מכון ויצמן למדע

WEIZMANN INSTITUTE OF SCIENCE

Thesis for the degree

Master of Science

Submitted to the Scientific Council of the

Weizmann Institute of Science

עבודת גמר (תזה) לתואר

מוסמך למדעים

מוגשת למועצה המדעית של

מכון ויצמן למדע

By

Shai Slomka de Oliveira

מאת

שי סלומקה דה אוליורה

חקר ההשפעות של רקומבינציה על האדפטציה של בצילוס סבטיליס בעזרת ניסויי אבולוציה במעבדה

Exploring the role of recombination in the adaptation of *B. subtilis* through lab evolution experiments

Advisor:

Prof. Yitzhak Pilpel

מנחה:

פרופ' יצחק פלפל

March, 2019

אדר ה'תשע"ט

Table of Contents

1. Abstract

Horizontal transfer of DNA within prokaryotes is thought to confer an improved ability to adapt to the constantly changing environment. Detected mainly by Bioinformatics approaches, HGT events are common in the evolutionary trajectories of bacteria, and they can occur between a wide variety of species. Though retrospective analysis of such events is a powerful tool, it bears some limitations, mainly since it lacks the temporal resolution which is needed in order to understand more about the outcomes of such integration events and the way in which they shape the adaptation of the species. A good way to study such aspects of horizontal DNA transfer, is to conduct in -lab controlled evolution experiments.

In this thesis, I present two lab evolution experiments of *B. Subtilis*, both designed to explore the role of HGT via natural competence in bacterial adaptation. In the first evolution experiment – the 'Foreign DNA' evolution, cells were evolved to a high salt condition, and were either provided with DNA from various phylogenetic sources, or were allowed to evolve without such external contribution, relying only on their own mutations. In this evolution, we were able to detect and measure relative abundances of both types of events; point mutations and horizontal DNA transfer from donor bacillus strains, to the acceptor genome of the *B. subtilis* 168. While certain point mutations were wide spread and were likely the drivers of the adaptation, HGT events also conferred a noticeable adaptation advantage, and mostly, introduced a striking amount of variation to the evolving population.

In a second evolution experiment, termed 'Self DNA' evolution, I examined the notion that even close-to-clonal population of a single strain may utilize the sharing of DNA with slight sequence variations to evolve. I compared the evolution of both competent and non-competent strains of *B. subtilis* that were evolved to a new medium. This evolution resulted in an accelerated adaptation (in terms of higher growth yields reached) of the competent populations, in comparison to the non-competent evolution lines. The experiment indicated the role of recombination within populations as it suggested that DNA that is released to the medium, and taken up by other, nearly-genetically identical cells, may boost evolution.

The two experiments together, uncover some of the dynamics of DNA uptake in evolution. They gave a sense to their extant and frequency of occurrence, and demonstrate their advantages, both when it occurs between species or among cells of the same species.

2. ***Introduction***

The driving force of adaptation is generation of genetic variation. While genetic variation can be generated from within cells, by means of mutations, indels, chromosomal duplications and rearrangements, other mechanisms have evolved in order to utilize existing variations between cells. In the eukaryotic domain, it is sexual recombination that is the predominant mechanism for inter-cellular genomic mixing. This mechanism has been studied for decades, and was shown both theoretically[1],[2] and experimentally[3] to accelerate evolution by breaking existing linkage disequilibrium in the genome, bringing beneficial mutations together on the same genome and preventing fixation of deleterious mutations. In the prokaryotic domain, in which organisms reproduce by asexual means only, many other systems have evolved in order to utilize inter-cellular variation.

The mechanisms of DNA transfer between different prokaryotic organisms open up a whole spectrum of possibilities. Modern science is completely reliant on these mechanisms (such as plasmids and the CRISPR system) and at the same time is constantly threatened by their outcomes (such as antibiotic resistance acquisition and immune system evasion). This showcases the great potency of those systems and their vital role in adaptation. Such a powerful force of adaptation is likely to generate many interesting phenomena in the evolution of prokaryotes; hence the field of its study is very varied both in scope and in methods.

Our choice was to focus on the natural competence mechanism and its effect on the adaptation of the organism possessing it. Natural competence is possessed by a wide variety of Gram positive and Gram negative bacteria[4]. Our model organism for natural competence is *B. subtilis,* a Gram positive, non-pathogenic, soil bacterium. Genetic competence in *B. subtilis* refers to a physiological state in which uptake of single strand exogenous DNA is allowed. This DNA, after converted within the cell into a double strand, can then be integrated into the genome in a form of homologous recombination. This physiological state is subject to three types of regulation – cell type specific, growth-stage specific and nutritional regulation. Competence depends on the growth media, and is usually occurring post-exponentially and in a minority of the cells (5-15%)[5]. The genetic regulatory network of competence is composed of almost 40 proteins. Roughly

half of them have a role in development of competence, with a key master regulator – comK. Another half of the genes, called late competence genes, are expressed mainly in competent cells, and encode for DNA uptake and transformation machinery[6].

Among the late competence genes, is the comGA gene. Part of the comG operon, is a traffic ATPase which localizes to the cell poles, and is involved in the uptake of DNA[7]. It was also suggested in an early work that comGA is responsible for growth inhibition[8] in competent cells, but a later work attributes this growth inhibition to a downstream protein, Maf[9].

While the natural competence of *B. subtilis* has been studied extensively, it was studied mainly by bio-chemical assays and characterizations, and its roles in the evolution of *B. subtilis* have remained un-explored. Other works, both theoretical[10], and experimental – on pathogenic species of bacteria[11],[12] explored the evolutionary potential of recombination in prokaryotes.

Our aim was to emulate in the lab, under highly controlled settings, evolution of *B. subtilis* and explore the manners in which its natural competence influences its adaptation. For that, two main evolution experiments were conducted. The first, 'Foreign DNA' evolution experiment was started prior to this thesis and is presented in the introduction, though further analysis of its result is conducted in this thesis.

This previous serial dilution evolution experiment done in our lab, attempted to compare evolution which is governed mainly by mutations, to evolution which is also allowed recombination with foreign DNA. In these two setups we aimed to explore the extent to which adaptation of *B. subtilis* to a novel environment occurs through self-mutations vs. acquisition of foreign DNA from the medium. The experimental setup consisted of four different evolution lines, all starting from a common ancestor – *B. subtilis* (strain 168), grown in a medium containing high salt concentration. The lines differed in the sources of the foreign DNA that was introduced to them, and its phylogenetic distance to the recipient strain, except for one line that was given no foreign DNA in the medium (no DNA line). All DNA donors, from which DNA was purified and supplemented to the medium of the evolving cells, shared in common their previous short – or long term adaptation to high osmotic concentrations. The D-Bacilli line was given

foreign DNA from four different bacteria strains, all part of the Bacilli class, that were pre-evolved in the same evolution conditions. A third line – D-Bacteria, was given foreign DNA from non-bacilli Halophilic bacteria, and a fourth line – D-Archea was given foreign DNA from extreme halophile Archea. The lines, each containing 3 repeats, were serially diluted for more than 500 generations. The evolved populations were subjected to growth assays and whole genome sequencing of last time point of the evolution. These methods allowed for quantification of the extent of improvement of each line in the medium, by extracting yield and growth rate parameters from their growth curves, and for analysis of the genomic drivers of this adaptation.

Results showed that all evolved lines significantly improved their growth throughout their evolution (Figure 2-1). Slight advantage in growth was observed for the D-bacilli line of the evolution in comparison to the other lines. Interestingly, this was the only line in which fragments of horizontally transferred donor DNA were observed in the population. These events of integration of donor DNA into the *B. subtilis* 168 cells were detected in the last time point of the evolution in the 3 repeats of the line in various frequencies, together with mutations, that were observed in all lines in high frequencies.

A further, more comprehensive analysis of those events in the D-Bacilli line, in additional time points of the evolution, is presented in the results section of this thesis, together with follow-up experiments that were performed on a specific donor fragment detected in the evolution.

In addition, a second evolution experiment was conducted and is presented in this thesis, aimed to compare the evolution of a competent strain of *B. subtilis*, to this of a non-competent mutant strain. In this evolution experiments I aim to assess whether recombination within cells of the same population can lead to faster adaptation, and if so, find the possible means by which this acceleration is achieved.

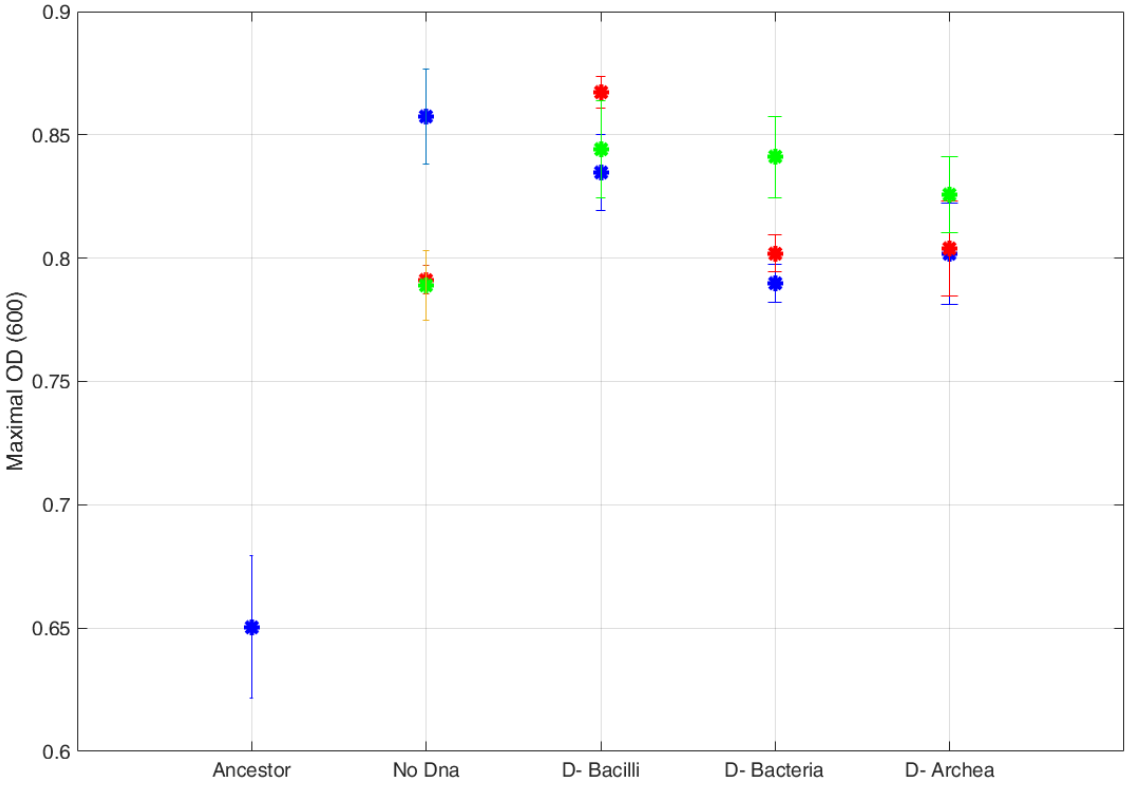


Figure 2-1. Maximal OD extracted from the growth curves of the evolved populations of the 'foreign DNA' evolutions as well as the ancestor. For each line, means and stds of each repeat are presented in different colors.

Overall, in my two main evolution experiments and complementary simulation, I aim to explore and characterize two main modes of action of recombination. One, which will be assessed through the first evolution experiment, is the ability of recombination to generate variation. I will compare the extant and contribution of variation introduced by recombination with foreign DNA in comparison to mutations. The second aspect that I will assess through the second evolution experiment and the simulation, is the ability of recombination to refine and accelerate the propagation of generated variation in the population during adaptation. These two complementary scopes, alone, and combined, will hopefully contribute to the understanding of the conditions and the means by which natural competence and recombination can affect the evolution of the species.

3. ***Materials and Methods***

3.1 Strains and Media

Table 3-1. list of strains and plasmids

Species/Plasmid	Strain	Genotype	Received from	Reference	comments
Bacillus Subtilis	168-comp	168 His met srfA-lacZ [tet] amyE:: xylR Pxyl-comK [ery] (comQ comX ComP replaced by B.mojavenesis RO-H-1 homologs)	Dr. Avigdor Eldar, TAU		
Bacillus Subtilis	RS-D-2	WT	Dr. Avigdor Eldar, TAU		
Bacillus Subtilis	168	TrpC2 LacA:: erm[ery]	WT Strain received from Bacillus Genetic Stock Center (BGSC)		Strain was constructed using the WT strain and PAX01 plasmid
Bacillus Subtilis	168 ΔcomGA	TrpC2 ΔcomGA::erm	Bacillus Genetic Stock Center (BGSC)	[13]	
Bacillus Subtilis	168	RFP (Phleomycin), YFP (Chloramphenicol)	Dr. Avigdor Eldar, TAU		Used for competence assay
Pax01 plasmid	--	Integrative plasmid to lacA locus in *B.subtilis* containing Erm resistance gene	Prof. Rotem Sorerk, Weizmann		
PDG1731 Plasmid	--	Integrative plasmid to threonine operon locus in *B.subtilis* containing Spec resistance gene	Dr. Avigdor Eldar, TAU		

'Self DNA' Evolutions and growth experiments were done on competence media (10.7 g/L K2HPO4, 5.2 g/L KH2PO4, 20 g/L glucose, 0.88 g/L trisodium citrate dihydrate, 0.022 g/L ferric ammonium citrate, 2.5 g/L potassium aspartate, 10 mM MgSO4, 150 nM MnCl2, 40 mg/l L-tryptophan, 0.05% yeast extract). 'Boot camp' growth and overnight growth for DNA purification was done on LB medium (tryptone 10g/L, yeast extract 5g/L, Nacl 5g/L). 'Foreign DNA' evolution and 'Fit Seq' competition experiments were

done on LB medium with either 0.17M (Normal Salt) Nacl (only in Fit-Seq) or 0.8M Nacl (High Salt). For transformations, MC medium (10.7 g/L K2HPO4, 5.2 g/L KH2PO4, 20 g/L glucose, 0.88 g/L trisodium citrate dihydrate, 0.022 g/L ferric ammonium citrate, 1 g/L casamino acids, 2.2 g/L potassium glutamate monohydrate, 20 mM MgSO4, 300 nM MnCl2, 20 mg/L L-tryptophan) was used.

For UV irradiation, minimal saltsX5 medium (K2HPO4 14g/L, KH2PO4 6g/L, Ammonium sulfate 2g/L, Sodium citrate 1g/L, MGSO4*7H20 0.2g/L), and GO starvation medium (Glucose 20% 50ml/L, minimal saltsX5 200ml/L) were used.

3.2 Bacterial DNA purifications

All bacterial DNA was purified using Wizard Genomic DNA Purification Kit (Promega) according to the manufacture protocol.

3.3 Transformation of Bacillus Subtilis cells

All Transformations in the thesis were performed as following: cells were streaked on a fresh LB agar plate and grown over night in 30°c. A single colony was picked to a tube containing 900µl DDW, 100µl MC medium and 10µl $MgSO_4$ 1M. Cells were vortexed and incubated for 2-4 hours in 30°c until the liquid became turbid. Each tube was divided into 5 reactions of 200µl volume. DNA was added to each reaction followed by a 3 hours' incubation in 37°c with aeration. In cases where the DNA for transformation contained no selective marker, PDG1731 plasmid containing Spec resistance gene was added as well, in a ratio of 1ng plasmid: 10 ng DNA for transformation. 100-200 µl of the reactions were plated on selective LB agar plates using glass beads, and incubated overnight in 30°c.

