, dass die Transformation zwischen eng verwandten Arten stark von der Sequenzidentität abhängt und von Ersetzungen dominiert wird.

Im zweiten Projekt werden die Fitnesseffekte der Transformation charakterisiert. Zu diesem Zweck wird eine Hybridbibliothek mit *B. subtilis* als Rezipienten und *B. vallismortis* als Donor erstellt. Die Verteilung der Fitnesseffekte (DFE, engl. *distribution of fitness effects*) der Hybridbibliothek wird durch Kompetitionsexperimente unter verschiedenen Bedingungen gemessen. Dazu gehören Experimente, die bei unterschiedlichen Temperaturen, mit und ohne Durchlaufen der Latenzphase und in mehr oder weniger komplexen Medien durchgeführt werden. Zwei zusätzliche Bibliotheken, eine mit zufälligem Ersatz ganzer Gene und eine mit *B. spizizenii* als Donor, werden erzeugt, unter einer Bedin-

gung gemessen und mit der ersten Bibliothek verglichen. Im Gegensatz zu bereits für Mutationen bekannten DFEs wird unter den meisten Bedingungen keine Verschiebung zu negativen Fitnesswerten hin beobachtet. Stattdessen wird in allen charakterisierten Umgebungen außer einer eine überwiegend neutrale DFE mit einigen vorteilhaften und nachteiligen Ausreißern gefunden. Einige dieser Ausreißer zeigen antagonistische pleiotropische Effekte unter verschiedenen Wachstumsbedingungen. Das unterstützt die Idee, dass ein Vorteil durch einen gemeinsamen Genpool zwischen eng verwandten Arten entsteht. Darüber hinaus wird die Vorhersagekraft der DFEs verifiziert, indem ein Evolutionsexperiment unter zwei der gemessenen Bedingungen durchgeführt wird. Es wird geschlussfolgert, dass zum einen DFEs ein geeignetes Instrument für die Vorhersage der Evolution sind und zum anderen die Transformation das Potenzial hat, die Anpassung an neue Umgebungen zu beschleunigen.

Im dritten Projekt wird die Rolle der vorangegangenen Anpassung auf die Fitnesseffekte der Transformation untersucht. Zu diesem Zweck werden zwei Hybridpopulationen mit unterschiedlichen Anpassungsgeschichten geschaffen, eine, deren Rezipient gut an das Wachstum in flüssigen Umgebungen angepasst ist, und eine andere, deren Rezipient gut an das Wachstum in strukturierten Umgebungen angepasst ist. Mit Hilfe dieser Hybridpopulationen wird zwei Fragen nachgegangen. Erstens wird untersucht, ob die Transformation mit genomischer DNA von *B. vallismortis* die Fitness von *B. subtilis* erhöht, die bereits gut an eine bestimmte Wachstumsbedingung angepasst sind. Mit Hilfe der Laborevolution wird ein zusätzlicher Nutzen der Transformation während des exponentiellen Wachstums in Flüssigkeit gefunden. Zweitens wird untersucht, ob die Transformation die Anpassung an neue Wachstumsbedingungen beschleunigt. Wir stellen fest, dass die Transformation während der Anpassung zu einem Anstieg der Fitness in beiden schlecht angepassten Populationen führt. Interessanterweise sind in den transformierten Zellen andere Gene von Ersetzungen oder Mutationen betroffen als in den nicht transformierten Kontrollpopulationen. Daraus wird geschlossen, dass die Transformation neue Wege der adaptiven Evolution eröffnet.

In dieser Arbeit wird gezeigt, dass ein umfangreicher Austausch zwischen eng verwandten Arten, eine weitgehende Neutralität für die Fitness von Genersetzungen und die Möglichkeit für neue Wege der adaptiven Evolution festgestellt werden. Dies deutet darauf hin, dass HGT zwischen eng verwandten Arten häufig vorkommt und einen großen Einfluss darauf hat, wie sich Bakterien an eine sich verändernde Umwelt anpassen.

Contents

1. Introduction

In nature, bacteria live in communities with other organisms [1]. In these communities, bacteria rely not only on spontaneous mutations when they need to adapt rapidly to changing environments. As explained in the following section, a large number of species is able to transfer genes horizontally between organisms. This leads to an increase in genotype variability in non-sexually reproducing organisms such as bacteria. Moreover, horizontal gene transfer (HGT) facilitates beneficial mutations or accessory genes present in a subset of cells to spread throughout the community and accelerate adaptive evolution.

1.1. Horizontal gene transfer

Horizontal gene transfer is the passing on of genetic material not from parent to child but between organisms. In the last decades, the increase in multi-resistant strains of pathogenic species has created a threat for the global health system [2–5]. HGT mediates the exchange of resistance genes between various species leading to multi-resistance. Irrespective of this, a high number of HGT events is found in phylogenetic studies even across species boundaries which lead to the suggestion to further develop Darwin's idea of a tree of life into a more interconnected web of life [6–9]. HGT shaped evolution significantly in the past. In order to explain and predict evolution it is therefore necessary to understand the underlying mechanisms, costs and benefits better.

Different mechanisms are known to enable horizontal gene transfer in bacteria (Fig. 1.1, [8, 10]): Transduction is the transfer of DNA via phages [11]. In conjugation, plasmids are transported from one cell to another via direct cell-to-cell contact through a "conjugative" pilus [12, 13]. Additionally, other so-called "non-canonical" ways are known: Some bacteria can build nanotubes through which they transfer molecules from the cytoplasm and non-conjugative plasmids [14, 15]. There are phage-like gene transfer agents (GTAs) [16] and membrane vesicles [17, 18] transporting genetic material from one cell to the other. Another mechanism that is common in many bacterial species is transformation (Fig. 1.1 **C**).

Natural transformation involves the uptake of extracellular DNA (eDNA) as single-stranded DNA (ssDNA) into the cell and the subsequent integration in the genome via homologous recombination. It was first discovered by Frederick Griffith in 1928 [19]. Nowadays, more than 80 species are known to be competent for transformation at least transiently [20–22]. Once a cell is in the so-called competent state, it is able to take up eDNA as ssDNA into the cytoplasm [23, 24]. The responsible uptake machineries differ for gram-positive and gram-negative cells and will be described in detail for the gram-positive bacterium *B. sub-*

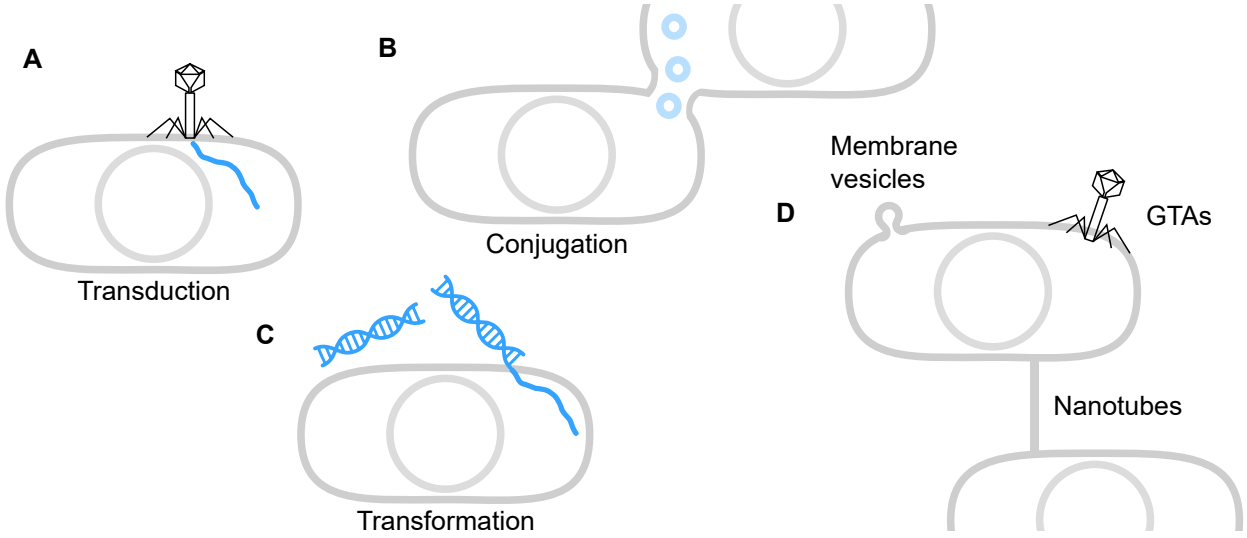


Figure 1.1.: The different modes of horizontal gene transfer. Several mechanisms are known through which bacteria transfer genes horizontally. **A** In transduction, phages transport genetic material from their former host cell into a new cell. **B** Conjugation describes the transfer of conjugative plasmids via direct cell-to-cell contact. **C** In transformation, cells take up external DNA from the surrounding. **D** DNA can also be transfered via ”non-canonial” mechanisms such as membrane vesicles, GTAs and nanotubes.

tilis in section 1.2.1. After the uptake, various proteins including RecA bind to the ssDNA and perform a homology search along the chromosome of the cell. If a homologous DNA segment is found, homologous recombination is initiated. During recombination, a stable heteroduplex is formed between the chromosome and the incoming single strand only minutes after the addition of eDNA [25, 26]. The heteroduplex is resolved during replication [27]. Genes are already expressed shortly after replication even before separation of cells [27].

There are three non-mutually exclusive hypotheses, why transformation has evolved and is conserved in so many species [23, 28]. For all three concepts evidence is found empirically. The first idea is that the cells degrade the DNA and use the nutrients as food [29, 30]. This does not explain why only the single strand is important and thus, half of the nutrients are wasted [23]. Second, the DNA templates are used to repair mismatches or damaged parts of the genome [31–33]. Michod et al. showed that survival of cells increases after UV radiation when cells are supplied with DNA [31]. However, competence is not induced by DNA damage [34]. In the third hypothesis, cells use the eDNA as a source for genetic diversity [35, 36]. On the one hand, this means that transformation can protect against diversity reductions due to bottleneck effects or solve clonal interference by spreading beneficial mutations [23, 37]. On the other hand, in a new niche, genes can be transferred from other organisms that are already adapted to that special growth environment.

1.1.1. Limits and barriers of transformation

Bacteria evolved various mechanisms to protect themselves against the uncontrolled uptake of DNA. In some bacterial species, like *Neisseria gonorrhoeae* and *Haemophilus influenzae*, only DNA containing a specific DNA uptake sequence (DUS) is imported [38–40]. The genus-specific DUS ensures the recognition of homotypic DNA. In *N. gonorrhoeae*, the minor pilin ComP is responsible for the recognition of the DUS in eDNA [41]. DNA uptake in *B. subtilis* is not restricted by a DUS. But unlike *Neisseria*, *B. subtilis* is not permanently competent. Only during late exponential growth phase a subset of cells (5 - 20 %) enters competent state transiently [25, 42, 43]. That way, the bulk remains unaffected by transformation, whereas a small fraction of the population may benefit from the genetic material in the surrounding. Restriction modification (RM) systems depict another mechanism of self-defense against foreign DNA [44, 45]. After uptake, DNA that is unmethylated at the recognition sites of the RM systems is cleaved. RM systems have been found in about 90 % of all sequenced bacteria [46]. Finally, the probability of homologous recombination depends strongly on the sequence divergence between the transforming DNA and the homologous DNA on the chromosome. In the level of single genes, it has been shown that the integration probability decreases exponentially with increasing sequence divergence [47, 48]. Additionally, Carrasco et al. suggest a divergence threshold, above which no more recombination happens [48].

1.1.2. Benefits and costs of transformation

The following section will list a number of important costs and benefits of bacterial transformation. The integration of random donor segments can lead to various costs for the transformant [49, 50]. One obvious consequence is the increased need for resources by transcription and translation when the number of genes increases [51]. Second, replacing genes or operons partially can potentially interrupt functional networks [52]. Changes in protein folding can lead to a loss of function or a reduced functionality [53]. Moreover, a change in expression level of a protein, up- or downregulation, or a codon usage mismatch between donor and recipient can cause severe costs [54–56].

Nevertheless, the fact that competence is preserved in many species suggests that it provides advantages [20, 21]. In the absence of foreign DNA, naturally competent bacteria have been shown to have a benefit by curing deleterious mutations and combining beneficial ones that would otherwise have undergone clonal interference [37, 57–59]. Moreover, transformation can be helpful in adapting to a new niche when the bacterium needs to adjust to a novel carbon source [60], higher salinity [61] or ground water contaminated with heavy metals [62].

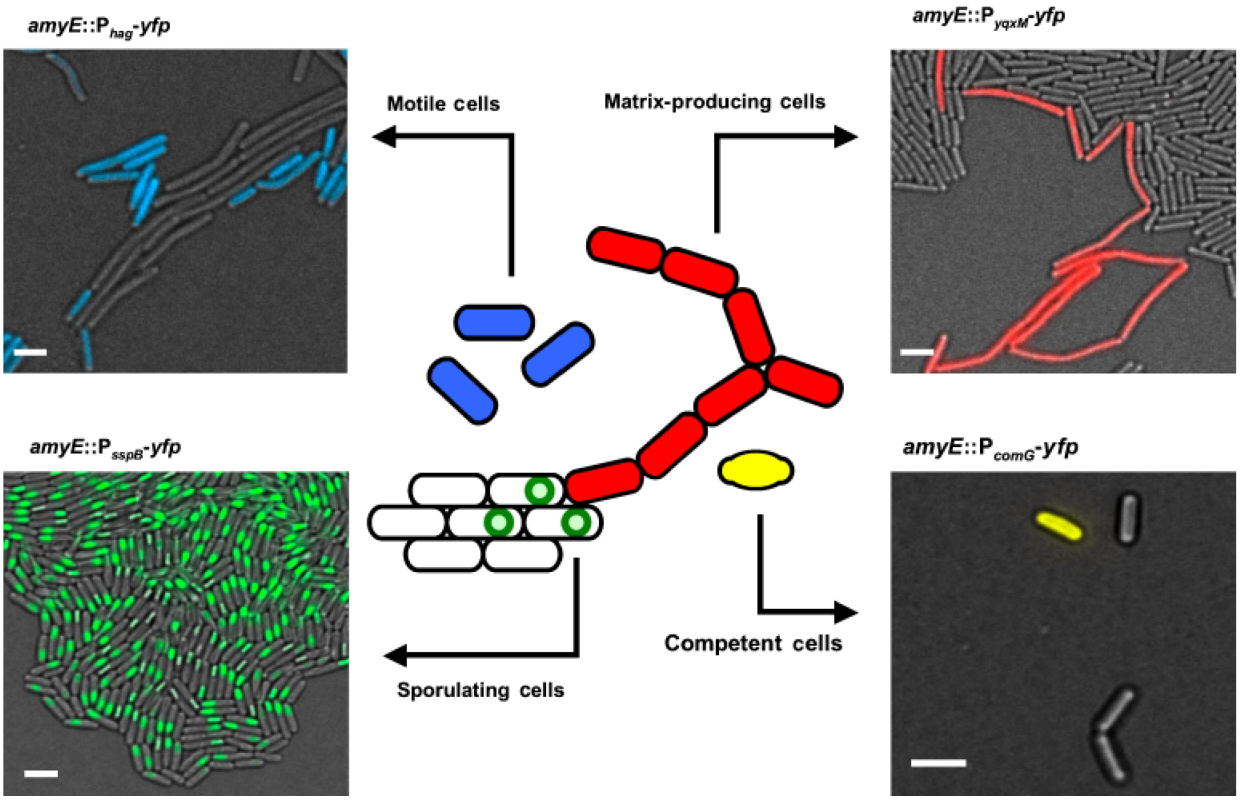


Figure 1.2.: Phenotypic states of *B. subtilis*. Although being genetically identical, *B. subtilis* cells differentiate into diverse cell types. Motile cells (blue) produce flagella, sporulating cells (green) are metabolically inactive and highly resistant to extreme environments. Competent cells (yellow) are able to take up free DNA from their surrounding and a subset of cells in a biofilm produces extracellular matrix (red). Reprinted with permission from [64].

1.2. The gram-positive model organism *Bacillus subtilis*

B. subtilis is a gram-positive soil bacterium that is found on plant roots but also in the gastrointestinal tract of mammals and in marine habitats [63]. Populations of *B. subtilis* cells, even when genetically identical, can differentiate into diverse phenotypic states (Fig. 1.2, [64–66]): Motile, flagellated and filament-forming, sessile, competent, spore-forming and extracellular matrix-producing.

In this study, we investigate the impact of competence on adaptive evolution. In our evolution experiments, the phenotypes that are found within biofilms (matrix producing and sporulating) and in a liquid environment (motile and sessile) are of particular interest.

1.2.1. Transformation in *B. subtilis*

In *B. subtilis*, about 5 - 20 % of the population enter compentent state in late exponential growth phase [25, 42, 43, 67]. Recently it was shown that this frequency depends on cell envelope stress that occurs if a non-kin strain is close by [68]. In the competent state, a

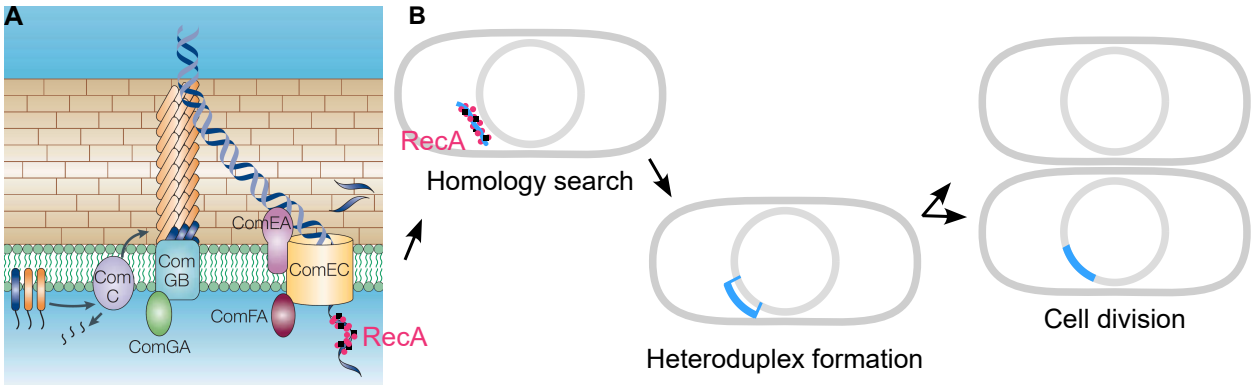


Figure 1.3.: DNA uptake and homologous recombination. A Current model of the DNA uptake machinery of *B. subtilis*. The pseudopilus is involved in the recruitment of eDNA to the cell envelope, where the protein ComEA binds the dsDNA. Subsequently, one strand is imported through a pore formed by ComEC, while the other strand is degraded outside the cell. **B** When the single strand enters the cytoplasm, RecA and other proteins bind and perform a homology search. If a homologous segment is detected, homologous recombination is initiated. The ssDNA and the chromosome build a stable heteroduplex [27]. This heteroduplex is solved during replication. Image **A** adapted from [28] with permission.

pseudopilus is involved in recruiting double-stranded DNA (dsDNA) to the cell envelope, where ComEA binds the dsDNA without sequence specifity (Fig. 1.3, [24, 69]). Each *B. subtilis* cell has 20 - 50 DNA binding sites [70]. One strand of the bound DNA is transported into the cytoplasm in a linear fashion with about 80 nt/s, through a pore formed by ComEC [22, 23, 71, 72]. The other strand is degraded outside the cell. When the ssDNA enters the cytoplasm, RecA and other proteins belonging to the recombination machinery bind to the imported single strand and perform a homology search. As soon as a homology is found homologous recombination is initiated as described in the previous section 1.1. In the competent state, *B. subtilis* is growth-arrest. Depending on the environmental conditions, growth-arrest lasts for 2 - 4 hours [25, 26, 73].

1.2.2. Exponential growth phase in liquid environment

Kearns et al. described that *B. subtilis* exists in two distinct phenotypes during mid-exponential growth phase [74]. They identified single flagellated, motile cells and filament forming, sessile cells. In the first case, the sigma factor of the flagellum-chemotaxis (fla-che) operon, *sigD*, was active, while in the second case *sigD* was inactive. Moreover, it has been shown that *sigD* mutants (*sigD* inactive) are deficient in autolysis, grow in long chains and have a higher fitness in a liquid environment than single motile cells [75].

1.2.3. *Bacillus* biofilms

In nature, bacteria often form surface-bound communities called biofilms [1]. In a biofilm, cells are held together by an extracellular matrix produced by a subset of the cells [76]. The extracellular matrix of *Bacillus* biofilms mainly consists of exopolysaccharides (EPS)

and proteins [77, 78]. Biofilm formation can increase protection against antimicrobial substances like antibiotics [79, 80]. Another subpopulation of cells in the biofilm forms spores that are metabolically inactive and can persist under extreme conditions [81].

The *B. subtilis* strain used in the following study as an ancestor builds flat and smooth colonies on agar plates as it has been described for matrix-deficient mutants [77].

1.3. Bacterial fitness

In an evolving population, the rate of mutations and gene transfer events, natural selection, and genetic drift determine population dynamics [82]. While genetic drift is a stochastic effect, natural selection is driven by the fittest clones in a population. Bacterial fitness is a measure for how well a strain can survive in a particular environment [83].

1.3.1. Experimental characterization of bacterial fitness

One way to measure fitness is to determine the growth rate (or Malthusian fitness) λ of the strains of interest by counting colony forming units or by tracking the optical density of the bacterial suspension [84, 85]. Cells with a constant growth rate λ grow exponentially with $N(t) = N_0 \cdot e^{\lambda t}$ where $N(t)$ is the number of cells at time t and N_0 the initial number of cells. When measuring growth rates, the relative fitness is usually expressed as

$$(\lambda_{\mathrm{mutant}}/\lambda_{\mathrm{ancestor}}) - 1. \tag{1.1}$$

When very small changes in fitness have to be characterized, head-to-head competition experiments are more accurate than direct comparison of growth rates between ancestor and evolved strain [86–88]. In a competition experiment, two strains grow within the same environment and compete for space and nutrients. In the simplest case, both competitors A, B grow exponentially during the whole of the experiment with

$$N_A(t) = N_{0,A} \cdot e^{\lambda_A t} \tag{1.2}$$

where N_A is the total number of cells of strain A at time t, λ_A the growth rate of strain A and $N_{0,A}$ the number of cells of strain A at the start. If the two strains do not interact, the selection coefficient $s_{A,B}$ is just the difference in growth rates of the two strains

$$\frac{N_A}{N_B} = \frac{N_{0,A}}{N_{0,B}} \cdot \frac{e^{\lambda_A t}}{e^{\lambda_B t}}$$

$$\Leftrightarrow \frac{\frac{N_A}{N_B}}{\frac{N_{0,A}}{N_{0,B}}} = \frac{e^{\lambda_A t}}{e^{\lambda_B t}}$$

$$\Rightarrow \frac{1}{t} \cdot ln(\frac{\frac{N_A}{N_B}}{\frac{N_{0,A}}{N_{0,B}}}) = \lambda_A - \lambda_B = s_{A,B} \tag{1.3}$$

and can be easily calculated by measuring the start and end fraction of the strain mixture. What distinguishes a competition experiment from growth rate determination via optical density is that the two strains share the same environment. Therefore, not only the difference in growth but also interactions between the cells like killing via toxins or other weapons, or support e.g. via cross-feeding come into play and influence the outcome of the experiment [1]. The selection coefficient $s_{A,B}$ then not only represents the difference in growth rate but also contains information about interactions of the competitors. The fractions of the competitiors at the start and end of a competition experiment can be determined via plating [86, 89] or by using a flow cytometer [88].

Other ways to determine relative fitness are via barcode sequencing [90–92], at the single cell level [93], via range expansion on agar plates [94, 95] or indirectly by Bayesian inference [96].

1.3.2. Fitness landscape

For the visualization of genotypic variants and their fitness, Wright established the metaphor ”fitness landscape” in 1932 [97]. The fitness landscape relates the genotype to its corresponding fitness [98, 99]. The space of genotypes is multidimensional, but is commonly represented as a two-dimensional surface (Fig. 1.4). A peak corresponds to a genotype with high fitness, a valley to one with low fitness. Evolution can be thought of as a walk towards higher fitness. The landscape is not static but changes with changing conditions [100]. Sign epistasis occurs when a mutation is detrimental on its own but becomes beneficial in presence of another mutation. Sign epistasis leads to rugged landscapes with multiple peaks (Fig. 1.4 **A**, [101, 102]). A fitness landscape without sign epistasis is called smooth (Fig. 1.4 **B**). Due to its complexity, the fitness landscape has rarely been studied empirically [103–105].

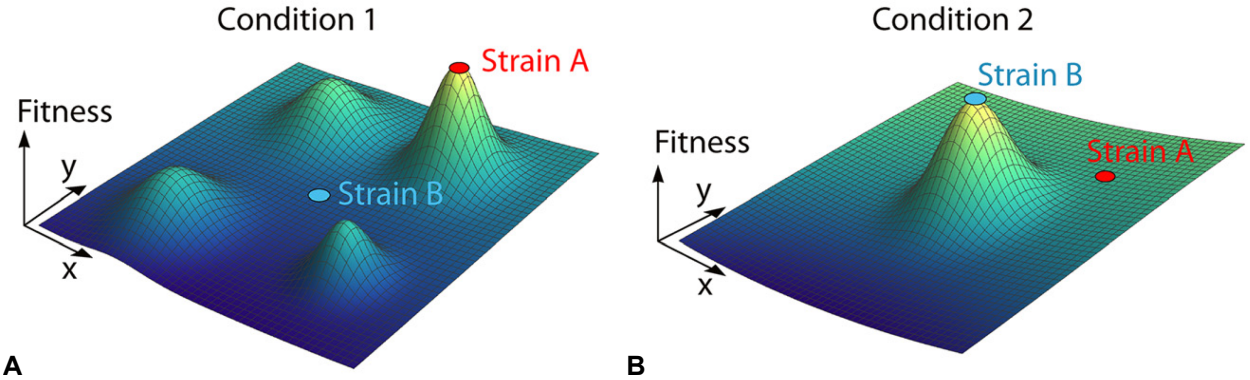


Figure 1.4.: Fitness landscapes. The xy-plane represents the genotype space, the z-axis the corresponding fitness. The fitness landscape changes with changing conditions. In condition 1, strain A is fitter than strain B. In condition 2, it is the other way around. Adapted from [100] with permission.

1.3.3. Distribution of fitness effects

The outcome of an evolutionary process depends on the genetic variations that are accessible and the selective pressure acting on them. The genetic variation can have different sources such as mutations, deletions or horizontal gene transfer [82]. The distribution of fitness effects (DFE) describes the fitness effects of mutational events on the fitness of an organism. Because of the sheer size of the possibilities, a small subset of the affected cells is analyzed in the hope that it represents the totality. The shape of a representative DFE can be used to predict evolution.

Multiple studies investigated the DFE of mutations [106–108]. For single mutations, Eyre-Walker et al. found a complex DFE containing mostly deleterious and neutral mutations with some rare beneficial mutations [106]. Similar results have been shown for several mutations accumulated within one strain in different environments [107]. However, when changing the mutation spectrum e.g. via deletion of a DNA repair gene or by applying stress, the DFE of single mutations was shifted in positive direction and thus, an increase in beneficial mutations was found [84, 109]. Since the beneficial mutations can drive adaptation, the DFE of beneficial mutations was investigated and found to be context-dependent with mostly small effect mutations in benign conditions, but with large effect mutations dominating in stressful environments [110, 111]. Rozen et al. showed theoretically and empirically that mainly large effect beneficial mutations drive adaptation [112].

Another source of genetic variation are deletions and insertions. Making use of the Keio collection of *Escherichia coli* knockout strains, it has been shown for non-essential gene deletions that the DFE is shifted to negative values if no stress is applied [113]. However, when different antibiotics were added, the width of the DFEs changed drastically. By inserting random DNA segments from different sources (bacteria and phages) into the *Salmonella* chromosome, Knöppel et al. found the fitness effect of horizontally transferred inserts to be mostly neutral [114]. Only 8 out of 98 inserts showed a disadvantage, suggesting that new genetic material usually has little effect on fitness and can therefore be retained in the chromosome.

DFEs are a widely used and appropriate tool for predicting evolution, but because the number of representatives measured for the DFEs is small, mutants with large effects may be overlooked. To scan a much larger pool, evolution experiments can be carried out.

1.4. Laboratory evolution

Evolution experiments are useful for studying fast-growing, small organisms such as bacteria [115]. Within days or weeks, hundreds of generations can be run through in diverse environments with various selective pressures and evolution can be tracked via fitness measurements or sequencing.

Serial transfer and continuous culture are the two most commonly used setups for experimental evolution (Fig. 1.5, [100, 116]). In serial transfer, cell cultures are regularly supplied with fresh medium, while old medium and some cells are discarded at the same time (Fig. 1.5 **A**). The total volume is kept constant, whereas the population size and nutrient availability are constantly changing. For this reason, the environment is continuously varying during serial dilution. The number of individuals remaining in the batch culture after dilution determines the genetic drift during the evolution experiment. The higher the cell number the less significant genetic drift is compared to the effects of natural selection. At a cell number of 1, e.g. at a single-cell bottleneck, genetic drift dominates the outcome of the evolution experiment. Single-cell bottlenecks are used to accumulate random mutations. Moreover, serial transfer setups can be used for studying evolution in more drastically changing environments by periodically applying stress or increasing stress at each dilution step. The most famous serial transfer experiment is the long-term evolution experiment that Richard Lenski started in 1988 and which is still running today [117]. In Lenski's experiment, 12 parallel *E. coli* populations are cultured in minimal medium with glucose as the only carbon source that *E. coli* can use. Every day, 1 % of each population is transferred to fresh medium. Recently, the experiment reached 75,000 generations [118]. For comparison, the species *Homo sapiens* is only about 7,500 generations old [119]. Over the course of the experiment, various evolutionary dynamics were

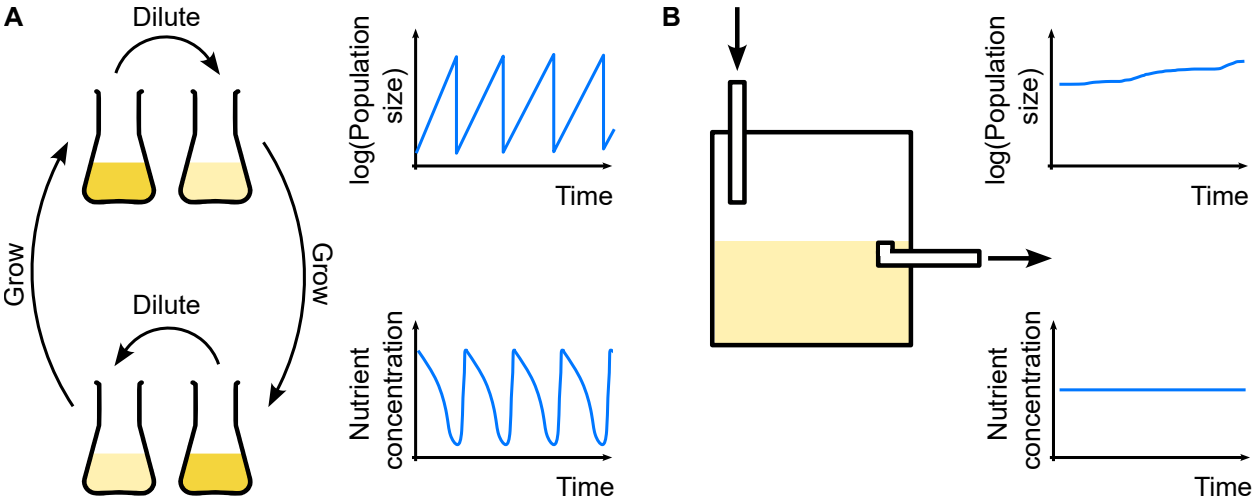


Figure 1.5.: Types of evolution experiments. A In serial transfer, cells are diluted regularly. Population size and nutrient availability are constantly changing. **B** In continuous culture, a permanent flow of nutrients and outflow of cells keeps the population size constant until adaptation leads to an increase.

observed across all years: an increase in the growth rate and cell size [120], a shortening of the lag phase [121] and clonal interference between competing mutants [122]. Moreover, one population acquired a new trait: the use of citrate as a carbon source [123].

In continuous culture, a constant flow ensures a constant nutrient availability (Fig. 1.5 **B**, [124]). The size of the populations remains the same until adaptation leads to an increase in size. By using a feedback loop to adjust the flow, the populations can be kept at a constant size permanently. This setup is called turbidostat. A continuous culture setup provides a very well-defined environment for the evolution experiment, but it is also quite far removed from natural conditions [125].

1.5. Population dynamics

In evolving populations, adaptation is driven by beneficial variants arising from (i) *de novo* mutations such as single nucleotide polymorphisms (SNPs), insertions, deletions and duplications or (ii) from horizontal gene transfer events like replacements with transforming DNA. Moreover, replacements can facilitate the insertion of novel genes or deletion of existing ones.

1.5.1. Adaptive evolution

In the course of evolution, favourable mutations or beneficial replacements with DNA of another species lead to gradually increasing fitness e.g. through mutations affecting gene expression levels or the efficiency of a metabolic pathway often referred to as genetic tinkering (Fig. 1.6 **A**). This results in a continuous optimization of the evolving population. The accumulation of beneficial variants often leads to diminishing-returns epistasis, i.e. the fitness increase of a beneficial event is lower in a strain that already has another beneficial event than on the background of the ancestor (Fig. 1.6 **C**, [102, 126, 127]).

Innovations are more dramatic changes (Fig. 1.6 **B**). A mutation or horizontal gene transfer event gives the strain a new function that is advantageous under the current conditions [123]. Innovations and optimization are often entangled. An innovation could only be possible due to the accumulated mutations, or mutations optimize an innovation [116].

In an adapting strain, advantageous mutations or transfer events can occur within the same time frame (Fig. 1.6 **D**). Before the one variant has swept through the entire population, the other occurs on the same background. As both are beneficial, they compete to take over the population. This is known as clonal interference [112, 128]. Clonal interference can transiently increase the genetic diversity of a population [122]. Beneficial mutations remain longer without fixation and can collect neutral mutations in the meantime. As discussed above (Sec. 1.1.2), transformation can help resolve clonal interference by merging the competing variants into one strain. Moreover, transformation can cure strains of detrimental mutations and lead to innovations through transfer of new genetic material.

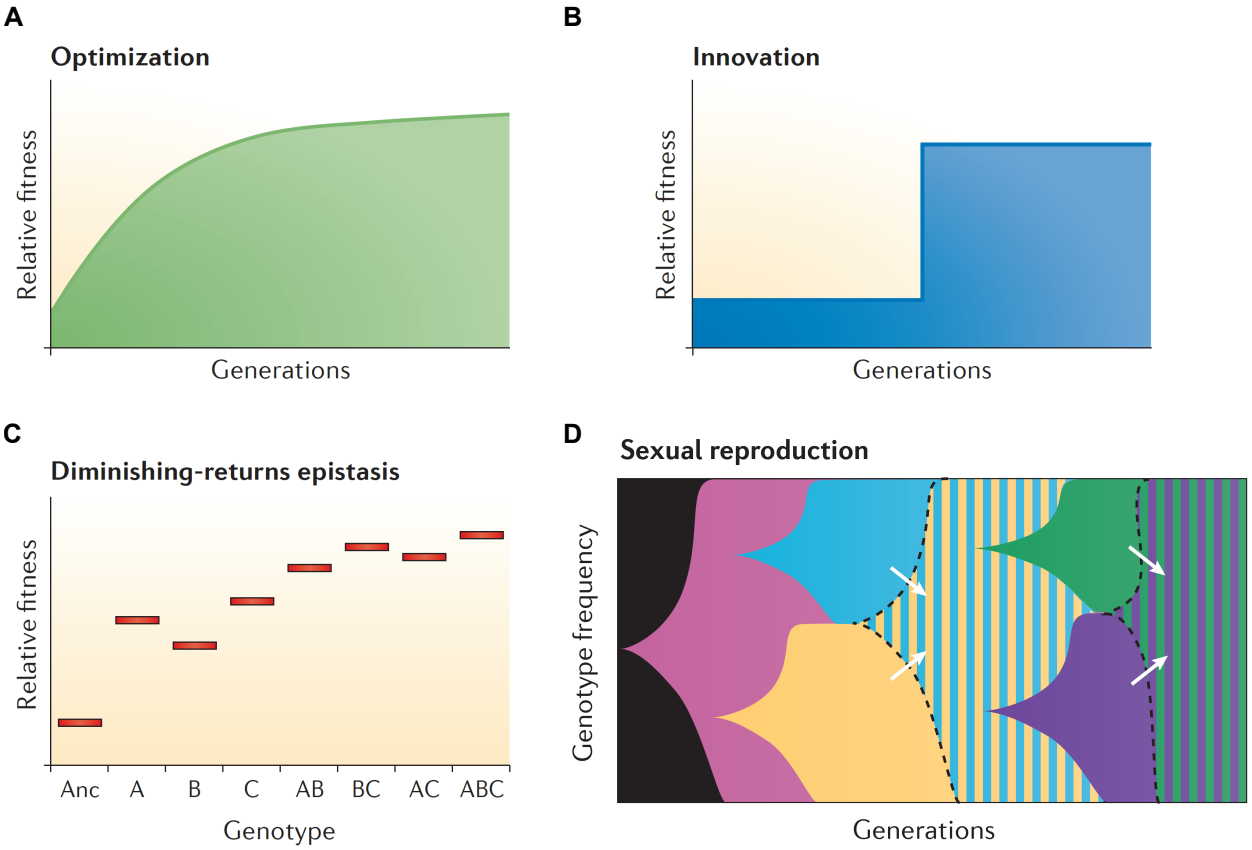


Figure 1.6.: Population dynamics. Optimization leads to a steady increase in relative fitness of the population (**A**), while innovation causes a sudden increase in relative fitness (**B**). During adaptation to a specific growth condition, diminishing-returns epistasis is observed frequently (**C**). The increase in fitness caused by mutation ”A”, ”B” or ”C” when arising in the ancestor is much higher than that of B when A was already present in the strain (”AB”), for example. At the start, the population is clonal (**D**, black). A favorable variant arises (magenta). On the magenta background, two competing beneficial variants appear simultaneously (blue and yellow). In a population in which the cells can exchange genetic information through transformation, both beneficial alleles can be recombined into one strain (blue/yellow mix). Images adapted from [116] with permission.

1.6. Scientific objectives

Since the discovery of transformation in 1928, much has been learned about the process itself and the obstacles that stand in its way. An exponential relationship between sequence divergence and the probability of replacement has been demonstrated for individual genes. Advances in whole genome sequencing in recent decades have opened up the possibility of a genome-wide view on the barriers and mechanisms of transformation. This also offers the opportunity to identify functional barriers.

The usefulness of DFEs for predicting the evolutionary dynamics of mutations has long been recognised. For transformation, another mechanism that leads to variants in bacterial genomes and thus to adaptive evolution, the DFE is unknown. A DFE has been measured for insertions, but since the main mode of natural transformation is replacement, this DFE lacks representative power.

In the first project, we investigate the barriers to transformation at the whole genome level. Therefore, we perform parallel replacement accumulation experiments with *B. subtilis* as recipient and other *Bacillus* and a *Geobacillus* species as donors. Using whole genome sequencing data, we identify orthologous replacements, deletions, duplications and insertions across the entire chromosome and relate their properties to the sequence divergence of the recipient-donor pairs. Additionally, we look for functional barriers and cold spots of transformation in closely related species.

In the second project, we measure the distribution of fitness effects of transformation under different conditions. To this end, we create hybrid libraries with *B. subtilis* as recipient and *Bacillus vallismortis* or *Bacillus spizizenii* as donors. For the fitness quantification, we design a high-throughput competition experiment. To test whether DFEs are suitable to predict the outcome of evolution, we perform evolution experiments with transformed populations under two different conditions.

In the third project, we investigate the effect of the history of preceding growth environments of the cells on evolution. We conduct evolution experiments with transformed *B. subtilis* populations in a structured environment and in a liquid, well-mixed environment. The *B. subtilis* ancestor of the transformed population is either well or poorly adapted to the environment in which the populations evolve. We develop various methods for fitness quantification and use the fitness as well as whole genome sequencing to characterize the dynamics of adaptive evolution in both environments and with both bacterial backgrounds.

In the following projects, we aim to better understand the mechanisms of integration and the fitness effects of transformation in order to predict how HGT affects adaptive evolution.

2. Experimental methods, strains and media

All experimental methods we use in this study are explained in detail in this section. For chapter 4, the replacement accumulation assay (Sec. 2.4) is most important. In chapter 5 the competition experiment (Sec. 2.6), the creation of hybrid libraries (Sec. 2.5) and finally the evolution experiment (Sec. 2.7) are used. In chapter 6, we perform evolution experiments (Sec. 2.7) and test the fitness of the evolved strains (Sec. 2.6).

2.1. Strains

Table 2.1 contains all strains used in this study. For all *Bacillus subtilis* strains, Bs166 (*amyE::PhscomK(spc)*, *comK::kan*, P_{comK}*gfp*) is the ancestor. To make the competence of the wild type *Bacillus subtilis* 168 controllable, Gabriele Schneider (GS) produced Bs166 for the publication by Jeffrey J. Power et al. on evolution in transformed *B. subtilis* cells under selection for fitness in stationary phase [52]. In Bs175, Melih Yüksel (MY)

Strain	Data base name	Relevant genotype	Source/Reference
Bacillus subtilis 168	Bs166	*amyE::spc-PhSPANK-comK comK::kan* P_{comK}*gfp*	GS
	Bs175	Bs166 *lacA::PrrnE-gfp* (erm)	MY
	Bs210	*sigD*-frame shift (Phe151fs)	evolution exp IR/MF
	Bs213	Bs210 *lacA::PrrnE-gfp* (erm)	MY
	Bs224	pre-adapted to agar plates [1]	evolution exp MF
	Bs226	Bs224 *lacA::PrrnE-gfp* (erm)	AL/MF
Bacillus spizizenii	Ob005	wild type	BGSC
Bacillus vallismortis	Ob018	wild type	DSMZ
Bacillus mojavensis	Ob003	wild type	DSMZ
Bacillus atrophaeus	Ob006	wild type	BGSC
Geobacillus thermoglucosidasius	Ob011	wild type	DSMZ

Table 2.1.: All strains used in this study, their characteristics and sources.

[1] Single nucleotide polymorphisms (SNPs) in *rapA* (Pro261Pro), *gudB* (Phe396Lys) and *epsC* (Glu328Val), frame shift mutation in *gudB* (Phe395fs)

added a gene to Bs166 that produces a green fluorescent protein so that the strain can be used as a reporter in competition experiments (*lacA::PrrnE-gfp* (erm)). Bs210 and Bs224 arose from Bs166 after pre-adaptation to a certain condition (Sec. 2.7). Bs210 is pre-adapted to exponential growth in liquid environment and Bs224 to growth on complex medium agar plates. Bs213 and Bs226 are their respective green-fluorescent reporter strains. All other *Bacillus* strains as well as the *Geobacillus* strain are wild type strains and were ordered from either the German Leibniz Institute in Braunschweig (DSMZ) or the Bacillus Genetic Stock Center (BGSC). The NCBI dictionaries of the strains used for genome analysis are specified in the Supplements (Tab. A.3).

2.2. Media

All products used are supplied from VWR, Merck, Carl Roth or Gibco unless otherwise stated. The cells are cultivated in lysogeny broth (LB), complex medium (CM), defined medium on glycerol basis (DM_{gly}) or defined medium on glucose basis (DM_{glu}). Cell handling is done in medium or phosphate buffered saline (PBS). For freezing at -80 °C, 10 % dimethyl sulfoxide (DMSO) is added to cell cultures.

If not stated differently, cells are grown at 37 °C. Liquid cultures are placed in a shaker at 250 rpm. For growth on plates, 1.5 % agar is added to the respective medium. After pouring, the plates are dried for approximately 30 minutes without a lid, then closed and turned over. Agar plates used in the agar plate handling system are not inverted.

2.2.1. Complex medium

As complex medium (CM) we use the ”competence medium” by Dubnau et al. [129]. For the complex medium, we prepare 10x Spizizen’s salts (SpS, Tab. 2.2) and 100x CM-master mix (Tab. 2.3) and sterilize them by autoclaving. By filtration, we sterilize 50% glucose and a 100x amino acid mix containing 5 mg/ml of methionine, leucine and histidine solved in water. For 500 ml of complex medium, we add 50 ml of 10x SpS and 5 ml of each, 100x CM master mix, 100x amino acid mix and 50 % glucose to 435 ml sterile MQ (Tab. 2.4).

2.2.2. Defined medium

In this study, we use two different chemically defined media, one based on glucose (DM_{glu}) and one with glycerol as the only carbon source (DM_{gly}). For DM_{glu}, 1x SpS (Tab. 2.2) is supplemented with 0.5 % glucose, 0.5 mg/ml sodium glutamate and 0.2 mg/ml $MgSO_4$. 1x amino acid mix, as used for the complex medium, is added. For DM_{gly}, SpS without trisodium citrate ($Na_3C_6H_5O_7 \cdot H_2O$) is supplemented with 0.2 mg/mg $MgSO_4$, 10 % minimum essential medium amino acid (MEM-AA) solution, 5 % MEM-non-essential amino acid (MEM-NEAA)-solution and 1x amino acid mix.

Table 2.2.: 10x Spizizen's salts

Agent	Amount [g]
KH_2PO_4	30
K_2HPO_4	70
$(NH_4)_2SO_4$	10
$Na_3C_6H_5O_7 \cdot H_2O$	5
MQ	fill to 500 ml

Table 2.3.: 100x CM master mix

Agent	Amount [g]
Casamino Acids	1
Yeast Extract	5
$MgCl_2 \cdot 6H_2O$	2.5
MQ	fill to 50 ml

Table 2.4.: 500 ml complex medium

Agent	Amount [ml]
10x SpS	50
100x CM master mix	5
100x amino acid mix	5
50 % glucose	5
MQ	435

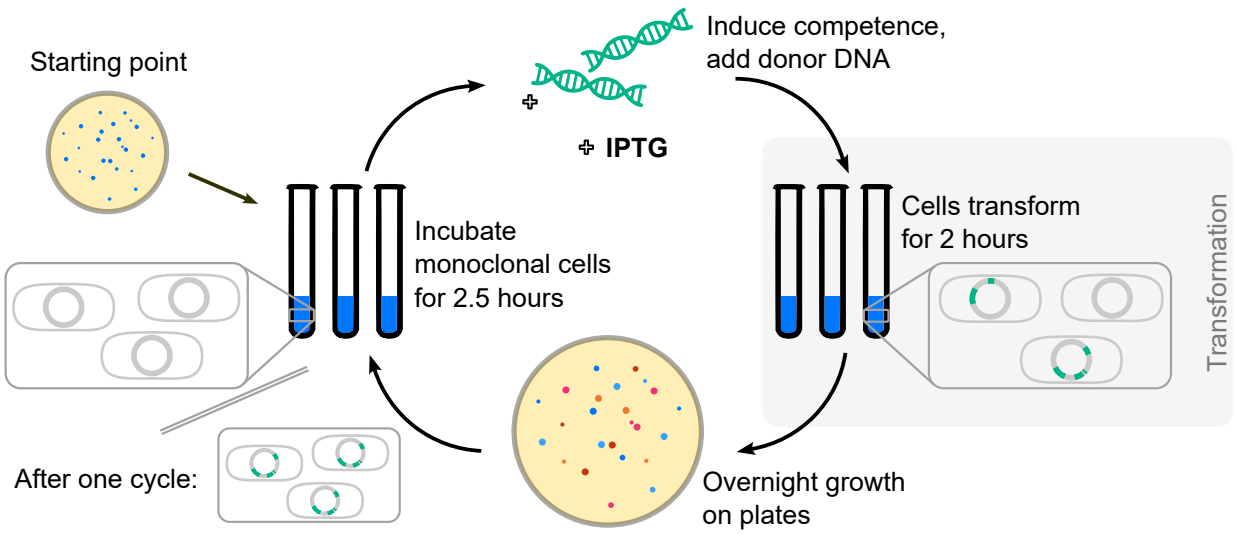


Figure 2.1.: Replacement accumulation assay. At the beginning, a single colony is picked from an LB agar plate, resuspended in complex medium and distributed over all start wells. The monoclonal cells are incubated for about 2.5 hours so that they can escape the lag phase and grow exponentially to an OD of approximately 0.1. Then, we add IPTG to induce competence and genomic DNA from a closely related species (donor). Together with the donor DNA, the cells transform for 2 hours while shaking at 37 ℃. After transformation, the cells are washed and plated on LB agar plates, where they grow overnight at 37 ℃ in the incubator. The next morning, single colonies are picked for the start of the new cycle (single-cell bottlenecking).

2.3. Data availability

All primary and processed data obtained in the following projects are stored on the servers of AG Maier in agreement with Prof. Dr. Berenike Maier.

2.4. Replacement accumulation assay

During transformation, cells can take up DNA from their surrounding and integrate it into their genome. To better understand the genome dynamics and barriers to integration, we conduct a replacement accumulation experiment. We gradually exchange the ancestor genome with DNA from a closely-related donor by repeating a transformation step followed by a single-cell bottleneck 20 times (Fig. 2.1).

At the beginning, we pick a single colony of Bs166 from an LB agar plate to ensure a monoclonal starting point. The colony is resuspended in complex medium and distributed over 8 to 15 starting wells (depending on the number of desired parallel experiments) of a 24 well plate. After 2.5 hours of incubation at 37 ℃ and shaking at 250 rpm, 600 μM isopropyl-beta-D-thiogalactopyranoside (IPTG), a synthetic analogue of lactose, is added to the cells in order to induce competence. Additionally, genomic DNA of a closely related species is added at a concentration of approximately 1 genome/cell. With IPTG and DNA, the cells are shaken for two hours at 37 ℃. During this time, the competent cells undergo transformation. The cells are then diluted in PBS with a factor of 10^{-5} - 10^{-6} and plated onto an LB agar plate using sterile glass beads. The next morning, we induce a single-cell bottleneck by randomly picking one single colony to restart the cycle.

Depending on their respective donor, hybrids that have undergone 20 cycles of transformation are referred to as BspizHyb (donor: *B. spizizenii*), BvalHyb (donor: *B. vallismortis*), BatroHyb (donor: *B. atrophaeus*) and GeoHyb (donor: *G. thermoglucosidasius*) in Publication 4.1.

2.5. Hybrid library preparation

To learn about the distribution of fitness effects of transformation, we create a hybrid library performing one cycle of the previously explained ”replacement accumulation assay” (Sec. 2.4). Starting with monoclonal Bs166 picked from an LB agar plate, we resuspend the colony in complex medium and fill 10 wells in a 24-well microtiter plate with the cells (Fig. 2.2). Four of these are used for the hybrid library cells, the remaining wells serve as references. After 2.5 hours, when the cells have reached an OD of about 0.1, we add IPTG and gDNA of either *B. vallismortis* or *B. spizizenii* and let them transform for 2 hours. Instead of plating them immediately, we wash the cells thrice in PBS to remove IPTG and DNA, dilute them 1:100 in fresh complex medium and then allow the cells to escape from the competent state. After the cells have resumed growth and the OD has at least doubled, we induce competence again by adding IPTG, this time to resolve the

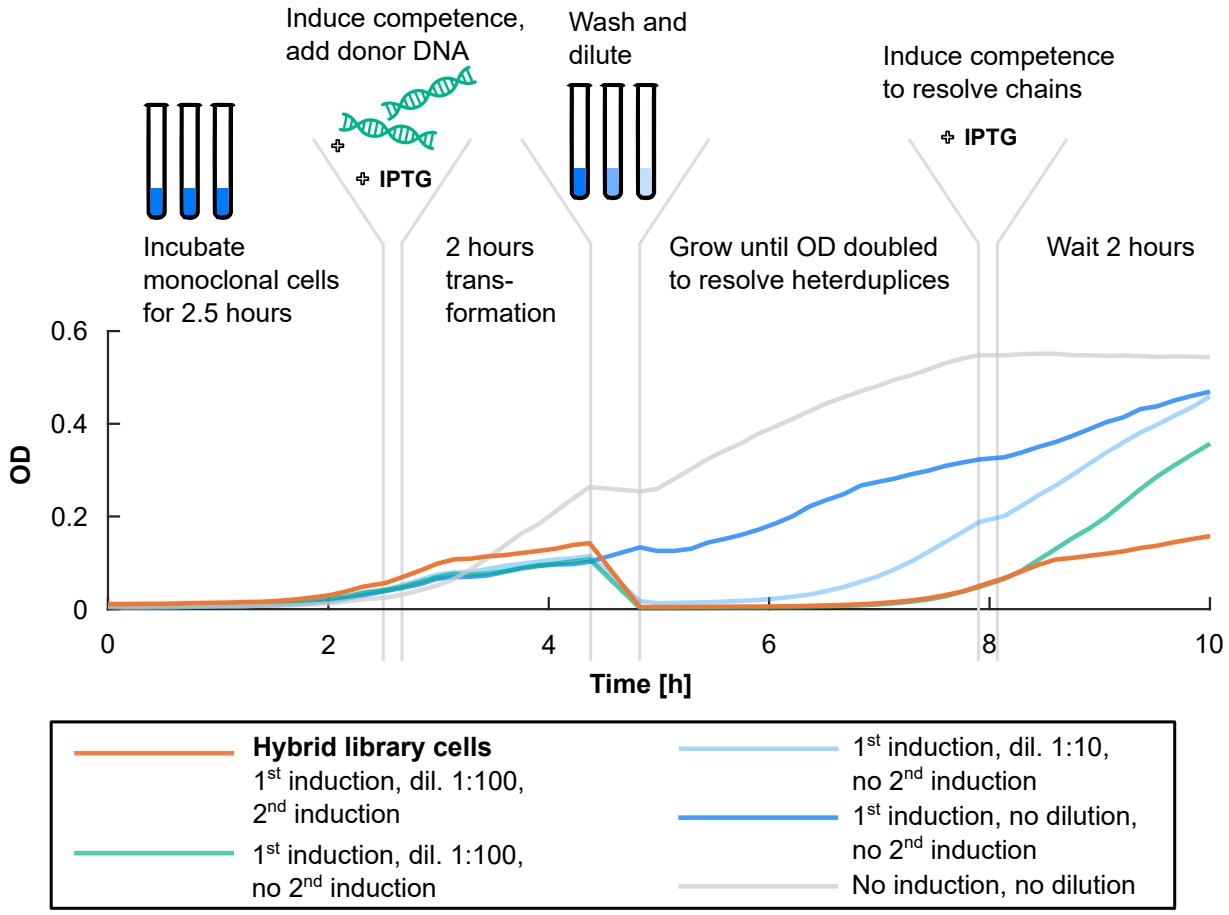


Figure 2.2.: Hybrid library preparation. A Bs166 colony is resuspended in complex medium in the morning (orange). After 2.5 hours, the cells escape lag phase and start to grow exponentially. We add IPTG to induce competence and genomic DNA of either *B. vallismortis* or *B. spizizenii*. To see the effect of growth arrest, a reference is grown to which no IPTG is added (gray). Cells are transformed for 2 hours. After this time, the cells are washed three times in PBS and then diluted 1:100 in complex medium. To track growth and escape from competent state, undiluted (dark blue) and 1:10 diluted (light blue) wells are also inoculated. To resolve the heteroduplexes formed during transformation, we wait until all cells have escaped competent state and divided at least once. When the OD has doubled after about 3 hours, we induce competence a second time, this time to resolve chains that form during growth. Again, a reference is grown without the addition of IPTG (green). Two hours after induction, cells are plated on LB agar plates.

chains that form during exponential growth. Finally, we plate the cells being sure that both heteroduplexes and chains have been resolved. In this way, we ensure that the growing colonies are monoclonal. The next morning, we pick 88 of these monoclonal hybrids to build up for the hybrid library. In the publication 5.1, the hybrid library with *B. spizizenii* as donor is referred to as BSPIZ. Likewise, the hybrid library for which *B. vallismortis* served as donor is referred to as BVAL.

2.6. Fitness via high-throughput competition experiments

In our competition experiment, we first dilute overnight cultures of the strains of interest (SOI) on a 96-well microtiter plate and of a fluorescent competitor in an Erlmeyer flask separately by a factor of 10 (Fig. 2.3 **A**). Overnight cultures and dilutions both happen in the medium and under the condition we use during competition. To eliminate systematic biases due to strain positioning on the 96-well plate, we randomly shift the columns of

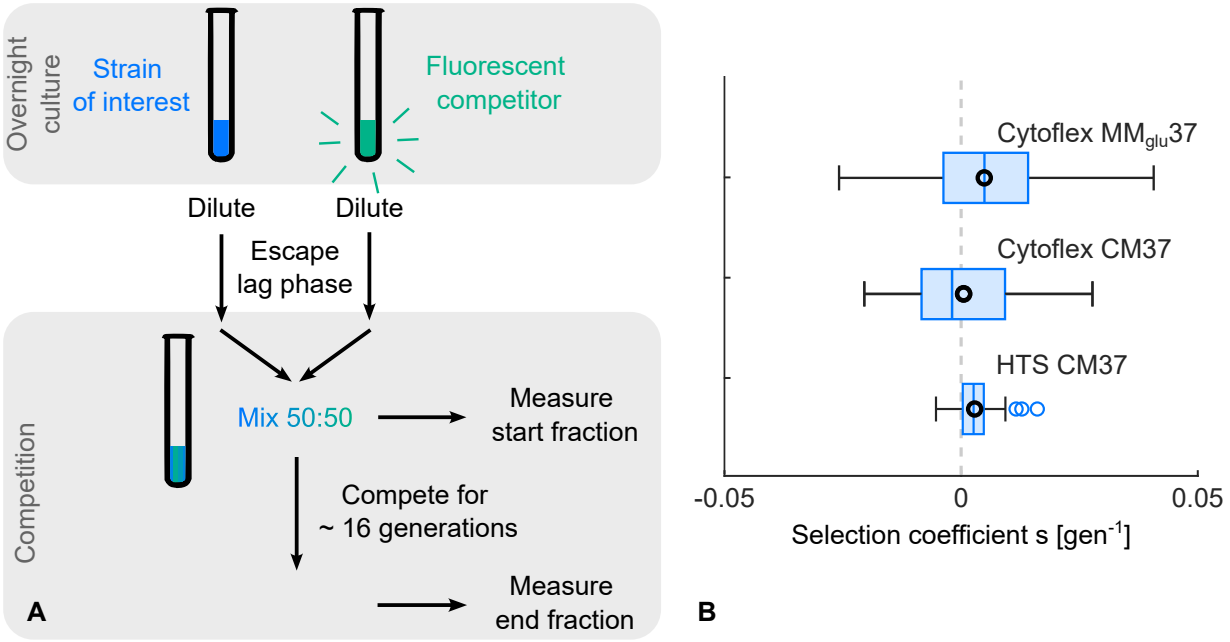


Figure 2.3.: Fitness via competition. **A** Scheme of a competition experiment in exponential growth phase. The overnight cultures of a strain of interest (blue) and a fluorescent competitor (green) are diluted and shaken in an incubator until they escape the lag phase. The exponentially growing strains are mixed in a 50:50 ratio. The exact proportion is measured with a flow-cytometer before and after about 16 generations of competition. **B** The precision of the competition experiment depends on both the used setup as well as the condition and can be measured by performing the competition experiment with the competitor but without fluorescence as SOI and compete it against its fluorescent proteins producing-self. In a perfect system, we would measure a selection coefficient of zero. In our case, we use two different flow cytometers, Cytoflex by Beckman Coulter and Canto II (HTS) by BD. The measured selection coefficients for the two setups are shown for two conditions: 18 hour competition in MM_{glu} at 37 ℃ ($MM_{glu}37$) and/or 4 hour competition in CM at 37 ℃ (CM37).

our plate every morning before dilution. Using a plate reader, we track the optical density until the cells escape the lag phase and growth becomes visible. For complex medium at 37 ℃, this is the case after about 2 hours. For less rich medium or lower temperatures, the time increases. The exponentially growing cells are diluted in an OD dependent manner and mixed in a 50:50 ratio in PBS at an OD of 0.01. This mix is used in the flow cytometer to determine the exact starting fraction. A 1:10 dilution of the mix in growth medium is grown under the selected conditions for about 16 generations. After competition, the mix is diluted 1:5 in PBS and the end fraction is measured with the flow cytometer.

