Thesis for the degree

Master of Science

Submitted to the Scientific Council of the

Weizmann Institute of Science

Rehovot, Israel

By

Itamar Françoise

עבודת גמר (תזה) לתואר

מוסמך למדעים

מוגשת למועצה המדעית של

מכון ויצמן למדע

רחובות, ישראל

מאת

איתמר פרנסואז

Experimental evolution of horizontal gene transfer

אבולוציה ניסיונית בהעברה גנטית אופקית

Advisor:

Prof. Itzhak Pilpel

מנחה:

פרופ' יצחק פלפל

March, 2016

אדר א, תשע"ו

1. Abstract

Horizontal gene transfer (HGT) allows organisms to evolve by up-taking foreign DNA from their environment and incorporating it into their genome. By acquiring DNA through HGT, organisms can potentially adapt faster and find innovative solutions to cope with the environmental conditions to which they are exposed. Although evidence for HGT can be found across the all domains of life there is still much to learn more about how it affects evolution. So far most HGT research utilized primarily bioinformatics approaches, which are very useful but bare some limitations. HGT events are identified only in retrospect by examining their remnant in extant species. Bioinformatics of extant genomes might reveal which genes have been transferred into a given genome but it is typically hard to determine who was the species from which the foreign DNA was acquired, when did the transfer event take place, what was the environmental challenge that the event helped solving. In this research we aimed to try and broaden our understanding on the conditions driving successful HGT by conducting HGT evolution experiments within the lab. We evolved on high salt *Bacillus subtilis*, a bacterium with a natural ability to uptake and integrate foreign DNA from the environment. We evolved *Bacillus subtilis* either with or without foreign DNA originating from a span of different halophile bacteria. We then tested how well the bacteria evolved. We found that after ~420 generations, the bacteria showed improved fitness on high salt, as seen by examining growth assays that compared evolved strains to its ancestor. The improvement appeared in all lines of evolution including the line of evolution were no DNA was supplemented. However, we found that lines of evolution, in which *Bacillus subtilis* was exposed to DNA originating from a set of phylogenetically-close salt tolerant bacilli, showed better improvement than the other evolution groups in the experiment. Next generation sequencing of the evolved strains has revealed HGT events in two out of three evolution repeats in a segment of several kbs. This segment includes genes involved in metabolism of lipids that are known to be elevated in the membranes of halotolerant bacteria. We now focus on finding more HGT events that occurred in the evolution and deciphering how did they contribute to the increased salt tolerance.

Contents

1. Abstract 1
2. Introduction 4
3 Materials and Methods 8
3.1 Strains and media 8
3.2 DNA purification 9
3.2.1 Bacterial DNA purification 9
3.2.2 Archeal DNA Purification 9
3.3 Lab evolution in serial dilution 10
3.3.1 Lab evolution at the presence of foreign DNA 10
3.4 Lab evolution in a Chemostat 10
3.5 Growth Experiments 12
3.5.1 Growth curves analysis using parameter fitting algorithm – "Curveball" 13
3.6 Competence Assay 14
3.7 Fluorescent microscopy 14
3.8 Phylogenic calculation 14
3.9 CFU 15
3.10 Next Generation Sequencing 15
3.10.1 *De-Novo* assembly of DNA donors' genomes 15
4 Results 16
4.1 Experimental design 16
4.1.1 Recipient strain, '168-comp', competence evaluation 17
4.1.2 Foreign DNA pool 18
4.1.3 Salt concentration calibration 18
4.2Phylogenetic characterization of the donor Bacilli 20
4.3 Evolution of the donor bacilli to high salt 21
4.4 HGT experiment in serial dilution 24
4.4.1 Growth analysis 25
4.4.2 Colony forming unit assay 27
4.4.3 Cells morphology analysis using microscopy 29
4.4.4 Identification of HGT through next generation sequencing 31

4.5 Evolution in a chemostat 33
5 Discussion 35
6 Acknowledgments 40
7 Bibliography 41

2. Introduction

The ability of organisms to change their genomes and adapt to new environments is a key driving force in evolution. While most adaptations are due to random mutations that occur in already existing DNA, it is quite rare for de-novo genes to arise solely from random mutations[1,2]. One way for an organism to enlarge its landscape of possible adaptations, and acquire new genes into its genome is through Horizontal Gene Transfer (HGT). In HGT organisms acquire DNA from other species and integrate it into their own gene repertoire. Thus, they can potentially both speed up the rate in which they adapt and also reach innovation via acquiring de-novo function genes. HGT is considered to have a big impact on microorganisms[3–7] though it also occurs in eukaryotes[8,9].

There are a few known mechanisms for HGT[10]: transformation, conjugation, transduction, gene transfer agents and nanotubes mediated DNA transfer. Transformation involves the uptake of naked DNA from the extracellular environment via specialized pathways. Conjugation is the transfer of DNA (primarily plasmids) from one organism to another via cell-to-cell contact. Transduction is DNA acquisition that is following phage infection. Gene transfer agents are phage-like DNA-vesicles that are produced by a donor cell, are loaded with random pieces of the donor's chromosome and are released to the environment. Lastly, it was recently showed that bacteria can form nanotube that bridge adjacent cell and facilitate transfer of DNA and proteins[11]

As mentioned above, the ability to acquire new genes in a single transfer can increase the genetic landscape that a given organism can explore. However, HGT may also come with a cost, since uptake of foreign DNA may consume energy and can be growth reducing[12,13]. Furthermore, newly acquired genes might result in toxicity[14]. Extensive bioinformatics studies had shown that there are certain parameters that may determine which genes are horizontally transferred, such as: number of protein partners a gene has, niche similarity between donor and recipient species, evolutionary distance between the donor and recipient species, and nucleotide composition and codon usage similarity between the two species[10,15–19]. Importantly, bioinformatics studies that aimed at deciphering HGT bare some limitations. HGT events are identified only in retrospect by examining the remnant HGT in extant species and thus a lot of information might be missed. Information such as the functionality of the horizontally transferred gene, in which environmental conditions did it transfer and what environmental challenge it helped overcoming.

We thus felt motivated to try and develop a platform by which to answer these questions through performing lab HGT evolution experiments. While most lab evolution experiments revolve around growing some model organism in a completely isolated environment, we were interested in evolving an organism while exposing it to DNA from different origins.

Our model organism to study HGT in the lab is *Bacillus subtilis* (*B. subtilis*) which is a gram positive bacterium found mainly in soil. *B. subtilis* population is heterogeneous and at a given time different cells may behave differently. Under certain stressful conditions a portion of the population (5-10%)[13] enters a state termed the 'competence state'. While competent, cells express a conserved machinery of proteins responsible for the uptake of extracellular DNA followed by integration through homologous recombination[20,21]. It is estimated that 12.8 percent of B. subtilis' genome is of foreign origin[22], emphasizing the impact of the acquisition of foreign genes on this organism's evolution.

In its natural habitat, *B. subtilis* is often exposed to rapid changes in osmolality due to flooding and drying of the soil[23]. Although having some mechanisms to cope with changes in osmolarity[23] it is generally considered not an halophile (adapted to high osmolality environments) organism. We were interested in asking how *B. subtilis* will evolve on high salt if it will be exposed to foreign DNA that was extracted from different halophile/halotolerant organisms.

Halotolerant and halophile organisms are generally classified into three groups according to the level of salt they can endure/they require. Slight, Moderate and Extreme halotolerants can endure up to 1.2M, 3M and 5.2M (saturation) of NaCl respectively[24]. Slight, Moderate and Extreme Halophiles thrive on salt concentrations of 0.3-0.8M, 0.8-3.4M and 3.4-5.2M respectively[25]. In addition, halophile organisms often require a specific salt to be present in the environment in order for them to survive, depending on the mechanism by which they cope with the salt level[26,27]. In general, one can separate the high-osmolality resistance adaptations into two groups. In the first, the resistance is due to the function of a defined low number of genes that lower the effects of high salinity. In the second, broad adaptation across the entire genome lowers the effect of high osmolality on the different components of the cell. We will elaborate on each separately.

Among the specific mechanisms to deal with high osmolality, organisms have found ways to cope both with the ionic strength posed by the environment and with the high osmolality level that threatens to destroy the cell. Halophiles change the ionic composition in the cell membrane so as to lower the ionic strength of their environment[24]. They do so by changing the composition of their membrane- lipids and proteins. Also, halophiles counter high osmolality by accumulating large quantities of small organic compounds termed "osmolytes"[27–29] . These osmolytes act as osmoprotectants as they help balance the osmolality inside the cell while not disturbing protein function. Osmolytes are composed of two groups: inorganic and organic. Inorganic compounds are composed mainly from potassium and chloride and they generally provide relatively narrow adaptation to high salt concentration. Organic compounds are composed of many types of molecules such as amino acids, amino sulfone acids, sugars, inositols and betains. All of these compounds have low molecular weight and high polarity.

Among the broader adaptation towards high osmolality, it was found that halophile bacteria tend to have elevated frequencies of negatively charged residues on protein surfaces that help stabilize protein folding[30,31]. Earlier works also pointed out towards low hydrophobicity as another adaptation to high salt[32]. On the genome level, it was shown that halophile organisms tend to have high level of GC content[33,34]. Altogether, one can speculate that specific mechanisms for high osmolality adaptations, such as network that facilitate the production of different types of osmolytes, will have a higher chance to be horizontally transferred than broader adaptations that exist across the genome.In this work, we aimed to study HGT by conducting lab HGT experiments. We were interested in asking: To what extent HGT is utilized as an adaptive means compared to other sources of genetic variation? How phylogenetic relatedness affect chance of successful transfer? What are the dynamics of HGT evolution? And how to identify all horizontally transferred elements in a genome with minimal false positives and false negatives? We established an experimental setup in which *B. subtilis*, a non-halophile bacterium, was cultured in high osmolality environment. *B. subtilis* was either exposed or not to different sources of DNA from diverse halophile bacteria. The sources of DNA were clustered into three groups, according to the phylogenetic distance between the donors to the recipient. We then examined phenotypically how well it adapted and are sequencing the genome of the evolving strains to detect HGT events. We suspect that in one line of evolution, in which *B. subtilis* was exposed to DNA originating from other halophile bacilli, the bacteria have evolved through

HGT. Notably, the DNA in this line of evolution is from the donor DNA group with the least phylogenetic distance from the recipient. Sequencing of the evolved strains is underway and we hope that soon we will analyze this evolution on the genomic level.

3 Materials and Methods

3.1 Strains and media

Table 1- List of strains

DNA group	Species	Strain	Halophile traits	Genotype	Received from	Reference
Recipient strain	*Bacillus subtilis*	168-comp	Non	168 His met srfA-lacZ [tet] amyE::xylR Pxyl-comK [ery] (comQ comX comP replaced by B. mojavensis RO-H-1 homologs)	Dr. Avigdor Eldar, TAU	
D-Bacilli	*Bacillus subtilis*	RO-FF-1	Showed tolerance for 0.8M NaCl LB	WT	Dr. Avigdor Eldar, TAU	35
		RS-D-2	Non			
		RO-E-2				
	Bacillus Mojavensis	RO-H-1				
D-Bacteria	*Halobacillus halophilus*	Unknown	Moderate	WT	Prof. Aharon Oren, Huji	
	Halomonas elongata		Extreme			
	Vibrio fischeri		Slight		Prof. Uri Gophna, TAU	
D-Archea	*Haloarcula marismortui*	Unknown	Extreme	WT	Prof. Uri Gophna,	

	Haloferax dentrificans	ATCC 35960			TAU	36
	Haloferax mediterranei	ATCC 33500				37
	Haloferax volcanii	WFD11				38

All evolution and growth analysis were done on LB media supplemented with a total of 0.8M NaCl. For the purpose of DNA extraction, vibrio fischeri was grown on Marine broth media (Difco). Halobacillus halophilus, Halomonas elongata and all four archea were grown on YPC media as specified in the Halohandbook[39]

3.2 DNA purification

3.2.1 Bacterial DNA purification

Bacterial DNA was purified using Wizard Genomic DNA Purification Kit (Promega) according to the manufacture guidelines. The protocol was adjusted for purifications of large volumes (bacterial culture volumes of 10 mL and above). Incubation with Lysozyme was for 2 Hours. All other incubations in 80^{o}c were for 20 minutes. Samples were centrifuged for 10 minutes each time.