3.4 Serial Dilution evolutions and Competitions

All lab evolutions and competition assays were performed in a manner of serial dilution. The assays were done in a 24 wells plate, each containing up to 5 repeats and one negative control containing no bacteria to control for contamination. The plates were incubated with 900 rpm shaking in 30°c. Each 24 hours, each repeat was diluted (1:120) into 1.2 ml of fresh medium. Every 3 or 6 dilutions, cells were stored in glycerol stocks (30%) in -80°c.

3.5 Competence Assay

Cells were grown overnight in 3 ml of LB medium in 30°c. Cells were than diluted (1:120) into fresh medium containing ~ 2ng/µl of genomic DNA containing Chloramphenicol resistance gene in a final volume of 1.2 ml. For negative control, cells were also diluted to fresh medium containing no DNA. Cells were than incubated overnight in 24 wells plate in 30°c. The following day, 200 µl of the cultures were plated on LB plates containing Chloramphenicol (35 ng/µl). Plates were incubated overnight in 30°c and resistant colonies were counted.

3.6 Growth Assay

Cells from glycerol storage in -80°c were inoculated into 3ml of fresh medium and grown overnight in 30°c. The following day, the OD of the grown strains were measured. The strain with the highest OD was diluted 1:50 and the rest of the strains were diluted respectively to reach similar final OD after dilution. The experiments were done in 96 wells plates each containing between 2-6 strains with technical repeats arranged in a checkboard manner, including repeats of medium without bacteria, in a volume of 150 µl per well. Growth of the strains during the assay and OD measurements every 30 minutes were done automatically using a robotic system of Thermo and plate reader of Tecan (Spark).

3.7 Foreign DNA evolution analysis and detection of donor fragments

The analysis of the Foreign DNA evolution sequencing was conducted by Gil Hornung from the INCPM bioinformatics unit. The analysis consisted of integration of two methods. In the first, reads from all samples were aligned to a merged genome of the recipient and donor strains using bwa mem[14]. Alignments with the highest scores were used, and reads that aligned with the same alignment score to the *B. subtilis* 168 ref. genome were discarded. Reads with zero mapping quality were processed in parallel, to account for cases where an HGT segment could arise from more than one donor. Reads that were mapped to the donor genomes (coverage>2) were used to define putative HGT regions. Putative regions from all samples of the same lineage were merged together. In parallel, all reads were aligned to the recipient genome using bwa mem[14], and variant calling was performed using VarScan2[15]. A list of positions of variation was generated

for each sample and compared with the variants contained by the putative HGT regions from the previous method. The combined lists were then used to better define borders of HGT fragments, and the frequency of each of the fragments was assigned by averaging the frequency of its respective positions of variation from the variant calling. Variants that could not be assigned to HGT regions were regarded as mutations.

3.8 Evolution validation Sequencing

A NGS library was prepared and sequenced as follows: 9 HGT regions that were detected in D-Bac3 based on WGS were selected. Amplicons of length 170-200 bp were designed inside those regions in a manner which allows for amplification of both the acceptor and donor DNA at that region. Amplicons were designed such that they will contain at least 6 defined positions of variability between the acceptor and donor fragment. Primers with Ilumina adaptors were designed for each amplicon. 10 time points in the evolution – from day 18 to day 72 of the evolution were chosen for sequencing, as well as the ancestor of the line. Each time point was sampled 3 times (except for the ancestor which was sampled once) from the glycerol stock, and inoculated into fresh LB media for overnight growth and DNA purification. After purification of DNA from all samples, each sample was served as template for PCR amplification of the 9 different amplicons, in 3 different PCR technical repeats to control for PCR biases. After the PCR amplification, all PCR technical repeats were pooled together, cleaned with SPRI beads (AMPure XP), and their concentration was measured with Qubit. For indexing, all amplicons of the same sample were pooled together at equal concentration, and served as template for PCR amplification for Ilumina indexes addition. The products of the 2nd PCR were cleaned with spri beads, and their concentration was measured with Qubit. Then, each sample was diluted to 4ng/µl and all samples were pooled together. The pool was sequenced with mini-seq, Ilumina. The sequencing reads were then aligned with bwa[14] to a merged genome containing the WT *B.subtilis 168* sequence of all 9 amplicons as separate chromosomes. Bam files were generated, and SAMTOOLS[16] were used to create a pileup file containing all variants. Then, the frequencies of all donor variants at the known variable positions of each amplicon were counted.

3.9 Fit Seq strains transformation:

168 comp cells, which were used as ancestor for the 'foreign DNA' evolution were transformed with 400ng RS-D-2 donor strain DNA fragment (5,621 kb), together with PDG1731 plasmid. One transformation reaction (20 colonies) was used to check the integration efficiency of the desired fragment using colony PCR and sanger Sequencing.

For the competition, a bulk transformation of 41 reactions was performed. Each plate yielded around 100 colonies. Colonies from all plates were scraped and pooled together using approximately 2 ml of LB per plate, reaching a final volume of 100 ml. The pool was then centrifuged and resuspended in 6 ml of LB. 10 µl of the pool was used for each of the 9 repeats in the Fit-Seq competition and the rest was stored in glycerol stocks (30%) in -80°c.

3.10 Fit Seq competition sequencing

DNA from samples of day 0, 6 and 10 of the Fit seq competition was purified. The genomic region of interest (~5kb) for sequencing was amplified, cleaned and sent to the INCPM for library preparation and sequencing. The 5kb DNA fragments were sheared mechanically to smaller fragments of around 500bp and sequencing of 300bp paired end reads was performed using the Next-Seq Ilumina in deep coverage. Results were aligned to the reference WT sequence using Bowtie2[17]. Bam files were generated from the alignment using SamTools[16]. Bam files were analyzed using a custom pipeline for generating a frequency matrix of the known variable positions in the segment, across all samples.

3.11 Epistasis analysis of Fit Seq competition results

The epistasis analysis was done on all couples of positions of variance between the WT and the donor genotype in the fragment, that were observed on the same sequencing read or paired end read couple. A pipeline was generated to count the frequency of each combination for each couple of positions. The 4 possible combinations that were considered were:

WT nucleotide on both first and second position (AA)

WT nucleotide on the first position and donor nucleotide on the second position (AD)

Donor nucleotide on the first position and WT on the second (DA)

Donor nucleotide on both first and second (DD)

After calculation of the frequencies of each of the combination, an estimate for fitness(f) of this combination was made, according to the following formula:

$$(1)\ f(AA/AD/DA/DD) = \log\left(\frac{Freq\ (AA\backslash AD\backslash DA\backslash DD)\ Day\ 10}{Freq\ (AA\backslash AD\backslash DA\backslash DD)\ Day\ 1}\right)$$

Epistasis was defined as the following inequality[18]:

$$(2)\ \Delta W_{AA \rightarrow DA} \neq \Delta W_{AD \rightarrow DD}$$

Where ΔW is the fitness difference between two mutational states.

3.12 Measuring extracellular DNA during growth of B. subtilis with and without DNAse

Since LB medium contains high levels of DNA, cells of *B. subtilis* 168 were grown in 30ml of MOPS defined medium, in 125ml flasks in 30°c with aeration. Different flasks contained different amounts of DNAse I. In addition, 1 flask was served as a negative control, containing only media. At each time point, 1ml of culture was taken, pelleted, and 50 µl were taken from the sup. Then, after filtration through a 0.4µm pore size syringe filter, 20µl were added to Qubit high sensitivity DS-DNA mix and the DNA concentration was read according to the Qubit protocol.

3.13 UV irradiation and survival assay

The protocol for UV irradiation of the strains was modified from the literature[19] and was the following: colonies of WT and Δ*comGA* were inoculated into 5ml of LB medium and were grown overnight in 30°c with aeration until reaching stationary phase. Cells were pelleted and resuspended in 5ml of GO medium, for an hour incubation in 30°c. 5ml of the culture was diluted in minimal salts to a final volume of 50 ml. For each round of irradiation, 5ml of the culture was diluted with DDW to reach 10 ml and was poured on a sterile petri dish. The dish was then irradiated without lid by a UV lamp with a flux of 1J/m^2/sec. The radiation doses for both WT and Δ*comGA* were 5,10,15,20 and 25 J/m^2. 200µl of each irradiated culture, and pre-irradiated cultures were plated in different dilutions for calculating survival. The rest of the irradiated cultures was stored in glycerol stocks. The plates for survival measurement were incubated overnight in 30°c and the

number of colonies on each plate of irradiated cells was counted and divided by the number of colonies of the pre-irradiated culture for estimation of the percent survival. Around 14 single colonies from each radiation treatment (5,10,15,20, and 25 J/m^2) were picked and stored in glycerol stock.

3.14 Growth Assay parameter fitting algorithm – 'Curveball'

The results of the growth experiment were analyzed using a parameter fitting software ("Curveball") created by Yoav Ram from Lilach Hadany's lab at Tel-Aviv University[20]. Curveball fits a mathematical model to growth curves and outputs biological parameters: maximal populatioan density (maximal OD), lag time and maximal growth rate for the fitted curve. The parameter fitting is based on Baranyi-Roberts model[21] which is defined by the set of differential equations:

$$(3)\ \frac{dN}{dt} = r\alpha(t)N(1 - (\frac{N}{K})^V$$

$$(4)\ N(0) = N_0$$

$$(5)\ \alpha(t) = \frac{q_0}{q_0 + e^{-mt}}$$

Where N is the population density, t is time, r is the per capita growth rate, is the adjustment function for the lag phase, K is the maximal population density, is a de-acceleration parameter, is the initial population density, is the initial amount of a needed molecule in the media (such as a nutrient) which is required in the cell, m is the rate by which is accumulated.

The solution for the set of the differential equations is (6-7):

$$(6)\ N(t) = \frac{K}{(1-(1-\left(\frac{K}{N_0}\right)^V e^{-rvA(t)})^{\frac{1}{v}}}$$

$$(7)\ A(t) = \int_0^t \alpha(x)dx = t + \frac{1}{m}\log(\frac{e^{-mt}+q_0}{1+q_0})$$

From the model, we can extract the biological parameters: Maximal OD-yield (parameter K in the model) and maximal growth rate.

3.15 Barcoded evolution competition

A 30 bp barcode was cloned into the PDG1731 plasmid and was transformed into Bacillus subtilis strain 168 WT. this barcoded strain was used as a reference strain for a

series of pairwise competitions with the evolved WT uv and Com uv lines (day 42) and their ancestors for their fitness estimation. Each competition was started by mixing stationary cultures of the reference strain with one of the evolution repeats, in a 1:1 ratio. Then, 3X 10 µl were taken from each mix and were diluted into 3 technical repeats containing 1.2 ml of competence medium each, in a 24 wells plate. The rest of the mix was pelleted and frozen for DNA purification. Competitions were performed in the same manner as the serial dilution evolutions. In total, 22 different pairwise competitions (wt uv A-J, com uv A-J, wt and com uv ancestors), with 3 technical repeats each resulted in 66 competitions. Glycerol stocks were frozen every 3 days and the competition ended after 12 days.

3.16 'Evolved Moran' simulation

The simulation relies on some basic principles of the Moran Process[22], and the measure for evolvability presented in a previous lab evolution experiment on yeast[23]

The simulation assumes a finite population of size N. each individual in the population is assigned a binary genotype of size n = 10. NK model was used, to generate fitness look up tables for each possible genotype (2^n) at different K values. Another two parameters that were set at each run of the simulation were pMut – probability for mutation at each iteration, and pHGT - Probability for HGT event at each iteration. The simulation begins with N individuals with identical genotypes (set to be all zero genotype). All individuals are labeled with a binary label, 'id', (either 0 or 1) independent of their genotypes, which cannot be mutated or pass in an HGT, such that ½ N has a '1' id and ½ N has a '0' id.

At each iteration of the simulation, one individual is chosen to replicate. The probability for each individual to be chosen for replication is given by the following equation:

$$\textbf{(8)}\ P(x) = \frac{i(x)*r(x)}{\sum_{x=1}^{N} i(x)*r(x)}$$

Where p(x) is the probability to choose individual x out on N individuals, i(x) is the frequency of x at the current iteration and r(x) is the fitness of x. once an individual is

chosen to replicate is has pMut probability to receive a mutation in a random position of its genome. It also has pHGT probability to receive 2 adjacent nucleotides from a dead genotype (randomly chosen from all dead genotypes). The replicated individual receives the same label as its parent individual, and if its genotype has changed during the replication, it receives a new fitness based on the look up table. In the same iteration, the population is kept at a constant number by choosing one individual to die at random. The genotype of the dead individual is stored in memory for HGT events.

The simulation ends when 95% of the population has the same id – either ‘1’ or ‘0’.

4. Results

4.1 'Foreign DNA' Evolution Experiment

4.1.1 Time points analysis

As described in the Introduction, a previous evolution experiment was done in our lab in which HGT events were detected from the last time point (following 500 generation of evolution). A further sequencing was conducted, and included 12 time points throughout the evolution of the 3 repeats of the D-Bacilli (from now will be referred to as D-Bac) lines. I continued with analysis of the evolution time course. HGT events and spontaneous mutations were detected, and the frequencies of events and mutations that exceeded a maximal frequency of at least 0.1 were plotted as a function of time (Figure 4-1). The visualization reveals that many genomic events which have occurred cluster to similar trajectories of frequencies. This is true most predominantly for HGT events and to a lesser extent for mutations. Several of such clusters can be observed in the different repeats. Some clusters within each repeat behave in an opposing manner – reflecting a competition between clones in the population in which one clone declines in frequency while the other rises. Some clusters are composed of mutations only, and some are composed of both mutations and HGT events. This might suggest that HGT events rose on background of mutations. In fact, all clusters of HGT events appear to occur consequent to mutations and converge to a similar frequency of at least one mutation. The analysis also shows that the different repeats of the evolution line vary in their number of HGT events- while in D-Bac 2 only one event succeeded to rise in frequency, in the other two repeats, and especially the third, much more events took place. However, our quantitative measure of HGT events is limited by our method of HGT events detection (see Discussion).

The fragments of DNA that were horizontally transferred in the evolution originated from the DNA of three different donors out of the four that were supplemented and contained a striking amount of positions of variation to the acceptor DNA (Table 7-1). Though the percent of identity between the donors and the acceptor DNA was very high (above 95%), the high amount of variation introduced by the fragments can be

explained by looking at their length distribution (Figure 4-2), which reveals that most fragments spanned a great length of several kilo base pairs.