To test the effects of the lag phase on fitness, the first step, the dilution of the overnight culture and the outgrowth of the lag phase, is skipped. The overnight cultures of SOI and the fluorescent competitor are mixed directly.

In project 5.1, the OD-dependent dilution of each individual well was not yet used. Instead, we created 3 - 4 plates with different ratios of non-fluorescent and fluorescent cells and measured them all. This allowed us to ensure that we achieved the 50:50 ratio for the starting fraction for each well on at least one of the plates. All these plates were diluted in growth medium and grown. After 16 generations, the end fraction was determined.

The precision of the competition experiments depends on the flow cytometer used and the condition (Fig. 2.3 **B**). It is determined by performing the measurement with a whole plate of ancestors as SOI. The distribution of selection coefficients of the ancestor is in agreement with zero for both used flow cytometers (Cytoflex by Beckman Coulter and Canto II by BD(HTS)).

2.7. Evolution experiments

Our laboratory strain Bs166 is first pre-adapted and then evolution experiments with the pre-adapted Bs166 are performed according to the following protocols for evolution experiments in a liquid or in a structured environment.

Evolution experiment in liquid environment To ensure competition without nutrient or space limitations, we first designed an evolution experiment in which cells are kept in exponential growth phase by diluting them to a predefined OD every few hours (Fig. 2.4 **A**). In this way, effects from the stationary or lag phase are largely excluded and the maximum number of generations is achieved. A liquid handling system is used to incubate the cells at 37 ℃. All instruments integrated into the robotic system are specified in the Supplements (Tab. A.1). Every 30 minutes the optical density is measured, and every four hours this OD measurement and the liquid handling robot are used to dilute each well with fresh medium back to an optical density of 0.001. All materials used, such as tips and media, are renewed every 24 hours to avoid contaminations.

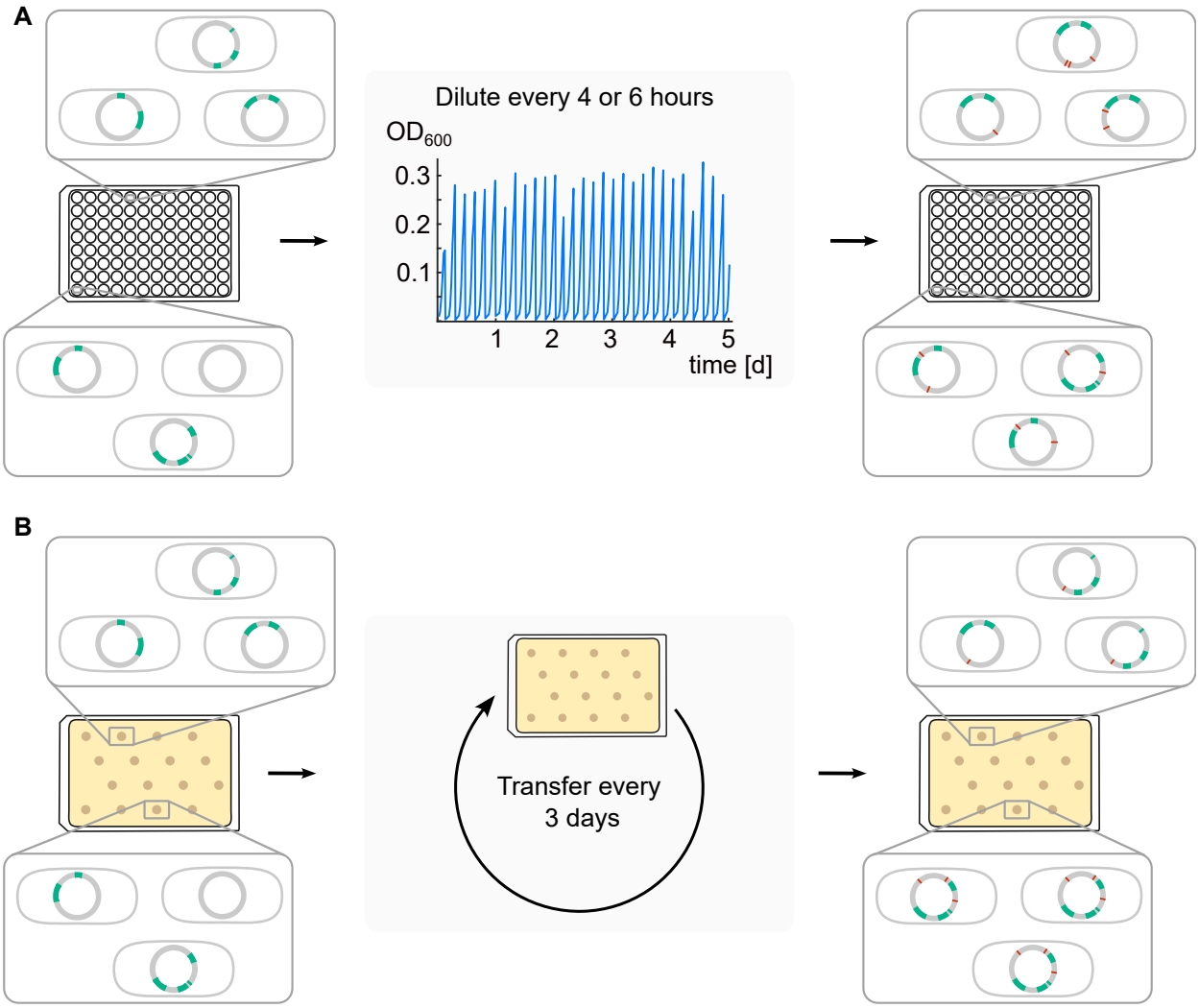


Figure 2.4.: Evolution experiments in liquid and in structured environments. **A** 88 wells of a 96-well microtiter plate are filled with transformed populations with a starting OD of about 0.01. The green parts of the genome depict regions affected by horizontal gene transfer. By diluting the cells down to an OD of 0.001 every four or six hours (depending on the growth rate of the cells in the respective condition, Sec. 2.8), the growth is kept exponential over several days. During growth, the fraction of hybrids shifts and *de novo* mutations arise atop (red). Every 24 hours, a copy of the current plate is mixed with DMSO and stored at -80 ℃. **B** For the evolution experiments in a structured environment, 1 μl of the transformed populations is dropped on agar plates at an OD of 1 and incubated at 30 ℃. Per plate, 16 populations evolve in parallel. Every three days, 7 μl medium is dropped on top of the biofilms to soften the colonies and test for drop collapse. To transfer the cells to fresh plates, we use a 10 μl inoculation loop to scratch the colony from the plate and resuspend the cells in medium. A 1 μl drop of these cells is plated on a fresh agar plate. The rest of the cells is stored at -80 ℃.

Evolution experiment in a structured environment Since uninterrupted exponential growth is far from what we expect in nature, we also designed an evolution experiment on agar plates (Fig. 2.4 **B**). Initially, cells are inoculated as 1 μl drops with an OD of approximately 1. The colonies grow at 30 ℃ on complex medium plates supplemented with 1.5 % agar. Every three days, 7 μl of complex medium is dropped onto the colony to test for drop collapse and to moisten the biofilm. With a 10 μl inoculation loop, the softened colony is scratch off the agar and transferred to liquid complex medium. A 1 μl drop of this liquid is inoculated on fresh plates for the next three-day growth period. The rest is frozen at -80 ℃.

2.8. Measuring generation times

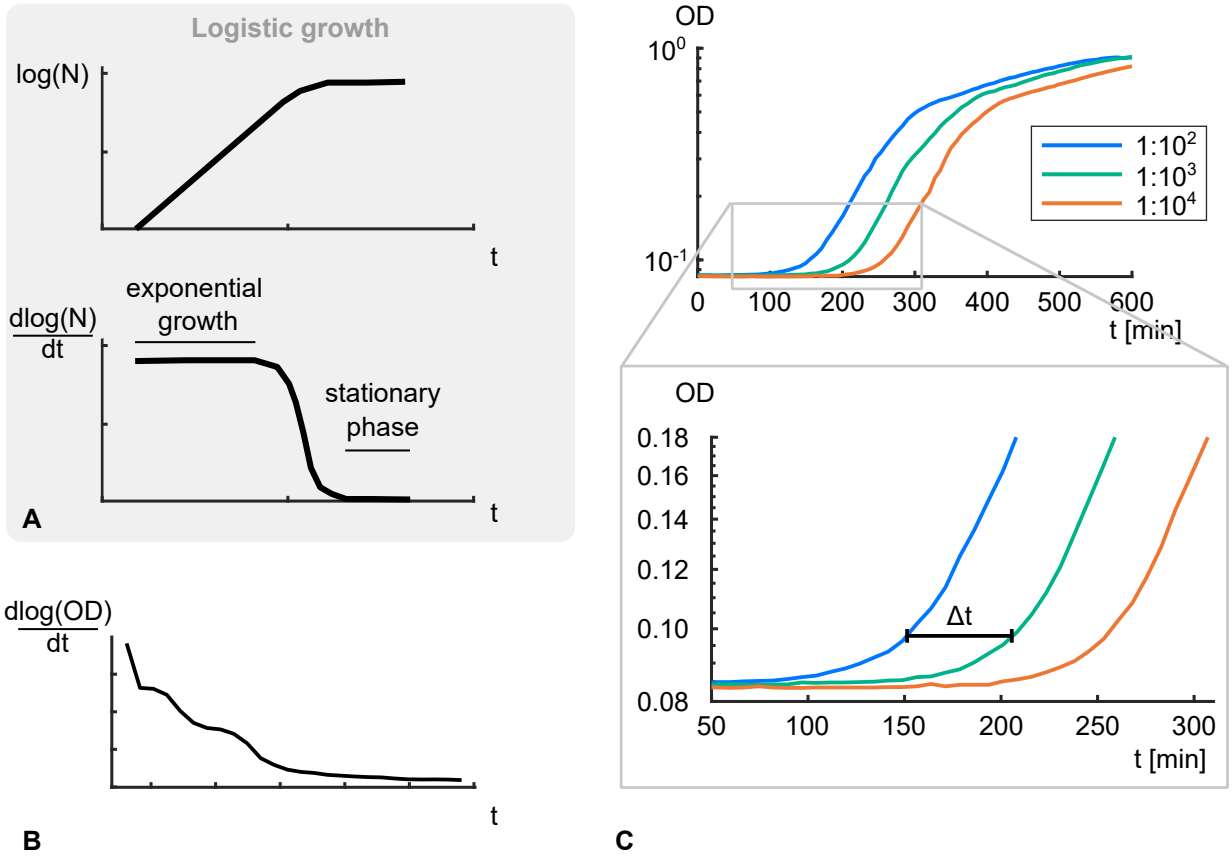
Cell growth approximately follows a logistic function with number of cells N at time t

$$N(t) = \frac{N_0 \cdot \exp(kt)}{1 - (N_0/N_{\max})(1 - \exp(kt))} \tag{2.1}$$

where N_0 is the number of cells at time 0, k the maximum growth rate and $N_{\max}$ the maximum cell number (Fig. 2.5 **A**). Using a plate reader, the cell number can be estimated from the measured optical density within a certain range. During exponential growth, the slope of the logarithmic OD is constant. The generation time of cells in exponential growth phase can be determined by finding the plateau of the time derivative of the log(optical density) (Fig. 2.5 **A**). For *B. subtilis*, this plateau seems to lie outside the resolution of the plate reader (Fig. 2.5 **B**). Using the slope of this OD curve would lead to a underestimation of exponential growth rate since the cells are already in the transition between exponential and stationary phase. So instead of using the visible area of the OD curve directly, we make a dilution series with a factor of 10 to find the rate. We use the OD curve of the cells and their 1:10 dilution to determine the time required for the dilution to reach the same OD value as the undiluted culture (close to the resolution minimum, Fig. 2.5 **C**). In this way, we do not know the maximum speed but we calculate the mean of generation time t_{gen} to that point including the invisible phase

$$t_{\text{gen}} = \frac{1}{k} = \frac{\Delta t}{\ln(\text{dilFact})} \tag{2.2}$$

where Δt is the time needed to reach the same OD after dilution with dilution factor dilFact. The detection algorithm was implemented by Jeffrey J. Power. Every experiment day, about 20 - 30 dilutions are measured per strain and per condition. Each experiment is repeated at least three times. The average generation times and standard errors for several strains and conditions are listed in figure 2.5 **D**.



D

Strain	Generation time [min]	Growth condition
Bs210	16.1 ± 0.2	in CM at 37 °C
Bs224	16.2 ± 0.3	in CM at 37 °C
Bs166	17.1 ± 0.4	in CM at 37 °C
Bs166	14.9 ± 0.2	in CM at 42 °C
Bs166	39.0 ± 0.4	in MM_{glu} at 37 °C
Bs166	48.3 ± 0.6	in MM_{gly} at 37 °C

Figure 2.5.: Generation times via dilution. For bacteria in liquid, we expect a logistic growth function (**A**). The time-derivative of the logarithm of the cell number shows a plateau during exponential growth and goes down to zero when reaching stationary phase. From the plateau the generation time in exponential growth phase can be calculated. In our experiments, where we estimate the cell number by measuring the optical density, we cannot see this plateau (**B**). Instead, we use a dilution series to estimate the mean generation time from the start of growth to visible growth (**C**). Repeating the measurement three times for each strain, we can estimate the generation times of our strains in various conditions (**D**, mean ± standard error).

2.9. Transformation efficiency

To assess how comparable the transformation rates are in our experiments, the transformation efficiency of each strains is determined before they are used in an experiment. We test it by letting the cells transform in the presence of gDNA of the same strain but with a point mutation that leads to rifampicin resistance. Cells that integrated the part with the point mutation grow on LB agar plates with rifampicin. This assay can only be used because the transformation rates are much higher than the mutation rates of our strains.

Cells from an overnight culture are diluted to an optical density of 0.1 in a total volume of 10 ml. After 2.5 hours of growth at 37 ℃ in the shaker (250 rpm), 500 μl of the cell culture is transferred to a 15 ml Falcon tube. 500 ng of gDNA of the rifampicin resistant strain is added to the cells. After shaking for two hours, the cells are washed and diluted in PBS. After dilution at a ratio of 1:10^1, 1:10^2 and 1:10^3, the cells are plated on LB agar plates with rifampicin (5 μg/ml) to count the resistant cells. Additionally, cells diluted 1:10^4, 1:10^5 and 1:10^6 are plated on LB agar plates without antibiotics to count the total number of cells in the culture. The fraction of resistant cells in all cells can be calculated (Tab. 2.5). For each strain, the experiment is repeated at least three times on three different days. As can be seen from table 2.5, the number of transformants is comparable for all three strains.

Strain	# transformants / total # cells
Bs166	0.001 $\pm$ 0.000(4)
Bs210	0.002 $\pm$ 0.001
Bs224	0.001 $\pm$ 0.001

Table 2.5.: Fraction of transformants for all strains that were transformed in the course of this project (mean $\pm$ standard deviation).

2.10. DNA isolation

For whole genome sequencing (WGS) and for gDNA preparation for all transformation assays, we grow monoclonal strains overnight in LB shaking at 37 ℃. 2 ml of the cells are pelleted in the morning at 16,700$\times$g for 3 minutes and the supernatant is discarded. Using the DNeasy Blood & Tissue Kit from Quiagen, gDNA with an average strand length of 30 kbp is extracted and sent for Illumina short read sequencing to Eurofins Genomics or is used as donor DNA.

3. Computational methods

The following projects require several computational methods for the analysis of genome characteristics and changes (Sec. 3.2), flow cytometer data (Sec. 3.3) and biofilm images (Sec. 3.4). In the following, the pre-processing of raw data from whole genome sequencing (WGS) (Sec. 3.2.1) and the subsequent detection of genomic variations like duplications, deletions or mutations (Sec. 3.2.3) will be described in detail. In particular, we describe how we detect horizontal gene transfer events (Sec. 3.2.4). We also use WGS data and the NCBI dictionaries for our strains to identify multimapping areas (Sec. 3.2.5) and distinguish between accessory and core regions (Sec. 3.2.2). In addition to the bioinformatic analysis of large sets of sequencing data, algorithms for fitness characterization have been developed as part of this thesis. This includes the analysis of flow cytometry data for the characterization of selection coefficients (Sec. 3.3). Since *B. subtilis* grows in filaments during exponential growth phase, we apply a filament correction to the flow cytometry data (Sec. 3.3.1). For evolution experiments on agar plates, we have developed image processing methods to capture the changes in colony size over time (Sec. 3.4).

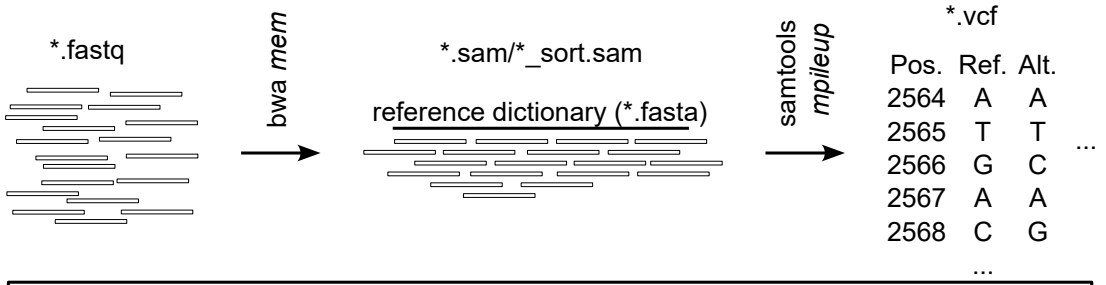
3.1. Data availability

All scripts used in Section 3.2 can be found under https://doi.org/10.5281/zenodo.8273150 and are referred to as ”folder number/script name” [130]. The scripts are co-authored by Isabel Rathmann and Mona Förster and were published as part of the Publication 4.1. The scripts for the Sections 3.3 and 3.4 can be found on GitHub in the folder ”mofoe/phdThesis”.

3.2. Genome analysis

3.2.1. Processing Illumina reads

From Eurofins Genomics we receive paired Illumina reads with a length of 2×150 bp and an average coverage of about 500 reads per position (in fastq-format). All used commands for the following software are shown in Figure 3.1. The reads are aligned to a reference dictionary downloaded from NCBI (dictionaries specified in the Supplements, Tab. A.3) using the *mem*-function of the Burrow-Wheeler Aligner (bwa) and saved in sam-format (bwa v0.7.17) [131]. For faster processing, the reads in the sam-file are sorted and indexed with samtools *sort* and samtools *index* (samtools v1.13) [132]. Then, the per-read-information is translated into per-position-information with the help of bcftools




```
bwa mem -t 8 -B 1 -O 1 *.fasta forward-reads.fastq.gz reverse-reads.fastq.gz > *.sam

samtools sort *.sam --threads 8 --reference *.fasta -o *_sort.bam
samtools index -b *_sort.bam > *_sort.bam.bai

bcftools mpileup -e 10 -F 0.00001 -h 80 -L 10000 -o 20 -a FORMAT/AD -d 8000
                                                    -f *.fasta *_sort.sam > *.vcf

bcftools call -vc *.vcf > *_bcfcall.vcf
```

```
bedtools genomeCoverageBed -d -ibam *_sort.sam > *_coverage.txt
```

Figure 3.1.: Raw data processing of Illumina reads. The 150 bp long paired-end reads in fastq-format are aligned along a reference dictionary in fasta-format with the help of Burrow Wheeler Aligner's (bwa) *mem* function. Since we want to detect replacements in which nucleotides differ in various positions along the genome, we have to choose very loose mapping constraints. The aligned reads are preprocessed with samtools *sort* and *index*. With the information about the most likely position on the reference genome, the read information can be transferred into per-position information via bcftools *mpileup*. To filter out the unchanged positions, the vcf-file is post-processed with bcftools *call*. The coverage can be extracted from the sorted sam-file using bedtools *genomeCoverageBed*.

mpileup-function and saved in vcf-format (bcftools v1.13). For each position, the vcf-file contains information about the raw read depth, the quality and the expected and found nucleotide. Furthermore, small inserts and deletions with the size of a few base pairs can be found in the file. To reduce the size of the vcf-file, only the variants are filtered out using bcftools *call* [133]. With bedtools *genomeCoverageBed* the per-position coverage can be extracted from the sorted sam-file (bedtools v2.31.0) [134]. The coverage determined using bedtools and the raw read depth determined using samtools are comparable measures, but not identical. The versions specified are the ones currently used. In the Publications 4.1 and 5.1, older version of the tools were used in some cases (see methods of the publications). The script for WGS raw data processing shown in figure 3.1 is also available under "0/2_CombiScriptSamtools.sh" [130].

3.2.2. Accessory and core regions

Accessory and core regions are defined by comparing two strains of interest nucleotide by nucleotide. If regions between the two strains are homologous, they are defined as part of the core genome. Parts that are unique in one of the two genomes are defined as

accessory genome. Please note that this definition differs from the definition of core and accessory genomes often used when comparing multiple strains [135, 136].

Methodically, the two regions are identified by aligning reads of the one strain along the dictionary of the other strain and determining the raw read depth, as explicitly explained in the prior section. Regions where the raw read depth is lower than 50 (approximately 10 % of the average raw read depth of our reads) for more than 150 bp in a row, are part of the accessory genome of the strain whose the dictionary was used (cf. Fig. 3.2 **C**). All other regions are core genome for both strains. The accessory regions of Bs166 when compared to *B. spizizenii*, *B. vallismortis*, *B. atrophaeus* and *G. thermoglucosidasius* are shown in the Publication 4.1 in Figure 3. The Matlab script for identifying accessory regions is available under ”1/A0c_AccessoryGenome.m”.

3.2.3. Detecting genomics variations via coverage

In the following projects, we use the coverage of the aligned sequencing data to identify duplications, deletions and insertions (Fig. 3.2). For deletions and duplications, we analyze the coverage when aligning along the recipients dictionary. At least ten positions in a row need to have a coverage of zero for a deletion (Fig. 3.2 **A**). Using an artificially created mock genome, we can verify the algorithm to identify deletions reliably down to

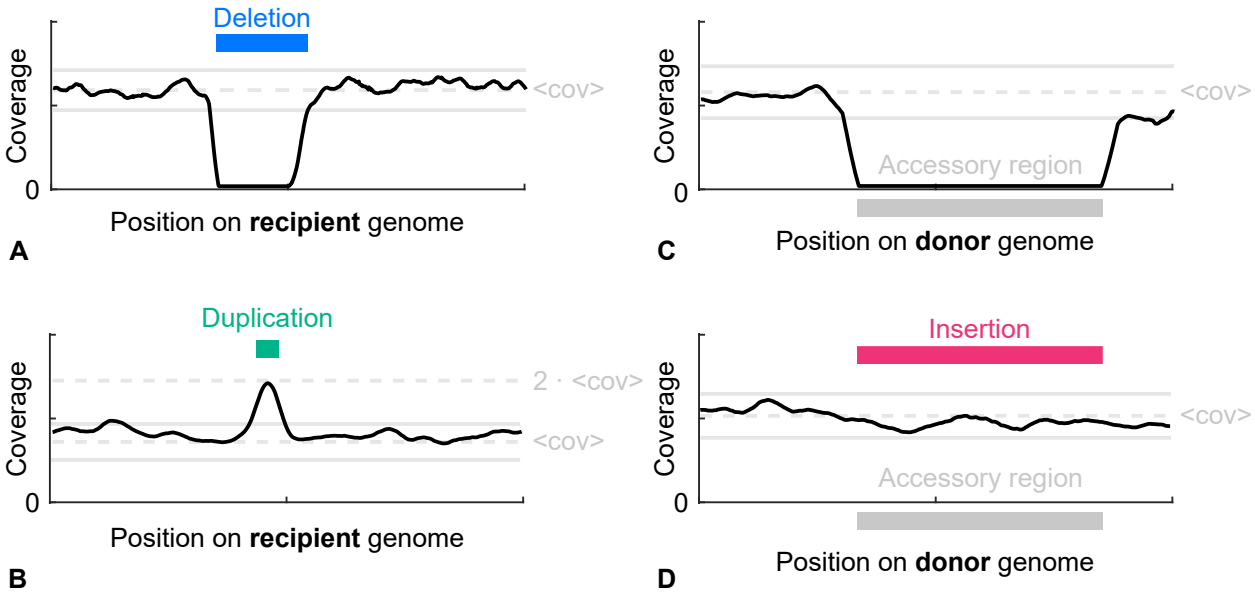


Figure 3.2.: Genomic variations via coverage. Deletions and duplications can be detected by analyzing the read coverage on the recipient dictionary (**A, B**). In the case of deletions, the coverage goes down to zero (**A**). In a duplication, the coverage reaches twice the average coverage (**B**). We test the accuracy of our predictions by generating strains with artificial deletions and duplications (Supplements A.4). Aligning the reads along the donor dictionary is used to identify insertions from the accessory regions of the donor (**C, D**). Before an insertion occurs, the accessory regions look like a deletion (**C**). The coverage goes down to zero. If this region is inserted in the hybrid strain, we detect coverage within this region (**D**).

a deletion length of 200 bp. The results of the mock genome analysis for deletions are shown in the Supplements (Fig. A.1 **B**). If twice the average coverage ⟨cov⟩ is hit for at least ten positions in a row, we infer a duplication (Fig. 3.2 **B**). Again, we use a mock genome to show that duplications longer than 200 bp can be detected reliably (Supplements Fig. A.1 **A**).

Insertions of donor accessory parts are also detected via coverage. In contrast to deletions and duplications, the hybrid reads are not aligned along the recipient genome, but along the donor genome. The alignment of a strain that has no insertions leads to stretches with zero coverage, the so-called accessory regions (Fig. 3.2 **C**). If such a region shows a non-zero coverage, we can assume that this part has been inserted during transformation (3.2 **D**). However, this analysis method does not provide any information about the position of integration. The scripts for deletion, duplication and insertion detection are available under ”1/A3_Cov2DelDup.m” and ”1/A4_Cov2Ins.m”.

3.2.4. Detecting replacements

The algorithm for detecting replacements in transformation hybrids is adapted from Power et al. [52]. First, a so-called master list is created (Fig. 3.3). The reads of the donor strain are aligned along the recipient genome. All single nucleotide polymorphisms (SNPs) of homologous regions between recipient and donor can be identified. These differences will be called master list SNPs (mlSNPs) in the following. Subsequently, the SNPs that we detect when aligning hybrid reads along the recipient genome can be compared to the master list. SNPs detected in hybrid strains which are absent from the master list did

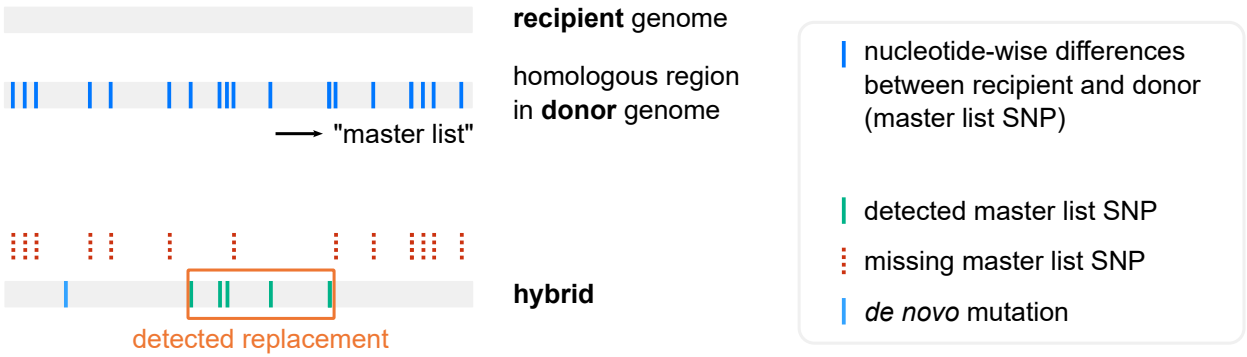


Figure 3.3.: Detection of replacements. In order to detect homologous recombination of parts of the donor genome into the recipient, a ”master list” is created. For this purpose, all homologous regions of recipient and donor are compared. The detected single nucleotide polymorphisms (SNPs) are called master list SNPs (mlSNPs, dark blue). The SNPs found in the hybrids are then compared to the master list. SNPs that do not fit the master list and thus most likely did not arise from horizontal gene transfer are assigned as *de novo* mutations (light blue). SNPs whose position and allele matches the master list are detected mlSNPs (green). Two detected mlSNPs start a ”cluster” (detected replacement). This cluster is extended as long as less than five mlSNPs are consecutively missing (dotted red). Five or more consecutively missing mlSNPs end the cluster.

most likely not arise from replacement with genetic material from the donor. We denote those SNPs as *de novo* mutations. SNPs where the position and the allele matches the master list are detected mlSNPs. Two detected mlSNPs start a cluster (detected replacement). A cluster ends when five or more mlSNPs are missing in a row. The detection is independent of direction. The script for detecting replacements can be found under ”1/A1_SNP2CNP.m” [130]. For pre-processing, the output file from Section 3.2.1 (*_bcfcall.vcf) needs to be run through ”1/A0_VariantFiltering.m”. To create the master list use ”1/A0b_MasterListFiltering.m”.

3.2.5. Multimapper regions

Genomic regions with copy numbers greater than one pose a problem for alignment when short-read sequencing is used. Due to the shortness of the reads, we cannot determine to which copy a read belongs if the region is longer than 300 bp. Therefore, we identify these regions and deliberately exclude them from our analyses. For identification, we use NCBI *blastn* [137] to align the dictionary of a strain to itself and evaluate the results with a Matlab script. This script is available under ” 1/A0d_Multimapper.m” [130]. A list of all multi mapping regions we detected in Bs166 and the genes within these regions can be found in the Supplements (Tab. A.2).

3.3. Evaluation of flow cytometry data

The raw data from the flow cytometer contains information about the forward and side scatter of each event as well as the fluorescence. For the analysis, we create a gate for forward and side scatter heights (FSC-H/SSC-H) to filter out events that do not fall within the typical properties of our cells (Fig. 3.4 **A**). To this end, we perform a measurement with fluorescent cells only. Without doubt, all fluorescing events are cells. The gate has to be adjusted if different strains are used or if the flow cytometer settings change. Within this gate, the fraction of fluorescent versus non-fluorescent cells is calculated (Fig. 3.4 **B**) and used for further determination of the selection coefficients as it is described in Section 1.3.1. The scripts used to process the raw data and analyze the fractions are available on GitHub under ”FC_01_rawData.m” and ”FC_02_analysis.m”

3.3.1. Filament correction of flow cytometry data

B. subtilis is known to form filaments during the exponential growth phase. When using a flow cytometer, filament formation can lead to an underestimation of cell number, as a filament comprising multiple cells is detected as a single event (cell). If SOI and competitor have comparable tendency to form filaments, the problem averages out. However, during adaptation to exponential growth in liquid, a frameshift mutation in *sigD* favors filamentation (Publication 5.1). Then, the filaments measured in the flow cytometer contain on average more cells per event than their unadapted competitors. This leads to a severe underestimation of the cell number of the SOI. In the following, we describe how we use the width of the forward scatter signal FSC-W to correct for the effect of filamentation.

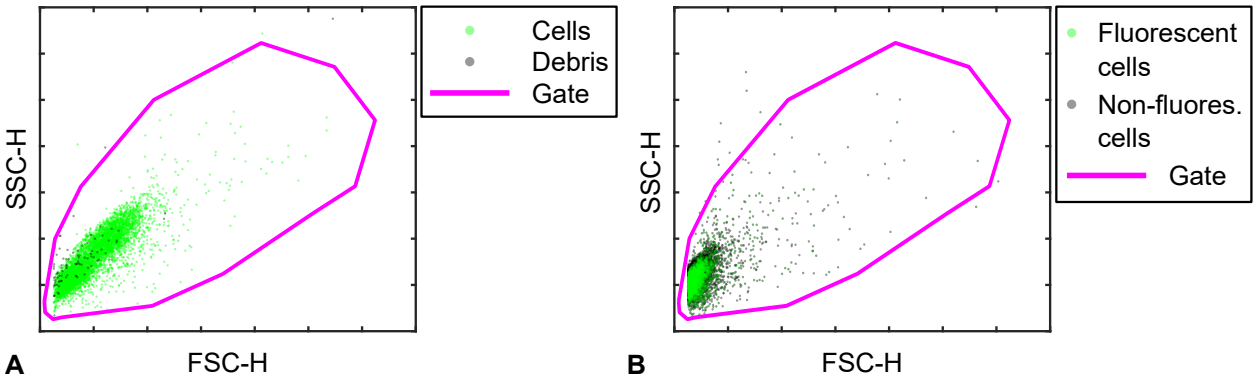


Figure 3.4.: Processing of flow cytometry data. A To create a gate, only fluorescent cells are measured with the flow cytometer. Thus, we can distinguish between cells (green) and debris (non-fluorescent). **B** When the gate is set, only events within this gate are considered cells, and the fraction of fluorescent and non-fluorescent cells within the gate is calculated.

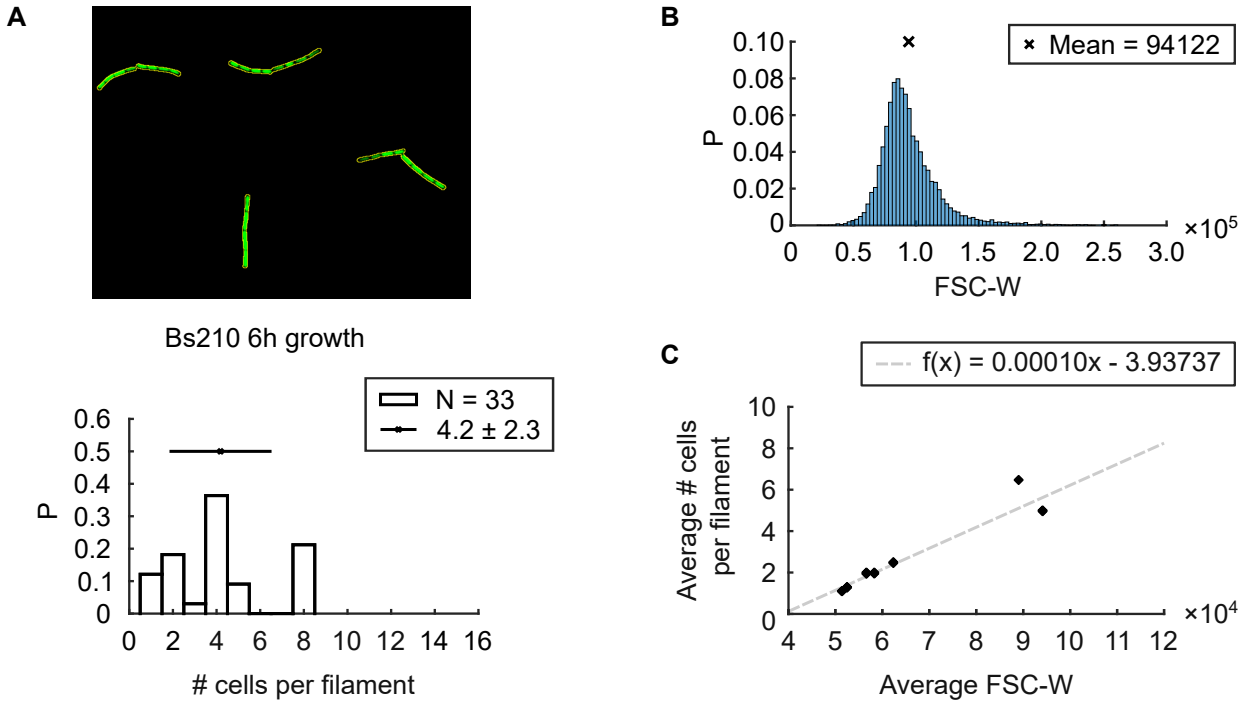


Figure 3.5.: Filament correction of flow cytometry data. For the ancestor strains Bs210 and Bs224 in different growth phases, the average number of cells per filament is counted (**A**). To this end, cells are stained with Syto9 and visualized under the microscope. Cells from the same culture are measured with the flow cytometer and the average FSC-W is determined (**B**). We find a linear relationship between the average number of cells per filament and the average FSC-W (**C**) and use it to correct our measurements.

At a constant flow through the cytometer, the FSC-W contains information about the event length. To test whether the FSC-W correlates with the cell number of an event, we measure cells with different filament lengths with the flow cytometer and additionally visualize them with the microscope (Fig. 3.5). To reliably count the average number of cells per filament, we label the cells with Syto9, a green-fluorescent chromosome stain (Fig. 3.5 **A**). Then, we correlate the average cell number with the average FSC-W of the sample (Fig. 3.5 **B**). By evaluating Bs210 and Bs224 at different growth phases (stationary phase, after 2 hours of growth, after 6 hours of growth), we can fit a linear function and use it to correct all flow cytometer raw data (Fig. 3.5 **C**). The analysis of the different strains in different growth phases is shown in Supplements D (Fig. E.23).

3.4. Image analysis of biofilms growing on agar plates

While the populations are growing on agar plates (Sec. 2.7), images are taken every hour using a plate imaging robot (BM3-BC, S&P Robotics Inc.) kindly provided by the group of Tobias Bollenbach. These images are processed with Matlab [139]. First, the images are converted from colored to gray with the Matlab function *rgb2gray* (Fig. 3.6 **A**,**B**). To find a threshold that distinguishes between background and colony, the *adaptthresh*-function with sensitivity 0.59 is used (Fig. 3.6 **C**). With this threshold, the Matlab function *imbinarize* converts the image into a binary image (Fig. 3.6 **D**). We remove the frame of the agar plate with a mask (Fig. 3.6 **E**) and then use the *bwboundaries*-function with the

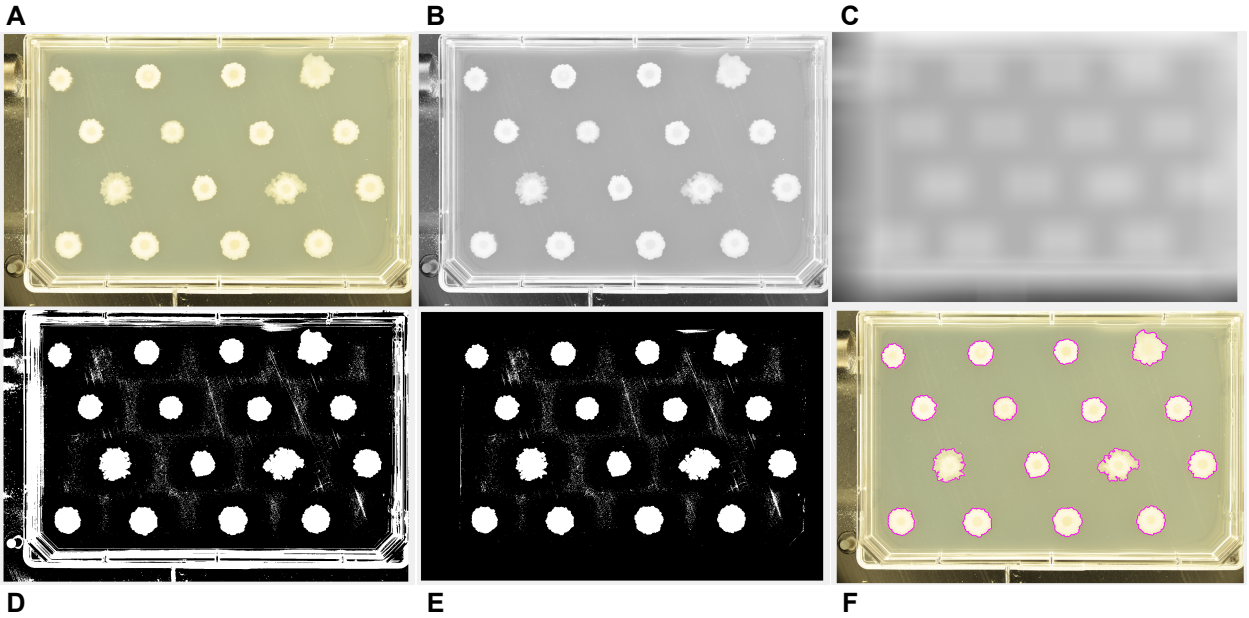


Figure 3.6.: Image analysis of biofilms on agar plates. The original image (**A**) is converted from RGB to gray (**B**). An adaptive threshold is applied to distinguish between colonies and background (**C**). The threshold is used by the *imbinarize*-function to convert the image into a binary image (**D**). With a mask, we remove the agar plate frame from the image (**E**) and then use the *bwboundaries*-function to find the contours of the biofilms (**F**).

”noholes”-option to obtain the contours from the binary image. Combined with the approximate expected positions of the colonies, we can identify the contours that belong to our biofilms (Fig. 3.6 **F**). Examples of detected contours are shown in Supplements D (Fig. D.22). Based on the contours, we can calculate the perimeter and area of the colonies. The script is available on GitHub under the name ”IA_pinningRobot2colonyProps.m”.

4. Genome-wide transformation reveals extensive exchange across closely related *Bacillus* species

4.1. Publication

This project was published with the title ”Genome-wide transformation reveals extensive exchange across closely related *Bacillus* species” in Nucleic Acids Research, Volume 51 in November 2023 under the Creative Commons CC-BY-NC license [140]. This publication reports the results of a joint project between me (MF), Isabel Rathmann (IR), Melih Yüksel (MY), Jeffrey J. Power (JJP) and Berenike Maier (BM), MF and IR being shared first author.

4.1.1. Contributions

The replacement accumulation experiment was designed by JJP, MY and BM and performed by MF and IR under supervision of JJP. MY conducted one of the replacement accumulation assays followed by selection with *B. spizizenii* as donor that was used for the pooling analysis. The rest of the experimental work was divided evenly between MF and IR. Both, MF and IR, isolated the genomic DNA and analysed the WGS data of half the samples each. The basic pipeline for WGS analysis was set up by JJP. The scripts for the advanced analysis were written by MF and/or IR. MF did the statistical analysis on mosaic events. IR and BM wrote the manuscript and designed the figures. MF wrote additions and was responsible for the paper revision. All authors read and discussed the article.

4.1.2. Key findings

Horizontal gene transfer plays an important role in bacterial evolution. To understand the barriers to transformation, an important mode of HGT, on an whole-genome level we perform an replacement accumulation assay with minimal selection. As recipient, *B. subtilis* is supplied with genomic DNA of other *Bacillus* species and a *Geobacillus* species. We find

(i) nearly no restrictions of replacements in the core genome of closely related species. More than 96 % of the core genes of *B. subtilis* are replaced at least partially by *B. spizizenii* homologues when pooling the WGS data of over 40 independent hybrid strains.

(ii) Orthologous replacements are dominating the genome dynamics during transformation and also govern the gain and loss of accessory genes which occur at a much lower rate.

(iii) The exponential relation found for the transfer probability of genes and their sequence divergence in previous studies holds for the transfer rate at the whole core genome level (Fig. 4.1, [47, 48]).

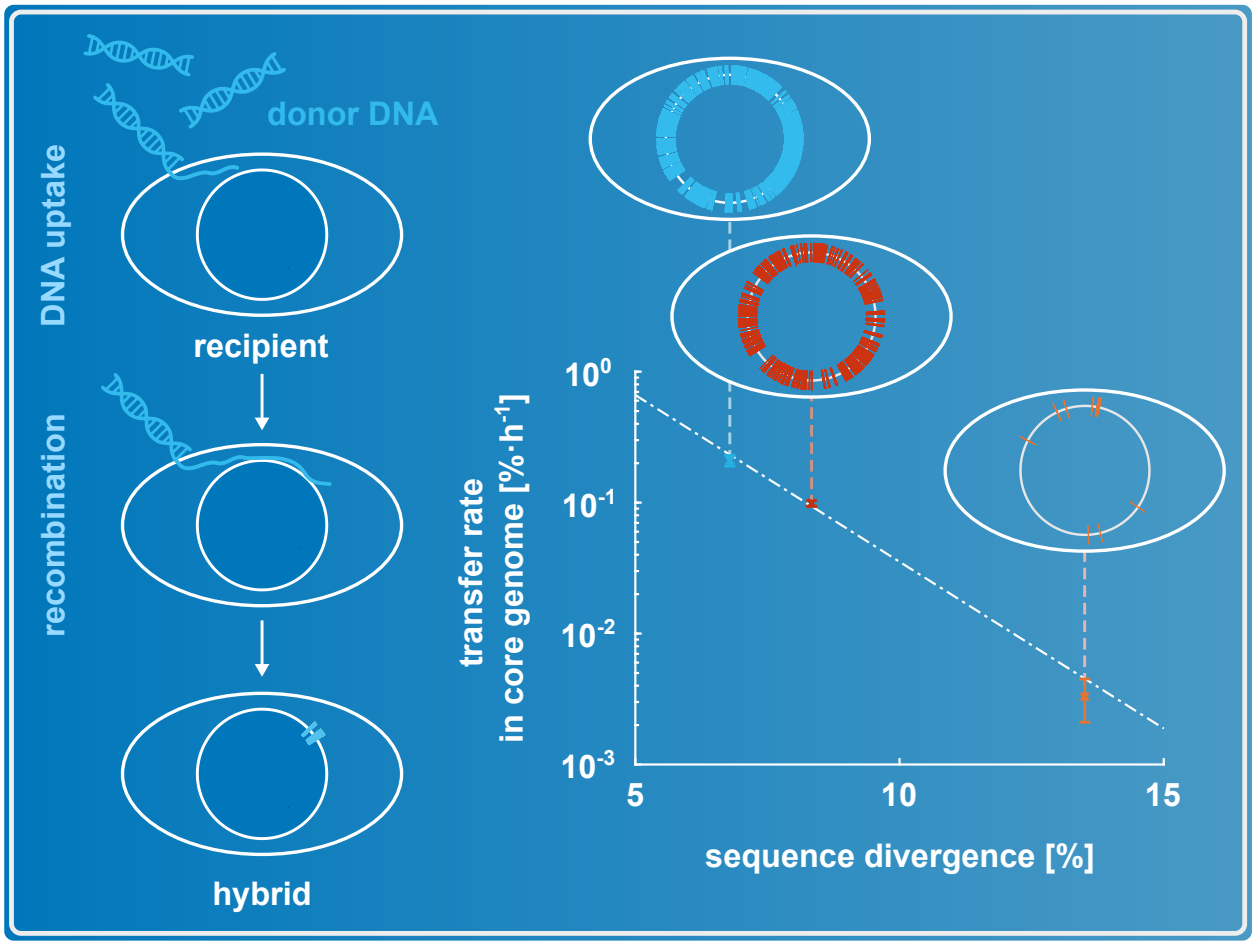


Figure 4.1.: Graphical abstract of the following publication [140]. In the replacement accumulation experiment, competent recipient cells are supplied with gDNA of other *Bacillus* species or a *Geobacillus* species. Through transformation, hybrids between donor and recipient are formed. Using different *Bacillus* species as donor, we show that the rate of core genome transfer decreases exponentially with increasing sequence divergence of the donor-recipient pair. No transfer is found using the *Geobacillus* species as donor. Image used under CC-BY-NC license.

Nucleic Acids Research, 2023, **51**, 12352–12366
https://doi.org/10.1093/nar/gkad1074
Advance access publication date: 16 November 2023
Genomics

Genome-wide transformation reveals extensive exchange across closely related *Bacillus* species

Mona Förster[1,†], **Isabel Rathmann**[1,†], **Melih Yüksel**[1], **Jeffrey J. Power**[1] **and Berenike Maier**[1,2,*]

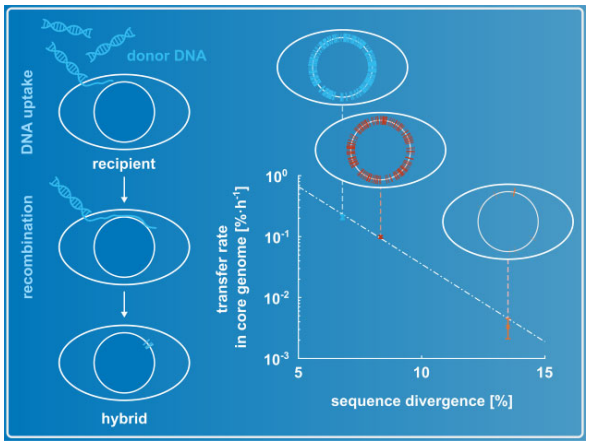
[1]Institute for Biological Physics, University of Cologne, Zülpicherstr. 47a, 50674 Köln, Germany
[2]Center for Molecular Medicine Cologne, University of Cologne, Cologne, Germany

*Correspondence may also be addressed. Tel: +47 221 470 8046; Email: berenike.maier@uni-koeln.de
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Bacterial transformation is an important mode of horizontal gene transfer that helps spread genetic material across species boundaries. Yet, the factors that pose barriers to genome-wide cross-species gene transfer are poorly characterized. Here, we develop a replacement accumulation assay to study the effects of genomic distance on transfer dynamics. Using *Bacillus subtilis* as recipient and various species of the genus *Bacillus* as donors, we find that the rate of orthologous replacement decreases exponentially with the divergence of their core genomes. We reveal that at least 96% of the *B. subtilis* core genes are accessible to replacement by alleles from *Bacillus spizizenii*. For the more distantly related *Bacillus atrophaeus*, gene replacement events cluster at genomic locations with high sequence identity and preferentially replace ribosomal genes. Orthologous replacement also creates mosaic patterns between donor and recipient genomes, rearranges the genome architecture, and governs gain and loss of accessory genes. We conclude that cross-species gene transfer is dominated by orthologous replacement of core genes which occurs nearly unrestricted between closely related species. At a lower rate, the exchange of accessory genes gives rise to more complex genome dynamics.

Graphical abstract



Introduction

Horizontal gene transfer (HGT) is an important factor in bacterial evolution. It plays a major role in providing non-sexually reproducing organisms with genetic variability. Phylogenetic studies have shown that a surprisingly large fraction of bacterial genomes and gene classes have been affected by HGT over the course of evolution (1–4). However, the genome-wide dynamics of HGT are poorly characterized.

In bacteria, one of the main mechanisms mediating HGT is natural transformation. In this process, cells take up DNA from the surrounding and integrate it into their genome in a process called homologous recombination. *Bacillus subtilis* is among the 80 species known to be competent for transformation (5,6). After external DNA has been taken up into the cytoplasm, the recombination machinery performs a homology search on the genome (7). Microhomologies as short as 8 nt are recognized and further binding of neighboring nucleotides can initiate branch invasion of the genome (8). If branch migration becomes stabilized, a heteroduplex can form as a three-stranded D-loop (9). Several studies report the formation of mosaic patterns between donor and recipient alleles resulting from this process (10–14). It remains unclear by

Received: July 3, 2023. **Revised:** September 7, 2023. **Editorial Decision:** October 24, 2023. **Accepted:** November 2, 2023

which mechanism and how frequently these mosaic patterns are formed.

Various processes limit the efficiency of HGT by transformation. At the level of DNA uptake, competence for transformation is tightly controlled in some species including *B. subtilis* (15,16). Once inside the cell, DNA from different strains or species is subject to degradation by restriction modification systems (17,18). Furthermore, the sequence divergence between the donor and recipient alleles has been established as a main barrier to recombination. Multiple studies have approached this dependence by investigating the rate and outcome of transformation for sets of predefined, single genes (19–22). An exponential decrease in the replacement probability was found for different bacterial species (22,23), including *B. subtilis* (19,24) and *Saccharomyces cerevisiae* (20). Likewise, increasing sequence divergence was found to cause decreasing integration lengths (19). Studying the effects of sequences divergence between single orthologous genes is well suited for understanding specific details in the process, yet it falls short of capturing the whole genome transfer dynamics and effects of accessory genes. Genome-wide recombination between different strains and species has been investigated under selective conditions for *B. subtilis* (18,25), *Streptococcus pneumoniae* (13,14) and *Haemophilus influenzae* (26). In these studies, the probability of detecting replacement of a specific gene was dependent on local sequence identity and on its fitness effects. Their disentanglement requires extensive modeling (25). Under minimally selective conditions, the distribution of fitness effects of transformation has been investigated (27), but the underlying genome dynamics were not systematically studied.

Here, we develop a replacement accumulation assay under minimal selection to investigate how the genomic distance between donor and recipient affects genome dynamics. We reveal that the exponential relation between transfer rate and sequence divergence holds for orthologous replacement within the core genomes of *Bacillus* species. Furthermore, we show that the dynamics of the accessory genomes are linked to the rates of orthologous replacement. By pooling the data from a large set of transformation experiments, we investigate cold spots of orthologous replacement. For two closely related *Bacillus* species we demonstrate that nearly the entire core genome is accessible to replacement. Taken together, our work contributes to understanding the barriers to horizontal gene transfer at the level of the entire genome.

Materials and methods

Strains, media and cultivation

We derive the recipient Bs166 strain (*his leu met, amyE::PhscomK(spc), comK::kan,* P_{comK}*gfp*) (25) from *B. subtilis* BD630. In this strain, the master regulator for competence *comK* is deleted and placed under an IPTG-inducible promoter into the *amyE* locus. As donor species, we use *B. spizizenii* 2A9, *Bacillus vallismortis* DSM 11031, *B. atrophaeus* 11A3, *Geobacillus thermoglucosidasius* 2542 wild types obtained from the BGSC and DSMZ. For the competition experiments, a GFP reporter strain (Bs175) was created based on Bs166 (*his leu met, amyE::PhscomK(spc), comK::kan, lacA::PrrnE-gfp (erm)*) (25).

In the replacement accumulation experiment and competition assay, cells were cultivated at 37°C in competence medium (CM) (28) which was based on Spizizen's salts (6 g/l KH_2PO_4, 14 g/l K_2HPO_4, 2 g/l $(NH_4)_2SO_4$, 1 g/l tri-sodium citrate dihydrate) to which we added 0.5% D-glucose, 50 μg/ml L-histidine, L-leucine and L-methionine, 0.02% casamino acids, 0.1% yeast extract, and 0.5 mg/ml $MgCl{\cdot}6H_2O$. Liquid cultures were incubated at 37°C and a shaking frequency of 250 rpm. The Infinite M200 plate reader (Tecan, Männedorf, Switzerland) was used to monitor bacterial growth. For picking single clones (single cell bottlenecking), bacteria were grown on solidified plates of lysogeny broth (LB) supplemented with 1.5% agar. Bacteria were inoculated onto the plate and spread by adding small glass beads and shaking for about 15 s.

Genomic DNA needed for the replacement accumulation experiment and for whole genome sequencing was isolated using the DNeasy Blood & Tissue Kit by Qiagen. This yielded DNA fragments with varying lengths, dominated by approximately 30 kb long fragments.

Replacement accumulation experiment

In the replacement accumulation experiment, we perform a separate experiment over 20 independent transformation cycles for each of four donor species. For the first cycle of each experiment, we plate Bs166 recipient cells on an LB agar plate, incubate overnight, and pick 8 colonies for the hybrid lineages and 4 as controls. Cells are resuspended in CM in separate wells on a microtiter plate and grown for 2.5 h at 37°C. We add genomic DNA from the donor species (∼1 genome equivalent per recipient cell) together with IPTG to induce competence. Cells take up DNA for 2 h and integrate it into their genome, becoming so-called hybrids. The transformation duration mimics the duration of the competent state in wt *B. subtilis* (29). We end the DNA uptake process by diluting the cells and immediately plating on LB agar plates. Plates are incubated overnight and one random colony is picked from each plate to start the next transformation cycle. For the control samples, we perform the same experiment but without adding donor DNA. Every other day, samples are frozen as backups and for later sequencing. 20 transformation cycles are performed consecutively and hybrid strains are whole genome sequenced (Illumina HiSeq, Eurofins) after cycle 10 and 20. We aim at sequencing monoclonal samples to resolve single hybrid lineages. Colonies picked from the plates are not necessarily monoclonal as heteroduplexes formed through transformation are only resolved after plating through cell division (9). For cycle 10 hybrids, detected recombination events pass through an additional filter in the analysis pipeline (look-ahead filter). This filter only retains changes that are also present at a later sequenced time point. For cycle 20 samples, hybrids are plated and picked a second time. The hybrid strains are named according to the supplemented donor DNA. The final set of whole genome sequences comprises eight independent replicates of each BspizHyb, BvalHyb, GeoHyb and seven independent replicates of BatroHyb. Additionally, we analyse seven representative replicates from the control experiments.

In this experiment, we apply minimal selective pressure as the different hybrids do not compete at any stage. During the transformation step, competent cells are growth arrested (30) and they are plated immediately after the transformation step. From the plates, one single colony is picked randomly and only non-viable variants are lost from the hybrid pool. This

experimental design allows us to accumulate unbiased transformation results that represent the general transformation process nearly independently of selection. We note, however, that hybrids that do not form colonies overnight are not presented in this study.

Detection of genomic replacements through orthologous replacement

For all strains, nucleotide sequences are obtained from NCBI database in fasta-format. For the recipient Bs166 we use NC_000964.3, and for the donors we use NC_014479.1 (*B. spizizenii*), NZ_CP026362.1 (*B. vallismortis*), NC_014639.1 (*B. atrophaeus*), NZ_CP012712.1 (*G. thermoglucosidasius*). Additionally, for the *B. subtilis* recipient, the gene annotation was downloaded in gff3-format. Genes with non-unique locus-tags were removed, leaving 4534 annotated genes. The reference fasta was adjusted to the lab strain by replacing differing SNPs and Indels.

We isolate genomic DNA from all hybrid samples, the recipient and donor species, and obtain whole genome paired-end sequencing reads with a length of 150 bp and an average coverage of 400×. We analyse the data by performing quality control with FastQC (v0.11.7, (31), http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), trimming with Trimmomatic (v0.36, (32)) and mapping to the appropriate reference with the Burrows-Wheeler alignment tool bwa mem (v0.7.17, (33)). Data is processed with the samtools' sort function (v1.16, (34)) and variant calling is done with the bcftools' mpileup and call function (v1.16, (35)). Variants are filtered by a read depth and base quality of at least 50. This pipeline is mostly adapted from the work of Power *et al.* (25).

With the pipeline, we call variants of the hybrids and donor species against the Bs166 recipient reference. By comparing the single nucleotide changes in the hybrids to the changes between donor and recipient, we detect orthologous replacement as clusters of donor nucleotide alleles (25). These clusters consist of at least two donor alleles and end when five or more consecutively missing donor alleles are detected. Here, we leave out multi-mapping regions to which reads map inconclusively. We detect these regions beforehand by using Blast (blastn) (36) to align the Bs166 reference genome to itself. We find 10 regions with a total length of about 50 kbp, containing rRNA and tRNA loci. For obtaining the length of the replaced segments, the recombination clusters are extended by half the distance to the next unchanged recipient allele on both sides of the segment. By doing so, we account for the fact that segment integration might initiate at completely identical stretches in which we cannot detect replacement. For each segment, the identity is computed as the fraction of alleles that are unchanged between recipient and donor. For the distribution of sequence identity of the replaced segments, we determine the 95% confidence intervals of the mean by performing a bootstrap analysis with 10^4 resampled data sets.

The presence of fully identical stretches prevents us from detecting the exact start and end positions of replacements and thus influences the measurement of segment length and identity. The deviation is strongest for short segments and we thus exclude segments below 100 bp from length and identity analyses.

Detecting replaced genes and their statistical overrepresentation

For all replaced segments, we use the start and end positions of genes from the annotation of the recipient to detect the affected genes. These gene lists are used to identify overrepresented functional categories with PANTHER GO (37,38). Here, *Bacillus subtilis* is selected as organism and we perform a statistical overrepresentation test with a Fisher's Exact test and a Bonferroni correction on the PANTHER protein classes.

Accessory genome and core genome identity

In this study, the genomic distance for each pair of recipient and donor species is quantified by the fraction of shared core genome, the sequence identity within this core genome, and the complimentary accessory genome. These definitions are different from those used in pan-genome analyses, where many genome sequences are compared from strains of the same species (39,40). There, genes shared by all strains make up the core genome, and genes only present in a few strains form the accessory genome (39). Here, we are interested in pairwise comparisons of recipient and donor and, therefore, the accessory genome refers to pairs of species only.

We detect accessory regions that are unique to the recipient on the basis of our pipeline as follows. First, donor sequencing reads are aligned to the Bs166 reference. Then, we detect regions of at least 150 bp length in which the read depth is <50. The remaining genomic regions are defined as the core genome in which sequences between donor and recipient are similar and differ only by SNPs. Orthologous replacement can only be detected in the recipient's core genome. We define accessory genes as those genes that are fully accessory and thus inaccessible to orthologous replacement. Complementary to this, core genes are those that can be affected by recombination and that at least partially contain core genome.

