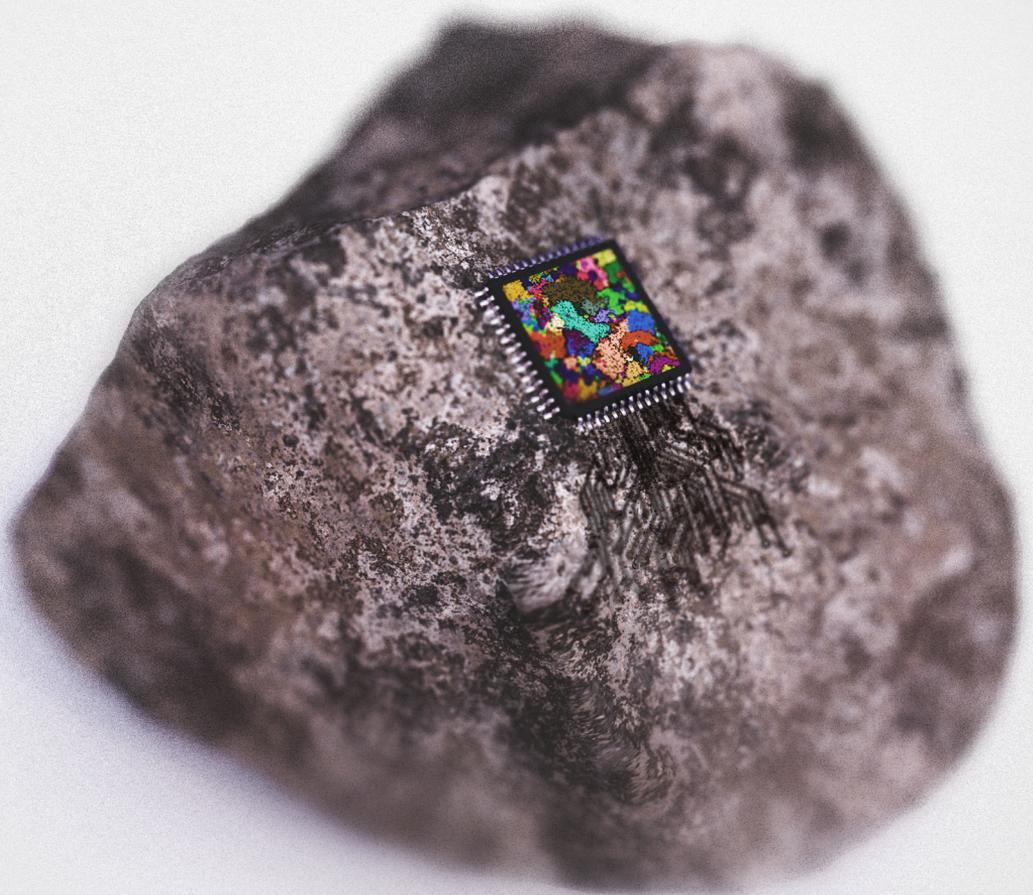


Microbial Evolution on a Virtual Grain of Sand



Bram van Dijk

ISBN: 978-94-6375-979-3

Cover: “*In silico on silico*” by Bram van Dijk, 3D render provided by Team3D

Print: Ridderprint | www.ridderprint.nl

Copyright © 2020 B. van Dijk. All rights reserved. No part of this publication may be reproduced in any form, by print or photo print, microfilm or any other means, without written permission by the publishers.

Microbial Evolution on a Virtual Grain of Sand

Microbiële Evolutie op een Virtuele Zandkorrel

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 16 september 2020 des middags te 2.30 uur

door

Bram van Dijk

geboren op 22 mei 1988
te Apeldoorn

Promotor:

Prof. dr. P. Hogeweg

The work presented in this thesis was financially supported by the European Commission 7th Framework Programme (FPFP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427).

Members of the dissertation committee:

Prof. dr. Martin Ackermann
ETH Zurich and Eawag, Switzerland

Prof. dr. Guillaume Beslon
Université de Lyon, France

Prof. dr. Santiago F. Elena
Universitat de València, Spain

Prof. dr. Paul G. Higgs
McMaster University, Canada

Prof. dr. Paul B. Rainey
MPI for Evolutionary Biology, Germany

“HE TOOK HIS TIME IN THE SUN,
HAD A DREAM TO UNDERSTAND,
A SINGLE GRAIN OF SAND.”

— TUOMAS HOLOPAINEN

Preface

Rather late to the game, I was 20 years old when I first read a proper textbook on the topic of biology. Where my rather basic biological education in secondary school was little more than a four-year-long anatomy lesson, this book discussed broad questions ranging from the stability of entire ecosystems, down to the finer details of microbial life. When I started studying biology a year later, it was understanding the latter microscale that inspired me most. However, it quickly became clear to me that the entire concept of “understanding something” required a lot of unpacking. One course, in particular, made it crystal clear to me that just knowing and naming all the components of a system is not enough. Instead, biological systems have a lot of moving parts and self-organised structures across different scales, all of which have been selected through Darwinian evolution through natural selection. Clearly, to grasp what is happening at the microscale, the process of evolution cannot simply be tossed aside as something that has happened in the past. Instead, how microbes have evolved, and how microbes are still evolving in the soils and lakes around us, is key to understanding them.

In this thesis titled ‘Microbial Evolution on a Virtual Grain of Sand’, I report on the research we have done that shines light on life at the microscale. With a variety of models, we tackle questions like: How can microbial diversity emerge, and how is it maintained in the long term? How is diversity impacted by Horizontal Gene Transfer? How predictable is evolution? How can microbes avoid accumulating deleterious mutations? We address these questions, and more, all by designing computational models in which bacterial ecosystems live *in silico*, on a Virtual Grain of Sand.

Contents

1	Introduction	15
1.1	Microbial diversity on a single grain of sand	15
1.2	Life at the microscale	18
1.2.1	What does a microbe need <i>to do</i> ?	18
1.2.2	Environmental interplay and microbial interactions	19
1.3	The flexible gene content of microorganisms	21
1.3.1	Horizontal gene transfer and gene-level selection	21
1.3.2	The processes shaping microbial gene content	23
1.4	Microbial evolution: lessons learned thus far	25
1.4.1	Emergent diversity in the lab	25
1.4.2	Interference competition and the maintenance of diversity	27
1.4.3	The impact of horizontal gene transfer in a constant environment	28
1.5	This thesis	31
1.5.1	Chapter introductions	31
2	Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives	35
2.1	Introduction	36
2.2	Results	38
2.2.1	Evolving Virtual Microbe “wild types”	38
2.2.2	<i>In silico</i> serial transfer evolution experiment	40
2.2.3	All wild types evolve to anticipate the serial transfer protocol	40
2.2.4	Evolution toward a growth-yield trade-off	42
2.2.5	Anticipating as a collective	43
2.2.6	Individual anticipation by tuning and trimming the gene regulatory network	46
2.2.7	Individual and collective solutions have similar macro-level observables	47
2.3	Discussion	49
2.3.1	Who is anticipating what?	51
2.3.2	Moving towards an eco-evolutionary understanding	52
2.4	Conclusions	52
2.5	Methods	54
2.5.1	Finding generic patterns of evolution	54
2.5.2	Model overview	54
2.5.3	Experimental setup	60
2.6	Supplementary Material	62

3	Contingent evolution of alternative metabolic network topologies determines whether cross-feeding evolves	71
3.1	Introduction	72
3.2	Results	74
3.2.1	Diverse metabolic strategies evolve on a single resource	75
3.2.2	The evolution of cross-feeding is not explained by protein cost	78
3.2.3	Trade-offs emerge during the evolution of metabolic networks	80
3.2.4	The evolution of cross-feeding requires spatial structure	81
3.2.5	Metabolic strategies are an evolutionary contingency	83
3.3	Discussion	85
3.4	Methods	87
3.4.1	Classification of metabolic types and construction of Muller diagrams	87
3.4.2	Lineage tracking with renewing markers	87
3.4.3	Analysis of within-population diversity (PCA of single-cell proteomics)	87
3.4.4	Testing metabolic dependency	87
3.4.5	Analysis of between-population diversity (PCA of metagenomics)	88
3.4.6	Full model description	88
3.4.7	Code availability	92
3.5	Supplementary Material	93
4	In silico gene-level evolution explains microbial population diversity through differential gene mobility	107
4.1	Introduction	108
4.2	Methods	110
4.2.1	General Description	110
4.2.2	Population Dynamics	110
4.2.3	Gene-Level Dynamics	111
4.2.4	Visualising Fitness Landscapes	112
4.2.5	Strain Diversity and Phylogenetic Reconstruction	112
4.2.6	Parameters used in the model	113
4.3	Results	114
4.3.1	Evolution of differential gene mobility	114
4.3.2	Differential gene mobility is required for phenotypic diversity	116
4.3.3	Evolution of the fitness landscape	116
4.3.4	Core resistance repertoires and accessory toxins	118
4.3.5	Strains are cohesive units of interaction	119
4.3.6	Diversity is robust to selective sweeps	120
4.3.7	Differential gene mobility evolves only in a closed ecosystem	122
4.4	Discussion	122
4.4.1	Closed ecosystems with high diversity	122
4.4.2	Rarely beneficial genes evolve high mobility	123
4.4.3	Who wants to take up eDNA?	124
4.5	Supplementary Material	126
5	Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements	133
5.1	Introduction	134
5.2	Results	136
5.2.1	Slightly beneficial genes fall into distinct gene classes	137

5.2.2	HGT is an evolutionarily stable strategy, but cannot evolve to ‘rescue’ rescuable genes	139
5.2.3	Spatial structure hinders the maintenance of genes, making HGT adaptive for a wider range of genes	140
5.2.4	Bacteria evolve DNA uptake for rescuable genes only in a spatially structured population	142
5.2.5	HGT is evolutionarily maintained in the presence of harmful SGEs	145
5.3	Discussion	148
5.3.1	Horizontal Gene Transfer: rescue or catastrophe?	150
5.4	Methods	151
5.4.1	General overview	151
5.4.2	ODE model(s)	151
5.4.3	Individual-based model	152
5.4.4	Software used	155
5.5	Supplementary Material	157
6	The evolution of alternative genome architectures through Horizontal Gene Transfer and Gene Duplications	177
6.1	Introduction	178
6.2	Results	180
6.2.1	Exploring the effects of gene duplications and additive HGT in Virtual Microbes	180
6.2.2	Gene duplication and gene transfer yield distinct genome size dynamics	181
6.2.3	Initially brittle genomes become robust to null-mutations in the long term	182
6.2.4	The evolution of expression rates underpins different genome size dynamics	184
6.2.5	A minimal model of the evolution of gene copy number and expression rate	184
6.3	Discussion	187
6.3.1	Mutations as drivers of genome size evolution	187
6.4	Conclusion	188
6.5	Methods	189
6.5.1	Overview of the Virtual Microbes model	189
6.6	Supplementary Material	192
7	Discussion	195
7.1	Chapter reviews	195
7.2	Models and unforeseen insights	197
7.2.1	Building multilevel models from the bottom up	198
7.2.2	Biological basic principles	198
7.2.3	The unique exploratory potential of complex models	199
7.2.4	Structures, processes, and levels of selection	200
7.3	Outlook	200
7.3.1	Embracing HGT in evolutionary theory	201
7.3.2	What is the impact of nonrandom mutations?	202
7.3.3	Is evolution predictable?	203
7.3.4	Concluding remarks	203

Appendix	205
References	206
Glossary	223
Samenvatting in het Nederlands	229
Dankwoord (acknowledgements)	237
Curriculum vitæ	241
List of publications	243

“WE ARE HERE AND IT IS NOW.
THE WAY I SEE IT IS, AFTER THAT,
EVERYTHING TENDS TOWARDS GUESSWORK.”