3.2.2 Archeal DNA Purification

Archeal DNA was purified according to the protocol specified in The Halohandbook[39]. The protocol was adjusted for purifications of large volumes (Archea culture volumes of 10 mL instead of 1 mL as listed in the protocol). All buffers used in the experiment were larger by a factor of 10 then what is mentioned in the protocol.

All other incubations in 80^{o}c were for 20 minutes. Samples were centrifuged for 10 minutes each time.

3.3 Lab evolution in serial dilution

Lab evolution in serial dilution was done on LB media with NaCl concentration of 0.8M. Evolution was done in 30˚c which is ideal for *Bacillus subtilis* growth. Each evolution line was conducted in three replicates. Cells were grown in liquid media, 1.2ml each, in a 24-well plate. Cells were diluted daily (1:120) into fresh media; dilution was done once per day. Every 6 dilutions cells were frozen in 30% Glycerol and are kept in -80˚c.

3.3.1 Lab evolution at the presence of foreign DNA

Serial dilution Lab evolution with foreign DNA was done as described in section 4.3.3 however the 0.8M NaCl LB media was supplemented with ~2 µg of foreign DNA mixture (Table 1) all mixtures were composed of equal amounts of the various DNA sources. In addition, to reduce the chance for contaminations ~20$\frac{\mu g}{mL}$ of erythromycin was added to the growth media so to avoid contaminations. The grown *Bacillus subtilis* strain was a modified 168 strain (see strains section). It is modified to be xylose induced competent and is erythromycin resistant.

3.4 Lab evolution in a Chemostat

Chemostat experiments were performed using a DASBox Bacterial Fermentation system (DASGIP, Eppendorf). Bacterial cells were grown in the fermenters at 30^0C in 100ml LB media containing 0.8M NaCl supplemented with ~20$\frac{\mu g}{mL}$ erythromycin. Strain 168-comp was initially grown for ~200 generations inside one chemostat during calibrations of the growth parameters. After which it was extracted and used as the ancestor strain for the experiment. Cultures were initially grown in a batch mode for 6 hours, and were then switched to chemostat mode with a doubling time of 4hrs (dilution of 0.173/hour, flow rate of 17.3 ml/hour). During the run various variables were monitored in the chemostat including the temperature, pH, DO and the OD of the cultures. Once a week cell samples were collected from the chemostat for downstream laboratory analysis and storage.

During the run the two cultures were treated seven times as follows: Both cultures were switched from chemostat mode to batch mode. The experimental culture was challenged with foreign DNA while the control culture was challenged with the vehicle only (TE Buffer) (see 4.5 Evolution in a chemostat). for the exact DNA injection regime). After four hours both cultures were switched back to chemostat mode in the same conditions as before.

Table 2 – list of DNA pulses

Gnerations	DNA source	Mass (µg)	Volume (ml)	Final chemostat DNA concentration ($\frac{\mu g}{ml}$)
~215	Pre-evolved *RS-D-2*	27	1.5	$\frac{(27+75)\mu g}{100\ ml} = 1.02$
	Pre-evolved *RO-H-1*	75	1.5	
~260	Pre-evolved *RO-H-1*	293	1.1	3.95
	Pre-evolved *RS-D-2*	102	1.1	
~305	Pre-evolved *RO-H-1*	270	1	4.83
	Pre-evolved *RS-D-2*	67	1	
	Pre-evolved *RO-E-1*	95	1	
	Haloarcula marismortui	51	1	
~345	Pre-evolved *RO-H-1*	123	1	2.29
	Pre-evolved *RS-D-2*	44	1	
	Pre-evolved *RO-E-1*	62	1	
~390	Pre-evolved *RO-H-1*	189	0.7	4.2
	Pre-evolved *RS-D-2*	136	1	
	Pre-evolved *RO-E-1*	77	1	
	Haloarcula marismortui	20	0.4	
~455	*Halomonas elongata*	21	0.27	3.65
	Halobacillus halophilus	31	0.27	
	Vibrio fischeri	50	0.27	
	Pre-evolved *RO-H-1*	8	0.27	
	Pre-evolved *RS-D-2*	13	0.27	
	Pre-evolved *RO-E-1*	15	0.27	
	Pre-evolved *RO-FF-1*	34	0.27	
	Haloarcula marismortui	43	0.27	
	Haloferax dentrificans	22	0.27	
	Haloferax mediterranei	40	0.27	

	Haloferax volcanii	88	0.27	
~515	*Halomonas elongata*	21	0.27	4.89
	Halobacillus halophilus	31	0.27	
	Vibrio fischeri	69	0.27	
	Pre-evolved *RO-H-1*	63	0.27	
	Pre-evolved *RS-D-2*	40	0.27	
	Pre-evolved *RO-E-1*	15	0.27	
	Pre-evolved *RO-FF-1*	57	0.27	
	Haloarcula marismortui	43	0.27	
	Haloferax dentrificans	22	0.27	
	Haloferax mediterranei	40	0.27	
	Haloferax volcanii	88	0.27	
~555	*Halobacillus halophilus*	34	0.3	4.16
	Vibrio fischeri	77	0.3	
	Pre-evolved *RO-H-1*	9	0.3	
	Pre-evolved *RS-D-2*	14	0.3	
	Pre-evolved *RO-E-1*	45	0.3	
	Pre-evolved *RO-FF-1*	37	0.3	
	Haloarcula marismortui	33	0.2	
	Haloferax dentrificans	24	0.3	
	Haloferax mediterranei	45	0.3	
	Haloferax volcanii	98	0.3	
~580	Chloramphenicol resistant *B. subtilis 168*	271	1.5	2.71

3.5 Growth Experiments

Cells were grown in 3 ml of LB media in 30˚c for 2 days, to reach deep-stationary phase. Cells were then diluted into fresh LB 0.8M NaCl (1:100). Growth experiments were done in 96-well plates, 150µl per well. Each plate contained two strains that were arranged in a checkerboard pattern across the plate. Plates were shaken in an incubator set to 30˚c. OD 600 was measured

every 1 hour for ~24 hours by a plate reader (infinite 500 Tecan). All measurements were done automatically using a Hamilton robotic system.

3.5.1 Growth curves analysis using 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[40]. Curveball fits a mathematical model to growth curves and outputs biological parameters: maximal population density (maximal OD), lag time and maximal growth rate for the fitted curve. The parameter fitting is based on Baranyi-Roberts model[41] which is defined by the set of differential equations (1-3):

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

$$(2) \quad N(0) = N_0$$

$$(3) \quad \alpha(t) = \frac{q_0}{q_0+e^{-mt}}$$

Where N is the population density, t is time, r is the per capita growth rate, $\alpha(t)$ is the adjustment function for the lag phase, K is the maximal population density, v is a de-acceleration parameter, N_0 is the initial population density, q_0 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 q_0 is accumulated.

The solution for the set of the differential equations is (4-5):

$$(5) \quad N(t) = \frac{K}{(1-(1-(\frac{K}{N_0})^{v}e^{-rvA(t)})^{\frac{1}{v}}}$$

$$(6) \quad A(t) = \int_0^t \alpha(x)dx = t + \frac{1}{m}\log(\frac{e^{-mt}+q_0}{1+q_0})$$

From the model we can extract the biological parameters: Maximal OD (parameter K in the model), maximal growth rate ($\max(\frac{1}{N} * \frac{dN}{dt})$) and lag time (time t where tangent line in the point of maximal growth rate intersects $y = N_0$).

3.6 Competence Assay

Cells were grown overnight in 3ml of LB at 30°C to reach stationary phase. Cells were then diluted into 1.2 mL of fresh 0.8M NaCl LB media (1:100) containing ~2$\frac{ng}{\mu L}$ of genomic DNA that was purified from Chloramphenicol resistant Bacillus subtilis strain. As a control another dilution was done to 1.2 mL of the same media with the exclusion of the DNA. Cells were incubated overnight at 30°C. Then 200 µL of each aliquot was plated into a Chloramphenicol+ petri dish (35$\frac{ng}{\mu L}$) that was then incubated over night at 30˚c. The day after, the number of cells on each plate were counted.

3.7 Fluorescent microscopy

Cultures were grown overnight in 0.8M NaCl LB media in a 24 wells plate and then a sample was taken for OD measurement. For microscopy analysis a sample (20 µl) was washed with PBS, then centrifuged (8000 RPM, 2 minutes), suspended in 20 µl PBS. The sample was recentrifuged and then suspended in 5 µl of PBS supplemented with 1µg/ml membrane stain 4-64 (Thermo Fisher). Then the cells were placed on a microscope slide and covered with a poly-L-Lysine (Sigma)-treated coverslip. Cells were visualized and photographed using an Axioplan2 microscope (Zeiss) equipped with a high-resolution microscopy Axiocam camera, as required. Data were captured using Axiovision suite software (Zeiss).

3.8 Phylogenic calculation

Previous phylogeny calculation was done to strains RS-D-2, RO-FF-1, RO-E-2 and RO-H-1 using the concatenation of three genes: dnaJ, gyrA and rpoB [35]. We recalculated the phylogenetic distance using the same sequenced data of those genes and by adding the data of *Bacillus subtilis* 168 homologous genes. The data for strains RS-D-2, RO-FF-1, RO-E-2 and RO-H-1 was downloaded from the website http://www.straininfo.net. The data for strain 168 was downloaded together with the complete genome sequence of strain 168 from NCBI (http://www.uniprot.org/uniprot/P08622). Phylogeny was calculated in a few steps. First, the three genes were concatenated (1406 nucleotides combined). Second, unaligned sections (/gaps) were eliminated. Third, a pairwise distance was calculated between each strain's concatenated genes. The pairwise distance calculates the proportions of sites at which two nucleotides are different.

3.9 CFU

To determine the number of live cell count, culture were grown overnight in 0.8M NaCl LB media in a 24 wells plate. OD measurement was taken and then cells were diluted in phosphate-buffered saline (PBS) (1:10^6). Then 200µl from the diluted culture were plated on LB-plates. Plates were incubatied at 30^{o}c for ~24 hours and colony forming units (CFU) were counted.

3.10 Next Generation Sequencing

Samples of both the evolved strains and its ancestor were sequenced using HiSeq Illumina with reads of 125 bp paired ends. For each sample, the reads were aligned using Bowtie2[42] against the genome of Bacillus subtilis 168 WT[43]. Alignment results were viewed using IGV alignment viewer[44]. Horizontally transferred DNA section was found by browsing the aligned reads of all the evolution lines in the search for areas in which mismatches appeared in the evolved's alignment but not in the ancestor's. In order to retrieve the sequence of the evolved strain (within the suspected region for horizontal gene transfer) we used the pileup format generated using SamTools[45]. The sequence (4Kb long) was then aligned to the different donors' genomes (for donors with unknown sequenced genome see section 3.10.1). The alignment was done using blast-2.3.0+ (legacy version)[46].

3.10.1 *De-Novo* assembly of DNA donors' genomes

Each DNA donor organism in the HGT experiment with no known sequenced genome in the literature was sequenced using HiSeq illumine with reads of 125 bp paired ends. Then, the genome of each organism was assembled using SPAdes assembler[47] (version 3.6.2) with 'careful' option.