We evaluate the sequence identity between the recipient and donor's core genome by calling the variants on the whole core genomes and calculating the percentage of similar single nucleotides.

Detection of deletions, insertions and mixed events

We detect deletions in the hybrids by extracting the coverage with which reads map to the reference with the bedtools2's function genomeCoverageBed (41). A deletion is detected when the local coverage drops to a coverage of 0 over at least 10 consecutive positions.

Additionally, we are able to detect inserted segments from the donor's accessory genome in the hybrids. These insertions are facilitated by homologous flanking regions (42). We align the hybrid reads to the donor genome and identify insertions longer than 150 bp through increased coverage in accessory regions.

Deletions of recipient's accessory parts, insertions of donor's accessory regions, and replacements can occur alongside each other in the same, mixed integration event. To detect mixed events, the direct proximity of individual events is investigated. For deletions and replacements this is simple, because they are detected with respect to the recipient's genome. As this is not the case for insertions, we first align the insertion's homologous flanking regions from the donor to the recipient's genome. This reveals the most probable entry point for the insertion. Finally, all events are combined, their positioning is analysed and mixed events are identified. In a second

approach to connect replacements with insertions and deletions, we align the complete donor genome to the recipient using the Blast algorithm (blastn) and visualize the similarities in so-called dot plots (Figure S1). Adding the regions of the detected replacement and deletion events on the x-axis (recipient genome), and of insertions on the y-axis (donor genome), we are able to identify mixed events that occurred together due to the positioning and architecture of core and accessory regions on the genomes.

Detection and analysis of mosaic recombination events and length distributions

We found that multiple orthologously replaced segments are located in very close proximity. This phenomenon has been previously described for other species (13,14) and here we follow a similar approach to show that the clustered replacements most likely did not occur independently. Then we determine a distance threshold between two consecutive replacements beyond which we assume that the recombination events have occurred independently of each other. Finally, to determine the length distribution of replaced segments, we merge the clustered segments.

We derived the distances between neighboring clusters for the BspizHyb, BvalHyb and BatroHyb samples after 10 and 20 cycles. Here, we determine the transferred lengths by assuming clusters to start and end at the next neighboring missing donor allele (maximum possible extension). On a logarithmic scale, the distance distribution reveals a bimodal behavior (short-distance mode, long-distance mode) for BspizHyb and BvalHyb (Figure 5A, B). For BatroHyb there is not enough data to evaluate bimodality.

We hypothesized that the short-distance mode represents mosaic replacements that occur together. To dismiss the possibility that these replacement events are independent, we investigate a Monte Carlo null model, as proposed previously by Croucher et al. and others (13,14). We generate a set of distances between replaced segments from different hybrid strains to ensure that the events have occurred independently of each other. Specifically, for each donor species and time point, we pool the shortest distances between two events in independent hybrids. From this pool, we collect 10^4 bootstrap samples that each have the same size as the experimentally detected distance data set, estimate the probability density using a log-normal kernel (to account for the logarithmic x-axis), and count the number of peaks of the probability density. We observe that the distribution of independent distances only rarely shows bimodal behaviour (Figure S2). By contrast, using the empirical data, most bootstrap runs show bimodal behavior.

Then, we perform the analysis to detect the location of the local minimum in the experimental distributions. Again, we use a bootstrapping approach, but this time we draw from the empirical data, perform a kernel density estimation with a log-normal kernel on 10^4 bootstrap samples, where each sample has the same size as the original data set. For all distributions that are found to be bimodal (for fractions compare Figure S2), we collect the position of the local minimum between them (Figure S3). The minimum is used as a threshold to distinguish the short- and long-distance regime. These thresholds are 1340 bp for BspizHyb, 1940 bp for BvalHyb and 10 935 bp for BatroHyb. Segments from the short-distance regime are subsequently merged to derive the corrected length distribution of (independent) replacement events.

The length distribution of the recombination events is assumed to follow an exponential probability distribution, described by the single parameter μ, as probability density $p = \mu * exp(-\mu x)$. Here, we exclude lengths below 100 bp and obtain μ and the error of the fit by fitting the cumulative probability distribution. With μ, we determine the characteristic length, at which the probability is reduced to half of its initial value.

Obtaining the transfer rates and fitting the exponential relation to the core genome sequence identity

For every transformation hybrid, we fit a linear model to the trajectory of the replaced core genome over time. This is done with a least square fit on a linear model with the intercept set to 0. We test how well the linear model fits the data with the R^2 quantity that reflects the goodness of fit. For BspizHyb, R^2 is close to 1 for 4 trajectories and exceeds 0.85 for the remaining samples, except one outlier. For BvalHyb, the value exceeds 0.9 for 6 samples and 0.74 for the remaining two. For BatroHyb, R^2 cannot be determined, because too many trajectories have values at 0. For each species, we average the slopes of the trajectories. The resulting quantity is called the species transfer rate. Transfer rates r are fitted against the sequence divergence (1-identity) to the exponential relation $r(x) = Aexp(B \cdot x)$, where the standard deviation serves as weights and A and B are fit parameters. This is done using the scipy.optimize.curve_fit function from python. We obtain the fit parameters, their standard deviation, and the R^2 value with and without considering the weights.

Detecting replacement cold spots with pooled data sets

In order to identify genes in Bs166 that have never been replaced by the *B. spizizenii* donor, we draw on additional data sets from hybrids created in different transformation experiments. In total, we use 42 sequenced hybrid organisms. We extend the 8 samples from this study by 12 additional samples created in 20 cycles of the same replacement accumulation experiment. An additional 7 *B. spizizenii* hybrids are included from Power et al. (25). Here, a transformation assay was performed over 21 cycles in which cells were exposed to ultraviolet (UV) radiation on the first day, plated onto agar plates and then transformed with donor DNA on the second day. After this, cells were grown to stationary phase overnight (about 18 h) and plated for single-cell bottlenecking on the next day. The pool is expanded by another 15 hybrids created in 20 cycles of a transformation assay that is similar to the one used in Power *et al.*, but which leaves out the UV radiation step.

For all hybrids, genome replacements are detected as described before. In order to identify gene cold spots, we combine replaced segments from all pooled samples and detect genes that were never replaced. In this analysis, we exclude all recipient genes that contain any accessory or multi-mapping parts, which leaves 3806 core genes. Additionally, we ignore completely identical genes. The length distribution of the cold spots is exponential: $P_{CS}(x;\ \lambda_{CS}) = C \cdot e^{-\lambda_{CS}x}$, where C is a normalization constant and $1/\lambda_{CS} = 952.4$ bp is determined by fitting the empirical data. Following the approach de-

scribed in Power *et al.* (25), the significance of outliers is evaluated from the Gumbel distribution (with parameters 5030, 954) by drawing 196 samples from $f_{CS}(x;\ \lambda_{CS})$ corresponding to the number of empirical cold regions.

High-throughput competition assay

We measure the relative fitness of the hybrids compared to the Bs166 recipient by performing a recently established high-throughput competition assay (27). This setup allows us to characterize all transformation hybrids in one experimental run and we repeat the experiment on four different days. First, overnight cultures of hybrid strains and the *gfp*-expressing reporter strain (Bs175) are diluted into fresh CM and grown for 2.5 h at 37°C. After entering the exponential growth phase, the hybrids are mixed in a 1:1 ratio with Bs175 and grown in competition to each other for 4 hours (∼14 generations (27)). With a flow cytometer, we measure the fraction of reporter strain x_{RS} and hybrid strain x_i before and after competition, at time points t_0 and t. Fitness is characterized as selection coefficient s with

$$s_{i,RS} = \frac{t_g}{t - t_0} ln\left(\frac{x_i\,(t)\,/x_{RS}\,(t)}{x_i\,(t_0)\,/x_{RS}\,(t_0)}\right),$$

where t_g is the generation time of the recipient. By measuring s of the recipient Bs166 in each run, we can evaluate the selection coefficient of hybrids with respect to the recipient as $s_{i,Bs166}$. In total, we measure s for eight BspizHyb strains, seven BvalHyb strains and five BatroHyb strains. Additionally, we perform the assay on eight control strains that obtained no donor DNA. We use a recently published recipient reference distribution that represents the resolution of the competition assay (27). The data set was obtained by measuring the selection coefficient of 80 ancestor replicates on four different days. The resulting distribution is centered around $s = 10^{-4}$ and has a standard deviation of $\sigma = 0.0031$.

Results

The replacement accumulation experiment is fitness neutral

We sought to find out how genomic DNA from different donor species introduces genetic variations to transforming *B. subtilis*. To this end, we set up a replacement accumulation experiment, using *B. subtilis* (Bs166) as recipient. As donor species, we chose two closely related species, *B. spizizenii* and *B. vallismortis*, one more distantly related *Bacillus* species, *B. atrophaeus* (43), and one species belonging to a different genus, *G. thermoglucosidasius* (Figure 1A). The fraction of the core genome and its sequence identity are used as a measure of the genetic distance between recipient and donor species. In this study, the core and accessory genomes are determined by pair-wise comparison between the recipient genome and each of the donor genomes. Specifically, we determine the fractions of core genomes by aligning donor sequencing reads to the recipient reference and detecting regions where alignment occurs (Materials and Methods). Within these aligning regions, the sequence identity is the fraction of identical base pairs. The remaining accessory genome is unique to each of the species (Fig S1, S6). For the four donor-recipient pairs, the fractions of core genome are 87% for *B. spizizenii*, 86% for *B. vallismortis*, 75% for *B. atrophaeus*, and 11% for *G. thermoglucosidasius* (Figure 1A). The sequence identities of the core genomes range from 93% to 80%. In the following, these distance measures will allow us to detect species-specific differences in the transfer characteristics.

The laboratory strain *B. subtilis* 168 develops competence only within a subpopulation comprising up to 10% of the population. We use a *comK*-inducible strain such that the entire population develops competence upon induction (15). We employ a replacement accumulation experiment (Figure 1B, C), because the total fraction of donor sequence in the hybrids is low after a single cycle of transformation (27). This experiment allows us to scale up the fraction of genome being replaced, which is particularly important for distant donor species. The set of hybrids created in the different experiments are called according to the donor organism, BspizHyb, BvalHyb, BatroHyb and GeoHyb (Figure 1C). After cycles 10 and 20, the hybrids are subjected to whole genome sequencing.

In this experiment, we apply minimal selective pressure as the different hybrids do not compete at any stage. To verify that the setup indeed is fitness neutral on average, we measured the selection coefficient s of hybrids after cycle 10 and 20 with a high-throughput competition assay (Materials and methods, Table S1) (27). In this analysis, only strains that contained at least one transfer event were included. As described in the next chapter, none of the GeoHyb strains and not all of the BatroHyb strains showed transfer. The selection coefficient s describes the difference of growth rates per generation between the hybrid strain and the recipient Bs166. For all sets of hybrids, we compare the mean fitness to a recently published recipient reference distribution ($n = 80$) (27). We find that the mean values do not significantly differ from the reference fitness (Figure 2, Welch's t-test with a P-value of 5%) at either time point. Inspecting individual hybrid lineages (Figure S4 A, B), we also find no net increase in fitness over time. Nevertheless, at the level of individual hybrid strains, transformation shows beneficial or deleterious fitness changes in some of the strains. We identify large fitness effects for hybrids by testing against the recipient reference distribution using a Z-test and Bonferroni correction. The number of hybrids showing a strong fitness effect increases from cycle 10 to 20. Overall, transformation causes the fitness distributions to broaden for later time points and higher donor sequence identity, including the BspizHyb and BvalHyb samples (Figure 2). There is no obvious correlation between the selection coefficients and the fractions of replaced genomes (Figure S5). Taken together, we find no net increase in mean fitness of the hybrid strains, indicating that selection is negligible, and we can study the effects of transformation on genome dynamics independent of selective effects. We note, however, that non-viable hybrids or hybrids that grow too slowly to form a visible colony overnight are ignored in this study.

The replacement probability clusters locally for distantly related *Bacillus* species but not for closely related species

We analyze the outcome of the 20 cycles of the replacement accumulation experiment by detecting three different types of genomic variations (Figure 3) (25,27). Most frequently, a segment of the recipient's genome is replaced by the orthologous segment from the donor species. This process is called orthologous replacement and can only be detected within the core genome. Additionally, insertions of segments of the donor's

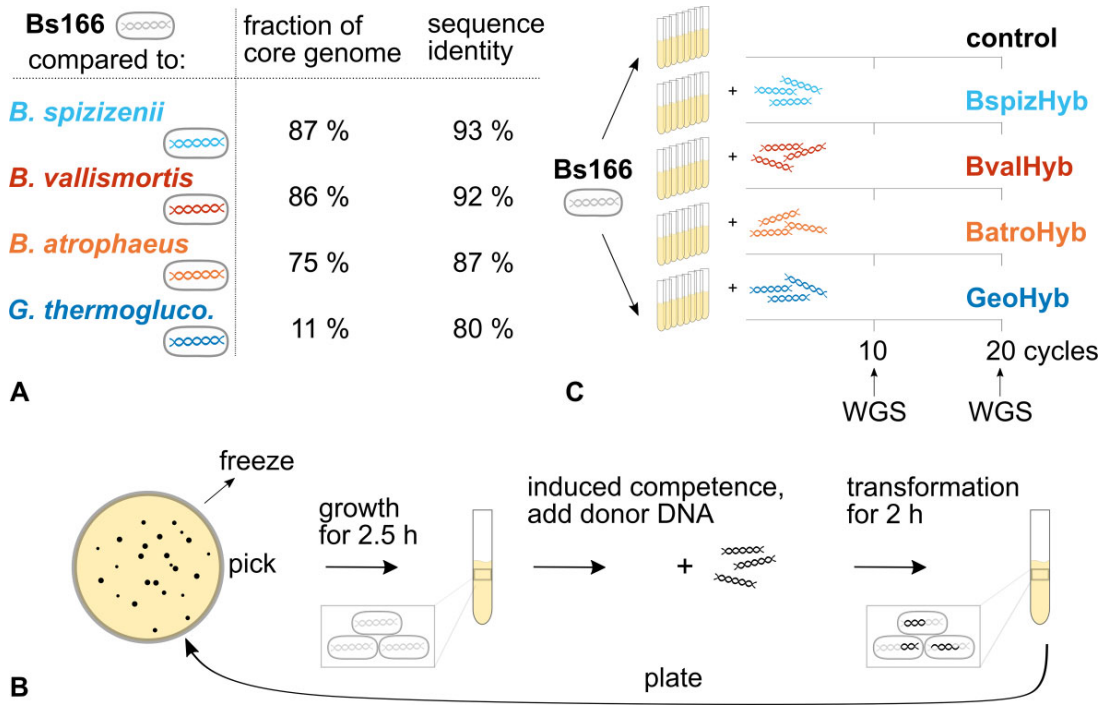


Figure 1. Replacement accumulation experiment performed with four donor species *B. spizizenii*, *B. vallismortis*, *B. atrophaeus* and *G. thermoglucosidasius*. (**A**) Genomic distance between recipient Bs166 and donor species is measured as fraction of core genome and its sequence identity. (**B**) In each cycle of the experiment, one random colony is picked from an agar plate, cells are grown to exponential growth phase, transformed for 2 h with genomic DNA from a donor species, and plated overnight. (**C**) For each donor species, the experiment is performed for eight replicates in parallel. Additionally, we run a control experiment where cells do not transform with donor DNA. Starting with the recipient Bs166, single transformation cycles are performed consecutively.

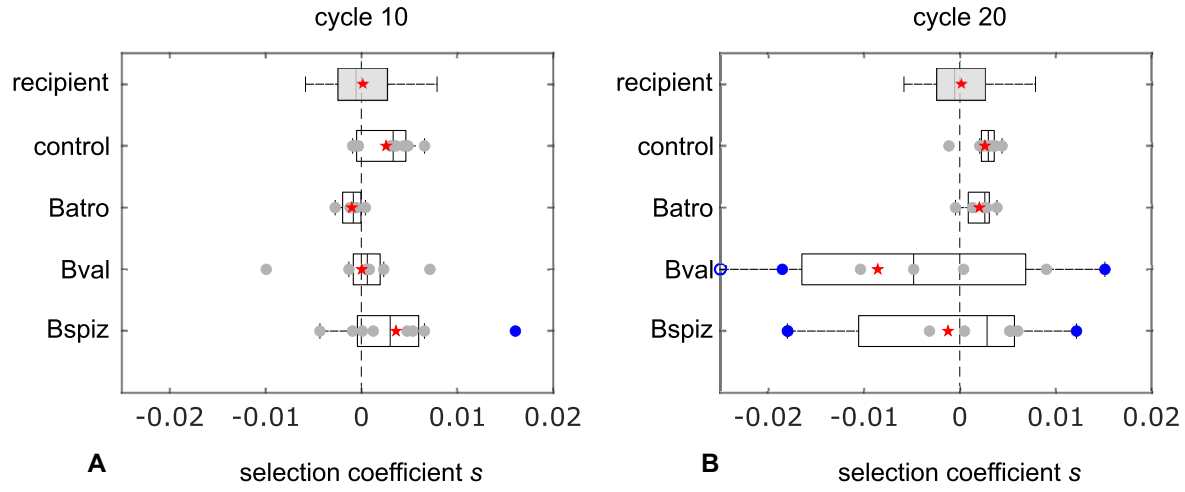


Figure 2. Replacement accumulation does not increase the mean fitness of the hybrid strains. Selection coefficient *s* is shown for hybrids from (**A**) cycle 10 and (**B**) cycle 20. Only strains that contain least one transfer event are shown. The recipient reference distribution ($n = 80$) (gray box) represents the resolution of the method and is taken from a recent publication (27). For each condition, the box plot with the median (vertical line) and the mean is shown (red star) along with the individual samples (gray circles). The box depicts the first to third quartile of the data. Hybrids with significantly strong effects (blue closed circle) when compared to the recipient's distribution are highlighted. For BvalHyb at cycle 20, the extreme negative outlier is depicted on the outer edge of the plot (empty blue circle).

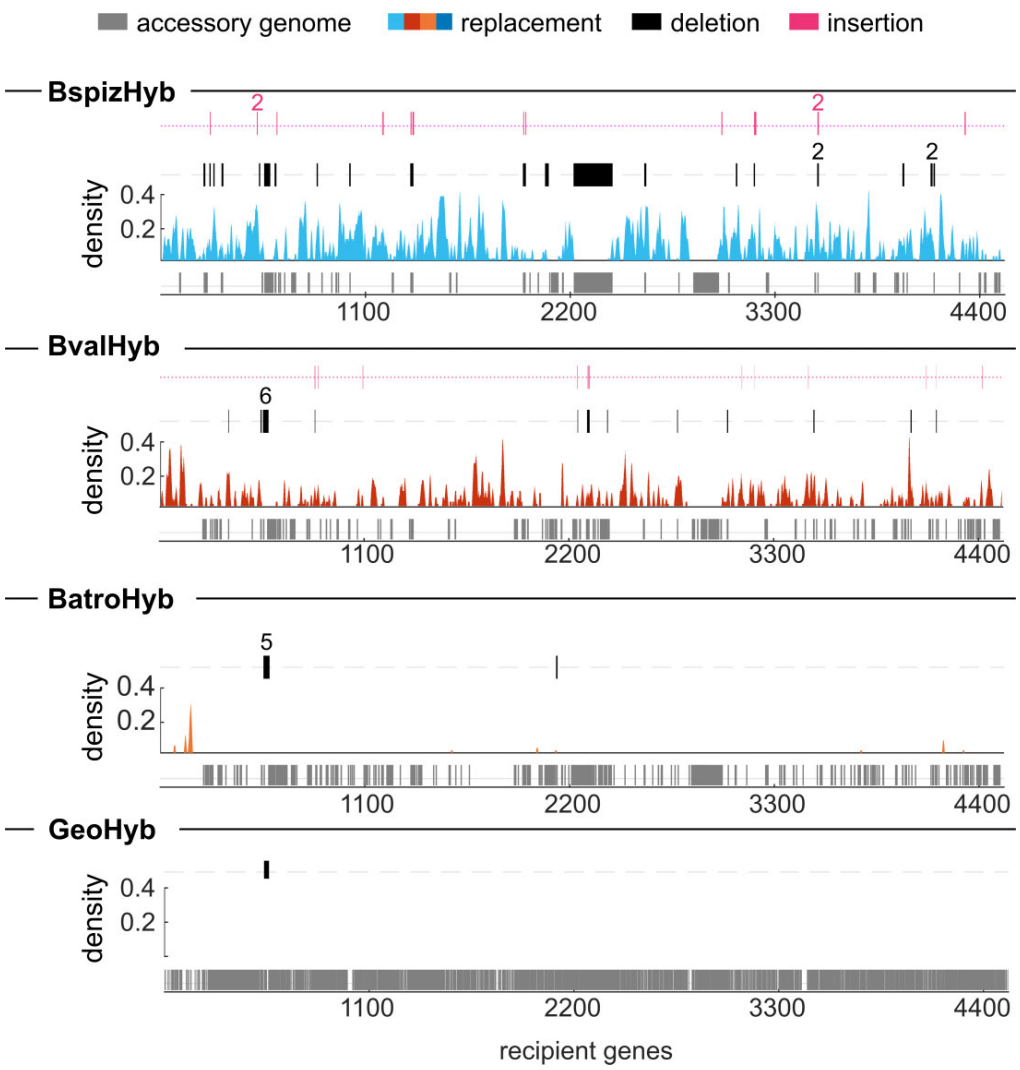


Figure 3. Genomic variations in transformation hybrids depend on the sequence identity between donor and recipient. For all transformation hybrids after 20 cycles, orthologous replacements, deletions and insertions are detected in the whole genome sequencing data. Across all recipient genes, the recipient's accessory genes for each donor species are shown in the lowest line (gray bar). The density of genes affected by orthologous replacement is calculated as the average of replaced genes in a sliding window of 10 genes over all replicates. Genes deleted (black bar) or inserted (magenta bar) in at least one hybrid strain are shown, whereby numbers indicate multiple hits.

accessory genome as well as deletions of segments of the recipient's genome are detected.

For BspizHyb and BvalHyb, we find that orthologous replacement occurs in abundance along the whole genome, excluding accessory genes (Figure 3, Figure S6). Using PANTHER GO, we find no class of replaced genes to be overrepresented. For BatroHyb, very few replacement events are detected in five out of eight strains. Most of these replacements occur close to the origin of replication, where the sequence identity is particularly high (Figure S7). Ribosomal genes are overrepresented among the group of replaced genes at a 5% significance level (using PANTHER GO with Fisher's Exact test and Bonferroni correction). No orthologous replacement was detected for GeoHyb. The seven control strains that did not receive donor DNA showed no orthologous replacement. Deletions and insertions occur, but their abundance is considerably lower compared to the abundance of orthologous replacement and decreases with increasing distance between donor and recipient (Figure 3). Since most of the insertions and deletions are related to the dynamics of the accessory genomes, we will describe them in detail in a separate paragraph.

In summary, we find that orthologous replacement dominates horizontal gene transfer between different *Bacillus* species. While we observe genome-wide replacements when *B. spizizenii* and *B. vallismorts* are used as donors, for *B. atrophaeus* replacement clusters close to the origin of replication where ribosomal gene are located.

For distantly related species, segments with a higher sequence identity than the genome-wide average are preferably transferred

We address the question whether the measured sequence identities of replaced segments reflect the average sequence identity between the donor and the recipient's core genome. To

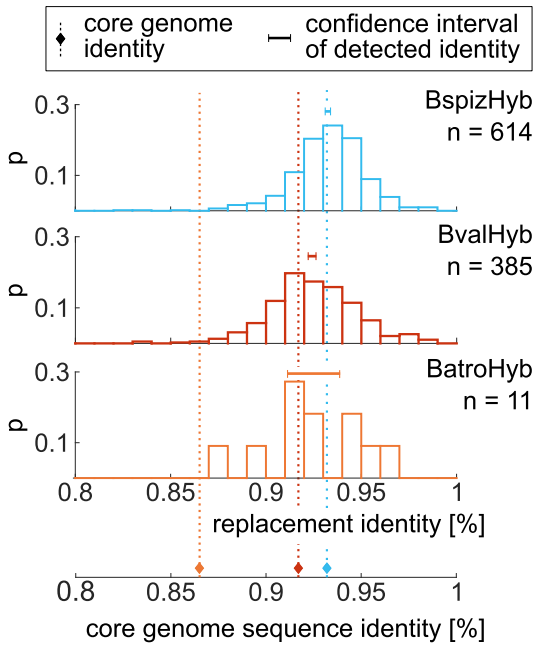


Figure 4. Replaced segments have a higher identity than the average core genome. Normalized probability distributions of the identity of transferred segments are shown for the cycle 20 transformation hybrids BspizHyb, BvalHyb, and BatroHyb. Only segments longer than 100 bp are taken into consideration and the sample size *n* differs strongly between the data sets. For each distribution, the 95% confidence interval for the mean identity is shown (obtained with bootstrap analysis, Methods, Figure S8). On the bottom, the core genome sequence identity is depicted (diamonds, dotted lines).

investigate this, we obtain the identity distributions of the orthologously replaced segments after cycle 20 and evaluate the confidence intervals for the mean with a bootstrap analysis (Figure S8). We then compare the measured identity to the expected core genome sequence identity (Figure 4). For BspizHyb strains, the replaced segments have an identity of 93.3% [93.1%, 93.4%] and are thus in accordance with the donor's identity of 93.2%. The numbers in brackets denote the confidence interval of the mean. For BvalHyb, the segment identity of 92.5% [92.2%, 92.7%] is higher than the donor's identity of 91.7%. Thus, segments with a sequence identity above average are preferentially replaced. This shift between the identity of replaced segments of 92.5% [91.1%, 93.9%] and the core genome identity of 86.5% is even more pronounced for BatroHyb. Examining segments across all samples, we find the minimum detected identity to be around 80%.

We note that we cannot detect recombination taking place in regions with 100% identity between donor and recipient. However, most completely identical regions are short and replacements detected in our analysis bridge them. Additionally, completely identical regions cause an uncertainty for the localization of recombination start and end sites. As the relative uncertainty is larger for short segments, we exclude those below 100 bp (Materials and methods) in the analysis. Nevertheless, we expect to slightly underestimate the average sequence identity of the replaced segments in Figure 4.

In conclusion, for all three donors studied, we find that replaced segments have an average sequence identity of around 93%. The identity of replaced segments is higher than the core genome sequence identity for the more distantly related donors.

Transfer lengths are longer for donors with a higher sequence identity

In the next step, we investigate whether the donor species influences the length distribution of transferred segments. When analyzing individual orthologous replacements, we notice some to occur in very close proximity to one another. This is most conspicuous for samples with very few replacements that occur close by. To elucidate this phenomenon, we investigate the distances between neighboring replacements (Figure 5A). On a logarithmic x-scale, we find a bimodal distribution of transfer distances for BspizHyb and BvalHyb after 20 cycles of recombination (Figure 5B). The long-distance mode contains more data points than the short-distance mode. Due to the small sample size, the effect is not detectable for BatroHyb. For the two regimes, we assume the large one to contain independently occurring events. We show that the events corresponding to the short-length mode most likely have not occurred independently (Methods, Figure S2), indicating that these imports originate from single recombination events that create a mosaic pattern between donor and recipient sequences, as described before (10,11,13). For each donor, 25–30% of all detected replacement events are mosaic events (BspizHyb: 29%, BvalHyb: 25%, BatroHyb: 30%). This suggests that the occurrence of mosaics is independent of donor-recipient identities. Most mosaic events contain two replaced segments, and a few have more, ranging up to a maximum of eight segments in one event.

Here, we are interested in the length distribution of full recombination events and not in the effects of additional processes that act upon recombination. Thus, we merge the mosaic segments that fall underneath a cut-off value in the distance distribution (Figure 5B, Figure S3, Materials and methods). The length distribution agrees with an exponential decay as a function of segment length (Figure 5C). The characteristic transfer length increases with increasing donor sequence identity (Figure 5D). Qualitatively, we find the same relation in the uncorrected length distribution, where mosaic replacements are considered as independent replacement events (Figure S9). This exponential distribution is expected, if we assume that the length of a segment is determined by the rate of RecA filament elongation and a constant probability of termination (11,44). Similar distributions have been reported for intra-species recombination between different strains of *H. influenzae*, *S. pneumoniae*, *N. meningitidis* (13,45,46). We find that with increasing sequence divergence, the characteristic length of the replaced segments decreases (Figure 5D). It has been suggested previously, that recombination is aborted at short stretches of heterology (47). The density of these stretches increases as the mean sequence divergence increases and if we average over the entire genome, the net abortion rate increases. We also note that the probability density of segments shorter than 100 bp deviates from the exponential relation. It has been suggested earlier that these short segments are integrated by an alternative mechanism (17,48).

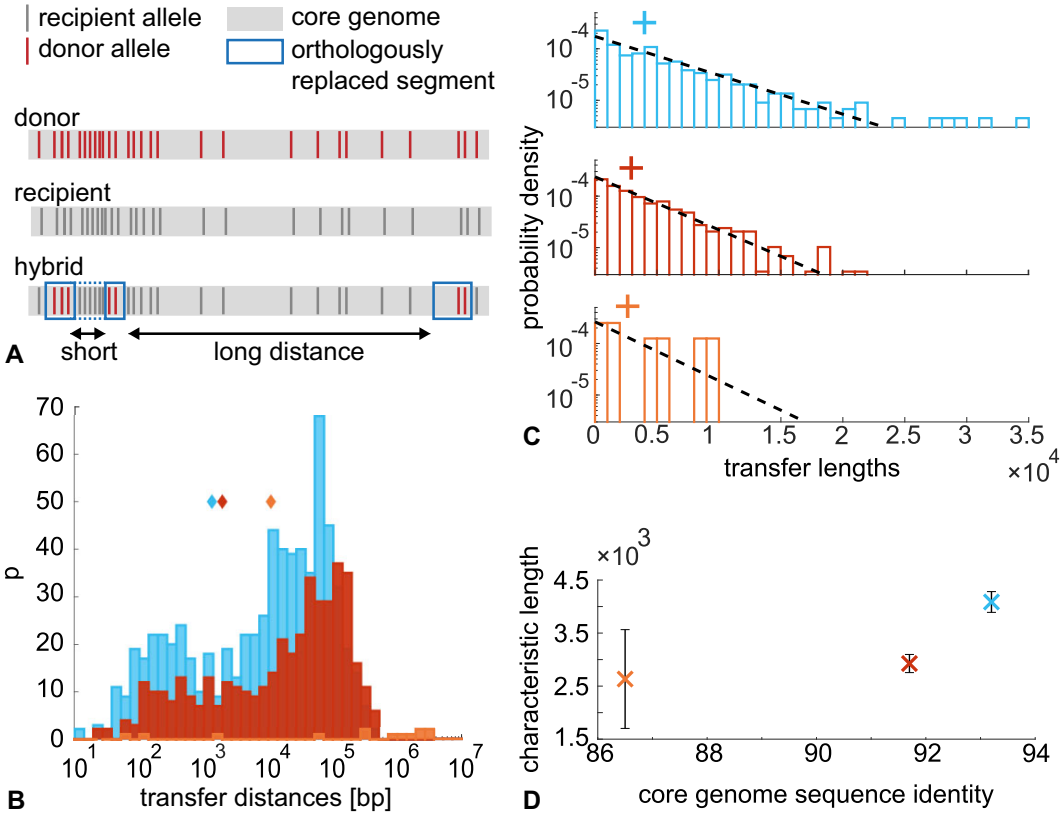


Figure 5. The characteristic transfer length of recombination events increases with increasing core genome sequence identity. (**A**) Scheme of mosaic pattern of orthologous replacement. (**B**) The distribution of distances between replacements, where the cut-off between short and long-distance regime is detected (diamond) using a kernel density estimation (Materials and methods, Figure S3). (**C**) The length distribution of merged events is fitted with an exponential probability distribution (dashed line) for events longer than 100 bp. The characteristic length is determined (cross) which increases (**D**) with increasing core genome sequence identity (error bars depict the standard deviation). Cyan: BspizHyb, red: BvalHyb, orange: BatroHyb.

Gene transfer rate decreases exponentially with core sequence divergence

We address the question how the rate of orthologous replacement (aka transfer rate) depends on core genome divergence. The transfer rate is measured using our time-resolved sequencing data of the replacement accumulation experiments. For each hybrid lineage at cycle 10 and 20 (Figure 1C), we calculate the fraction of core genome replaced by recombination (excluding GeoHyb, where no transfer was detected). Here, each cycle includes 2 h of DNA uptake. The depicted hybrid uptake trajectories (Figure 6A) reveal an approximately linear increase over time. Even though transfer rates differ strongly, the linear characteristics of the uptake process are preserved between donor species (Methods). We obtain a transfer rate for each trajectory as the slope from a linear fit.

The species transfer rate r is determined as the mean of the individual rates and the error is evaluated as the standard error. For BspizHyb we find a rate of $r = (0.210 \pm 0.020)\%\ \mathrm{h}^{-1}$. The rate of $r = (0.1000 \pm 0.005)\%\ \mathrm{h}^{-1}$ is slightly lower for BvalHyb, and as low as $r = (0.003 \pm 0.001)\%\ \mathrm{h}^{-1}$ for BatroHyb. Taken together, we find an exponential decrease of transfer rate with increasing sequence divergence (Figure 6B). By fitting this exponential relation (Methods), we detect the exponential decay parameter as $-0.59\%^{-1}$ (standard deviation of $0.11\%^{-1}$).

In summary, hybrids accumulate orthologous replacements linearly over time and the transfer rate decreases exponentially as a function of average sequence divergence within the core genome.

At least 96% of the *B. subtilis* core genes are accessible to orthologous replacement by *B. spizizenii* alleles

We sought to find out whether large portions of the recipient's core genome are inaccessible to orthologous replacement from a specific donor. To this end, we pool sequencing data from 42 hybrid genomes between *B. subtilis* and *B. spizizenii* that we obtained in different transformation experiments (Methods). All of the strains had undergone 20 or 21 cycles of transformation. On average, 12.5% of the 3806 full core genes are affected at least partially by orthologous replacement within each strain (Figure S10A). About half of the strains (20) were transformed under minimal selective conditions as described in Figure 1. The other strains were grown for about 18 hours after each transformation cycle, allowing for competition between the strains as described in the Methods. We expect that some transfers are selected for or against in this set of strains. Thus, we will only determine an upper limit for the fraction of genes that are not accessible to replacement. Also, the selective pressure is not expected to be strong, because the probability

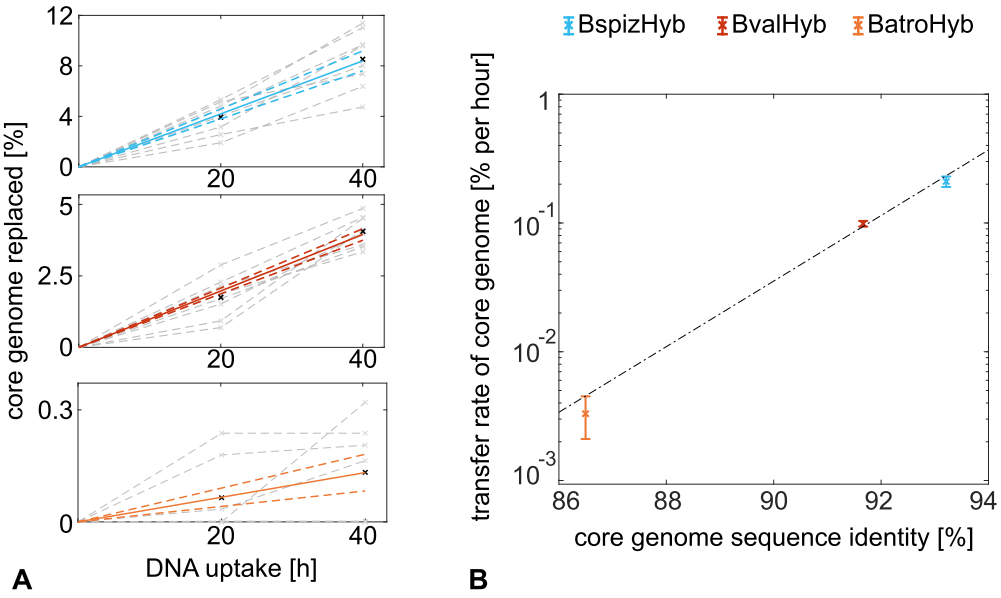


Figure 6. The transfer rate depends exponentially on the average sequence divergence between donor and recipient. (**A**) For different donors in the replacement accumulation experiment, the fraction of core genome replaced increases linearly with hours of DNA uptake. One cycle of transformation takes 2 h (Figure 1B); 40 h of DNA uptake correspond to 20 cycles. Each hybrid's trajectory is shown separately (gray line/cross) together with the mean (black cross). For each hybrid, the data is fitted and the fit mean ± standard error is shown (solid and dashed line). (**B**) Transfer rates decrease exponentially with the sequence divergence of the core genome. Error bars correspond to standard errors. An exponential curve is fitted to the data, yielding an exponential coefficient of $0.59\%^{-1}$ with a standard deviation of $0.11\%^{-1}$ and an intercept of $12.45\%\ h^{-1}$ with a standard deviation of $10.85\%\ h^{-1}$ ($R^2 = 0.974$, weighted $R^2 = 0.999$).

to find genes replaced in a certain number of samples depends only weakly on the selective conditions (Figure S10B).

Out of 3806 genes that entirely belong to the core genome, only 137 were never affected by orthologous replacement. We will call these genes 'putative cold spots' of orthologous replacement in the following. 96.4% of the genes were at least partially replaced and 91.8% were fully replaced in at least one out of 42 strains. At the level of base pairs, 94.5% of base pairs of the core genome were replaced in at least one strain. We note that replacement cannot be detected if the identity of the replaced segment is 100%. However, out of 93 completely identical genes, only 3 are detected as cold spots, indicating that this detection limit has no strong influence on the detected cold spots.

The putative cold spots are distributed throughout the genome (Figure 7). We find that the distribution of sequence identities of the cold spot genes is shifted towards lower identities as compared to the replaced genes (Figure S11). Most cold spots are flanked by accessory parts of the genome (Figure 7, Figure S12), indicating that the probability that a recombination event is initiated is affected by the close proximity to non-homologous regions. Thus, not only the local sequence identity, but also the distribution of the accessory genome affects the replacement probability.

We labeled the cold spot regions as 'putative', because the fraction of replaced genome is not yet saturating in our 42 hybrid strains (Figure S13). We identify one region as a 'genuine' cold spot, because it is a significant outlier from the length distribution of all putative cold spot regions. The length distribution of the putative cold spot regions is exponential with a characteristic length of 660 bp (Figure S14A). From this length distribution, we find one significant outlier with a length of ∼ 12 kb containing 13 genes (Figure S14B). The average sequence identity of this region is 0.95. Even though it is flanked by accessory genome at both sides, the average sequence identity and the length should enable initiation of homologous recombination. Therefore, we interpret this region as a genuine cold spot for replacement.

We conclude that the largest part of the *B. subtilis* core genome is accessible to recombination with the core genome of *B. spizizenii*.

The genome architecture influences the transfer of accessory genome parts

Recombination supports replacement of a core segment by the respective donor part. Additionally, it can also affect the dynamics of the accessory genome depending on its individual architecture (Figure 8A). Accessory genes of the recipient can be purged through recombination initiated between their flanking regions and donor DNA (Figure 8B). Similarly, accessory genes of the donor may be inserted into the recipient (Figure 8C). Finally, accessory recipient genes can be replaced by accessory donor genes and this would be detected as a combined insertion/deletion event (Figure 8D). In the sequenced hybrids, we find examples for all of the three possibilities (Figure S1). Their occurrence correlates with the frequency of orthologous replacement (Figure 8E, F), indicating that they are driven by recombination. Yet they occur at considerably lower frequencies.

Most deletions purge accessory parts from the genome, as is the case for about 87% of deleted genes in BspizHyb and 68% in BvalHyb. For example, in one BspizHyb strain, SPβ prophage comprising about 200 genes on a length of 134 kbp

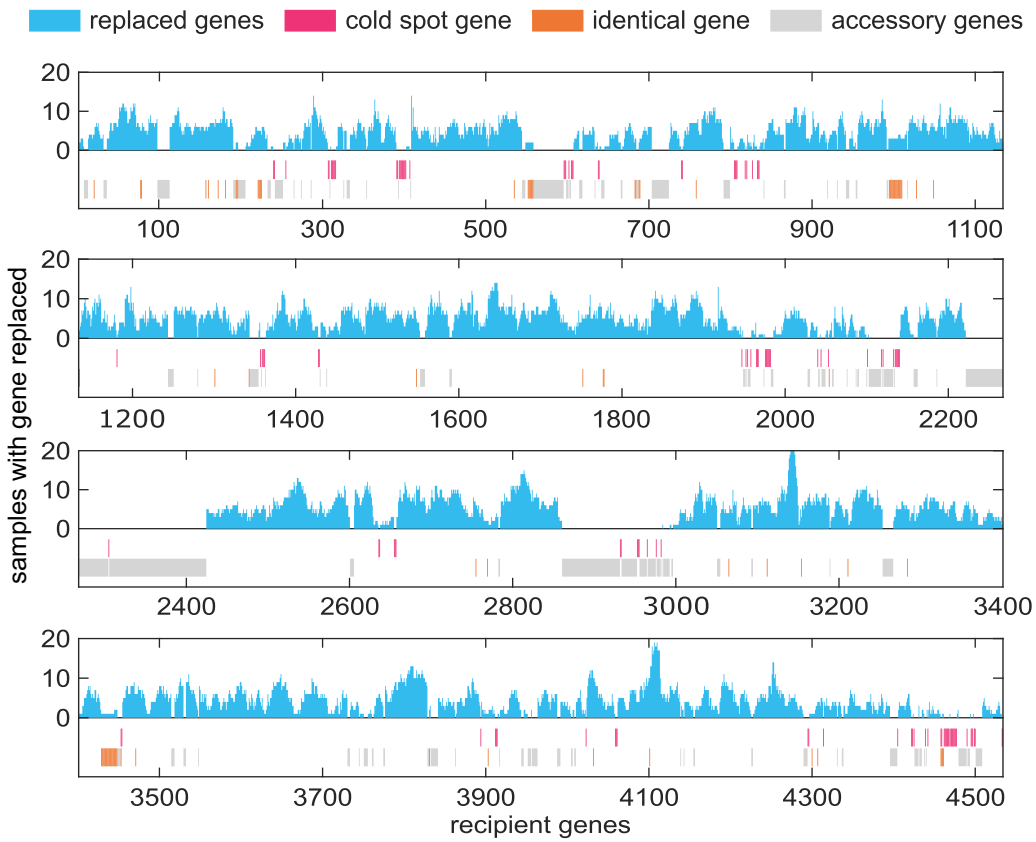


Figure 7. Detection of putative cold spots of recombination with the donor *B. spizizenii* by pooling 42 hybrid strains. Genes that were affected by orthologous replacement in at least 1 out of 42 strains (cyan), genes that were affected in 0 out of 42 strains (pink), fully identical genes (orange), accessory genes (grey).

is fully deleted. This most likely happens through recombination that we detect on the much shorter 17 kbp and 7 kbp long flanking regions. Nevertheless, we also find deletions of non-accessory parts when the donor and recipient genomes have different local arrangements, which is the case for both detected deletions found in BatroHyb. Additionally, the mobile element ICEBs1 is lost from the recipient genome under all conditions, but not in all hybrids (Figure 3, Figure S6). Even in some GeoHyb samples, where no orthologous replacement is observed, we find occasional loss of ICEBs1. The ICEBs1 element comprises 27 genes and excises itself independently of recombination. Therefore, loss of ICEBs1 is not considered in the analysis of deletions in this study. Excision is known to happen during SOS response after DNA damage due to up-regulation of RecA (49,50), which is also up-regulated during the transformation step in our experiment. Notably, the control samples that did not receive donor DNA did not lose ICEBs1, indicating that the process of transformation induces loss of this mobile element.

We detect all insertions from the accessory genome of the donor to happen alongside orthologous replacement (Figure 8F) and their frequency is lower compared to the frequency of deletions (Figure 8E). As the recipient and donor genomes can be rearranged in complicated ways, we find co-occurrence of deletions, insertions and recombination (Methods, Figure S1) in mixed events (Figure 8D). For example, in two BspizHyb samples, the stretch containing *comX*, the pheromone involved for example in quorum sensing (51), is replaced by its non-homologous counterpart. These mixed events are rare, yet they reflect the specific architecture of the genomes involved in the recombination process.

We conclude that the arrangement of core and accessory genome influences the dynamics of the accessory genome through transformation. The frequencies of insertions and deletions are low compared to orthologous replacement.

Discussion

For eukaryotes, mating occurs only between individuals of the same species. This is different for bacteria where genes can be transferred horizontally between species, yet the rates and restrictions of this process are poorly understood at the whole genome level. Here, we employed a replacement accumulation assay for characterizing the effects of sequence divergence on genome-wide transfer between *Bacillus* species. We show that orthologous replacement dominates genome dynamics and that the sequence identity between donor and recipient is crucial for the dynamics of both core and accessory genomes.

Remarkably, nearly the entire core genome of *B. subtilis* is accessible to orthologous replacement by *B. spizizenii* core genes. In the 42 pooled hybrid strains, only 4% of all core genes were never hit by replacement. In the following, we dis-

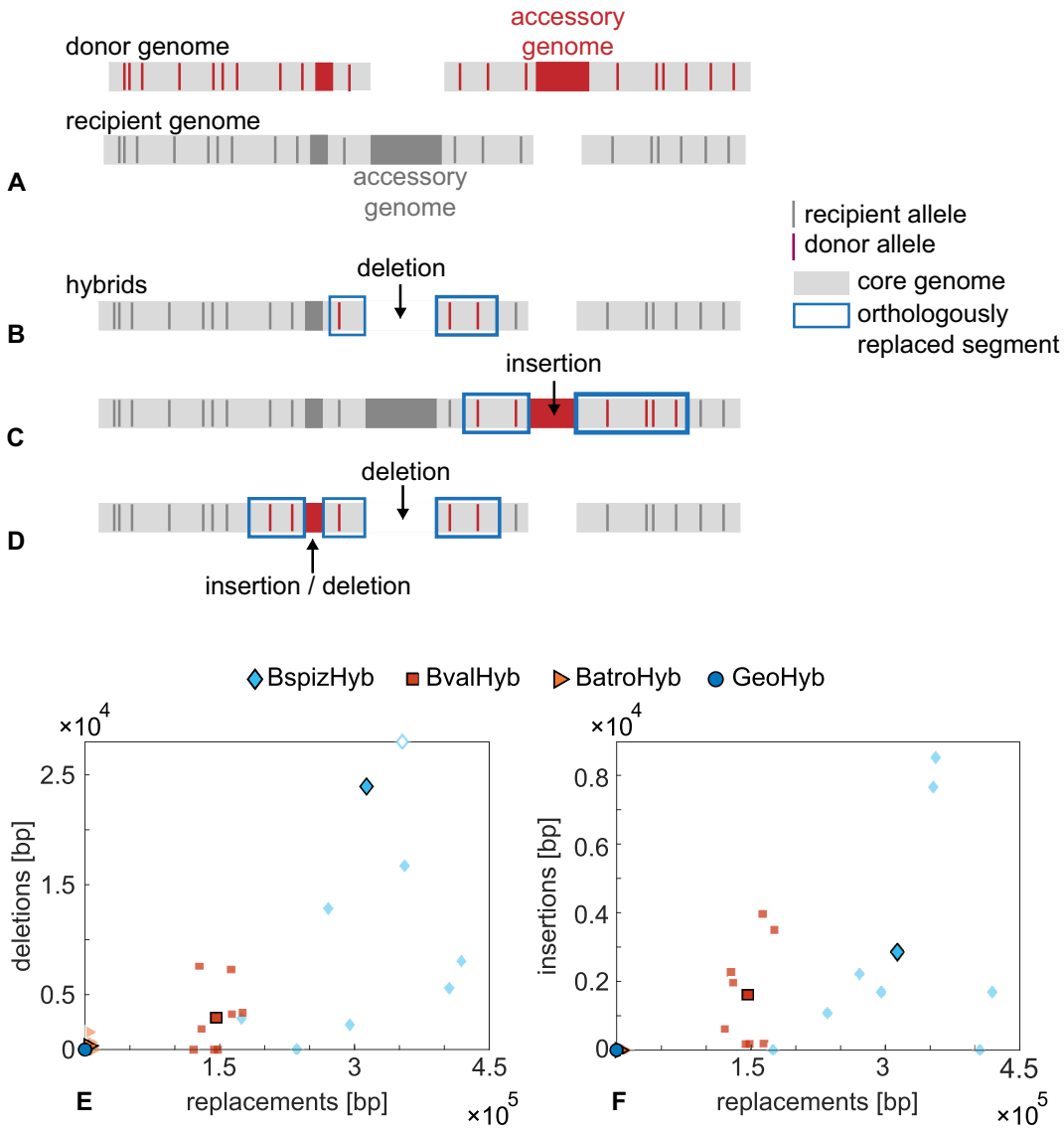


Figure 8. Deletions and insertions occur through orthologous replacements. (A–D) Potential rearrangements of accessory genomes through recombination on flanking regions. (**A**) Scheme of donor and recipient genomes. (**B**) Deletion of an accessory segment of the recipient. (**C**) Insertion of an accessory segment of the donor. (**D**) Complex rearrangement in a mixed event including a swap of an accessory segment of the recipient by an accessory segment of the donor and an additional deletion of an accessory segment of the recipient. (E, F) For each hybrid after 20 cycles, the number of base pairs deleted (**E**) and inserted (**F**) in the transformation hybrids is plotted against the number of base pairs replaced by orthologous replacement. BspizHyb: diamond, BvalHyb: square, BatroHyb: triangle, GeoHyb: circle; light symbols: individual hybrids, dark symbols: average, open diamond: extreme outlier depicted on the axis. Mobile element ICEBs1 is excluded from the analysis.

cuss all 137 cold spot genes. Lack of replacement within specific genes can be caused (i) by flanking regions that belong to the accessory genome and inhibit recombination, (ii) by the high local sequence divergence within the putative cold spot, or (iii) by lethality (or strong reduction of fitness). To address the first point, we calculated the fraction of cold spots that are flanked by accessory genome and found that 29% and 46% of the regions are flanked by accessory genome at one side or both sides, respectively (Figure S12, 7). This indicates that being flanked by accessory genome reduces the probability of orthologous replacement and we conclude that the architecture of the accessory genome affects the rate of core genome replacement. Next, we compared the distributions of mean identities of genes that have been fully replaced in at least one hybrid strain with cold spot genes (Figure S11C). The identity distribution of cold spot genes is shifted towards lower values, indicating that lower-than-average sequence identity contributes to lack of recombination within these regions. However, multiple putative cold spot genes have high sequence identity (Figure S11B). We speculated that replacement within a specific class of these genes may cause a strong reduction of fitness, but using Panther GO, we found no gene class overrepresented within the putative cold spots with a sequence identity exceeding 95%.

Recent reports focused on genome-wide recombination between different strains of the same species. Even though the experimental details vary strongly between the different studies, there are multiple characteristics of orthologous replace-

ment that tend to hold in general. For *S. pneumoniae* (13), *H. influenzae* (26), *Helicobacter pylori* (17) and *B. subtilis* (18) replacements were scattered throughout the genome without obvious hot-spots apart from the sites that were selected for. Here, using *B. subtilis* as recipient, we also find that gene transfer occurs across the entire genome, when *B. spizizenii* or *B. vallismortis* served as donors. By contrast, for the *B. atrophaeus* donor, we observed transfer only within a few genome regions. The average core genome identities of all donor-recipient pairs described above exceed 92%, whereas the average identity between *B. subtilis* and *B. atrophaeus* is considerably lower at 87%. For BatroHyb, transfer was observed only within regions of higher-than-average identity, where ribosomal genes are overrepresented. While other genes with high identity exist (Figure S7), none of them is overrepresented using a Panther GO analysis. It remains difficult to disentangle whether the high sequence identity of these essential genes is a cause or a consequence of enhanced gene transfer rate. These genes may be functional only if their sequence is highly conserved. On the other hand, frequent gene transfer across species may be involved in maintaining the integrity of the essential genes.

We found that the rate of replacement decreases exponentially as a function of sequence divergence of the core genomes of the *Bacillus* donors. At the level of single gene replacement, it has been shown that the probability for recombination decreases rapidly as a function of sequence divergence (19,22–24). In those studies, the transformation rate was determined as the number of clones that developed antibiotic resistance due to replacement of a single nucleotide. In the present study, the transfer rate is defined very differently, i.e. as the fraction of recipient core genome replaced by donor genome per time. This rate depends on the local sequence divergence of the replaced segment (comparable to the previous studies), but also on the length distribution of the integrated segments as well as the architecture of the accessory genomes. Therefore, it is remarkable that the exponential relation holds for genome-wide transfer. Given the exponential relation, we can estimate a transfer rate of 1.2×10^{-4} % h^{-1} for GeoHyb. This corresponds to a number of transferred bp not significantly different from zero during the whole of the 40 h experiment. Due to the large error, we cannot determine whether this donor deviates from the exponential behavior and supports the idea of an identity cut-off for transformation as previously proposed (48). Nevertheless, it is interesting to note that we observed no gene transfer across the different genera. The decrease of replacement probability with average sequence divergence between core genomes is consistent with earlier findings that show larger-than-average sequence identity close to the endpoints of the replacements (17,25). On the other hand, several studies report a lack of dependence between the replacement probability and sequence identity (13,18). This difference is particularly striking in a study by Vasileva *et al.* (18) that reports on recombination between *B. subtilis* and donor species whose sequence divergences are comparable to the ones used in our study. Vasileva *et al.* (18) used protoplast fusion to import donor DNA. Therefore, the substrate for recombination in that study was double-stranded DNA while single-stranded DNA is imported by transformation. It is conceivable that the substrate affects the recombination process.

By comparing the results of the replacement accumulation experiment described in this study with results from our previous study (25), we can assess effects of competition between hybrids on the genome dynamics. That earlier study differed in two important aspects. We let *B. subtilis* transform with genomic DNA of *B. spizizenii*, running 21 cycles. By contrast to the present study, however, during each cycle newly formed hybrid strains were allowed to compete against each other for extended periods of time. As a consequence, the fitter hybrids were selected for. Second, during each cycle, cells were treated with UV light (25). We determined the core genome transfer rate in both studies. Here, we determined a transfer rate of 0.21% h^{-1} for BspizHyb which is lower than the rate found by Power et. al of approximately 0.28% h^{-1}. This indicates that high transfer rates were beneficial in the evolution experiment in agreement with the net fitness increase found in Power *et al.* (25). The length distributions of integrated segments were exponential in both studies, but the characteristic length was higher in the earlier study with selection and irradiation than in the replacement accumulation experiment studied here, suggesting that competition has selected for the integration of extended segments. This observation is interesting in the context of disruptive epistasis; with increasing length of the integrated segments, the probability of partially replacing operons decreases. In summary, comparison between two replacement accumulation experiments in the presence and absence of selection, we find a tendency, that selection and/or irradiation enhances the rate of orthologous replacement and the integration of extended segments.

In conclusion, we show that for species with low sequence divergence, nearly all core genes are accessible to gene transfer. The gene replacement rate between *Bacillus* species decreases exponentially with genome divergence between donor and recipient. Recombination between orthologous genes also mediated dynamics of the accessory genomes, however that rate of insertions and deletions of accessory genome was more than a magnitude lower compared to the replacement rate. We propose that cross-species transformation affects bacterial evolution in two ways. First, there is a very frequent transfer between orthologous genes of closely related species. The fitness effects of most of these transfers are fitness-neutral, yet there is potential for beneficial transfer (27). This frequent exchange represents genomic 'tinkering'. On the other hand, accessory genes are exchanged at a lower rate, yet we expect them to have larger fitness effects because they enable gain and loss of entire genes or operons.

Data availability

The sequencing data that was analysed for this study is made available as a SRA BioProject at NCBI under PRJNA987405. The programming scripts used to analyse the raw data and detect different genomic variations are freely available under https://doi.org/10.5281/zenodo.8273150.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank Viera Kovacova and Leon Haffmans for support with WGS analysis, Lucas Horst and Tobias Bollenbach for help with the liquid handling robot, Gerrit Ansmann for advice on parts of the statistical analysis, and the Maier lab for helpful discussions.

Funding

Deutsche Forschungsgemeinschaft [CRC 1310] and Deutsche Forschungsgemeinschaft [INST216/512/1FUGG] granted to the Regional Computing Center of the University of Cologne (RRZK) for the High Performance Computing (HPC) system CHEOPS. Funding for open access charge: Deutsche Forschungsgemeinschaft [CRC 1310].

Conflict of interest statement

None declared.

References

1. Soucy,S.M., Huang,J. and Gogarten,J.P. (2015) Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.*, **16**, 472–482.
2. Rivera,M.C., Jain,R., Moore,J.E. and Lake,J.A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 6239–6244.
3. Brito,P.H., Chevreux,B., Serra,C.R., Schyns,G., Henriques,A.O. and Pereira-Leal,J.B. (2018) Genetic competence drives genome diversity in *Bacillus subtilis*. *Genome Biol. Evol.*, **10**, 108–124.
4. Popa,O., Landan,G. and Dagan,T. (2017) Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. *ISME J.*, **11**, 543–554.
5. Johnston,C., Martin,B., Fichant,G., Polard,P. and Claverys,J.P. (2014) Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat. Rev. Micro.*, **12**, 181–196.
6. Maier,B. (2020) Competence and transformation in *Bacillus subtilis*. *Curr. Iss. Mol. Biol.*, **37**, 57–76.
7. Bell,J.C. and Kowalczykowski,S.C. (2016) RecA: regulation and mechanism of a molecular search engine. *Trends Biochem. Sci*, **41**, 491–507.
8. Qi,Z., Redding,S., Lee,J.Y., Gibb,B., Kwon,Y., Niu,H.Y., Gaines,W.A., Sung,P. and Greene,E.C. (2015) DNA sequence alignment by microhomology sampling during homologous recombination. *Cell*, **160**, 856–869.
9. Dalia,A.B. and Dalia,T.N. (2019) Spatiotemporal analysis of DNA integration during natural transformation reveals a mode of nongenetic inheritance in bacteria. *Cell*, **179**, 1499–1511.
10. Kennemann,L., Didelot,X., Aebischer,T., Kuhn,S., Drescher,B., Droege,M., Reinhardt,R., Correa,P., Meyer,T.F., Josenhans,C., *et al.* (2011) *Helicobacter pylori* genome evolution during human infection. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 5033–5038.
11. Mostowy,R., Croucher,N.J., Hanage,W.P., Harris,S.R., Bentley,S. and Fraser,C. (2014) Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution. *PLoS Genet.*, **10**, e1004300.
12. Lin,E.A., Zhang,X.S., Levine,S.M., Gill,S.R., Falush,D. and Blaser,M.J. (2009) Natural transformation of *Helicobacter pylori* involves the integration of short DNA fragments interrupted by gaps of variable size. *PLoS Pathog.*, **5**, e1000337.
13. Croucher,N.J., Harris,S.R., Barquist,L., Parkhill,J. and Bentley,S.D. (2012) A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog.*, **8**, e1002745.
14. Gibson,P.S., Bexkens,E., Zuber,S., Cowley,L.A. and Veening,J.W. (2022) The acquisition of clinically relevant amoxicillin resistance in *Streptococcus pneumoniae* requires ordered horizontal gene transfer of four loci. *PLoS Pathog.*, **18**, e1010727.
15. Maamar,H. and Dubnau,D. (2005) Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. *Mol. Microbiol.*, **56**, 615–624.
16. Leisner,M., Sting,K., Frey,E. and Maier,B. (2008) Stochastic switching to competence. *Curr. Opin. Microbiol.*, **11**, 553–559.
17. Bubendorfer,S., Krebes,J., Yang,I., Hage,E., Schulz,T.F., Bahlawane,C., Didelot,X. and Suerbaum,S. (2016) Genome-wide analysis of chromosomal import patterns after natural transformation of Helicobacter pylori. *Nat. Commun.*, **7**, 11995.
18. Vasileva,D.P., Streich,J.C., Burdick,L.H., Klingeman,D.M., Chhetri,H.B., Brelsford,C.M., Ellis,J.C., Close,D.M., Jacobson,D.A. and Michener,J.K. (2022) Protoplast fusion in Bacillus species produces frequent, unbiased, genome-wide homologous recombination. *Nucleic Acids Res.*, **50**, 6211–6223.
19. Carrasco,B., Serrano,E., Sanchez,H., Wyman,C. and Alonso,J.C. (2016) Chromosomal transformation in Bacillus subtilis is a non-polar recombination reaction. *Nucleic Acids Res.*, **44**, 2754–2768.
20. Datta,A., Hendrix,M., Lipsitch,M. and JinksRobertson,S. (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 9757–9762.
21. Li,L.L., Jean,M. and Belzile,F. (2006) The impact of sequence divergence and DNA mismatch repair on homeologous recombination in Arabidopsis. *Plant J.*, **45**, 908–916.
22. Majewski,J., Zawadzki,P., Pickerill,P., Cohan,F.M. and Dowson,C.G. (2000) Barriers to genetic exchange between bacterial species: streptococcus pneumoniae transformation. *J. Bacteriol.*, **182**, 1016–1023.
23. Vulic,M., Dionisio,F., Taddei,F. and Radman,M. (1997) Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 9763–9767.
24. Roberts,M.S. and Cohan,F.M. (1993) The effect of DNA-sequence divergence on sexual isolation in Bacillus. *Genetics*, **134**, 402–408.
25. Power,J.J., Pinheiro,F., Pompei,S., Kovacova,V., Yuksel,M., Rathmann,I., Forster,M., Lassig,M. and Maier,B. (2021) Adaptive evolution of hybrid bacteria by horizontal gene transfer. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2007873118.
26. Mell,J.C., Shumilina,S., Hall,I.M. and Redfield,R.J. (2011) Transformation of natural genetic variation into *Haemophilus influenzae* genomes. *PLoS Pathog.*, **7**, e1002151.
27. Rathmann,I., Forster,M., Yuksel,M., Horst,L., Petrungaro,G., Bollenbach,T. and Maier,B. (2023) Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution. *ISME J.*, **17**, 130–139.
28. Dubnau,D. and Davidoff-Abelson,R. (1971) Fate of transforming DNA following uptake by competent Bacillus subtilis. I. Formation and properties of the donor-recipient complex. *J. Mol. Biol.*, **56**, 209–221.
29. Leisner,M., Stingl,K., Radler,J.O. and Maier,B. (2007) Basal expression rate of comK sets a 'switching-window' into the K-state of *Bacillus subtilis*. *Mol. Microbiol.*, **63**, 1806–1816.
30. Dubnau,D. and Blokesch,M. (2019) Mechanisms of DNA uptake by naturally competent bacteria. *Annu. Rev. Genet.*, **53**, 217–237.
31. Andrews,S. (2010), FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
32. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
33. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
34. Danecek,P., Bonfield,J.K., Liddle,J., Marshall,J., Ohan,V., Pollard,M.O., Whitwham,A., Keane,T., McCarthy,S.A., Davies,R.M., *et al.* (2021) Twelve years of SAMtools and BCFtools. *Gigascience*, **10**, giab008.
35. Danecek,P., Auton,A., Abecasis,G., Albers,C.A., Banks,E., DePristo,M.A., Handsaker,R.E., Lunter,G., Marth,G.T., Sherry,S.T., *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
36. Camacho,C., Coulouris,G., Avagyan,V., Ma,N., Papadopoulos,J., Bealer,K. and Madden,T.L. (2009) BLAST plus : architecture and applications. *BMC Bioinf.*, **10**, 421.