—Terry Pratchett

Chapter 1

Introduction

1.1 Microbial diversity on a single grain of sand

Across all scales, biology is bursting with diversity. While the number of different species visible to the naked eye is already baffling, the diversity of life that is too small to directly observe is truly astonishing. We know this, not only by eagerly looking through our microscopes, but also by taking samples from a wide variety of environments, and studying the occurrence of unique nucleic acid sequences (DNA and RNA) found within those samples. Indeed, a great assortment of microorganisms are found in lakes (Jones et al., 1998; Sorokin et al., 2014), oceans (Sogin et al., 2006; Rusch et al., 2007; Pedrós-Alió et al., 2015), soils (Torsvik and Øvreås, 2002; Roesch et al., 2007), but also habitats closer to home such as the human skin (Costello et al., 2009), and inside shower heads (Feazel et al., 2009). Diverse microorganisms are even being observed in extremely cold or hot environments (Steven et al., 2006; Coman et al., 2013), thriving at extremes of -20°C (Staley and Gosink, 1999) to above the boiling point of water (Blöchl et al., 1997). But in terms of sheer diversity within single samples, especially terrestrial soils are remarkable (Ranjard and Richaume, 2001; Nannipieri et al., 2003). When the automated sequencing of DNA became more commonplace in the 90s, early estimates suggested that a single gram of soil contains thousands of different microbial species (Torsvik, 1994; Ritz, 1998). More recently it became clear that even within a single genus, an individual grain of soil can be a display of tens or hundreds of different strains of *Streptomyces* (Vetsigian et al., 2011), forming complex interaction networks where strains promote or inhibit one another (**Figure 1.1**). How can this astonishing diversity exist on the scale of a single grain of sand?

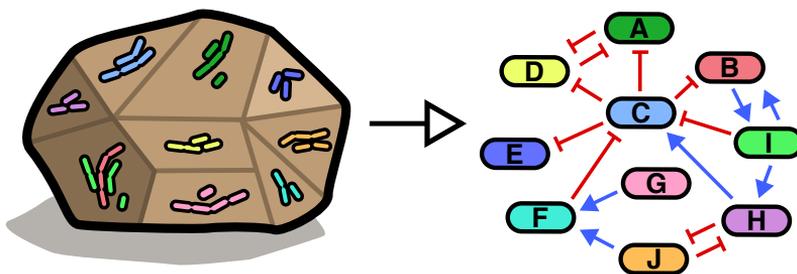


Figure 1.1: Even an individual grain of soil contains many different microbial strains. Vetsigian et al. (2011) have shown that individual grains of soil, hand-picked with the tip of a needle, contain many different strains of *Streptomyces*, either promoting or inhibiting the growth rates of one another.

All the endless forms of life on this planet are ultimately generated by Darwinian evolution by means of natural selection (Darwin, 1859). The microbial world is no different. Thus, in order to shine light on the diversity of microorganisms, we have to improve our understanding of evolution. Conversely, much of our understanding of evolution is being advanced by studying microorganisms, as their evolution can be observed in a matter of days or weeks. Besides experimental evolution, other approaches like bioinformatic analysis of genomes and metagenomes provide treasures of empirical data on what has happened with microbes in the past, and what is happening with them right now. The goal of this thesis is to build upon these empirical approaches and genomic data by designing and analysing mechanistic evolutionary models. By building multilevel biological systems from the BOTTOM UP, we can study the ECO-EVOLUTIONARY DYNAMICS of microbes in what could be called a “virtual laboratory” (**Figure 1.2**), and further our understanding of what these systems look like when the wheels are actually turning.

Before we plunge into these virtual laboratories however, I first aim to lay the groundwork for microbial ecology and evolution, giving an idea of which details are important to consider in modelling these systems. Briefly, we aim to understand the evolution of microbes as the multi-levelled process that it is, with the chemistry of life and individual cells at the lower level, and communities and ecosystems at the higher levels. Without explicitly modelling higher level structures such as ecological interactions, we instead focus on how they emerge from, and feedback on, the lower levels. To design such a multi-levelled system however, we first need to consider the important details of life at the microscale, *i.e.* what it actually is that a microbe is doing. Where much theoretical work has already been developed to understand microbial evolution, the actual intricacies of microbes have remained somewhat in the background. This first chapter gives a basic introduction of the intricacies we think are important to understand microbial evolution, and which we will attempt to incorporate into our models while taking care that our models remain simple enough. First, in **Section 1.2**, we will provide an overview of the context in which microbes evolve, discussing what a microbe must do to survive and reproduce, and discussing the importance of environmental interplay through the uptake and

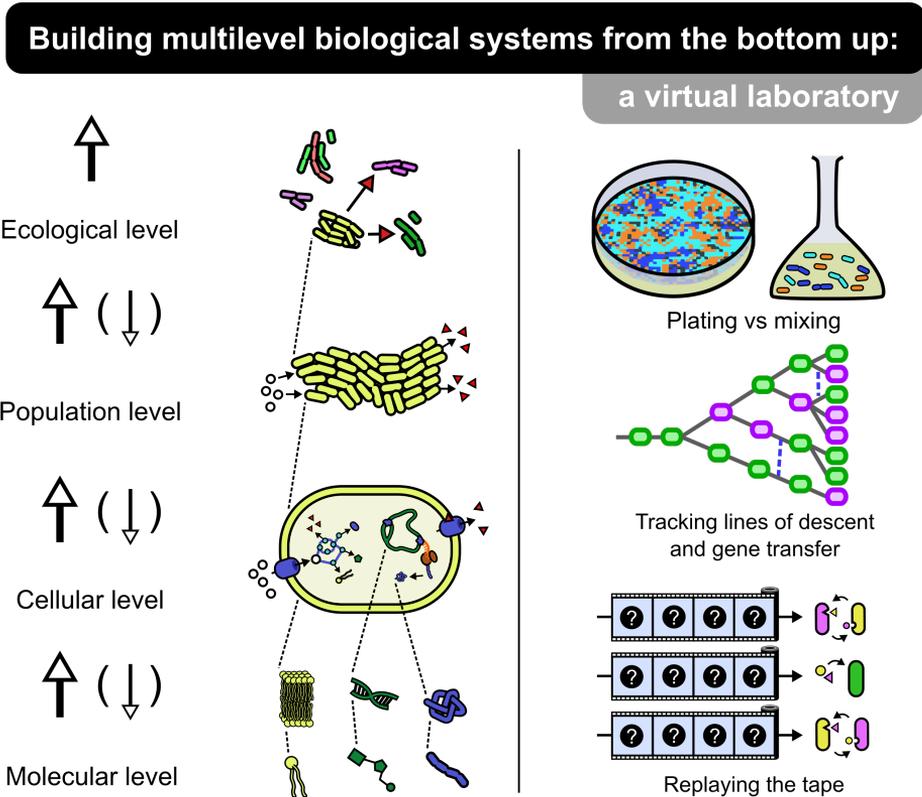


Figure 1.2: By building multilevel biological systems from the bottom up, we can study the eco-evolutionary dynamics of microbes in a “virtual laboratory”.

exchange of metabolites. Next, **Section 1.3** will discuss how microorganisms evolve primarily through rapid changes in their gene content, and which processes are important in shaping microbial genomes. **Section 1.4** will discuss important lessons learned about microbial evolution, reviewing how diversity has repeatedly emerged in experimental evolution in the lab, how hostile interactions appear to play an important role in maintaining diversity, and discuss some insights of earlier studies on the role of horizontal gene transfer in microbial evolution. Finally **Section 1.5** introduces the research questions and modelling approaches used throughout the different chapters.

1.2 Life at the microscale

Like all life on Earth, microbes need to survive and reproduce to persist. These processes are often summarised in concepts like instantaneous growth rates, rates of reproduction, or competitive fitness. Although these concepts have kickstarted our thinking about population dynamics and evolution, we should realise that these numbers depend strongly on ever-changing ecological contexts. For example, investing all resources into the production of one's offspring is only a viable strategy if those offspring will also have access to sufficient resources. If resources are scarce, investing resources into survival strategies may be more important. While this is a relatively simple example, Darwin himself already realised that the ecological contexts in which species evolve can be “infinitesimally complex” (Darwin, 1859). Therefore, if we want to know what exactly underlies concepts like fitness and growth rates, we need to better understand the principles that drive them at bottom. Instead of thinking about microbial evolution as an optimisation process, it is useful to consider what a microbe needs *to do* in order to survive and reproduce (*i.e.* the TODO-principle of Hogeweg and Hesper, 1991).

1.2.1 What does a microbe need *to do*?

Where large organisms like plants and animals may take years or decades to procreate their entire complex body plans, the reproduction of single-celled microorganisms can be observed within the hour. Although we can describe the growth of a population with a simple model of logarithmic growth, this ignores underlying processes by simply assuming cell division happens. In other words, such a simple model fails to address *how* it is that microbes grow and divide. Indeed, from the perspective of the microbe, the process of growth and division entails a lot of complex chemistry, because the required BUILDING BLOCKS to make a cell (lipids, nucleic acids, amino acids, *etc.*) are not directly available in the environment. Instead, natural environments like soils contain many precursor metabolites (Young and Crawford, 2004; Torsvik and Øvreås, 2002), all of which still need to be processed. To drive all this chemistry, a microbe must express a variety of different enzymatic proteins that catalyse reactions to import the available metabolites into the cell and convert them into the essential building blocks. These building block metabolites can then be invested to grow the cell and its contents, including the genome, new proteins, the membrane, and the cell wall. The remaining by-products can be recycled or discarded to prevent them from accumulating to toxic concentrations. Many metabolic processes furthermore require energy to drive them. Finally, all of these processes need to be tuned to one another, or growth may be unsustainable (*e.g.* when the DNA is not replicated if the cell divides). Even with a very simple “toy chemistry” as the one depicted in **Figure 1.2**, we can imagine how the goal of producing building blocks can be achieved in a variety of different ways. Indeed, endless metabolic strategies are observed in nature, differing not only in which pathways are used, but also in the fluxes of metabolites through these pathways within and between organisms. Thus, what seems like a simple enough task when described from the top down, entails a lot of orchestrated chemistry at bottom.

Much knowledge on the above-mentioned metabolic processes, gene expression, and DNA replication, are established by studying fast-growing model organisms

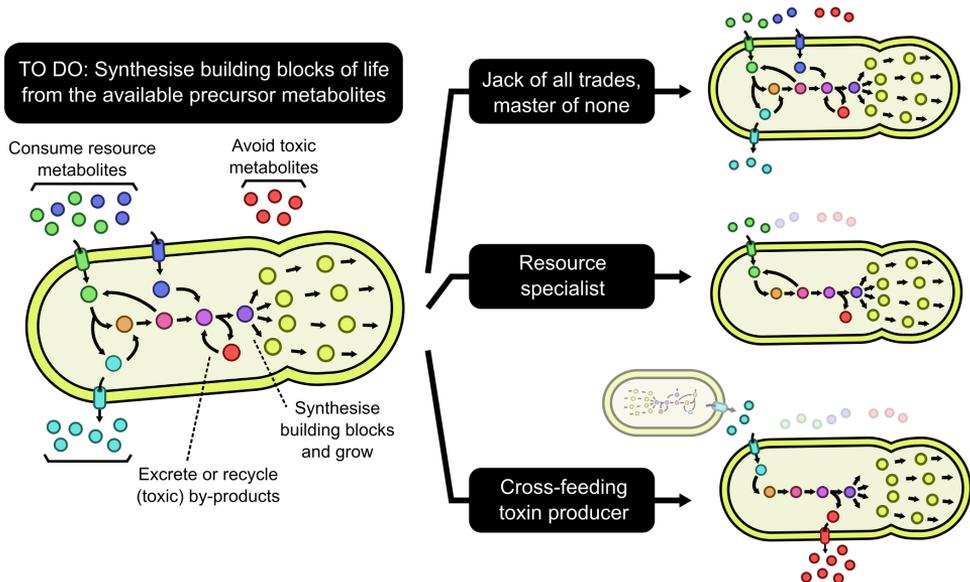


Figure 1.3: Even with a very simple “toy chemistry”, the goal of producing building blocks may be achieved in a variety of different ways.

like *Escherichia coli*. However, as pointed out by Bergkessel et al. (2016), the exponential growth rates of these organisms would allow them to grow to the mass of the Earth in less than two days, which clearly does not occur. It is therefore safe to conclude that microbes in nature are often facing the challenge to simply survive, perhaps only growing exponentially in short bursts when resources are abundant. Moreover, even with abundant resources, microbes still need to maintain HOMEOSTASIS, making sure not to consume toxic levels of resources and prevent the accumulation of by-products, and maintaining a steady supply of energy. Indeed, microbes in nature display a wide variety of survival mechanisms, ensuring for example regulated growth arrest (Finkel, 2006) or dormancy (Higgins and Dworkin, 2012) during phases of starvation. Even when resources are abundant, bacterial populations are known to contain persister cells that do not grow (Shah et al., 2006), potentially functioning to promote long-term survival (Lewis, 2007). These dormant or persister phenotypes may be the result of stochastic switching between gene expression patterns (Schultz et al., 2009), but might also be driven by purely physiological effects (van Heerden et al., 2014). Although many of the details of the survival strategies of microbes remain unresolved, it is clear that the success of a microbial lineage is not only determined by its rate of growth.