4 Results

4.1 Experimental design

Emulating evolution with HGT was done by conducting in-lab evolution for highly competent *B. subtilis* in high salt (0.8M NaCl). The bacteria were grown either in batch mode (followed with a process of daily dilutions) or in a chemostat (Figure 1). In the serial dilution experiment, four evolution groups were done in-parallel. In three groups the bacteria were exposed to foreign DNA that was extracted by me from a span of different donor species that range in phylogenetic distance from the recipient. We termed the three groups as 'D-Bacilli' –DNA was originated from bacilli that were pre-adapted to high salt media, 'D-Bacteria' – DNA was originated from non-bacilli bacteria with different halophile traits (slight, moderate or extreme halophile) and 'D-Archea' – DNA was originated from extreme halophile archea (Table 1). In these three evolution groups the evolving bacteria had the chance to evolve by acquiring genes from the provided external DNA. As a control (termed 'No DNA') we grew *B. subtilis* in the same conditions without exposing it to foreign DNA but only to the buffer in which we extract DNA. We performed this line of evolution in order to compare evolution with HGT to one without. In addition, we cultured a second control group (termed 'Self DNA') in which the bacteria were exposed to its ancestral DNA. This evolution line was created to control for the possible effect extracellular DNA on the evolving bacteria regardless for its gene content; for example, external DNA can be used as a nutrition source. During phenotypic analysis of the evolution of the different evolution lines, no significant difference was found between No DNA to Self DNA and so we decided to focus more on No DNA. Thus, this work we will not include analysis done on "Self DNA".

In a second evolutionary setup we grew *B. subtilis* in a chemostat on the same high salt media (0.8M NaCl). The bacteria grew in two chemostats. In the first, bacteria were not exposed to foreign DNA. In this line of evolution bacteria could have evolved only by acquiring mutations in its genome. In the second line of evolution bacteria were exposed to foreign DNA from D-Bacilli, D-Bacteria and D-Archea combined. In this line of evolution, the evolving *B. subtilis* could have evolved also by integrating and utilizing foreign DNA fragments.

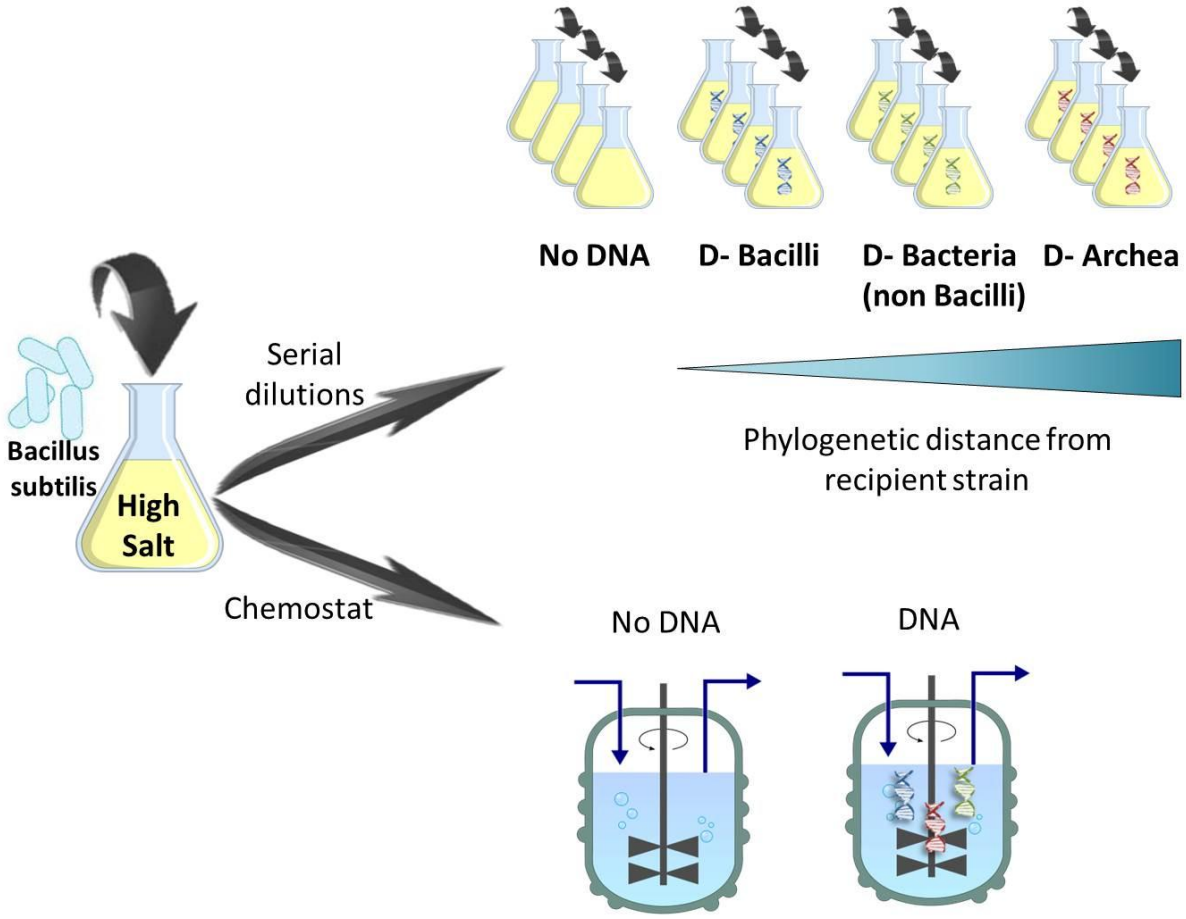

Figure 1 - Experimental design

A- Bacillus subtilis was grown in high salt media either in a chemostat or in serial dilution. In the serial dilution setup four evolution groups were created. The first group, termed 'No DNA', B. subtilis grew in high salt media containing no foreign DNA. In the latter three evolution groups foreign DNA that was supplemented to the evolution experiment was composed different groups of DNA donors, ranging in phylogenetic distance from the recipient. 'D-bacilli' was composed of bacilli that were adapted a priori to high salt media, 'D-bacteria' was composed of non-bacilli halophile bacteria and 'D-Archea' was composed of extreme halophile Archea. In the chemostat experiment two lines of evolution were created. In the first, No DNA was supplemented to the chemostat. In the second DNA was supplemented once a week from the combined pool of D-Bacilli, D-Bacteria and D-Archea.

4.1.1 Recipient strain, '168-comp', competence evaluation

The chosen model for the project was *B. subtilis* which is a gram positive bacterium that possess conserved machinery for the insertion and integration of foreign DNA. We acquired a specific strain of *B. subtilis* 168 (one of the most studied *B. subtilis* strains) with improved competence traits. This strain, which we termed '168-comp' had a second copy of the gene comK (the key

regulator for competence) under a leaky xylose promoter. We assayed the competence of this strain on high salt media (0.8M NaCl LB) and found that the strain is competent even without xylose induction. During the HGT evolution experiments we performed routine competence evaluation to the evolved strain to make sure that it maintained its competence. In all lines of evolution in the HGT experiments, the evolved strains kept their competence behavior throughout.

4.1.2 Foreign DNA pool

Donor DNA was composed of three different groups of donors with ranging phylogenetic distance from the *Bacillus subtilis* 168-comp (Table 1). A group of extreme halophile archea (termed 'D-Archea'), a group of bacteria with different halophile traits (termed 'D-Bacteria') and a group of bacilli (most are non-halophile) that I evolved in-lab (see section 3.2 in the results) under high salt prior to extracting their DNA ('D-Bacilli'). Bioinformatics studies showed that there is a positive correlation between successful HGT events to phylogenetic similarity between donor and recipient[10,16,18]. Also the integration of foreign DNA by *Bacillus subtilis* is believed to rely primarily on homologous recombination[48]. We therefore hypothesized that the closer an organism is to *B. subtilis*, the higher would be the chance for a successful HGT event to occur, including in our evolutionary setup. On the other hand, the less similar an organism is to *Bacillus subtilis*, the higher are the chances that a successful HGT event provides an innovative gene for the recipient.

4.1.3 Salt concentration calibration

To determine the salt concentration to use in the evolution experiments we looked into the effect of various salt concentrations on both 168-comp and on the non-halophile bacteria. We wanted to maximize the stress that strain 168-comp will be exposed to so to have bigger selection pressure in its evolution. But, we also wanted the stress to be mild enough so the strain will reach stationary phase in ~24 hours. In this way we could set up the serial dilution experiment on a daily dilution basis. Also, we planned to evolve the non-halophile bacilli under the same high salt conditions prior to extracting their DNA that would later serve as a donor. Therefore, we had to choose a condition with the same constraints for the other bacilli as for 168-comp. We grew the 168-comp strain and the non-halophile bacilli on ranging concentrations of NaCl. OD was measured a day after, upon reaching stationary phase (Figure 2-A). As can be expected most

bacilli strains showed a decline in OD upon increase in NaCl concentration. Interestingly, one species/strain, *B. subtilis* RO-FF-1 showed a slight increase in OD with increase in salt concentration. After this crude screening we continued in performing complete growth assays for the 168-comp strain in some of the above NaCl concentrations. We selected LB-0.8M NaCl to be used in the experiments (Figure 2-B) since in this concentration the non-halophile bacilli were still viable while 168-comp showed a significant reduction in fitness in comparison to growth on regular LB.

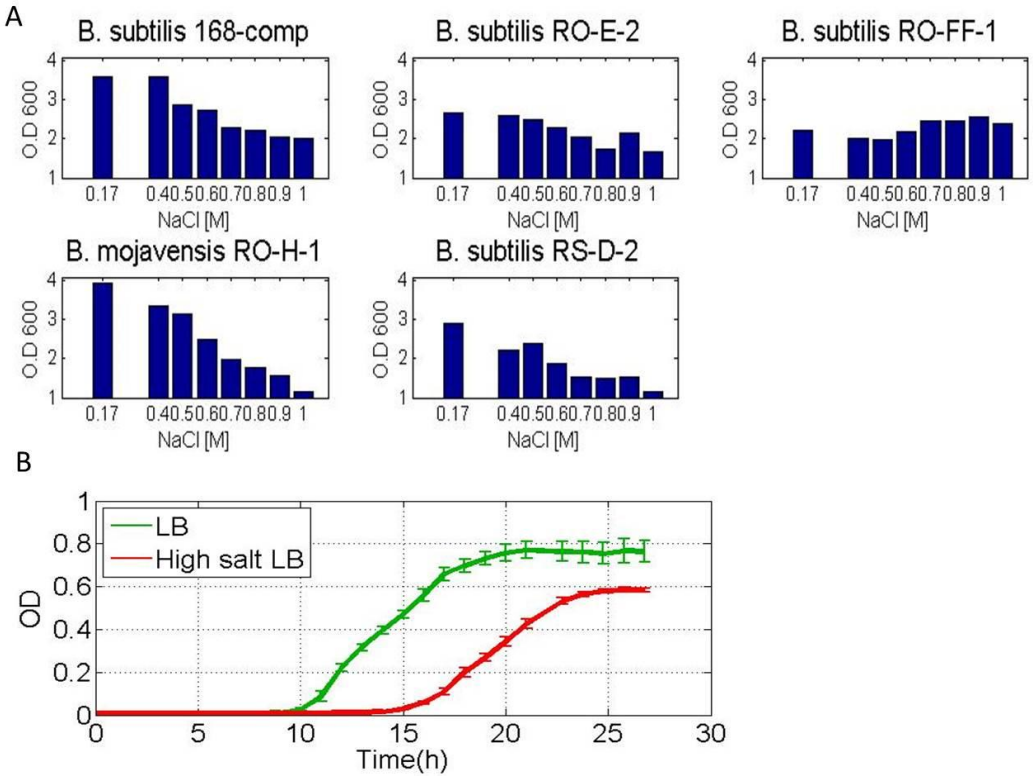

Figure 2 – Salt calibration

Screening for optimal NaCl concentration was done by analyzing the growth of the different strains on different salt concentrations. A – Strains/species were cultured for ~24 hours in ranging NaCl concentration. After which a sample was taken for each sample and its OD was measured. 0.17M NaCl is the concentration in standard LB medium. Except *for B. subtilis* RO-FF-1 all strains showed reduction in maximal OD as NaCl Concentration increased. B- Growth curve for the recipient strain, 168-comp, on 0.8M NaCl LB (red) and on LB (green).