37. Thomas,P.D., Ebert,D., Muruganujan,A., Mushayahama,T., Albou,L.P. and Mi,H.Y. (2022) PANTHER: making genome-scale phylogenetics accessible to all. *Protein Sci.*, **31**, 8–22.
38. Mi,H.Y., Muruganujan,A., Huang,X.S., Ebert,D., Mills,C., Guo,X.Y. and Thomas,P.D. (2019) Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.*, **14**, 703–721.
39. Kim,Y., Gu,C.D., Kim,H.U. and Lee,S.Y. (2020) Current status of pan-genome analysis for pathogenic bacteria. *Curr. Opin Biotech.*, **63**, 54–62.
40. Perrin,A. and Rocha,E.P.C. (2021) PanACoTA: a modular tool for massive microbial comparative genomics. *Nar. Genom. Bioinform.*, **3**, lqaa106.
41. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.
42. Polz,M.F., Alm,E.J. and Hanage,W.P. (2013) Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.*, **29**, 170–175.
43. Fan,B., Blom,J., Klenk,H.P. and Borriss,R. (2017) *Bacillus amyloliquefaciens*, *Bacillus velezensis*, and *Bacillus siamensis* form an "Operational Group B. amyloliquefaciens" within the *B. subtilis* species complex. *Front. Microbiol.*, **8**, 22.
44. Philips,R., Kondev,J., Theriot,J., Gercia,H.G. and Orme,N. (2013) *Physical Biology of the Cell*. Garland Science.
45. Mell,J.C., Lee,J.Y., Firme,M., Sinha,S. and Redfield,R.J. (2014) Extensive cotransformation of natural variation into chromosomes of naturally competent Haemophilus influenzae. *G3*, **4**, 717–731.
46. Alfsnes,K., Frye,S.A., Eriksson,J., Eldholm,V., Brynildsrud,O.B., Bohlin,J., Harrison,O.B., Hood,D.W., Maiden,M.C.J., Tonjum,T., *et al.* (2018) A genomic view of experimental intraspecies and interspecies transformation of a rifampicin-resistance allele into Neisseria meningitidis. *Microb. Genomics*, **4**, e000222.
47. Serrano,E., Ramos,C., Alonso,J.C. and Ayora,S. (2021) Recombination proteins differently control the acquisition of homeologous DNA during Bacillus subtilis natural chromosomal transformation. *Environ. Microbiol.*, **23**, 512–524.
48. Carrasco,B., Serranolt,E., Martin-Gonzalez,A., Moreno-Herrero,F. and Alonso,J.C. (2019) Bacillus subtilis MutS modulates RecA-mediated DNA strand exchange between divergent DNA sequences. *Front. Microbiol.*, **10**, 237.
49. Auchtung,J.M., Aleksanyan,N., Bulku,A. and Berkmen,M.B. (2016) Biology of ICEBs1, an integrative and conjugative element in Bacillus subtilis. *Plasmid*, **86**, 14–25.
50. Auchtung,J.M., Lee,C.A., Monson,R.E., Lehman,A.P. and Grossman,A.D. (2005) Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 12554–12559.
51. Kalamara,M., Spacapan,M., Mandic-Mulec,I. and Stanley-Wall,N.R. (2018) Social behaviours by *Bacillus subtilis*: quorum sensing, kin discrimination and beyond. *Mol. Microbiol.*, **110**, 863–878.

Received: July 3, 2023. **Revised:** September 7, 2023. **Editorial Decision:** October 24, 2023. **Accepted:** November 2, 2023

5. Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution

5.1. Publication

This project was published with the title ”Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution” in ISME Journal, Volume 17 in January 2023 under a Creative Commons Attribution 4.0 International License. The publication reports the results of a joint project between me (MF), Isabel Rathmann (IR), Melih Yüksel (MY), Lucas Horst (LH), Gabriela Petrungaro (GP), Tobias Bollenbach (TB) and Berenike Maier (BM), MF and IR being shared first author.

5.1.1. Contributions

MF, IR, BM, and MY designed the project. The high-throughput competition assay and the evolution experiments with transformation hybrids were developed and performed by MF and IR. Work load was splitted evenly between MF and IR. MF and IR also made the hybrid libraries with donor species *B. vallismortis* (IR) and *B. spizizenii* (MF). MY created the defined growth medium on glycerol basis and genetically modified the used strains. He also generated the single-gene replacement library with genes from *B. vallismortis*. IR isolated genomic DNA for sequencing. MF processed the sequencing raw data with scripts that were developed by MF and IR. Matlab scripts for basic flow cytometry analysis converting events to fractions were written by MF. IR wrote a script to find and optimize the gating that was used on the data. MF created an algorithm to calculate the selection coefficients and IR and MF both optimized the visualization of data. Most statistical analysis regarding the DFEs was performed by IR. LH, GP and TB helped to handle and program the robotic system for the competition and the evolution experiment. MF, IR and BM wrote the manuscript. All authors read and discussed the article.

5.1.2. Key findings

The costs and benefits of transformation at a whole genome level were poorly characterized. We create a hybrid library and measure the distribution of fitness effects (DFE) of transformation to predict conditions in which transformation is beneficial. We find

(i) unlike shown for mutations [106], that the DFE of transformation is not shifted to negative fitness but is on average fitness neutral for most tested conditions. Additionally, we see strongly beneficial fitness outliers that reveal the potential for rapid adaptation (Fig. 5.1).

(ii) Pleiotropy of hybrids in changing conditions supports the idea of a benefit through a shared gene pool for closely related species.

(iii) The results of the evolution experiments provide evidence that DFEs are a suitable tool for predicting evolution of hybrid populations.

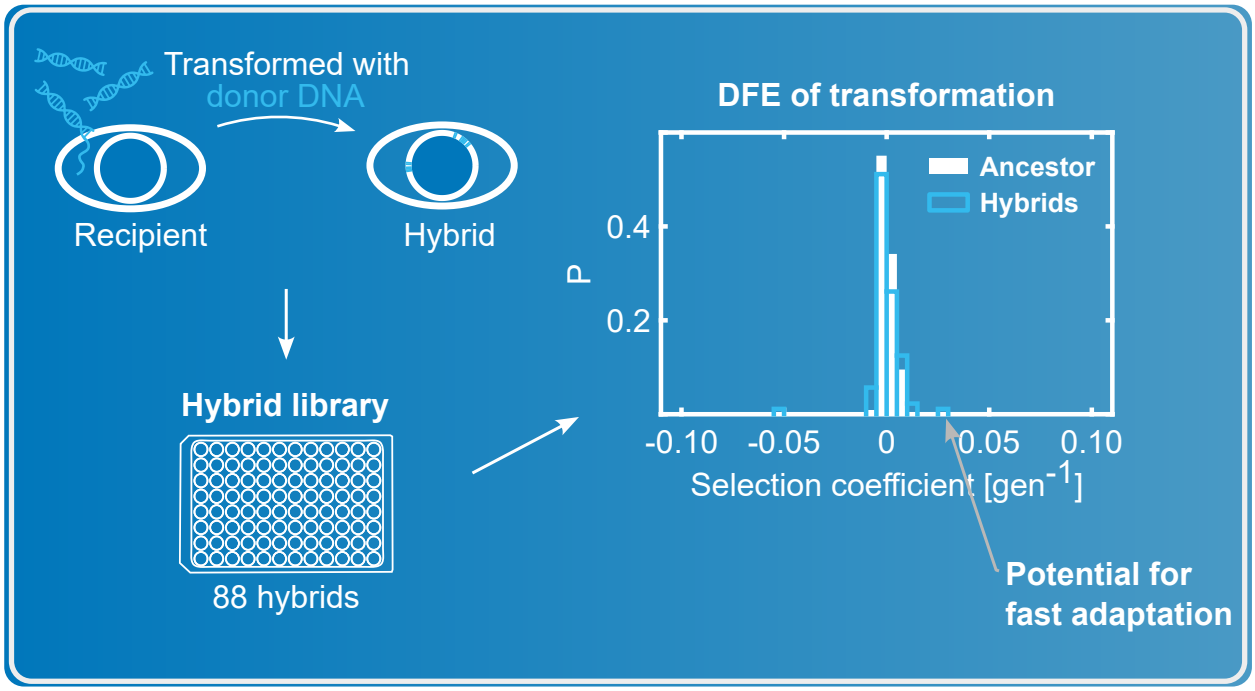


Figure 5.1.: Graphical abstract of the following publication. To predict conditions in which transformation is beneficial, we create a hybrid library and measure the DFEs in several conditions. Positive outlier reveal the potential for fast adaptive evolution. Data from publication used under CC BY 4.0 DEED license.

ISME

www.nature.com/ismej

ARTICLE OPEN

Check for updates

Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution

Isabel Rathmann[1,4], Mona Förster[1,4], Melih Yüksel[1], Lucas Horst[1], Gabriela Petrungaro[1], Tobias Bollenbach[1,2] and Berenike Maier[1,3]✉

Bacterial transformation, a common mechanism of horizontal gene transfer, can speed up adaptive evolution. How its costs and benefits depend on the growth environment is poorly understood. Here, we characterize the distributions of fitness effects (DFE) of transformation in different conditions and test whether they predict in which condition transformation is beneficial. To determine the DFEs, we generate hybrid libraries between the recipient *Bacillus subtilis* and different donor species and measure the selection coefficient of each hybrid strain. In complex medium, the donor *Bacillus vallismortis* confers larger fitness effects than the more closely related donor *Bacillus spizizenii*. For both donors, the DFEs show strong effect beneficial transfers, indicating potential for fast adaptive evolution. While some transfers of *B. vallismortis* DNA show pleiotropic effects, various transfers are beneficial only under a single growth condition, indicating that the recipient can benefit from a variety of donor genes to adapt to varying growth conditions. We scrutinize the predictive value of the DFEs by laboratory evolution under different growth conditions and show that the DFEs correctly predict the condition at which transformation confers a benefit. We conclude that transformation has a strong potential for speeding up adaptation to varying environments by profiting from a gene pool shared between closely related species.

The ISME Journal; https://doi.org/10.1038/s41396-022-01325-5

INTRODUCTION

Horizontal gene transfer (HGT) can enhance the speed of bacterial adaptation to new environments and generate intra-species diversity [1–4]. The simplest mechanism of HGT is transformation. During transformation, bacteria take up DNA from the environment and integrate segments of the newly acquired DNA into their chromosomes by homologous recombination [5, 6]. While the mechanism of transformation is well characterized, there is an ongoing debate about costs and benefits of bacterial transformation as detailed in the following.

Benefits of transformation include the acquisition of novel functions like antibiotic resistance [7–9] or adaptation to novel carbon sources [10]. Transformation can also benefit bacteria by purging the genome from deleterious mutations and mobile genetic elements, or by recombining beneficial mutations that would compete in asexual populations [11–14]. On the other hand, transformation may introduce various costs including reduced RNA stability, protein activity, codon usage mismatch, or disruptive epistasis at the network level [15–17]. Various laboratory evolution experiments have addressed costs and benefits of transformation. These experiments were carried out under different experimental conditions and yielded a broad range of fitness effects of gene transfer [13, 18–22]. In an early study, Baltrus et al. found that the rate of adaptation of transformable *Helicobacter pylori* to a novel environment was higher than the rate of transformation-inhibited bacteria [13]. Other studies report that fitness effects of transformation were dependent on external stress [18] or growth phase [19]. We have studied the evolution of competent *Bacillus subtilis* in the presence of genomic DNA from *Bacillus spizizenii* (formerly *B. subtilis* subsp. *spizizenii*) [23]. Transformation benefitted *B. subtilis* during stationary phase and we found evidence for genome-wide positive and negative selection [23]. Therefore, it is important to understand and even predict conditions, where transformation by a specific donor benefits the recipient.

One central ingredient for predicting evolution is the distribution of fitness effects (DFE) [24]. It characterizes the spectrum of beneficial and deleterious genomic changes available to the evolving organism. This spectrum is usually represented by a library of strains with different specific genomic sequence modifications. So far, libraries containing strains with single mutations [25], multiple mutations [26, 27], and gene deletions [28] have been used to characterize the DFEs. The fitness effects were quantified either by measuring growth rates [25], by determining the selection coefficients in competition experiments [27, 29], indirectly by Bayesian inference [26], or at the single cell level [30]. In general, mutations or deletions occurring in the absence of selection shifted the DFE towards decreased fitness and only few mutations or deletions increased the fitness compared to their ancestors. In contrast, the DFE of

[1]Institute for Biological Physics, University of Cologne, Zülpicherstr. 47a, 50674 Köln, Germany. [2]Center for Data and Simulation Science, University of Cologne, Köln, Germany. [3]Center for Molecular Medicine Cologne, University of Cologne, Köln, Germany. [4]These authors contributed equally: Isabel Rathmann, Mona Förster.
✉email: berenike.maier@uni-koeln.de

Received: 8 October 2021 Revised: 14 September 2022 Accepted: 20 September 2022
Published online: 12 October 2022

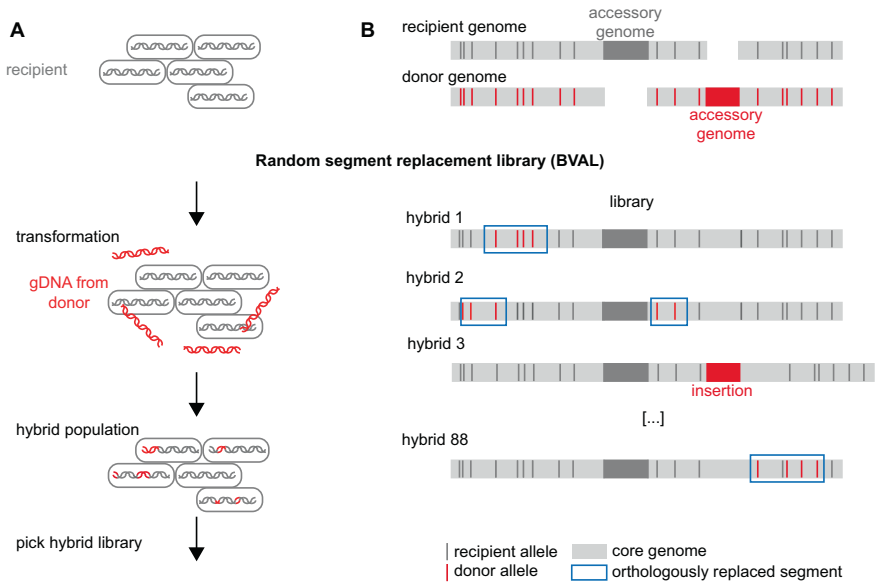


Fig. 1 Library preparation. A The BVAL random segment replacement library was generated by transforming the recipient with genomic DNA of *B. vallismortis* for 2 h, plating, and picking of single colonies each consisting of a monoclonal hybrid. **B** The library consists of hybrids between donor and recipient. Random segments carry the donor alleles of core genes, deletions, and insertions from the donor's accessory genome.

transformation is poorly characterized. One study shows that inserting randomly chosen genes from different species into *Escherichia coli* caused no or mildly deleterious fitness effects [31]. Another study shows that insertion of different genes from *Salmonella typhimurium* into *E. coli* caused strong fitness costs that were dependent on the expression levels of the inserted genes [32]. The fitness effects of orthologous replacement have been investigated at the level of a single gene and shown to be mostly deleterious for different donor species [15, 33, 34]. To the best of our knowledge, the DFE of genome-wide transformation has not been characterized so far.

In this study, we characterized the DFE of cross-species transformation. *B. subtilis* was transformed by genomic DNA from *Bacillus vallismortis* to generate hybrid libraries. By competition between the hybrids and the recipient, we determined the DFE. In complex growth medium, we found strongly beneficial transfers that bear potential for rapid adaptation. In different growth environment, we found evidence for positive and negative synergistic pleiotropy as well as fitness trade-offs. Finally, we scrutinized the predictive value of our DFEs by laboratory evolution and found that the net fitness effects of transformation agree well with our predictions. Thus, our study is a significant step towards making the effects of transformation on bacterial evolution more predictable.

MATERIAL AND METHODS

Strains and media

All experiments were performed with recipient strain Bs166, derived from *B. subtilis* BD630, and reporter strain Bs175, carrying an additional gene encoding GFP (RS). Donor strains are *B. spizizenii* NRRL B-14472/W23 (hybrid library BSPIZ) and *Bacillus vallismortis* DV1-F-3 (hybrid libraries BVAL, BVAL_single and evolved hybrid libraries). Experiments were performed in either complex medium (CM), defined medium (DM) or defined medium with glycerol as sole carbon source ($DM_{glycerol}$).

We used whole genome sequencing data to detect orthologous recombinatios, insertions, deletions and duplications [23]. Reads are mapped using Burrows-Wheeler Aligner (v.0.7.17) [35], then processed with the mpileup function from samtools (samtools 1.8) and the variants are called with the call function from bcftools (bcftools 1.8) [36]. Detailed information to strains, media and sequencing data analysis can be found in Supplementary Methods.

Generation of hybrid libraries

Random replacement hybrid libraries BVAL and BSPIZ were created by transforming recipient Bs166 cells with genomic DNA from *B. vallismortis* or *B. spizizenii*, respectively, and subsequently picking 88 hybrids from single colonies.

For the BVAL_single library, randomly picked single donor genes from *B. vallismortis* were replaced in the recipient entirely without selective markers [37] (details in Supplementary Methods, Fig. S1). BVAL_single consists of 24 strains each having a different gene fully replaced by the donor's ortholog (Dataset S1) and additional 19 strains with a partial replacement.

Experimental evolution. An evolution experiment was performed with 88 hybrid populations created from one transformation step with *B. vallismortis* DNA. Populations were grown in exponential phase in parallel for ~450 generations in either CM or DM. Experiments ran for different periods of time as generation time depended on the growth medium (Table S3). After ~450 generations, one clone per population was picked and collected in the monoclonal hybrid libraries BVALevoCM and BVALevoDM. As a reference for both libraries, 88 wells of the recipient Bs166 were evolved and libraries RECevoCM and RECevoDM were generated. The evolution experiment was performed on an automated system integrated by the company HighRes Biosolutions. Detailed information available in Supplementary Methods.

Determination of selection coefficients. For each created library and the respective control, selection coefficients are measured and represented as distribution of fitness effects (DFE). Strains were competed against the reporter strain (RS) in different growth media and under different temperatures and growth conditions (Table S2). At the start t_0 and end time point t of competition, fraction x_i of the strain of interest and x_{RS} of the reporter strain are determined with a flow cytometer. The selection coefficient was calculated as $s_{i,RS} = \frac{t_g}{t-t_0} ln\left(\frac{x_i(t)/x_{RS}(t)}{x_i(t_0)/x_{RS}(t_0)}\right)$, where t_g is the generation time of the recipient in the respective media (Table S3). For the DFEs, the selection coefficients $s_{i,r}$ were calculated relative to the recipient's fitness measured on the same experimental plate (Details to experimental procedure and analysis of flow cytometry data are explained in Supplementary Methods).

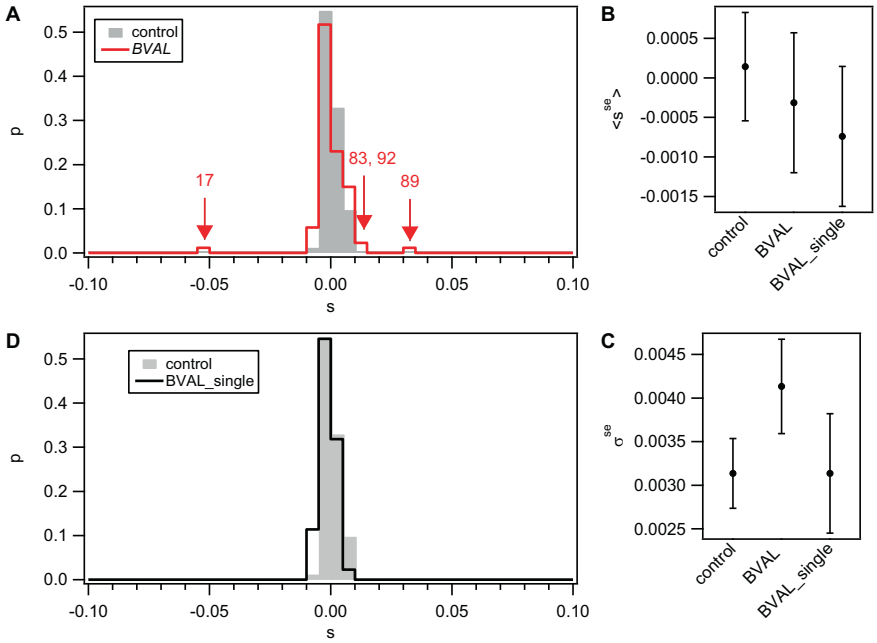


Fig. 2 Distribution of fitness effects of BVAL and BVAL_single libraries. DFE was determined by competition against *gfp* reporter strain (Bs175) during exponential phase in complex medium for 4 h. Grey: Control DFE obtained from competition experiments between recipient (Bs166) and *gfp* reporter strain (grey). **A** Distribution of selection coefficients *s* of the BVAL library. Arrows denote outliers (large effect transfers) from the distribution derived at a significance level of $\alpha = 0.05$ after Bonferroni correction. Numbers denote respective strains of BVAL. **B** Mean selection coefficients of core distributions excluding outliers $\langle s^s \rangle$. Error bars: confidence intervals obtained from bootstrap analysis. **C** Standard deviation of selection coefficients of core distributions excluding outliers σ^{se}. Error bars: confidence intervals obtained from bootstrap analysis. **D** Distribution of selection coefficients *s* of the BVAL_single library.

RESULTS

Characterization of hybrid libraries formed between *B. subtilis* and *B. vallismortis*

In this study, *B. subtilis* served as a recipient species and the closely related *B. vallismortis* as a donor for gene transfer. *B. subtilis* and *B. vallismortis* share a core genome of 3.5 Mbp with an average sequence divergence of 7.4%. Additionally, *B. subtilis* and *B. vallismortis* both have an 0.7 Mbp accessory genome. By creating hybrids between these two species and determining the DFE, we ultimately aimed at understanding how bacterial transformation can drive adaptive evolution.

The hybrid library BVAL consists of 87 strains in which the recipient is genetically modified through transformation by genomic donor DNA (Fig. 1A). Between 0.01 and 0.66% of the core genome were orthologously replaced, i.e. a DNA segments belonging to the core genome of the recipient strain were replaced by a segment of the donor with similar sequence (Fig. 1B). On average, orthologous recombination replaced $(0.1 \pm 0.2)\%$ of the core genome (including all sequenced strains) (Table S1) and affected 6 ± 11 genes at least partially. Besides orthologous replacements, deletions and de novo SNPs were detected. The library was designed to reflect the population of hybrids generated after 2 h of transformation by donor DNA and thus not all hybrid clones were genetically different from the recipient. In the hybrid library BVAL_single, random genes were orthologously replaced (Dataset S1, Fig. S1). By contrast to the BVAL library, intergenic regions were not affected by gene transfer in the BVAL_single library.

Cross-species transfer has potential to enhance fitness

We determined the DFE of transformation in competence medium, a complex medium (CM) used for studying competence and transformation. We obtained the DFE by conducting competition experiments between each strain of the libraries and a fluorescent reporter of the recipient and measuring the selection coefficients as described in the Methods.

All DFEs were be compared to a control DFE which characterizes the resolution of our setup. For the control DFE, we determined the selection coefficients of 82 independently growing recipients competing against their own reporter. The control distribution obtained is centered around $s_{control} = 0.0001 \pm 0.0007$ (mean ± confidence interval) (Fig. 2). Using a Kolmogorov-Smirnov-test (KS-test), we find that the distribution is consistent with a normal distribution. The standard deviation is $\sigma_{control} = 0.0031 \pm 0.0004$.

We measured the DFE of the BVAL library (Fig. 2A). By resampling we show that our sample size adequately captures the global shape of the underlying DFE (Fig. S2). Three strains had strongly positive selection coefficients and one strain had a strongly negative selection coefficient. We defined outliers from the control distribution, i.e. strains with large fitness effects, by a significance level of $\alpha = 0.05$. To account for multiple testing, the Bonferroni correction [38] was applied. Using this criterion, we find three positive and one negative outlier. Henceforth, "large effect transfers" will denote transfers that cause these outliers.

The central portion of the distribution appeared to be broader than the control distribution (Fig. 2A). This could indicate "small effect transfers", whereby individual strains would not show significant fitness effects but the distribution as a whole would still differ from the control. To analyze this, we obtained the core distribution by removing the large effect transfer outliers, which left us with a DFE dominated only by small effects (se), DFE^{se}. We find that the mean of the DFE^{se} is comparable to the control (Fig. 2B). By contrast, the DFE^{se} of BVAL shows a significantly larger standard deviation (Fig. 2C), indicating the existence of multiple small effect transfers with positive and negative fitness effects. An individual strain with a small effect transfer is not significantly

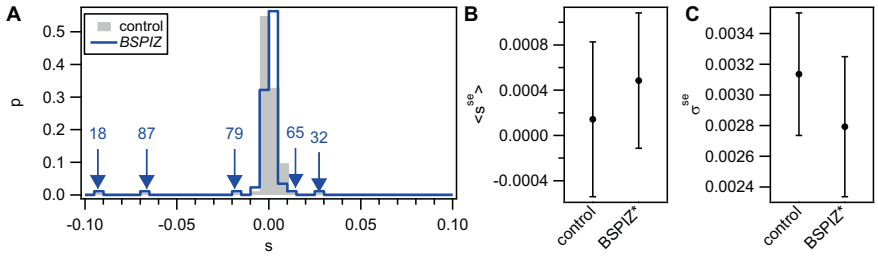


Fig. 3 Distribution of fitness effects of BSPIZ library. DFE was determined by competition against *gfp* reporter strain (Bs175) during exponential phase in complex medium for 4 h. Grey: Control DFE obtained from competition experiments between recipient (Bs166) and *gfp* reporter strain (grey). **A** Distribution of selection coefficients *s*. Arrows denote outliers (large effect transfers) from the distribution derived at a significance level of $\alpha = 0.05$ after Bonferroni correction. Numbers denote respective strains of BSPIZ. **B** Mean selection coefficients of core distributions excluding outliers $\langle s^{se} \rangle$. Error bars: confidence intervals obtained from bootstrap analysis. **C** Standard deviation of selection coefficients of core distributions excluding outliers σ^{se}. Error bars: confidence intervals obtained from bootstrap analysis.

different from the control distribution, but DFE^{se} as a whole is significantly different. This result was also confirmed by a Bartlett test.

The DFE of the BVAL_single library showed a small shift to negative selection coefficients but no outliers from the control distribution (Fig. 2D). Neither the mean selection coefficient nor the standard deviation was significantly different from the control distribution (Fig. 2B, C). When strains with full gene replacements and partial replacements were analyzed separately, no significant difference was found (Fig. S3). Taken together our data do not reveal strong fitness effects of single gene replacement. We conclude that in complex medium the DFE of transformation shows large effect and small effect transfers. In particular, 3 out of 88 transformants have strongly enhanced fitness.

Transfers from more closely related donor species create higher genomic variability but weaker fitness effects

Different donors are expected to generate different DFEs of transformation. Here, we investigate the DFE of transformation using *B. spizizenii* as donor. *B. subtilis* and *B. spizizenii* share a core genome of 3.6 Mbp with an average sequence divergence of 6.8 %. The size of the core genome is comparable to *B. vallismortis*, but the sequence divergence is lower. Additionally, *B. subtilis* has 0.6 Mbp accessory genome and *B. spizizenii* has 0.4 Mbp. The library BSPIZ was generated using the same method as for BVAL. The mean fraction of orthologous replacement was (0.5 ± 0.8) % (Table S1). On average, 21 ± 31 genes were hit by orthologous replacement. Next to orthologous replacement, insertions from the accessory genome of the donor and deletions were detected. In total, the strains of BSPIZ showed more gene transfer compared to BVAL.

We measured the DFE of the BSPIZ library in complex medium (Fig. 3A). Two strains were characterized as positive outliers and three strains as negative outliers. After removing these large effect transfer outliers, we investigated whether the core distribution was influenced detectably by small effect transfers. We found that neither the mean selection coefficient of the core distribution nor the standard deviation was significantly different from the control (Fig. 3B, C). In summary, a more closely related donor generated more gene transfer. The DFE showed no detectable small effect transfers, but multiple beneficial and deleterious strong effect transfers.

Different growth conditions strongly affect the DFE showing different types of pleiotropy

The heterogeneity generated by transformation likely enables adaptation to fluctuating environments. We expect different hybrids to show fitness effects under different experimental and environmental conditions. To find out how growth conditions affect the DFE, we measured the selection coefficients of the BVAL hybrids against the recipient under different conditions (Table S2). First, we assessed the influence of different growth conditions in complex medium. In particular, we addressed the effects of the lag phase and of increased temperature. Second, we maintained the original growth conditions but varied the carbon sources and amino acid compositions.

The lag phase introduces fitness effects that are distinct from changes in growth rate during exponential phase. The duration of the lag phase is a selective trait, for example in the presence of antibiotics [39]. In our experimental setup, fast escape from the lag phase will be detected as a benefit. To assess the fitness effects of the lag phase, we compared the selection coefficients of BVAL determined excluding (Fig. 2) and including (Fig. 4A) the lag phase. To include the lag phase, the competitors were harvested from stationary phase, and mixed with the reporter immediately. When the lag phase was excluded, the competitors were harvested from stationary phase, grown to exponential phase, and subsequently mixed. When we included the lag phase, we find six strong effect transfers (Fig. 4A, Fig. S4A) of which three are positive outliers and three are negative outliers. Using a scatter plot, we investigate correlations between conditions with and without lag phase (Fig. 4A). We find different types of pleiotropy between both conditions. Regarding outliers only, one strain (BVAL_89) shows beneficial synergistic pleiotropy, i.e. fitness is strongly increased under both conditions. Another strain (BVAL_17) shows deleterious synergistic pleiotropy. One strain (BVAL_92) shows antagonistic pleiotropy, suggesting that a trait has been transferred that confers a benefit during exponential growth in complex medium but a cost during escape from the lag phase. Several strains were defined as fitness outliers only under one of both conditions. To evaluate whether transfers with small fitness effects were prominent, we analysed the DFE^{se}_{lag} after removing the outliers. We found that the mean selection coefficient was positive (Fig. S4B) and both the mean and standard deviation were significantly higher than in the control (Fig. S4C).

B. vallismortis has been isolated from Death Valley soil and it is conceivable that it is well adapted to higher temperatures. To investigate whether the hybrids benefit from the donor's trait, we performed the competition experiment (excluding the lag phase) at a temperature of 42 °C. At this temperature, the generation time of the recipient was decreased slightly to (14.9 ± 0.2) min. Five strong effect transfers were detected; two positive outliers and three negative outliers (Fig. 4B, Fig. S4D). The positive outliers (BVAL_4, 35) are different from the positive outliers under the previously studied conditions, showing that different horizontally acquired genes confer a benefit in different environments. Again, BVAL_17 shows deleterious synergistic pleiotropy and BVAL_92 shows antagonistic pleiotropy (Fig. 4B). After removing the outliers, the standard deviation was significantly higher than the control standard deviation (Fig. S4F).

We investigated effects of different carbon sources on the DFE. First, we used a chemically defined medium (DM) where glucose,

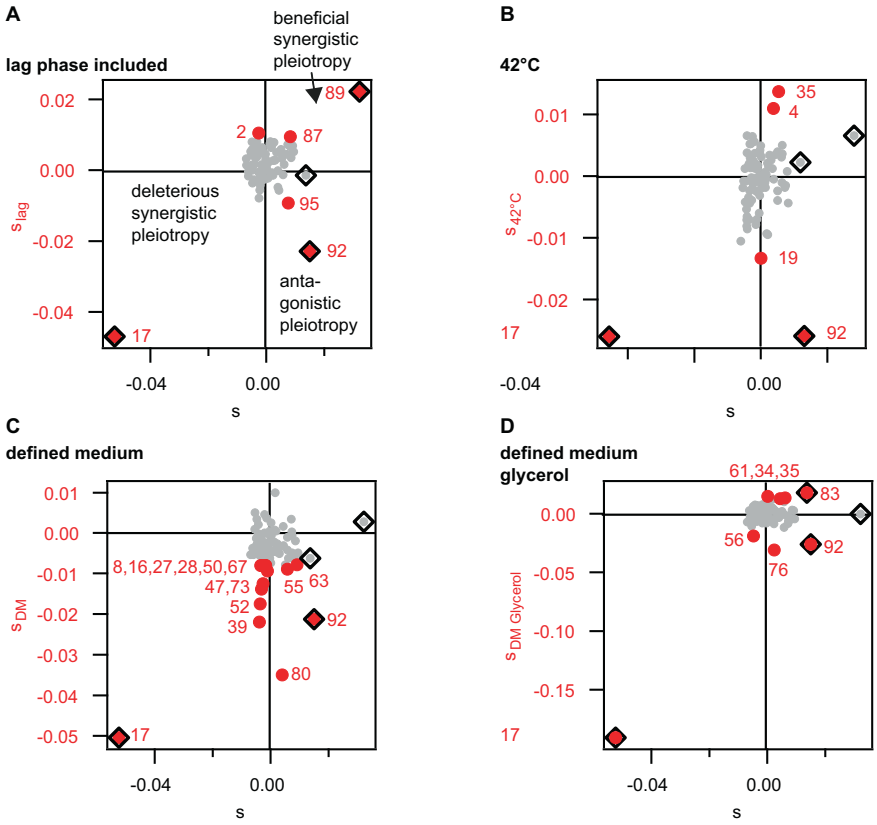


Fig. 4 Fitness effects depend on growth conditions. Scatter plots of selection coefficients of BVAL determined under different growth conditions plotted against the selection coefficients in complex medium at 37 °C °C as shown in Fig. 2A. Red circles: outliers from DFE under different conditions, black diamonds: outliers from DFE in complex medium at 37 °C, grey: strains with small effect transfers. **A** Complex medium including the lag phase at 37 °C. **B** Complex medium excluding the lag phase at 42 °C. **C** Defined medium excluding lag phase at 37 °C. **D** Defined medium with glycerol as only carbon source excluding lag phase at 37 °C.

glutamate, and citrate were used as carbon source. The generation time in DM was increased to (39.0 ± 0.2) min. In DM, the DFE was shifted towards negative selection coefficients (Fig. 4C, Fig. S5A). We found 15 strong effect transfers and all of them were outliers towards negative fitness (Fig. S5A). Without outliers, the mean selection coefficient was negative with $\langle s^{se}_{DM} \rangle = (-0.0026 \pm 0.0008)$ (Fig. S5B). Thus, exponential growth at 37 °C in defined medium is the only condition under which we see a significant decrease in mean selection coefficients.

We characterized the fitness effects in defined medium with glycerol as the only carbon source ($DM_{Glycerol}$). The generation time in $DM_{Glycerol}$ was (48.3 ± 0.3) min. We found eight strong effect transfers, four of them (BVAL_34, _35, _61, _83) having a positive fitness effect (Fig. 4D, Fig. S5D). Compared to competition in complex medium, BVAL_83 showed positive synergistic pleiotropy.

In summary, the DFEs revealed beneficial transfers from *B. vallismortis* to *B. subtilis* under most but not all conditions studied. When comparing fitness effects of transformation under different conditions, we found evidence for synergistic and antagonistic pleiotropy. Several transfers were beneficial only under a single condition.

Transformation confers a benefit in complex medium but not in defined medium

Reviewing the measured DFEs of gene transfer from *B. vallismortis* to *B. subtilis*, we identified two conditions that allow for a prediction on the course of adaptive evolution. First, we note that in complex medium, we found large effect beneficial transfers. Opposed to this, in defined medium, the mean selection coefficient was shifted to a negative value and no large effect beneficial transfers were found. Using this information, we predicted that bacteria benefit more from transformation in complex medium but less so in defined medium. To scrutinize this prediction, we designed a laboratory evolution experiment (Fig. 5A) that ran in both growth media. First, the recipient *B. subtilis* was transformed by gDNA from the donor *B. vallismortis* (Fig. 1A). For both CM and DM, the freshly generated hybrids were split into 88 wells. Thus, at the beginning of the evolution experiment, we had 88 populations each containing ~10^5 different hybrid clones. For each condition, these populations evolved independently by growing exponentially for ~450 generations (Fig. 5A), i.e. 5 days in CM and 12.5 days in DM. The same experiment was performed with the untransformed recipient *B. subtilis* so that we could compare hybrid populations to populations without prior transformation. After ~450 generations, we assessed how the fitness in CM and DM had changed during evolution. To this end, we generated the evolved strain libraries by picking a random clone from each of the 88 populations. The evolved hybrids were represented by the libraries BVALevoCM and BVALevoDM for evolution in complex and defined medium, respectively. The distribution of selection coefficients relative to the recipient was determined using competition experiments.

In complex medium (CM), the mean selection coefficient of the hybrid library BVALevoCM $\langle s^{CM}_{BVALevo} \rangle = (0.008 \pm 0.001)$ was

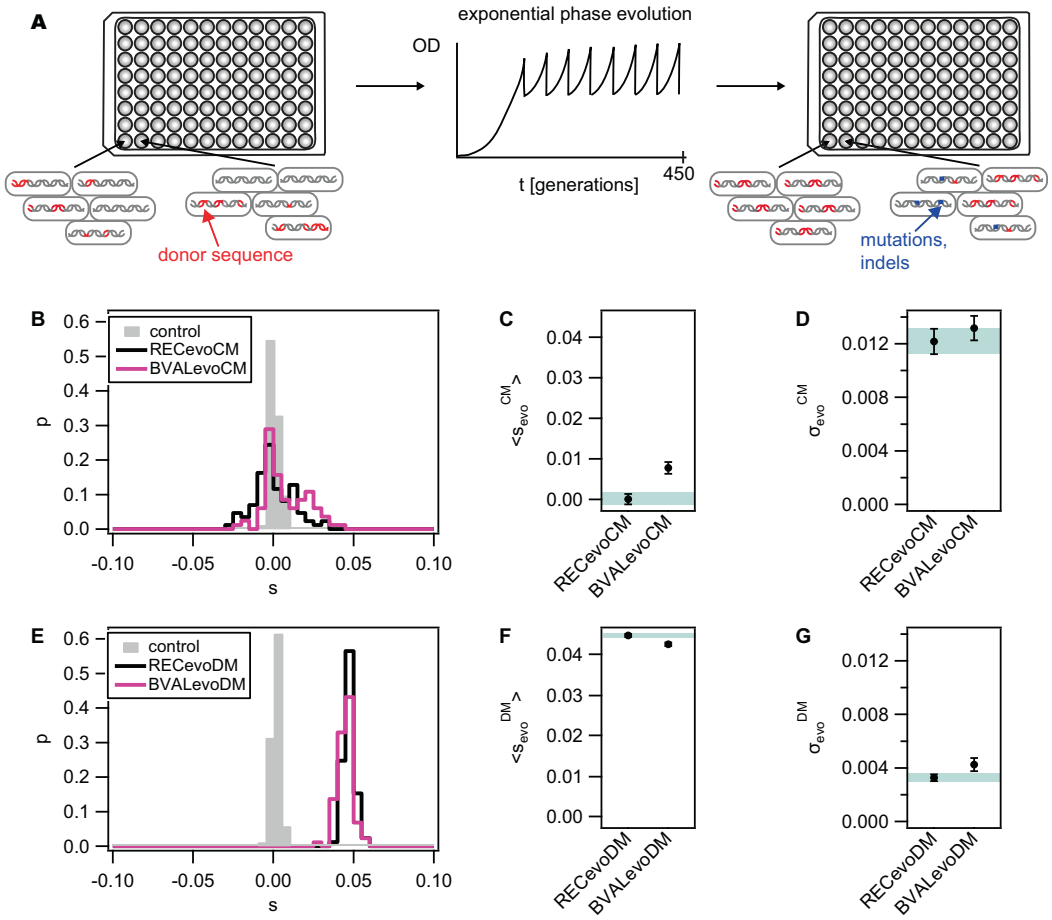


Fig. 5 Fitness effects of transformation in evolving populations. A Sketch of the evolution experiment. 88 batches of the recipient (REC, Bs166) were transformed with donor DNA prior to the evolution experiment, creating a population consisting of different hybrids within a single batch. During laboratory evolution, the OD was monitored, and the batches were diluted repeatedly to maintain exponential growth. After ~450 generations (5 days in CM and 12.5 days in DM), a single clone was picked from each batch, creating a library of evolved strains (BVALevoCM in CM and BVALevoDM in DM). The same procedure was repeated using the untransformed recipient, creating RECevoCM and RECevoDM. The selection coefficients of the libraries were determined by competition against the reporter Bs175. **B**, **E** Distribution of selection coefficients in **B** complex medium and **E** defined medium. Grey: control (Bs166) without evolution, black: evolved recipient strains, pink: evolved hybrid strains. **C**, **F** Mean selection coefficients and **D**, **G** standard deviation in **C**, **D** complex medium and **F**, **G** defined medium. Error bars: confidence intervals obtained from bootstrap analysis. Shaded area: confidence level of evolved recipient data.

significantly higher than the mean selection coefficient of the recipient library RECevoCM $\langle s^{CM}_{RECevo}\rangle = (0.000 \pm 0.001)$ (Fig. 5B, C). This result shows that the transformation caused a fitness increase in CM as predicted by the DFE of BVAL. A KS-test shows that the distributions of selection coefficients of BVALevoCM and RECevoCM are significantly different ($p = 0.003$). The distributions of selection coefficients showed comparable broadening relative to the control distribution (Fig. 5B, D).

In defined medium the mean selection coefficient of the untransformed library RECevoDM $\langle s^{DM}_{RECevo}\rangle = (0.0447 \pm 0.0004)$ was slightly but significantly higher than the selection coefficient of the hybrid library BVALevoDM with $\langle s^{DM}_{BVALevo}\rangle = (0.0425 \pm 0.0004)$ (Fig. 5E, F). This result is consistent with the prediction based on the DFE of BVAL that transformation is not beneficial in defined medium. While the standard deviation of selection coefficient of RECevoDM was not significantly different from the standard deviation of the control distribution (Fig. 5G, Fig. 2C), the standard deviation of BVALevoDM was slightly increased (Fig. 5G). A KS-test shows that the distributions of selection coefficients of BVALevoDM and RECevoDM are significantly different ($p = 0.005$). We conclude that transformation by DNA from *B. vallismortis* confers a net benefit in complex medium, but introduces a cost in defined medium as predicted by the respective DFEs.

Genetic variability is higher for bacteria evolved in complex medium than in defined medium

We assessed repeatability of orthologous recombination and de novo mutations. During the evolution experiment, de novo mutations occur and selection leads to their fixation. For the evolved recipients, this is the only source of genetic variation. For the evolved hybrids on the other hand, mutations occur against the background of the orthologous recombination. The distributions of selection coefficients in Fig. 5C, F suggested high repeatability in defined medium and lower repeatability in complex medium. To assess repeatability, we performed whole genome sequencing on the ten fittest hybrids of each evolved library (Dataset S2). We detected orthologous recombination as well as de novo mutations, including de novo indels and de novo SNPs. Here, we report genes that genetically changed in at least two strains as hotspots (Fig. 6).

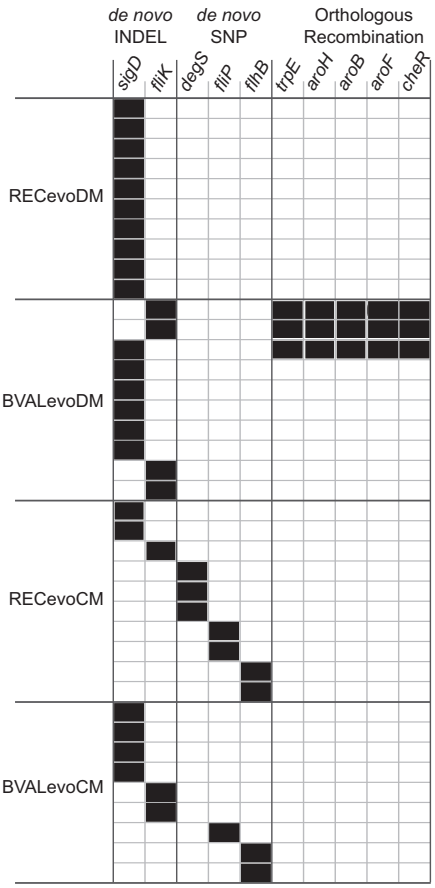


Fig. 6 Hotspots of mutations, indels, and orthologous recombination. A hotspot was defined as a gene that showed sequence changes in at least two of the evolved strains. De novo INDEL: small insertion or deletion of bases within the gene, de novo SNP: single nucleotide polymorphism in the gene that does not arise from orthologous replacement with a donor segment, Recombination: recipient gene replaced by donor gene through orthologous recombination.

In defined medium, we found that both evolved libraries, BVALevoDM and RECevoDM, had considerably increased fitness (Fig. 5E–G). This suggests that the better part of the fitness increase occurred independently of transformation and was caused by a few shared de novo mutations. Indeed, we found that all of the sequenced strains of RECevoDM had frame-shift mutations in the *sigD* gene encoding for an alternative sigma factor involved in flagella synthesis (Fig. 6) [40, 41]. The indels occurred within a poly-A stretch (Dataset S2) suggesting that *sigD* is phase-variable. No further hotspots were found for RECevoDM. Out of the ten strains sequenced for BVALevoDM (hybrids evolved in defined medium), six showed an indel in *sigD* and four had an indel in *fliK* encoding for a protein involved in controlling the hook length of the flagellum [42]. No further mutational hotspots were found. In addition to the de novo mutations, we found a recombination hotspot. In three out of ten strains, the operon *cheR-aroF-aroB-aroH*, and the adjacent gene *trpE* were affected by orthologous recombination.

For complex medium, we found a broader distribution of selection coefficients of BVALevoCM and RECevoCM (Fig. 5B–D), suggesting that evolution is less repeatable. In accordance with this, we found a higher diversity of hotspots in the sequenced strains. Nevertheless, all of the sequenced strains show indels or SNPs in flagella-related genes (Fig. 6). These include indels in *sigD* and *fliK*, and additionally, SNPs in genes responsible for forming the basal complex of flagella, *fliP*, and *flhB* as well as the two-component sensor kinase *degS* which affects the *sigD* regulon [43]. Even though orthologous recombination took place, we did not detect a recombination hotspot. Instead, many different genes were affected by recombination.

In summary, we found that repeatability of evolution is higher in defined medium than in complex medium consistent with the high selection coefficients in this medium.

DISCUSSION

In this work, we systematically address the effects of cross-species transformation on bacterial fitness. We show how its benefit depends on the growth context and use the DFEs to make predictions about environments that favour transformation during adaptive evolution. Results obtained by laboratory evolution support the predictive value of the DFEs.

We found qualitative differences between the DFEs of transformation and the previously characterized DFEs of single mutations and single gene deletions show qualitative differences [24–27]. For the latter, the centres of the DFEs were shifted towards deleterious fitness effects with few exceptions, including a mutator strain with a specific mutation spectrum [25]. Few mutations or deletions had positive fitness effects. For orthologous replacement investigated in this study, we expected that fitness effects were deleterious on average as a consequence of hybrid incompatibilities. Disruptive epistasis at the level of functional networks [23] and suboptimal gene expression levels [15, 33, 44] were likely to reduce the fitness of the hybrid strains. Here, we found little support for this expectation with one exception discussed below. Using two different donor strains and four different growth conditions, our data indicate that transformation by DNA from closely related species is fitness neutral. The DFE has a core distribution around $s = 0$ that tends to broaden as a result of gene transfer. For all tested conditions 2–5% of the hybrid strains had selection coefficients significantly higher than the control distribution, i.e. large effect beneficial transfers. 1–5% of the hybrid strains showed large effect deleterious transfers. The large effect beneficial transfers have potential for increasing the speed of adaptation. This suggests that *B. subtilis* can use a shared gene pool with closely related species to adapt to a variety of new environments. We found one interesting exception from the general trend described so far. In defined medium (DM), we identified fifteen large effect deleterious transfers and the core distribution was shifted to a negative value. We conclude that gene transfer from *B. vallismortis* to *B. subtilis* tends to confer a net fitness cost in DM. Currently, we can only speculate why the DFE in defined medium is qualitatively different from all other conditions probed in this study. The difference between complex medium and defined medium could be explained by the complexity hypothesis [17]; most likely, bacteria adapt "add-on" functions that work independently of strongly interconnected networks belonging to the central metabolism. In defined medium, bacteria metabolize glucose, glutamate, and citrate. The involved pathways are likely to be similar between the species and, therefore, orthologous replacements have small fitness effects. In complex environments, different alternative and poorly linked pathways enable bacteria to adapt to different growth environments. Therefore, genetic exchange in complex environment can confer a higher fitness advantage than in environments of low complexity. However, the $DFE_{glycerol}$ in defined medium with glycerol as the only carbon source does not support this explanation, since we found strong effect beneficial transfers.

Based on the results obtained from the DFEs, we predicted that transformation conveys a fitness benefit in CM but less so in DM. An

evolution experiment supported this prediction. Additionally, we found that repeatability of fitness effects, de novo mutations, and orthologous recombination was higher in defined medium. In DM, all sequenced strains carry an indel in *sigD* or *fliK* disrupting stable flagella formation. The latter gene plays an important role during the formation of the flagellum for robustly activating *sigD*, and we conclude that its frame-shift mutation confers the same functional change as inactivation of *sigD*. Inactivation of *sigD* has been shown to confer a benefit previously [45]. Even though we sequenced only one clone of each population, it is most likely, that this indel became fixed and conferred the fitness increase. This also explains why we see no strain with negative fitness effect, as evolution has most likely led to fixation in most populations. For the hybrid strains evolved in defined medium, we also find a hotspot of gene transfer encompassing the *cheR-aroF-aroB-aroH* operon, and the adjacent gene *trpE*. These genes are involved in the shikimate pathway responsible for synthesis of tryptophan, phenylalanine and tyrosine [46]. Defined medium does not contain these amino acids, and the transfer may enhance their production rate. While in DM transformation caused a cost at the genome-wide scale, this specific orthologous recombination was most likely selected for.

In complex medium, beneficial evolved hybrids have a higher genetic variety. Again, all of the mutational hotspots occurred within flagella-related genes. However, the diversity of flagella-related genes mutated in complex medium is higher than in defined medium. Unexpectedly, the untransformed recipient showed no net fitness increase in complex medium after ~450 generations. We suggest that the recipient is already well-adapted to the complex medium prior to the evolution experiment and mutations do not likely convey large benefits but rather broaden the DFE. Despite the net benefit of gene transfer, we found no hotspot of recombination in complex medium. We argue that this is due to the fact that the initial hybrid populations likely contain multiple different beneficial transfers. Each hybrid population initially comprises about 10^5 cells and we have even observed beneficial transfers in the BVAL library made up by only 87 strains. Additionally, we hypothesize that fitness differences between beneficial transfers are smaller in complex medium and that this slows down fixation time. Consequently, beneficial hybrids might not yet have reached fixation in the population. For both the recipient and the control, we found multiple strains that even show decreased fitness. We suggest that complex medium sets the stage for more complicated population dynamics in which fixation is delayed and even hybrids with negative fitness effects can persist. It is also conceivable that interactions between different clones of one population increase the population fitness. By picking a single clone from each evolved population, we did not account for this possibility. Future experiments will have to address within population diversity as a function of time.

Here, we have designed the hybrid libraries to reflect fitness effects after one single cycle of transformation in *B. subtilis*. *B. subtilis* switches stochastically into the state of competence at high cell density and remains competent for ≈2 h [6, 47]. Therefore, our libraries were generated by transforming with donor DNA for this period of time. We show that within one cycle, different donor species create different degrees of genetic variation. The fraction of replaced genome is considerably (>2 fold) higher with *B. spizizenii* than with *B. vallismortis*, yet the fitness effects are stronger with *B. vallismortis*. While both DFEs contain strongly beneficial outliers, the DFE of *B. vallismortis* reveals additional small effect transfers, suggesting that fitness effects increase with increasing sequence divergence between donor and recipient. The mean lengths of the replaced segments were considerably shorter in BVAL (1.3 kbp) compared to BSPIZ (4.0 kbp). Whereas the mean segment length of the BVAL strains hardly exceeds the mean length of a single gene, the mean segment length of the BSPIZ strains encompasses multiple average-sized genes. Therefore, the probability that operons including their promoters are fully replaced is higher and the probability of disruptive epistasis of functional networks is lower, suggesting that the fitness effects of gene transfer from *B. spizizenii* are lower.

In this work, cells are generally kept in exponential growth phase during evolution and competition experiments. In this well-defined and reproducible condition, the number of cells increases rapidly compared to other growth phases and selection mainly acts on fast growth. Thus, restricting the experiments to this single phase facilitates mapping of genetic variations to their fitness effects. We show that adding another growth phase, such as the lag phase, to the growth condition has an impact on the hybrid fitness. We expect the same to be true if including the stationary growth phase, especially for *B. subtilis* where processes like cannibalism, sporulation, and biofilm formation will potentially have an impact on fitness effects [48]. We envision that our method can be used in the future to systematically assess the DFEs in different growth phases.

In conclusion, the DFE of transformation is systematically different from the DFE of mutations investigated previously. By contrast to the latter, there is no net shift to reduced fitness in most growth conditions. This difference may be explained by the fact that the exchanged sequences were functional in the donor. Our study corroborates the idea that a shared gene-pool between closely related species enables rapid adaptation to changing environments. In support of this idea, recent work showed that *B. subtilis* enters the state of competence more frequently if a closely related species is present [8]. In future studies, it will be interesting to find out how the prolonged presence of other species affects the fitness and genome dynamics of the recipient. In particular, it is currently unclear whether fitness effects of multiple transfers are additive or whether epistatic effects dominate. In terms of application, the predictive value of the DFEs will most likely become useful for predicting effects of transformation on the speed of antibiotic resistance evolution.

DATA AVAILABILITY

The datasets generated during and/or analysed during the current study are available as Supplementary data and at NCBI SRA (BioProject PRJNA877563).

REFERENCES

1. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet 2015;16:472–82. https://doi.org/10.1038/nrg3962.
2. Brito PH, Chevreux B, Serra CR, Schyns G, Henriques AO, Pereira-Leal JB. Genetic competence drives genome diversity in Bacillus subtilis. Genome Biol Evol 2018;10:108–24. https://doi.org/10.1093/gbe/evx270.
3. Pang TY, Lercher MJ. Each of 3,323 metabolic innovations in the evolution of E-coli arose through the horizontal transfer of a single DNA segment. P Natl Acad Sci Usa 2019;116:187–92. https://doi.org/10.1073/pnas.1718997115.
4. Popa O, Landan G, Dagan T. Phylogenomic networks reveal limited phylogenetic range of lateral gene transfer by transduction. ISME J 2017;11:543–54. https://doi.org/10.1038/ismej.2016.116.
5. Dubnau D, Blokesch M. Mechanisms of DNA uptake by naturally competent bacteria. Annu Rev Genet. 2019;53:217–37. https://doi.org/10.1146/annurev-genet-112618-043641.
6. Maier B. Competence and transformation in Bacillus subtilis. Curr Issues Mol Biol. 2020;37:57–76. https://doi.org/10.21775/cimb.037.057.
7. Woods LC, Gorrell RJ, Taylor F, Connallon T, Kwok T, McDonald MJ. Horizontal gene transfer potentiates adaptation by reducing selective constraints on the spread of genetic variation. Proc Natl Acad Sci Usa 2020;117:26868–75. https://doi.org/10.1073/pnas.2005331117.
8. Stefanic P, Belcijan K, Kraigher B, Kostanjsek R, Nesme J, Madsen JS, et al. Kin discrimination promotes horizontal gene transfer between unrelated strains in Bacillus subtilis. Nat Commun 2021;12:3457. doi: ARTN 345710.1038/s41467-021-23685-w.
9. Rouquette-Loughlin CE, Reimche JL, Balthazar JT, Dhulipala V, Gernert KM, Kersh EN, et al. Mechanistic basis for decreased antimicrobial susceptibility in a clinical isolate of neisseria gonorrhoeae possessing a mosaic-like mtr efflux pump locus. Mbio 2018;9:e02281.