1.2.2 Environmental interplay and microbial interactions

Microbial growth and metabolism do not happen in isolation, but in the context of other cells. Even when a cell is initially alone, clonal reproduction will quickly result in many competitors that are also competing for resources. Moreover, as resources in nature are often confined to certain habitats (*e.g.* marine food particles, grains of soil, the gut lumen), cell densities within these relatively small habitats can be very

high. On the one hand, these high cell densities intensify the competition between microbes, resulting in an impressive range of chemical and mechanical strategies to kill off competitors (Granato et al., 2019). On the other hand, these high cell densities also open the door to cooperative behaviour that would be inviable for single cells, such as metabolic CROSS-FEEDING (Germerodt et al., 2016; D'Souza et al., 2018) and the collective degradation of large macromolecules by secreting enzymes into the environment (Allison, 2005; Datta et al., 2016; Hehemann et al., 2016). Thus, microbial cells strongly influence one another through the uptake of nutrients, the release of metabolites, and the secretion of enzymes into the shared environment.

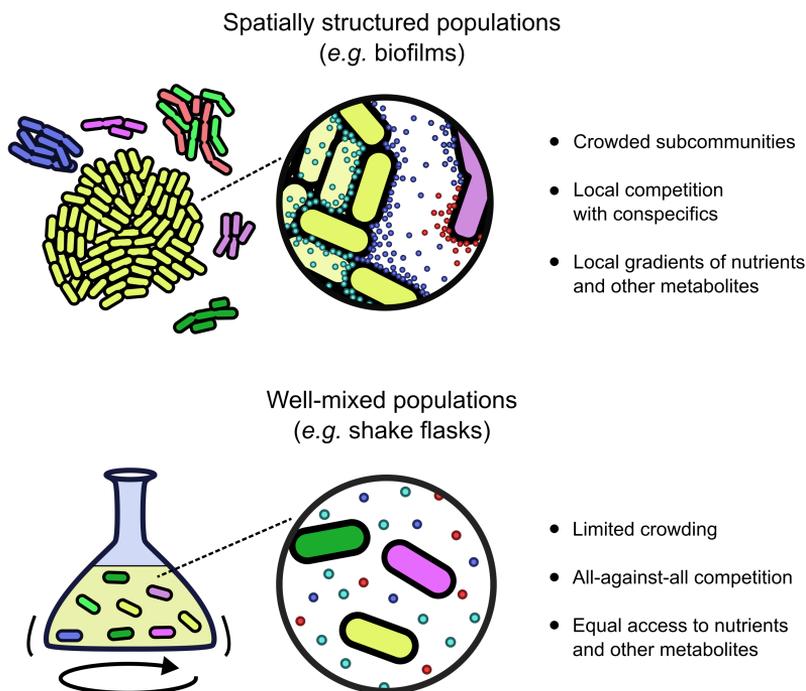


Figure 1.4: Microbes grow, survive, metabolise resources, and excrete by-products or toxins in the context of other cells. This is especially relevant in spatially structured populations, where cells compete for limited reproduction space and dynamically shape their local environment.

When considering the environments that microbial cells share, it is important to take into account the fact that the motility of microorganisms is often limited. Although most microbial species either actively move towards resources (*i.e.* chemotaxis) or passively drift towards them (*e.g.* during the planktonic phase of species growing on oceanic food particles), they typically stay put when resources are available. With minimal motility, microbial populations will self-organise into spatially separated communities, strongly influencing the conditions in their local environment (**Figure 1.2.2**). Under these conditions, competition between and within these spatially structured communities may depend more on the collective properties of a cohort than on any given individual (van Gestel and Kolter, 2019). These dynamics are

in stark contrast with many experimental protocols and theoretical models that use an assumption of a well-mixed population of cells (*e.g.* shaken cultures and simple mathematical models), conditions which are furthermore nearly impossible to realise (Herrerías-Azcué et al., 2018). Indeed, taking these facts to heart has greatly improved our understanding of microbial biodiversity (Rainey and Travisano, 1998; Czárán et al., 2002; Reichenbach et al., 2007; van Vliet et al., 2019), the evolutionary outcome of ANTAGONISTIC INTERACTIONS (Pagie and Hogeweg, 1999; Vetsigian, 2017; Granato et al., 2019), MUTUALISTIC INTERACTIONS (Colizzi and Hogeweg, 2016; Gorter et al., 2019), and the strong resilience microbial populations can have with respect to nutrient switches and antibiotics (Nadell et al., 2016; Dal Co et al., 2019a; Van Vliet et al., 2019). Throughout the chapters in this thesis we will often compare populations under well-mixed conditions with spatially structured populations like biofilms, in order to reshape our intuitions on which outcomes are attainable, and maintainable, by evolution.

1.3 The flexible gene content of microorganisms

With a clear picture of what microbes have to do to survive and reproduce, and in which possible contexts this may happen, let us next consider how they evolve. Although microbes adapt to their environment by means of the same process as other lifeforms on Earth (*i.e.* through mutation and selection), their evolution is in many ways different from that of larger organisms like plants and animals. Rather than through (gradual) sequence evolution, bacteria and other prokaryotes appear to adapt primarily through rapid changes in their gene content (Kolstø, 1997; Snel et al., 2002; Puigbò et al., 2014), quickly losing genes on the one hand (Van Ham et al., 2003; Barrick et al., 2009; Kuo and Ochman, 2009; Morris et al., 2012; Albalat and Cañestro, 2016), and gaining new genes on the other hand (Puigbò et al., 2014; Soucy et al., 2015). What entails such a rapid turn-over of genes?

1.3.1 Horizontal gene transfer and gene-level selection

Considering that bacteria and other microorganisms evolve primarily through rapidly changing their gene content, how then do these changes happen? Of course, losing genes is a relatively “easy” mutation which can happen through small inactivating point mutations, replication slippage, intrachromosomal crossover, or through the loss of chromosomes or plasmids. Gene gain on the other hand is far less likely to continuously occur, as back mutations or the *de novo* discovery of genes are rare events. Nevertheless, gene gain is very common in microorganisms, occurring primarily through a process called HORIZONTAL GENE TRANSFER (HGT). Where VERTICAL GENE TRANSFER refers to how genes are inherited from parent to offspring, HGT entails all other processes where genes are transferred from one genome to another. HGT can happen through many different mechanisms (see **Figure 1.5**), *e.g.* by bacteriophages (TRANSDUCTION), the uptake of extracellular DNA from the environment (TRANSFORMATION), or the transfer of plasmids between neighbouring cells (CONJUGATION). More recent studies have revealed that certain bacteria also produce phage-like particles called GENE TRANSFER AGENTS (GTAs) (Stanton, 2007; Lang et al., 2012), whereas other bacteria produce vesicles derived from the

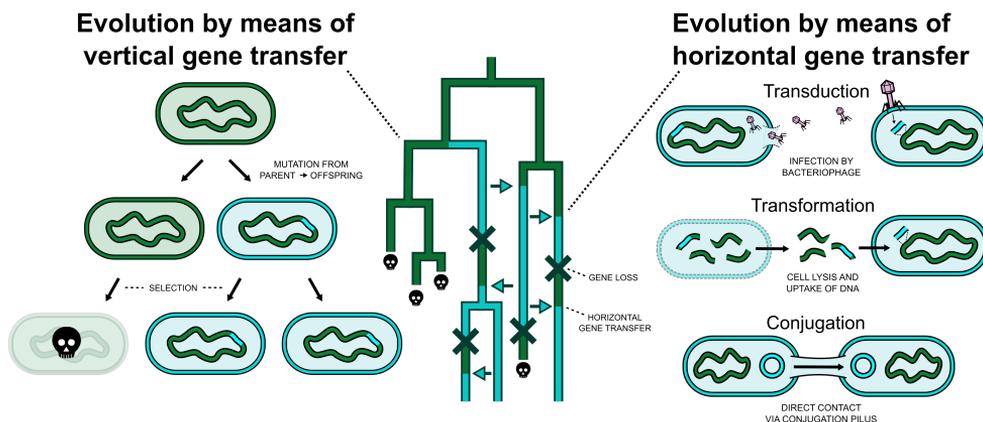


Figure 1.5: Evolution by means of vertical and horizontal gene transfer Where vertical gene transfer refers to the (regular) process where genes are inherited from parent to offspring, a variety of mechanisms allow horizontal gene transfer to occur between individuals that do not have a parent-offspring relationship (see Thomas and Nielsen, 2005).

outer membrane that appear to play an important role in transferring genes across great distances (Biller et al., 2014; Soler et al., 2015; Tashiro et al., 2017).

Not only is HGT an important driver of microbial evolution (Ochman et al., 2000; Doolittle and Zhaxybayeva, 2009), it also appears to completely override smaller events like point mutations. For example, staggeringly high rates of HGT have been recently reported for *Escherichia coli* (Frazão et al., 2019; Sakoparnig et al., 2019), *Streptomyces* (Tidjani et al., 2019), and for prokaryotes in general (Puigbò et al., 2014). Instead of having to wait for rare beneficial point mutations, HGT enables microbes to quickly adapt to new niches by picking up “tried-and-true” genes from a mobile gene pool (Jain et al., 2003; Wiedenbeck and Cohan, 2011; Casacuberta and González, 2013; Mell and Redfield, 2014; Niehus et al., 2015; Lopatkin et al., 2016). Furthermore, the GENE MOBILITY in this pool is not identical for all genes, where some genes transfer much more rapidly than others by being associated with MOBILE GENETIC ELEMENTS (MGEs) like transposons, plasmids, or bacteriophages. In general, important housekeeping genes appear to be vertically inherited, while genes that carry more specialised functions transfer horizontally (Hacker and Carniel, 2001; Norman et al., 2009; Nogueira et al., 2009; Rankin et al., 2011).

While taking up genes may be very beneficial when adapting to a new environment, HGT also comes with many burdens for microbes. For example, MGEs may disrupt essential housekeeping genes or replicate selfishly, at which point they are referred to as SELFISH GENETIC ELEMENTS (SGEs). As it is generally accepted that SGEs require HGT to persist (Bergstrom et al., 2000; Lili et al., 2007; Slater et al., 2008), the host cell may be better off avoiding HGT. Clearly, when HGT is beneficial, and for whom it is beneficial, may change depending on the ecological conditions and the precise mechanism of HGT. For example, while transduction by bacteriophages is generally outside of the host’s control, the domestication of phages in the form of the aforementioned GTAs shows that there are plausible exceptions to this rule. *Vice versa*, while the transfer of plasmids through conjugation is

primarily determined by the host, self-promoting plasmids (Harrison et al., 2014) and “addictive” toxin-antitoxin systems (Leplae et al., 2011) highlight the importance of GENE-LEVEL SELECTION in determining microbial gene content (Werren et al., 1988; Werren, 2011). In short, there are many mechanisms for HGT, which do not only differ in their physiological details, but also in terms of who or what is selected for, opening up a wide continuum of possible outcomes in the co-evolution of genes and their hosts.

Clearly, HGT is a major player in bacterial evolution. Because classical evolutionary studies have exclusively looked at evolution through vertical inheritance, the theory to describe the key principles of microbial genome evolution is therefore still lacking. Given the range of advantages and disadvantages of HGT (Vogan and Higgs, 2011), is it merely something bacteria have to cope with, or are there conditions where it benefits them? How can the conflict between the host and SGEs be resolved? What are the short- and long-term consequences of introducing novel genes into a microbial gene pool?

1.3.2 The processes shaping microbial gene content

As mentioned in the previous section, the flexible gene content of microbes is not only shaped by HGT, but also by gene loss. To take this into account, we have to consider which genes can be (temporarily) discarded, and which genes must always be retained. Genes that are not adaptive in the current environment may either neutrally get lost, or may adaptively get lost if the gene or its function has a certain cost. The latter adaptive gene loss can furthermore be empowered by the fact that other community members may still perform the costly function, a process that was coined the BLACK QUEEN HYPOTHESIS (Morris et al., 2012). By means of this process, even genes with functions that appear vital may get lost from a microbial genome. Finally, beneficial genes can also be lost due to the stochastic or deterministic processes brought forth by ongoing mutations (Muller, 1964; Eigen, 1971). All these reductive genome processes can potentially be countered by HGT, allowing species to retrieve lost genes from a shared ‘mobile gene pool’. What does such a mobile gene pool look like, for any one microbial species?