4.2 Phylogenetic characterization of the donor Bacilli

For the purpose of the HGT experiment we wanted to acquire DNA of bacilli which are closely related to *B. subtilis* 168-comp and are also fitted to grow on high salt. We acquired four types of Bacilli that were not characterized as halophiles (three *Bacillus subtilis* strains and one *Bacillus mojavensis*) (Table 1). We checked the growth of all four strains on ranging concentrations of salt and concluded that they are non-halophile except for *B. subtilis* RO-FF-1 which seems to reach higher maximal OD when grown on high salt (Figure 2-A). Since most of our chosen Bacilli were found not to be halophiles we decided to evolve these bacilli to high salt so to later on use their DNA as a possible source for HGT evolution of our recipient strain, 168-comp.

Since we planned to detect HGT events through sequencing, we became aware that knowing how different these strains' genomes are from 168 is crucial. If the recipient and donor genomes are very close to one another then it might be hard to distinguish mutations from HGT events in our evolving 168-comp strain. Three of the donating bacilli strains/species are without a sequenced genome, yet the phylogeny of all four was established based on the sequence of three highly conserved genes shared among them (dnaJ, gyrA, rpoB)[35]. In order to roughly estimate how 168-comp's genome differs from that of the four donating bacilli we constructed a phylogenetic tree using the sequenced dnaJ, gyrA, and rpoB genes (Figure 3). Among the four strains, strains RS-D-2 and RO-FF-1 were more similar to the 168-comp strain (with ~0.01 probability for a base difference in each position). Strains RO-E-2 and RO-H-1 were less similar to the 168-comp strain (with ~0.06 probability for a base difference in each position). Note that these are highly conserved genes, so the rate of sequence divergence in other genes is likely to be even higher. Thus it appears that these substantial distances are likely to allow easy and safe detection of HGT events: deep sequencing reads of about 100 nucleotides are likely to differ by more than one, or six (depending on donor) between the ancestral 168-comp strain and the evolved strains if evolution occurred through HGT. We thus concluded that HGT events will be detectable through deep sequencing and we have thus proceeded to adapt these strains to high salt.

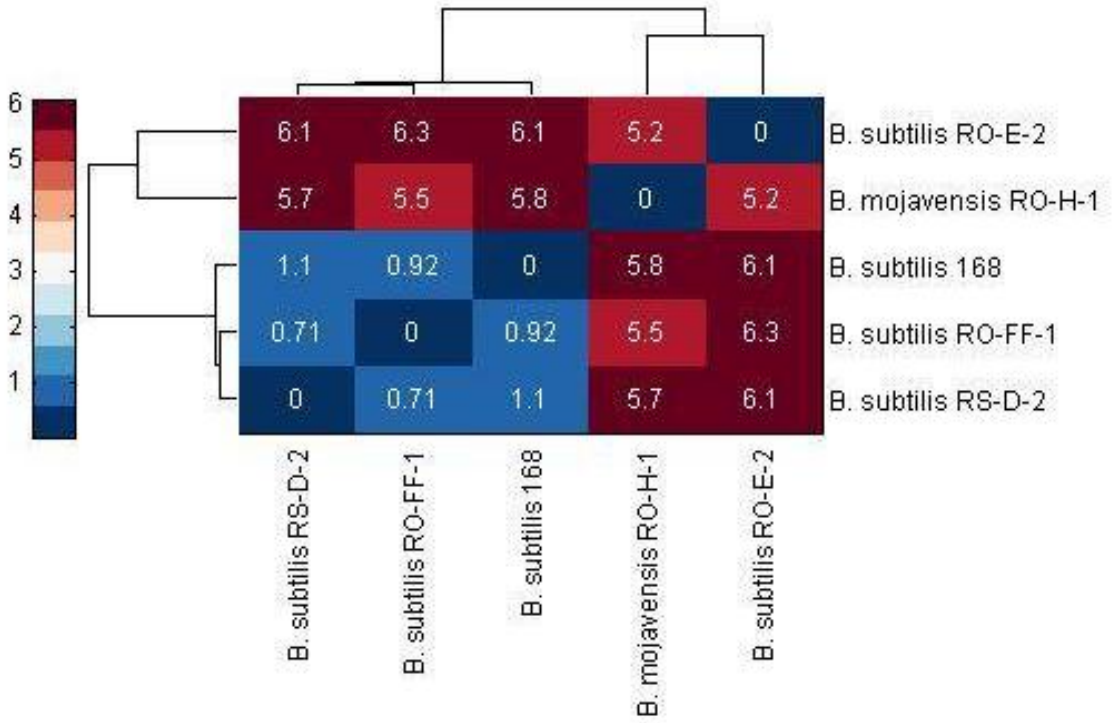

Figure 3 – Percent nucleotide difference between the recipient strain (*B. subtilis 168*) to the other bacilli

Phylogenetic distance between strain *B. subtilis 168* to the other bacilli was done by concatenating the sequence of three genes: dnaJ, gyrA and rpoB. The concatenated sequences were aligned and pairwise distance was measured between the different strains. The colors and numbers (white) represent the percent nucleotide difference between two strains.

4.3 Evolution of the donor bacilli to high salt

In order to adapt the four bacilli to high salt, we conducted a serial dilution evolution experiment in high salt media. The four strains/species grew on LB-0.8M NaCl media until reaching stationary phase and diluted daily 1:120 (estimated 7 generations a day). Samples were taken for growth assay analysis after ~200 generations (Figure 4- A). Strains *B. subtilis* RO-E-2, *B. mojavensis* RO-H-1 and *B. subtilis* RS-D-2 showed fitness improvement (higher maximal OD and/or bigger maximum growth rate) when grown on LB-0.8M NaCl and no apparent fitness improvement when grown on LB. surprisingly, *B. subtilis* RO-FF-1 strain showed no improvement, in fact actually a decrease in fitness to high salt, and no apparent change to regular LB.

In order to quantitate the growth assays' results we analyzed each growth assay with a parameter fitting program (in a compute package called 'Curveball', see Methods). This application fits a mathematical model to growth curves and outputs biological parameters for that curve[40]. For each growth assay we analyzed how well did the evolved strains behave in comparison to its ancestor in terms of maximal OD, lag time and maximal growth rate (Figure 4-B). The quantification of the biological parameters of the growth curves strengthened what we observed by eye when examining the growth curves; Improvement in maximal OD and/or maximal growth rate in LB-0.8M NaCl for all evolved strains/species except for *B. subtilis* RO-FF-1. *The* quantification also revealed that lag time have decreased (and thus, improved) in *B. subtilis* RO-E-2 and *B. subtilis* RS-D-2 when grown in LB as well as in LB-0.8M NaCl.

The evolution continued for another ~300 generations. During which, no further significant growth adaptation was detected in growth assays (data not shown). We therefore concluded that the evolution lines reached their maximal improvement under high salt. The DNA of the samples in generation ~500 was used as a DNA source for the HGT experiments.

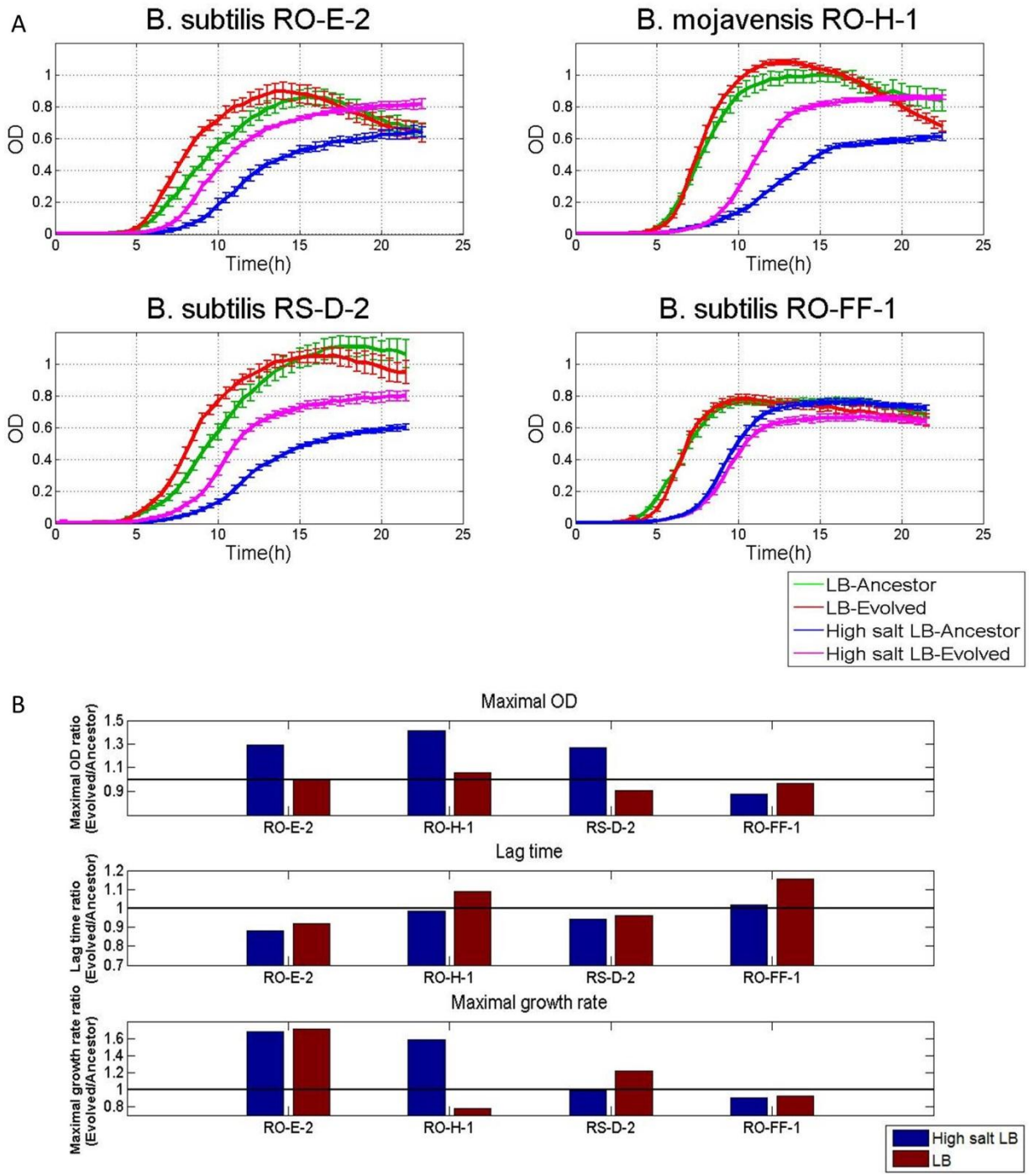

Figure 4 – Growth assays of the Bacilli evolution on high salt

Evolved strains (~200 generation) and their ancestor were grown on either LB or LB-0.8M NaCl. For each, OD over time was measured. A- Growth curves for each strain, before and after evolution, in regular and high salt concentrations. B- For each curve in A, biological parameters (maximum OD, lag time, maximum growth) were extracted using a parameter fitting algorithm ("Curveball"). Each bar plot represents the ratio of the values between the evolved to the ancestor for one of those parameters; both on high salt (blue) and normal salt (red). Black line

indicates the height in which there is no difference between the evolved to the ancestor. Note that for lag time and max growth lower values mean higher fitness.

4.4 HGT experiment in serial dilution

In order to emulate HGT evolution we set to evolve *B. subtilis* in high salt with the presence of foreign DNA. Since the acquisition of foreign DNA in *B. subtilis* is considered to be primarily through homologous recombination, we hypothesized that DNA from closely related donors will have more chance of being integrated into the evolving bacteria genome. On the other hand, we postulated that DNA from more distant species might have a higher chance to provide an innovative new gene to the recipient. This line of thinking led us to decide to split the evolution of *B. subtilis* into a few parallel lines of evolution. In each evolution line the evolving bacteria were exposed to a different pool of DNA fragments. Each DNA pool was composed of donors with different phylogenetic distance from the recipient.