10. Chu HY, Sprouffske K, Wagner A. Assessing the benefits of horizontal gene transfer by laboratory evolution and genome sequencing. BMC Evol Biol. 2018;18:54. Epub 2018/04/21. doi: 10.1186/s12862-018-1164-7.
11. Muller HJ. Some genetic aspects of sex. Am Nat. 1932;66:118–38.
12. de Visser JAGM, Elena SF. The evolution of sex: empirical insights into the roles of epistasis and drift. Nat Rev Genet 2007;8:139–49. https://doi.org/10.1038/nrg1985.
13. Baltrus DA, Guillemin K, Phillips PC. Natural transformation increases the rate of adaptation in the human pathogen helicobacter pylori. Evolution 2008;62:39–49. https://doi.org/10.1111/j.1558-5646.2007.00271.x.
14. Croucher NJ, Mostowy R, Wymant C, Turner P, Bentley SD, Fraser C. Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. Plos Biol 2016;14:e1002394. doi: ARTN e100239410.1371/journal.pbio.1002394.
15. Bershtein S, Serohijos AWR, Bhattacharyya S, Manhart M, Choi JM, Mu WM, et al. Protein homeostasis imposes a barrier on functional integration of horizontally transferred genes in bacteria. Plos Genet. 2015;11:e1005612. doi: ARTN e100561210.1371/journal.pgen.1005612.
16. Bloom JD, Labthavikul ST, Otey CR, Arnold FH. Protein stability promotes evolvability. P Natl Acad Sci Usa 2006;103:5869–74. https://doi.org/10.1073/pnas.0510098103.
17. Hall JPJ, Brockhurst MA, Harrison E. Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria. Philos T R Soc B 2017;372:20160424.
18. Engelmoer DJP, Donaldson I, Rozen DE. Conservative sex and the benefits of transformation in Streptococcus pneumoniae. Plos Pathog. 2013;9:e1003758.
19. Utnes ALG, Sorum V, Hulter N, Primicerio R, Hegstad J, Kloos J, et al. Growth phase-specific evolutionary benefits of natural transformation in Acinetobacter baylyi. Isme J 2015;9:2221–31. https://doi.org/10.1038/ismej.2015.35.
20. Hulter N, Sorum V, Borch-Pedersen K, Liljegren MM, Utnes ALG, Primicerio R, et al. Costs and benefits of natural transformation in Acinetobacter baylyi. Bmc Microbiol 2017;17:34. doi: ARTN 3410.1186/s12866-017-0953-2.
21. Ambur OH, Engelstadter J, Johnsen PJ, Miller EL, Rozen DE. Steady at the wheel: conservative sex and the benefits of bacterial transformation. Philos T R Soc B 2016;371:20150528. doi: ARTN 2015052810.1098/rstb.2015.0528.
22. Slomka S, Francoise I, Hornung G, Asraf O, Biniashvili T, Pilpel Y, et al. Experimental evolution of Bacillus subtilis reveals the evolutionary dynamics of horizontal gene transfer and suggests adaptive and neutral effects. Genetics 2020;216:543–58. https://doi.org/10.1534/genetics.120.303401.
23. Power JJ, Pinheiro F, Pompei S, Kovacova V, Yuksel M, Rathmann I, et al. Adaptive evolution of hybrid bacteria by horizontal gene transfer. P Natl Acad Sci Usa 2021;118:e2007873118. doi: ARTN e200787311810.1073/pnas.2007873118.
24. Eyre-Walker A, Keightley PD. The distribution of fitness effects of new mutations. Nat Rev Genet. 2007;8:610–8.
25. Sane M, Gaurav DD, Bhat BA, Wahl LM, Agashe D. Shifts in mutation spectra enhance access to beneficial mutations. bioRvix. 2020. https://doi.org/10.1101/2020.09.05.284158.
26. Bondel KB, Kraemer SA, Samuels T, McClean D, Lachapelle J, Ness RW, et al. Inferring the distribution of fitness effects of spontaneous mutations in Chlamydomonas reinhardtii. PLoS Biol. 2019;17:e3000192 https://doi.org/10.1371/journal.pbio.3000192.
27. Dillon MM, Cooper VS. The fitness effects of spontaneous mutations nearly unseen by selection in a bacterium with multiple chromosomes. Genetics 2016;204:1225–38. https://doi.org/10.1534/genetics.116.193060.
28. Chevereau G, Dravecka M, Batur T, Guvenek A, Ayhan DH, Toprak E, et al. Quantifying the determinants of evolutionary dynamics leading to drug resistance. Plos Biol 2015;13:e1002299.
29. Gallet R, Cooper TF, Elena SF, Lenormand T. Measuring selection coefficients below 10(-3): method, questions, and prospects. Genetics 2012;190:175–U287. https://doi.org/10.1534/genetics.111.133454.
30. Robert L, Ollion J, Robert J, Song X, Matic I, Elez M. Mutation dynamics and fitness effects followed in single cells. Science 2018;359:1283–6. https://doi.org/10.1126/science.aan0797.
31. Kassen R, Bataillon T. Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat Genet 2006;38:484–8. https://doi.org/10.1038/ng1751.
32. Kirit HA, Lagator M, Bollback JP. Experimental determination of evolutionary barriers to horizontal gene transfer. Bmc Microbiol 2020;20:326. doi: ARTN 32610.1186/s12866-020-01983-5.
33. Sandberg TE, Szubin R, Phaneuf PV, Palsson BO. Synthetic cross-phyla gene replacement and evolutionary assimilation of major enzymes. Nat Ecol Evol 2020;4:1402–9. https://doi.org/10.1038/s41559-020-1271-x.
34. Dalia AB, Dalia TN. Spatiotemporal analysis of DNA integration during natural transformation reveals a mode of nongenetic inheritance in bacteria. Cell 2019;179:1499–511. https://doi.org/10.1016/j.cell.2019.11.021.
35. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–60. https://doi.org/10.1093/bioinformatics/btp324.
36. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience 2021;10:giab008.
37. Patrick JE, Kearns DB. MinJ (YvjD) is a topological determinant of cell division in Bacillus subtilis. Mol Microbiol 2008;70:1166–79. https://doi.org/10.1111/j.1365-2958.2008.06469.x.
38. Dunn OJ. Multiple Comparisons among Means. J Am Stat Assoc. 1961;56:52–64. https://doi.org/10.2307/2282330.
39. Fridman O, Goldberg A, Ronin I, Shoresh N, Balaban NQ. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. Nature 2014;513:418–21. https://doi.org/10.1038/nature13469.
40. Kearns DB, Losick R. Cell population heterogeneity during growth of Bacillus subtilis. Genes Dev 2005;19:3083–94. https://doi.org/10.1101/gad.1373905.
41. Pedreira T, Elfmann C, Stulke J. The current state of SubtiWiki, the database for the model organism Bacillus subtilis. Nucleic Acids Res. 2022;50:D875–D82. https://doi.org/10.1093/nar/gkab943.
42. Calvo RA, Kearns DB. FlgM is secreted by the flagellar export apparatus in Bacillus subtilis. J Bacteriol 2015;197:81–91. https://doi.org/10.1128/Jb.02324-14.
43. Mader U, Antelmann H, Buder T, Dahl MK, Hecker M, Homuth G. Bacillus subtilis functional genomics: genome-wide analysis of the DegS-DegU regulon by transcriptomics and proteomics. Mol Genet Genomics 2002;268:455–67. https://doi.org/10.1007/s00438-002-0774-2.
44. Lind PA, Tobin C, Berg OG, Kurland CG, Andersson DI. Compensatory gene amplification restores fitness after inter-species gene replacements. Mol Microbiol 2010;75:1078–89. https://doi.org/10.1111/j.1365-2958.2009.07030.x.
45. Nicholson WL. Increased competitive fitness of bacillus subtilis under non-sporulating conditions via inactivation of pleiotropic regulators AlsR, SigD, and SigW. Appl Environ Microb 2012;78:3500–3. https://doi.org/10.1128/Aem.07742-11.
46. Liu DF, Ai GM, Zheng QX, Liu C, Jiang CY, Liu LX, et al. Metabolic flux responses to genetic modification for shikimic acid production by Bacillus subtilis strains. Micro Cell Fact. 2014;13:40 https://doi.org/10.1186/1475-2859-13-40.
47. Yuksel M, Power JJ, Ribbe J, Volkmann T, Maier B. Fitness trade-offs in competence differentiation of Bacillus subtilis. Front Microbiol. 2016;7:888.
48. Graumann PL. Bacillus, cellular and molecular biology. Graumann PL, editor. Caister Academic Press; 2017.

ACKNOWLEDGEMENTS

We thank Gabriele Schneider for support with the generation of hybrid strains and Isabel Gordo and the Maier and Bollenbach labs for helpful discussions. This work was supported by the Deutsche Forschungsgemeinschaft through grant CRC 1310. We furthermore thank the Regional Computing Center of the University of Cologne (RRZK) for providing computing time on the DFG-funded (Funding number: INST 216/512/1FUGG) High Performance Computing (HPC) system CHEOPS as well as support.

AUTHOR CONTRIBUTIONS

IR, MF, MY, and BM designed research; IR, MF, and MY developed, performed, and analyzed the experiments; LH, GP, and TB designed and set up the protocols for the liquid handling robot; IR, MF, MY, and BM wrote the article; all authors discussed and approved the article.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

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/s41396-022-01325-5.

Correspondence and requests for materials should be addressed to Berenike Maier.

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.

6. Transformation opens up new paths for evolving *B. subtilis* populations

6.1. Manuscript

This manuscript reports the results of an on-going project between me (MF), Ariana Leu (AL), Melih Yüksel (MY), Isabel Rathmann (IR) and Berenike Maier (BM).

6.1.1. Contributions

The evolution experiment in liquid was developed by MF, IR and BM. MF adapted the lab strain to growth in structured environments. MY and AL established DNA barcoding. MF and AL shared the work on the evolution and the competition experiments in liquid medium equally. Both analysed the results. MF developed and performed the evolution experiment on plates and analysed the results. MF and IR isolated the gDNA for whole genome sequencing (WGS). MF analysed the WGS data. The manuscript was written by MF with corrections of BM.

6.1.2. Key findings

In the previous publication we demonstrated the importance of HGT during adaptive evolution (Sec. 5.1, [138]). In this project, we pre-adapt *B. subtilis* cells via mutations to two different conditions, growth in a liquid and in a structured environment, to understand the role of the adaptive history of the cells. We generate hybrid populations with both pre-adapted strains as recipient and perform evolution experiments with them in the two conditions, with one strain serving as a well-adapted strain and the other as a poorly-adapted strain. We find

(i) that transformation in liquid medium confers a fitness benefit to both, the poorly-adapted and the well-adapted populations without a replacement hotspot. This suggests that there are multiple beneficial replacements in both backgrounds that undergo clonal interference.

(ii) two exclusive ways to adapt to growth in a structured environment, by transformation and by mutations, respectively. In structured environments, we detect a fitness benefit of transformation for poorly-adapted populations with a replacement hotspot comprising seven genes, including *ymfK*, which is a pseudogene in *B. subtilis* but not in the donor. This makes it a good candidate for explaining the fitness advantage. In the poorly-adapted

recipient populations, we identify a mutational hotspot in *gudB*. None of the sequenced hybrid clones shows a mutation in *gudB*. In the well-adapted, transformed populations which have a mutation in *gudB* from the beginning, no replacement is found in the replacement hotspot. We hypothesize epistatic effects between the replacement hotspot and the mutation hotspot.

(iii) that none of the cells have lost their adaptive trait from the previous environment, suggesting rapid adaptation when the environment changes back to the old conditions.

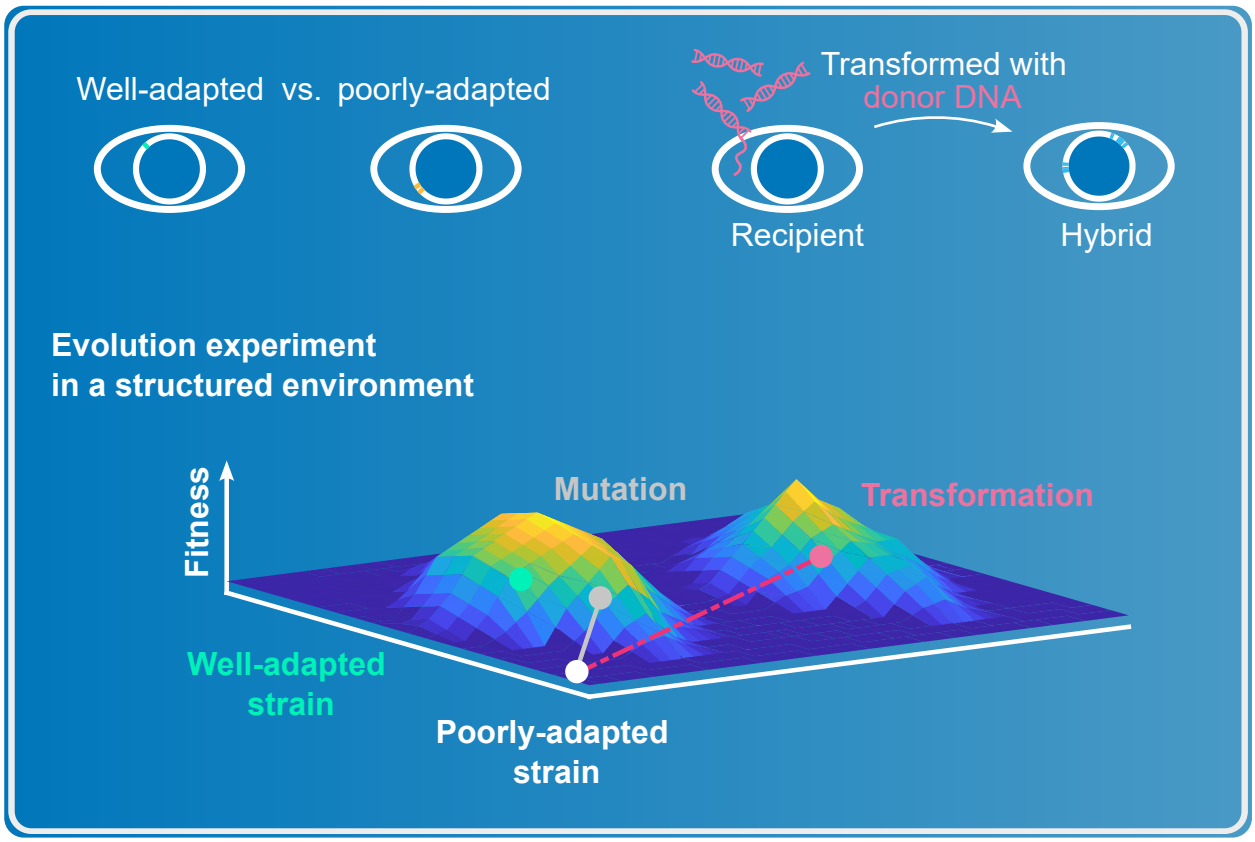


Figure 6.1.: Graphical abstract of the following manuscript. We pre-adapt the recipient to growth in a liquid and in a structured environment. In an evolution experiment in the structured environment, we find two exclusive ways in which the poorly-adapted recipients adapt via mutation and transformation, respectively. In none of the sequenced clones, the mutation and the transformation event are found together.

MANUSCRIPT

Transformation opens up new paths for evolving *Bacillus subtilis* populations

Mona Förster, Ariana Leu, Melih Yüksel, Isabel Rathmann, Berenike Maier

Abstract Horizontal gene transfer (HGT) is a key mechanism driving bacterial evolution. Nevertheless, the benefits of transformation, a mode of HGT, in changing environments are barely understood. In particular, it is unclear how the recent history of the cells influences the fitness effects of transformation. In this study, we aim at investigating the effect of pre-adaptation by mutations on the fitness benefit of HGT. To this end, we pre-adapt the recipient to exponential growth in liquid and to growth in a structured environment, and subsequently study the effects of HGT during evolution of well-adapted versus poorly-adapted bacteria in both environments. By comparing the distributions of selection coefficients between hybrid populations and recipient populations, we evaluate the benefit of HGT during evolution in the exponential growth phase in liquid environment. We find that HGT benefits both, well-adapted and poorly-adapted bacteria. To evaluate fitness effects of HGT during evolution in structured environments, we use the colony area as a measure of fitness and show that HGT benefits poorly-adapted bacteria. After evolution in structured environment, the poorly-adapted hybrid strains show a hotspot of orthologous replacement comprising seven genes. These genes do not overlap with the mutation hotspot discovered during pre-adaptation in the absence of HGT. We conclude that transformation has the potential to speed up adaptation to varying environments and opens up a new pathway for adaptive evolution.

Introduction

Horizontal gene transfer plays an important role in evolutionary dynamics. For all three domains of life, HGT has been detected, even some examples of cross-boundary transfer have been reported [1, 2]. Nevertheless, the impact of the cells' history of adaptive evolution on the fitness effects of transformation, a common mechanism of HGT, is poorly understood.

During natural transformation, competent cells take up DNA from their surroundings and integrate it into their genome via homologous recombination [3, 4]. At times, this can lead to the acquisition of novel genes but in the majority of cases a donor segment replaces a homologous region of the recipient [5]. In the arising hybrid, those donor segments appear in groups along the recipient genome and thus are often referred to as mosaic events [5–9]. In *Bacillus subtilis*, the main barrier to transformation is the sequence divergence between donor and recipient [5, 10, 11]. The transformation efficiency decreases exponentially with increasing sequence divergence. The integration of random segments of donor DNA can cause diverse costs for the transformant [12, 13] including the interruption of functional genes or networks [14], changes in protein folding [15] or in protein abundance [16, 17]. But still, the ability to transform is conserved in many species [18] and advantages of transformation are recorded repeatedly. In a clinical context, pathogens are able to escape treatments by transferring genes that confer antibiotic resistance via HGT [19]. When naturally competent cells growth in absence of other strains, transformation was shown to benefit evolving bacteria by combining beneficial mutations in one individual or by reducing the mutational load [20–23]. Various studies report fitness effects of external DNA during laboratory evolution of naturally competent bacteria and reach different conclusions. Utnes et al. showed that the additional supply of DNA of other species has no additive advantage [22]. By contrast, others show that cells benefit from transformation with DNA of closely related species when they are introduced to a new environment [14, 24, 25]. Slomka et al. found a benefit for naturally competent *B. subtilis* cells that evolved in high-salt containing liquid with DNA of closely related *Bacillus* species [24]. Because of the high number of replacements per replicate they suggest transfer events to have adaptive or neutral effects. Given the large variety of reported fitness effects under different growth conditions, we systematically characterized the distribution of fitness effects (DFE) of transformation hybrids of *B. subtilis* and closely related *Bacillus* species in various growth conditions [25]. We showed that transformation has the potential to speed up adaptation in most but not all growth environments tested. However,

it is unknown how the history of previous growth conditions affects the benefit of transformation. *B. subtilis* is a soil-dwelling species and as such faces strong seasonal variations of growth conditions.

In this study, we address the question how the growth history of B. subtilis affects the beneficial effect of transformation. Specifically, we pre-adapt the recipient to two different growth conditions through mutations. Growth in well-mixed environments with endless nutrient supply is far from bacterial growth in nature. In this study, we compare evolution in liquid and in structured environments of transformed and non-transformed cells that have been pre-adapted to one of the two conditions before (Fig. 1).

Material and methods

Strains

All used strains in this study are derived from strain Bs166 (*his leu met*, *amyE::PhscomK(spc)*, *comK::kan*, P_{comK}*gfp*) [14]. In this strain, the master regulator of competence, *comK*, is under the control of an IPTG-inducible promoter. As reference dictionary the NCBI dictionary NC_000964.3 is used. Bs210 is generated by adapting Bs166 to exponential growth phase in liquid medium (Fig. 1**A**). The only difference to Bs166 is a frame shift mutation in the coding-region of the flagellar gene sigma factor *sigD* (NC_000964.3:p.Phe151fs). Bs224 is generated by adapting Bs166 to growth on plate (Fig. 1**B**). Compared to the ancestor Bs166, Bs224 has single nucleotide polymorphisms (SNPs) in *rapA* (NC_000964.3:p.Pro261Pro), *gudB* (NC_000964.3:p.Phe396Lys) and *epsC* (NC_000964.3:p.Glu328Val) and an indel in *gudB* (NC_000964.3:p.Phe395fs) causing a frame shift. We generate the respective green-fluorescent reporter Bs211 (Bs210 *lacA::PrrnE-gfp* (erm)) and Bs226 (Bs224 *lacA::PrrnE-gfp* (erm)). These strains are used for fitness determination via competition.

Complex medium and culturing

The medium that serves as complex medium in this study is the "competence medium" for *B. subtilis* by Dubnau et al. [26]. For 500 ml of complex medium, 50 ml of 10x Spizizen's salts (60 g/l KH_2PO_4, 140 g/l K_2HPO_4, 20 g/l $(NH_4)_2SO_4$ and 10 g/l $Na_3C_6H_5O_7{\cdot}H_2O$) are supplemented with 0.1 g casamino acids, 0.5 g yeast extract, 0.25 g $MgCl_2{\cdot}6H_2O$ and 5 ml of 50 % glucose and filled up to 500 ml with sterile water. Because of histidine, leucine and methionine auxotrophy of our strains, media are supplied with 25 mg of each of theses amino acids. For liquid overnight cultures (ONC), cells are inoculated in complex medium and are shaken at 250 rpm at 37 ℃. Cells are stored at -80 ℃ mixed with 10 % DMSO.

Barcoding cells

Before Bs210 and Bs224 are used in the evolution experiments, we integrate random barcode sequences for tracking of population dynamics. The barcoding procedure was established by Ariana Leu and Melih Yüksel. Due to the barcoding, the cells contain an additional erythromycin resistance. The barcodes are not used for analyses in this manuscript.

Generation of hybrid populations

We generate hybrid populations with the barcoded strains Bs210 and Bs224, respectively. Therefore, the ONC of each of the strains is diluted 1:100 in complex medium and distributed into five wells of a 24-well microtiter plate containing 1 ml each. The cells are shaken in a Infinite M200 plate reader (Tecan, Männedorf, Switzerland) at 37 ℃ at 250 rpm. Growth is monitored by measuring the OD_{600}. After 2.5 hours, the cells escape the lag phase and begin to grow exponentially. For transformation, 600 μM IPTG is added to the cells to four out of five wells of the microtiter plate to induce competence. Additionally, genomic DNA (gDNA) of *Bacillus vallismortis* with one genome equivalent per recipient cell is added to two out of five wells. The gDNA is extracted beforehand using the Qiagen DNeasy Blood & Tissue Kit. The fragments have a length of predominantly 30 kbp. The cells transform for 2 hours. The populations in the two wells that receive donor DNA form the hybrid population. The populations in the two wells in which competence is induced but no DNA is added are the recipient populations in the following.

Evolution in liquid environment

For the evolution experiment in liquid environment, a high throughput experiment is set up at a automated system that allows to dilute each well of a 96-well plate independently (Fig. 1**C**). The automated system is integrated by the HighRes Biosolutions (HRB) company and consists of the following devices: A robotic arm (Acell, HRB), an incubator (StoreX STX44, Liconic), a liquid-handling device (Lynx LM900, Dynamic Devices), a plate reader (Synergy, BioTek Instruments), a shaker (BioShake 3000, QInstruments), a plate storage (NanoServe, HRS) and a delidding device (LidValet, HRS). At the start of the evolution experiments, the cells are inoculated in liquid complex medium at an optical density of about 0.1. Four 96-well microtiter plates run simultaneously: One plates contains

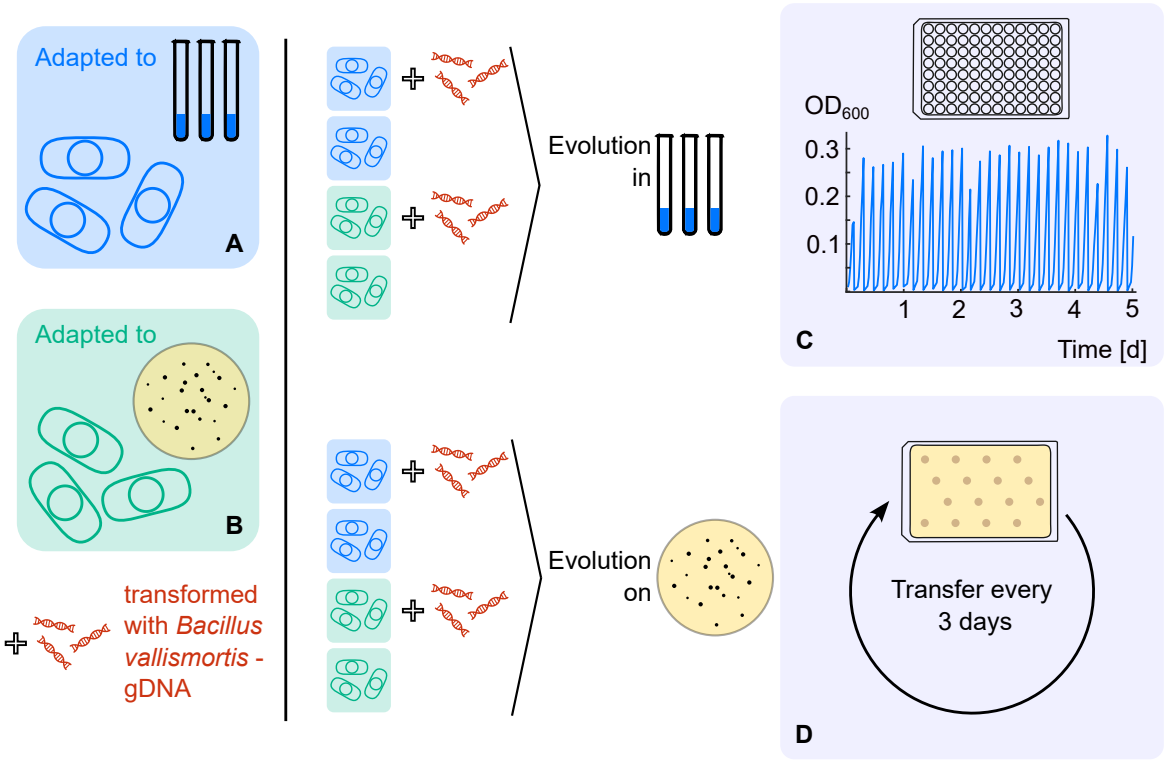


Figure 1: Evolution assay in liquid and on agar plates with cells adapted to growth in liquid environment (blue, **A**) and cells adapted to growth on agar plate (green, **B**). Both strains are transformed with *B. vallismortis* genomic DNA (gDNA) and then evolved in liquid environment (top, **C**) and on agar plates (bottom, **D**). Simultaneously, non-transformed recipient populations are evolved. This results into eight parallel evolution experiments, four in liquid environment and four on agar plates. In liquid environment, the four populations are spread over a 96-well microtiter plates each facilitating 88 parallel evolving populations (+ 8 blanks) per condition (**C**). Every four hours, the optical density per well is measured and each well is diluted back individually to an OD of 0.001. Thus, cells grow exponentially for five days. In the structured environment, 16 populations per plate evolve at the same time (**D**). Every three days, the biofilm is scratched from the plate and transferred to a fresh plate.

hybrid populations with barcoded Bs210 as ancestor (Bs210hyb), one plate contains recipient populations of barcoded Bs210 (Bs210ctl), the third plate contains hybrid populations with barcoded Bs224 as ancestor (Bs224hyb) and the fourth plate contains recipient populations of barcoded Bs224 (Bs224ctl). See the previous section for the description of how theses populations are generated. After every four hours of incubation at 37 ℃, the optical density of each well is measured and diluted likewise down to an OD of 0.001. This is how cells are kept in exponential growth phase over a period of five days. With the generation time of our strains of about 16 min (Supplements A, Fig. A.1), this equals 450 generations.

For pre-adaptation to growth in liquid environment, the same experimental setup was used [25]. In complex medium, a 96-well microtiter plate filled with Bs166 recipient populations was diluted every 4 hours down to an OD of 0.0008 for five days in a row ([25]: Supplementary methods - Generation of libraries of evolved hybrid populations).

Preparation of agar plates

As structured environment, we use complex medium agar plates supplemented with 1.5 % agar. For the agar plate handling robot, a rectangluar petri dish with 127.8 mm length and 85.5 mm heights is required. Each plate is filled with 38 ml.

Evolution in structured environments

For the evolution experiment in a structured environment, 1 μl drops of cells with an OD of 0.1 are plated on complex medium agar plates (1.5 % agar) and grown for 3 days at 30 ℃ (Fig. 1**D**). Per plate, 16 colonies grow in parallel. Simultaneously, four plates are handled with a agar plate handling system (BM3-BC, S&P Robotics Inc.): One plate with drops of hybrid populations with barcoded

Bs210 as ancestor (Bs210hyb), the second with recipient populations of barcoded Bs210 as control (Bs210ctl), the third plate with hybrid populations with barcoded Bs224 as ancestor (Bs224hyb) and their respective control, recipient populations of barcoded Bs224, on plate 4 (Bs224ctl). The colonies are delidded and imaged every 2 hours by the agar plate handling robot. After three days, 7 μl of complex medium is dropped on top of the biofilms to soften them. For the transfer of cells to fresh plates, we scratch the colony from the plate with the help of a 10 μl inoculation loop and resuspend the cells in liquid complex medium. 1 μl of this suspension is inoculated onto a fresh complex medium agar plate. This cycle is repeated five times. Time-lapse images of all colonies can be found in the Supplements (Supplements C, Fig. C.19 - C.20).

Pre-adaptation to growth in structured environments

The ancestor Bs166 is pre-adapted to growth on complex medium agar plates by dropping 1 μl cells with an OD of 0.1 onto the agar plates. Two plates with 16 colonies each are incubated simultaneously. The colonies grow at 30 ℃ for 3 to 4 days. They are imaged at intervals of 1-2 hours using the agar plate handling system. Time-lapse images of the pre-adaptation can be found in the Supplements (Supplement C, Fig. C.16 - C.17). Subsequently, the colonies are transferred to fresh CM agar plates as described in the previous section. For pre-adaptation, the cycle is repeated twelve times. We sequence six individual clones of the 32 pre-adapted populations and take one with representative genetic changes as "pre-adapted to growth in structured environment" (Fig. 2**B** hotspot table, all genetic variations in Supplements B, Fig. B.2). This clone is referred to below as Bs224.

Fitness of populations that evolved in liquid via competition experiments

Overnight cultures of the strains of interest (SOI) and a fluorescent competitor (GFP) are prepared by scratching cells from frozen stocks and inoculating them into complex medium. The cultures are incubated overnight at 37 ℃ shaking at 250 rpm. In the morning, they are diluted in fresh medium with a factor of 10. Overnight cultures and dilutions both happen in complex medium. To remove systematic biases from the positioning of the strain on the microtiter plate, we shift the columns of our plates randomly each morning before dilution. With the help of a plate reader the optical density is tracked until the cells escape lag phase and growth is visible. For complex medium at 37 ℃, this happens after 2 - 2.5 hours. Subsequently, the cells are diluted in PBS with the help of the robotic system to a fixed OD and mixed in a 50:50 ratio with the fluorescent reporter strain at an OD of 0.01. The exact starting fraction of SOI and fluorescent competitor in the mixes $N_{\mathrm{SOI},0}/N_{\mathrm{GFP},0}$ are measured using the flow cytometer. A 1:10 dilution of the mix in growth medium is grown under the selected conditions for four hours (about 16 generations). Subsequently, the cell mix is diluted 1:5 in PBS and the flow cytometer is used to measure the end fraction $N_{\mathrm{SOI}}/N_{\mathrm{GFP}}$. The selection coefficient of the SOI and the GFP $s_{\mathrm{SOI,\ GFP}}$ is calculated via $s_{\mathrm{SOI,\ GFP}} = \frac{t_{\mathrm{gen}}}{t_{\mathrm{meas}}} \cdot \frac{N_{\mathrm{SOI}}/N_{\mathrm{GFP}}}{N_{\mathrm{SOI},0}/N_{\mathrm{GFP},0}}$ where $N_{\mathrm{SOI/GFP}}$ is the number of SOI or GFP cells at time point t_{meas}, $N_{\mathrm{SOI,0/GFP,0}}$ at time point zero and t_{gen} the generation time of the ancestor. To correct for the difference of fitness between fluorescent competitor and ancestor (anc), also five ancestor samples are measured each run. The mean is used for correction: $s_{\mathrm{SOI,\ anc}} = s_{\mathrm{SOI,\ GFP}} - \langle s_{\mathrm{anc,\ GFP}} \rangle$. Since we barcoded the cells, we also need to correct for the difference of fitness between ancestor (anc) and barcoded ancestor (barAnc). Therefore, we measure the fitness difference of multiple barcodes and the ancestor in three independent experiments. As before, the mean is used for correction: $s_{\mathrm{SOI,\ barAnc}} = s_{\mathrm{SOI,\ anc}} - \langle s_{\mathrm{barAnc,\ anc}} \rangle$.

Filament correction of flow cytometry data

A subset of *B. subtilis* cells grow in filaments during the exponential growth phase [27]. Cells that are adapted to growth in liquid with a frame shift in *sigD* show a defect in autolysis. Unresolved filaments containing multiple cells are counted as one event by the flow cytometer. To correct for this, we identify a linear correlation between the parameter forward scatter width (FSC-W) and the average number of cells per filament (ACF) for the used setup: $\mathrm{ACF} = 0.0001 \cdot \mathrm{FSC\text{-}W} - 3.9661$ and use it to correct all flow cytometer raw data. In Supplements E, microscopy images of different strains in different growth phases are shown as well as their respective average FSC-W and number of cells per filament.

Colony size of populations that evolved on plates as a measure of fitness

Three-day-old biofilms are not easily solvable in liquid. Therefore, after competition on plates, flow cytometry is not an applicable method to measure the fraction of cells. Instead of competing cells with their ancestor, the increase of colony size during the evolution experiment is measured. For each cycle, the average occupied area per colony after 72 hours is evaluated. In Supplements D, examples of colony

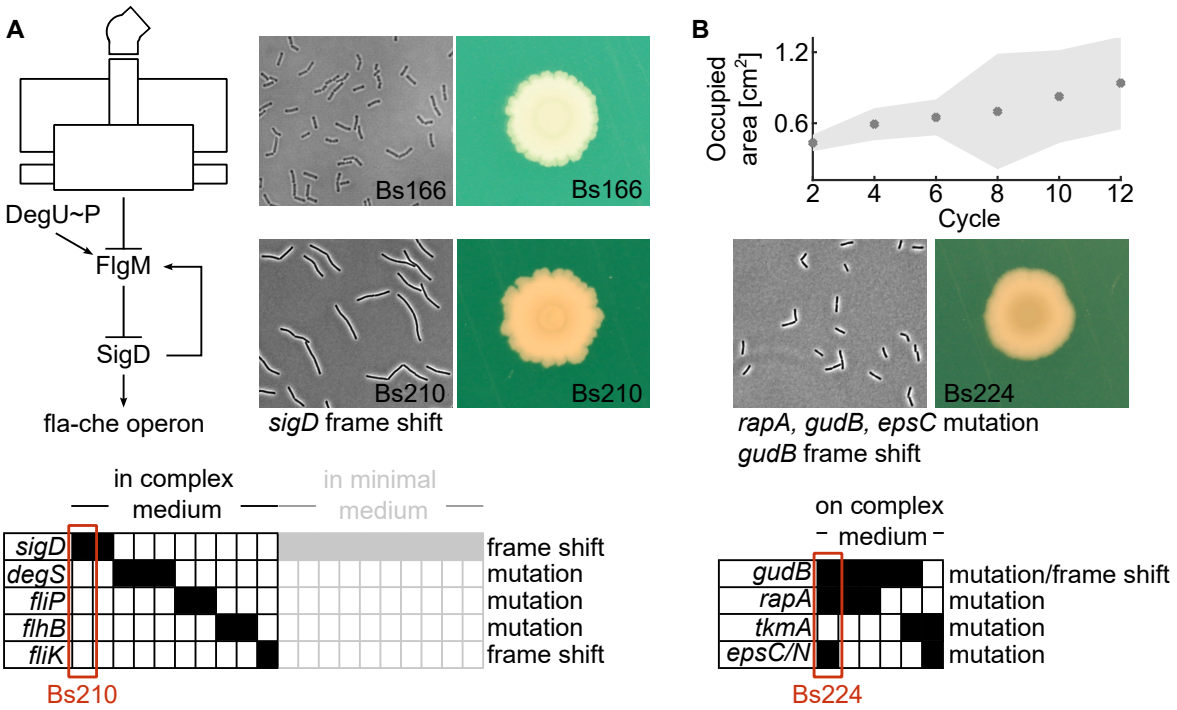


Figure 2: Adaptation to exponential growth in liquid (**A**) and to growth on agar plates (**B**). **A** The hotspot table at the bottom shows genetic variants found in an evolution experiment in one of our earlier publication where cells grew exponentially for 5 days in two different media [25]. All cells either showed a frame shift or a point mutation in the listed genes. All genes found are involved in the regulatory network of the flagellum-chemotaxis (fla-che) operon. Additionally, microscopy images of Bs166, the ancestor, and Bs210, the strain with *sigD* frame shift and their appearance on agar plates are shown. **B** During adaption to growth on plates, the average occupied area per colony increases with time. After twelve cycles, we sequence a clone of six of 32 pre-adapted colonies. The hotspot table at the bottom shows genes that are hit in at least two of the six clones by a mutation or frame shift. We choose Bs224 for the following experiment as "pre-adapted to growth on plates" which has mutations in *gudB*, *rapA* and *epsC* and a frame shift mutation in *gudB*. Additionally a microscopy image and an image of an agar plates are shown for Bs224.

images and their segmentation are shown.

DNA sequencing and analysis

For whole genome sequencing, gDNA of monoclonal samples is isolated with the Qiagen DNeasy Blood & Tissue-Kit and then sent to Eurofins Genomics, Germany, for 150 bp paired-read Illumina sequencing.

Results

Pre-adaptation of Bs166 to exponential growth in liquid environment targets regulatory genes of the flagellum-chemotaxis operon

In the course of the study "Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution" published in ISME Journal in 2023 [25], we let 20 samples of Bs166 grow exponentially for five days, 10 of them in complex medium and 10 in a minimal medium. In all strains that evolved for five days in liquid medium we find frame-shift or point mutations in the regulatory network of the flagellum-chemotaxis (fla-che) operon (Fig. 2**A**). Most prominent is a frame shift in flagellar gene-specific sigma factor σ_{D}. The frame shift occurs in all 10 sequenced strains that evolved in minimal medium and in two of 10 that evolved in complex medium. The σ_{D} frame shift leads to the loss of flagella and a defect in autolysis of the cells and consequently to the formation of extended filaments (Fig. 2**A**: compare microscopy images of Bs166 and Bs210). An inactivation of *sigD* has been shown to increase fitness in *Bacillus subtilis* before [28]. Other mutations occur in *degS*, *fliP*, *fliK* and *flhB* in two or more samples that evolved in parallel. *fliP*, *fliK* and *flhB* are known to be essential for the secretion of FlgM which inhibits the activity of σ_{D} [29], [30]. A secretion stop would increase the concentration of FlgM in the cell and hence, anticipate σ_{D}-dependent transcription. DegS also regulates the expression of the fla-che operon by phosphorelation of DegU [31]. DegU$\sim$P inhibits the fla-che operon. Therefore, we hypothesize

the mutations to be redundant [25]. For further experiments, a strain with *sigD*-frame shift is chosen as "pre-adapted to growth in liquid" (Bs210).

Mutations during adaptation of Bs166 to growth in structured environments target gudB and regulatory genes of sporulation and eps genes
For this study, we pre-adapt Bs166 to growth in a structured environment. Each cycle, we find increasing average colony size indicating the ongoing adaptation to the new growth environment (Fig. 2**B**). The colony size over time of the single colonies is shown in Supplements C (Fig. C.18). From six of the 32 populations that pre-adapted in parallel, we sequenced one clone each (Fig. 2**B**, all genetic variations in Supplements B, Fig. B.2). In five of six clones (83 %), we find a point mutation and/or a frame shift mutation in *gudB*. The glutamate dehydrogenase GudB is known to degrade glutamate [32]. In the ancestor, the gene is in frame suggesting that it is functional. A frame shift by one nucleotide suggests that functionality is lost. Additionally, we find mutations in *rapA* or *tkmA* in five of six sequenced clones. The rap (= response-regulator aspartyl phosphate) family is responsible for regulating phenotypic differentiation in *Bacillus* [33]. RapA dephosphorelates Spo0F∼P, an early sporulation gene of *B. subtilis*. Both proteins, TkmA and Spo0F are involved in the regulation of biofilm-associated sporulation [34, 35]. We suggest the mutations in those two genes to be redundant. Two of six sequenced clones have a mutation in an *eps* gene, namely *epsC* and *epsN*. Both have recently been shown to be involved in biosynthesis of bacillosamine [36, 37]. Bs224, the strain that is used for the following evolution experiments, has point mutations in *rapA*, *gudB* and *epsC* and a frame shift mutation in *gudB*. The total genetic differences between the two strains we use in the following, Bs210 (pre-adapted to growth in liquid) and Bs224 (pre-adapted to growth on plates), are 5 mutations (*rapA*, 3x *gudB*, *epsC*) and 2 frame shifts (*sigD*, *gudB*).

We aim at finding out how the history of previous growth conditions affects adaptation to new growth environments under HGT. To this end, we perform laboratory evolution experiments using the strains that are pre-adapted to growth in liquid and growth in structured environment, respectively.

Evolution in liquid vs. in structured environments
Transformation increases fitness in liquid environment
To find out how cross-species transformation affects adaptation to the exponential growth phase, strains Bs210 and Bs224 are grown exponentially for roughly 450 generations in liquid culture (Fig. 1**C**). Both strains are transformed with donor DNA prior to the evolution experiment and, therefore, thousands of different hybrid clones compete against each other during the experiment. For each recipient strain, we run a control experiment in which the bacteria are treated equally, yet without addition of donor DNA. This experiment results in 88 populations of each Bs210 hybrids and Bs224 hybrids, as well as the respective recipient populations. For each population, the selection coefficient relative to their respective ancestor is determined as described in the Methods (Fig. 3). The pre-adapted recipient populations Bs210ctl show a narrow distribution of selection coefficients with an average close to zero (Fig. 3**A-C**). The distribution of poorly-adapted recipient populations Bs224ctl shows a net shift towards positive selection coefficients (Fig. 3**D**) and the mean selection coefficient is significantly larger than zero (Fig. 3**E**). This indicates that the average fitness of the Bs224ctl populations has increased. Interestingly, the evolved, well-adapted Bs210hyb populations show a significantly broader distribution of selection coefficients with higher mean compared to the Bs210ctl recipient populations (Fig. 3**A-C**). This result shows that even after adaptation by mutation, transformation confers a benefit during adaptation to exponential growth. The distribution of selection coefficients of Bs224hyb shows a clear shift to positive values and its mean is significantly larger compared to the Bs224ctl distribution (Fig. 3**D-F**). This demonstrates that transformation accelerated adaptation of the strain that is poorly-adapted at the beginning of the evolution experiment. In summary, we find that transformation speeds up adaptation to exponential growth, both for well-adapted and poorly-adapted strains.

Next, we investigate the genome dynamics of the populations that evolved in liquid environment. To this end, a single, randomly picked clone of five populations is whole genome sequenced for each of the four conditions (Fig. 1**C**, all genetic variations in Supplements B, Fig. B.10 - B.15). This limited set of sequences does not show any hotspot of replacement, i.e. no gene shows signatures of orthologous replacement in more than one strain. On average, we find transformation events in 20 % of the Bs224hyb clones with 15 genes hit per clone and 40 % of the Bs210hyb clones with on average 11 genes affected. Next to orthologous replacement, we investigate *de novo* mutations

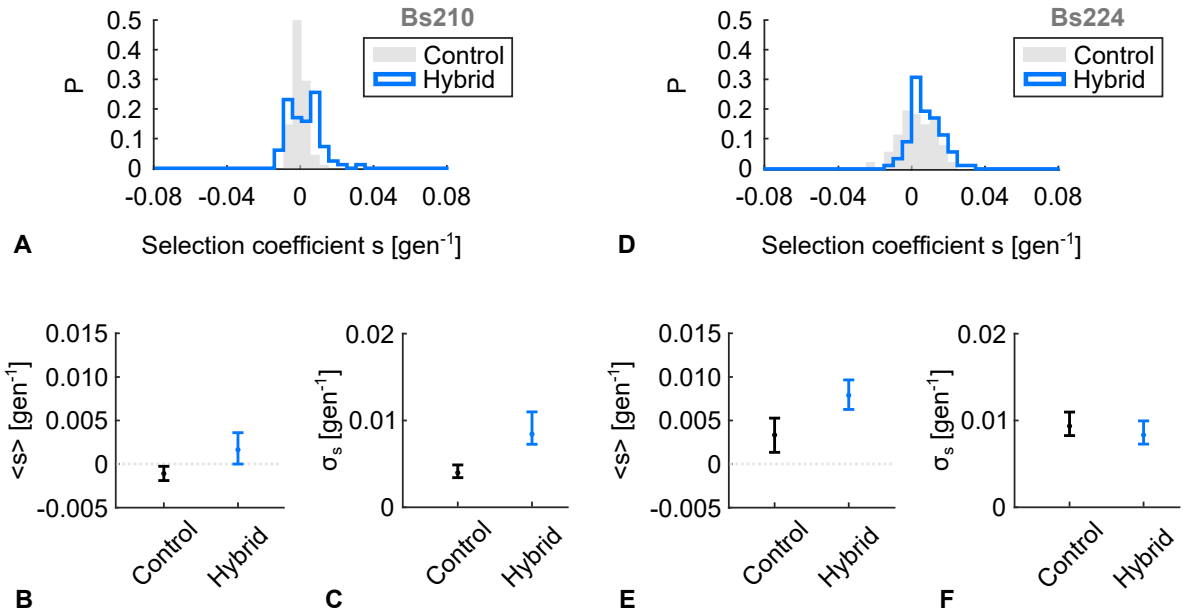


Figure 3: Fitness of populations that evolved in liquid medium for 5 days. **B**, **C**, **E** and **F** show means and 95 % bootstrap confidence intervals for the mean and the standard deviation of the distributions from **A** and **D**. Transformed populations (blue) of cells that were already well-adapted to growth in liquid **A** - **C** and cells that were pre-adapted to growth in structured environment **D** - **F** are compared to their respective controls (gray, no transformation).

and as during the pre-adaptation [25], we find that the poorly-adapted populations collect mutations in the fla-che operon during the evolution in liquid environment (Supplements Fig. B.12 - B.15). However, only 40 % of the control and 60 % of the sequenced hybrid clones show such a mutation. Rathmann, Förster et al. found mutations in all sequenced clones after five days of exponential growth in complex medium [25]. Finally, we find that none of the mutations related with pre-adapation to the respective growth conditions are lost, i.e. the *sigD* frame shift is present in all of the sequenced Bs210 strains and the Bs224 retained the mutations in *gudB*, *epsC* and *rapA*.

During growth in structured environments, transformation benefits poorly-adapted strain

Then we address the question how pre-adaptation affects evolution of transformed populations in structured environment. For 5 cycles, hybrid and recipient colonies of well-adapted and poorly-adapted populations are grown on complex medium plates at 30 ℃ for 3 days before they are transferred to fresh plates (Fig. 4**A**). At the start of the experiment, the pre-adapted strain Bs224 covers a considerably larger area than the poorly-adapted strain Bs210 (Fig. 4**B,C**). The average colony size of well-adapted cells (Bs224ctl) does not increase during evolution on plates (Fig. 4**B**). Even hybrid populations (Bs224hyb) do not increase their average size, suggesting that the hybrids do not benefit from having transformants among them. Consistently, whole genome sequencing (WGS) shows neither orthologous replacement in any of the 5 sequenced hybrid strains nor mutational hotspots in any of the 5 sequenced recipient strains (Supplements Fig. B.8 - B.9). By contrast, the average colony size of the poorly-adapted cells increases for both, hybrid (Bs210hyb) and recipient populations (Bs210ctl, Fig. 4**C**). For the hybrids, the increase starts immediately. The colony size over time of the single colonies is shown in Supplements C (Fig. C.21).

Hybrid populations use a different pathway to increase fitness during evolution on plates

Again, randomly picked clones from 5 populations are sequenced for each condition (all genetic variations in Supplements B, Fig. B.3 - B.9). The clones from the poorly-adapted recipient populations Bs210ctl show a hotspot of mutation in *gudB* (Fig. 4**D**, Supplements Fig. B.3), indicating that pre-adaptation to liquid has little effect on the benefit of this mutation for growth in structured environments. Importantly, this mutation is absent in all of the sequenced Bs210hyb strains (Supplements Fig. B.4 - B.7). Instead, in all five strains a replacement covering the same seven genes is detected. We conclude that transformation enables a different evolutionary trajectory during adaptation to growth in the structured environment. For the

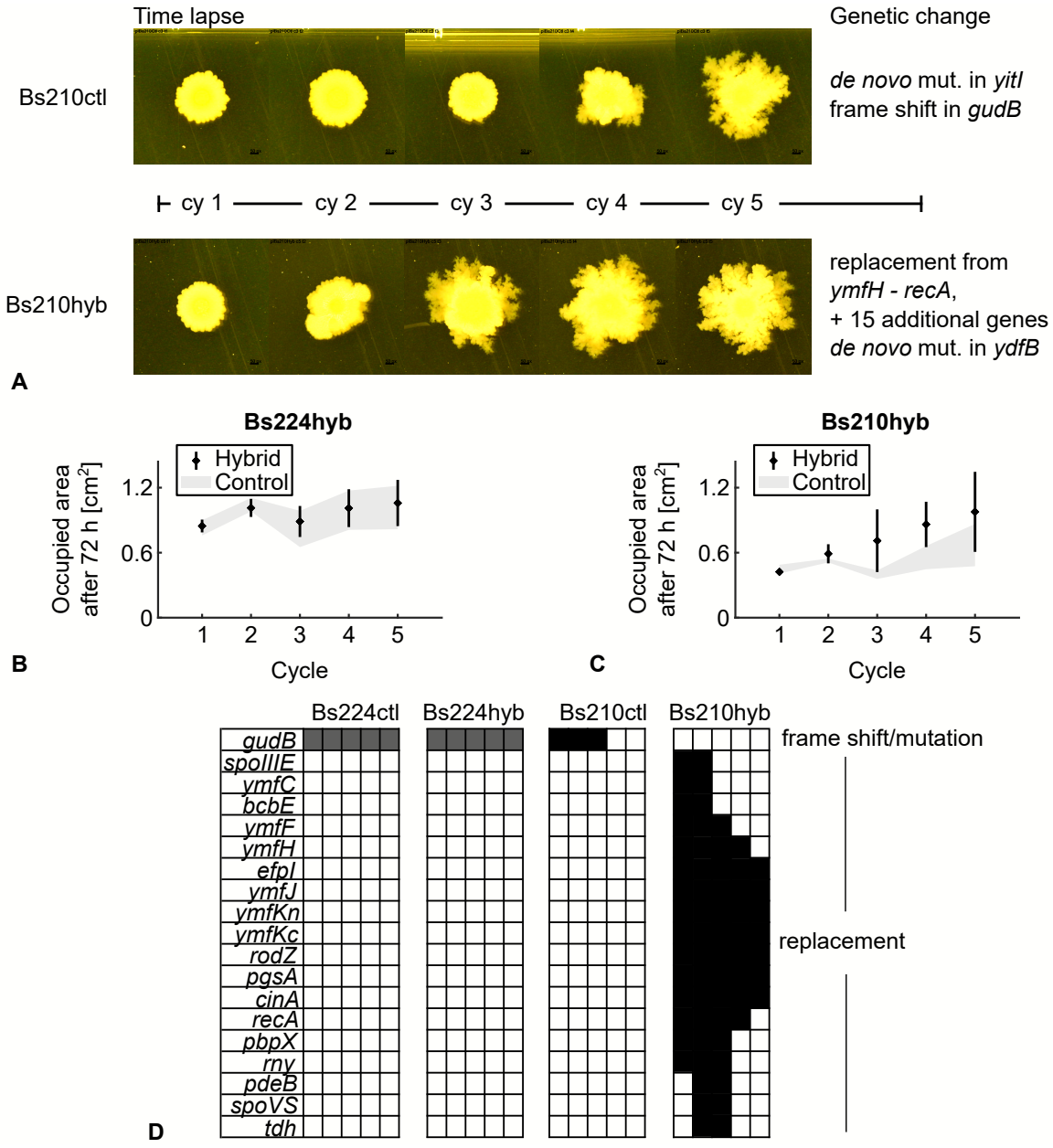


Figure 4: Evolution in structured environment. A shows one example for Bs210ctl and Bs210hyb populations each evolving on plates. The colonies are imaged after 72 hours of growth before they are transferred to fresh plates. Time lapse images for all Bs210hyb/ctl and Bs224hyb/ctl samples can be found in Supplements C. **B** and **C** show the temporal development of the average occupied area per colony after 72 hours on plate. The size of the recipient population colonies is shown in gray, the hybrid population colonies in black (mean + standard deviation). **D** shows the hotspots of frame-shift mutations, SNPs and orthologous replacements during the evolution experiment (black). We define a hotspot as a gene that has been affected by mutation or replacement in more than one sample. Gray boxes mark mutations that are already present from the pre-adaptation.

pre-adapted strain, no benefit of transformation is found. Interestingly, none of 20 sequenced clones looses the prior adaptive trait after changing the environment. After evolution in the environment they are not adapted to, 20 of 20 sequenced clones still have the mutations they got when adapting to the other environment (*sigD*/*gudB*,*rapA*,*epsC*).

Discussion

Adaptation to structured environment follows different trajectory in the presence of cross-species transformation

As the populations evolved in a structured environment we discovered two highly repeatable but distinct pathways of adaptation by mutation and by transformation. In the transformed populations, seven genes were replaced by *B. vallismortis* DNA in all sequenced replicates. Two potential candidates among the seven genes within the replacement hotspot, to increase fitness, are *ymfKn* and *ymfKc* which are the two ends of a *ymfK* gene. As they are split, they are pseudo genes in Bs166 [38, 39]. In the donor *B. vallismortis*, however, the gene is not interrupted. The function of YmfK is unknown, but it has an ACT domain, a binding domain for small molecules that is characteristic of metabolic enzymes regulated by amino acids [40, 41]. Kabisch et al. suggest that *ymfK* plays a role in more efficient utilization of amino acids [38]. Currently, we cannot exclude the possibility that other genes contribute to the fitness advantage. In the non-transformed populations, a mutation hotspot was found in the gene *gudB*. In none of the poorly-adapted, transformed populations which have the replacement hotspot a mutation in *gudB* was detected. By contrast, in the well-adapted, transformed strains, that already have the *gudB* mutation, none of the seven hotspot genes were replaced. Potentially, an epistatic effect leads to a remarkable decrease in the fitness advantage of the replacement and the mutation respectively if the other is already present. One reason for this could be that the two genetic variations are redundant. GudB is one of two glutamate dehydrogenases in the *B. subtilis* genome, RocG and GudB [42]. In several common laboratory strains, GudB is inactive due to a 9 bp repeat in the gene [43] and is reactivated when *rocG* is mutated [42, 44]. In a strain in which *rocG* is active, an active *gudB* leads to a more efficient utilization of amino acids of the glutamate family (arginine, ornithine and glutamine) [42]. Gunka et al. hypothesized that a cryptic *gudB* results in a selective growth advantage of bacteria under laboratory conditions [45]. The recipient used in this project does not show the 9 bp repeat. That suggests that *gudB* is active in the recipient at the start and potentially becomes cryptic due to the mutations during the evolution experiment. This does not indicate a redundancy of the mutation in *gudB* and the replacement of *ymfK*. Thus, we suggest that for the adaptive pathway of the cell it is important which genetic variant occurs first.

Colony size as fitness measure

As the colonies grown on agar plates are not soluble in liquid medium, competition experiments cannot be used to determine fitness of populations that have adapted to growth in structured environments. In this project, colony size is used as the measure of fitness. Miller et al. proposed to measure fitness via colony size for *Saccharomyces cerevisiae* [46]. They correlated the colony size with the cell number. However, this method has several short-comings. For colonies with inhomogeneous densities, the colony size is not proportional to the cell number. In addition, in our project cells are not diluted to a specific cell number when the cells are transferred to fresh agar plates, which leads to different cell numbers in each colony at the beginning of a growth period [47]. To estimate growth rate on agar plates, a method called phenotypic array analysis can be used [48, 49]. The cells are spotted on agar plates at low cell density. During the first generations, the intensity of the spots is in a linear relationship to the cell number. In our case, the change in colony size does not equal growth rate but during adaptation to a structured environment, there appears to be selection for larger colony sizes. Therefore, we suggest that the determination of the colony size in this context is a appropriate measure of the degree of adaptation.

No replacement hotspot in liquid

The poorly-adapted, transformed populations that grew on agar plates showed a replacement hotspot, whereas the poorly-adapted, transformed populations that grew in liquid medium did not. Either no replacement led to a high fitness increase/selective pressure was too low or clonal interference between multiple beneficial replacements prevents the domination of a specific replacement. In a recent publication, we measured the DFE of transformation in liquid complex medium with a poorly-adapted strain [25]. The DFE was distributed around zero and showed some positive and negative fitness outliers, suggesting a large number of potentially beneficial replacements. To identify highly beneficial transfers, we performed an evolution experiment with consecutive whole genome sequencing. Each sequenced strain contained a mutation in the fla-che operon that probably led to loss of flagellation. In

both experiments, this seems to be the predominant genetic variation, accompanied by some replacements. The lack of a hotspot is probably explained by several beneficial replacements undergoing clonal interference.

The adaptive history of the cells is remembered
In well-mixed, liquid environments, the cells with the highest growth rate are selected, whereas in structured environments like biofilms, the economical use of available resources can lead to decreased individual fitness [50]. In biofilms, the positioning of a cell plays an important role [51]. Inner cells are protected but cut off from nutrients. Cell that live in a biofilm must cooperate for the common good. This is what makes the two chosen environments so different. Nevertheless, all cells that switched from one environment to the other retained their adaptive traits. This suggests that the fitness cost of the prior adaptive trait is lower than the fitness advantage that an additional mutation or replacement brings. For constantly changing environments, this may give the cells an advantage as they still have the adaptive trait from the previous environment.

This study gives a first indication that the history of growth environments affects the benefit of transformation. The used methods show strong potential for investigating the role of gene transfer in fluctuating environments. In future studies, it would be interesting to investigate the effect of adaptive history in more and more complex and in frequently changing conditions.

Acknowledgments
We would like to thank Tobias Bollenbach and his group for making their agar plate handling system available to us, Janina Müller for her help in carrying out experiments with this system, and the Maier lab for helpful discussions.

References

[1] T. Dagan, Y. Artzy-Randrup, and W. Martin. “Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.29 (2008), pp. 10039–10044.

[2] L. Olendzenski and J. P. Gogarten. “Evolution of genes and organisms: the tree/web of life in light of horizontal gene transfer”. In: *Annals of the New York Academy of Sciences* 1178 (2009), pp. 137–145.

[3] D. Dubnau and M. Blokesch. “Mechanisms of DNA Uptake by Naturally Competent Bacteria”. In: *Annual review of genetics* 53 (2019), pp. 217–237.

[4] B. Maier. “Competence and Transformation in Bacillus subtilis”. In: *Current issues in molecular biology* 37 (2020), pp. 57–76.

[5] M. Förster, I. Rathmann, M. Yüksel, J. J. Power, and B. Maier. “Genome-wide transformation reveals extensive exchange across closely related Bacillus species”. In: *Nucleic acids research* 51.22 (2023), pp. 12352–12366.

[6] N. J. Croucher, S. R. Harris, L. Barquist, J. Parkhill, and S. D. Bentley. “A high-resolution view of genome-wide pneumococcal transformation”. In: *PLoS pathogens* 8.6 (2012), e1002745.

[7] E. A. Lin, X.-S. Zhang, S. M. Levine, S. R. Gill, D. Falush, and M. J. Blaser. “Natural transformation of helicobacter pylori involves the integration of short DNA fragments interrupted by gaps of variable size”. In: *PLoS pathogens* 5.3 (2009), e1000337.

[8] R. Mostowy, N. J. Croucher, W. P. Hanage, S. R. Harris, S. Bentley, and C. Fraser. “Heterogeneity in the frequency and characteristics of homologous recombination in pneumococcal evolution”. In: *PLoS genetics* 10.5 (2014), e1004300.

[9] L. Kennemann, X. Didelot, T. Aebischer, S. Kuhn, B. Drescher, M. Droege, R. Reinhardt, P. Correa, T. F. Meyer, C. Josenhans, D. Falush, and S. Suerbaum. “Helicobacter pylori genome evolution during human infection”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.12 (2011), pp. 5033–5038.

[10] M. S. Roberts and F. M. Cohan. “The effect of DNA sequence divergence on sexual isolation in Bacillus”. In: *Genetics* 134.2 (1993), pp. 401–408.

[11] B. Carrasco, E. Serrano, H. Sánchez, C. Wyman, and J. C. Alonso. “Chromosomal transformation in Bacillus subtilis is a non-polar recombination reaction”. In: *Nucleic acids research* 44.6 (2016), pp. 2754–2768.