The gene pool of microorganisms is often categorised by sampling and sequencing multiple genomes, and counting how often different genes occur. By doing so, the total collection of genes, the PANGENOME, can be partitioned into CORE GENES and ACCESSORY GENES. When sampling multiple strains from the same microbial species, core genes are the genes that are found in all strains, while accessory genes are present in only a subset of strains. Interestingly, the ratio between these two types of genes can differ strongly across species (**Figure 1.6**). For some species, we find what is referred to as an OPEN PANGENOME, meaning that for every new genome that is sampled the number of accessory genes increases while the core genome decreases (Welch et al., 2002; Lefébure and Stanhope, 2007; Touchon et al., 2009; Kim et al., 2015). On the other hand, other species appear to have a CLOSED PANGENOME, with only a limited number of accessory genes. As a general rule, it appears that free-living species have more open pangenomes, while species with a more restricted lifestyle (*e.g.* intracellular parasites or obligate pathogens) have closed pangenomes (Lukjancenko et al., 2010; McInerney et al., 2017). Besides

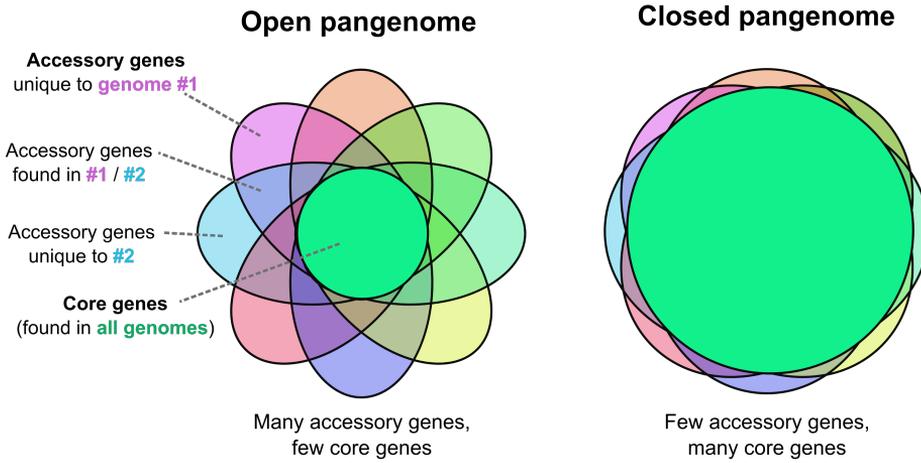


Figure 1.6: When sampling multiple microbial genomes from the same strain or species, the total collection of genes ranges from “open” to “closed” pangenomes.

these general patterns, the exact processes shaping pangenomes are not entirely established. It is however becoming increasingly clear that HGT, *i.e.* the ability of microbes to take up genes from their neighbours or from the environment, plays a crucial role.

1.4 Microbial evolution: lessons learned thus far

Where the previous sections have given an overview of the context in which microbes evolve, and the mechanisms by which they evolve, I will next address the questions with which we started: how does microbial diversity emerge, and how is it maintained in the long term? I will give a brief overview of experimental work on microbial evolution and diversity, focusing primarily on those studies that are relevant to the work presented in this thesis. I will then discuss some theoretical work that has studied the impact of gene loss and HGT (*i.e.* rapid gene content evolution) on the maintenance of diversity, and on the maintenance of genetic information in general.

1.4.1 Emergent diversity in the lab

Due to their short generation times and large population sizes, microorganisms are ideal to study in the lab (Dykhuizen, 1990). They are easily propagated, for example with a simple serial transfer protocol using a single carbon source. Using this protocol, the long-term evolution experiment (LTEE) has subjected *E. coli* to over 10,000 transfers on a minimal glucose medium (Lenski et al., 1991). While this exploratory experiment has resulted in many insights and search images, one initially surprising result was the evolution of novel ecological interactions. Due to the continuous cycles of feast and starvation, a BALANCED POLYMORPHISM has been shown to evolve consisting of one lineage that grows fast on the primary resource, and a second lineage that grows on the metabolic byproducts of the first (Rozen and Lenski, 2000; Rozen et al., 2009) (see **Figure 1.7A**). More recently, a metagenomic analysis has revealed that more polymorphisms like this may be present across the populations of the LTEE (Good et al., 2017). Where classical evolutionary theory predicts that no two species can stably exist on a single resource due to the COMPETITIVE EXCLUSION principle (Gause, 1934; Hardin, 1960; MacArthur, 1970; Tilman, 1977, 1982; Sommer, 1985), metabolic dependencies like cross-feeding are now widely understood to make room for microbial diversity (Treves et al., 1998; D'Souza and Kost, 2016; San Roman and Wagner, 2018; D'Souza et al., 2018).

Another important factor that may drive the emergence or maintenance of ecological diversity is spatial structure (Korona et al., 1994; Bell, 1997). This effect was explicitly investigated in a study by Rainey and Travisano (1998), where the only difference between two sets of evolving *Pseudomonas fluorescens* populations was the level of mixing. Where the mixed populations showed no genetic or phenotypic differentiation, the spatially heterogeneous environment gave rise to self-adhering mutants that colonize the liquid-surface interface (having access to both nutrients and oxygen) in a matter of days (see **Figure 1.7B**). This mutant however does not entirely replace its ancestor, as high densities of this newly emerged morphotype cause the emergence of non-sticky “cheaters” that cause the colony to sink. These dynamics result in a fluctuating, yet stable, coexistence of morphotypes, each having their own distinct ecological niche. Similar forms of coexistence have more recently been observed in *Saccharomyces cerevisiae* (Frenkel et al., 2015), similarly driven by mechanisms of crowding and competition avoidance (Dieckmann and Doebeli, 1999; Holt, 2001). Subsequent evolution may even cause the members of such communities to adapt to one another, making the community more stable and productive (Hansen et al., 2007). Clearly, ecological opportunities are quickly

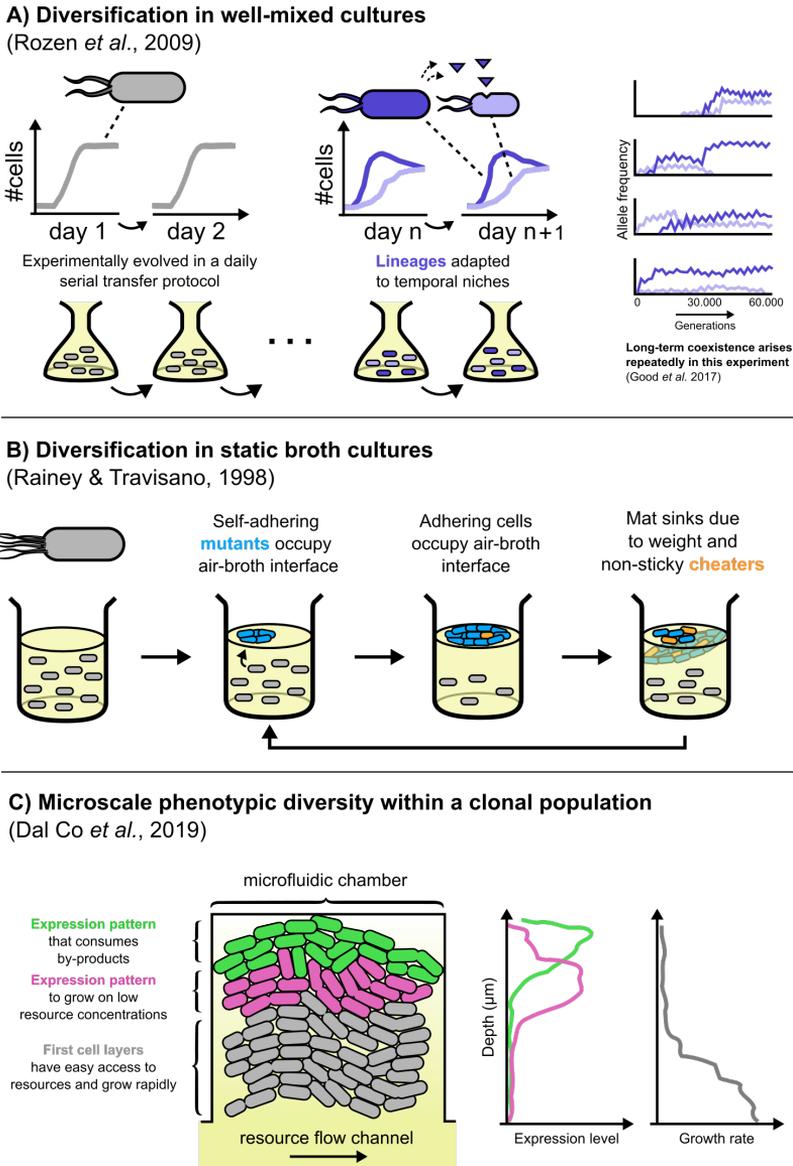


Figure 1.7: Evolutionary and ecological experiments in the lab have shown that, in the absence of ecology, ecology still happens.

seized upon by microorganisms, and we cannot reasonably exclude these ecological opportunities when trying to understand evolutionary processes and community function.

While the examples of microbial diversity mentioned above are both observed on the macroscale, having distinct colony morphologies on agar plates, progress in experimental techniques now also allows us to observe microbial diversity at the microscale. By using microfluidic chambers that are open to resources on one side,

small biofilm populations can be realised in which we can observe the behaviour of individual cells (Dal Co et al., 2019a). These studies have revealed that even within biofilms of genotypically identical *E. coli* cells, different ecological opportunities are exploited (see **Figure 1.7C**). As cells closest to the chamber opening consume the resource, cells further away express genes to adapt to lower resource concentrations, and cells in the back of the chamber feed primarily on the remaining by-products (Dal Co et al., 2019b). While this differentiation is caused by the fact that resources only come in on one side, a follow-up study has shown that two neighbouring cells even impact one another directly, where cross-feeding genotypes grow quicker when directly next to their interaction partner (Dal Co et al., 2020). In summary, whether it is at the macroscale or at the microscale, through genotypic variation or phenotypic plasticity, all of these experimental studies show that in the absence of ecology, ecology still happens.

1.4.2 Interference competition and the maintenance of diversity

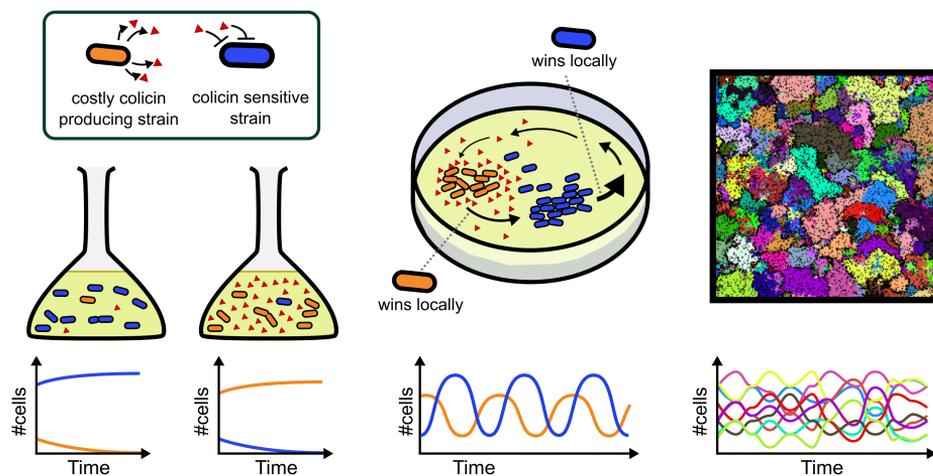


Figure 1.8: The high diversity of toxins and antibiotics in nature can be better understood in terms of spatially structured populations.

So far, we have extensively discussed how microbes living in dense populations compete for resources (*i.e.* nutrients, space, light, oxygen), which is known as **EXPLOITATIVE COMPETITION**. However, microorganisms are notorious for imposing direct, negative effects on another in what is called **INTERFERENCE COMPETITION** (see Granato et al., 2019). With respect to this so-called “microbial warfare”, spatial structure is known to play an important role in maintaining diversity. One very intuitive example of the importance of spatial structure is found in studies related to the production of colicins, toxic substances produced by *E. coli*. These colicins are surprisingly diverse (Riley, 1993; James et al., 1996; Dykes and Hastings, 1997), resulting in many combinations of toxin production and / or sensitivity. Under well-mixed conditions (*e.g.* a shaken flask), a single mutant producing these costly colicins cannot overtake a sensitive population of cells, as the effect

of the colicin toxicity would be diluted. *Vice versa*, a sensitive mutant cannot invade a population of colicin-producing cells (Chao and Levin, 1981). Thus, under well-mixed conditions, no diversity of colicins is expected (see **Figure 1.8**). In spatially structured environments however, the toxins can have high concentrations locally (favouring the producer), and low concentrations elsewhere (favouring the sensitive cells). Indeed, a spatially heterogeneous environment does not only make room for these two genotypes (Riley, 1998), but can explain the existence of very large sets of different colicins even in relatively small populations (Pagie and Hogeweg, 1999). Thus, not only is interference competition very common in microbial ecosystems, it also appears to be one of the most important factors in maintaining bacterial diversity.