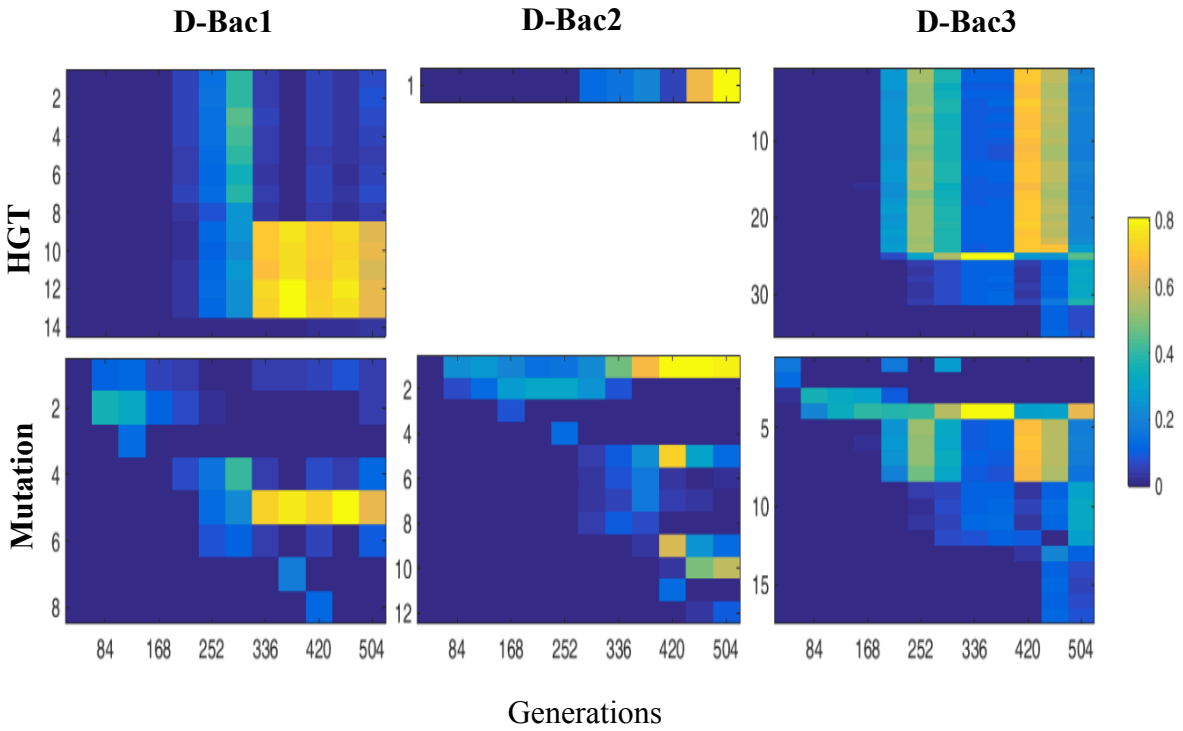


Figure 4-1. Frequency of HGT events (top) and mutations (bottom) as a function of time (generations) in the three repeats of the D-Bac lines (left to right)

The steepness of frequency changes between consequent time points raised the concern that some frequencies are over or under estimated due to sampling biases (see Discussion). This behavior was strongest in the D-Bac 3 repeat (especially between generation 378 and 420). It was decided to perform another sequencing of selected regions. 9 genomic regions in which HGT events were detected from the last sequencing of the D-Bac3 repeat were chosen for sequencing. They were selected such that they will represent all four unique trajectories of frequencies that were observed in this repeat. They were amplified from the ancestor sample and each of the 10 latest time points. To check for sampling biases each time point was sampled in 3 independent repeats. The fragments were sequenced and the frequency of donor variants in each selected fragment was counted in each time point sample, in order to examine whether sampling biases caused the rapid changes in frequencies. The results show (Figure 4-3) that for most time

points, the frequencies of the different samples were consistent between time points and repeated the dynamics observed in the previous sequencing, though one time point – from day 60 (generation 420) showed large variation in frequencies between its 3 repeats.

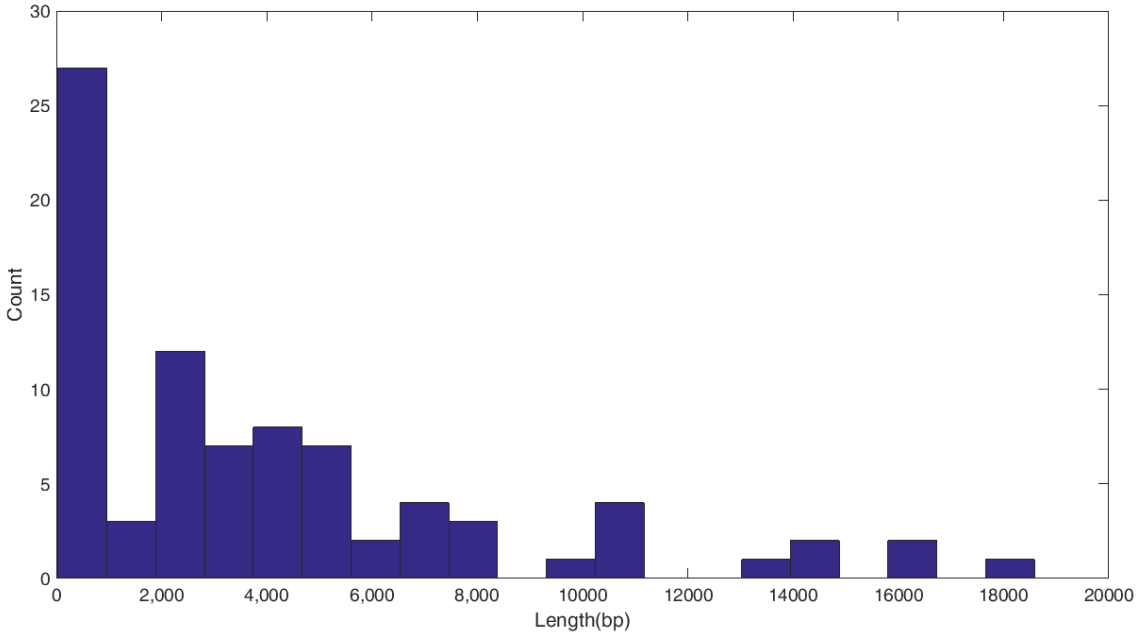


Figure 4-2. Length distribution of HGT events in all D-bac lines

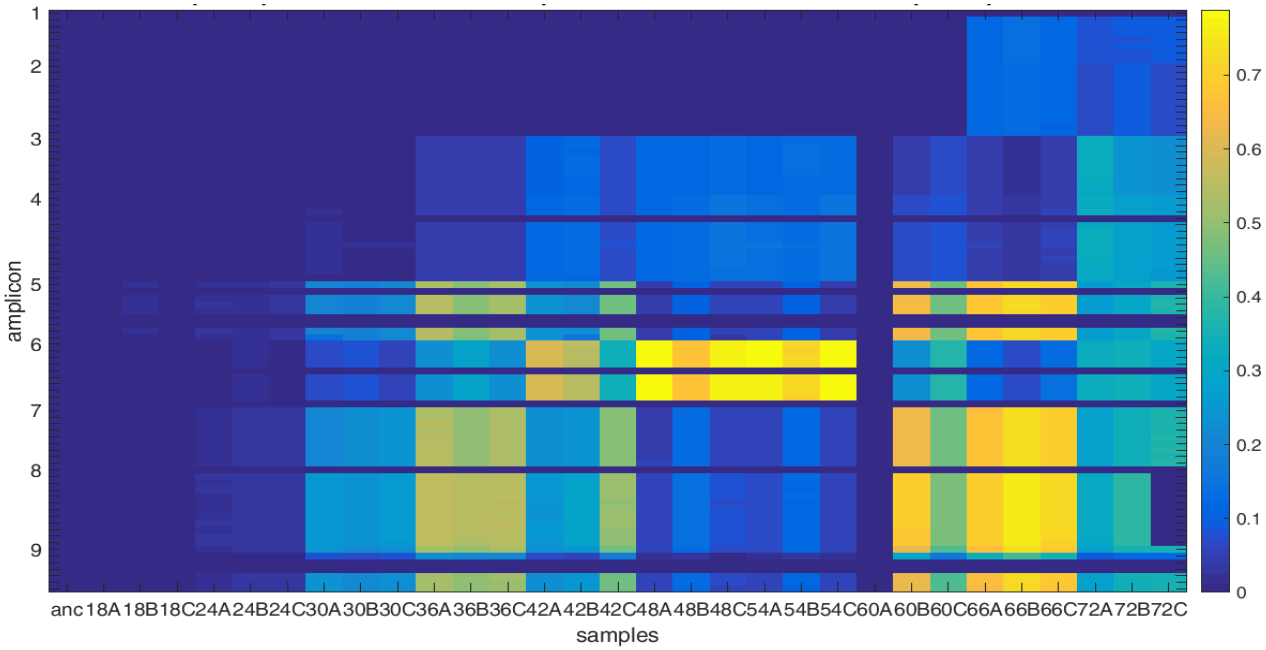


Figure 4-3. Frequencies of each variable position in the 9 selected HGT regions, in different time points and each in 3 independent sampling repeats.

4.1.2 Fit Seq competition

Since the D-Bac evolution repeats contained both mutations and HGT events, we wanted to disentangle the two types of events and check whether the HGT fragments provide an advantage in growth. It was decided to test the fitness contribution of one donor fragment from the evolution, by introducing it back to the ancestor, and following its frequency during a competition with the untransformed ancestor. We decided to select the only HGT fragment that was observed for D-Bac line 2, and reached a final frequency of 0.76 (Table 7-1). The high frequency the fragment reached in D-Bac2 repeat and the appearance of a longer version of that fragment in the D-Bac 3 lines suggested to us that it poses a fitness advantage. The 5.4 kb long donor version of that fragment was transformed into our ancestor. It was decided to conduct many transformations of the fragment and pool a large number of colonies (around 400), to one population that is composed of variants containing different parts of that donor fragment. This method allowed us to not only ask whether the fragment as a whole was beneficial but also narrow down the beneficial regions and learn about genetic interactions between positions in the fragment. Since transformation efficiency was not high, the pool contained mostly un-transformed cells of the ancestor, which served as competitors against the donor versions of the fragment. This population was then subjected to a 10 days' competition in similar conditions to the evolution (6 repeats) and also in regular LB medium without high salt stress (3 repeats). The populations from day 6 and 10 of the competition were sampled and the 5.4 kb region from all samples was sequenced. The frequency of each position of variation in the fragment was counted. Plots of those frequencies at day 10 across the repeats as well as the ancestral population (Figure 4-4) show that all positions of donor variants were represented at day 0, though not uniformly. Their initial frequency varied between 0.1-0.2, these initial frequencies represent the transformation efficiencies of the various regions of the fragment. Results also show that across all repeats, donor variants rose in frequency by day 10. This increase in frequency was not homogenous across the fragment since some sub-regions showed a greater incline than others. Interestingly, most of those positions that show the highest increase are ones that changed the coding sequence of the acceptor in a non-synonymous manner. To our surprise, the frequency increase of all positions was greater in the repeats of low

salt condition. Another difference between the conditions is the difference in variance between the different repeats- while in low salt, repeats behave similarly, in the high salt, there is more variation between repeats.

Overall, we can conclude that the fragment did pose a fitness advantage as it rose in frequency throughout the competition. This advantage was most likely driven by the certain positions in the fragment which manifested the highest frequency increase. Their contribution was most predominant in the regular salt repeats, though still apparent also in the high salt condition.

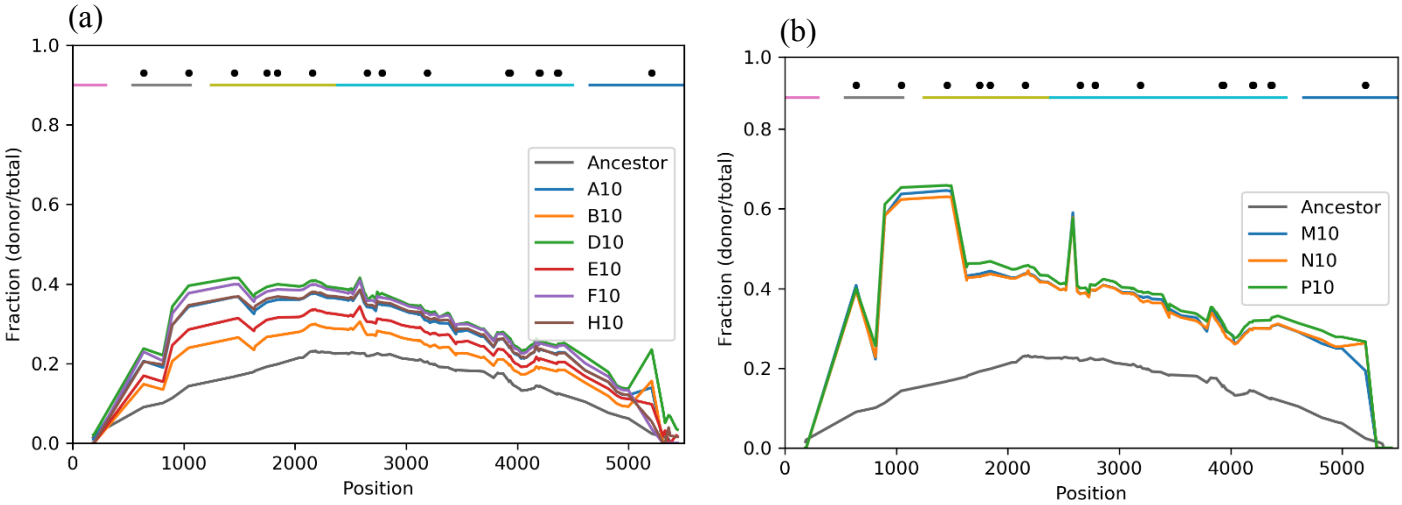


Figure 4-4. Frequencies of the donor variants of the transformed fragment in the fit-seq competition, before (Ancestor) and after (A-P 10)10 days' competition. The colored top lines correspond to the gene- coding sequences present in the fragment, and the black dots represent positions in which the donor variant is non-synonymous to the acceptor variant. (a). frequencies in the repeats of the high salt competition (A-H). (b). frequencies in the repeats of the low salt competition (M-P).

4.1.3 Epistasis estimation on pairs of donor variants in the fit-Seq fragment

In an attempt to uncover more information from the Fit-Seq competition results, a further analysis on pairs of positions of variance in the sequenced fragment was made. Our aim was to estimate how much of the contribution of the individual donor positions was influenced by their neighboring positions of variation. For each pair of positions we could detect together on the same read pair (indicating they appeared in the same individual),

we calculated the fitness of all 4 possible combinations of acceptor and donor identity and plotted the fitness difference between the donor and acceptor identity of one position under either donor or acceptor background of the second position. Since the different repeats show a similar trend, Figure 4-5 presents this analysis on two repeats out of the 9, one repeat of high salt and one of low salt. The equality line plotted represents independency between the two positions (see equation 2 in materials and methods). The results show that while most of the position pairs cluster in one cloud on the equality line, 2 pairs that are far from the cloud are observed in all repeats, on two opposing sides. This implies that these pairs harbor a certain level of epistatic interaction. However, linkage between positions limits our possibility to deduce conclusively about such pairwise interactions (see Discussion).