[12] D. A. Baltrus. “Exploring the costs of horizontal gene transfer”. In: *Trends in ecology & evolution* 28.8 (2013), pp. 489–495.

[13] J. P. J. Hall, M. A. Brockhurst, and E. Harrison. “Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372.1735 (2017).

[14] J. J. Power, F. Pinheiro, S. Pompei, V. Kovacova, M. Yüksel, I. Rathmann, M. Förster, M. Lässig, and B. Maier. “Adaptive evolution of hybrid bacteria by horizontal gene transfer”. In: *Proceedings of the National Academy of Sciences of the United States of America* 118.10 (2021).

[15] S. Bershtein, W. Mu, A. W. R. Serohijos, J. Zhou, and E. I. Shakhnovich. “Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness”. In: *Molecular cell* 49.1 (2013), pp. 133–144.

[16] S. Bershtein, A. W. R. Serohijos, S. Bhattacharyya, M. Manhart, J.-M. Choi, W. Mu, J. Zhou, and E. I. Shakhnovich. “Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria”. In: *PLoS genetics* 11.10 (2015), e1005612.

[17] C. Park and J. Zhang. “High expression hampers horizontal gene transfer”. In: *Genome biology and evolution* 4.4 (2012), pp. 523–532.

[18] M. G. Lorenz and W. Wackernagel. “Bacterial gene transfer by natural genetic transformation in the environment”. In: *Microbiological reviews* 58.3 (1994), pp. 563–602.

[19] N. A. Lerminiaux and A. D. S. Cameron. “Horizontal transfer of antibiotic resistance genes in clinical environments”. In: *Canadian journal of microbiology* 65.1 (2019), pp. 34–44.

[20] D. A. Baltrus, K. Guillemin, and P. C. Phillips. “Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori”. In: *Evolution; international journal of organic evolution* 62.1 (2008), pp. 39–49.

[21] D. J. P. Engelmoer, I. Donaldson, and D. E. Rozen. “Conservative sex and the benefits of transformation in Streptococcus pneumoniae”. In: *PLoS pathogens* 9.11 (2013), e1003758.

[22] A. L. G. Utnes, V. Sørum, N. Hülter, R. Primicerio, J. Hegstad, J. Kloos, K. M. Nielsen, and P. J. Johnsen. “Growth phase-specific evolutionary benefits of natural transformation in Acinetobacter baylyi”. In: *The ISME journal* 9.10 (2015), pp. 2221–2231.

[23] R. J. Redfield. “Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all?” In: *Genetics* 119.1 (1988), pp. 213–221.

[24] S. Slomka, I. Françoise, G. Hornung, O. Asraf, T. Biniashvili, Y. Pilpel, and O. Dahan. “Experimental Evolution of Bacillus subtilis Reveals the Evolutionary Dynamics of Horizontal Gene Transfer and Suggests Adaptive and Neutral Effects”. In: *Genetics* 216.2 (2020), pp. 543–558.

[25] I. Rathmann, M. Förster, M. Yüksel, L. Horst, G. Petrungaro, T. Bollenbach, and B. Maier. “Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution”. In: *The ISME journal* 17.1 (2023), pp. 130–139.

[26] D. Dubnau and R. Davidoff-Abelson. “Fate of transforming DNA following uptake by competent Bacillus subtilis. I. Formation and properties of the donor-recipient complex”. In: *Journal of molecular biology* 56.2 (1971), pp. 209–221.

[27] D. B. Kearns and R. Losick. “Cell population heterogeneity during growth of Bacillus subtilis”. In: *Genes & development* 19.24 (2005), pp. 3083–3094.

[28] W. L. Nicholson. “Increased competitive fitness of Bacillus subtilis under nonsporulating conditions via inactivation of pleiotropic regulators AlsR, SigD, and SigW”. In: *Applied and environmental microbiology* 78.9 (2012), pp. 3500–3503.

[29] Q. O. Ababneh and J. K. Herman. “CodY Regulates SigD Levels and Activity by Binding to Three Sites in the fla/che Operon”. In: *Journal of bacteriology* 197.18 (2015), pp. 2999–3006.

[30] T. Caramori, D. Barilla, C. Nessi, L. Sacchi, and A. Galizzi. “Role of FlgM in sigma D-dependent gene expression in Bacillus subtilis”. In: *Journal of bacteriology* 178.11 (1996), pp. 3113–3118.

[31] R. Belas. “When the swimming gets tough, the tough form a biofilm”. In: *Molecular microbiology* 90.1 (2013), pp. 1–5.

[32] A. R. Fernie and Y. Zhang. “The Bacillus subtilis glutamate anti-metabolon”. In: *Nature metabolism* 4.2 (2022), pp. 161–162.

[33] R. Gallegos-Monterrosa and Á. T. Kovács. “Phenotypic plasticity: The role of a phosphatase family Rap in the genetic regulation of Bacilli”. In: *Molecular microbiology* 120.1 (2023), pp. 20–31.

[34] J. Gerwig, T. B. Kiley, K. Gunka, N. Stanley-Wall, and J. Stülke. “The protein tyrosine kinases EpsB and PtkA differentially affect biofilm formation in Bacillus subtilis”. In: *Microbiology (Reading, England)* 160.Pt 4 (2014), pp. 682–691.

[35] J. Gerwig and J. Stülke. “Far from being well understood: multiple protein phosphorylation events control cell differentiation in Bacillus subtilis at different levels”. In: *Frontiers in microbiology* 5 (2014), p. 704.

[36] C. R. Kaundinya, H. S. Savithri, K. K. Rao, and P. V. Balaji. “EpsN from Bacillus subtilis 168 has UDP-2,6-dideoxy 2-acetamido 4-keto glucose aminotransferase activity in vitro”. In: *Glycobiology* 28.10 (2018), pp. 802–812.

[37] C. R. Kaundinya, H. S. Savithri, K. Krishnamurthy Rao, and P. V. Balaji. “In vitro characterization of N-terminal truncated EpsC from Bacillus subtilis 168, a UDP-N-acetylglucosamine 4,6-dehydratase”. In: *Archives of biochemistry and biophysics* 657 (2018), pp. 78–88.

[38] J. Kabisch, A. Thürmer, T. Hübel, L. Popper, R. Daniel, and T. Schweder. “Characterization and optimization of Bacillus subtilis ATCC 6051 as an expression host”. In: *Journal of biotechnology* 163.2 (2013), pp. 97–104.

[39] T. Pedreira, C. Elfmann, and J. Stülke. “The current state of SubtiWiki, the database for the model organism Bacillus subtilis”. In: *Nucleic acids research* 50.D1 (2022), pp. D875–D882.

[40] G. A. Grant. “The ACT domain: a small molecule binding domain and its role as a common regulatory element”. In: *The Journal of biological chemistry* 281.45 (2006), pp. 33825–33829.

[41] D. M. Chipman and B. Shaanan. “The ACT domain family”. In: *Current opinion in structural biology* 11.6 (2001), pp. 694–700.

[42] B. R. Belitsky and A. L. Sonenshein. “Role and regulation of Bacillus subtilis glutamate dehydrogenase genes”. In: *Journal of bacteriology* 180.23 (1998), pp. 6298–6305.

[43] D. R. Zeigler, Z. Prágai, S. Rodriguez, B. Chevreux, A. Muffler, T. Albert, R. Bai, M. Wyss, and J. B. Perkins. “The origins of 168, W23, and other Bacillus subtilis legacy strains”. In: *Journal of bacteriology* 190.21 (2008), pp. 6983–6995.

[44] K. Gunka, S. Tholen, J. Gerwig, C. Herzberg, J. Stülke, and F. M. Commichau. “A high-frequency mutation in Bacillus subtilis: requirements for the decryptification of the gudB glutamate dehydrogenase gene”. In: *Journal of bacteriology* 194.5 (2012), pp. 1036–1044.

[45] K. Gunka and F. M. Commichau. “Control of glutamate homeostasis in Bacillus subtilis: a complex interplay between ammonium assimilation, glutamate biosynthesis and degradation”. In: *Molecular microbiology* 85.2 (2012), pp. 213–224.

[46] J. H. Miller, V. J. Fasanello, P. Liu, E. R. Longan, C. A. Botero, and J. C. Fay. “Using colony size to measure fitness in Saccharomyces cerevisiae”. In: *PloS one* 17.10 (2022), e0271709.

[47] A. Blomberg. “Measuring growth rate in high-throughput growth phenotyping”. In: *Current opinion in biotechnology* 22.1 (2011), pp. 94–102.

[48] J. L. Hartman and N. P. Tippery. “Systematic quantification of gene interactions by phenotypic array analysis”. In: *Genome biology* 5.7 (2004), R49.

[49] N. A. Shah, R. J. Laws, B. Wardman, L. P. Zhao, and J. L. Hartman. “Accurate, precise modeling of cell proliferation kinetics from time-lapse imaging and automated image analysis of agar yeast culture arrays”. In: *BMC systems biology* 1 (2007), p. 3.

[50] J.-U. Kreft. “Biofilms promote altruism”. In: *Microbiology (Reading, England)* 150.Pt 8 (2004), pp. 2751–2760.

[51] C. D. Nadell, K. Drescher, and K. R. Foster. “Spatial structure, cooperation and competition in biofilms”. In: *Nature reviews. Microbiology* 14.9 (2016), pp. 589–600.

7. Conclusion and outlook

In the first project, we found that the most frequent mode of transformation is orthologous replacement. Gene deletions occur at a much lower frequency and insertions even less frequently. Replacements facilitate deletions and insertions, and there are almost no restrictions on replacements found between closely-related *Bacillus* species. For this reason, we know that the DFE of an insertion library does not represent the DFE of transformation well [114]. In our second project, we measured DFEs of transformation under various conditions and found that the majority of hybrids are fitness neutral under most conditions. Despite the small sample size, we detected deleterious and beneficial fitness outliers and identified synergistic and antagonistic pleiotropy, suggesting that transformation accelerates adaptation and closely related species benefit from a shared gene pool. All DFEs were measured in well-mixed environments with poorly-adapted strains as recipient. To learn how the adaptive history of the cells affects the benefit of transformation and to gain a first insight into more natural growth environments, we conducted evolution experiments comparing poorly- and well-adapted strains in a liquid and a structured environment in the third project. We found a fitness advantage of transformation for poorly- and well-adapted strains in the liquid environment and in the structured environment, we found two exclusive pathways for adaptation by transformation and mutation, respectively. We hypothesize that transformation opens new pathways for the adaptation of cells to changing environments.

7.1. Testing species boundaries with replacement accumulation assays

When we investigated the barriers to transformation on the whole-genome level, we found exponentially decreasing transfer rate with increasing core genome sequence divergence of the donor-recipient pair (Fig. 7.1 **A**). This fits with the exponential decrease in the transformation efficiency of a gene with increasing sequence divergence [47, 48]. Carrasco et al. speculated about a divergence cut-off at higher sequence divergence of the genes [48]. When we used *G. thermoglucosidasius* as donor, no orthologous replacement was detected, although some regions of the *G. thermoglucosidasius* genome have homologues in the recipient genome. We have to consider that if the exponential relation still holds, we would probably have to run the assay with *G. thermoglucosidasius* as donor much longer to detect a replacement. We have calculated the theoretical transfer rate of *G. thermoglucosidasius* DNA to be approximately 0.57 $\frac{\mathrm{bp}}{\mathrm{h}}$ (Sec. 4.1, Fig. 7.1 **B**). To replace 100 bp of the recipient would then require around 175 h to arise. This results in

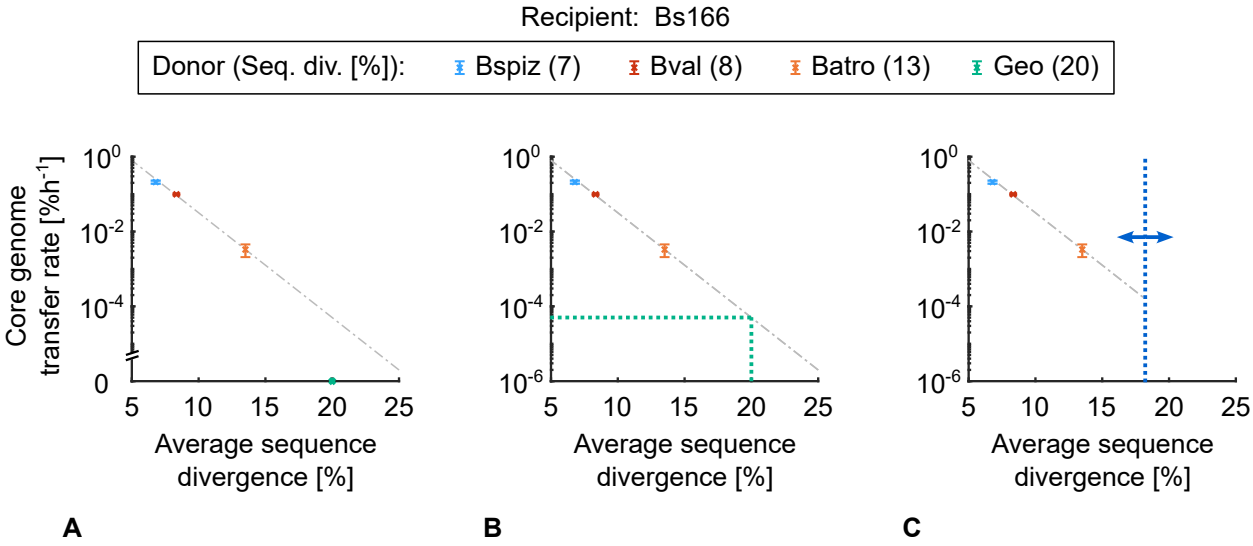


Figure 7.1.: Exponential relationship between core genome transfer rate and average sequence divergence of the donor-recipient pair. In project 1, we found a exponential relation between core genome transfer rate and average sequence divergence of the donor-recipient pair for three of four pairs (**A**, light blue: Bspiz donor, red: Bval donor, orange: Batro donor). When *G. thermoglucosidasius* was used as donor, no transfer was detected (green). If the exponential relation holds, the theoretical transfer rate of *G. thermoglucosidasius* DNA is approximately $1.2 \cdot 10^{-4}$ % h^{-1} of the core genome which corresponds to 0.57 $\frac{bp}{h}$ (**B**). If there is still no replacement after a sufficiently long run of the replacement accumulation assay, a potential species divergence barrier (dark blue) for transformation could be identified with further experiments (**C**).

a total of 88 cycles of the replacement accumulation assay to detect an average of 100 bp replaced. If we do not find a replacement, we could test whether we can identify a cut-off, as it was proposed by Carrasco et al. for genes, but between species (Fig. 7.1 **C**, [48]). Therefore, we could repeat the replacement accumulation assays with donor species that have an average sequence divergence higher than *B. atrophaeus* and smaller than *G. thermoglucosidasius* and the recipient.

Another open question is how far we can take the replacement accumulation. For the experiments we conducted, we found a still linearly increasing percentage of genome replaced (Sec. 4.1). We expect this relationship to flatten out with increasing amount of exchange, as the chance of replacing an already replaced region increases. Follow-up experiments could test when the curve starts to flatten and whether it is possible to exceed the 50 % (more donor, than recipient) or maybe even reach the 100 % (turn one species into the other). For this experiment, the donor with the highest transfer rate, namely *B. spizizenii*, is best suited.

7.2. Does transformation with gDNA from a different species protect from reintegration of phage elements?

In project 1, we found the spβ-prophage with a length of 134 kbp, comprising about 200 genes, deleted in one BspizHyb strain (Sec. 4.1, Fig. 7.2 BspizHyb). On both sides of the deleted segment, the flanking regions were replaced by a 17 kbp and a 7 kbp *B. spizizenii* homologue. The donor does not have the spβ-prophage [141]. Similarly, Croucher et al. found evidence for the removal of a mobile genetic element (MGE) through a transformation event in *Streptococcus pneumoniae* and suggest that transformation frequently cures strains of MGEs [142, 143]. In our BspizHyb, the integrating donor segment either led to the cut-out of the phage element or the phage excised itself and the replacement occurred on the remaining parts [144]. This part is difficult to retrace, but an interesting open question is why the strain was not reinfected with the spβ-phage. In further experiments, we want to use a phage deletion strain and the BspizHyb strain lacking the spβ-phage and test whether the replacement of the usual integration side (within the *spsM* gene) not only cures from the phage element but also protects the hybrid strain from reintegration (Fig. 7.2, [144, 145]).

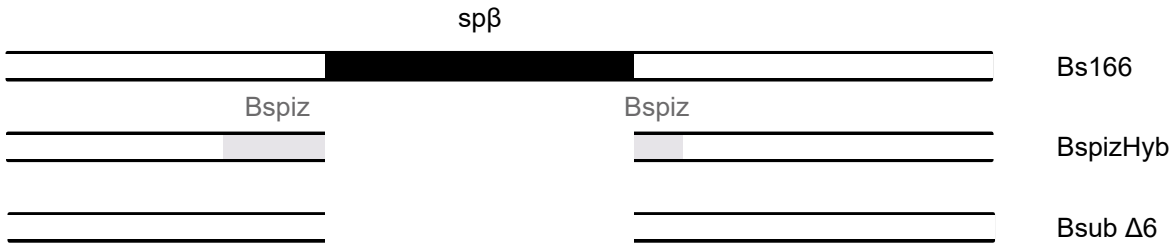


Figure 7.2.: The spβ-prophage in the *B. subtilis* genome. In one of the BspizHyb strains, the spβ-prophage (black) has been deleted from the genome and we detected replacements by the donor homologues in the flanking regions (light gray). Using a phage deletion strain from Westers et al. [145], we want to find out whether the replacement only removes the phage element from the hybrid or also protects it from reintegration.

7.3. The distribution of fitness effects of bacterial transformation and its potential for predicting evolutionary processes

In project 2, we measured DFEs of transformation under different conditions with libraries containing about 90 replicates. We found several conditions under which the DFE was centered around zero, with some additional positive and negative outliers. In one condition, the DFE was shifted to negative values with no positive outlier. By means of evolution experiments, we showed that the DFEs were highly useful for predicting in which condition transformation is beneficial.

A smooth DFE can be well characterized with our method. An improvement in precision could be achieved by increasing the number of replicates by an order of magnitude,

as it was done for gene deletions [113]. On the other hand, the characterization of a DFE with rare beneficial outliers with large fitness effects pose a problem. These remain undiscovered if only a small subset of all possible configurations is measured, but have a major impact on adaptive evolution. To reveal beneficial large effect outliers, evolution experiments must be performed such as those conducted in project 3. By means of WGS, these beneficial transfers are then detected as hotspots.

To predict the effect of transformation on an evolution experiment, it is not only important to know the DFE, but also the rate at which transformation occurs, as determined in project 1. If the transfer of only one gene is responsible for the largest fitness effect, the rate can be used to estimate, for example, the minimum size of the hybrid populations to start with in order to have at least one hybrid with this transfer in the population. This minimum size ranges from a few thousand hybrids when *B. spizizenii* is the donor to hundreds of thousands with *B. atrophaeus* as the donor. It must also be remembered that in the less closely related species, transformation was not found everywhere in the core genome, but only in regions with above-average sequence identity. In the structured environment, we found the same replacement hotspot in all sequenced replicates. This indicates that the hybrid number was high enough so that in all populations a hybrid with the largest effect outlier was present. The large effect tail of the corresponding DFE can be characterized in more detail when starting with a lower number of hybrids. In the liquid environment, we do not see a hotspot. Either the diversity of hybrids was lower at the start compared to the experiment in structured environment or the beneficial outliers have similar fitness advantages, undergo clonal interference and the populations are diverse. Random barcoding will provide information about the dynamics of population diversity in future experiments.

7.4. Population dynamics of transformed populations via barcode sequencing

In projects 2 and 3, we conducted evolution experiments. In project 2, the fitness of 88 individual clones from the evolved populations was measured via competition. Ten of these clones were whole genome sequenced. In project 3, we moved from measuring the fitness of individual clones to measuring the population fitness of all evolved populations. That improved the representativeness of the measured fitness value. Again, single clones were sent for whole genome sequencing. This time there were five clones per experiment. If one clone has overtaken the whole population during the evolution experiment, sequencing single clones gives a good insight into the genetic changes. But we have no information about population diversity and dynamics. To find this out, we could use time-resolved barcode sequencing methods [90–92]. Time-resolved barcode sequencing provides information about the diversity of the population, but can also be used to calculate the fitness of the most abundant clones in the population. Using barcode sequencing, we could learn about the clonal interference and the timing of fixation of transfer events and mutations. In project 3, we have barcoded the cells before running

the laboratory evolution experiment but we have not sequenced the barcodes yet.

At the moment, we can only speculate on the diversity of populations. A broad distribution of selection coefficients like we found in project 2 after the evolution experiment in liquid complex medium, could indicate a high cell diversity, but it can also mean that we started with a low number of hybrids and different fittest clones have overtaken the populations. A small variance in selection coefficients like we found in project 2 after the evolution experiment in liquid minimal medium, could be explained by a low diversity, where the same replacement or mutation swept through the populations. The sequencing data provides more information about the population diversity. The discovery of a hotspot of genetic variation in the sequenced clones suggests a variant that has already spread through the entire population. As discussed before (Sec. 7.3), the lack of a hotspot, as in both projects in liquid competence medium, could mean that multiple hybrids have similar fitness advantages and therefore still rival. This would lead to a high diversity. Or the cell number at the start was so low that each population was overtaken by another fittest clone. Since competence in our strain is inducible and does not occur randomly, beneficial transfers cannot be merged into one hybrid by transformation. It would be interesting to see how population dynamics change when transformation is induced from time to time, either with no additional DNA so that rivaling genetic variants can merge, or with addition of donor DNA.

7.5. Adaptation to frequently changing environments

In project 3, we found that all pre-adapted strains that changed environments retained their adaptive trait from the previous environment. This seems to be very helpful in environments in which conditions change back and forth, such as seasonal changes. In further experiments, the conditions could be changed more frequently and the ”memory” of the cells could be tested further. Is it really just the fitness that makes the cells retain the adaptive property of their previous environment or can they somehow remember that this genetic change was beneficial a short time ago and could be useful when the condition returns?

In this work, we have answered many open questions about transformation and its effects on adaptive evolution, but we have also encountered many more questions that we will try to answer in the future.

Bibliography

[1] S. Mitri and K. R. Foster. “The genotypic view of social interactions in microbial communities”. In: *Annual review of genetics* 47 (2013), pp. 247–273.

[2] N. A. Lerminiaux and A. D. S. Cameron. “Horizontal transfer of antibiotic resistance genes in clinical environments”. In: *Canadian journal of microbiology* 65.1 (2019), pp. 34–44.

[3] C. E. Rouquette-Loughlin, J. L. Reimche, J. T. Balthazar, V. Dhulipala, K. M. Gernert, E. N. Kersh, C. D. Pham, K. Pettus, A. J. Abrams, D. L. Trees, S. St Cyr, and W. M. Shafer. “Mechanistic Basis for Decreased Antimicrobial Susceptibility in a Clinical Isolate of Neisseria gonorrhoeae Possessing a Mosaic-Like mtr Efflux Pump Locus”. In: *mBio* 9.6 (2018).

[4] D. R. Evans, M. P. Griffith, A. J. Sundermann, K. A. Shutt, M. I. Saul, M. M. Mustapha, J. W. Marsh, V. S. Cooper, L. H. Harrison, and D. van Tyne. “Systematic detection of horizontal gene transfer across genera among multidrug-resistant bacteria in a single hospital”. In: *eLife* 9 (2020).

[5] S. Peter, M. Bosio, C. Gross, D. Bezdan, J. Gutierrez, P. Oberhettinger, J. Liese, W. Vogel, D. Dörfel, L. Berger, M. Marschal, M. Willmann, I. Gut, M. Gut, I. Autenrieth, and S. Ossowski. “Tracking of Antibiotic Resistance Transfer and Rapid Plasmid Evolution in a Hospital Setting by Nanopore Sequencing”. In: *mSphere* 5.4 (2020).

[6] T. Dagan, Y. Artzy-Randrup, and W. Martin. “Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.29 (2008), pp. 10039–10044.

[7] L. Olendzenski and J. P. Gogarten. “Evolution of genes and organisms: the tree/web of life in light of horizontal gene transfer”. In: *Annals of the New York Academy of Sciences* 1178 (2009), pp. 137–145.

[8] S. M. Soucy, J. Huang, and J. P. Gogarten. “Horizontal gene transfer: building the web of life”. In: *Nature reviews. Genetics* 16.8 (2015), pp. 472–482.

[9] C. Darwin. *Darwin C. 1859. On the origin of species by means of natural selection, On the origin of species by means of natural selection or preservation of favoured races in the struggle for life.* London, United Kingdom: J Murray, 1859.

[10] B. J. Arnold, I.-T. Huang, and W. P. Hanage. “Horizontal gene transfer and adaptive evolution in bacteria”. In: *Nature reviews. Microbiology* 20.4 (2022), pp. 206–218.

[11] N. D. ZINDER and J. LEDERBERG. “Genetic exchange in Salmonella”. In: *Journal of bacteriology* 64.5 (1952), pp. 679–699.

[12] J. LEDERBERG and E. L. TATUM. “Gene recombination in Escherichia coli”. In: *Nature* 158.4016 (1946), p. 558.

[13] M. Llosa, F. X. Gomis-Rüth, M. Coll, and F. de La Cruz Fd. “Bacterial conjugation: a two-step mechanism for DNA transport”. In: *Molecular microbiology* 45.1 (2002), pp. 1–8.

[14] G. P. Dubey and S. Ben-Yehuda. “Intercellular nanotubes mediate bacterial communication”. In: *Cell* 144.4 (2011), pp. 590–600.

[15] G. P. Dubey, G. B. Malli Mohan, A. Dubrovsky, T. Amen, S. Tsipshtein, A. Rouvinski, A. Rosenberg, D. Kaganovich, E. Sherman, O. Medalia, and S. Ben-Yehuda. “Architecture and Characteristics of Bacterial Nanotubes”. In: *Developmental cell* 36.4 (2016), pp. 453–461.

[16] A. S. Lang, O. Zhaxybayeva, and J. T. Beatty. “Gene transfer agents: phage-like elements of genetic exchange”. In: *Nature reviews. Microbiology* 10.7 (2012), pp. 472–482.

[17] S. N. Chatterjee and J. Das. “Electron microscopic observations on the excretion of cell-wall material by Vibrio cholerae”. In: *Journal of general microbiology* 49.1 (1967), pp. 1–11.

[18] A. Matin and W. N. Konings. “Transport of lactate and succinate by membrane vesicles of Escherichia coli, Bacillus subtilis and a pseudomonas species”. In: *European journal of biochemistry* 34.1 (1973), pp. 58–67.

[19] F. Griffith. “The Significance of Pneumococcal Types”. In: *The Journal of hygiene* 27.2 (1928), pp. 113–159.

[20] M. G. Lorenz and W. Wackernagel. “Bacterial gene transfer by natural genetic transformation in the environment”. In: *Microbiological reviews* 58.3 (1994), pp. 563–602.

[21] C. Johnston, B. Martin, G. Fichant, P. Polard, and J.-P. Claverys. “Bacterial transformation: distribution, shared mechanisms and divergent control”. In: *Nature reviews. Microbiology* 12.3 (2014), pp. 181–196.

[22] B. Maier. “Competence and Transformation in Bacillus subtilis”. In: *Current issues in molecular biology* 37 (2020), pp. 57–76.

[23] D. Dubnau. “DNA uptake in bacteria”. In: *Annual review of microbiology* 53 (1999), pp. 217–244.

[24] D. Dubnau and M. Blokesch. “Mechanisms of DNA Uptake by Naturally Competent Bacteria”. In: *Annual review of genetics* 53 (2019), pp. 217–237.

[25] B. J. Haijema, J. Hahn, J. Haynes, and D. Dubnau. “A ComGA-dependent checkpoint limits growth during the escape from competence”. In: *Molecular microbiology* 40.1 (2001), pp. 52–64.

[26] E. W. Nester and B. A. STOCKER. “Biosynthetic latency in early stages of deoxyribonucleic acid transformation in Bacillus subtilis”. In: *Journal of bacteriology* 86.4 (1963), pp. 785–796.

[27] A. B. Dalia and T. N. Dalia. “Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria”. In: *Cell* 179.7 (2019), 1499–1511.e10.

[28] I. Chen and D. Dubnau. “DNA uptake during bacterial transformation”. In: *Nature reviews. Microbiology* 2.3 (2004), pp. 241–249.

[29] R. J. Redfield. “Evolution of natural transformation: testing the DNA repair hypothesis in Bacillus subtilis and Haemophilus influenzae”. In: *Genetics* 133.4 (1993), pp. 755–761.

[30] R. J. Redfield. “Do bacteria have sex?” In: *Nature reviews. Genetics* 2.8 (2001), pp. 634–639.

[31] R. E. Michod, M. F. Wojciechowski, and M. A. Hoelzer. “DNA repair and the evolution of transformation in the bacterium Bacillus subtilis”. In: *Genetics* 118.1 (1988), pp. 31–39.

[32] M. F. Wojciechowski, M. A. Hoelzer, and R. E. Michod. “DNA repair and the evolution of transformation in Bacillus subtilis. II. Role of inducible repair”. In: *Genetics* 121.3 (1989), pp. 411–422.

[33] M. A. Hoelzer and R. E. Michod. “DNA repair and the evolution of transformation in Bacillus subtilis. III. Sex with damaged DNA”. In: *Genetics* 128.2 (1991), pp. 215–223.

[34] R. J. Redfield. “Genes for breakfast: the have-your-cake-and-eat-it-too of bacterial transformation”. In: *The Journal of heredity* 84.5 (1993), pp. 400–404.

[35] J. S. Kroll, K. E. Wilks, J. L. Farrant, and P. R. Langford. “Natural genetic exchange between Haemophilus and Neisseria: intergeneric transfer of chromosomal genes between major human pathogens”. In: *Proceedings of the National Academy of Sciences of the United States of America* 95.21 (1998), pp. 12381–12385.

[36] S. Suerbaum, J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunstmann, I. Dyrek, and M. Achtman. “Free recombination within Helicobacter pylori”. In: *Proceedings of the National Academy of Sciences of the United States of America* 95.21 (1998), pp. 12619–12624.

[37] R. J. Redfield. “Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all?” In: *Genetics* 119.1 (1988), pp. 213–221.

[38] D. B. Danner, R. A. Deich, K. L. Sisco, and H. O. Smith. “An eleven-base-pair sequence determines the specificity of DNA uptake in Haemophilus transformation”. In: *Gene* 11.3-4 (1980), pp. 311–318.

[39] S. D. Goodman and J. J. Scocca. “Identification and arrangement of the DNA sequence recognized in specific transformation of Neisseria gonorrhoeae”. In: *Proceedings of the National Academy of Sciences of the United States of America* 85.18 (1988), pp. 6982–6986.

[40] O. H. Ambur, S. A. Frye, and T. Tønjum. “New functional identity for the DNA uptake sequence in transformation and its presence in transcriptional terminators”. In: *Journal of bacteriology* 189.5 (2007), pp. 2077–2085.

[41] A. Cehovin, P. J. Simpson, M. A. McDowell, D. R. Brown, R. Noschese, M. Pallett, J. Brady, G. S. Baldwin, S. M. Lea, S. J. Matthews, and V. Pelicic. “Specific DNA recognition mediated by a type IV pilin”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.8 (2013), pp. 3065–3070.

[42] C. Hadden and E. W. Nester. “Purification of competent cells in the Bacillus subtilis transformation system”. In: *Journal of bacteriology* 95.3 (1968), pp. 876–885.

[43] F. H. Cahn and M. S. Fox. “Fractionation of transformable bacteria from ocompetent cultures of Bacillus subtilis on renografin gradients”. In: *Journal of bacteriology* 95.3 (1968), pp. 867–875.

[44] S. Ikawa, T. Shibata, T. Ando, and H. Saito. “Host-controlled modification and restriction in Bacillus subtilis: Bsu 168-system and BsuR-system in B. subtilis 168”. In: *Molecular & general genetics : MGG* 170.2 (1979), pp. 123–127.

[45] S. Jentsch. “Restriction and modification in Bacillus subtilis: sequence specificities of restriction/modification systems BsuM, BsuE, and BsuF”. In: *Journal of bacteriology* 156.2 (1983), pp. 800–808.

[46] C. D. Johnston, S. L. Cotton, S. R. Rittling, J. R. Starr, G. G. Borisy, F. E. Dewhirst, and K. P. Lemon. “Systematic evasion of the restriction-modification barrier in bacteria”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.23 (2019), pp. 11454–11459.

[47] M. S. Roberts and F. M. Cohan. “The effect of DNA sequence divergence on sexual isolation in Bacillus”. In: *Genetics* 134.2 (1993), pp. 401–408.

[48] B. Carrasco, E. Serrano, H. Sánchez, C. Wyman, and J. C. Alonso. “Chromosomal transformation in Bacillus subtilis is a non-polar recombination reaction”. In: *Nucleic acids research* 44.6 (2016), pp. 2754–2768.

[49] D. A. Baltrus. “Exploring the costs of horizontal gene transfer”. In: *Trends in ecology & evolution* 28.8 (2013), pp. 489–495.

[50] J. P. J. Hall, M. A. Brockhurst, and E. Harrison. “Sampling the mobile gene pool: innovation via horizontal gene transfer in bacteria”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 372.1735 (2017).

[51] J. G. Bragg and A. Wagner. “Protein material costs: single atoms can make an evolutionary difference”. In: *Trends in genetics : TIG* 25.1 (2009), pp. 5–8.

[52] J. J. Power, F. Pinheiro, S. Pompei, V. Kovacova, M. Yüksel, I. Rathmann, M. Förster, M. Lässig, and B. Maier. “Adaptive evolution of hybrid bacteria by horizontal gene transfer”. In: *Proceedings of the National Academy of Sciences of the United States of America* 118.10 (2021).

[53] S. Bershtein, W. Mu, A. W. R. Serohijos, J. Zhou, and E. I. Shakhnovich. “Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness”. In: *Molecular cell* 49.1 (2013), pp. 133–144.

[54] S. Bershtein, A. W. R. Serohijos, S. Bhattacharyya, M. Manhart, J.-M. Choi, W. Mu, J. Zhou, and E. I. Shakhnovich. “Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria”. In: *PLoS genetics* 11.10 (2015), e1005612.

[55] C. Park and J. Zhang. “High expression hampers horizontal gene transfer”. In: *Genome biology and evolution* 4.4 (2012), pp. 523–532.

[56] T. Tuller, Y. Girshovich, Y. Sella, A. Kreimer, S. Freilich, M. Kupiec, U. Gophna, and E. Ruppin. “Association between translation efficiency and horizontal gene transfer within microbial communities”. In: *Nucleic acids research* 39.11 (2011), pp. 4743–4755.

[57] D. A. Baltrus, K. Guillemin, and P. C. Phillips. “Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori”. In: *Evolution; international journal of organic evolution* 62.1 (2008), pp. 39–49.

[58] D. J. P. Engelmoer, I. Donaldson, and D. E. Rozen. “Conservative sex and the benefits of transformation in Streptococcus pneumoniae”. In: *PLoS pathogens* 9.11 (2013), e1003758.

[59] A. L. G. Utnes, V. Sørum, N. Hülter, R. Primicerio, J. Hegstad, J. Kloos, K. M. Nielsen, and P. J. Johnsen. “Growth phase-specific evolutionary benefits of natural transformation in Acinetobacter baylyi”. In: *The ISME journal* 9.10 (2015), pp. 2221–2231.

[60] H. Y. Chu, K. Sprouffske, and A. Wagner. “Assessing the benefits of horizontal gene transfer by laboratory evolution and genome sequencing”. In: *BMC evolutionary biology* 18.1 (2018), p. 54.

[61] S. Slomka, I. Françoise, G. Hornung, O. Asraf, T. Biniashvili, Y. Pilpel, and O. Dahan. “Experimental Evolution of Bacillus subtilis Reveals the Evolutionary Dynamics of Horizontal Gene Transfer and Suggests Adaptive and Neutral Effects”. In: *Genetics* 216.2 (2020), pp. 543–558.

[62] C. L. Hemme, S. J. Green, L. Rishishwar, O. Prakash, A. Pettenato, R. Chakraborty, A. M. Deutschbauer, J. D. van Nostrand, L. Wu, Z. He, I. K. Jordan, T. C. Hazen, A. P. Arkin, J. E. Kostka, and J. Zhou. “Lateral Gene Transfer in a Heavy Metal-Contaminated-Groundwater Microbial Community”. In: *mBio* 7.2 (2016), e02234–15.

[63] Á. T. Kovács. “Bacillus subtilis”. In: *Trends in microbiology* 27.8 (2019), pp. 724–725.

[64] D. Lopez, H. Vlamakis, and R. Kolter. “Generation of multiple cell types in Bacillus subtilis”. In: *FEMS microbiology reviews* 33.1 (2009), pp. 152–163.

[65] N. Mirouze and D. Dubnau. “Chance and Necessity in Bacillus subtilis Development”. In: *Microbiology spectrum* 1.1 (2013).

[66] R. M. Losick. “Bacillus subtilis: a bacterium for all seasons”. In: *Current biology : CB* 30.19 (2020), R1146–R1150.

[67] H. Maamar and D. Dubnau. “Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop”. In: *Molecular microbiology* 56.3 (2005), pp. 615–624.

[68] P. Stefanic, K. Belcijan, B. Kraigher, R. Kostanjšek, J. Nesme, J. S. Madsen, J. Kovac, S. J. Sørensen, M. Vos, and I. Mandic-Mulec. “Kin discrimination promotes horizontal gene transfer between unrelated strains in Bacillus subtilis”. In: *Nature communications* 12.1 (2021), p. 3457.

[69] R. Provvedi and D. Dubnau. “ComEA is a DNA receptor for transformation of competent Bacillus subtilis”. In: *Molecular microbiology* 31.1 (1999), pp. 271–280.

[70] R. N. Singh. “Number of deoxyribonucleic acid uptake sites in competent cells of Bacillus subtilis”. In: *Journal of bacteriology* 110.1 (1972), pp. 266–272.

[71] B. Maier, I. Chen, D. Dubnau, and M. P. Sheetz. “DNA transport into Bacillus subtilis requires proton motive force to generate large molecular forces”. In: *Nature structural & molecular biology* 11.7 (2004), pp. 643–649.

[72] I. Draskovic and D. Dubnau. “Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation of disulphide bonds”. In: *Molecular microbiology* 55.3 (2005), pp. 881–896.

[73] B. A. STOCKER. “Transformation of Bacillus subtilis to motility and prototrophy: micromanipulative isolation of bacteria of transformed phenotype”. In: *Journal of bacteriology* 86.4 (1963), pp. 797–804.

[74] D. B. Kearns and R. Losick. “Cell population heterogeneity during growth of Bacillus subtilis”. In: *Genes & development* 19.24 (2005), pp. 3083–3094.

[75] W. L. Nicholson. “Increased competitive fitness of Bacillus subtilis under nonsporulating conditions via inactivation of pleiotropic regulators AlsR, SigD, and SigW”. In: *Applied and environmental microbiology* 78.9 (2012), pp. 3500–3503.

[76] H. Vlamakis, C. Aguilar, R. Losick, and R. Kolter. “Control of cell fate by the formation of an architecturally complex bacterial community”. In: *Genes & development* 22.7 (2008), pp. 945–953.

[77] S. S. Branda, S. Vik, L. Friedman, and R. Kolter. “Biofilms: the matrix revisited”. In: *Trends in microbiology* 13.1 (2005), pp. 20–26.

[78] H. Vlamakis, Y. Chai, P. Beauregard, R. Losick, and R. Kolter. “Sticking together: building a biofilm the Bacillus subtilis way”. In: *Nature reviews. Microbiology* 11.3 (2013), pp. 157–168.

[79] J. C. Nickel, I. Ruseska, J. B. Wright, and J. W. Costerton. “Tobramycin resistance of Pseudomonas aeruginosa cells growing as a biofilm on urinary catheter material”. In: *Antimicrobial agents and chemotherapy* 27.4 (1985), pp. 619–624.

[80] M. E. Davey and G. A. O’toole. “Microbial biofilms: from ecology to molecular genetics”. In: *Microbiology and molecular biology reviews : MMBR* 64.4 (2000), pp. 847–867.

[81] P. J. Piggot and D. W. Hilbert. “Sporulation of Bacillus subtilis”. In: *Current opinion in microbiology* 7.6 (2004), pp. 579–586.

[82] I. Gordo, L. Perfeito, and A. Sousa. “Fitness effects of mutations in bacteria”. In: *Journal of molecular microbiology and biotechnology* 21.1-2 (2011), pp. 20–35.

[83] H. A. Orr. “Fitness and its role in evolutionary genetics”. In: *Nature reviews. Genetics* 10.8 (2009), pp. 531–539.

[84] M. Sane, G. D. Diwan, B. A. Bhat, L. M. Wahl, and D. Agashe. “Shifts in mutation spectra enhance access to beneficial mutations”. In: *Proceedings of the National Academy of Sciences of the United States of America* 120.22 (2023), e2207355120.

[85] M. Lukačišinová and T. Bollenbach. “Toward a quantitative understanding of antibiotic resistance evolution”. In: *Current opinion in biotechnology* 46 (2017), pp. 90–97.

[86] M. J. Wiser and R. E. Lenski. “A Comparison of Methods to Measure Fitness in Escherichia coli”. In: *PloS one* 10.5 (2015), e0126210.

[87] S. F. Elena and R. E. Lenski. “Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation”. In: *Nature reviews. Genetics* 4.6 (2003), pp. 457–469.

[88] R. Gallet, T. F. Cooper, S. F. Elena, and T. Lenormand. “Measuring selection coefficients below 10(-3): method, questions, and prospects”. In: *Genetics* 190.1 (2012), pp. 175–186.

[89] H. Maughan and W. L. Nicholson. “Increased fitness and alteration of metabolic pathways during Bacillus subtilis evolution in the laboratory”. In: *Applied and environmental microbiology* 77.12 (2011), pp. 4105–4118.

[90] A. M. Smith, L. E. Heisler, J. Mellor, F. Kaper, M. J. Thompson, M. Chee, F. P. Roth, G. Giaever, and C. Nislow. “Quantitative phenotyping via deep barcode sequencing”. In: *Genome research* 19.10 (2009), pp. 1836–1842.

[91] T. van Opijnen, D. W. Lazinski, and A. Camilli. “Genome-Wide Fitness and Genetic Interactions Determined by Tn-seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms”. In: *Current protocols in microbiology* 36 (2015), 1E.3.1–1E.3.24.

[92] K. M. Wetmore, M. N. Price, R. J. Waters, J. S. Lamson, J. He, C. A. Hoover, M. J. Blow, J. Bristow, G. Butland, A. P. Arkin, and A. Deutschbauer. “Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons”. In: *mBio* 6.3 (2015), e00306–15.

[93] L. Robert, J. Ollion, J. Robert, X. Song, I. Matic, and M. Elez. “Mutation dynamics and fitness effects followed in single cells”. In: *Science (New York, N.Y.)* 359.6381 (2018), pp. 1283–1286.

[94] K. S. Korolev, M. J. I. Müller, N. Karahan, A. W. Murray, O. Hallatschek, and D. R. Nelson. “Selective sweeps in growing microbial colonies”. In: *Physical biology* 9.2 (2012), p. 026008.

[95] R. Zöllner, E. R. Oldewurtel, N. Kouzel, and B. Maier. “Phase and antigenic variation govern competition dynamics through positioning in bacterial colonies”. In: *Scientific reports* 7.1 (2017), p. 12151.

[96] K. B. Böndel, S. A. Kraemer, T. Samuels, D. McClean, J. Lachapelle, R. W. Ness, N. Colegrave, and P. D. Keightley. “Inferring the distribution of fitness effects of spontaneous mutations in Chlamydomonas reinhardtii”. In: *PLoS biology* 17.6 (2019), e3000192.

[97] S. Wright. “The Roles of Mutation Inbreeding, Crossbreeding and Selection in Evolution”. In: *Proc. 6th Int. Cong. Genetics* 1 (1932), pp. 356–366.

[98] P. F. Stadler and C. R. Stephens. “Landscapes and Effective Fitness”. In: *Comments on Theoretical Biology* 8.4-5 (2003), pp. 389–431.

[99] S. Gavrilets. *Fitness landscapes and the origin of species*. Princeton and Oxford: Princeton University Press, 2004.

[100] B. van den Bergh, T. Swings, M. Fauvart, and J. Michiels. “Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution”. In: *Microbiology and molecular biology reviews : MMBR* 82.3 (2018).

[101] J. A. G. M. de Visser and J. Krug. “Empirical fitness landscapes and the predictability of evolution”. In: *Nature reviews. Genetics* 15.7 (2014), pp. 480–490.

[102] D. J. Kvitek and G. Sherlock. “Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape”. In: *PLoS genetics* 7.4 (2011), e1002056.

[103] B. A. Malcolm, K. P. Wilson, B. W. Matthews, J. F. Kirsch, and A. C. Wilson. “Ancestral lysozymes reconstructed, neutrality tested, and thermostability linked to hydrocarbon packing”. In: *Nature* 345.6270 (1990), pp. 86–89.

[104] J. A. G. M. de Visser, R. F. Hoekstra, and H. van den Ende. “TEST OF INTERACTION BETWEEN GENETIC MARKERS THAT AFFECT FITNESS IN ASPERGILLUS NIGER”. In: *Evolution; international journal of organic evolution* 51.5 (1997), pp. 1499–1505.

[105] D. W. Hall, M. Agan, and S. C. Pope. “Fitness epistasis among 6 biosynthetic loci in the budding yeast Saccharomyces cerevisiae”. In: *The Journal of heredity* 101 Suppl 1 (2010), S75–84.

[106] A. Eyre-Walker and P. D. Keightley. “The distribution of fitness effects of new mutations”. In: *Nature reviews. Genetics* 8.8 (2007), pp. 610–618.

[107] M. M. Dillon and V. S. Cooper. “The Fitness Effects of Spontaneous Mutations Nearly Unseen by Selection in a Bacterium with Multiple Chromosomes”. In: *Genetics* 204.3 (2016), pp. 1225–1238.

[108] S. Trindade, A. Sousa, and I. Gordo. “Antibiotic resistance and stress in the light of Fisher’s model”. In: *Evolution; international journal of organic evolution* 66.12 (2012), pp. 3815–3824.

[109] R. Kishony and S. Leibler. “Environmental stresses can alleviate the average deleterious effect of mutations”. In: *Journal of biology* 2.2 (2003), p. 14.

[110] R. C. MacLean and A. Buckling. “The distribution of fitness effects of beneficial mutations in Pseudomonas aeruginosa”. In: *PLoS genetics* 5.3 (2009), e1000406.

[111] T. Bataillon, T. Zhang, and R. Kassen. “Cost of adaptation and fitness effects of beneficial mutations in Pseudomonas fluorescens”. In: *Genetics* 189.3 (2011), pp. 939–949.

[112] D. E. Rozen, J. A. G. M. de Visser, and P. J. Gerrish. “Fitness effects of fixed beneficial mutations in microbial populations”. In: *Current biology : CB* 12.12 (2002), pp. 1040–1045.

[113] G. Chevereau, M. Dravecká, T. Batur, A. Guvenek, D. H. Ayhan, E. Toprak, and T. Bollenbach. “Quantifying the Determinants of Evolutionary Dynamics Leading to Drug Resistance”. In: *PLoS biology* 13.11 (2015), e1002299.

[114] A. Knöppel, P. A. Lind, U. Lustig, J. Näsvall, and D. I. Andersson. “Minor fitness costs in an experimental model of horizontal gene transfer in bacteria”. In: *Molecular biology and evolution* 31.5 (2014), pp. 1220–1227.

[115] D. Dykhuizen. “Thoughts Toward a Theory of Natural Selection: The Importance of Microbial Experimental Evolution”. In: *Cold Spring Harbor perspectives in biology* 8.3 (2016), a018044.

[116] J. E. Barrick and R. E. Lenski. “Genome dynamics during experimental evolution”. In: *Nature reviews. Genetics* 14.12 (2013), pp. 827–839.

[117] R. E. Lenski. “Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations”. In: *The ISME journal* 11.10 (2017), pp. 2181–2194.

[118] R. E. Lenski. “Revisiting the Design of the Long-Term Evolution Experiment with Escherichia coli”. In: *Journal of molecular evolution* 91.3 (2023), pp. 241–253.

[119] I. McDougall, F. H. Brown, and J. G. Fleagle. “Stratigraphic placement and age of modern humans from Kibish, Ethiopia”. In: *Nature* 433.7027 (2005), pp. 733–736.

[120] M. T. Stanek, T. F. Cooper, and R. E. Lenski. “Identification and dynamics of a beneficial mutation in a long-term evolution experiment with Escherichia coli”. In: *BMC evolutionary biology* 9 (2009), p. 302.

[121] F. Vasi, M. Travisano, and R. E. Lenski. “Long-Term Experimental Evolution in Escherichia coli. II. Changes in Life-History Traits During Adaptation to a Seasonal Environment”. In: *The American Naturalist* 144.3 (1994), pp. 432–456.

[122] R. Maddamsetti, R. E. Lenski, and J. E. Barrick. “Adaptation, Clonal Interference, and Frequency-Dependent Interactions in a Long-Term Evolution Experiment with Escherichia coli”. In: *Genetics* 200.2 (2015), pp. 619–631.

[123] Z. D. Blount, C. Z. Borland, and R. E. Lenski. “Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli”. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.23 (2008), pp. 7899–7906.

[124] A. NOVICK and L. SZILARD. “Description of the chemostat”. In: *Science (New York, N.Y.)* 112.2920 (1950), pp. 715–716.

[125] D. Gresham and M. J. Dunham. “The enduring utility of continuous culturing in experimental evolution”. In: *Genomics* 104.6 Pt A (2014), pp. 399–405.

[126] A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper. “Negative epistasis between beneficial mutations in an evolving bacterial population”. In: *Science (New York, N.Y.)* 332.6034 (2011), pp. 1193–1196.

[127] H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx. “Diminishing returns epistasis among beneficial mutations decelerates adaptation”. In: *Science (New York, N.Y.)* 332.6034 (2011), pp. 1190–1192.

[128] C. A. Fogle, J. L. Nagle, and M. M. Desai. “Clonal interference, multiple mutations and adaptation in large asexual populations”. In: *Genetics* 180.4 (2008), pp. 2163–2173.

[129] D. Dubnau and R. Davidoff-Abelson. “Fate of transforming DNA following uptake by competent Bacillus subtilis. I. Formation and properties of the donor-recipient complex”. In: *Journal of molecular biology* 56.2 (1971), pp. 209–221.

[130] Easybel and ACCakut. *Easybel/DetectionGV: Detection of genetic variants*. Zenodo, 2023.

[131] H. Li. “Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM”. Mar. 16, 2013.

[132] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. “Twelve years of SAMtools and BCFtools”. In: *GigaScience* 10.2 (2021).

[133] H. Li. “A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data”. In: *Bioinformatics (Oxford, England)* 27.21 (2011), pp. 2987–2993.

[134] A. R. Quinlan and I. M. Hall. “BEDTools: a flexible suite of utilities for comparing genomic features”. In: *Bioinformatics (Oxford, England)* 26.6 (2010), pp. 841–842.

[135] Y. Kim, C. Gu, H. U. Kim, and S. Y. Lee. “Current status of pan-genome analysis for pathogenic bacteria”. In: *Current opinion in biotechnology* 63 (2020), pp. 54–62.

[136] A. Perrin and E. P. C. Rocha. “PanACoTA: a modular tool for massive microbial comparative genomics”. In: *NAR genomics and bioinformatics* 3.1 (2021), lqaa106.

[137] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. “Basic local alignment search tool”. In: *Journal of molecular biology* 215.3 (1990), pp. 403–410.

[138] I. Rathmann, M. Förster, M. Yüksel, L. Horst, G. Petrungaro, T. Bollenbach, and B. Maier. “Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution”. In: *The ISME journal* 17.1 (2023), pp. 130–139.

[139] The MathWorks Inc. *MATLAB*. Version version: 9.13.0 (R2022b). Natick, Massachusetss, United States: The MathWorks Inc., 2022.

[140] M. Förster, I. Rathmann, M. Yüksel, J. J. Power, and B. Maier. “Genome-wide transformation reveals extensive exchange across closely related Bacillus species”. In: *Nucleic acids research* 51.22 (2023), pp. 12352–12366.

[141] D. R. Zeigler. “The genome sequence of Bacillus subtilis subsp. spizizenii W23: insights into speciation within the B. subtilis complex and into the history of B. subtilis genetics”. In: *Microbiology (Reading, England)* 157.Pt 7 (2011), pp. 2033–2041.

[142] N. J. Croucher, R. Mostowy, C. Wymant, P. Turner, S. D. Bentley, and C. Fraser. “Horizontal DNA Transfer Mechanisms of Bacteria as Weapons of Intragenomic Conflict”. In: *PLoS biology* 14.3 (2016), e1002394.

[143] N. J. Croucher, P. G. Coupland, A. E. Stevenson, A. Callendrello, S. D. Bentley, and W. P. Hanage. “Diversification of bacterial genome content through distinct mechanisms over different timescales”. In: *Nature communications* 5 (2014), p. 5471.

[144] K. Kohm and R. Hertel. “The life cycle of SPβ and related phages”. In: *Archives of virology* 166.8 (2021), pp. 2119–2130.

[145] H. Westers, R. Dorenbos, J. M. van Dijl, J. Kabel, T. Flanagan, K. M. Devine, F. Jude, S. J. Seror, A. C. Beekman, E. Darmon, C. Eschevins, A. de Jong, S. Bron, O. P. Kuipers, A. M. Albertini, H. Antelmann, M. Hecker, N. Zamboni, U. Sauer, C. Bruand, D. S. Ehrlich, J. C. Alonso, M. Salas, and W. J. Quax. “Genome engineering reveals large dispensable regions in Bacillus subtilis”. In: *Molecular biology and evolution* 20.12 (2003), pp. 2076–2090.

Acknowledgement

First of all, I would like to thank **Berenike** for all the years of support and great advice, for giving me the chance to work in your lab, and for all the help during the writing process.

Thanks to **Isabel Gordo** for reading through this thesis and for taking the time to be my second examiner.

Thanks to the whole **Maier Lab**. You are one hell of a group. I enjoyed every moment with you guys:
Melih, I cannot count how many times you helped me with whatever I came to you. Thanks for every Mauerbier you animated the group for and for your unique laughter.
Thorsten, you are the one I ask when I cannot find a thing. Thanks for the great smell that always gives away where you are.
Ariana, you are a great robot partner. Thanks for spending so many hours with me in this beautiful, windowless 30 °C room. And for all the laughter during Mensa time.
Philipp, thanks for building the pinning tool with me, helping me with my trunk, ... I don't think I have ever met anyone as attentive as you are. Stay the way you are, but take care of yourself too.
Paul, I cannot image a better office partner, guard of the candy drawer. You are the perfect mix of quiet, intense work time and interesting conversations.
Isa, my new office mate. A pity that we did not do that earlier. I enjoy to talk with you a lot.
Sebi, thanks for being you. For me, you are the hardest working person in the lab, but still you always find the time to help everybody else. Special thanks for the parties you and Micha gave. I always had so much fun.
Aurelio, thanks for your supergreat humor. You always make me laugh.
Stephan, **Ilke** and **Mahak**, thanks for always being kind and having a smile on your lips.
Simone, thank you for reorganizing the lab from scratch.
Lucas, thanks for the endless hours you spent in the robot room helping us solve problems without ever complaining.
And last but not least, **Isabel**, I think I do not have to tell you that you made my time in the Maier lab. Thank you for all our discussions about work, but especially for everything that had nothing to do with work. I have found a true friend in you.

Most importantly, I would like to thank my **Mama** and **Papa**, who always believe in me, and my siblings, **Carina** and **Micha**, who have been there for me all my life. As well as **Chris**, thank you for reading through my whole thesis without understanding a word, for the endless support you give me every day, and for the love.

Used resources

The sequencing data were processed on the High Performance Computing (HPC) system CHEOPS of the Regional Computing Center of the University of Cologne (RRZK). To improve the wording within this thesis, the DeepL translator was used.

A. General supplementary information

A.1. List of used instruments

Instrument	Name	Producer
Microplate reader	Infinite M200 Pro	Tecan
Flow cytometer 1	Cytoflex	Beckman Coulter
Flow cytometer 2	Canto II	BD Biosciences
Agar plate handling system	BM3-BC	S&P Robotics Inc.
Used within the robotic system:		
Robotic arm	Acell	HighRes Biosolutions
Liquid-handling device	Lynx LM900	Dynamic Devices
Microplate reader	Synergy	BioTek Instruments
Shaker	BioShake 3000 elm	QInstruments
Incubator	StoreX STX44	Liconic
Plate storage	Nanoserve	HighRes Biosolution
Delidding device	LidValet	HighRes Biosolutions

Table A.1.: Instruments used in this study.

A.2. Multimapper genes

Start pos.	End pos.	Length [bp]	Hit genes
9811	14845	5035	*rrnO-16S, trnO-Ile, trnO-Ala, rrnO-23S, rrnO-5S*
30279	35153	4875	*rrnA-16S, trnA-Ile, trnA-Ala, rrnA-23S*
90446	95452	5007	*rrnJ-16S, rrnJ-23S, rrnJ-5S, trnJ-Val*
95954	101248	5295	*trnJ-Arg, trnJ-Pro, trnJ-Ala, rrnW-16S, rrnW-23S, rrnW-5S*
160804	165712	4909	*rrnI-16S, rrnI-23S, rrnI-5S*
166063	171314	5252	*trnI-Arg, trnI-Pro, trnI-Ala, rrnH-16S, rrnH-23S, rrnH-5S*
171329	176351	5023	*rrnG-16S, rrnG-23S, rrnG-5S*
635281	640255	4975	*rrnE-16S, rrnE-23S, rrnE-5S*
946543	951388	4846	*rrnD-16S, rrnD-23S*
3173719	3178792	5074	*trnB-Val, rrnB-5S, rrnB-23S, rrnB-16S*

Table A.2.: List of multi mapper regions in Bs166.

A.3. NCBI dictionaries

Strain	Data base name	NCBI dictionary
Bacillus subtilis 168	Bs166	NC_000964.3
	Bs175	NC_000964.3
	Bs210	NC_000964.3
	Bs213	NC_000964.3
	Bs224	NC_000964.3
	Bs226	NC_000964.3
Bacillus spizizenii	Ob005	NC_014479.1
Bacillus vallismortis	Ob018	NZ_CP026362.1
Bacillus atrophaeus	Ob006	NC_014639.1
Geobacillus thermoglucosidasius	Ob011	NZ_CP012712.1

Table A.3.: List of NCBI dictionaries used.