Clearly, spatially structured environments make room for many ecological opportunities, both in terms of exploitative competition (*e.g.* by avoiding direct competition) and interference competition (*e.g.* by producing antibiotics). Moreover, as microbes are continuously reshaping their environment, these selection pressures may continuously change depending on neighbouring cells. In light of earlier discussed rapid changes in gene content, this may result in so-called RED QUEEN DYNAMICS, a term that is borrowed from Alice Through the Looking-Glass (1871):

“IT TAKES ALL THE RUNNING YOU CAN DO, TO KEEP IN THE SAME PLACE.”

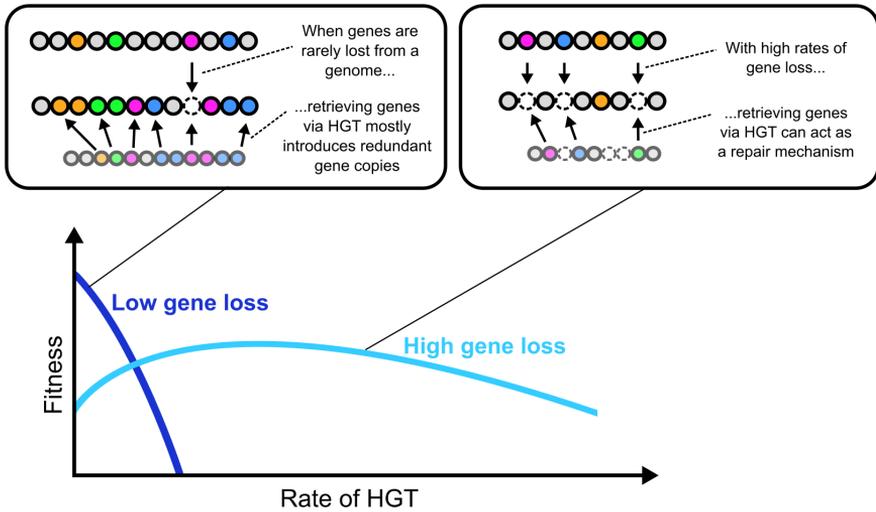
—Lewis Carroll

Red Queen dynamics are often considered for organisms that evolve through rapid sequence evolution, *e.g.* viruses which continuously escape their host’s immune system. In the absence of such rapid sequence evolution however, the continuous changes in gene content discussed in **Section 1.3** may enable similar processes in microbial evolution. Many microbial strains are for example engaged in a continuous arms race to best their competitors, rapidly changing with respect to antibiotic production and resistance. These dynamics result in many phenotypes which often only poorly correlate with their phylogeny (Bruce et al., 2017; Kinkel et al., 2014; Vos and Velicer, 2009; Cordero et al., 2012b). Moreover, these rapid dynamics do not only change the *tempo* of evolutionary processes, but can also alter the *outcome*, for example by attaining levels of diversity that cannot be reached by introducing one mutant at a time (Vetsigian, 2017). How exactly the rapid turnover of genes that is mediated by HGT interlocks with these Red Queen dynamics, and how it may make or break diversity, is still poorly understood.

1.4.3 The impact of horizontal gene transfer in a constant environment

HGT can clearly be beneficial when adapting to changing environments, for example by improving the ability to adapt to alternating selection pressures (Mozhayskiy and Tagkopoulos, 2012). Besides its role in adapting to new selection pressures however, recent modelling work has shown that HGT can also be beneficial in the absence of changing environments. In 2011, Vogan and Higgs have shown that HGT can have a positive impact on fitness (**Figure 1.9a**) when gene loss is high. In this model, bacteria can retrieve genes from the gene pool by interacting with

A) HGT improves fitness at high rates of gene loss
(Vogan & Higgs, 2011)



B) HGT helps to alleviate Muller's ratchet
(Takeuchi *et al.*, 2014)

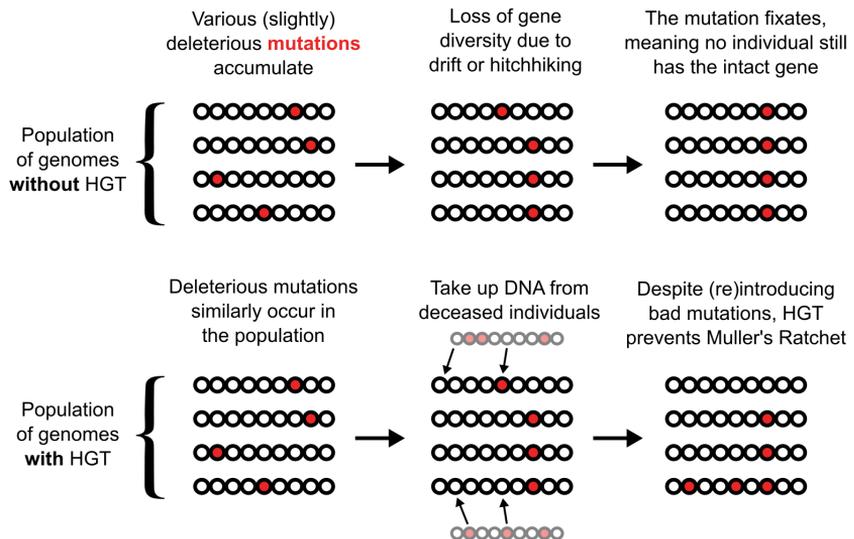


Figure 1.9: In light of ongoing mutations, HGT can have a positive effects on bacterial fitness by preserving genetic information.

recently deceased individuals (Vogan and Higgs, 2011). When gene loss occurs rarely however, this process merely introduces many redundant copies into the bacterial genome, negatively impacting fitness. In a similar study, Takeuchi *et al.* have shown that HGT can act to rescue bacteria from accumulating slightly deleterious mutations (Takeuchi *et al.*, 2014), therewith preventing what is known as MULLER’S RATCHET (**Figure 1.9b**) (Muller, 1964). Although this is analogous to how recombination impacts the evolution of sexual organisms (Lynch *et al.*, 1995; Schultz and Lynch, 1997; Lynch *et al.*, 2016; Vos *et al.*, 2019), it is important to note that HGT is quite distinct from sex. First, HGT in the models shown in **Figure 1.9** is implemented through interactions with deceased individuals, which are necessarily less fit. This has sometimes been referred to as “sex with dead cells” (Redfield, 1988). Besides this dubious nomenclature however, HGT also allows genes to spread independently of the host’s cell cycle, which may have important consequences for evolution (Shapiro *et al.*, 2012; Takeuchi *et al.*, 2015).

Finally, it is interesting to consider how HGT is a mutation that is itself very much the product of natural selection. Not only do microbes vary in their rates of HGT (Ochman *et al.*, 2000; Vos *et al.*, 2015), but different genes also transfer with different rates (Rankin *et al.*, 2011; Cordero *et al.*, 2012b; Hehemann *et al.*, 2016). Moreover, what exactly is selected for (and to what end) changes depending on the conditions and mechanisms that entail the transfer of genes (Redfield, 2001; Thomas and Nielsen, 2005; Vogan and Higgs, 2011). Within the last two decades, it has become clear that it is important to take into account how evolution can *itself* evolve (Beslon *et al.*, 2018), *e.g.* by shaping mutation rates, the phenotypic effects of mutations, mutational hotspots, or genome flexibility. The impact of the evolution of evolution (EVOEVO) has for example been demonstrated in transposon-mediated mutations (Crombach and Hogeweg, 2007), gene regulatory network architectures (Crombach and Hogeweg, 2008; Draghi and Wagner, 2009), and mutator strains (Matic, 2016; Rutten *et al.*, 2019). How this new concept of EvoEvo is impacted by HGT, which is of course a very particular mutation, remains as of yet poorly established.

1.5 This thesis

This introductory chapter has given an overview of the intricacies we think are important for microbial evolution. Clearly, we know a lot about the key players that are important for microbial ecosystems (cells, genes, mutations, metabolites, spatial structure, gene transfer, toxins, *etc.*), but it is still hard to imagine what these systems actually look like when the wheels are turning. This is where we rely on modelling, where the goal is to gain insights into a system by simplifying reality. However, how can we make sure we study biologically interesting simplifications? For example: given the goal of producing the required building blocks to grow, how can we understand which strategies evolution will adopt? Will a generalist strategy using all chemical reactions evolve, or will microbes specialise on specific pathways? Will bacterial populations become more diverse when subjected to HGT, or will HGT have a homogenising effect?

When asking questions like the ones above, it is important to think about what type of model is useful. When we already have an intuition on possible outcomes, how do we make sure that we do not unintentionally force the model to confirm our expectations? Throughout this thesis, we aim to prevent this hurdle by implementing biological models from the bottom up, without having too many expectations of what the model *should* do. By doing so, we have allowed our models to surprise us, the importance of which I revisit in the final chapter of this thesis. For now, I will briefly introduce the different topics of each chapter in order, including how we chose to explore the complexity of microbial evolution.

1.5.1 Chapter introductions

This thesis roughly consists of two parts. In the first part (**Chapter 2** and **3**), we take to heart all that a microbe has to do to survive and reproduce (*i.e.* what is described in **Section 1.2**) by studying the evolution of Virtual Microbes. This modelling platform — that was developed during the course of my PhD — is precisely the type of “virtual laboratory” that was mentioned at the beginning of this introductory chapter. By implementing a simple artificial biochemistry at the lower level, higher-level phenomena such as fitness, homeostasis, and community structure emerge. These higher-level phenomena moreover feed back on the lower level, producing a multilevel, eco-evolutionary system with which we can experiment. In the second part (**Chapter 4 – 6**), we focus our attention on HGT and the resulting interplay of genes, their microbial hosts, and selfish genetic elements. For this part, we implement models varying in complexity, ranging from simple ORDINARY DIFFERENTIAL EQUATIONS (ODEs) to multilevel INDIVIDUAL-BASED MODELS (IBMs).

Chapter 2 investigates how Virtual Microbes adapt to conditions similar to those found in the laboratory. For this, we first evolve complex “wild types” in a naturally fluctuating environment, which we then transfer to a much more predictable and clean serial transfer protocol like the one from **Figure 1.7A**. Because the wild types vary strongly in their gene expression patterns and gene regulatory mechanisms, we herewith address the question how future evolution is constrained or potentiated by prior adaptations. Thus, besides comparing our observations to studies like the long-term evolution experiment (Lenski et al., 1991), we also investigate the

predictability of evolution. In **Chapter 3** we continue with Virtual Microbes, but now study their evolution in a more complex, spatially structured environment. We repeatedly evolve microbes that can explore the possibilities of a toy chemistry, all the while reshaping their local environment by metabolite uptake and exchange. Within this toy chemistry, that contains many alternative pathways, we investigate the evolution of diversity, the predictability of evolution, and the stability of the evolved communities.

The second part of this thesis (**Chapter 4 – 6**) investigates the impact of HGT on microbial evolution, focusing on its impact on phenotypic and phylogenetic diversity, and its role in maintaining genetic information. We do so by explicitly studying both a gene-oriented and a cell-oriented point of view, to disentangle what is beneficial for the host, and what is beneficial for the genes. In **Chapter 4**, we first focus on the perspective of the genes by allowing individual genes to evolve their own rate of transfer. We study this in a microbial population undergoing antagonistic interactions via toxin production and resistance genes. While cells lose these genes with a fixed rate, they can be recovered by taking up extracellular DNA from the environment. As genes can evolve their individual rate of integrating into the host genome, this allows the host to retain more or less mobile genes, but it also opens the door for the emergence of selfish genetic elements. We study how the resulting interplay of gene- and cell-level selection shapes which genes become mobile, which genes minimise their mobility, and how differential gene mobility impacts the evolution of the pangenome. Instead of assuming that cells simply take up DNA from their environment at a fixed rate, **Chapter 5** investigates when bacterial cells may benefit from taking up foreign DNA despite the accompanying costs and risks. We first use a simple model to investigate the conditions under which bacteria can, in light of gene loss, evolve the ability to retrieve slightly beneficial genes from the gene pool. We next study how spatially structured populations like biofilms help or hinder the spread of genes, and how deleterious selfish elements can coexist with their cellular hosts. Finally, to further explore the effects of HGT, **Chapter 6** studies the potential impact of HGT on the evolution of Virtual Microbes. We do so by contrasting the ability of HGT to recover genes from the gene pool with the alternative of evolving backup genes by means of gene duplication. While gene gain in prokaryotes has been shown to happen primarily through HGT, the evolution of eukaryotes is strongly impacted by gene duplication. We investigate how these different mechanisms of gaining genes impact fitness, genome dynamics, and the maintenance of genetic information.