We conducted HGT evolution in four groups (Figure 1); all with a shared ancestor and each with three biological replicates (termed 1-3). In the first evolution group (Termed "No DNA"), evolving *B. subtilis* were not exposed to external DNA. This line of evolution was conducted in order for us to be able to compare evolution with HGT to one without. In the second evolution group (termed "D-Bacilli"), evolving *B. subtilis* were exposed to DNA originating from phylogenetic close bacilli that were pre-adapted to high salt (Table 1). In the third evolution group (termed "D-Bacteria"), evolving *B. subtilis* were exposed to DNA originating from non-bacilli bacteria with ranging halophile traits (Table 1). Lastly, in the fourth evolution group (termed "D-Archea"), *B. subtilis* were exposed to DNA originating from phylogenetic distant halophile Archea (Table 1). In addition to these four evolution groups we performed also a fifth group (termed "Self DNA"), in which the bacteria were exposed to DNA originating from the ancestral strain. This line of evolution served as additional control alongside "No-DNA". During fitness evaluation of all evolved strains (will be elaborated shortly) "Self DNA" evolved strains showed the same fitness improvement as No DNA (Data not shown) and so we decided to stop assaying this evolution line and to focus on the others. Therefore, we are presenting in this work only the analysis done on the evolution lines that are presented in Figure 1.

All lines of evolution went through serial dilution in LB-0.8M NaCl for a total of ~500 generations (seven generations per day) except for the D-Bacteria line which grew for ~330

generations (this line of evolution started after the other lines). During the evolution the bacteria were transferred daily to a fresh medium containing ~2 $\frac{\mu g}{ml}$ of foreign DNA – the concentration used in transformation protocols with *B. subtilis*.

4.4.1 Growth analysis

Our first assay for progress of evolution was growth analysis on evolved strains. When performing growth analysis we grew evolved strains in parallel to their ancestor so to define how well did they adapt. The growth assays were analyzed using a parameter fitting software (called "Curveball") so to quantitate the fitness improvement of the evolved strains in terms of maximal OD, lag time and maximal growth rate. We hypothesized that different adaptations, such as adaptations due to acquisition of DNA from different donors, could be detected by comparing the growth curves of evolved strains. For example, one strain could adapt by having increased maximal OD while another could adapt by having increased maximal growth rate. Furthermore, by comparing the growth curves of the No DNA experiments to the rest of the evolution lines, we could speculate in which evolution lines HGT occurred. We also aimed to capture the dynamics of the HGT evolution in comparison to the No-DNA evolution line by examining when each line of evolution had increased fitness.

When evolution groups: No DNA, D-Bacilli and D-Archea reached ~420 generations (D-Bacteria reached generation ~250), we performed a growth analysis assay on all evolved strains. We decided to perform only one technical repetition for D-Bacteria since it had fewer generations and we wanted to focus on evolution lines which are with the same generations and are 'older'.

Maximal OD improvement was shown in all biological repetitions of all evolution groups (Figure 5-A). Interestingly, the maximal OD of all D-Bacilli repetitions seemed slightly higher than in other evolution lines. Lag time showed no decrease (i.e. improvement) in any of the evolution lines (Figure 5-B) and an increase in maximal growth rate was shown in some of the samples regardless of their DNA group (Figure 5-C). Figure 5-D shows that the average maximum OD improvement of the D-Bacilli lines reached a higher OD value (1.179) in comparison to No-DNA (1.122), D-Bacteria (1.15) and D-Archea (1.127).

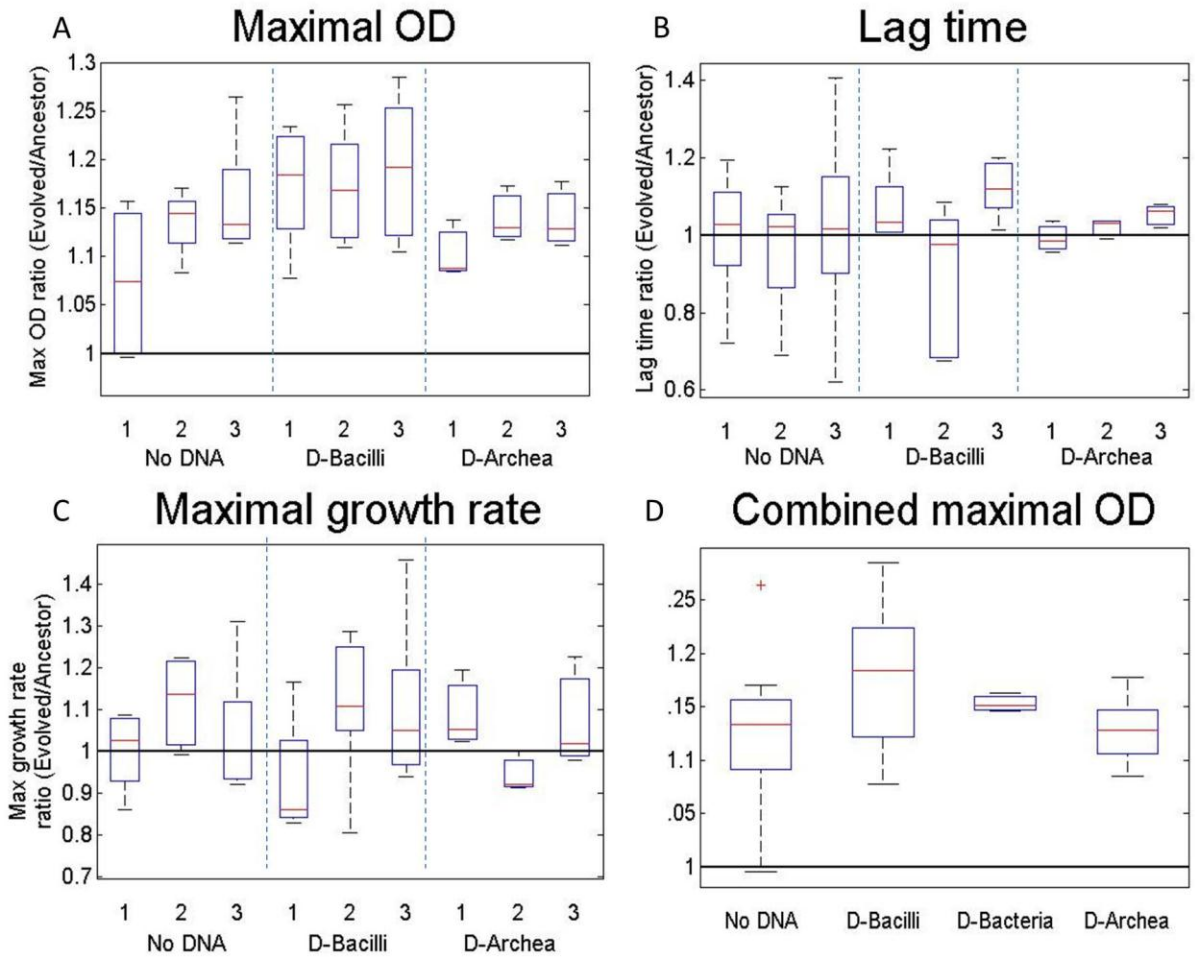

Figure 5 - Growth assays for the serial dilution experiment

Growth assay experiments were performed on evolved strains and their ancestor on LB-0.8M NaCl. For each evolution group, all three biological repetitions (1-3) were measured. The growth assays were for samples from generation ~420 of evolution groups No-DNA (N=5), D-Bacilli (N=6) and D-Archea (N=3). Generation ~250 of D-Bacteria (N=1) was also measured. The growth curves were analyzed using parameter fitting software ("Curveball") and three biological parameters were quantified: maximal OD, lag time and maximal growth rate. (A-C) Box plot representation for maximal OD, lag time and maximal growth rate for all repetitions of No-DNA, D-bacilli and D-Archea. Black lines indicate the height in which there is no difference between the evolved to the ancestor. (D) The combined maximal OD ratio between all biological repetitions of No-DNA, D-Bacilli, D-Bacteria and D-Archea. Black line indicates the height in which there is no difference between the evolved to the ancestor.

4.4.2 Colony forming unit assay

While OD measurement provides a good approximation for cell number it can be influenced by other parameters such as cell size, granularity and cell shape. In order to check whether the detected increase in maximal OD in D-Bacilli represents an increase in the number of viable cells we performed a colony forming unit assay (CFU) (Figure 6). Repetition #2 of each evolution group was grown over night on LB-0.8M NaCl to reach stationary phase. Then, the OD of each sample was measured so to later be used to calculate the live cell concentration. A similar trend in OD appeared here as in previous growth assay experiments; Ancestor has lower OD than the evolved strains while the D-Bacilli has higher OD compare to all other evolution lines. We then diluted all strains 1:10e6 and plated on LB plates. From the number of colonies counted on each plate we could extrapolate the concentration of live cell in each initial culture. D-Bacilli showed the highest number of live cell concentration (significance of 0.054 when tested against the ancestor). Surprisingly, the ancestor did not have the lowest concentration of live cell, although this result was not statistically significant when tested against No-DNA, D-Bacteria and D-Archea. These results further strengthens our speculation that in the D-Bacilli lines of evolution, the adaptation was different from in all other lines of evolution, possibly due to HGT.

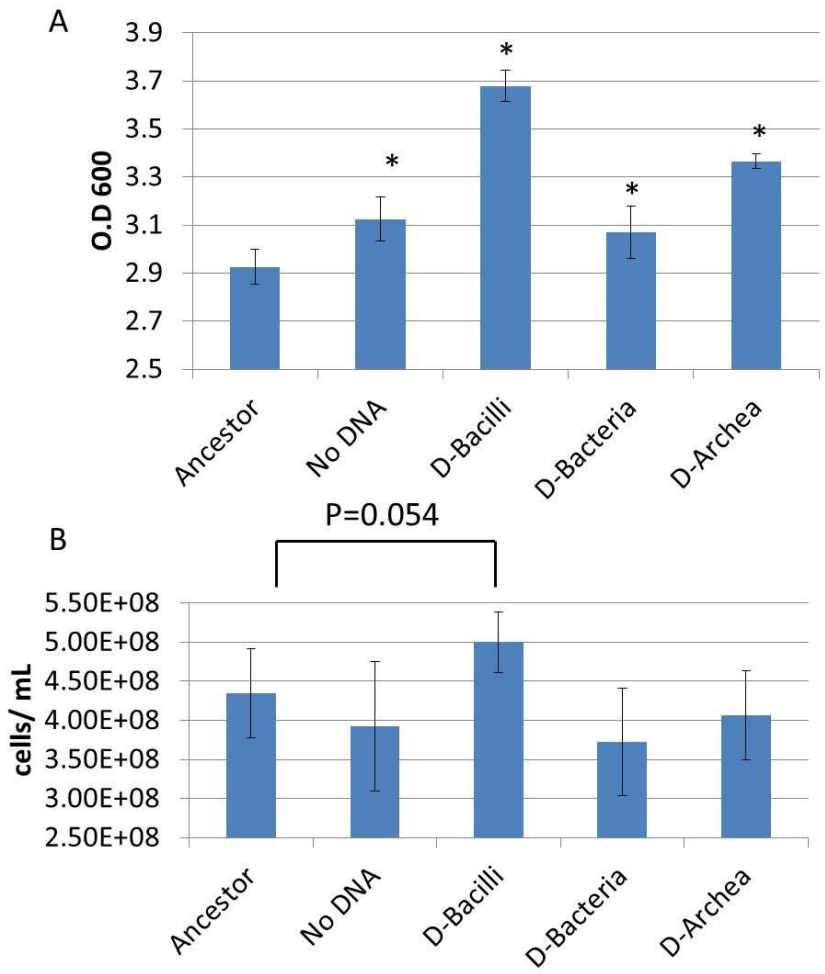

Figure 6 - Live count for the strains from the HGT evolution in serial dilution

The ancestor and repetition two of each evolution group (all at generation ~420 except D-Bacteria generation ~250) were grown for ~20 hours, so to reach stationary phase. OD was then measured and the samples were diluted and plated on LB plates so to determine live cell concentration. A- Four repetitions of culture were grown overnight to reach stationary phase. OD was measured for each. Standard deviations are shown. Star represents significant difference from ancestor. B- Cell Forming Units (CFU) was done for the samples of A. From the results the original concentration of the live cells was calculated and showed in this bar plot. Standard deviations are shown. The only significance between an evolved strain to the ancestor was in the D-Bacilli strain (p=0.054).