A.4. Mock genomes

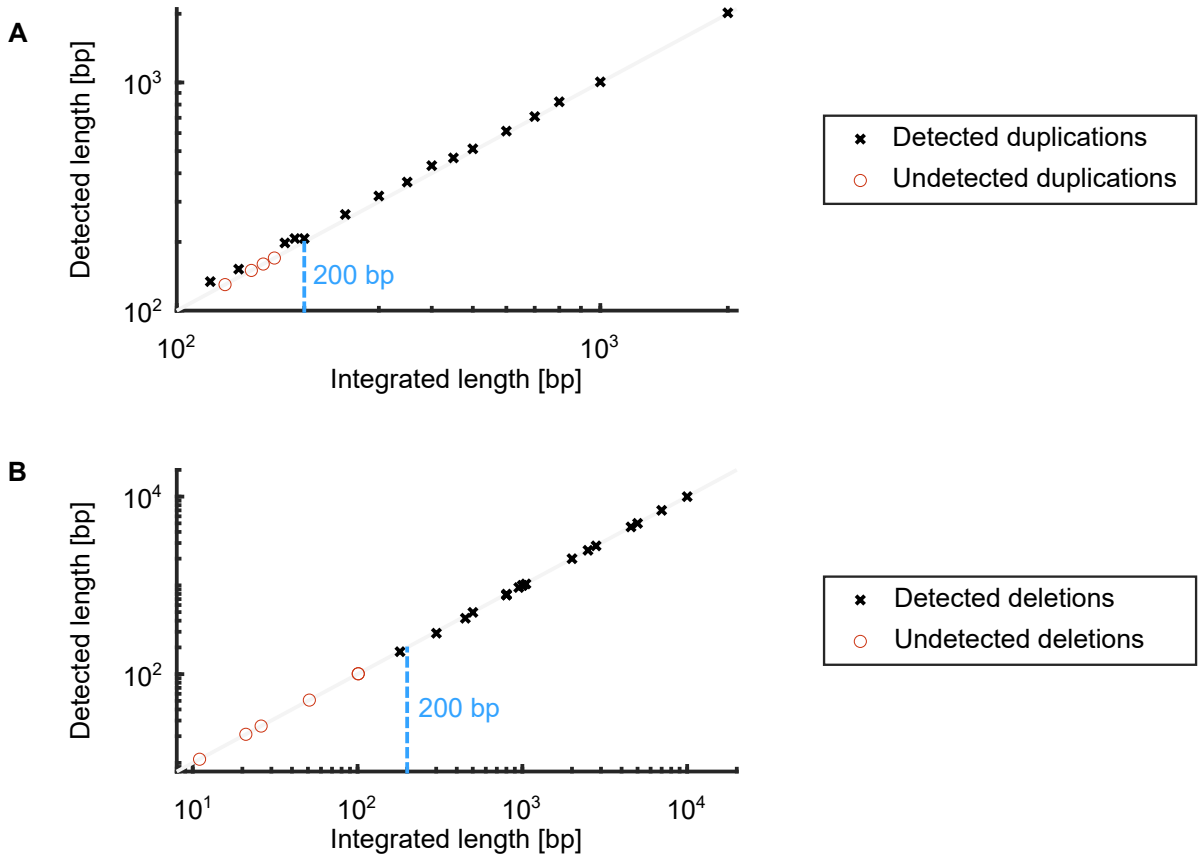


Figure A.1.: Analysis of mock genomes. We artifically created mock genomes for duplications (**A**) and deletions (**B**) to test how reliable the results of our algorithm are. Deletions as well as duplications longer than 200 bp are detected without problems.

B. Supplementary information on the publication: Genome-wide transformation reveals extensive exchange across closely related *Bacillus* species

Supplementary Figures and Tables for

Genome-wide transformation reveals extensive exchange across closely related *Bacillus* species

Mona Förster *[1], Isabel Rathmann *[1], Melih Yüksel[1], Jeffrey Power[1], Berenike Maier[1,3,4]

*contributed equally

[1] Institute for Biological Physics, Zülpicherstr. 47a, 50674 Köln, University of Cologne

[2] Center for Molecular Medicine Cologne, University of Cologne

[3] email: berenike.maier@uni-koeln.de

Supplementary Figures

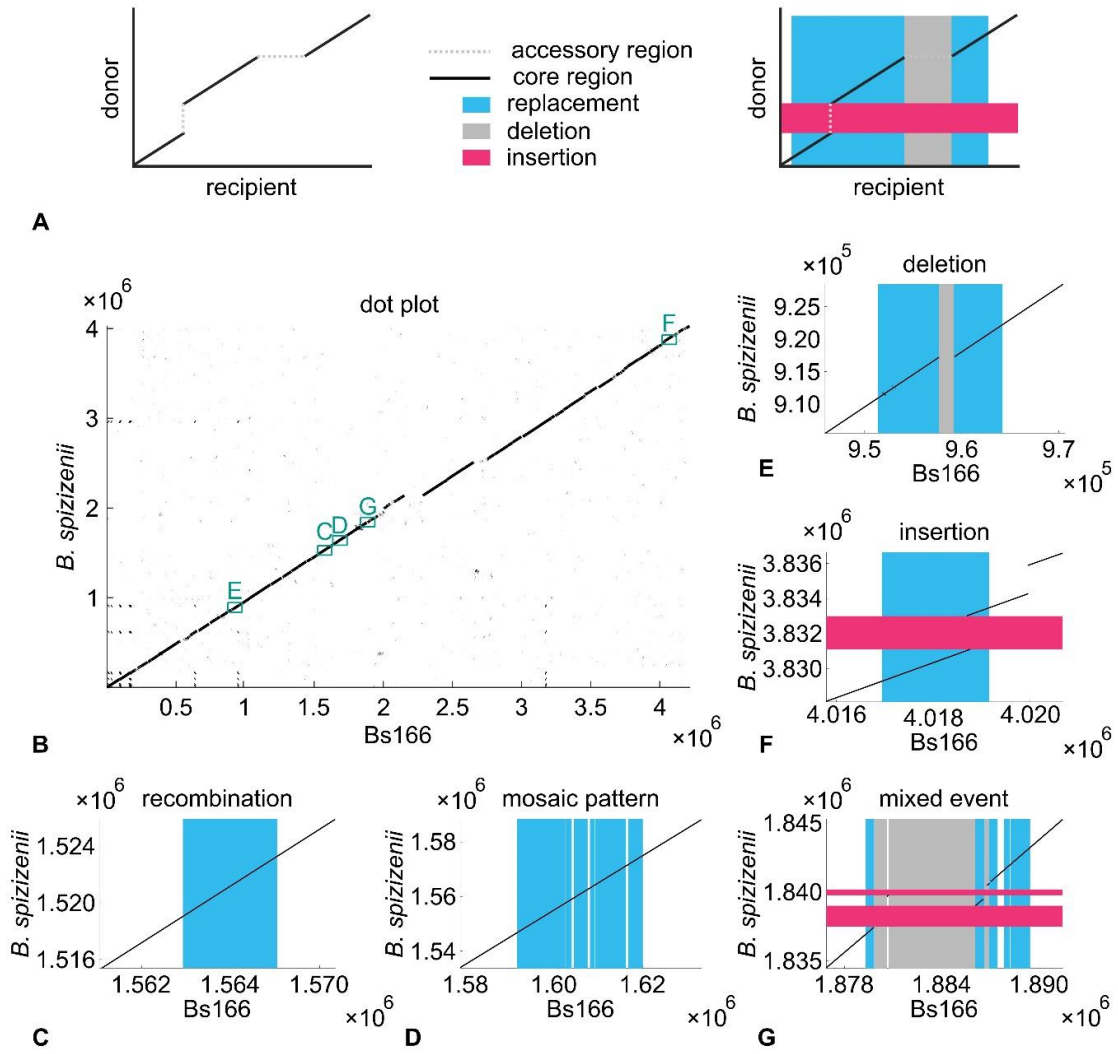


Fig. S1 Mixed integration events are detected by using dot plots to investigate the differences in genome architecture between strains. Dot plots are created by aligning the donor to the recipient with the blastn algorithm (Methods). A) Scheme of a dot plot in which homologous regions between donor and recipient genome (solid black lines) and accessory regions (gaps, dotted lines) are depicted. Adding detected replacements, deletions, and insertions visualizes the local dependence between events. For *B. spizizenii* and the ancestor Bs166, the whole genome dot plot is shown in B). Most of the homologous regions are present in the same order within both genomes (vertical black line). The green boxes indicate the approximate positions of examples C) – G) that depict genomic changes within the BspizHyb strains. C) contains a simple recombination, D) a recombination with mosaic pattern and, E) and F) a deletion and an

insertion both caused by a recombination of homologous flanking regions. In G), we find a mixed event that combines recombination, mosaic, insertion, and deletion.

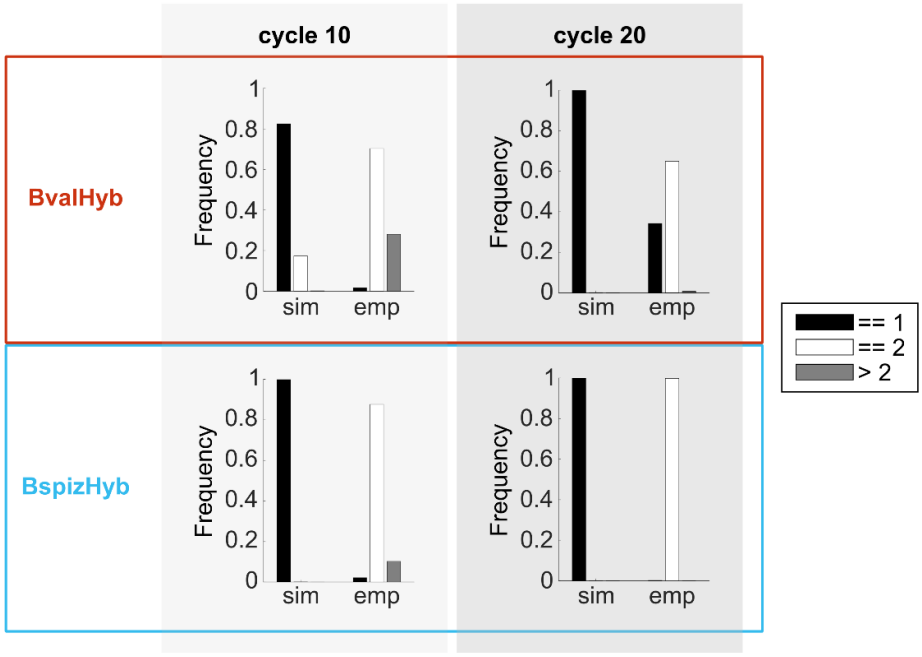


Fig. S2 Test for bimodality of the probability distribution of transfer distances of independent replacement events. We assess whether the long-distance regime arises from independently integrated segments whereas the short distances originate from single recombination events that generate a mosaic pattern between donor and recipient alleles. Fractions of bootstrap runs that show a single peak (black bar), bimodal behavior (white bar), more than two peaks (gray bar). Sim: independent transfers generated by the Monte Carlo null model (described in the Methods). Emp: transfers from empirical data.

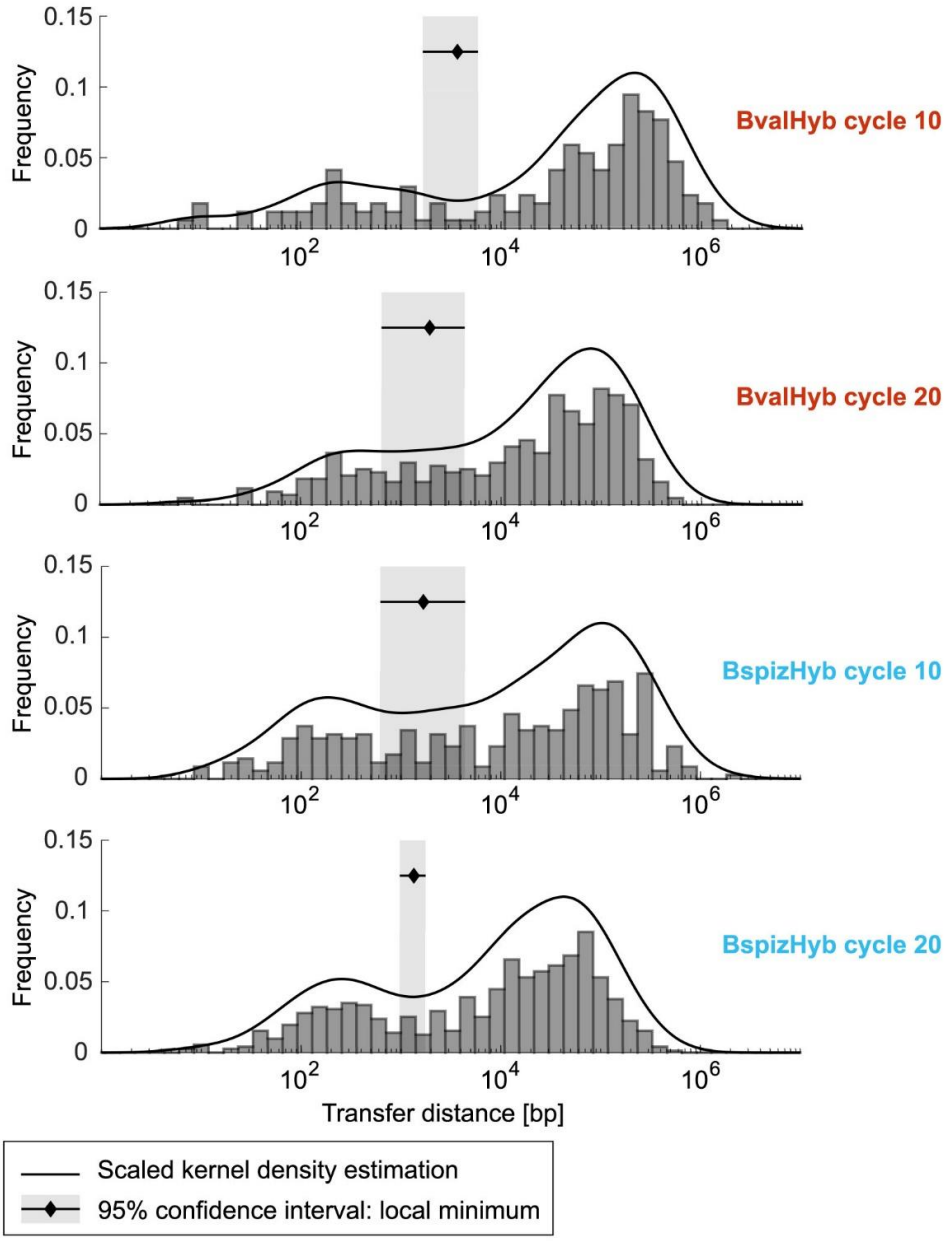


Fig. S3 Determination of threshold length between mosaic and independent replacement events. The empirical data set is represented as histograms (grey bars) and as probability density estimated with a log-normal kernel (black line) (shown in arbitrary units). The light gray area around the diamonds depicts the 95 % confidence interval of the local minima determined by bootstrapping (see Methods). Those minima are used as a cut-off between long- and short-distance regimes.

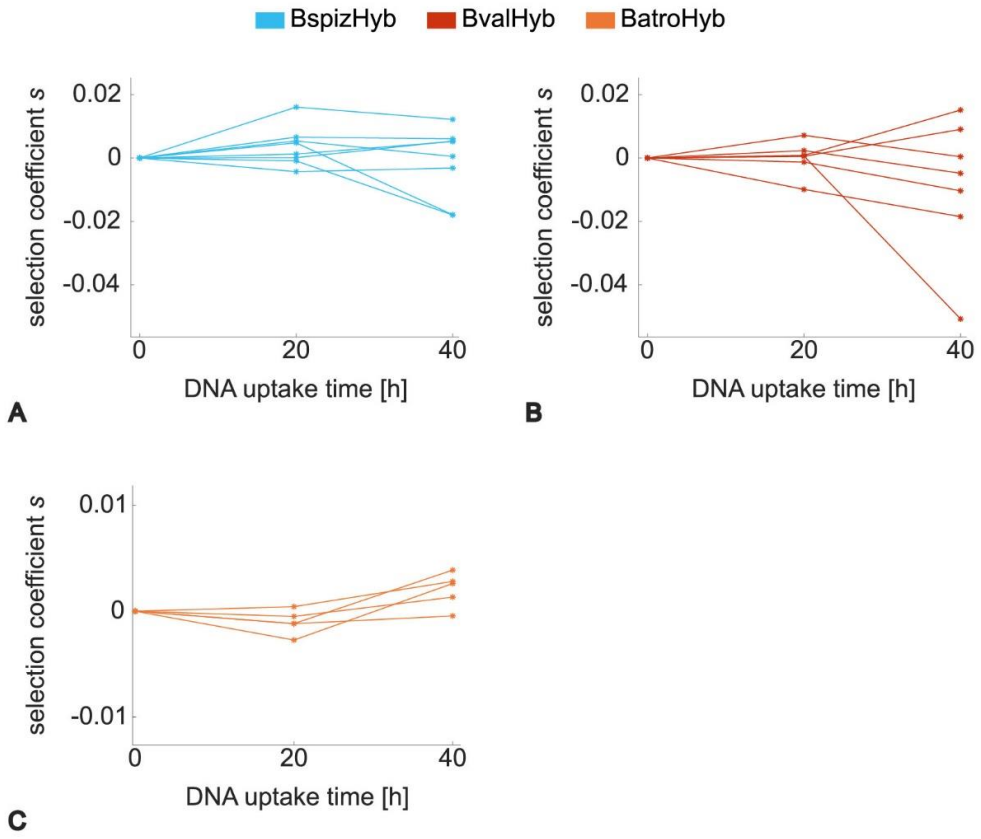


Fig. S4 Fitness trajectories of transformation hybrids show no net increase. For each hybrid, the fitness is depicted after cycle 10 and 20 as a trajectory. Neither for A) BspizHyb (N = 8), B) BvalHyb (N = 7), nor for C) BatroHyb (N = 5) do we find curves that systematically increase their fitness over time.

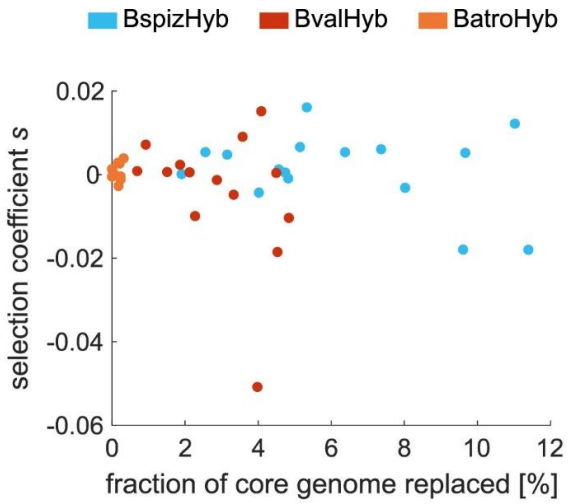


Fig. S5 Selection coefficient does not correlate with the fraction of replaced genome. The selection coefficients of the hybrid strains at cycles 10 and 20 (Fig. 2) are plotted against the fractions of the replaced core genomes.

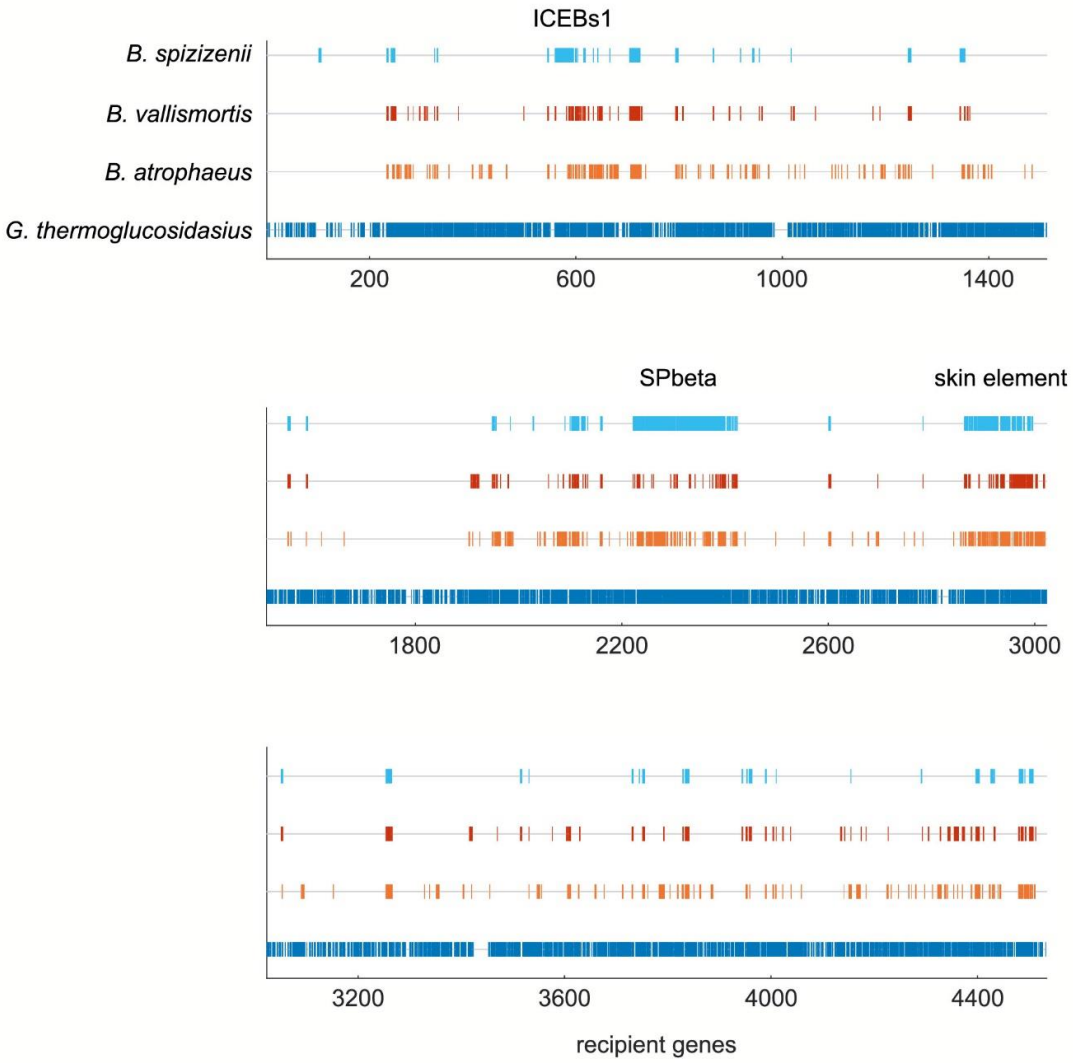


Fig. S6 Architecture of accessory genomes. The accessory genomes are shown for the Bs166 recipient with respect to the four donor species. Accessory genes are depicted as boxes and the names of important prophage elements and mobile elements are added, according to the annotation (Methods).

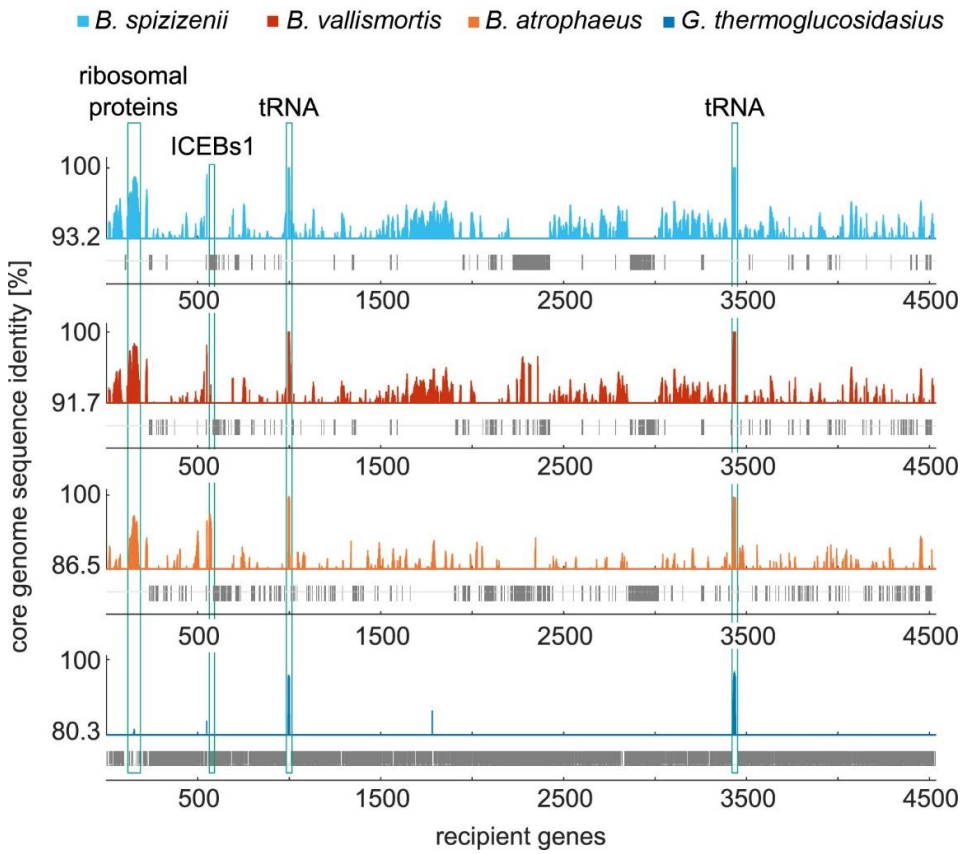


Fig. S7 The core genome sequence identity is distributed irregularly along the recipient genome. For four donor species with respect to the recipient, the gene-wise sequence identity of the core genome is shown, evaluated as an average in a sliding window of 10 genes. The accessory genes are depicted as gray boxes. The lower limits of the y-axis correspond to the average core genome identities of each species pair. Most prominent clusters of highly identical genes are highlighted.

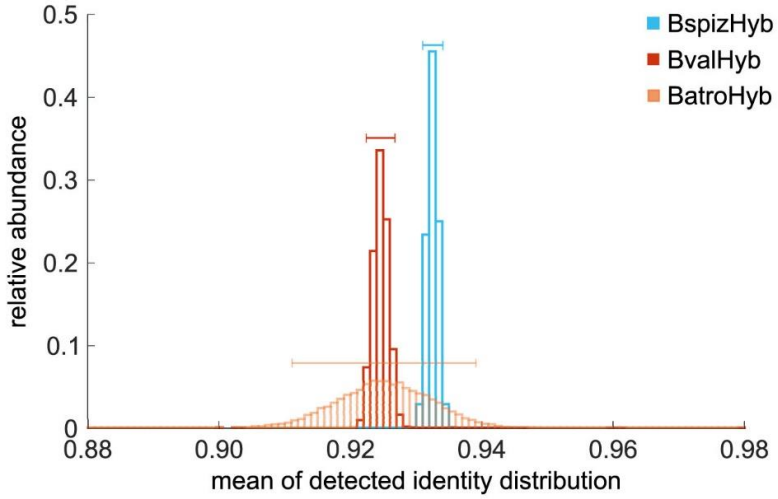


Fig. S8 A bootstrap analysis for identity distributions of orthologous replacements is performed on the mean of the identity distributions (Fig. 4) for BspizHyb (blue), BvalHyb (red) and BatroHyb (turquoise). Bootstrap samples were drawn 10^4 times and 95 % confidence intervals are depicted as horizontal lines.

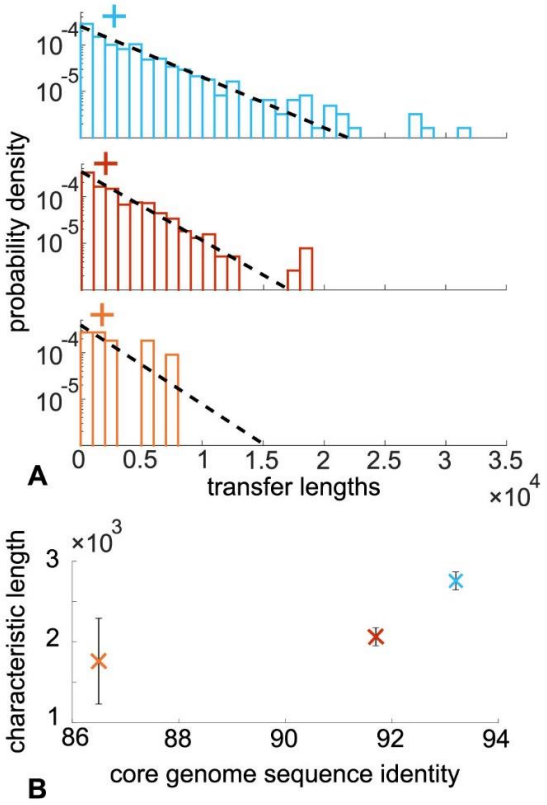


Fig. S9 Characteristic transfer lengths of replaced segments increase with increasing core genome sequence identity. Here, mosaic segments are not merged. A) The length distribution of segments is fitted with an exponential function (dashed line) for segments longer than 100 bp and the characteristic length is determined (cross). B) This quantity increases with increasing core genome sequence identity (error bars depict the standard deviation). Cyan: BspizHyb, red: BvalHyb, orange: BatroHyb.

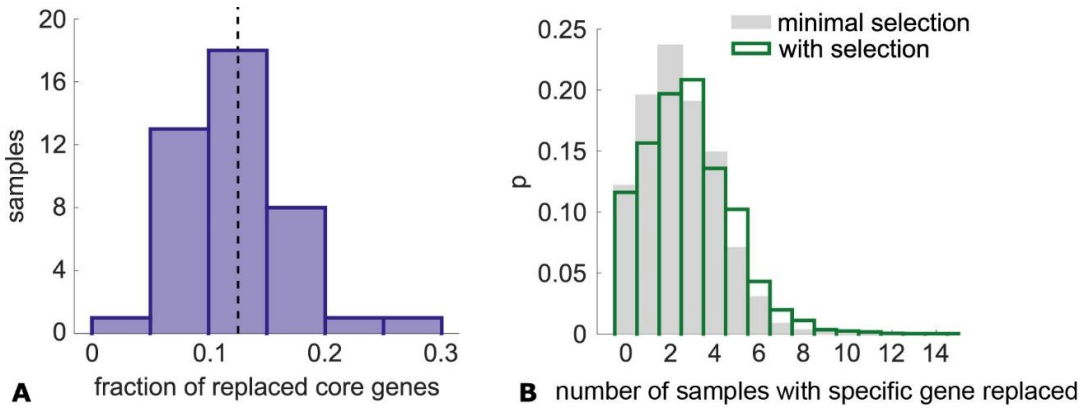


Fig. S10 Analysis of pooled data from 42 hybrid strains with *B. spizizenii* as donor. A) Distribution of fractions of fully or partially replaced core genes for all hybrids. B) Empirical probability distribution of finding genes replaced in a certain number of samples. Hybrids were obtained by single cell bottlenecking under minimal selection (grey) or including competition among different hybrid prior to bottlenecking (green) as described in the Methods.

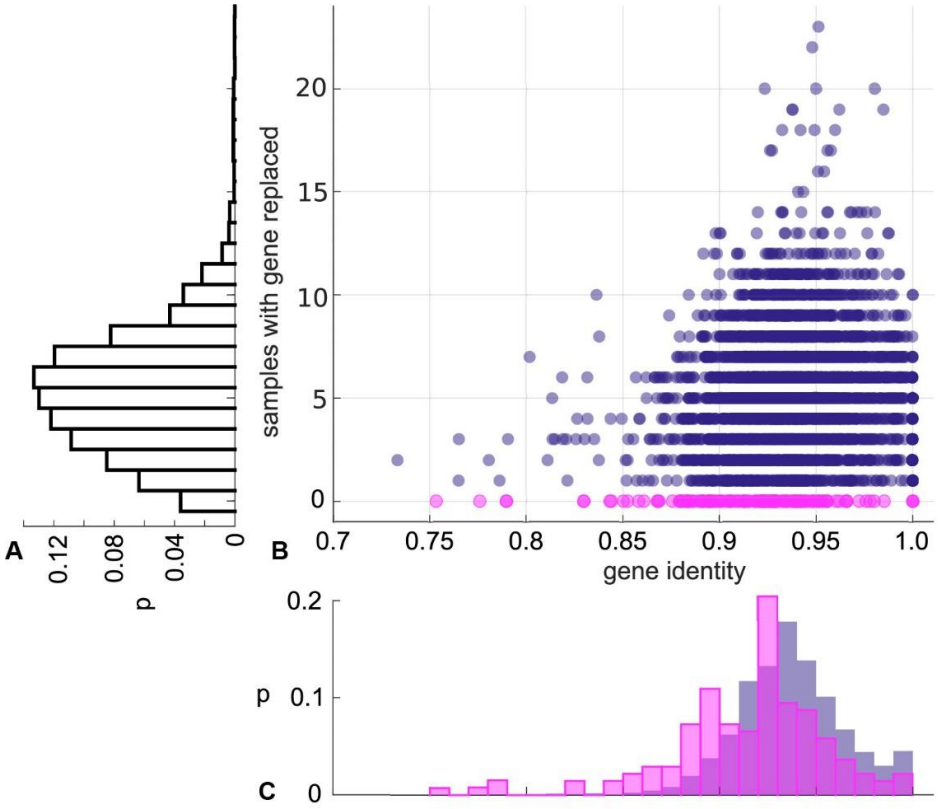


Fig. S11 Putative cold spots of orthologous replacement have a lower identity than replaced genes. Pooled data from 42 hybrid strains with *B. spizizenii* as donor was analyzed. A) Probability to find genes replaced in a certain number of samples. B) For each core gene, its average sequence identity is plotted against the number of samples it was replaced fully or partially. C) Distribution of gene identity of genes that were replaced in at least one strain (blue) and genes that were not replaced in any of the hybrids (pink). Based on a Welch's t-test, the distributions are different at a significance level of 0.05.

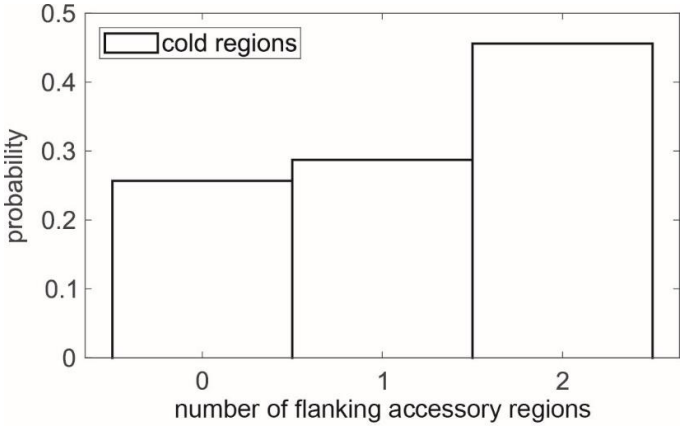


Fig. S12 Probability that a putative cold spot regions of 42 pooled *B. spizizenii* hybrids is flanked by 0, 1, or 2 accessory regions.

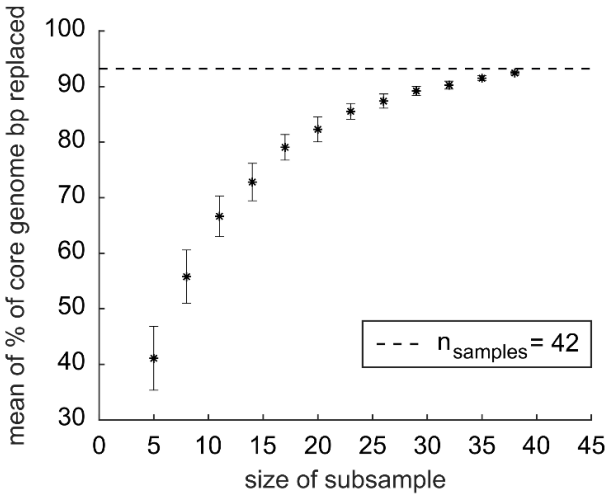
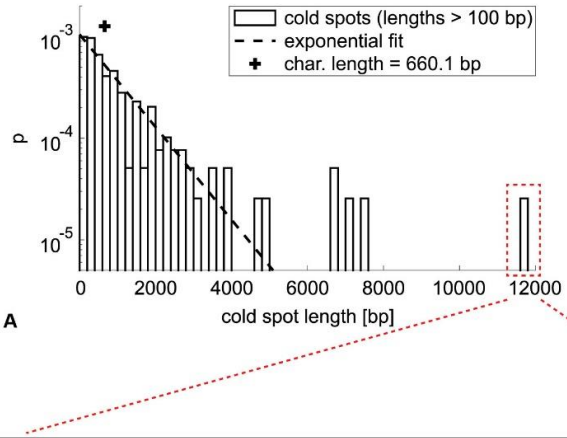


Fig. S13 Subsampling analysis of the fraction of replaced core genome. Random subsampling with the size of the sample indicated on the x-axis. For each size, 42 subsamples are drawn randomly without replacements and the percentage of the replaced genome (in bp) is determined. We find that the percentage does not yet saturate, indicating that a larger number of samples might show replacements within the putative cold spots.



gene name	function	start pos	end pos	ident	dir	hitFrac
yyaD	putative transporter	4203425	4204441	0.937	-	0.62
sprC	spore-specific protease	4204901	4205518	0.921	+	1
parB	site-specific DNA-binding protein	4205557	4206405	0.945	-	1
parA	chromosome partitioning protein%3B transcriptional regulator	4206398	4207159	0.949	-	1
yyaB	putative antibiotic immunity protein	4207407	4207847	0.896	+	1
nocA	DNA-binding protein Spo0J-like	4207898	4208749	0.967	-	1
rsmG	7-methylguanosine methyltransferase (16S rRNA%2C nucleotide G527)	4208871	4209590	0.956	-	1
trmF	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	4209604	4211490	0.957	-	1
mnmE	tRNA modification GTPase and tRNA-U34 5-formylation enzyme	4211511	4212890	0.959	-	1
jag	SpoIIIJ-associated RNA/ssDNA-binding protein	4213201	4213827	0.936	-	1
oxaAA	Sec-independent factor for membrane protein insertion (YidC/SpoIIIJ family)	4213824	4214609	0.971	-	1
rnpA	protein component of ribonuclease P (RNase P) (substrate specificity)	4214754	4215104	0.966	-	1
rpmH	ribosomal protein L34	4215256	4215390	0.985	-	1

B

Fig. S14 Length distribution of putative cold spot regions of pooled *B. spizizenii* hybrids. A) Distribution of the lengths of contiguous regions that have not been hit by replacements in any of the 42 pooled strains. Black dotted line: exponential fit with a characteristic length of 660.1 bp. Red dotted box: outlier from the exponential distribution ($p = 0.0008$) determined as described in the Methods. B) Genes residing within the cold spot of length 11796 bp.

Donor	Hybrid strain	$\langle s \rangle$	sd of s
B. spizizenii	Wns1110	-4.32E-03	1.20E-02
	Wns1210	1.61E-02	1.07E-02
	Wns1410	5.38E-03	4.47E-03
	Wns1510	1.06E-04	7.28E-03
	Wns1610	6.59E-03	1.68E-02
	Wns1710	4.78E-03	6.56E-03
	Wns1810	-9.20E-04	4.40E-03
	Wns1910	1.25E-03	5.57E-03
	Wns1120	-3.16E-03	4.94E-03
	Wns1220	1.22E-02	2.58E-02
	Wns1420	4.95E-04	6.80E-03
	Wns1520	5.35E-03	9.84E-03
	Wns1620	6.06E-03	2.16E-02
	Wns1720	-1.79E-02	1.57E-02
	Wns1820	-1.80E-02	1.49E-02
	Wns1920	5.21E-03	4.94E-03
B. vallismortis	Vns0110	6.25E-04	2.19E-03
	Vns0210	8.41E-04	7.12E-03
	Vns0310	2.34E-03	1.00E-02
	Vns0410	7.16E-03	2.11E-02
	Vns0510	-9.92E-03	3.18E-02
	Vns0610	5.44E-04	4.86E-03
	Vns0710	-1.30E-03	9.21E-03
	Vns0120	-5.08E-02	5.26E-02
	Vns0220	1.52E-02	9.85E-03
	Vns0320	-4.82E-03	1.06E-02
	Vns0420	3.93E-04	1.88E-02
	Vns0520	-1.85E-02	2.22E-02
	Vns0620	9.05E-03	9.08E-03
	Vns0720	-1.04E-02	7.61E-03
B. atrophaeus	Bans0110	-1.17E-03	5.01E-03
	Bans0210	-4.94E-04	1.44E-02
	Bans0410	4.28E-04	7.75E-03
	Bans0610	-1.18E-03	1.06E-02
	Bans0810	-2.73E-03	9.17E-03
	Bans0120	3.88E-03	7.31E-03
	Bans0220	1.33E-03	1.40E-02
	Bans0420	2.82E-03	1.51E-02
	Bans0620	-4.46E-04	7.08E-03
	Bans0820	2.62E-03	1.13E-02
controls, no DNA added	NoVns0110	-3.04E-04	6.68E-03
	NoWns0110	-9.04E-04	1.84E-03
	NoVns0210	3.62E-03	1.03E-02
	NoWns0210	-6.80E-04	4.93E-03
	NoVns0310	4.43E-03	1.08E-02
	NoVns0410	6.60E-03	9.05E-03
	NoWns0510	3.03E-03	4.69E-03
	NoWns1110	4.87E-03	4.18E-03

	NoVns0120	2.09E-03	1.21E-02
	NoWns0120	3.71E-03	9.05E-03
	NoVns0220	3.07E-03	6.15E-03
	NoWns0220	2.89E-03	5.01E-03
	NoVns0320	3.48E-03	1.33E-02
	NoVns0420	-1.13E-03	5.46E-03
	NoWns0520	2.43E-03	4.95E-03
	NoWns1120	4.41E-03	8.73E-03

Table S1 Selection coefficients of the hybrid strains and controls.

C. Supplementary information on the publication: Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution

Supplementary information for

Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution

Isabel Rathmann*[1], Mona Förster*[1], Melih Yüksel[1], Lucas Horst[1], Gabriela Petrungaro[1], Tobias Bollenbach[1,2], Berenike Maier[1,3]

*contributed equally

[1] Institute for Biological Physics, Zülpicherstr. 47a, 50674 Köln, University of Cologne

[2] Center for Data and Simulation Science, University of Cologne

[3] Center for Molecular Medicine Cologne, University of Cologne

Supplementary Methods

Strains, media, and cultivation.

The recipient strain Bs166 (*his leu met amyE::P$_{hs}$comK(spc) comK::kan, P$_{comK}$gfp (CBL,cat)* (24) was derived from *B. subtilis* BD630. In this strain, the master regulator for competence, *comK*, is under control of an IPTG-inducible promoter. The reporter strain Bs175 (*his leu met amyE::PhscomK(spc) comK::kan lacA::PrrnE-gfp* (erm)) [1] carries an additional GFP reporter. The donor strains are *B. spizizenii* NRRL B-14472/W23 (hybrid library BSPIZ) and *B. vallismortis* DV1-F-3 (hybrid libraries BVAL, BVAL_single and evolved hybrid libraries).

B. subtilis strains were grown either in complex medium (CM) at 37 or 42 °C or in chemically defined medium supplemented with glucose (DM) or glycerol (DM$_{glycerol}$) at 37°C. CM and DM are based on the Spizizen salts (6 g/l KH_2PO_4, 14 g/l K_2HPO_4, 2 g/l $(NH_4)_2SO_4$, 1 g/l tri-sodium citrate dihydrate). CM was supplemented with 0.5 % D-glucose, 50 µg/ml L-histidine, L-leucine, and L-methionine, 0.02 % casamino acids, 0.1 % yeast extract, and 0.5 mg/ml $MgCl{\cdot}6H_2O$. DM was supplemented with 0.5 % glucose, 0.5 mg/ml sodium glutamate, 50 µg/ml L-histidine, L-leucine, and L-methionine, and 0.2 mg/ml $MgSO_4$. DM$_{glycerol}$ was based on the Spizizen salts without tri-sodium citrate dihydrate added and was supplemented with 0.2 mg/ml $MgSO_4$, 50 µg/ml L-histidine, L-leucine, and L-methionine, and 10% MEM-AA-solution (Roth) and 5% MEM-NEAA-solution (Roth). Supplementation with amino acids was necessary to support growth in this medium. Growth was monitored by measuring the OD_{600} on an Infinite M200 plate reader (Tecan, Männedorf, Switzerland).

Measurement of generation times under different growth conditions

For the four different growth conditions that combine 3 growth media and 2 temperatures (Table S2), we determined the generation time of the recipient strain Bs166 during exponential growth phase by analysing OD curves measured with a plate reader (Tecan, infinite M200 Pro).

Prior to the measurement, the cells were grown overnight in the medium and temperature in question. Then, cells were diluted into columns of a 96-well microtiter plate at 1:100, 1:1000, 1:10000 and growth curves were recorded every 6 minutes with 25 flashes of light with 600 nm wavelength for ~12 h until they reached stationary phase. Shaking amplitude was 2.5 mm and frequency was 250 rpm. The generation time t_g was determined by measuring the time t cells in the dilution series needed to make up for a 10 fold higher dilution and calculating $t_g = t \cdot$

$\frac{log(2)}{log(10)}$. For each condition, at least 60 growth curves were recorded on 3 individual days. Results are shown in Table S3.

Generation of hybrid libraries BVAL and BSPIZ.

In wild type *B. subtilis*, competence for transformation is transient, i.e. bacteria stochastically switch into the state of competence, and exit this state after ~ 2 h [2]. To mimic this behaviour, we used strain Bs166 [1] which carries the master regulator for competence, *comK*, under the control of an IPTG-inducible promoter and induced competence for 2 h in the presence of donor DNA (Fig. 1a).

To generate the hybrid libraries BVAL and BSPIZ, genomic DNA of the donor species was extracted using the Qiagen DNA Blood & Tissue kit. The predominating DNA fragment length specified for this kit is ~ 30 kbp. The recipient Bs166 was grown overnight at 37 °C on LB agar plates, a colony was picked, resuspended in CM, and then grown for 2.5 h at 37 °C in the liquid medium. For transformation, competence was induced with 600 µM IPTG and genomic DNA from *B. vallismortis* or *B. spizizenii*, respectively, at one genome equivalent per recipient cell. After 2 h, cells were washed thrice with phosphate-buffered saline (PBS) and diluted 1:100 in CM. We obtain a population of transformation hybrids from which we randomly pick hybrids to generate monoclonal libraries BVAL and BSPIZ.

To ensure monoclonality, we added two more steps to the protocol. Competent *B. subtilis* are growth-arrested and require ~ 2 h to resume growth after the escape from the competent state. To resolve heteroduplexes formed during transformation, one cell division is required [3]. Therefore, cells were grown until the OD doubled. Exponentially growing *B. subtilis* form chains. To ensure that chain formation does not interfere with monoclonality, 600 µM IPTG was added for 2 h to resolve the chains. During this procedure, cells grew for only few generations, and therefore, selection is minimal.

Then, cells were diluted $1{:}10^5$, plated on LB agar, and grown over night at 37 °C. Next day, 88 single colonies were picked, grown in CM, mixed with DMSO (10% v/v), and stored at – 80 °C.

Generation of BVAL_single library. In the random replacement library BVAL, the orthologously replaced segments contain coding regions as well as regulatory regions of the genome. We addressed the question whether hybrids with individual donor genes replaced showed strong fitness effects. Genes for the BVAL_single library were randomly picked from

a list of *B. subtilis* genes having a homolog in *B. vallismortis.* Homologous genes were found using the Basic Local Alignment Search Tool (blast) by aligning *B. subtilis* genes to *B. vallismortis* and excluding those with 100 % sequence identity and all hits covering less than 90% of the gene's length [2].

When generating this library, we faced the challenge that the replacement had to be free of selective markers, as the latter might have introduced unwanted fitness effects. To this end, we adapted a method for generating marker-less (aka clean) deletion mutants [4] for generating clean replacements (Fig. S1)**.** Specifically, the construction of the BVAL_single library was performed by using the temperature-sensitive pMiniMad2 plasmid [4]. The method is based on single crossover integration of the plasmid into the chromosome at the target locus at 37°C in the presence of selection, where the vector replication is restricted and plasmid excision at the permissive temperature in the absence of selection.

Primers used for the generation of the BVAL_single library were designed by using the SnapGene® software (from Insightful Science) (Dataset S1). The gene replacement strategy is based on the following protocol. Regions of 500 bp up- (5'-UTR) and downstream (3'-UTR) of the target recipient gene and the donor gene were amplified by PCR under standard condition using Q5 polymerase (NEB). Purified DNA fragments containing overlapping parts (OP) were assembled with the NEBuilder Hifi DNA Assembly Master Mix (NEB) into the plasmid backbone, which was also amplified with the Q5 polymerase using the primers pMiniMad_fwd (5'-CACTGGCCGTCGTTTTAC-3') and pMiniMad_rev (5'-TGGCGTAATCATGGTCATAG-3'). The resulting plasmid was isolated from NEB10ß strain (NEB) and used to transform freshly prepared competent *B. subtilis* BD3836 [5]. After transformation cells were plated on LB agar plates supplemented with lincomycin (20µg/ml) and erythromycin (1µg/ml) and incubated overnight at 37°C. Eight single clones were randomly selected and grown in liquid LB medium with lincomycin and erythromycin at 37°C overnight. Next, transformants were cultured at room temperature for 2-3 days and diluted twice a day 1:30 in liquid LB medium. Several clones were tested via PCR for the gene replacement using *B. vallismortis* specific primers (Bval_spec), which bind only to the *B. vallismortis* gene. The loci of the positive clones were amplified by using primers seq_L and seq_R (Dataset S1) and sequenced by Eurofins (Ebersberg, Germany).

The BVAL_single library consists of 24 strains each having a different gene fully replaced by the donor's ortholog (Dataset S1) and additional 19 strains with a partial replacement.

Generation of libraries of evolved hybrid populations.

In the evolution assay, populations growing in parallel on 96-well microtiter plate were diluted regularly. For this, we used an automated system integrated by the company HighRes Biosolutions. The system consists of the following devices: Incubator (StoreX STX44, Liconic), liquid-handling device (Lynx LM900, Dynamics Devices), robotic arm (Acell, HighRes Biosolutions), plate reader (Synergy, BioTek Instruments), shaker (BioShake 3000 elm, QInstruments), plate storage (NanoServe, Acell, HighRes Biosolutions) and delidding device (LidValet, HighRes Biosolutions).

We started by performing one transformation step with *B. vallismortis* DNA, as explained for BVAL (Fig. 1a) resulting in a hybrid population. The population was washed, diluted to an OD of 0.0008 in either CM or DM and split up into 88 wells on a 96-well microtiter plate (~$1x10^5$ cfu/well). On this plate, hybrid populations were then grown in parallel at 37°C, at 500 rpm and 1 mm amplitude for ~450 generations, which amounted to 5 days in CM and 12.5 days in DM. The cells were continuously kept in exponential growth phase by individually diluting each well to OD 0.0008 after 4h in CM and 6h in DM, respectively. After ~ 450 generations, we plated each population on a LB agar plate and picked one hybrid per population, thus generating the monoclonal hybrid libraries BVALevoCM and BVALevoDM. As a reference for both libraries, 88 wells of the recipient Bs166 were evolved in the respective media and the libraries RECevoCM and RECevoDM were generated.

High-throughput competition assay.

For measuring the selection coefficients of different libraries under different growth conditions, and their respective controls, we developed high-throughput competition experiments by making use of the liquid-handling device (Lynx LM900, Dynamics Devices) and a flow-cytometer (Beckman Coulter) that reads out 96-well microtiter plates. To measure the relative fitness, strains are always competed against the reporter strain (RS), i.e. the *gfp*-expressing recipient Bs175.

To prepare for the competition experiment, all 88 strains of the library of interest were grown overnight in the medium later used for the competition on a 96-well microtiter plate together with 5 wells of the control strain (non-transformed Bs166), and 3 wells containing only medium. Additionally, RS was grown in an Erlenmeyer flask under the same conditions. For most competition experiments, we ensured that the bacteria were in exponential growth phase prior to the competition assay by diluting the overnight culture in medium and letting the cells

grow out of lag phase for 2 h and 6 h, in CM and DM, respectively. The only exception was the experiment in which we address the effect of the lag phase, where we immediately started the competition with the overnight culture (Table S2).

For the actual competition experiment, we diluted the cells in PBS to 0.01 OD and then mixed the cells on 96-well microtiter plates. We created 2-4 plates of mixes, aiming at a 1:1 ratio of competitors, by slightly varying the added volume of strain/RS cells. The actual starting fractions were determined with the flow-cytometer. All plates were diluted 1:10 in the growth medium of interest and grown for at least 14 generations in exponential growth phase (in CM at 37°C and 42°C: 4 h; in DM and $DM_{glycerol}$: 16 h) in a shaker at 37°C or 42°C, respectively. For the competition assay including lag phase, cells were grown for 6 h, ~ 2 h in lag and 4 h in exponential growth phase. Table S2 gives an overview of the different growth conditions. We prevented cells from reaching stationary phase. In CM, cells are still exponential after 4h. In the case of the 16h competition in DM and $DM_{glycerol}$, cells were diluted during competition to retain exponential growth. After the competition, all plates were diluted in PBS and the final ratio of strain/ RS cells was measured with the flow-cytometer.

The selection coefficient of the i-th strain was calculated with the fraction x_i of the strain of interest and x_{RS} of the reporter strain at the start t_0 and end time point t as follows

$s_{i,RS} = \frac{t_g}{t-t_0} ln\left(\frac{x_i(t)/x_{RS}(t)}{x_i(t_0)/x_{RS}(t_0)}\right)$. t_g is the generation time of the recipient in the respective media (Table S3). For the DFEs, the selection coefficients $s_{i,r}$ were calculated relative to the recipient's fitness measured on the same experimental plate (details in Supplementary Methods).

Each experiment ran independently on at least 3 days and the selection coefficients were averaged values over days and plates. Not all 88 library samples could be considered, as competitions with starting fractions of 40-60% measured on at least 2 days were required. Detailed information on the analysis of the flow-cytometry data is given below. For each library, the measured selection coefficients together represent the DFE for the applied conditions.

The control DFEs were determined for each growth condition for 82 samples of the recipient strain competing against RS with the same assay as used for the libraries.

Analysis of flow cytometry data. In the competition assay, we determined the fractions of the library strains i and the competitor, the fluorescent reporter strain (RS), at the start and end time

point of competition. We measured about 30000 events per sample at the flow-cytometer, exported the raw data in the FCS3.0 format and processed it with Matlab after importing it with the fca_readfcs function [6]. To include all events ascribed to cells and exclude debris, we created a gate on the forward and side scatter data (FSC and SSC) and counted the events in the gates to obtain the raw fraction of strain i and RS cells.

Two kinds of corrections were applied to the raw fractions determined by the flow-cytometer. First, during the competition experiment, three positions on the 96-well microtiter plate contained only RS cells that were used to correct for debris that may have been falsely detected as cells. We detected a fraction d of events that was falsely categorized as non-fluorescent cells per plate but that in fact were due to impurities in the solution or non-fluorescing RS cells. As these events would have falsely been interpreted as non-*gfp*-expressing cells in the mixed samples, we applied the following correction. For each competition sample, we computed the fraction d of measured RS and then increased the fraction of RS and decreased the fraction of strain i by this amount. We note that d, amounted to a minor correction only. For the further analysis, we only considered samples with starting fractions of 40-60 %.

Second, we corrected for the slightly different fitness of the reporter strain RS compared to the recipient r with the five wells on the 96-well plate that were used for recipient competition during every experiment. First, we measured for each strain i the selection coefficient relative to the direct competitor RS in the same well as $s_{i,RS} = \frac{t_g}{t-t_0} ln\left(\frac{x_i(t)/x_{RS}(t)}{x_i(t_0)/x_{RS}(t_0)}\right)$. To calculate the coefficient $s_{i,r}$ relative to the recipient, we determined the mean selection coefficient of the recipient and RS as $\overline{s}_{r,RS}$ from the 5 recipient wells on the same plate with $s_{r,RS} = \frac{t_g}{t-t_0} ln\left(\frac{x_r(t)/x_{RS}(t)}{x_r(t_0)/x_{RS}(t_0)}\right)$. For each strain i, the selection coefficients were finally determined as $s_{i,r} = s_{i,RS} - \overline{s}_{r,RS}$ with the average of recipient fitness per day and plate. Afterwards, for each library strain, for every day the values from different plates were averaged and the results averaged over the individual measurement days. Each library was measured on at least 3 days and only samples were considered in the DFE that were successfully measured on at least 2 days.

Statistical analysis of DFE.

For each distribution of fitness effects, we defined the strains with large fitness effects as outliers. For this, we used the DFE of the control measurement obtained with the same assay and performed a two-sided z-test with each strain of the hybrid library, obtaining p-values. We

then applied the Bonferroni correction for multiple testing and defined outliers at the significance level of $\alpha = 0.05$. The DFE without the large effect transfer outliers was assumed to be dominated by the remaining small effect transfers.

In order to compare the distribution of small fitness effects for the different conditions, we first excluded outliers from the DFEs and then performed bootstrap analysis with 10000 resamples on the sample statistics mean and standard deviation. From the sampling distributions of the mean and standard deviation, we obtained the 95% confidence intervals for the sample statistics.

Whole genome sequencing.

Clonal genomes of ancestral and evolved populations were obtained using next generation sequencing (NGS) methods, in particular Illumina HiSeq. Samples were grown overnight in CM at 37 °C. A 2 ml aliquot of that culture was pelleted at 16.7 xg for 3 min, decanted, and then frozen at - 20 °C. Genomic DNA was isolated from the frozen pellet using the Qiagen DNeasy Blood & Tissue Kit (Hilden, Germany) according to the manufacturer's instructions. A small aliquot of the isolated DNA was run on a 1% agarose gel with a 1 kb plus DNA Ladder (Thermo Scientific) to check for degradation. Non-degraded samples were sent to Eurofins Genomics (Ebersberg, Germany) for NGS. Sequencing was performed on an Illumina HiSeq 3000/4000 system with 150 bp paired-end reads and an average depth of ~ 450. Sequence analysis was performed using the pipeline described in Power et al [1] and orthologous replacements, insertions, deletions and duplications as well as mutations were detected as described in the Supplementary Methods.

Detection of orthologous replacement, insertion, deletion, and duplication. We used the protocol developed by Power et al for detecting orthologously replaced segments [1]. In short, we created a master list of positions and bases representing point differences between donor and recipient genomes. The sequencing reads of the library strains were mapped to the recipient (NCBI RefSeq NC_000964.3, [7]) using Burrows-Wheeler Aligner (v.0.7.17) [8], and variants were called with the mpileup function from samtools (samtools 1.8) and the call function from bcftools (bcftools 1.8) [9]. At each donor-recipient divergence site, the sequence of the library strain has a recipient (R) or donor (D) consensus. We inferred transfer segments in the evolved sequence symmetrically in both directions, using the following algorithm: (1) The 3'-end of each segment is marked by a D site that is followed by a sequence of 5 consecutive R alleles.

The 3'-end coordinate of the segment was then assigned to the midpoint between the last D site and first of the five R sites. (2) The 5'-end of each transfer segment was marked by a D site that is preceded by a sequence of five consecutive R alleles. The 5'-end coordinate of the segment was then assigned to the midpoint between the last of the five R sites and the first D sites. Putting the 3'-ends and the 5'-ends together as orthologously replaced segment, only those were accepted that contain a minimum of two D sites.

Allele differences between library sequences and R/D consensus sequences that were not explained by orthologous replacement were ascribed to de novo point mutations. Alleles inserted or deleted that were not explained by orthologous replacement were ascribed to indels.

Start and end positions of the accessory recipient/donor genes were defined using the coverage of reads when mapping the sequenced donor reads on the recipient and vice versa. We calculated the coverage per base from the mapped reads using bedtools (v2.26.0) [10]. Parts of the genome, with a coverage below 50 for 150 subsequent bases we define to be the accessory genome. All genes within these regions are accessory genes. Every part of the genome that does not belong to the accessory genome has a homologue in the other species' genome and is called core genome.

To identify deletions and duplications the genome coverage, per base, of library strains mapped to recipient was calculated. After smoothing, the coverage with a sliding window of 30 bps, all segments in which the coverage dropped down to zero for at least ten bases in a row were defined to be deletions. These can be detected all over the genome of the library strains, both the core and the accessory regions. To be accepted as duplication, the coverage has to hit twice the genome-wide average coverage for ten subsequent bases.

We detected integrations of donor specific genes by mapping the sequencing reads of the library strains stringently to the donor. Without integrations from the donor specific genes, we would expect no coverage in the accessory regions of the donor. In case we found coverage in those regions, and it is a maximum of 1.5 σ smaller than the genome-wide average coverage we assumed this part to be integrated into the library strain. Testing this method reveals a resolution down to 50 bp long integrations.

All protocols described above were tested using artificially designed sequencing data.