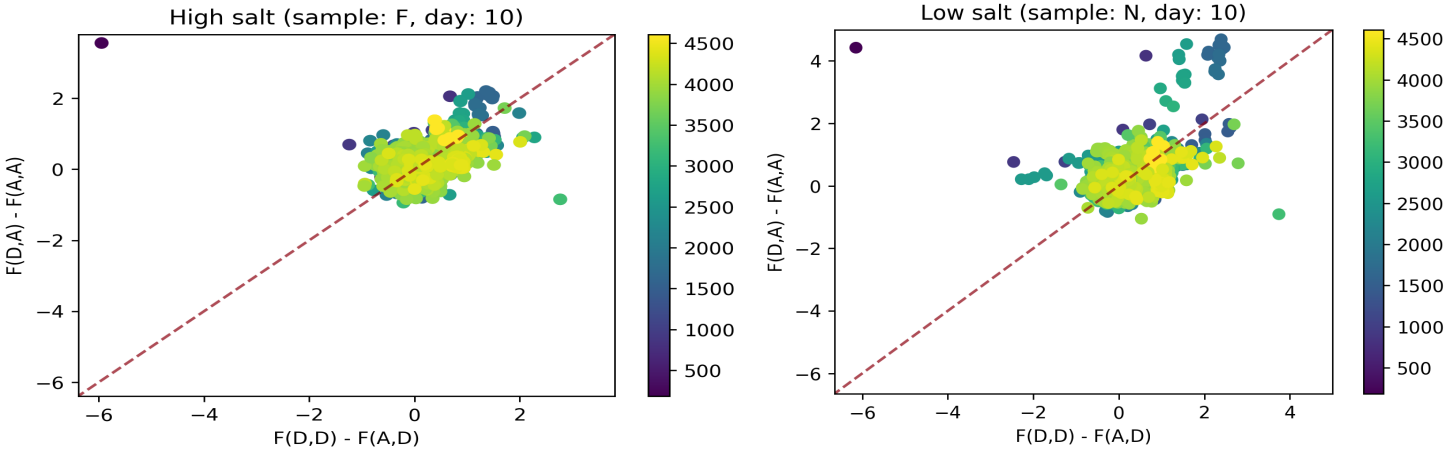


Figure 4-5. Epistasis analysis on pairs of variants in the fragment of two repeats of the fit-seq competition. (a). Epistasis analysis on repeat F – high salt competition, at day 10. (b). Epistasis analysis on repeat N – low salt competition, day 10.

Another observation is the difference in the pattern between the high and low salt repeats. In the low salt, more points seem to be diverging left-wards from the equality line, suggesting that in the different conditions, different genetic interactions between positions emerge.

The presented results of the analysis suggest that some genetic interactions are present between positions of the fragment. The two most predominant pairs which do not cluster together with the rest, represent two opposing interactions; one (the left pair)

suggests that the contribution of one donor positions is compromised in the presence of the second donor position. The other pair suggests the opposite interaction- having both donor positions on the same background confers a higher advantage to the first position than if it was introduced alone.

4.2 'Self DNA' Evolution

In the first experiment, we explored the effect of integration of foreign DNA on the evolution of *B. subtilis 168*. However, even without any addition of DNA from other species, natural competence allows for integration of DNA which is of the same population – self DNA. We hypothesized that the ability to recombine with DNA of the population, has implications on the adaptation of the species, as it breaks the clonality of the population, and allows for different clones to mix with each other. In order to study and understand the effects of recombination within the population, we decided to conduct a second evolution experiment, in which two lines will be evolved – one line which is able to utilize the self-external DNA (eDNA) and integrate it, and another line which will be compromised in this ability. We wanted this difference between the two populations to not affect their initial fitness, so that we could compare their rate and extant of adaptation from an equal starting point, and attribute any difference in the outcomes of their evolution to their inherent difference in the ability to utilize self DNA. In order to create this difference between the lines, we first turned to environmental factors which could eliminate either the DNA in the medium or the ability of the cells to utilize it.

4.2.1 eDNA measurement and DNAse/S. S treatment

I first sought to measure if and how much eDNA accumulates during growth of *B. subtilis*. I measured the concentration of eDNA accumulated in the medium during the growth of *B. subtilis* with or without DNAse I in the medium at two different concentrations. Results reveal (Figure 4-6) that while eDNA concentrations are low during the first hours of growth of *B. subtilis*, substantial amounts are accumulated already at 24 hours of growth. We hypothesized that addition of DNAse I to the medium of the growing cells will be able to eliminate the eDNA available for integration. However, as can be seen in Figure 5-6 a. DNAse I in the medium is possibly effective in the first hours of growth, though not effective at all after 24 hours, in eliminating eDNA.

In fact, in the last time point, DNAse treated growth medium contained even higher amounts of DNA than non-treated medium. Separately, a growth assay of *B. subtilis* cells with and without DNAse was conducted in order to assess the growth of the cells in the presence of DNAse (Figure 5-6 b.).

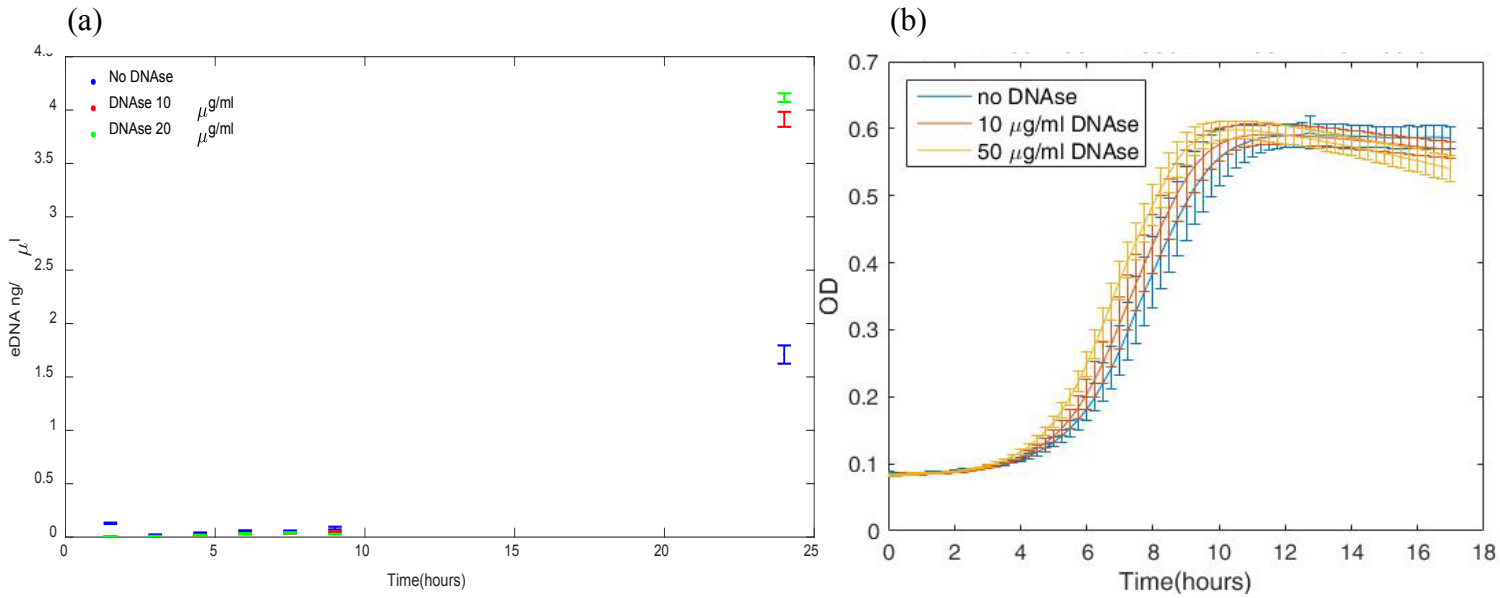


Figure 4-6. Effects DNAse I on growth of B. subtilis 168 and eDNA concentrations during growth in Mops medium. (a). eDNA measurements during growth of 168 with and without DNAse in mops medium. (b). growth assay of 168 with and without DNAse I in mops medium.

The growth assay revealed that cells treated with DNAse, show a slight improved growth rate than without DNAse, followed by a moderate decrease in OD upon reaching stationary phase.

Another attempt to eliminate competence of cells in a non-genetic manner, was done by addition of Salmon Sperm (S.S) to the media of growing cells (Data not shown). We hypothesized that addition of large amounts of DNA which is not available for recombination (due to the phylogenetic distance), will mask the DNA which is available for recombination and render the cells non-competent. While competence assay conducted with and without salmon sperm in different concentrations gave encouraging results, as cells with high amounts of S.S were non-competent, a growth assay with S.S revealed that

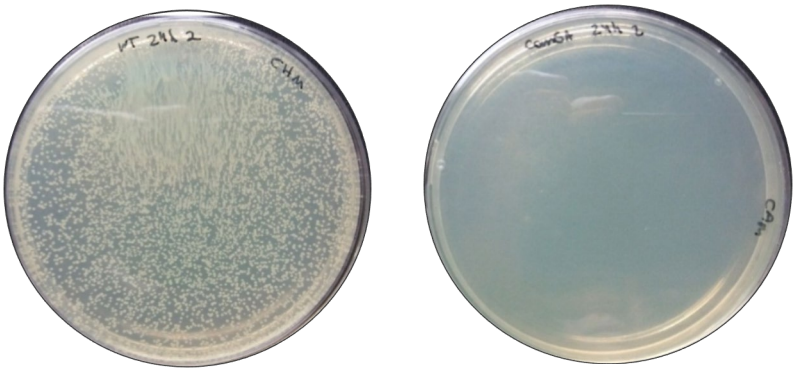
massive cell death occurred during growth with such high amounts of S.S. We, therefore, decided to turn to genetic manipulations of *B. subtilis* in order to eliminate its competence.

4.2.2 Phenotypic comparison of WT and ΔcomGA mutant and evolution calibrations

In attempts to find genetic manipulations which will eliminate the competence of B. subtilis, we searched for gene deletion candidates which are harmed in their competence but are downstream enough to not have a big growth difference in comparison to the WT strain. Δ*comGA B. subtilis* KO strain was ordered from the Bacillus Genetic Stock Center[13]. A growth and competence comparison was made between the strain and its WT *B. subtilis* (Figure 4-7). Competence assay on competence medium resulted in no resistant colonies of ΔcomGA after 24 hours in comparison to a grass of colonies for the WT strain. This confirms that KO of the comGA gene results in lack of competence. Despite the drastic difference in the competence level between the strains, the growth of the two strains was very similar in competence medium, and no differences were observed at any stage of the 24 hours' growth. Figure 4-7 shows results of one of such growth experiments, that was highly reproducible in repeated growth experiments.

In light of the phenotypic comparison of the two strains, it was decided to conduct the evolution experiment with the two strains as the ancestors for separate lines of the evolution experiment.

(a)



(b)

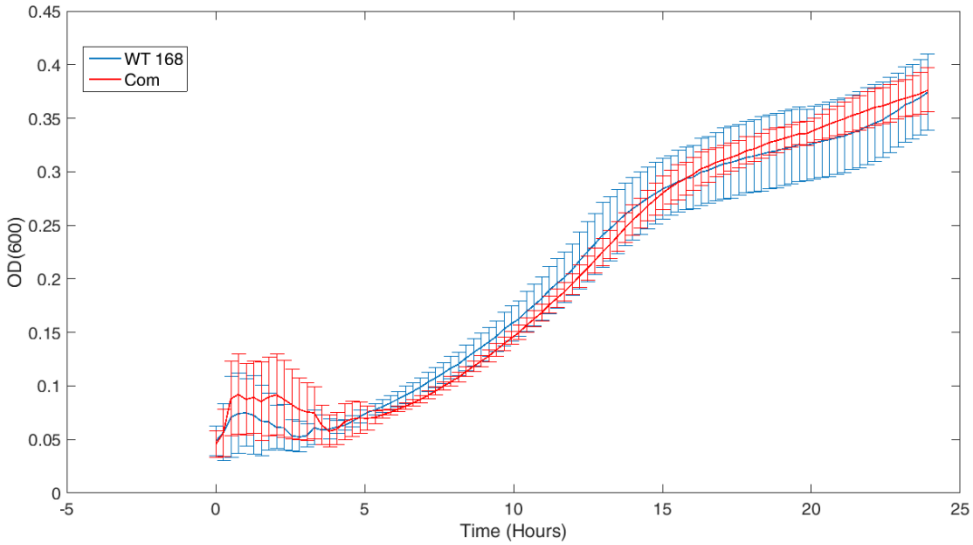


Figure 4-7. Phenotypic comparison between the WT B. subtilis 168 and the ΔcomGA mutant

(a). Results of one repeat of a competence assay of the two strains. WT, on the left, yielded a grass of resistant colonies. ΔcomGA, on the right, yielded no resistant colonies. (b). OD as a function of growth time of the two strains in competence medium with Erythromycin (5ng/µl) in 24 wells plate. No visible difference in growth is observed.

In order for recombination to exert any effect on the evolving genomes, some standing variation must exist in the population. Thus, in an attempt to increase the initial standing variation in the ancestral populations of the evolution, a UV radiation treatment was performed on the two strains. The radiation was done with different UV doses, and survival curves for each dose and strain were plotted (Figure 4-8). Overall, WT and Δ*comGA* survival curves gave similar values, and their survival decreased as the function of the UV dose used. It was decided to continue with the populations which have

undergone 25 J/m^2 UV radiation, and to add them as ancestors of additional evolution lines.

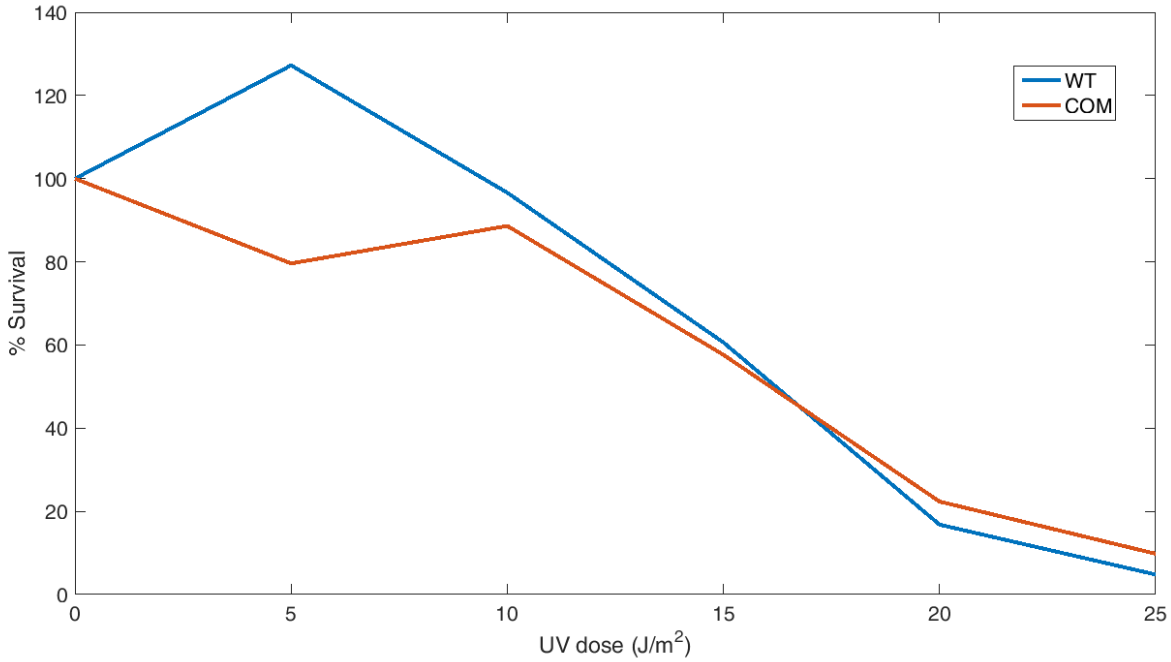


Figure 4-8. Survival curve of the WT and ΔcomGA mutant after UV radiation in different doses

A growth assay for all four ancestral populations was conducted in order to compare growth before and after UV irradiation (Figure 4-9). As can be observed from the results of the growth assay, the non-irradiated populations grow only marginally faster than the irradiated populations, but reach a similar final OD. Furthermore, in agreement with the results obtained from the survival curve, no difference was observed between the WT and comGA populations.