4.4.3 Cells morphology analysis using microscopy

As mentioned above, difference in maximal OD can occur due to change in cell morphology. To examine if cell morphology had occurred in these evolution groups we examined how the ancestor and the evolved strains: No DNA repetitions 1-2 and D-Bacilli repetitions 1-3 (generation 420), look under the microscope (Figure 7). Bactria tends to arrange in multi-cellular chains that many consist of a handful of cells connect to one another. To quantify the results the chain lengths of 300 cells from each sample were measured using graphical software ("imageJ"). It appears that in both evolution lines No DNA and D-Bacilli, the strains had evolved to have longer chains in stationary phase. It is interesting to ask what type of advantage, if at all, does longer chains confer in our evolutionary setup. Interestingly, two out of the three repetitions in D-Bacilli seemed to have longer chains in comparison to the No DNA strains.

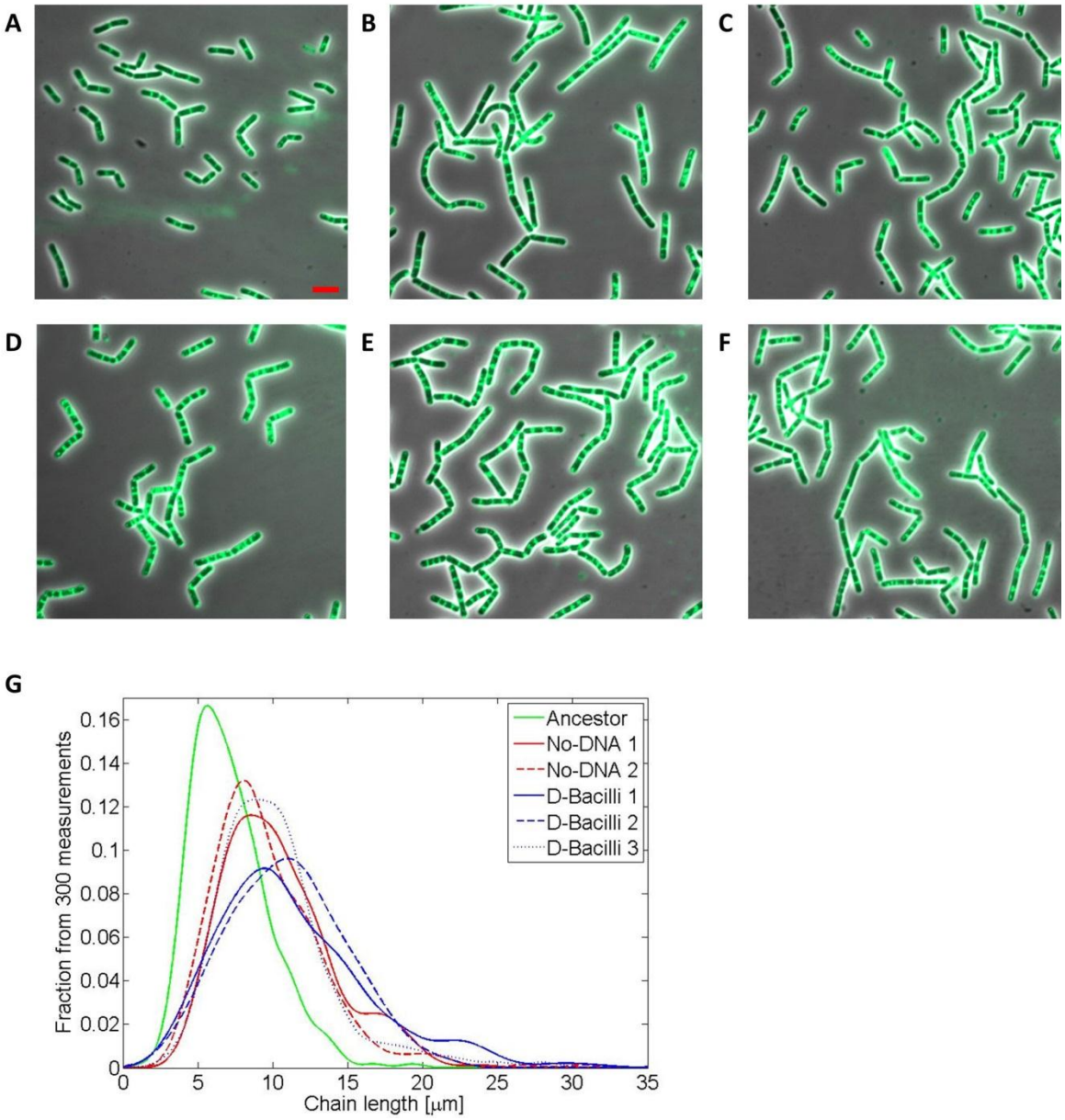

Figure 7 - Cell morphology examination for ancestor and evolved strains

Ancestor and five evolved strains (No-DNA 1-2, D-Bacilli 1-3) of generation ~420 were grown in high salt LB for ~20 hours, to reach stationary phase. Samples were then imaged and the length of 300 chains was measured for each sample using a graphical software ("imageJ"). (A-F) Microscopy images of ancestor, No DNA 1, No DNA 2, D-Bacilli 1, D-Bacilli 2, D-Bacilli 3 accrodingly. Cell membrane is labeled (green). Scale bar correspond to 5μm. (G) Fraction distribution of chain lengths for each sample.

4.4.4 Identification of HGT through next generation sequencing

In order to identify potential cases of horizontal gene transfer we sequenced all evolved strains and also their shared ancestor. We then aligned the reads of the sequencing to a reference genome of *Bacillus subtilis* 168 WT and looked for long DNA segments in which the evolved strains show difference in number of mismatches/gaps from the ancestor (Figure 8). We identified a possible HGT event that occurred in D-Bacilli 2 and D-Bacilli 3. This region was 4,306 and 10,535 bps in length in the two respective cases. We decided to focus on the shared segment between these two regions. In this segment the percent identity between reads and the ancestral genome dropped from close to 100% to a level of 97.629% (102 mismatches) – see (Table 3). To prove that this represents genuine HGT events, we aligned this shared DNA stretch to our partially assembled genome of the DNA donors in D- Bacilli evolution line. Through this analysis we found that the DNA 4 Kb segment matches 99.977% to the DNA in one donor bacilli (zero mismatches and one gap) (Table 3), demonstrating that this DNA segment came from that donor. Other than this segment I could not detect in this initial analysis additional such segments that appear to have been horizontally transferred to this or to any of the other evolutionary lines.

Table 3 - Suspected HGT DNA segment alignment to DNA donors

Alignment against reference genome:	Percent identity	Alignment length	Number of mismatches	Number of gap opens
Bacillus subtilis 168 WT	97.629	4302	102	0
Bacillus subtilis RS-D-2	99.977	4306	0	1
Bacillus subtilis RO-FF-1	97.84	4306	93	0
Bacillus mojavensis RO-H-1	88.152	4296	506	3
Bacillus subtilis RO-E-2	93.36	4292	285	0

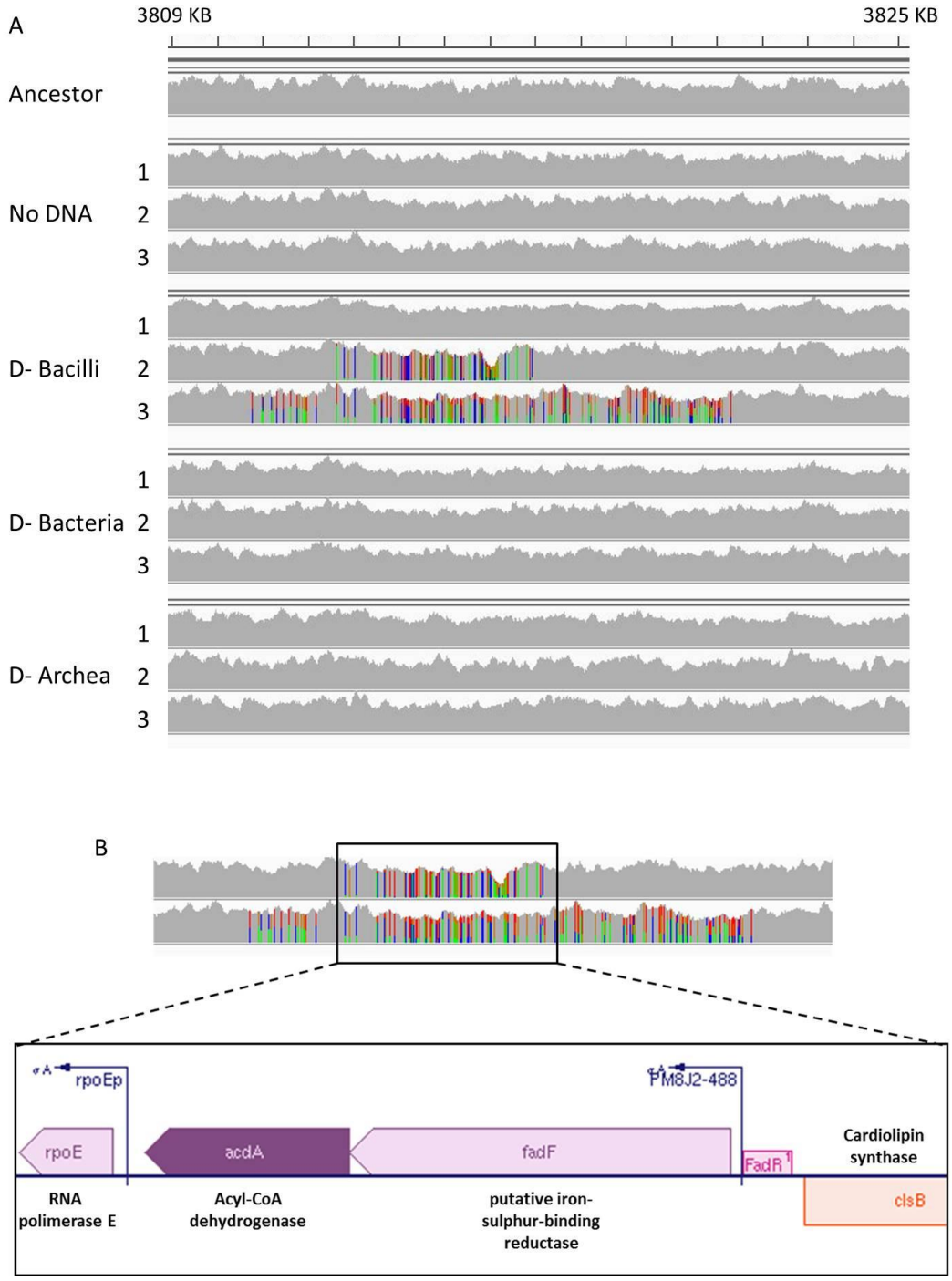

Figure 8 - Horizontal gene transfer detection through next generation sequencing

HGT event was found by comparing the number of mismatches in the evolved strains to that of the ancestor. A- Alignment representation of a ~16 Kb genomic segment in the ancestor and in all of the evolved strains. The position on the genome is represented on top. Gray landscape represents perfect matches of reads to the ancestor. Vertical colored bars represent mismatches. Gray landscape's height represents number of read that correspond to that match B- Gene annotation for the shared HGT DNA region in D-Bacilli 2 and D-Bacilli 3.