Supplementary Figures

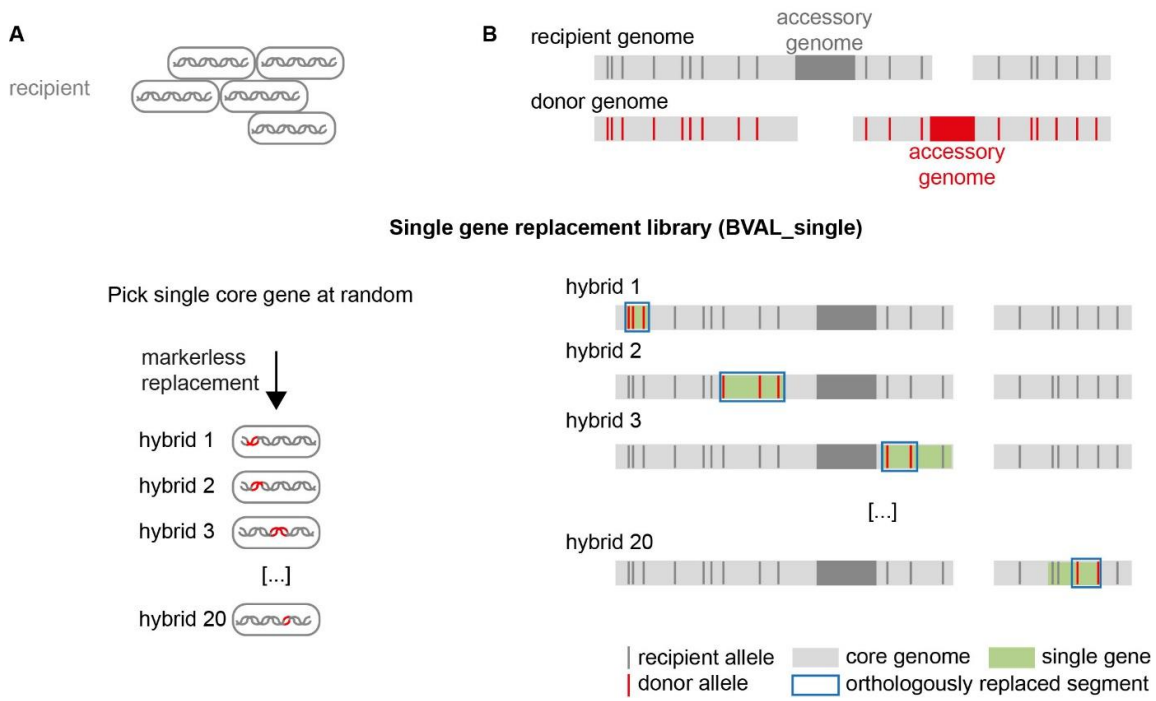


Fig. S1 Generation of the single gene replacement library BVAL_single. A) The single gene replacement libraries were generated by randomly selecting a core gene and fully or partially replacing the recipient allele by the donor allele by means of marker-less replacement. B) The library consists of hybrids in which a single core gene has been fully or partially replaced by its donor ortholog.

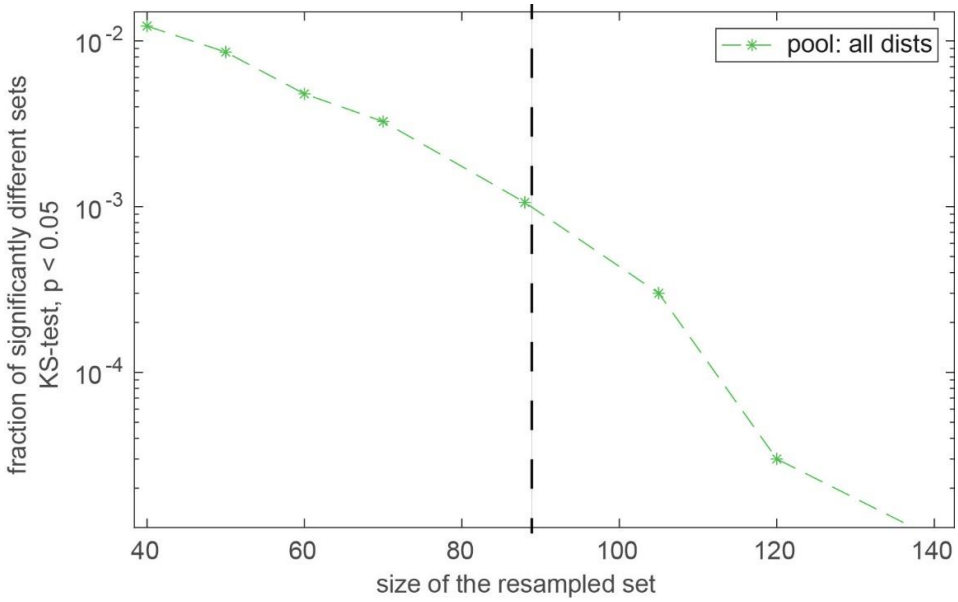


Fig. S2 Subsampling analysis of DFE. Random subsampling of the selection coefficients suggests that the experimentally determined DFEs are likely to be representative for the DFE of transformation. DFE data from the BVAL, BSPIZ and control measurement in complex medium are pooled and subsamples are randomly drawn from the DFE data 100000 times. Using a KS-test, each subsample is tested against the pooled data set and assigned a p-value. Fractions of resampled datasets that are significantly different from the pooled data set ($p < 0.05$) are shown as a function of sample size. We conclude that at a sample size of 88 strains, the probability that we do not describe the global shape of the "true" DFE correctly is on the order of 10^{-3}.

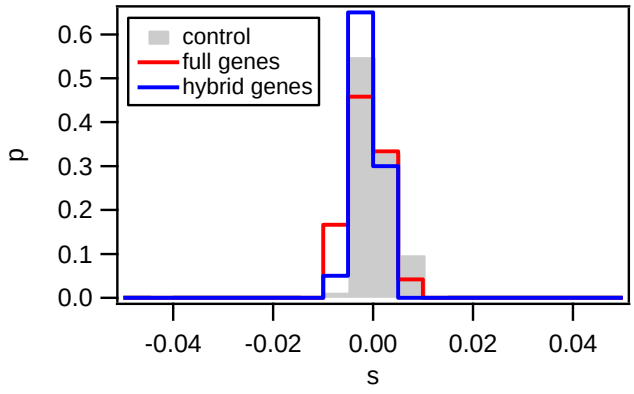


Fig. S3 Analysis of BVAL_single library. DFE form Fig. 2 whereby the DFEs of fully replaced genes and partially replaced genes are shown separately. Neither the mean nor the variance of the distributions is significantly different from the control distribution.

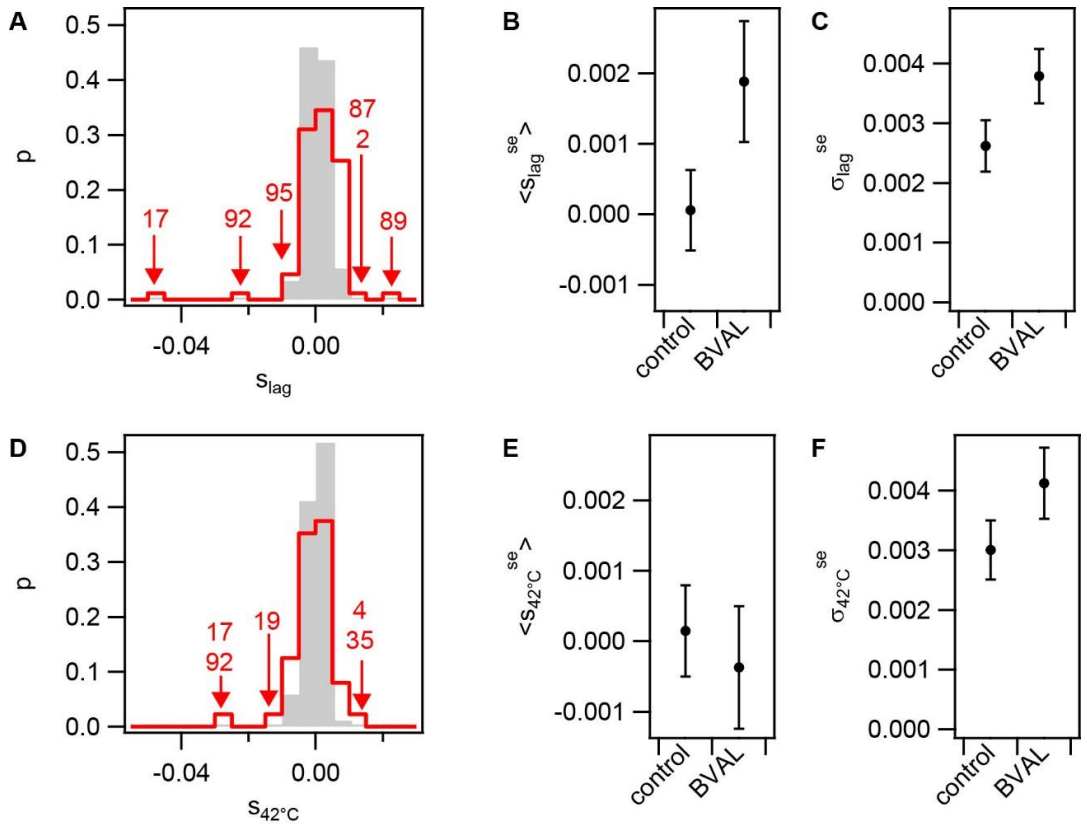
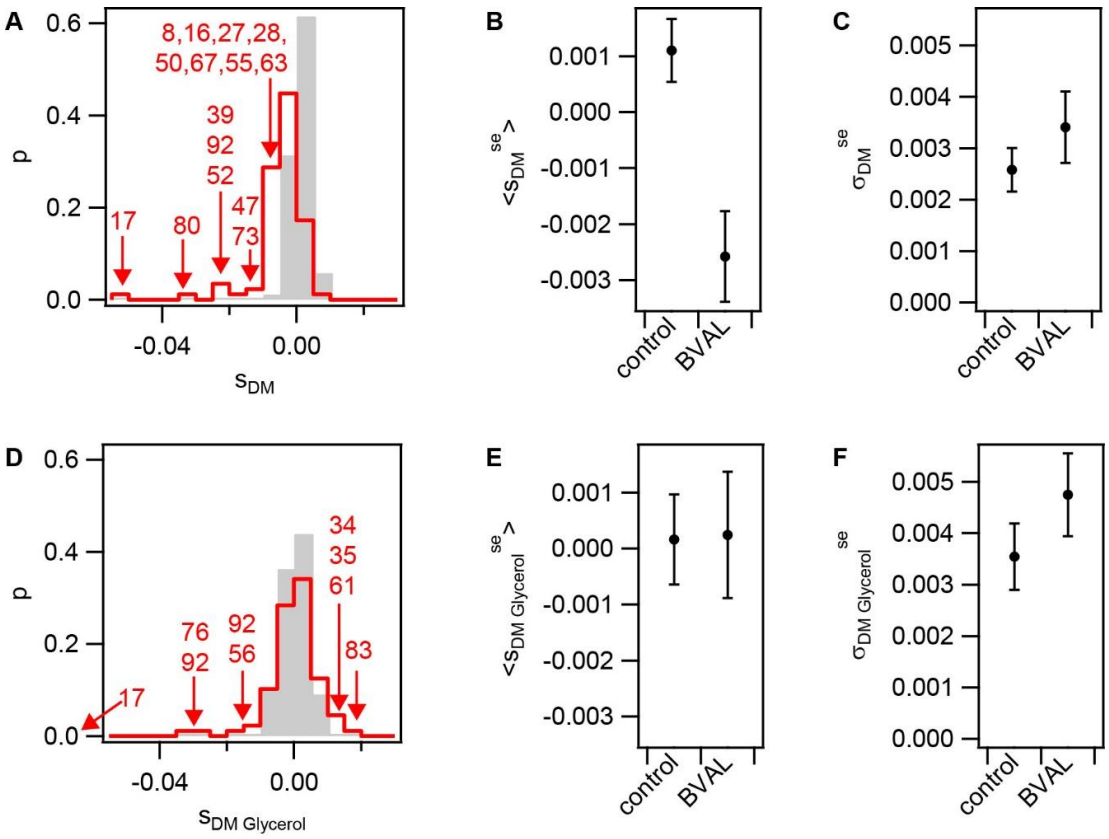


Fig. S4 Distribution of fitness effects of BVAL library with lag phase and at 42 °C. Selection coefficients of library BVAL were determined in complex medium including the lag phase A) - C) and at 42 °C D) – F). a) Distribution of selection coefficients s_{lag} resulting from competition experiments between single strains of the BVAL library and the recipient expressing *gfp* (Bs175) are shown (red). Control distribution (grey). B) Mean selection coefficients s_{lag}^{se} and C) standard deviation σ_{lag}^{se} of core distributions after removing outliers. D) Distribution of selection coefficients $s_{42°C}$ resulting from competition experiments between single strains of the BVAL library and the recipient expressing *gfp* (Bs175) are shown (red). Control distribution (grey). E) Mean selection coefficients $s_{42°C}^{se}$ and F) standard deviation $\sigma_{42°C}^{se}$ of core distributions after removing outliers. Error bars: confidence intervals obtained from bootstrap analysis.



312

313 **Fig. S5 Distribution of fitness effects of BVAL library in defined medium.** Selection
314 coefficients of library BVAL were were determined in complex medium. Competitors were
315 mixed in exponential phase and grown for 16 h. A) Distribution of selection coefficients s_{DM}
316 resulting from competition experiments between single strains of the BVAL library and the
317 recipient expressing *gfp* (Bs175) are shown (red). Control distribution (grey). B) Mean selection
318 coefficients s^{se}_{DM} and C) Standard deviation σ^{se}_{DM} of core distributions after removing outliers.
319 D) Distribution of selection coefficients $s_{DM\ Glycerol}$ resulting from competition experiments
320 between single strains of the BVAL library and the recipient expressing *gfp* (Bs175) are shown
321 (red). Control distribution (grey). E) Mean selection coefficients $s^{se}_{DM\ Glycerol}$ and F) Standard
322 deviation $\sigma^{se}_{DM\ Glycerol}$ of core distributions after removing outliers. Error bars: confidence
323 intervals obtained from bootstrap analysis.

324

Supplementary Tables

	B. spizizenii (BSPIZ)	*B. vallismortis (BVAL)*
# SNPs	1274 (1977)	334 (655)
fraction core genome replaced	0.5 (0.8) %	0.1 (0.2) %
length of replaced segment	4080 (4598)	1289 (1220)
# replaced segments	4 (4)	3 (4)
# hit genes	21 (31)	6 (11)
# insertions	0.5 (0.5)	0
# deletions	0.2 (0.2)	0.1 (0.3)
# de novo mutations	3 (8)	0.3 (0.7)

Table S1 Genomic changes of ten randomly chosen strains from BSPIZ and BVAL. A randomly chosen subset of ten strains of each library were sequenced. Mean (standard deviations) are shown.

Medium	Competition in exp. phase	Temperature	Library (sample size)
CM	4h	37°C	BVAL (87), BVAL_single (43), BSPIZ (87), BVALevoCM (83), RECevoCM (86)
	4h	42°C	BVAL (88)
	6h (incl. ~2h lag)	37°C	BVAL (87)
DM	16h	37°C	BVAL (88) BVALevoDM (88), RECevoDM (85)
DM_Glycerol	16h	37°C	BVAL (86)

Table S2: The hybrid libraries were examined under different growth conditions with the high-throughput competition assay. Cells were mostly competed in exponential growth phase and the only exception was the measurement where the ~ 2h long lag phase was included (incl. ~ 2h lag). Library sizes can be less than 88, as selection coefficients could not be measured for all samples. Each library was characterized on at least 3 days and for each condition, a control was included with the same assay.

Growth medium	Temperature [°C]	Generation time [min]
CM	37	17,1 ±0,2
CM	42	14,9 ± 0,3
DM	37	39,0 ± 0,2
$DM_{glycerol}$	37	48,3 ± 0,3

Table S3 Generation time of the recipient strain Bs166 in different growth conditions. Errors were calculated by performing error propagating with the standard error of the mean of each individual measurement.

Dataset S1 Genes replaced in BVAL_single

Dataset S2 Genetic changes of 10 fittest evolved strains

1. Power JJ, Pinheiro F, Pompei S, Kovacova V, Yuksel M, Rathmann I, et al. Adaptive evolution of hybrid bacteria by horizontal gene transfer. P Natl Acad Sci USA. 2021;118(10):e2007873118. doi: ARTN e200787311810.1073/pnas.2007873118. PubMed PMID: WOS:000627429100011.

2. Maier B. Competence and Transformation in Bacillus subtilis. Curr Issues Mol Biol. 2020;37:57-76. doi: 10.21775/cimb.037.057. PubMed PMID: WOS:000508654200005.

3. Dalia AB, Dalia TN. Spatiotemporal Analysis of DNA Integration during Natural Transformation Reveals a Mode of Nongenetic Inheritance in Bacteria. Cell. 2019;179(7):1499-511. doi: 10.1016/j.cell.2019.11.021. PubMed PMID: WOS:000502546200009.

4. Patrick JE, Kearns DB. MinJ (YvjD) is a topological determinant of cell division in Bacillus subtilis. Mol Microbiol. 2008;70(5):1166-79. doi: 10.1111/j.1365-2958.2008.06469.x. PubMed PMID: WOS:000261070300010.

5. Maamar H, Dubnau D. Bistability in the Bacillus subtilis K-state (competence) system requires a positive feedback loop. Mol Microbiol. 2005;56(3):615-24. doi: 10.1111/j.1365-2958.2005.04592.x. PubMed PMID: WOS:000228179100005.

6. fca_readfcs [Internet]. 2022.

7. O'Leary NA, Wright MW, Brister JR, Ciufo S, McVeigh DHR, Rajput B, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733-D45. doi: 10.1093/nar/gkv1189. PubMed PMID: WOS:000484575500002.

8. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60. doi: 10.1093/bioinformatics/btp324. PubMed PMID: WOS:000267665900006.

9. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008. doi: ARTN giab00810.1093/gigascience/giab008. PubMed PMID: WOS:000637191300010.

10. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2. doi: 10.1093/bioinformatics/btq033. PubMed PMID: WOS:000275243500019.

D. Supplementary information on the manuscript: Transformation opens up new paths for evolving *B. subtilis* populations

A. Generation times

Strain	Generation time [min]	Growth condition
Bs210	16.1 ± 0.2	in CM at 37 °C
Bs224	16.2 ± 0.3	in CM at 37 °C
Bs166	17.1 ± 0.4	in CM at 37 °C

Figure A.1: Generation times of Bs166 (ancestor), Bs210 (pre-adapted to growth in liquid environment) and Bs224 (pre-adapted to growth in structured environment).

B. Whole genome sequencing data

B.1. Pre-adaptation in structured environment

Strain	Pos	Ref	Alt	Mut	Gene
pl. 1 col. 11	1316064	G	A	SNP	*rapA*
	2403267	G	T	SNP	*gudB*
pl. 1 col. 9	1316653	C	T	SNP	*rapA*
	2403258	TG	T	indel	*gudB*
	2403261	T	A	SNP	*gudB*
	2403262	C	A	SNP	*gudB*
	2403263	C	A	SNP	*gudB*
	3527388	A	T	SNP	*epsC*
pl. 2 col. 4	328580	G	C	SNP	-
	2015523	G	A	SNP	*gltC*
	3152512	C	A	SNP	*menF*
	3515756	C	T	SNP	*epsN*
	3733060	G	A	SNP	*tkmA*
pl. 1 col. 4	118783	T	C	SNP	*rplK*
	954076	GT	G	indel	*spo0M*
	954079	T	C	SNP	*spo0M*
	1519136	AGCTGCT	AGCTGCTGCT	indel	*kinC*
	1918177	A	T	SNP	*lexA*
	2403237	A	C	SNP	*gudB*
	2403241	C	A	SNP	*gudB*
	2482645	T	G	SNP	*polyA*
	3732940	T	G	SNP	*tkmA*
pl. 2 col. 10	713796	GAAAAA	GAAAA	indel	*yerA*
	1316328	T	A	SNP	*rapA*
	1552907	TAAAAA	TAAAA	indel	*ftsW*
	2403057	G	A	SNP	*gudB*
pl. 2 col. 9	2403057	G	A	SNP	*gudB*
	1580033	GTTTTTT	GTTTTT	indel	-
	3628263	CGCAGGCGCCTGCAGGCGCC	CGCAGGCGCC	indel	-

Figure B.2: Genetic variants found in the sequenced clones of populations pre-adapted to growth in a structured environment.

B.2. Evolution experiment in structured environment

B.2.1. Bs210ctl

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210ctl_01	1716942			GTTTT	GTTTTT	indel	*sigD*
Bs210ctl_02	2403057			G	A	SNP	*gudB*
	1716942			GTTTT	GTTTTT	indel	*sigD*
Bs210ctl_03	1178348			C	T	SNP	*yitI*
	1580033			GTTTTTT	GTTTTT	indel	-
	1716942			GTTTT	GTTTTT	indel	*sigD*
Bs210ctl_04	166813			TCC	TC	indel	*rrnH-16S*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	2403094			GAAAA	GAAAAA	indel	*gudB*
	2523573			A	C	SNP	*yqxC*
Bs210ctl_05	863332			G	A	SNP	*rbn*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	2402134			C	CT	indel	*gudB*

Figure B.3: Bs210ctl - Genetic variants found in the sequenced clones grown for 5 cycles in structured environment. The variants that emerged already during pre-adaptation are shown in gray.

B.2.2. Bs210hyb

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210hyb_01	806208			G	A	SNP	*mtrA*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	1753279	1753546	268			replacement	*spoIIIE*
	1753640	1763605	9966			replacement	*spoIIIE*
							ymfC
							bcbE
							ymfF
							ymfH
							efpI
							ymfJ
							ymfKn
							ymfKc
							rodZ
							pgsA
							cinA
	1763914	1764322	409			replacement	*cinA*
	1764525	1768425	3901			replacement	*recA*
							pbpX
							rny
	1771942	1777367	5426			replacement	*kbl*
							miaB
							ricA
							cotE
							mutS
	3765022			ATTTT	ATTTTT	indel	-

Figure B.4: Genetic variants found in Bs210hyb_01 after 5 cycles on complex medium agar plates. The variants that emerged already during pre-adaptation are shown in gray.

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210hyb_02	1716942			GTTTT	GTTTTT	indel	*sigD*
	1756974	1757957	984			replacement	*ymfF*
	1758254	1770506	12253			replacement	*ymfF*
							ymfH
							efpI
							ymfJ
							ymfKn
							ymfKc
							rodZ
							pgsA
							cinA
							recA
							pbpX
							rny
							pdeB
							spoVS
							tdh
	3253635					SNP	*comP*
	3499066	3500469	1404			replacement	*yvfR*
							rsbQ
							rsbP

Figure B.5: Genetic variants found in Bs210hyb_02 after 5 cycles on complex medium agar plates. The variants that emerged already during pre-adaptation are shown in gray.

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210hyb_03	1716942			GTTTT	GTTTTT	indel	*sigD*
	1741396	1771039	29644			replacement	*ylxY*
							mlpA
							ymxH
							spoVFA
							spoVFB
							asd
							dapG
							dapA
							rnjB
							tepA
							tepJ
							spoIIIE
							ymfC
							bcbE
							ymfF
							ymfH
							efpI
							ymfJ
							ymfKn
							ymfKc
							rodZ
							pgsA
							cinA
							recA
							pbpX
							rny
							pdeB
							spoVS

Figure B.6: Genetic variants found in Bs210hyb_03 after 5 cycles on complex medium agar plates. The variants that emerged already during pre-adaptation are shown in gray.

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210hyb_04	1716942			GTTTT	GTTTTT	indel	*sigD*
	1760200	1764337	4138			replacement	*efpI*
							ymfJ
							ymfKn
							ymfKc
							rodZ
							pgsA
							cinA
Bs210hyb_05	582086			G	A	SNP	*ydfB*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	1759279	1764812	5534			replacement	*ymfH*
							efpI
							ymfJ
							ymfKn
							ymfKc
							rodZ
							pgsA
							cinA
							recA
	3125754	3127051	1298			replacement	*asnB*
	3127534	3128451	918			replacement	*asnB*
							metK
	3130323	3130488	166			replacement	*pckA*
	3130575	3130719	145			replacement	*pckA*
	3131193	3132280	1088			replacement	*ytmB*
							ytmA
	3132518	3136003	3486			replacement	*ytlA*
							ytlC
							ytlD
							rppG
							ytkC
	3320691	3322272	1582			replacement	*lipA*
							lytH
	3322738	3322837	100			replacement	*fisB*
	3323137	3323143	7			replacement	*fisB*
	3323342	3323440	99			replacement	*yunC*
	3323717	3324488	772			replacement	*yunD*

Figure B.7: Genetic variants found in Bs210hyb_04 and Bs210hyb_05 after 5 cycles on complex medium agar plates. The variants that emerged already during pre-adaptation are shown in gray.

B.2.3. Bs224ctl

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224ctl_01	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_02	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_03	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_04	1316653			C	T	SNP	*rapA*
	1859424			G	T	SNP	*pksS*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_05	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*

Figure B.8: Bs224ctl - Genetic variants found in the sequenced clones grown for 5 cycles in structured environment. The variants that emerged already during pre-adaptation are shown in gray.

B.2.4. Bs224hyb

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224hyb_01	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_02	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_03	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_04	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_05	1316653			C	T	SNP	*rapA*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*

Figure B.9: Bs224hyb - Genetic variants found in the sequenced clones grown for 5 cycles in structured environment. The variants that emerged already during pre-adaptation are shown in gray.

B.3. Evolution experiment in liquid environment

B.3.1. Bs210ctl

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210ctl_C02	211383			C	T	SNP	-
	1716942			GTTTT	GTTTTT	indel	sigD
Bs210ctl_C03	1716942			GTTTT	GTTTTT	indel	sigD
Bs210ctl_C04	1716942			GTTTT	GTTTTT	indel	sigD
Bs210ctl_C05	1716942			GTTTT	GTTTTT	indel	sigD
Bs210ctl_C06	1716942			GTTTT	GTTTTT	indel	sigD

Figure B.10: Genetic variants found in Bs210ctl clones after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

B.3.2. Bs210hyb

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs210hyb_C02	166813			TCC	TC	indel	*rrnH-16S*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	2239581			G	A	SNP	*yomW*
Bs210hyb_C03	1716942			GTTTT	GTTTTT	indel	*sigD*
	1868741			T	C	SNP	*nrdI*
	3239608			T	A	SNP	-
Bs210hyb_C04	1171029			C	T	SNP	*yitA*
	1625830	1625842	13			replacement	*pyrAB*
	1716942			GTTTT	GTTTTT	indel	*sigD*
	4106257	4109105	2849			replacement	*qdoI* *qdoR* *yxnA*
	4109136			C	T	SNP	-
	4109139			T	C	SNP	-
	4110908	4118330	7423			replacement	*yxaC* *yxaB* *glxK* *gntR* *gntK* *gntP* *gntZ*
Bs210hyb_C05	1711637	1711672	36			replacement	*flhG*
	1712495	1712549	55			replacement	*cheB*
	1714969	1717084	2116			replacement	*cheW* *cheC* *cheD* *sigD*
	1718872	1718922	51			replacement	*tsf*
	1719549	1722157	2609			replacement	*tsf* *pyrH* *frr* *uppS* *cdsA*
	1722480	1722600	121			replacement	*cdsA*
Bs210hyb_C06	1716942			GTTTT	GTTTTT	indel	*sigD*
	2152052			A	C	SNP	-

Figure B.11: Genetic variants found in Bs210hyb clones after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

B.3.3. Bs224ctl

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224ctl_C02	166813			TCC	TC	indel	*rrnH-16S*
	1316653			C	T	SNP	*rapA*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_C03	1316653			C	T	SNP	*rapA*
	2186959			T	C	SNP	-
	2186960			A	G	SNP	-
	2186965			T	A	SNP	-
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224ctl_C04	1316653			C	T	SNP	*rapA*
	1705733			ATTTTT	ATTTT	indel	*fliQ*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*

Figure B.12: Genetic variants found in the sequenced clones from colonies Bs224ctl_C02 - C04 after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224ctl_C05	166813			TCC	TC	indel	*rrnH-16S*
	1316653			C	T	SNP	*rapA*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	2518567			G	T	SNP	*spo0A*
	3527388			A	T	SNP	*epsC*
Bs224ctl_C06	1204698			C	A	SNP	-
	1316653			C	T	SNP	*rapA*
	1707430			TCACTA	T	indel	*flhB*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*

Figure B.13: Genetic variants found in the sequenced clones from colonies Bs224ctl_C05 and C06 after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

B.3.4. Bs224hyb

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224hyb_C02	1316653			C	T	SNP	*rapA*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	2503340			A	G	SNP	*bcd*
	3527388			A	T	SNP	*epsC*
	3638127			ATTTTT	ATTTT	indel	*flgL*
Bs224hyb_C03	1316653			C	T	SNP	*rapA*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_C04	1316653			C	T	SNP	*rapA*
	1699075			G	GAAAGCCAGCTG	indel	*fliK*
	2224910			T	C	SNP	*yonO*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*
Bs224hyb_C05	1316653			C	T	SNP	*rapA*
	1693682			CAAA	CAAAA	indel	*fliF*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3527388			A	T	SNP	*epsC*

Figure B.14: Genetic variants found in the sequenced clones from colonies Bs224hyb_C02 - C05 after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

Strain	Pos (Start)	End	Length	Ref	Alt	Mut	Gene
Bs224hyb_C06	43132	43697	566			replacement	*trmNF* *yazA*
	1128868	1129538	671			replacement	*ntdA*
	1316653			C	T	SNP	*rapA*
	1874994	1879685	4692			replacement	*spoVK* *hflX* *ynbB* *glnR* *glnA*
	2317441	2317600	160			replacement	*bpsA*
	2403258			T	TG	indel	*gudB*
	2403261			T	A	SNP	*gudB*
	2403262			C	A	SNP	*gudB*
	2403263			C	A	SNP	*gudB*
	3246542			A	G	SNP	-
	3527388			A	T	SNP	*epsC*
	3899893	3902235	2343			replacement	*ywdD* *ywzG* *pdxK* *ywdA* *sacA*
	3922941	3923060	120			replacement	*slrC*

Figure B.15: Genetic variants found in in the sequenced clone from colony Bs224hyb_C06 after 5 cycles in liquid medium. The variants that emerged already during pre-adaptation are shown in gray.

C. Timelapse images

C.1. Timelapse images of pre-adaptation to growth in structured environments

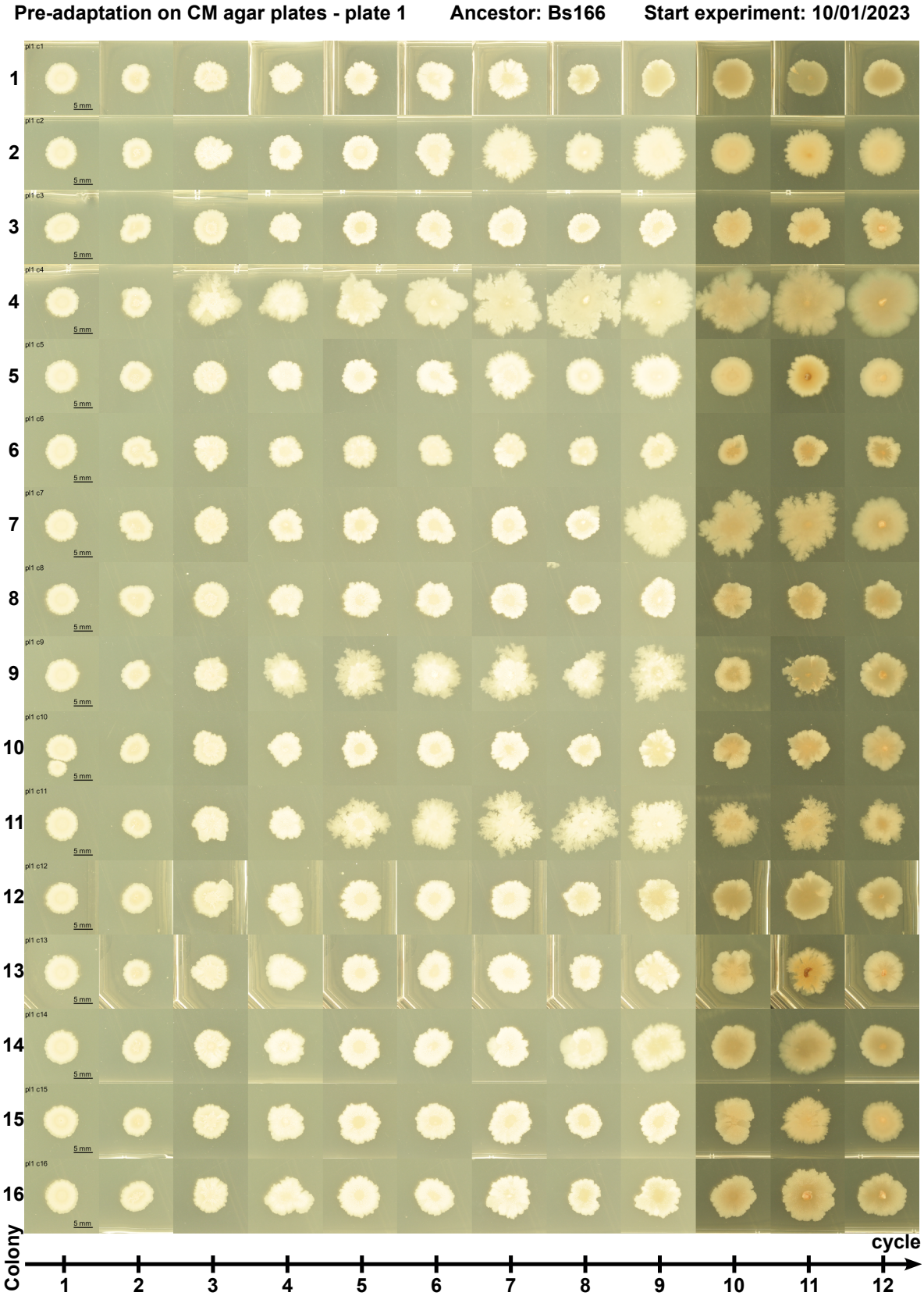


Figure C.16: Timelapse pre-adaptatation of Bs166 - plate 1.

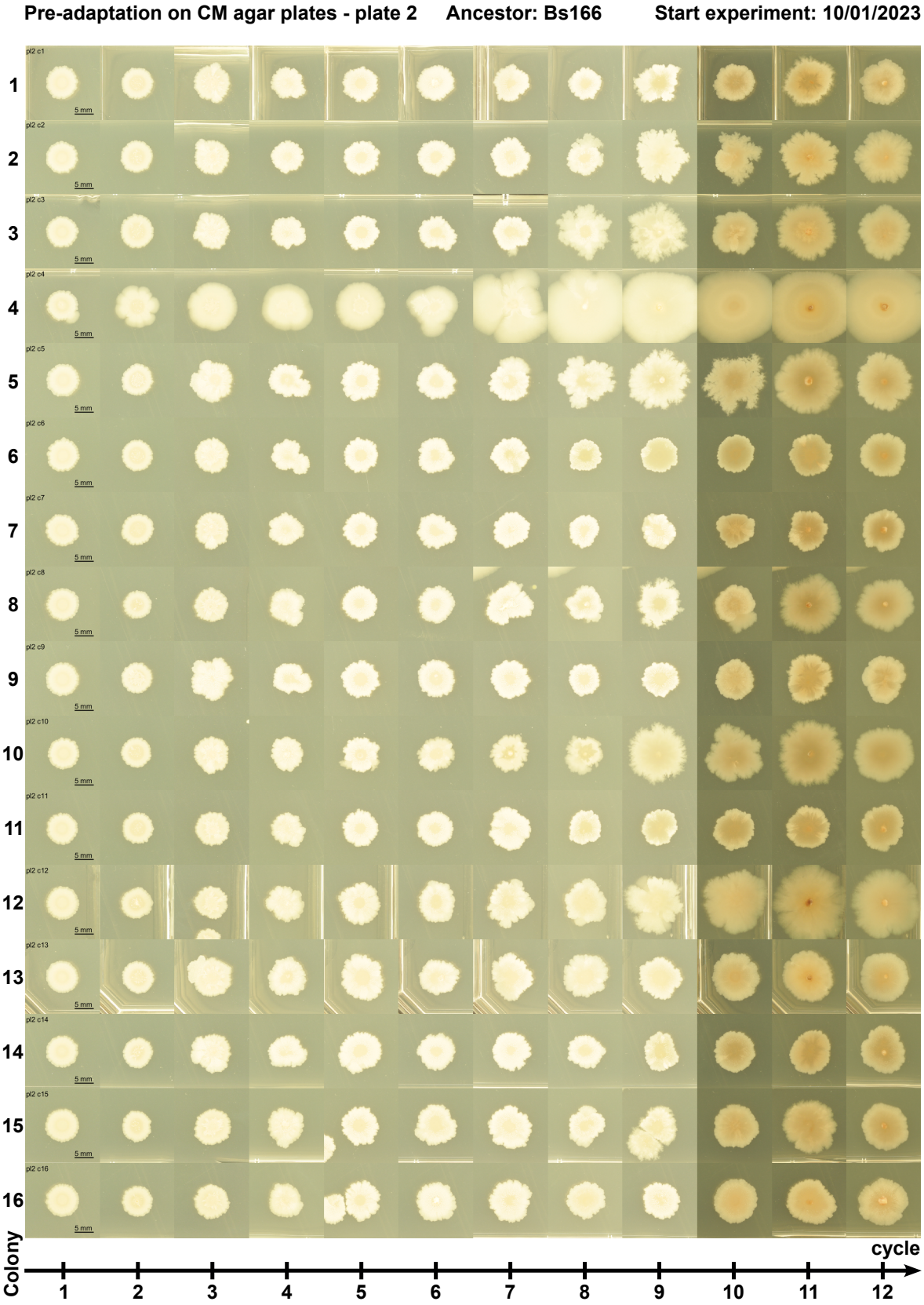


Figure C.17: Timelapse pre-adaptatation of Bs166 - plate 2.

C.1.1. Colony size over time during the pre-adaptation

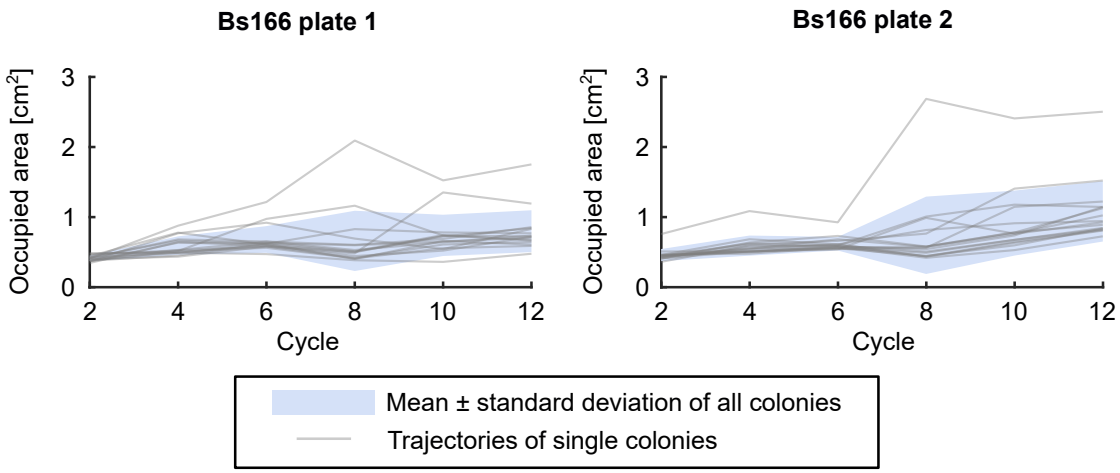


Figure C.18: Occupied area over time. In blue, the average colony size during adaptation is shown. The trajectory of each single colony is depicted in gray.

C.2. Timelapse images of evolution experiment in structured environments

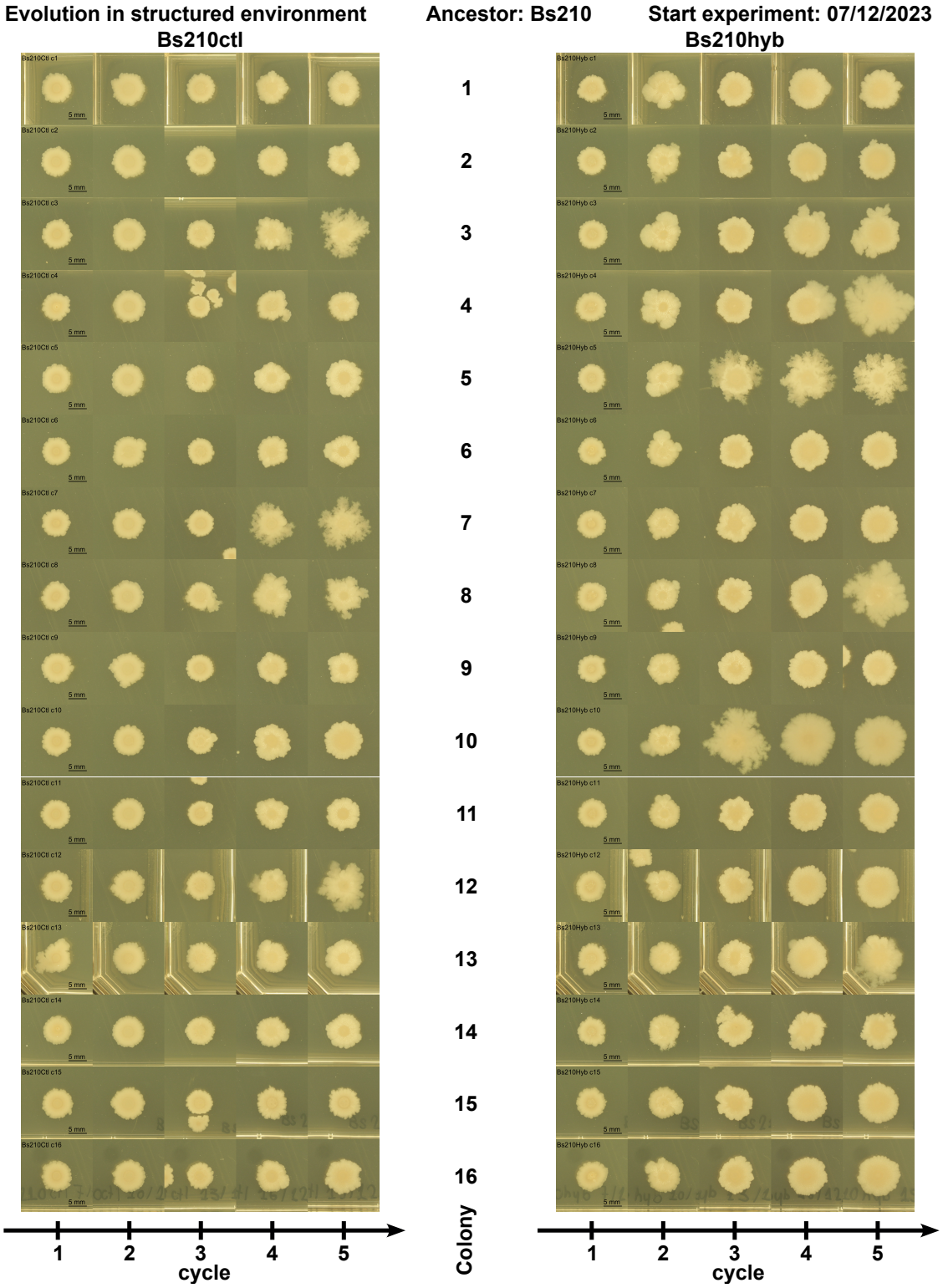


Figure C.19: Timelapse evolution experiment Bs210ctl/Bs210hyb.

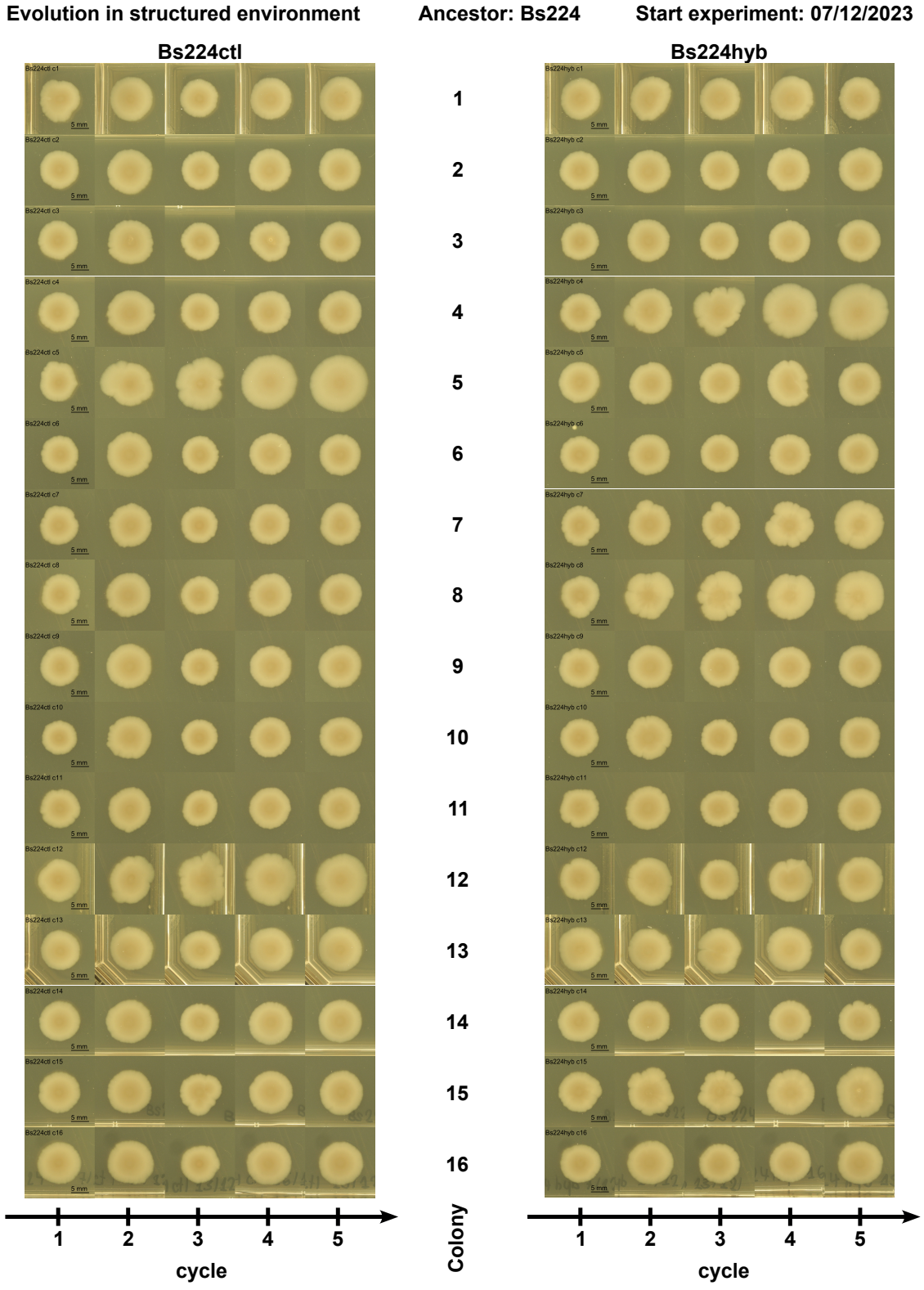


Figure C.20: Timelapse evolution experiment Bs224ctl/Bs224hyb.

C.2.1. Colony size over time during the evolution experiment

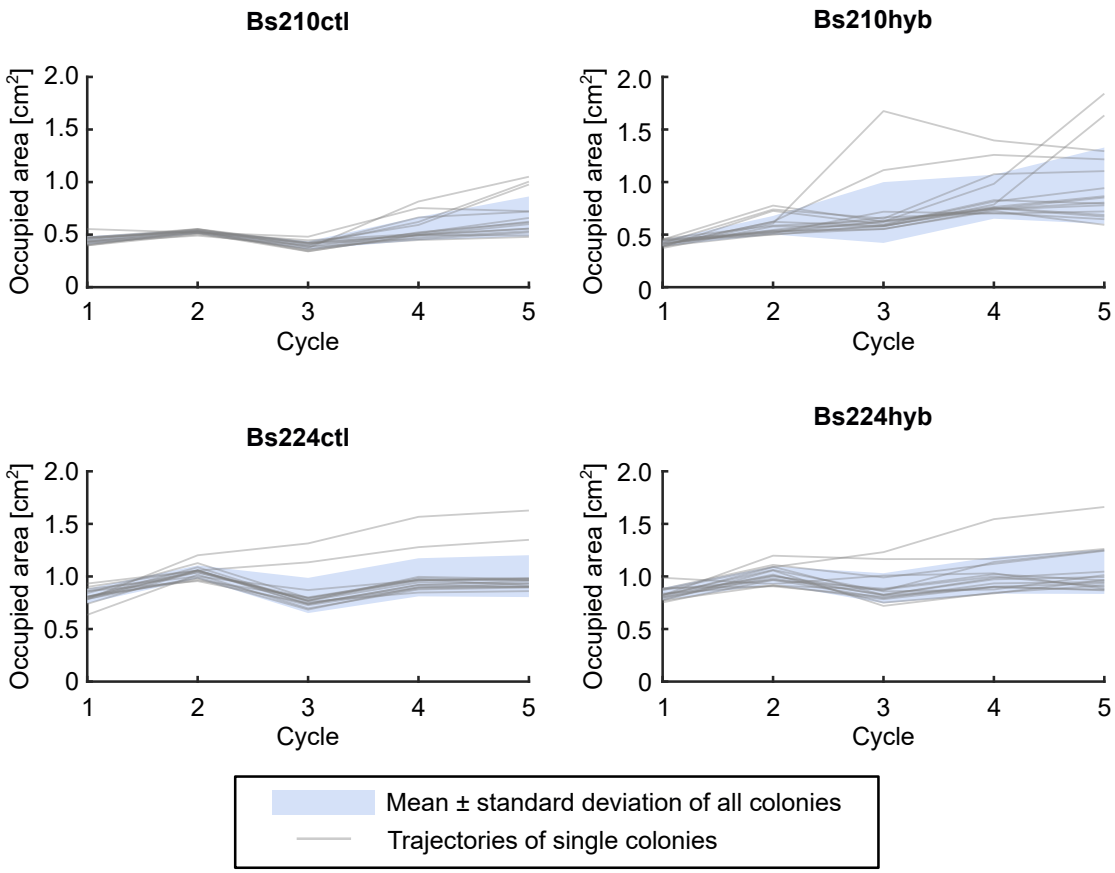


Figure C.21: Occupied area over time. In blue, the average colony size is shown. The trajectory of each single colony is depicted in gray.

D. Examples for the used image segmentation

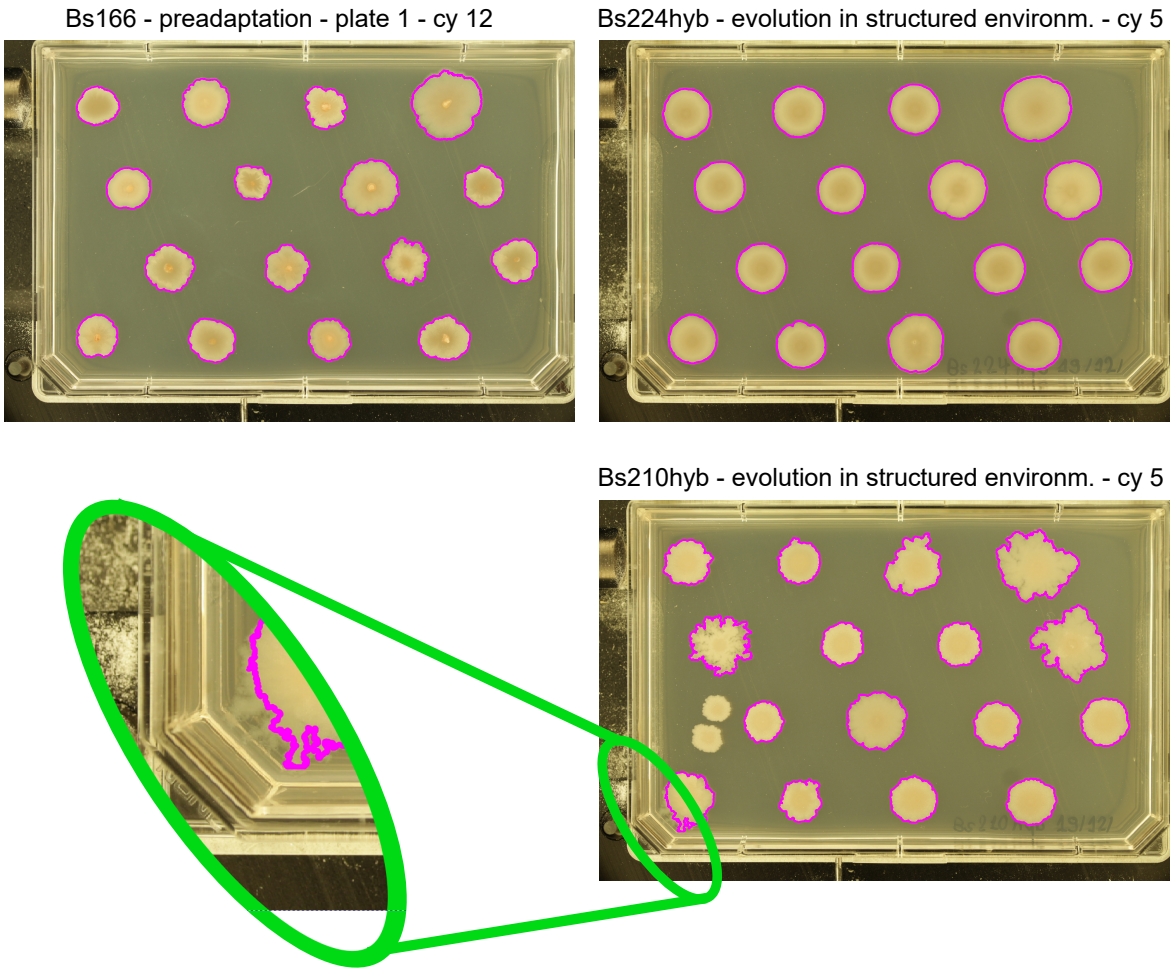


Figure D.22: Image segmentation with Matlab. We apply an adaptive threshold to find the contours of the colonies. In most cases, the contour lines are a good fit, but there are some exceptions (green).

E. Filament correction of flow cytometry data

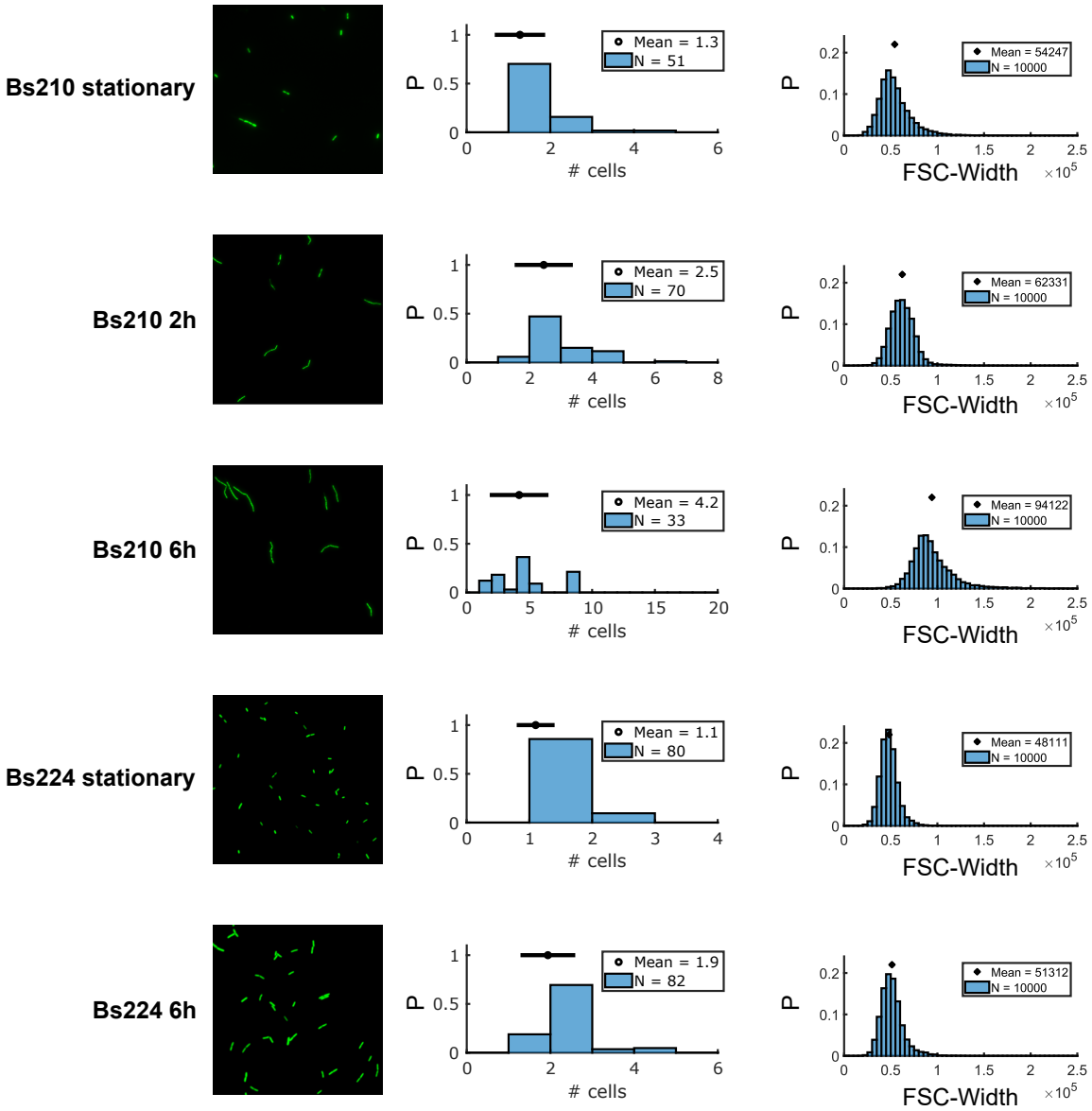


Figure E.23: Filament correction of flow cytometry data. For different strains (Bs210, Bs224) in different growth phases (stationary phase, early exponential growth phase (2h) and late exponential growth phase (6h)), the average number of cells per replicate is counted using microscopy images of cells stained with Syto9. Cells from the same cultures are measured with the flow cytometer to determine the average FSC-Width (FSC-W).

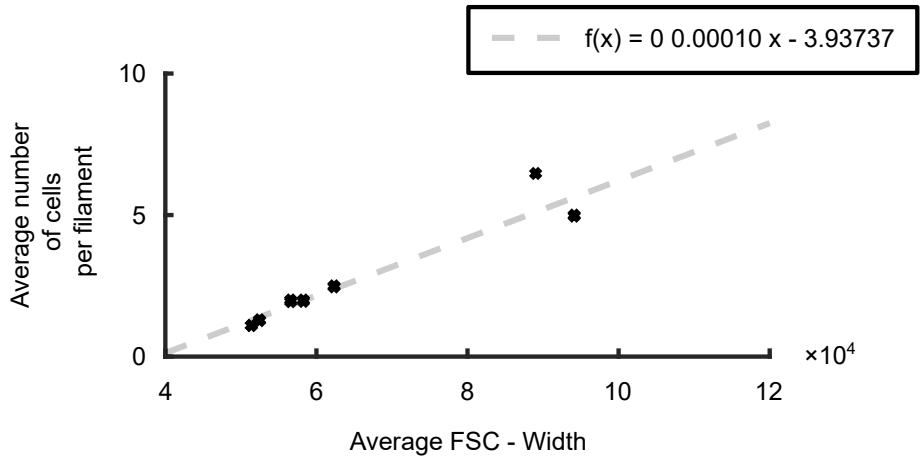


Figure E.24: Filament correction of flow cytometry data - Fit. The relation between the average cell number and the average FSC-W can be described by a linear function.

List of Tables

List of Figures

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen

- Förster, M., Rathmann, I., Yüksel, M., Power, J. J., & Maier, B. (2023). Genome-wide transformation reveals extensive exchange across closely related Bacillus species. *Nucleic Acids Research*, 51(22), 12352-12366.
- Rathmann, I., Förster, M., Yüksel, M., Horst, L., Petrungaro, G., Bollenbach, T., & Maier, B. (2023). Distribution of fitness effects of cross-species transformation reveals potential for fast adaptive evolution. *The ISME Journal*, 17(1), 130-139.

Weitere Publikationen

- Power, J. J., Pinheiro, F., Pompei, S., Kovacova, V., Yüksel, M., Rathmann, I., Förster, M., Lässig, M. & Maier, B. (2021). Adaptive evolution of hybrid bacteria by horizontal gene transfer. *Proceedings of the national Academy of Sciences*, 118(10), e2007873118.

Köln, den