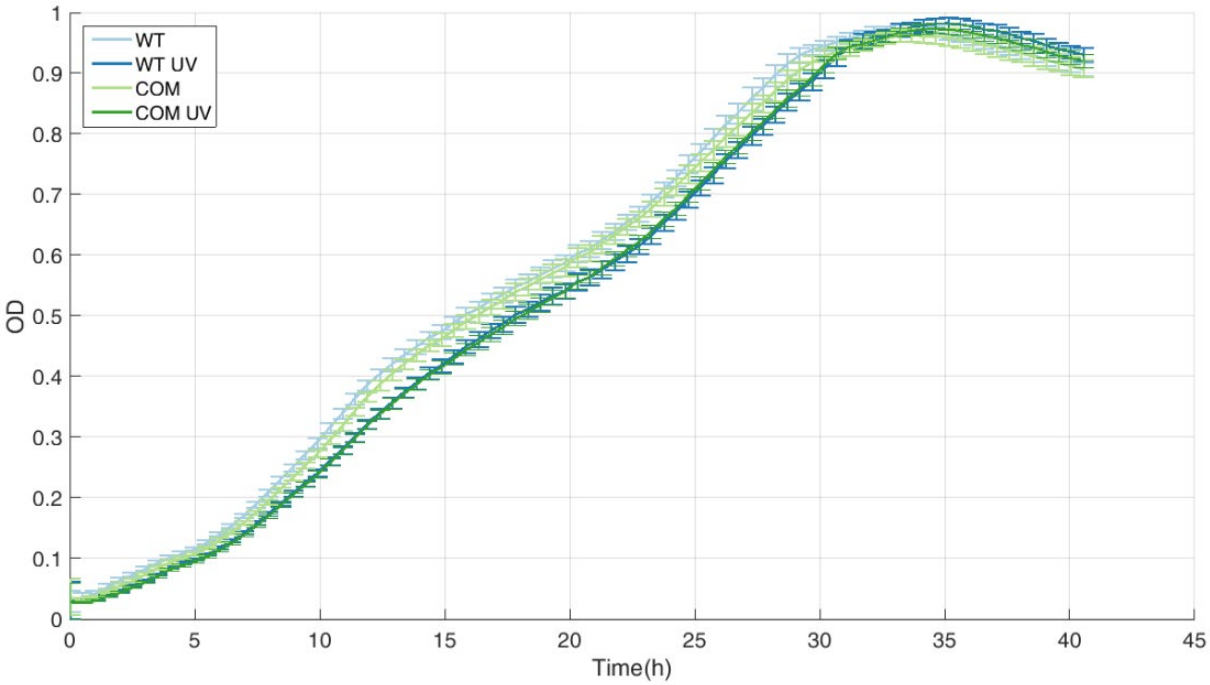


Figure 4-9. OD of the UV irradiated and non-irradiated WT and ΔcomGA strains as a function of time

4.2.3 'Self DNA' Evolution Set-Up

Following the above calibrations, we decided to compare the evolutionary dynamics of both WT strain and the comGA deletion strain that lack the ability to uptake DNA from the environment. We decided to explore both isogeneic populations of both strains as well as the irradiated populations. Each population was evolved in several independent evolutionary repeats, as described in Table 4-1. It was decided to do the evolution experiment in competence medium, in which all growth and competence calibrations were done. The evolution was carried out via daily serial dilutions and continued for 42 days (around 300 generations). Samples for sequencing were frozen every 6 days.

4.2.4 Growth analysis of evolved populations

In order to evaluate their fitness improvement following the evolution, the populations of all repeats of the evolution were subjected to a growth assay following their 300 generations of evolution. The results of the growth assay are shown for each line separately together with its ancestor (Figure 4-10).

Table 4-1. list of the different evolution lines in the 'self DNA' evolution

Evolution Lines	# of repeats	Line Name
WT *B.subtilis 168*	5	WT A-E
WT *B.subtilis 168,* uv irradiated at 25J/m^2	10	WT uv A-J
ΔcomGA *B.subtilis 168*	5	Com A-E
ΔcomGA *B.subtilis 168,* uv irradiated at 25J/m^2	10	Com uv A-J

It can be seen that all evolved lines improved their growth rate significantly in comparison to their ancestors and reach the stationary phase of their growth much faster than their ancestors.

Nonetheless, comparing the OD values reached at stationary phase reveled a difference between the WT and comGA deleted populations. Only in the WT and WT uv lines we see that all repeats have a higher OD than their ancestors after 24 hours of growth. In the com lines, some repeats exhibit a decrease in OD after reaching maximal yield.

In order to better compare the growth of the evolved lines, parameters of yield and maximal growth rate were extracted from each growth curve, using curveball – a growth curve parameters fitting algorithm[20] (figure 5-11). The yield distributions of the different repeats of each line reveal that WT uv line reached higher yields than all other lines. In pairwise rank sum tests, the yields of the WT uv, were found to be significantly higher than those of the com uv line. Without irradiation, the WT line was not significantly better than the Com line in terms of yield. Furthermore, the WT uv line reached significantly higher yields than its parallel non-irradiated line. The same thing cannot be said about the evolution lines of the competence impaired strain – com lines.

The line of Com uv reached the highest maximal growth rates, significantly more than the WT uv and com line.

It can be seen that the variation between repeats is highest for the irradiated lines, both in terms of yield and growth rate. If we compare between the different parameters, then it is apparent that there is more variation in WT and WT uv in the max growth rate parameter than in the yield parameter. The opposite is true for the com and com uv lines.

Taking the results into account, comparing the uv irradiated lines, we can conclude that the WT uv line, despite having a compromised maximal growth rate, has evolved to reach higher maximal yields than its parallel Com uv line. This difference in yield suggests an improved adaptation of the WT uv line. Such advantage in adaptation cannot be attributed to the WT isogenic line, as no significant differences in growth are found between the WT and COM lines.

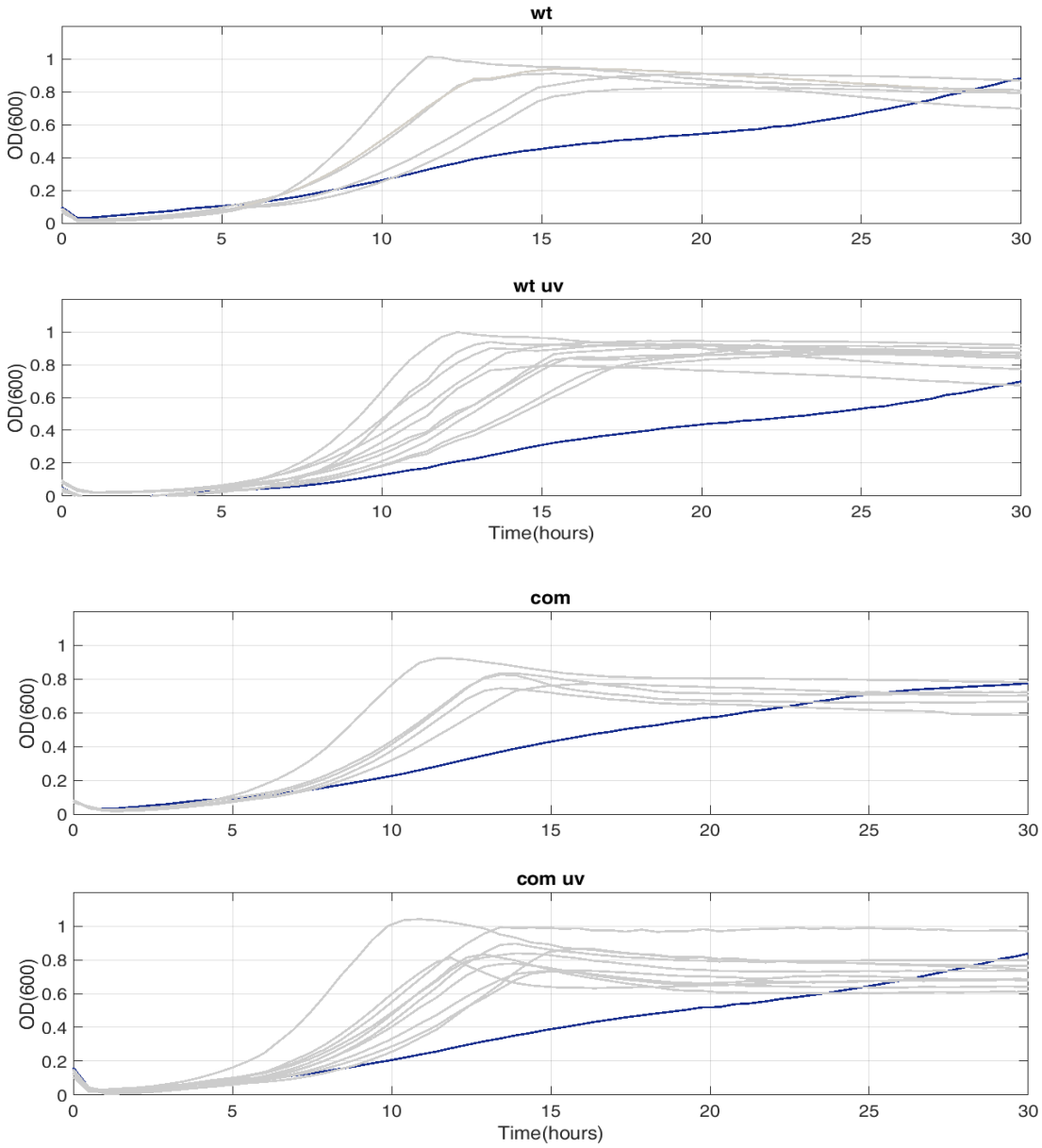


Figure 4-10. Growth curves of evolved repeats from day 42 and their ancestors. All evolution repeats in all lines of day 42(grey) improved their growth rate compared to their Ancestor (blue)

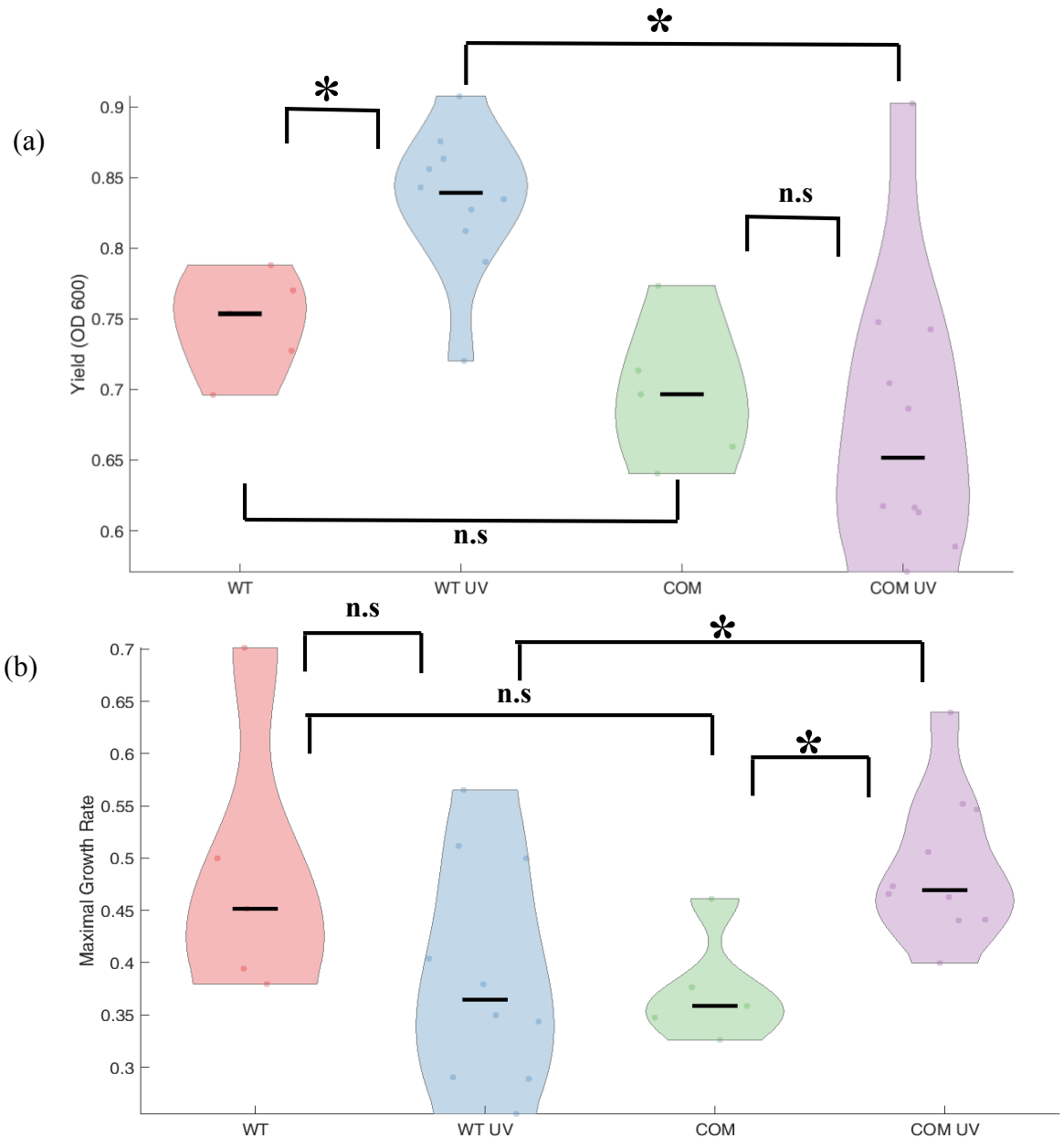


Figure 4-11. Curveball- derived growth parameters distribution plots of the different evolved lines. Statistically significant pairwise comparisons (rank sum tests) are marked with asterisk (*). (a). Yield parameter distribution in the different evolved lines. (b). Maximal Growth rate parameter distribution in the different evolution lines.

4.2.5 *Barcoded competition of the 'self DNA' evolved populations*

In addition to OD analyses of the 'Self DNA' evolution, we sought for another method to assess the fitness improvement of the evolved populations of the self DNA evolution. We have decided to barcode the ancestor WT strain with a 20bp barcode, and compete it in a 1:1 ratio with each of the evolved repeats of both WT uv line and COM uv lines. . We conducted 20 pairwise competitions – each competition with one repeat of the evolved populations against the barcoded strain. Each competition was conducted with 3 repeats. In addition, also the WT uv and COM uv ancestral populations were competed against the barcoded strain to control for slight differences in fitness between the ancestral populations. the competition was continued for 12 days. Results are not shown here as sequencing results of the competition are not yet available. We plan to determine the fitness of each evolution repeat by calculating the ratio between non-barcoded cells to barcoded cells at the first time point in the competition in comparison to a later time point.