4.5 Evolution in a chemostat

Unlike in batch growth, where organisms go through lag, logarithmic and stationary phases, bacterial culture in chemostat remain constantly in growth. In order to emulate another form of evolutionary environment we performed an HGT experiment using a chemostat. Two cultures of *Bacillus subtilis* 168-comp cultures were grown in parallel in small volume chemostats (100 mL). The bacteria were grown in the same medium as in the serial dilution experiment (0.8M NaCl LB) and with a doubling time of ~4 hours. Foreign DNA (combined pool of D-Bacilli, D-Bacteria and D-Archea – see table Table 1) was injected into one chemostat (termed 'DNA+') in several pulses roughly once a week. As a control, DNA-free buffer was injected into the second chemostat (termed 'DNA-'). The bacteria evolved for roughly ~700 generations, during that time the OD inside both chemostats was measured constantly (Figure 9-A). The constant OD measurement in the chemostat was crucial to detect when evolved species have appeared and fixated in the population, resulting in higher OD. Throughout the experiment both chemostats have had roughly the same OD except for generations ~360-420 where the OD measurement in the DNA- chemostat was higher than that of the DNA+ chemostat. In order to validate this small difference in OD, samples were taken out of the chemostat in order for their OD to be measured in another spectrophotometer and in that measurement no difference was measured between the two samples (data not shown). At generation ~400 samples were taken for growth analysis (Figure 9-B). It appears that both evolved strains had slightly decreased in maximal OD compared to the ancestor. Also, both strains had a slight increase in lag time; where the DNA- chemostat had higher increase than that of the DNA+ chemostat. These results might be explained by the fact that the strains evolved in a chemostat environment but their growth was checked in batch mode. It is possible that their fitness to batch growth was decreased although their fitness to chemostat growth had either increased or remained indifferent. After ~400 generations the experiment was temporarily stopped and the instruments inside the chemostat were re-calibrated. After which, the experiment continued with no apparent difference in OD between the strains in the two chemostats.

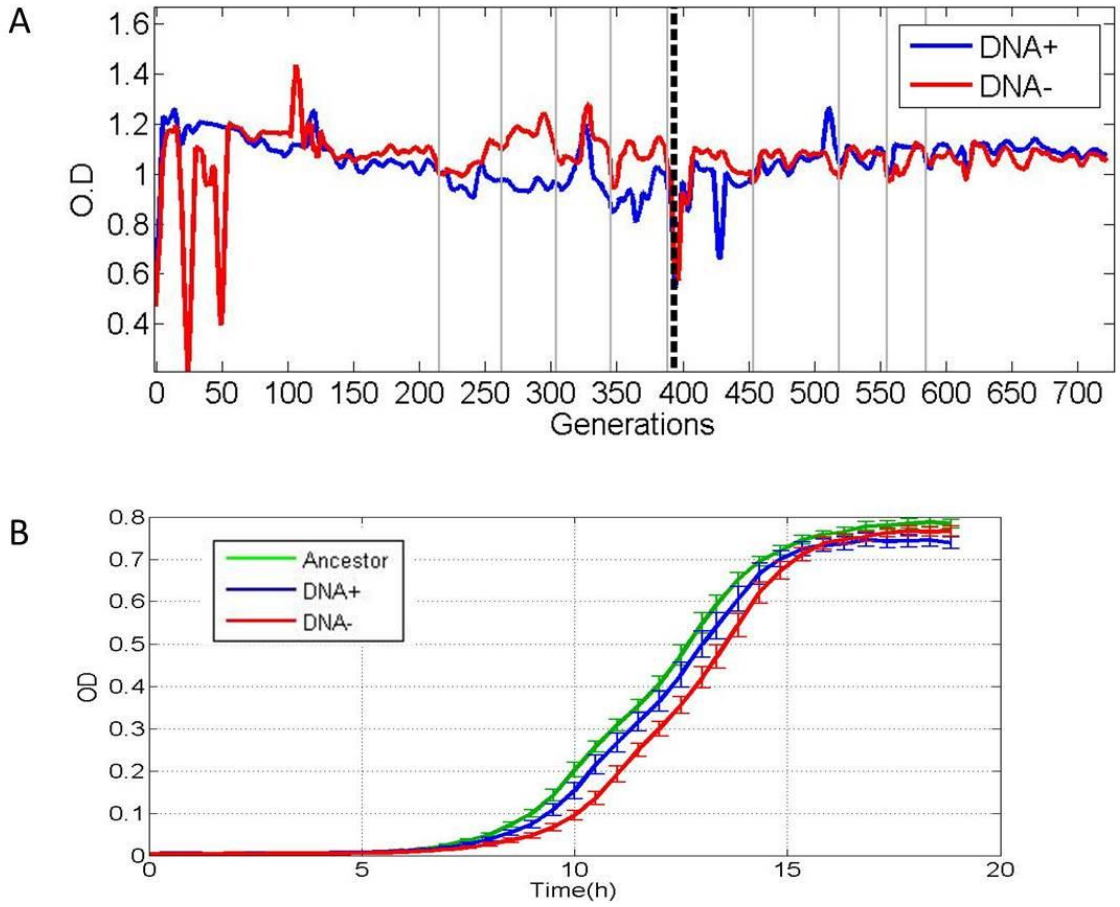

Figure 9 – Analysis for the chemostat HGT evolution

B. subtilis 168-comp was grown in parallel in two chemostats. In the first (termed 'DNA +'), DNA was supplement to the medium. In the second (termed 'DNA –') no DNA was supplemented to the media. A- OD measurements were taken inside the chemostat throughout the experiment and are shown here. The DNA+ chemostat is colored in blue. The DNA- chemostat is colored in red. Vertical grey lines represent time points in which DNA was supplemented. Dashed black line represents the time point in which the chemostat was decommissioned for routine maintenance; after which the experiment was continued. B – Growth experiments for the ancestor (colored green), and the evolved strains from the DNA+ chemostat (colored blue) and the DNA- chemostat (colored red). The evolved strains were from generation ~390.

5 Discussion

In this work we wanted to emulate lab HGT evolution so to compare evolution with and without the presence of foreign DNA. We focused on one bacterium, *Bacillus subtilis*, which is a competent species with the ability to integrate foreign DNA into its own genome. We decided to grow *Bacillus subtilis* on high salt since it may encounter this kind of environment in nature. Also, there are many anti-high salt mechanisms that are known and characterized. We hypostatized that *Bacillus subtilis* could acquire DNA segments that confer these mechanisms through HGT and thus gain improved fitness to high salt. We also hypothesized that by acquiring foreign DNA the bacillus will evolve faster than it would on evolution without the presence of foreign DNA.

The *Bacillus* was exposed to foreign DNA which was originated from a span of halophile organisms. We divided the foreign DNA organisms into three groups, based on phylogenetic distance from the recipient strain: extreme halophile Archea (D-Archea), non-bacilli halophile bacteria (D-Bacteria) and four supposedly non-halophile Bacilli that we evolved in the lab to have improved fitness to high salt (D-Bacilli). Bioinformatics studies have shown that horizontally transformed genes tend to arrive from phylogenetic close relatives[10,16,18]. Also, in *Bacillus subtilis* competence relies mainly on homologous recombination, further stressing the need for genome similarity between donor and recipient genome in our evolutionary setup. We thus speculated that in our evolutionary setup, genes from closely related organisms (such as the D-Bacilli group) will have a higher chance to be successfully horizontally transferred to the recipient strain. However, by definition the higher the distance between two organisms, the more different is their genome from one to another. Thus, one can postulate that more distant organisms will tend to have more non-homologous genes from the recipient strain. We thus hypothesized that the more distant an organism is from *Bacillus subtilis*, the higher chance for horizontally transferred gene to provide an innovative function that the recipient strain does not possess. We have planned our evolutionary setup in a way in which we spanned the different foreign DNA according to phylogenetic distance between donor and host. Thus, we created a tool to study *in vivo* the effects of phylogenetic distance on evolution with HGT. It is important to note that different donor organisms in our evolutionary setup might contain different mechanisms to cope with high salt, and that only some of them can be potentially transferred. For example, if one donor from our setup has a conserved machinery to cope only with extreme

salt concentrations only, it is possible that this machinery will have no effect in our low-medium salt concentrations.

We screened for a salt concentration that will reduce the fitness for the recipient strain. Also we required a salt concentration in which the strains from the D-Bacilli group are still viable because we wanted to adapt them to high salt. Out of the four supposedly non-halophile strains, we discovered that only one of the bacilli (strain *B. subtilis* RO-FF-1) showed no decrease in maximal OD in high salt concentrations. We thus suspected it to have halophile/halotolerant traits. Nonetheless we evolved it together with the other three bacilli to high salt so it could improve even more to this exact environment. After ~200 generations *B. subtilis* RO-FF-1showed no improvement on high salt as seen by comparing its growth curve to that of its ancestor. In fact it showed a bit of decrease in maximal OD. However the other three strains showed improvement to the high salt in lag time, maximal OD and/or maximal growth rate. All of the above strains showed no improvement to regular LB and so we could deduce that any fitness improvement that we detected was specific to the high saline condition and not to other aspects in the media (such as having an improved metabolism for the medium's carbon source). The species/strains evolution continued for a total of ~500 generation, after which, growth assays were measured again in high salt media (data not shown). In these measurements *B. subtilis* RO-FF-1 showed small indications for improvement on maximal OD but none of the other strains/species showed further increase in fitness.

In the main HGT experiment, we evolved *Bacillus subtilis* comp-168 in serial dilution in four evolution groups, exposing it to different pools of foreign DNA. After ~420 generations we performed several repetitions for each evolution line (except for D-Bacteria to which we performed only one repetition). Maximal growth rate improvement was shown in several biological repetitions in all evolutions groups. If the adaptation was due to HGT we would have suspected to see the same improvement trend in all biological repetitions of the same evolution line. However, that is not the case for the maximal growth improvement. So we assume that this fitness improvement was not due to HGT.

Analysis of maximal OD improvement revealed a trend in which the maximum OD of D-Bacilli's three biological repetitions showed a slight better improvement from that of the No-DNA, D-Bacteria and D-Archea. This trend, although weak, is prevalent in the three biological

repeats of D-Bacilli and thus we suspected that in this evolution group the recipient found different solutions to the stress other than what was found in the other evolution groups. Since the only difference between this evolution group to the others is the DNA source, we suspected that the origin for the difference is due to HGT. To verify this we sequenced all evolved strains through next generation sequencing and looked for HGT events in the genome. While analysis is not yet complete, we already found evidence for HGT event that occurred in a 4 Kb DNA segment in both D-Bacilli 2 and D- Bacilli 3. Curiously, both D- Bacilli 2 and D- Bacilli 3 showed increased maximal growth rate but D- Bacilli 1 did not. This raises the possibility that the HGT event found in this region is responsible for the improvement of the growth rate under high salt {do they grow better also under normal salt?} in the recipients. One interesting thing to do next is to check if these strains had gained improved growth rate on regular LB as well.

In this HGT region are four genes: RNA polymerase subunit, Acyl CoA that is responsible for lipid metabolism, a putative membrane iron-sulphur-binding reductase and cardiolipid synthase. As mentioned before, one of the high osmolality mechanisms is change in the membrane composition[24]. Furthermore, it was shown that elevated amounts of cardiolipids are found in the membrance of halotolerant organisms[24]. We can hypothesize that through this HGT event, the evolved strain gained improved fitness through a change in their membrane lipid composition. However, this hypothesis needs to be further examined by measuring membrane lipid composition of the evolved strain.