Finally, **Chapter 7** grants an overview of the results found in each chapter, discusses the unique exploratory potential of bottom-up models, and speculates on future directions and open questions that are still in need of answering.

“ANTICIPATING PAIN WAS LIKE ENDURING IT TWICE.
WHY NOT ANTICIPATE PLEASURE INSTEAD?”

—Robin Hobb

Chapter 2

Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives

BRAM VAN DIJK¹, JEROEN MEIJER¹, THOMAS D CUYPERS¹ AND PAULIEN HOGEWEG¹
1 UTRECHT UNIVERSITY, UTRECHT, THE NETHERLANDS
BMC EVOLUTIONARY BIOLOGY VOLUME 19, ISSUE 1, NOVEMBER 2019, PAGE 201

Abstract

Background: Experimental evolution of microbes often involves a serial transfer protocol, where microbes are repeatedly diluted by transfer to a fresh medium, starting a new growth cycle. This has revealed that evolution can be remarkably reproducible, where microbes show parallel adaptations both on the level of the phenotype as well as the genotype. However, these studies also reveal a strong potential for divergent evolution, leading to diversity both between and within replicate populations. We here study how *in silico* evolved Virtual Microbe “wild types” (WTs) adapt to a serial transfer protocol to investigate generic evolutionary adaptations, and how these adaptations can be manifested by a variety of different mechanisms.

Results: We show that all WTs evolve to anticipate the regularity of the serial transfer protocol by adopting a fine-tuned balance of growth and survival. This anticipation is done by evolving either a high yield mode, or a high growth rate mode. We find that both modes of anticipation can be achieved by individual lineages and by collectives of microbes. Moreover, these different outcomes can be achieved with or without regulation, although the individual-based anticipation without regulation is less well adapted in the high growth rate mode.

Conclusions: All our *in silico* WTs evolve to trust the hand that feeds by evolving to anticipate the periodicity of a serial transfer protocol, but can do so by evolving two distinct growth strategies. Furthermore, both these growth strategies can be accomplished by gene regulation, a variety of different polymorphisms, and combinations thereof. Our work reveals that, even under controlled conditions like those in the lab, it may not be possible to predict individual evolutionary trajectories, but repeated experiments may well result in only a limited number of possible outcomes.

2.1 Introduction

In order to see microbial evolution in action, we often rely on experimental evolution under controlled laboratory conditions. The Long-term Evolution Experiment (LTEE) (Lenski et al., 1991) and similar shorter studies (Dettman et al., 2007; Paterson et al., 2010) have, for example, evolved many generations of microbes using a serial transfer protocol, where microbes are repeatedly diluted and transferred to a fresh medium to start a new growth cycle. Conceptually, if we learn to understand how microbes adapt to such a resource cycle, we might one day be able to predict evolution in the lab and — ideally — also in nature. Indeed, a lot of evolution in the lab seems remarkably reproducible, where microbes show parallel adaptations both on the level of the phenotype as well as the genotype (Dunham et al., 2002; Cooper et al., 2003; Pelosi et al., 2006; Philippe et al., 2007; Hindré et al., 2012; Laan et al., 2015; Salverda et al., 2017; Consuegra et al., 2017). However, there also seems to be strong potential for divergent evolution, leading to diversity both between and within replicate populations (Rozen and Lenski, 2000; Rozen et al., 2009; Plucain et al., 2014). Diversification events within populations in serial transfer regularly show cross-feeding interactions (Treves et al., 1998; Rozen and Lenski, 2000; Rozen et al., 2005, 2009; Turner et al., 2015), where strains emerge that grow on metabolic by-products. These cross-feeding interactions are increasingly well understood with the help of metabolic modeling and digital evolution (Großkopf et al., 2016; Rocabert et al., 2017). A recent metagenomics study has revealed even more coexisting lineages in the LTEE than were previously reported (Good et al., 2017). It is however not yet clear whether all these polymorphisms are the result of uni-directional cross-feeding interactions, or if other mechanisms could drive coexistence in a simple experiment such as a serial transfer protocol. Furthermore, whether or not the diversified communities experience fundamentally different selection pressures and growth dynamics as a collective, is still an open question.

Prior to being subjected to lab conditions, the microbes used in the aforementioned experimental studies have all had a long evolutionary history in natural environments, experiencing harshly fluctuating and — more often than not — unfavourable conditions. While a serial transfer protocol at a first glance selects mostly for higher growth rates when resources are abundant (*i.e.* during the log phase), there is also selection to survive when resources are depleted and the population no longer grows (*i.e.* during the stationary phase). In fact, given the unpredictable conditions found in nature, some of the ancestors of *Escherichia coli* might have survived precisely because they diverted resources *away* from growth. Indeed, *E. coli* does exactly this during the stationary phase by means of the stringent response, regulating up to one third of all genes during starvation (Sarubbi et al., 1989). This response lowers the growth rate, but promotes efficiency and survival (*i.e.* a higher yield). While most microbes have ways to deal with starvation, the physiology of growth arrest varies a lot across different microbes, and especially display great variation in how long they can persist in the absence of nutrients (for an excellent review, see (Bergkessel et al., 2016)). After prolonged starvation, many species of bacteria go through even more physiological changes, such as the GASP response (Finkel, 2006), persistence (Balaban et al., 2004), and sporulation (Piggot and Hilbert, 2004). Bacteria have also been shown to employ bet-hedging strategies with respect to these physiological changes (Veening et al., 2008b; Solopova et al., 2014; Veening et al., 2008a), which

could help to adapt to unexpected environmental changes. Finally, it has been shown that microorganisms can even adjust to *expected* environmental changes, anticipating regularity in environmental changes (Balaban et al., 2004; Mitchell et al., 2009; Tagkopoulos et al., 2008), which usually entails using predictive cues from the environment. All these responses, as well as other features that organisms have acquired during their evolutionary history (gene clustering, gene regulatory network architecture, metabolic regulation, *etc.*), might strongly influence the adaptation and reproducibility we observe in the lab today.

What do we expect when a complex, “pre-evolved” organism adapts to serial transfer protocol in the lab, given how clean and extremely regular these conditions are? We here use Virtual Microbes in order to firstly mimic natural evolution, acquiring Virtual “wild types” (WTs), which we then expose to a serial transfer protocol (see methods). We do so in order to obtain a fresh perspective on *which generic adaptations* might appear in spite of evolutionary contingencies, and *how* these adaptations are achieved. We find that all the WTs — which are both genotypically and phenotypically diverse — evolve to anticipate the regularity of the serial transfer protocol by timing their growth rate, yield, and survival, to accurately fit the daily cycle. Yet, we observe many alternative paths in terms of growth dynamics trajectories, gene regulation, and diversification. Whereas some WTs adapt by means of clever gene regulation, others diverge into multiple strains with their own temporal niche, and others simply time their resource consumption as to not over-exploit the medium. In short, our WTs all recognized and exploited the regularity of the serial transfer protocol, having learned to trust the hand that feeds, but they solve this challenge by a variety of different mechanisms.

2.2 Results

In this study we use Virtual Microbes, a model of the eco-evolutionary dynamics of microbes (see **Figure 2.1** and methods). In short, the Virtual Microbe model is unsupervised, meaning that it aims to combine relevant biological structures (genes, genomes, metabolism, mutations, ecology, etc.), allowing us to study the emergent properties of fitness and evolution in an undirected system. In other words, by not explicitly defining what the model *should* do, we take a serendipitous approach to study microbial evolution. By modelling evolution with many degrees of freedom, the process can be seen as a “inventive” generator of attainable (and maintainable) adaptations (Takeuchi and Hogeweg, 2009), and can furthermore serve to debug false intuitions (Lehman et al., 2018). Our main objective in this study is to elucidate *generic adaptations* of evolution in a serial transfer protocol, to investigate how this is achieved, and to what extent it is constrained by prior evolution. In order not to lose track of the objective of *finding generic patterns*, we refrain from discussing and analysing every mechanistic detail, and instead focus on major observables and discuss some illustrative cases.

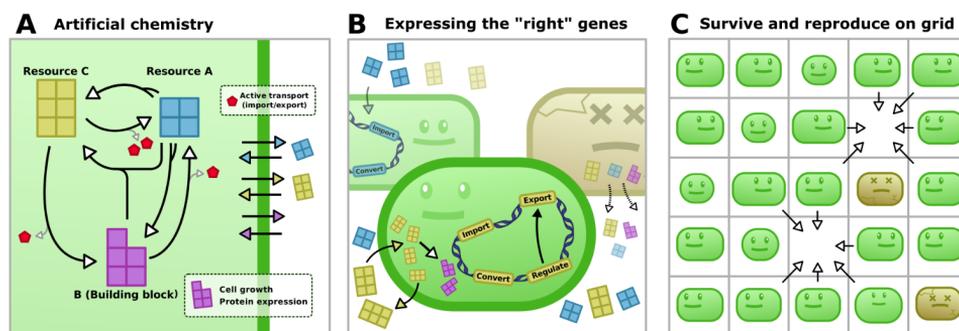


Figure 2.1: Virtual Microbes model overview. **A)** At the basis of the Virtual Microbe model is an artificial “metabolic universe”, describing all the possible reactions that can be catalysed. Resources (yellow and blue) are fluxed in, but building blocks (purple) and energy (red) must be synthesized to express proteins and transport metabolites across the membrane, respectively. **B)** A Virtual Microbe only needs to express a subset of all possible reactions to be viable, and no metabolic strategy is necessarily the “right” one. **C)** The individuals grow and reproduce on a spatial grid, and can only reproduce when there is an empty spot. Death happens stochastically or when a cell has accumulated toxicity by having excessively high concentrations of metabolites. Only cells that have grown sufficiently are allowed to reproduce, we simulate evolution with no prior expectation.

2.2.1 Evolving Virtual Microbe “wild types”

Before evolving Virtual Microbes in a serial transfer protocol, we first evolved a set of Virtual “Wild Types” (WTs). Instead of optimizing these WTs solely for high growth rates or optimal metabolic flux, we here mimic natural circumstances by fluctuating resource conditions (**Figure 2.2A**). When too little resource is available, the Virtual Microbes cannot grow, and can only stay alive for as long as their internal resources last. When too much resource is available however, the Virtual Microbes run the risk of accumulating too high concentrations of metabolites, resulting in increased

death rates due to toxicity. Furthermore, a stochastic death process is implemented, allowing even a maximally flourishing Virtual Microbes to only live 100 time steps on average. To avoid extinction, we divided the total grid into four sub-grids, where the two resource metabolites A and C (yellow and blue in Figure **Figure 2.1A**) independently change in their influx rates with probability 0.01 (see **Table 2.3**). Thus, on average, an individual will experience one fluctuation in resource conditions during its lifetime (see full configuration in S1). While both influxed resources can be converted into building blocks required for growth, the rates of influx span four orders of magnitude (10^{-5} – 10^{-1} , see **Table 2.3**), and conditions will thus vary from very favourable to very poor. Although poor conditions could cause a local population of microbes to go extinct due to limiting resources, total extinction is highly unlikely due to the 4 independent sub-grids. All this in turn depends on which resources the evolved Virtual Microbes like to consume (and at which rate), whether or not there is too much or too little resource, and whether or not space for reproduction is available. Finally, persisting in an unfavourable environment for a long time can be rewarding if conditions improve. All in all, this results in an unsupervised evolutionary process where there is no prior expectation of what metabolic strategy or gene regulatory networks might be best suited to survive. We study what will be the long-term target of the eco-evolutionary dynamics, not in terms of fitness, but in terms of what the Virtual Microbes evolve *to do*.