4.2.6 *'Boot camp' evolution*

UV radiation introduced variation to the ancestor strains of the evolution. It was suggested by us that recombination might speed up adaptation of the irradiated WT by reduction of this variation, either by purging of deleterious mutations from the genomes of the population, or by combining beneficial mutations from separate genomes to the same ones. In order to better understand which is the main mechanism that led to the faster improvement of the irradiated WT in the self DNA evolution, it was decided to complement the evolution, with a second evolution experiment, in which the irradiated populations (WT and Com uv) would be served again as ancestors. However, unlike their previous evolution, before evolving them in competence media, they were serially diluted for 5 days in LB medium, in which our previous competence assays showed (Data not shown) that the competence of the WT is compromised in comparison to competence medium by roughly 100 folds. This step was added since we hypothesized that 5 days in LB would be enough to reduce many of the deleterious mutations that were generated by the UV radiation in both lines. we assumed that after this period of time in LB, both populations would remain with mostly neutral or beneficial mutations, and if the WT line would still show faster adaptation in the consequent evolution after LB, it would most

likely be due to the shuffling of beneficial mutations. After 5 days in LB, the irradiated populations of WT and com UV were diluted into 9 repeats each, and were evolved for 36 more days, to a total of around 287 generations, similar to the previous evolution. The evolved lines of this evolution will be referred to as WT BC and com BC. Growth assay was performed on the evolved populations and their ancestors (Figure 4-12) .

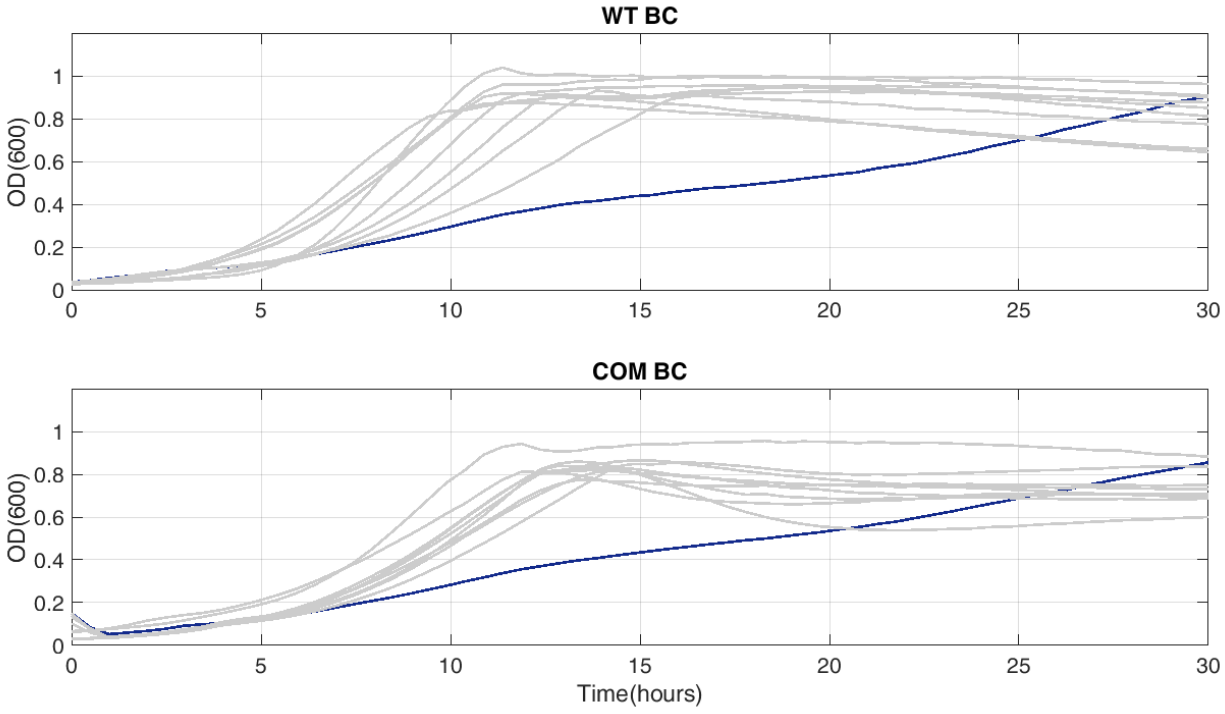


Figure 4-12. Growth curves of the evolved lines of the BC evolution after 42 days. The evolved repeats (gray) are presented for each line, as well as their respective ancestor(blue)

The growth parameters were extracted just as in the previous evolution (Figure 4-13). Similar to the previous evolution, WT BC reached higher yields by the end of the evolution than the Com BC line. However, in this evolution, WT BC has also reached significantly higher maximal growth rates than com BC, and contained more variation between repeats.

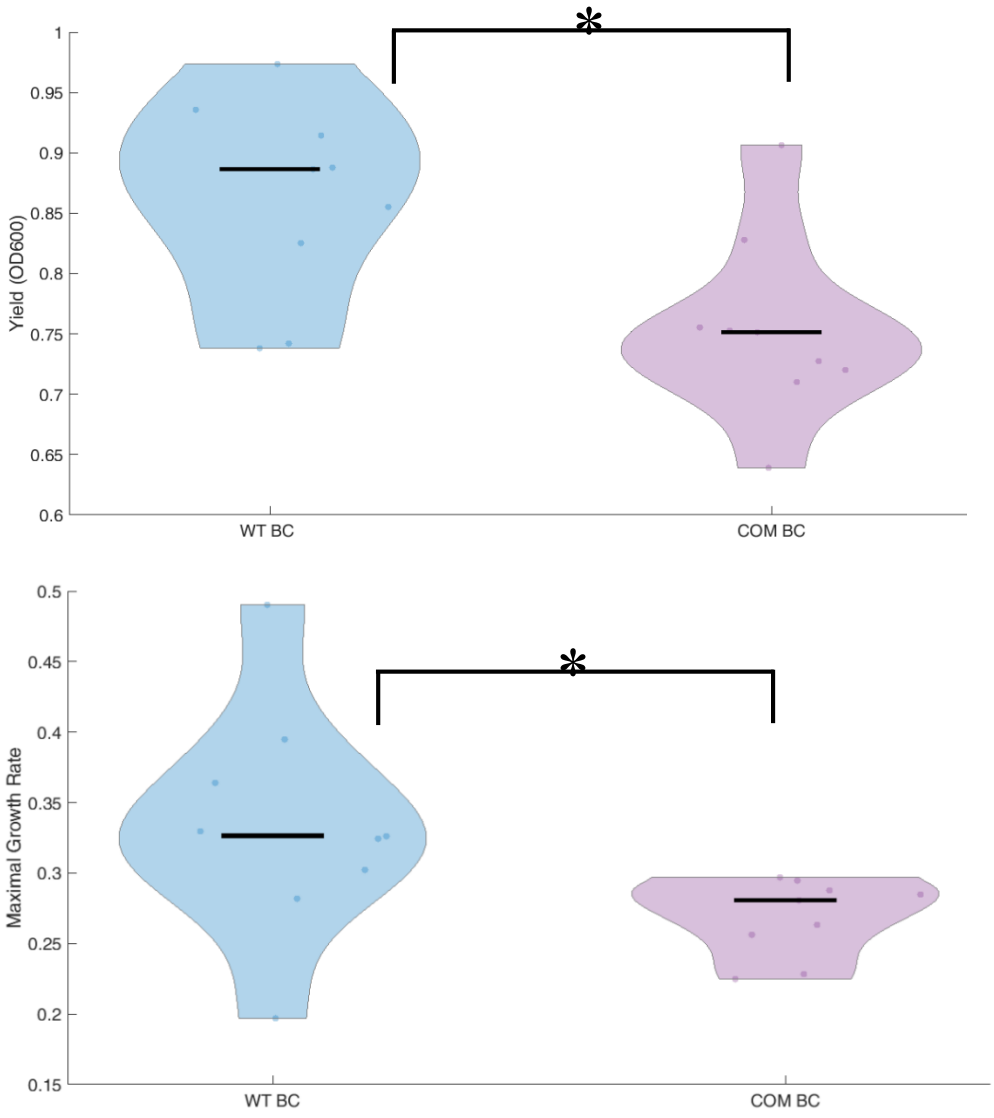


Figure 4-13. Curveball derived parameters of growth distribution in the evolved lines of the BC evolution. WT yields and maximal growth rates are significantly higher than the ΔcomGA mutant.

Overall, in both Self DNA evolutions, the irradiated WT evolved to reach higher yields than the irradiated ΔcomGA lines. In the boot camp evolution, this was also accompanied by increased maximal growth rate in comparison to the ΔcomGA, indicating that in those conditions in particular, the WT evolution was improved in comparison to the non-competent Com line.

4.3 'Evolved Moran' Simulation

In order to explore the theoretical possible effects of HGT on evolution, in comparison to mutations, a stochastic evolution simulation was written. In this simulation, a population of a finite size N, and of genome size n=10 for each individual, starts from a completely homogenous genome composition. Their genotype, and each possible genotype is assigned a fitness value decided by NK model [24], depending on the 'K' parameter set in the beginning of the simulation. A binary parameter (0 or 1), 'id', is also given randomly at the beginning of the simulation to each individual in the population, so that half of N have 'id' of 1 and half have 'id' of 0. The simulation runs such that every iteration an individual is chosen to reproduce in a manner which depends on the product of the fitness and frequency of its genotype. The chosen individual has a certain probability (PMut, and pHGT - set at the beginning of the simulation) to mutate one position or to receive by HGT 2 nucleotide positions from other genotypes. The progeny of that replication is given the same 'id' parameter as the parental individual. At the same iteration, another individual is chosen to die at random, so that the population size is kept constant. The simulation ends when the 'id' parameter of either 0 or 1 has reached over 95% of the population. For each set of values – N, K, pMut and pHGT chosen, 50 simulation runs were made. Different properties of the simulation were collected from each run; the number of iterations until fixation of 'id', the maximal fitness reached by any individual in the final iteration, the average fitness in the population, and the std of fitness values in the last iteration. Distributions of these parameters under different values of pMut, K and pHGT were plotted and analyzed (Figure 4-14). An apparent result is the lack of effect of HGT on any of the parameters. It can be concluded since runs of simulations in which the probability for HGT events were set to zero distribute similarly across the different parameters to runs in which HGT was permitted. The distribution of the number of iterations needed until convergence seems to be the most variable parameter, showcasing the stochastic nature of the simulation. It can be seen that in high mutation rates, the highest maximal fitness possible is found in most of the runs, however, as mutation rates decrease, the distribution becomes larger and in less runs the maximal fitness is found. This is true for all 3 K values. Similarly, the average fitness parameter seems to be decreasing as the mutation rate decreases. Another interesting trend is that the standard

deviations of the fitness of the populations seem to increase with the K value in the high mutation rate, but this increase becomes milder when mutation rate is lowered to 10^{-3}.

Altogether, the simulation in its current configuration suggests no difference between evolution with and without recombination. Time until fixation of the simulations does not seem to be dependent on recombination nor the ruggedness of the landscape (K) or the mutation rate, suggesting it is of a completely stochastic nature. the distribution of fitness values in the population upon fixation of the simulation seems to be dependent mostly on the mutation rate and the ruggedness of the landscape, implying these are the main constrains on the improvement of the population during the simulated evolution, in the current settings of the simulation.

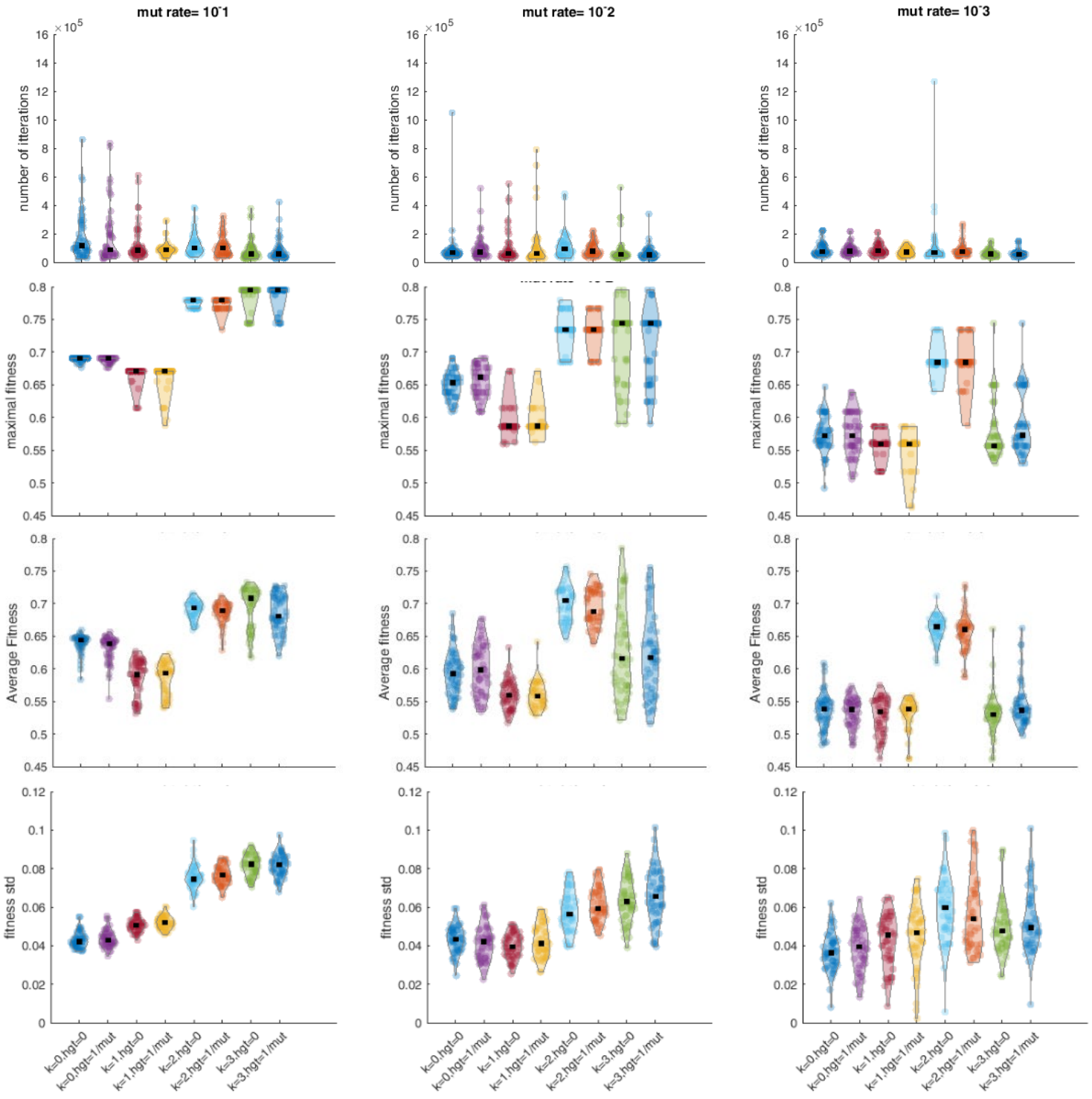


Figure 4-14. Distributions of different properties of the evolved Moran simulation for different mutation rates, K values and HGT rates (50 runs per each combination of parameters), where N=1000. (a). distributions of the run lengths (number of iterations). (b). distributions of the maximal fitness reached at the final iteration of each run. (c). distributions of the average fitness in the population in the final iteration of each run. (d). distributions of the standard deviation (std) of the fitness in the population in the final iteration of each run.

5. ***Discussion***

The natural competence possessed by our model organism, *B. subtilis* enables it to explore other genotypes and introduce variation in an additional, and more extensive manner than mutations. In my thesis, we have decided to explore the potency of this ability, by emulating the evolution of this organism in our lab under conditions which either do or do not permit this introduction of variation.