Because the sequencing analysis is still not complete, we still cannot say for sure which evolution lines evolved through HGT and which did not. If D-Bacilli indeed the only evolution line that improved via DNA acquisition, it supports the general notion that HGT has higher chances to occur between closely related organisms. However, it is possible that the donors in the other evolution lines had un-transferable solutions for the stress. Such as solutions that spread across the entire genome (like low hydrophobicity across protein repertoire) or anti salt mechanisms that cannot function in our evolutionary setup (give advantage only in higher salt concentrations). Importantly, since we performed the thorough growth assay only in generation 420 we could not assess the dynamics of the evolution. We did do growth analysis prior to generation 420(data not shown) but they always only one technical repetition; therefore we could

not conclude with confidence which evolution lines were better than the others in those time points.

It was shown before that experimental settings may affect the evolutionary path of the evolving organism. For example S. cerevisiae grown under constant heat stress have adapted by duplicating their chromosome[49]. However, when grown under gradual increase of the same stress there was no duplication. Therefore, we performed also HGT experiment in a different setup - a chemostat. There are many differences between chemostat to batch growth including, growth phase of the sample, oxygen availability, fixation time etc. We reasoned that these differences may influence the balance between solutions that could be acquired due to HGT to solutions that will arise due to random mutations. We grew bacteria culture in two chemostats for ~ 700 generations. In the first chemostat (DNA+) DNA pulses were injected roughly once a week while the other chemostat(DNA-) served as a control and was injected only with the vehicle. The foreign DNA composition differed from pulse to pulse but overall the bacteria culture was exposed to DNA originating from D-Archea, D-Bacteria and D-Bacilli combined. We monitored the OD inside the chemostats throughout the experiment and looked for a time point in which the OD will be dramatically increased so to indicate an improved fitness for the culture inside. Although at some time points the measurements in the chemostat indicated improvement of one culture over the other, re-measuring the OD outside the chemostat showed no change in OD between the bacterial cultures. Furthermore, a growth assay that was performed with culture samples of both chemostats in generation ~400 showed no significant growth improvement for either culture. We thus conclude that no adaptation has occurred in the chemostat strains in this evolution. However, it is important to note that the maximum OD measured for the ancestor in the growth assay of this experiment is higher than that of the ancestor in the serial dilution experiment. We speculate that during the 200 generations in which we calibrated the chemostat, the strain managed to evolve in the same manner as evolution has occurred in the serial dilution experiments. We conclude that during calibration of the chemostat, the 168-comp strain had evolved to have better yield. Later on, in the actual experiment where in one chemostat the bacteria were exposed to DNA, no further adaptation was captured. Why did evolution had occurred in serial dilution but not in the chemostat? One possible explanation is simply that in the chemostat there was not enough time for fixation. Another explanation might be the solutions found in the serial dilution experiment, could not improve fitness in the chemostat environment.

To conclude, in this thesis we aimed to capture in-lab evolution through HGT, to characterize it and compare it to evolution without HGT. We evolved non halophile bacilli to have improved fitness to salt and used their DNA, together with DNA from other halophile organisms, in evolutionary HGT experiments. *Bacillus subtilis* was grown on high salt either in serial dilution or in a chemostat and was exposed to foreign DNA. In the serial dilution evolution the recipient seemed to have improved maximum OD in all evolution groups. Furthermore, one specific evolution group (foreign DNA originated from pre-evolved Bacilli) showed better improvement than the other lines of evolution. This improvement was manifested both in maximal OD and in live count. Early sequencing analysis showed that two of the three repetitions in this evolution line had indeed at least one HGT event. This HGT event occurred in the same exact DNA region. While this HGT event is not enough to explain why all three biological repetitions had improved fitness, it might explain why only these two biological repetitions had increased maximal growth rate. These early findings show that the DNA was horizontally transferred only from the closest DNA donor groups. This finding concurs with our initial understanding that HGT evolution has higher chance to occur when the recipient and the donor strains have similar genomes, presumably due to reliance of homologous recombination.

6 Acknowledgments

I would like to thank my mentor, Yitzhak Pilpel, for his support and guidance – You taught me so much. I wish to thank all my lab members for many fruitful discussions. Special thanks go to Orna Dahan, & Ruthy Towers – for help on a daily basis and for Tammy Biniashvily for helping with the computational aspects in my work. I thank Ghil Jona from the Bacteriology and Genomic Repository units for helping in the design and for conducting the chemostat experiments. I wish to thank also Yoav Ram for his help with 'Curveball' software and for his always fruitful ideas. I would like to thank Dan Pollack, Gili Rozenberg and Ilana Kolodkin-Gal for their help with the microscopy and countless advices regarding *B. subtilis* culture. I would like to thank prof. Avigdor Eldar, Dr. Neta Altman, Prof. Uri Gophna and Prof. Aharon Oren for kindly providing me with the strains and species needed for my work. Lastly, I would like to acknowledge "The Nancy & Stephen Grand Israel National Center for Personalized Medicine". Shlomit Gilad and Shmulik Motola who worked on the sequencing, Barak Markus and Michael Gershovis who helped with the computational analysis of the sequencing.

7 Bibliography

1. Carvunis, A.-R. *et al.* Proto-genes and de novo gene birth. *Nature* 3–7 (2012). doi:10.1038/nature11184
2. Kaessmann, H. Origins, evolution, and phenotypic impact of new genes. **20,** 1313–1326 (2010).
3. McDaniel, L. D. *et al.* High frequency of horizontal gene transfer in the oceans. *Science* **330,** 50 (2010).
4. Liu, L. *et al.* The human microbiome: A hot spot of microbial horizontal gene transfer. *Genomics* **100,** 265–270 (2012).
5. Ochman, H., Lawrence, J. G. & Groisman, E. a. Lateral gene transfer and the nature of bacterial innovation. *Nature* **405,** 299–304 (2000).
6. Lurie-Weinberger, M. N., Peeri, M. & Gophna, U. Contribution of lateral gene transfer to the gene repertoire of a gut-adapted methanogen. *Genomics* **99,** 52–58 (2012).
7. Koonin, E. V, Makarova, K. S. & Aravind, L. H ORIZONTAL G ENE T RANSFER IN P ROKARYOTES : (2001).
8. Keeling, P. J. & Palmer, J. D. Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9,** 605–618 (2008).
9. Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16,** 472–482 (2015).
10. Popa, O. & Dagan, T. Trends and barriers to lateral gene transfer in prokaryotes. *Curr. Opin. Microbiol.* **14,** 615–23 (2011).
11. Dubey, G. P. & Ben-Yehuda, S. Intercellular nanotubes mediate bacterial communication. *Cell* **144,** 590–600 (2011).
12. Garcia-ojalvo, J. & Elowitz, M. B. Tunability and Noise Dependence in Differentiation Dynamics. *Science (80-.).* **315,** 1716–1719 (2007).
13. 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).
14. Sorek, R. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science (80-.).* **89,** (2007).
15. Gophna, U. & Ofran, Y. Lateral acquisition of genes is affected by the friendliness of their products. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 343–8 (2011).
16. Tuller, T. *et al.* Association between translation efficiency and horizontal gene transfer within microbial communities. *Nucleic Acids Res.* **39,** 4743–4755 (2011).
17. Wellner, A., Lurie, M. N. & Gophna, U. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.* **8,** R156 (2007).
18. Boto, L. Horizontal gene transfer in evolution: facts and challenges. *Proc. Biol. Sci.* **277,** 819–827 (2010).
19. Amorós-Moya, D., Bedhomme, S., Hermann, M. & Bravo, I. G. Evolution in regulatory regions rapidly compensates the cost of nonoptimal codon usage. *Mol. Biol. Evol.* **27,**

2141–51 (2010).

20. Jakobs, M. & Meinhardt, F. What renders Bacilli genetically competent? A gaze beyond the model organism. *Appl. Microbiol. Biotechnol.* (2014). doi:10.1007/s00253-014-6316-0

21. Hamoen, L. W. Controlling competence in Bacillus subtilis: shared use of regulators. *Microbiology* **149,** 9–17 (2003).

22. Garcia-Vallve, S., Guzman, E., Montero, M. a. & Romeu, a. HGT-DB: A database of putative horizontally transferred genes in prokaryotic complete genomes. *Nucleic Acids Res.* **31,** 187–189 (2003).

23. Hahne, H. *et al.* A comprehensive proteomics and transcriptomics analysis of Bacillus subtilis salt stress adaptation. *J. Bacteriol.* **192,** 870–82 (2010).

24. Russell, N. J. Adaptive modifications in membranes of halotolerant and halophilic microorganisms. *J. Bioenerg. Biomembr.* **21,** 93–113 (1989).

25. Ollivier, B., Caumette, P., Garcia, J. L. & Mah, R. a. Anaerobic bacteria from hypersaline environments. *Microbiol. Rev.* **58,** 27–38 (1994).

26. Saum, S. H. & Müller, V. Regulation of osmoadaptation in the moderate halophile Halobacillus halophilus: chloride, glutamate and switching osmolyte strategies. *Saline Systems* **4,** 4 (2008).

27. Lentzen, G. & Schwarz, T. Extremolytes: natural compounds from extremophiles for versatile applications. *Appl. Microbiol. Biotechnol.* **72,** 623–634 (2006).

28. Vargas, C. *et al.* Unravelling the adaptation responses to osmotic and temperature stress in Chromohalobacter salexigens, a bacterium with broad salinity tolerance. *Saline Systems* **4,** 14 (2008).

29. Galinski, E. a. & Trüper, H. G. Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol. Rev.* **15,** 95–108 (1994).

30. Eisenberg H, Mevarech M, Z. G. Life in unusual environments: progress in understanding the structure and function of enzymes from extreme halophilic bacteria.

31. Fukuchi, S., Yoshimune, K., Wakayama, M., Moriguchi, M. & Nishikawa, K. Unique Amino Acid Composition of Proteins in Halophilic Bacteria. *J. Mol. Biol.* **327,** 347–357 (2003).

32. Lanyi, J. K. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bacteriol. Rev.* **38,** 272–90 (1974).

33. Kennedy, S. P. Understanding the Adaptation of Halobacterium Species NRC-1 to Its Extreme Environment through Computational Analysis of Its Genome Sequence. *Genome Res.* **11,** 1641–1650 (2001).

34. Soppa, J. From genomes to function: haloarchaea as model organisms. *Microbiology* **152,** 585–90 (2006).

35. Stefanic, P. *et al.* The quorum sensing diversity within and between ecotypes of Bacillus subtilis. *Environ. Microbiol.* **14,** 1378–1389 (2012).

36. Tomlinson, G. a, Jahnke, L. L. & Hochstein, L. I. Halobacterium denitrificans sp. nov., an extremely halophilic denitrifying bacterium. *Int. J. Syst. Bacteriol.* **36,** 66–70 (1986).

37. Rodriguez-Valera, F., Ruiz-Berraquero, F. & Ramos-Cormenzana, a. SHORT COMMUNICATION Isolation of Extremely Halophilic Bacteria Able to Grow in Defined Inorganic Media with Single Carbon Sources. *Microbiology* **119,** 535–538 (1980).

38. Charlebois, R. L., Lam, W. L., Cline, S. W. & Doolittle, W. F. Characterization of pHV2 from Halobacterium volcanii and its use in demonstrating transformation of an archaebacterium. *Proc. Natl. Acad. Sci. U. S. A.* **84,** 8530–8534 (1987).

39. Nuttall, S. *et al.* The Halohandbook. 1–144 (2009).

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

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

42. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9,** 357–359 (2012).

43. Bolotin, A. & Borchert, S. The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. **390,** (1997).

44. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29,** 24–26 (2011).

45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

46. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27,** 2987–2993 (2011).

47. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19,** 455–477 (2012).

48. Dawit Kidane, Silvia Ayora, Joann Sweasy, Peter L. Graumann, J. C. A. The cell pole: The site of cross talk between the DNA uptake and genetic recombination machinery. **47,** 531–555 (2013).

49. Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 21010–5 (2012).