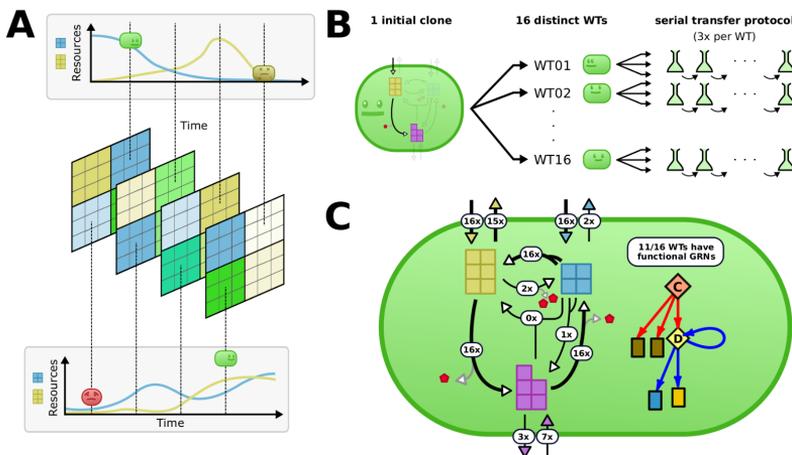


Figure 2.2: Evolution of Virtual “wild types” under naturally unpredictable and fluctuating resource conditions. **A)** Natural evolution is mimicked by (harshly) fluctuating resource conditions, resulting in a wide variety of resource conditions. The (actual) grid is 40x40, with four 20x20 subspaces where the rates of influx vary stochastically. These subspaces do not impede diffusion of metabolites or reproduction. The fluctuations of the A and C resource (blue and yellow respectively) are independent, resulting in a variety of different conditions. **B)** We repeat the evolution in natural conditions 16 times starting from the same (minimally viable) initial clone (varying the mutations that happen) yielding 16 distinct WT. These WT are later transferred to a serial transfer protocol. **C)** In the white labels we show how many of the evolved WT adapted to use particular reactions. The thicker arrows represent the shared core genome which consists of two resource importers, a metabolic cycle, and a C-exporter (yellow). Transcription factors (diamonds) were always present across WT, but only 11/16 WT visibly display changes in gene expression correlated with changes in the environment.

We evolved the same initial clone in the exact same “random” resource fluctuations, *only* varying the mutations that happened across $\sim 10,000$ generations of evolution. This produced 16 distinct WTs with their own evolutionary history, which we then expose to the serial transfer protocol (**Figure 2.2B**). Despite experiencing precisely the same fluctuations, no two WTs evolved to be the same. For example, we observe a great diversity in gene content, kinetic parameters of enzymes, gene regulatory networks and their complexity, and responses to environmental stimuli. The core metabolism is however strikingly similar across WTs, always consisting of a simple metabolic cycle. The rates of building block production and death rates are also very similar across all WTs (**Suppl. Figure S2.3**). In other words, it appears that there are many different ways to be fit, and that no solution is evidently better. The similarities and differences between our WTs are summarized in **Figure 2.2C**, but we discuss this in more detail in Supplementary Section 1.

2.2.2 *In silico* serial transfer evolution experiment

After evolving a variety of different WTs, we transfer the WTs to a serial transfer protocol. With regular intervals, all but 10 percent of the cells are removed, while at the same time refreshing the medium. Although time in Virtual Microbes has arbitrary units, we will refer to this process as the “daily” cycle from this point forward. Early in the day, during the log phase, high growth rates are very rewarding as there is a lot of opportunity to reproduce. However, once the population has reached stationary phase (having consumed all resources), it is favourable to survive and to not invest in growth any further. We will focus on how our WTs adapt to these alternating selection pressures. The results discussed here are found for a variety of different medium conditions (*e.g.* also see **Table S2**). In the main text however, we present the 50 time step serial transfer protocol where the medium contained both resources (A and C), as this was a condition on which all WTs could be cultivated, ensuring equal treatment. We focus on the generic adaptations towards this protocol first, and then show how specific WTs and contingent factors from their evolutionary history shape these outcomes.

2.2.3 All wild types evolve to anticipate the serial transfer protocol

After 800 days of evolving in a serial transfer protocol, we compare the ancestral WTs with the evolved populations. We first show some of the well-known growth dynamics of microbes: the lag-, log-, and stationary phase (**Figure 2.3A**). As most experimental evolutionary studies in the lab, we too observe a decreased lag phase and an increased growth rate. The increased growth rate in the evolved population results in an earlier onset of the stationary phase, which therefore takes much longer than for their WT ancestors. Eventually, this leads to a phase where the cell count decreases again (death phase), revealing a decrease in survival for the evolved populations. To further study how this decreased survival comes about, we next investigated the dynamics of average cell volumes. Cell volume is an indicator for the “health” of the population, determining the ability to divide (minimal division volume) and survive (minimal viable volume). A first interesting observation is an increase in average cell volume during the log phase (**Figure 2.3B-C**), which is also one of the first results from the LTEE (Lenski and Travisano, 1994). However, after

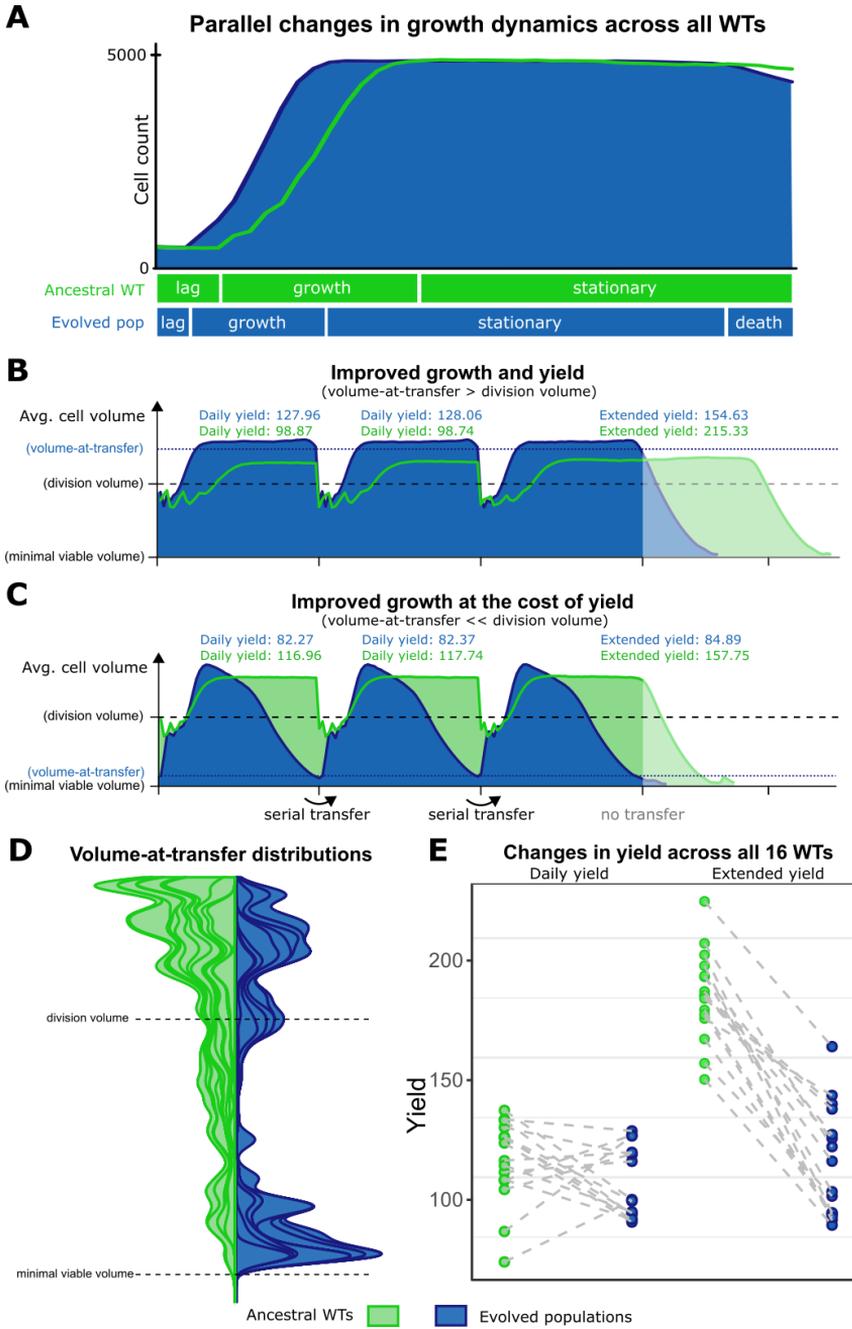


Figure 2.3: Virtual Microbes adapt to anticipate the regularity of a serial transfer protocol. **A)** Growth dynamics of early population (green) and evolved populations (blue) in terms of cell counts. (WT03#1 taken as an illustrative example). **B-C)** Two WTs (green) and the population after prolonged evolution in the serial transfer protocol (blue) are shown as an illustration of the anticipation effects. Over the course of 3 cycles, the average cell volume is plotted against time for the ancestral WT (green) and for the evolved population (blue). The y-axis (cell volume) indicates the minimal viable volume and division volume (which are fixed for the model), and the evolved volume-at-transfer (as measured at the end of the third cycle). Daily and extended yield are measured as defined in the method section. After the third cycle, serial transfer is stopped (transparent area), showing decreased survival of the evolved populations with respect to their ancestor. **D)** Stacked density distributions are plotted for the volume-at-transfer both early (transfer 0-40, green) and late (transfer 760-800, blue). **E)** The evolved changes in yield both “daily” (within one cycle of the protocol) and “extended” (after prolonged starvation) for all 16 WTs.

this increase in cell volumes during the log phase, evolved populations display a clear decrease in cell volumes, either at the end of the day (**Figure 2.3B**), or during the whole stationary phase (**Figure 2.3C**). Indeed, if we expose the populations to prolonged starvation by extending the day, the evolved populations die shortly after the anticipated serial transfer, while their WT ancestors survived for much longer (**Figure 2.3B-C**, right-hand side). Strikingly, we observed that the cell volume at the time of transferring the cells to a fresh medium (henceforth ‘volume-at-transfer’) fall into two distinct categories. In the high yield scenario (**Figure 2.3B**), cell volumes are maintained *above the division volume* until the very end of the day, whereas the low yield scenario, albeit having a higher growth rate, leads to a volume-at-transfer that is *just above minimal*. Indeed, the distribution of these observed volume-at-transfer across ancestral WTs are mostly high (**Figure 2.3D**, left-hand side), while the evolved cells clearly show a bimodal distribution (**Figure 2.3D**, right-hand side). Thus, all the populations evolved to either be ready to immediately divide at transfer (high yield mode), or exploit as much resource as possible while remaining above the minimal viable volume (high growth rate mode). Despite this difference in growth modes, both populations have evolved to accurately time the regularity of the serial transfer protocol. All evolved populations also show a consistent decrease in extended yield (**Figure 2.3E**) relative to the WTs, as long term yield is now masked from natural selection. Finally, we found that this anticipation effect did not depend on details in the protocol, such as the length of the daily cycle or the number of resources used (**Figure 2.5**, Table S2). This reveals that a key selection pressure in a serial transfer protocol is not only growth as fast as possible, but also remaining viable until the next day, anticipating the next supply of nutrients.

2.2.4 Evolution toward a growth-yield trade-off

The two extreme categories of cell volume dynamics from **Figure 2.3** illustrate a well-studied trade-off between growth and yield in microbial populations (Wortel et al., 2016; Li et al., 2019; Manhart et al., 2018). We next investigate how our different WTs evolve towards this trade-off, and how reproducible these trajectories are. For this, we repeated the serial transfer protocol 3 times for each WT, and

follow the trajectories over time. After ~ 800 serial transfers, all populations have adapted along a trade-off between growth and yield (**Figure 2.4A**). No trade-off was not observed during the first cycle of the protocol, which instead shows a positive correlation between growth and yield (**Figure 2.4B**), revealing how both growth and yield could initially be improved for most WT. Evolution towards the trade-off, by improving both growth and yield by *e.g.* importing more resources or producing more building blocks, is similar across all WTs, although not all WTs approach it with the same angle (also see **Figure 2.6**). Subsequent evolution on the trade-off diverges into two distinct clusters, representing the two aforementioned modes of high yield and high growth rate. This divergence is not only seen between different WTs (**Figure 2.4C-D**), but also occurs in replicate experiments of the same WT (**Figure 2.4E**, **Figure 2.6**). Finally, specific WTs appear to more readily give rise to certain outcomes, having specific adaptations in their “mutational neighbourhood”. This is for example illustrated by two WTs (5 and 11) that repeatedly gave rise to mutants with extremely high, but unsustainable growth rates, causing populations to go extinct repeatedly (black crosses in **Figure 2.4**). In summary, some WTs adapt in a similar way to the serial transfer protocol, while others (that have experienced the same amount of prior evolution) have diverging evolutionary trajectories and can reach different solutions, especially after having adapted towards the trade-off.