Two main experiments were conducted, each with a different scope. One aimed to highlight the possible advantages of utilizing DNA from distant sources – different strains ('foreign DNA evolution'), while the other sought to explore the advantages of utilizing almost self DNA, from within the same strain ('self DNA evolution').

It is important to note that beyond the difference in the extant of variation between the acceptor (the evolved population) and donor DNA, another difference between the evolutions is the nature of this variation. In the first experiment, the large distance between the genomes reflected to some extant (but not entirely) their different adaptation potential, as we chose the donor genomes to be pre-adapted to the evolution condition, a condition which was novel to the acceptor genome. In the second evolution experiment, the only variation introduced by us by the UV radiation, was not reflective of any adaptation history. This enabled us to separate those two properties of DNA uptake – introduction of random variation, and introduction of 'selected' variation, and to ask how each of them separately and jointly influenced the adaptation of *B. subtilis* to a novel environment.

5.1 'Foreign DNA' evolution

In the 'foreign DNA' evolution experiment we were able to confirm that *B. subtilis* indeed utilize their competence in order to generate variation. Using Next Generation Sequencing (NGS) of the various time points in evolution, we were able to observe events of donor DNA integration (Figure 4-1 and Table 7-1). We learned that the fragments that were integrated in those events were able to propagate in the population and reach significant frequencies. We also saw that such fragments do not appear singularly, but occur in bursts, as many of them have similar population -frequency dynamics. This might mean that many HGT fragments originated from a single event in

single clone. It is most predominant in the 3rd repeat of the D-Bac line in which we can observe 4 of such bursts and assume they represent 4 competing clones. It implies an 'all or nothing' nature to such events - once a cell is permitting the integration of fragments into its genome, it is likely to integrate more than one fragment, in different genomic locations (Figure 7-1). This can also explain the high variation in the number of HGT fragments detected between the different repeats.

We should note that the detection algorithm used for detection of horizontally transferred fragments, although very sensitive, has its limitations, since, in some cases, high similarity between donors and acceptor genome did not allow us to determine whether two adjacent fragments were integrated as a whole or separately. Our choice of thresholds in this algorithm could have caused a slight over estimation on the total number of fragments detected.

A puzzling result, was the speed at which, in some cases of the evolution, the frequencies change in a drastic manner. While in most events we see a gradual increase, and decrease in frequency between the time points, for some of the cases, especially in the D-Bac3 line, the frequency changes dramatically and rapidly. It was noticed that during the evolution, as cells evolved, they tended to cluster together and create some clumps. Samples for sequencing were stored in glycerol, and it is possible that if cells weren't mixed well enough, that clumps of cells would still be present, and might cause sampling biases. Another possible cause for sampling bias would be contamination by another species. In order to verify the reproducibility of our results, we conducted another sequencing, this time of selected fragments, and with 3 independent samples from each time points. Figure 4-3 revealed that repeats were significantly similar to each other, confirming the results that were obtained in the first sequencing, with one exception in the time point of day 60 (generation 420). In the previous sequencing sample from that time point we also found many reads mapping to the *E. coli* genome. It is most likely a contamination causes the over and under estimation of fragments in that time point.

Table 5-1. Cumulative amount of SNP – either from mutation or HGT source, detected in the different evolution repeats

Evolution Repeat	cumulative # snp from HGT fragments	cumulative # snp from mutations
D-Bac1	1288	19
D-bac2	106	28
D-bac3	3241	19

Table 5-1 presented in the discussion, summarizes the information from Table 7-1 on the total amount of positions of variation introduced by all fragments which were detected in the different time points of the different repeats. These values are compared to the total amount of such positions which were caused due to mutations (indels are included). The difference in the amount of variation that is introduced between the two mechanisms is striking. This comparison reveals the overwhelming difference between the two mechanisms. Mutations, which also appeared sooner, introduce small variation. This variation can either be neutral, deleterious or beneficial. For the mutations that come first we can say with confidence that they were beneficial, since they manage to propagate without any other background changes. However, in the case of foreign DNA integrations, which appear later, since such a large amount of variation is introduced at once, their classification as either neutral, deleterious or beneficial is not so straight forward and can also be influenced by genetic interactions between the different positions. In fact, it most likely that they contain all three possible effects with potential epistatic interactions between them. Since they appear on background of mutations, it is hard to determine even what was their average contribution. We can assume by looking at the fragments' distribution across the genome (Table 7-1) that regions in which fragments within it were detected in more than one repeat are likely to have an average beneficial effect, since it is less expected to occur by chance. An alternative, not necessarily mutually exclusive, is that certain repeated integration events represent "hot-spots" for recombination. An example for such region, can be seen in the repeats of D-Bac 2 and 3. In D-Bac2, a large region of 5.4kb was able to reach a high frequency by the end of the evolution. A larger fragment from the same region and donor, which contains in it the fragment observed in that repeat, is also seen in the D-Bac3 repeat. In the latter

repeat it was less successful in taking over the population by the end of the evolution, but it was able to reach a high maximal frequency during the evolution time course. This dynamic suggested to us that at least a sub-region in this fragment does confer an advantage in adaptation. We therefore chose to integrate back to the ancestor the region that was observed in the D-Bac2 repeat. As described in the results and methods, we determined its contribution by conducting a competition between cells possessing parts of that donor fragment and cells not possessing it at all. This method enabled us to not only see if this entire region is beneficial on average – since all of it rose in frequency by the end of the competition, but to highlight the sub-regions that were the most beneficial – since they increased in frequency more than the average of the entire fragment. The large difference between the two conditions – the high and low salt, lead us to conclude that the advantage given by the donor fragment is not due to the salt stress rather to more general aspects of bacterial growth, perhaps despite the stress (as the advantage is compromised in the high salt repeats). What supports this notion is the difference in variance between the conditions. While in the low salt, results behave similarly across repeats of the competition, in the high salt there is more variation, suggesting that the fitness advantage in that condition, was not as robust.

By looking on the change in frequency of different mis-matches along the 5kb region, we learnt that some sub-regions were more beneficial than others, especially in low salt, and not by chance most of them caused non-synonymous changes to the coded proteins. We wanted to know whether the benefit gained by the positions was independent of the background they were in, or whether genetic interactions with other positions in the fragment shaped their fitness advantage. To ask that in the cleanest manner, we would have to have each combination of donor positions and all of them separately on different backgrounds. It was not possible for us to achieve such high resolution, but even though our constructs contain more than one donor positions, and some positions are more linked to each other than other (their distribution is not even on the fragments), we thought that analysis on the background effect, would still yield some interesting results that could be true, even if we could not associate them to specific positions. We set to ask this question on pairs of positions, which we were able to detect on the same read or read pair. We estimated the fitness of each possible combination of

those pairs in terms of their identity (donor/acceptor), and ask how did the background of one position affected the fitness increase or decrease of the other position. We based this on the definition of epistasis presented by Poelwijk[18]. This analysis (Figure 4-5) yielded interesting results. While most points cluster together indicating little epistasis, two pairs which are relatively far from the others, were found in all repeats. Those pairs represent different types of genetic interactions. The first pair, represented a negative interaction – the contribution of the first donor position (which was the first position in the fragment which rose to very high frequencies) was much greater on an acceptor background than on donor background. This suggests a disadvantage to the HGT mechanism, as if it was introduced alone, it would have probably been more successful. However, the second pair shows exactly the opposite, the contribution of the first donor position was greater when the background of the second position was also of donor identity. Even though this method has limitations (e.g. it can only detect interactions between pairs of positions that are located within the same Ilumina read length distance) and determining the exact positions that drive those interactions is hard, we do not need to know their identity in order to conclude that such interactions exist and they reflect the trade –off in the introduction of many mutations at once. Another conclusion brought by this analysis is that the fitness landscape of the low salt and the high salt is different – in the low salt, a sub-group of pairs drifts from the main cloud of pairs. It is hard to deduce on the drivers of this interactions, but we can appreciate the different interactions that emerge between positions when we alter the conditions.

5.2 'Self DNA Evolution'

In this evolution experiment, our aim was to determine whether the mechanism of competence is able to accelerate the adaptation of *B. subtilis*. Unlike the first experiment, which focused on introduction of variation, on this experiment, we focus on the possible role of recombination in the spread of variation in the population and the effect it has on the rate of adaptation. We hypothesized that this mechanism will allow the shuffling of the genotypes in the population and will help to distill the beneficial mutations from neutral or deleterious backgrounds. This hypothesis was based on theoretical works[10,25] and experimental work done on and yeast by Desai Lab[3] which showed that recombination can lead to accelerated evolution by lowering the total amount of fixed

mutations and increasing the proportion of beneficial mutations fixed . Although the latter experiment was performed on a eukaryotic species and aimed to explore the benefits of sexual recombination, we argue that although it is not similar mechanistically to natural transformation in prokaryotes, the outcomes of the two mechanisms can be similar.

It was decided then to use a competent deficient mutant, lacking the ComGA gene, which is part of the comG operon. The null mutant of this protein is not competent, however, its growth, which was examined in repeated experiments in comparison to WT is not changed. Therefore, we have decided to use both the mutant and WT as ancestors.

In order to emphasize the differences between the strains, we induced a standing variation in the ancestral populations by UV irradiation. We compared our survival curve and UV doses used to a previous experiment which reported mutation rates as a function of UV dose[19]. In that work, the mutation frequency was calculated by a suppressor mutation test. The frequency given, is the number of revertant colonies out of 10^8 cells. With this mutation frequency measure, we estimated that our uv dose of 25 J/m^2 has generated roughly 10 mutations per genome. We chose that amount of standing variation to be satisfactory and we used those irradiated populations as additional ancestors.

As for now, only results of growth parameters extracted from growth curves fitting are available for comparison of the different lines.

From extraction of the growth parameters of yield and maximal growth rates (Figure 4-11) from growth curves of the evolved populations, we can conclude that although compromised in maximal growth rate in comparison to the Com UV line, the WT uv line reaches the highest yields. In fact, if we look at the raw growth curves (Figure 4-10) only in the WT lines all repeats were able to increase their final OD after 24 hours (which is the evolution's time interval) in comparison to their ancestors. The Com lines, although maybe reaching the yield faster, are very close to the OD of their ancestor by 24 hours, suggesting that WT lines, and especially WT uv line evolved better.

The fact that a higher difference was observed between the UV lines rather than the non uv lines, confirms our hypothesis that only with some standing variation in the

population, natural transformation could manifest its effects. In fact, we see that introduction of standing variation enabled the WT uv line to reach higher yields than its parallel WT line. In the Com uv line we do not see this improved evolution, and although not significant, it appears to have even lower yields mean than the Com line. This suggests, that while variation can be used as an accelerator of adaptation if recombination is possible, when it is not, random variation alone is not able to exert such an effect and might even slow the adaptation.

We can also learn from comparing not only the means, but the variance in those two parameters of growth. We see that the variance in the distribution of the growth parameters is higher for the UV irradiated lines, although their sample size is larger. It is very possible that introduction of variation in the populations of both strains created more possible trajectories for their evolution, and hence we see more variation in their outcomes. However, it is possible that this variation in growth parameters, was present already in the ancestors and is not an outcome of evolution. An experiment not reported in the results, might disprove the second option, as we performed a growth assay to 30 single colonies of the WT uv ancestor and found no significant difference in their growth except for one colony which had higher growth rates. It is possible though, that 30 colonies are not a large enough sample size to capture the variance in growth in the ancestral population.

Another interesting difference in variance, is between the two parameters in each line – WT lines seem to have more variation in the growth rate parameter than in the yield parameter, and the opposite is apparent for the Com lines. It can be an outcome of their evolution; however, it can also be an outcome of the genetic difference between the two strains. A possible hypothesis comes from the fact that competence is linked to compromised growth rate in the competent cells. Since both ΔcomGA mutants and WT cells activate the competence system and pay the cost of its activation (especially in competence medium), perhaps selection was stronger for the com lines to lose any activation of competence, resulting in less variation in growth rates between repeats, but in the WT line, we see much more variation, since the cost might be traded off by the possibility of genetic exchange. This hypothesis, though encouraging, needs to be

confirmed once sequencing analysis will be done on the populations and will reveal whether indeed, mutations in activation of competence occurred in the different repeats. Another important control, would be to knock down the ComGA gene in the evolved repeats of the WT lines and examine whether their variation in growth remains the same. By doing so we could eliminate the concern that inherent differences in the strains other than DNA uptake caused these results.

It is also important to note that it is essential to compare the strains with different methods other than OD measurements, such as a competition which was conducted and described in the results. This is important since OD measurements can be influenced by cell morphology and formation of cell clusters. In fact, the ΔcomGA cells do have a slightly more elongated form. We also observed that throughout the evolution, cells from both lines though marginally more in the comGA lines, tended to lose motility and aggregate in the bottom. Though this aggregation occurs at the last stages of growth, it is possible, that some of the growth results could be affected from this behavior (for example, the decrease in OD upon entrance to stationary of some of the comGA repeats in Figure 4-10).

One of our hypotheses to the success of the WT uv line in comparison to the Com uv line, was that recombination allowed for better purging of deleterious mutations, following a uv radiation. It is possible that recombination allowed for faster purging of deleterious mutants, more than it acted towards recombining beneficial mutations. We decided therefore, to complement the evolution experiment with a second evolution, in which, prior to evolution in competence media, cells were grown for 5 days in LB medium (‘Boot camp’), in which the WT’s competence is compromised. We speculated that 5 days were enough to eliminate all highly deleterious clones, and mostly neutral or beneficial mutations will be present in the populations. If indeed most of the contribution of recombination was purging of deleterious mutations, the difference in the rate of adaptation after the boot camp evolution should be smaller between the lines.

The results of the boot camp evolution showed that there must be an additional effect to competence other than purging of deleterious mutations, as we still observed differences between the WT and Com uv lines. Furthermore, the WT uv line was, in

those conditions, better than Com uv line even in maximal growth rate. The advantage in adaptation due to natural competence was even stronger in these slightly changes conditions, implying that stronger effect of recombination emerges when the populations experience changing conditions.

My hope is that when sequencing results of both the competition and the whole genomes of the evolved populations will be received, we could better explain the mechanisms which led to the accelerated evolution of the WT line. For example, we aim to know whether recombination preserved the variation in the population, or whether it acted towards homogenization and reduction of the mutation load. We could learn about the composition of the mutations in the different lines, such as the ratio of synonymous/non synonymous mutations in each line.

While conducting the experiment, we have encountered two previous works[12,11] that aimed to ask the same question on the natural competence of two pathogenic bacteria. Both works found that natural competence accelerated adaptation in the strain which possessed it. The latter paper found this acceleration to be context dependent – they saw acceleration of adaptation only when populations were evolving under periodic stress rather than benign conditions. The two works were conducted on pathogenic Bactria, which have undergone severe selection toward thriving under different and very stressful conditions. It is encouraging to learn that their findings agree to some extant with our findings, on the competence of the non-pathogenic soil dwelling *B. subtilis* since it demonstrates that despite completely different niches and fitness landscapes, natural competence is found to be a strong force of adaptation and its effects are detectable even in short term adaptation experiments, and even (in our evolution) under stable conditions.

5.3 'Evolved Moran' Simulation

In an attempt to better understand the results from the Self DNA evolution, we turned to a simplistic model of evolution, in order to try and find possible effects of recombination on evolving populations. If those effects are found, we will be able to study their precise outcomes and these outcomes could be used to explain our evolution results. We decided to use a stochastic model of evolution, and we based it on the Moran Process[22]. The Moran process assumes no mutations, as it has two absorbing state in which alleles are

either extinct or take over the population. Since we wanted to allow for mutations and events of DNA transfer, we decided to keep the basic principles of Moran, but set different absorbing states to the simulation. We decided to rely on a previous published experiment[23], which compared the rate of evolution of yeast with different ploidy levels by labeling each strain with a certain ploidy level with two different fluorophores, and evolved each strain with equal ratios of the two fluorophores. The rate of adaptation for each strain was set to be the time it took for one fluorophore to take over another in the population. We decided that we would label our initial finite population in the simulation with two different labels, which are independent of their genotype and pass on in replication. This allowed for introduction of mutation and DNA transfer rates with a clear absorbing state to the simulation. Another integration to the stochastic model was the use of the NK model for determining fitness values. In this way, we could also observe differences in outcomes of the evolutions, as we change the ruggedness of the landscape (K).

We changed different parameters in different runs in order to compare the rate of adaptation, as well as the variation in fitness within the population, the maximal fitness reached by the end of the simulation and also the average fitness in the population. The results of the simulation reveled no difference between the runs in which HGT was allowed and those in which only mutations occurred. This could be due to many reasons; first, it could be that genomes of length 10, which were the case for our simulation are too small to generate an effect for recombination. It is also possible that choosing an absorbing state which requires the almost complete disappearance of one label, demands a long time to reach the absorbing state (notice the high amount of iterations to fixation in Figure 4-14), and perhaps at such a long time, differences between the different runs are eliminated. It is also possible that the population size was not appropriately adjusted to the genome size, as the population size was almost identical to all the possible genotypes. All these together might suggest that perhaps our conditions were not complex enough to exert an advantage to HGT. Nevertheless, we did see some interesting results from the simulation, mentioned in the results, but due to the scope of the thesis, they will not be discussed here.

5.4 Final integration

Combining the two evolutions, we see an interesting effect to recombination.

In the first evolution, we identified events of recombination from other strains which introduced large variation. These integrations did not yield a significant improvement in overall rate of adaptation for the line in which they occurred (though we did see a subtle advantage). This is probably due to the fact that those events were mainly of very long fragments, and such large variation introduced at once, resulted in a mixture of beneficial deleterious and neutral effects (as we saw better in the epistasis analysis). In the second evolution, we do see a significant difference between the lines that differ in their integration abilities. This difference was only observable for the lines with some standing variation in them, suggesting that this variation was the ground to this difference.

From these two results, we can deduce that recombination can exert an acceleration effect, but it depends on the extant of variation between the acceptor and donor genome and the nature of this variation. The large size of the HGT fragments detected in the first evolution and their bursty nature, suggests that once those events occur, they occur at a significant proportion of the genome. If a small amount of variation exists (such in the uv lines), these extensive events might be enough to catch even small variations by random, and shuffle them in the population. However, when variation is large (1-5%), such as in the first evolution experiment, such large integrations introduce a lot of mutations at once, and the effect of the possible beneficial mutations is compromised. It is possible that if we were to evolve the D-Bac repeats for more time without another addition of foreign DNA, recombination would have acted on those populations to lower the amount of mutations and refine the beneficial effects of the integrated fragments. Despite this large variation, it was shown for at least one of such events (in the fit-seq competition) that it was still able to confer an advantage to the cells possessing it. This advantage is conferred possibly only due to the fact that the large variation was not random, but was introduced from a relative species which have undergone similar selection.

The mechanism of competence in *B. Subtilis* works in a homologous recombination manner. It was evolved to prevent integration of too-distant genomes (we saw that for other lines of the first evolution in which we saw no DNA integration). This is traded off

with the lack of (or non-stringent) limitation on the fragments size (as we saw integration of even 20kb long fragments). It is possible that in this manner *B. subtilis* is able to maximize the chances of receiving small variations, probably within-strain or within-class variations, and avoiding integration of too distant fragments, and thus maximizing the potency of the recombination.

6. References

1. Society, T. A. & Press, C. Some Genetic Aspects of Sex Author (s): H . J . Muller Source : The American Naturalist , Vol . 66 , No . 703 (Mar . - Apr ., 1932), pp . 118-138 Published by : The University of Chicago Press for The American Society of Naturalists Stable URL : https. **66**, 118–138 (2019).

2. Joseph Felsenstein. the Evolutionary Advantage of Recombination. *Genetics* **78**, 737–756 (1974).

3. Mcdonald, M. J., Rice, D. P., Desai, M. M. & Biology, E. Sex Speeds Adaptation by Altering the Dynamics of Molecular Evolution. **531**, 233–236 (2016).

4. Stewart, G. The Biology of Natural Transformation. *Annu. Rev. Microbiol.* **40**, 211–235 (2002).

5. Dubnau, D. Genetic competence in Bacillus subtilis. *Microbiol. Rev.* **55**, 395–424 (1991).

6. Grossman, A. D. Genetic Networks Controlling the Initiation of Sporulation and the Development of Genetic Competence in Bacillus subtilis. *Annu. Rev. Genet.* **29**, 477–508 (2002).

7. Chung, Y. S., Breidt, F. & Dubnau, D. Cell surface localization and processing of the ComG proteins, required for DNA binding during transformation of Bacillus subtilis. *Mol. Microbiol.* **29**, 905–913 (1998).

8. Haijema, B. J., Hahn, J., Haynes, J. & Dubnau, D. A ComGA-dependent checkpoint limits growth during the escape from competence. *Mol. Microbiol.* **40**, 52–64 (2001).

9. Briley, K., Prepiak, P., Dias, M. J., Hahn, J. & Dubnau, D. Maf acts downstream of ComGA to arrest cell division in competent cells of B. subtilis. *Mol. Microbiol.* **81**, 23–39 (2011).

10. Fraser, C., Hanage, W. P. & Spratt, B. G. Recombination and the Nature of Bacterial Speciation Recombination and the nature of bacterial speciation. **476**, 476–481 (2007).

11. Engelmoer, D. J. P., Donaldson, I. & Rozen, D. E. Conservative Sex and the Benefits of Transformation in Streptococcus pneumoniae. *PLoS Pathog.* **9**, 1–7 (2013).

12. Baltrus, D. A., Guillemin, K. & Phillips, P. C. Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori. *Evolution (N. Y).* **62**, 39–49 (2008).

13. Typas, A. *et al.* Construction and Analysis of Two Genome-Scale Deletion Libraries for Bacillus subtilis. *Cell Syst.* **4**, 291–305.e7 (2017).

14. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

15. Shen, D. *et al.* VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).

16. Wysoker, A. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

17. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–9 (2012).

18. Poelwijk, F. J., Tănase-Nicola, S., Kiviet, D. J. & Tans, S. J. Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J. Theor. Biol.* **272**, 141–144 (2011).

19. Sung, H. M., Yeamans, G., Ross, C. A. & Yasbin, R. E. Roles of YqjH and YqjW, homologs of the Escherichia coli UmuC/DinB or Y superfamily of DNA polymerases, in stationary-phase mutagenesis and UV-induced mutagenesis of Bacillus subtilis. *J. Bacteriol.* **185**, 2153–2160 (2003).

20. Ram, Y. *et al.* Predicting competition results from growth curves. *BioRxiv* (2015).

21. Baranyi, J. & Roberts, T. A. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.* **23**, 277–294 (1994).

22. Fudenberg, D., Imhof, L., Nowak, M. A. & Taylor, C. Stochastic Evolution as a

Generalized Moran Process Drew Fudenberg*, Lorens Imhof,** Martin A. Nowak***, and Christine Taylor**** First draft: August 29, 2003 This version: September 9, 2004. 1–26 (2004).

23. Sorenson. Polyploidy can drive rapid adaptation in yeast : Nature : Nature Publishing Group. *Nature.Com* **519**, 349–352 (2015).

24. Kauffman, S. A. & Weinberger, E. D. The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).

25. Redfield, R. J. Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? *Genetics* **119**, 213–221 (1988).

7. ***Supplementary Material***

Table 7-1. List of HGT events that were detected in the different line of the HGT evolution

Evolution Line	168 start	168 end	donor	# pos. of variation	length	Maximal Frequency
D-Bac1	1139995	1145535	B31	99	5541	0.471
D-Bac1	1145495	1145535	B32	3	41	0.471
D-Bac1	3257317	3257351	B30	3	35	0.5329
D-Bac1	3257317	3260462	B30	192	3146	0.499
D-Bac1	3257317	3266918	B31	24	9602	0.4703
D-Bac1	3799559	3800400	B30	59	842	0.4128
D-Bac1	3800655	3813876	B30	566	13222	0.45685
D-Bac1	3810795	3812193	B31	17	1399	0.2916
D-Bac1	1749189	1767774	B31	257	18586	0.9007
D-Bac1	1754362	1754563	B31;B32	4	202	0.88175
D-Bac1	1755471	1755765	B31;B32	3	295	0.8892
D-Bac1	1761889	1762045	B31	4	157	0.94285
D-Bac1	1766473	1766731	B32	5	259	0.9529
D-bac2	3812641	3818047	B31	106	5407	0.7589
D-bac3	248684	251111	B32	59	2428	0.6834
D-Bac3	507400	510491	B32	41	3092	0.675
D-bac3	840398	842545	B30	139	2148	0.6577

D-bac3	894848	898769	B30	271	3922	0.6549
D-bac3	1033701	1033851	B30	6	151	0.69315
D-bac3	1435748	1438316	B30	148	2569	0.67375
D-bac3	1438544	1439015	B30	29	472	0.6567
D-bac3	1983698	1990002	B31	104	6305	0.6596
D-bac3	2960442	2967458	B31	120	7017	0.6697
D-bac3	3091816	3094462	B31	107	2647	0.6573
D-bac3	3091933	3092421	B32	28	489	0.6438
D-bac3	3092599	3092747	B31;B32	13	149	0.6577
D-bac3	3093294	3097284	B32	95	3991	0.6713
D-bac3	3093294	3097606	B31	96	4313	0.671
D-bac3	3095414	3111293	B31	242	15880	0.66855
D-bac3	3187379	3187781	B30	29	403	0.6419
D-bac3	3229881	3230526	B30	39	646	0.6708
D-bac3	3234892	3239242	B30	232	4351	0.67565
D-bac3	3304264	3314986	B31	133	10723	0.681
D-bac3	3759232	3760838	B32	18	1607	0.68235
D-bac3	3761295	3763668	B32	17	2374	0.6873
D-bac3	3799801	3807151	B30	339	7351	0.6743
D-bac3	3810795	3821329	B31	202	10535	0.68465
D-bac3	3812268	3815993	B30	52	3726	0.67965

D-bac3	1755993	1764229	B32	88	8237	0.78515
D-bac3	3116204	3119895	B31	85	3692	0.2944
D-bac3	2827458	2832381	B31	64	4924	0.30705
D-bac3	2829701	2829887	B32	4	187	0.30335
D-bac3	2832058	2832208	B32	4	151	0.3081
D-bac3	3081530	3081611	B30	8	82	0.2949
D-bac3	1761099	1766533	B31	57	5435	0.364
D-bac3	1217137	1219524	B31	39	2388	0.1039
D-bac3	1237833	1252435	B31	213	14603	0.1
D-bac3	2946937	2952156	B31	110	5220	0.10575
D-bac3	2953675	2954019	B31	10	345	0.11705

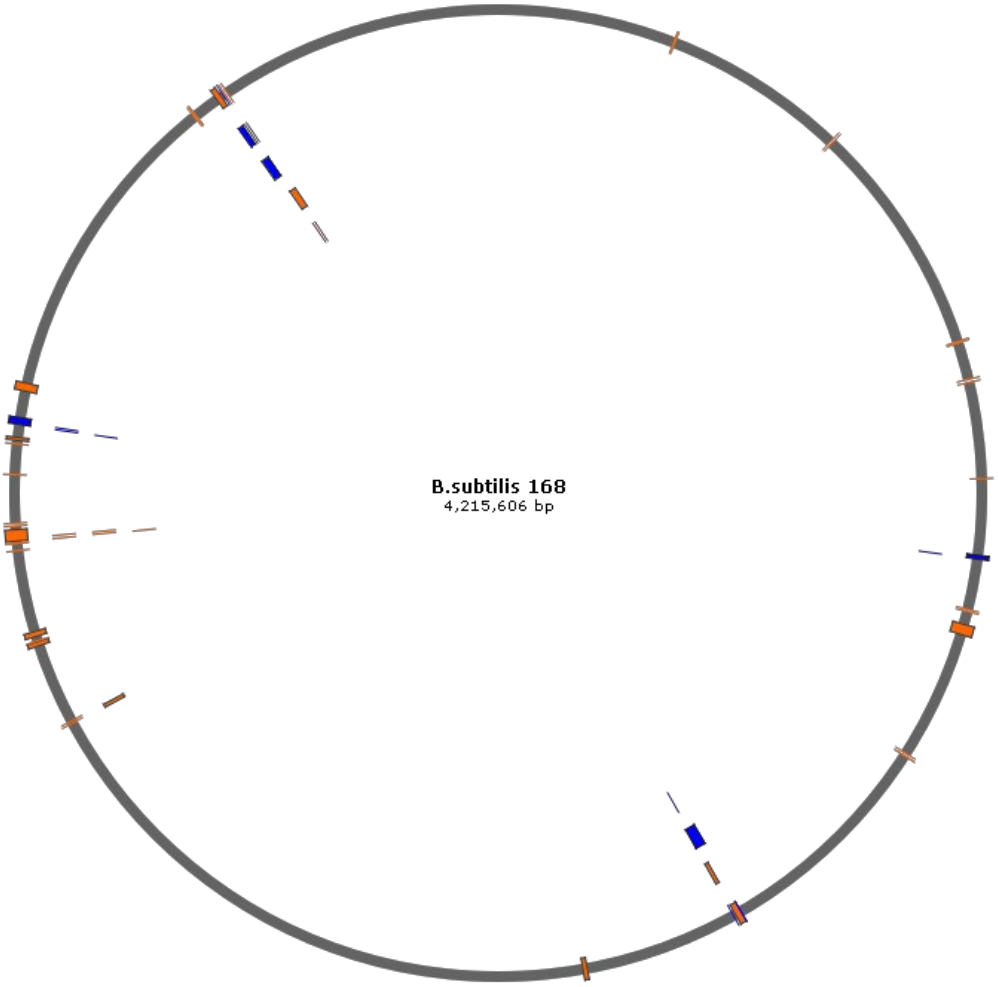


Figure 7-1. The spread of HGT events detected in all lines of the Foreign DNA evolution across the Bacillus genome. D-Bac 1 fragments are presented in blue, D-Bac 2 fragments are presented in red and D-Bac 3 fragments are presented in orange.

8. Acknowledgments

Much like my thesis project, my Msc was an adaptation experience.

I want to thank Tzachi and Orna for providing the right conditions for my growth with their excellent guidance and mentoring.

I want to thank the wonderful lab members for many great thoughts and ideas which I was able to integrate into my work.

Special thanks to Omer Asraf from the lab for helping and conducting some of the analyses in this thesis and for guidance in the coding process.

In addition, I would like to thank the Ilana Kolodkin and Rotem Sorek lab for the help in better understanding B. subtilis, to Gil Hornung from the INCPM for the analysis of the first evolution results, and to Dr. Shai Kovo, which helped me with the UV irradiations.

Lastly, I want to thank the Feinberg Graduate School and the Weizmann institute for posing the selective pressure.