2.2.5 Anticipating as a collective

So far we have only looked at population averages. Next, we study the dynamics of lineages and the evolved dynamics within cells. To track lineages we tag each individual in the population with a neutral lineage marker at the start of the experiment (analogous to DNA barcoding (Blundell and Levy, 2014)). When a single lineage reaches fixation, we reapply these neutral markers, allowing us to quickly detect long-term coexistence. Moreover, these neutral markers allow us to study which arising mutants are adaptive in the different phases of the growth cycle. In **Figure 2.5A** we show dynamics of neutral lineage markers that are frequently redistributed when one lineage fixates in the population, indicating that there is no long-term coexistence of strains. In contrast, **Figure 2.5B** displays repeatedly observed (quasi-)stable coexistence, where two lineages coexist for some time, but coexistence was not stable in the long-term. Lastly, **Figure 2.5C** shows stable, long-term coexistence, where the population sustains a balanced polymorphism until the end of the experiment. Based on these lineage markers (also see **Suppl. Figure S2.7**), coexistence (either quasi-stable or stable) was observed in 21 out of 44 extant populations (**Figure 2.5D**).

By zooming in on the dynamics of coexisting lineage markers over a shorter time span (**Figure 2.5B-C**, right-hand side), we can better understand how these lineages stably coexist. Notably, one lineage is dominating during log phase, while the other lineage performs better during stationary phase. In other words, the lineages have specialized on their own temporal niche. We find that these dynamics can be the result of three mechanisms (or combinations thereof): 1) cross-feeding on building block metabolites, 2) specialisation on either of the two resources, or 3) based on the growth vs. yield trade-off. Cross-feeding dynamics always resulted in quasi-stable coexistence (such as depicted in 2.5B), and never resulted

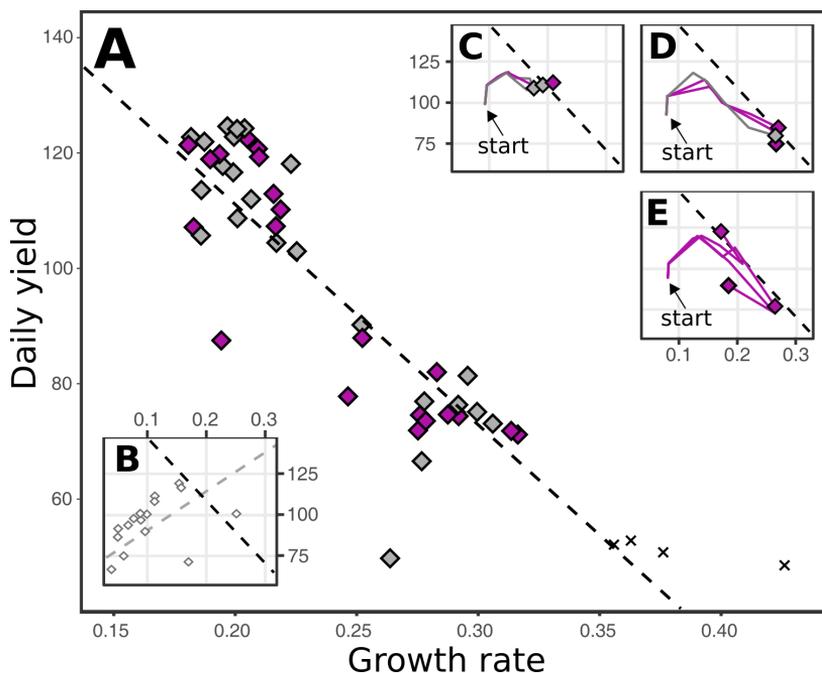


Figure 2.4: Trajectories towards a growth versus yield trade-off end in either the high growth rate mode or the high yield mode **A**) Growth rate (average building block production rate) is plotted against daily yield (average population biomass within a single cycle), for all the 48 experiments after adaptation to 800 serial transfers. The black dotted line is a linear regression model ($R^2 = 0.54$). **B**) Shows the initial points for all 16 WT's, which actually have a positive correlation between growth and yield ($R^2 = 0.32$) instead of the negative correlation (black dotted line). **C-E**) These insets display how the repeated evolution of certain WT's produce very similar trajectories towards the trade-off (time points are day 0, 20, 40, 100, 200 and 800), ending in either high daily yield (**C**) or low daily yield (**D**). Other WT's diverge after reaching the trade-off, and thus show more diverse trajectories when repeated (**E**). The colours of the end point symbols depict different modes of adaptation as discussed in the next paragraph (grey = no coexistence, purple = (quasi-)stable coexistence, black cross = extinction due to over-exploiting the medium).

in the balanced polymorphism as depicted in **Figure 2.5C**), while the other two mechanisms (resource specialisation and growth vs. yield differentiation) most often resulted in long-term coexistence where lineages perform better together than they do alone (**Suppl. Figure S2.9**).

While specialisation on different resources is a well known mechanism for negative frequency dependent selection, it is far less evident how a growth vs. yield trade-off would result in a fully balanced polymorphism. Mutants with higher growth rates but elevated death rates have a very distinct signature of increasing in frequency early in the daily cycle and decreasing to much lower frequencies during the stationary phase (**Figure 2.8A**), as apposed to lineages that increase in frequency throughout all phases of the cycle (**Suppl. Figure S2.8B**). While such mutants readily arise across our experiments, they often have difficulty rising to fixation due to the increased duration of the stationary phase, where they are unfit. In the meantime, a slower growing lineage with lower death rates can be optimized to utilize resources

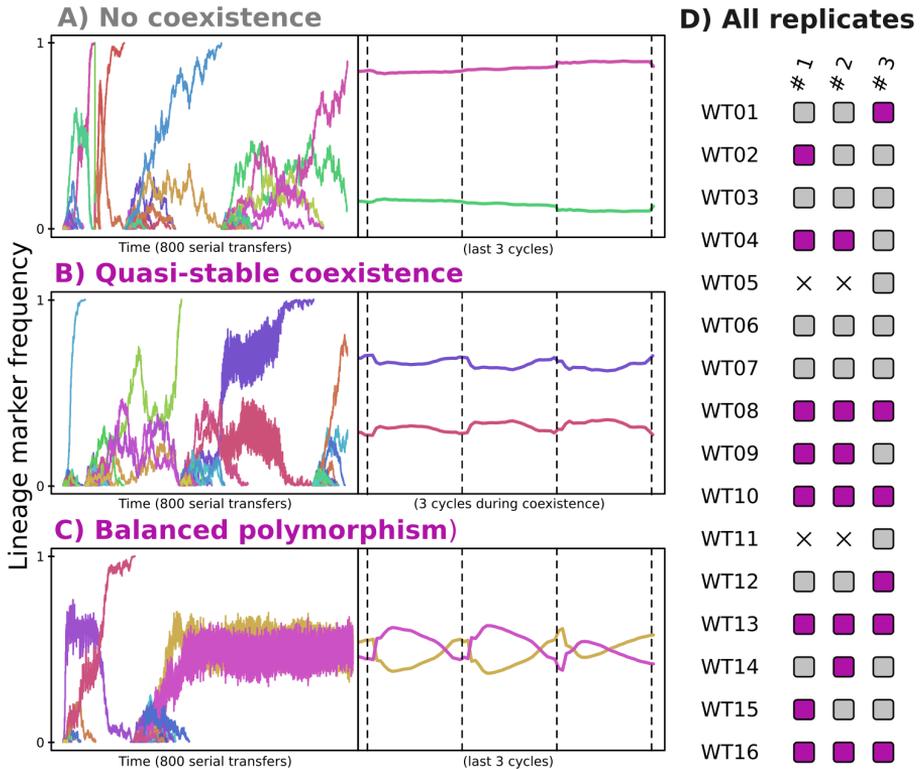


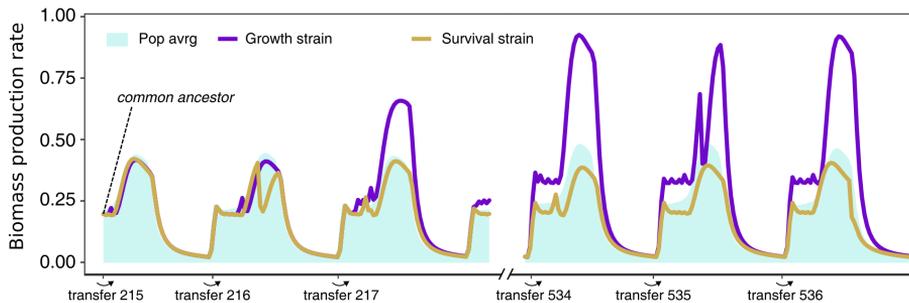
Figure 2.5: Dynamics of neutral lineage markers reveal balanced polymorphisms based on the daily cycle. **A-C)** Neutral lineage markers (random colours) frequencies are plotted along 800 serial transfers (left hand side) and along 3 cycles. Panel A shows an example with no coexistence which is found in 23 out of 44 replicates, and panel B and C show (quasi-)stable coexistence, found in the remaining 21 replicates. **D)** shows, for all 3 replicates of all WTs whether or not coexistence of neutral lineage markers was observed (grey = no coexistence, purple = (quasi-)stable coexistence, black cross = extinction due to over-exploiting the medium). Also see Supplementary **Suppl. Figure S2.7**.

at low concentrations during stationary phase. These dynamics can give rise to a balanced polymorphism that does not depend on resource specialisation or cross feeding, and it also observed in our experiments with a single resource (Table S2). Indeed, **Figure 2.5C** illustrates how two lineages with more than a three-fold difference in death rates (± 0.015 and ± 0.048) can stably coexist.

The different forms of polymorphism discussed above can differ strongly across WTs and replicated experiments. For example, since *de novo* gene discoveries were disabled during this experiment, cross-feeding on building blocks is only possible if the ancestral WT had the necessary importer for building blocks, which was true only for 6/16 WTs. Similarly, even though all WTs have the necessary importers for both the A and C resource, one WT consistently diverged into an A- and C-specialist (WT10). While other WTs have multiple gene copies for these importers, WT10 had only 1 copy of both genes, making the loss-of-function mutations readily accessible. In conclusion, although all polymorphic populations also anticipate the serial transfer protocol, they do so in a different way than populations consisting of a single lineage.

They all consist of strains which time growth and survival strategies in relation to each other in order to precisely finish the available nutrients by the end of the day.

A) Anticipation based on lineages specialising on growth rate and survival



B) Anticipation based on single lineage with two metabolic modes

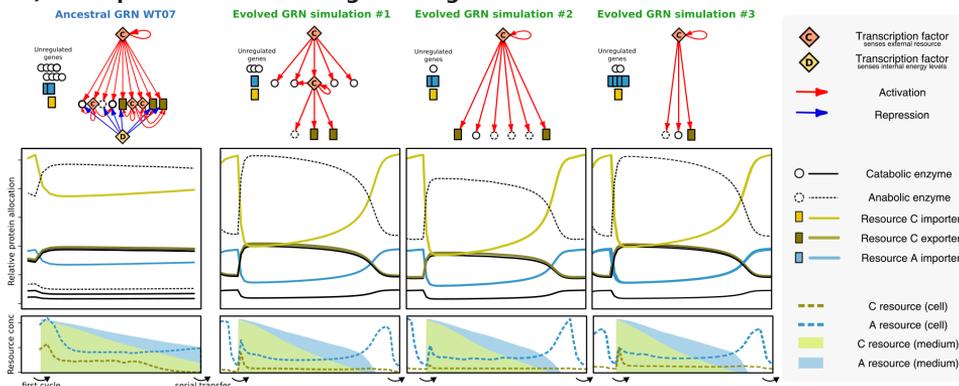


Figure 2.6: Anticipation can entail polymorphism or a single lineage that switches between two metabolic modes. **A)** Two lineages occupy different niches on the growth vs. yield trade-off WT02#01 diverges into a slow growing lineage (yellow lineage, average death rate ± 0.015) and a faster growing lineage with elevated death rates (blue lineages, average death rate ± 0.048), together anticipating the serial transfer protocol. **B)** A single-lineage anticipates the daily cycle by trimming and tuning the gene regulatory network. On the left the ancestral GRN, protein allocation dynamics, and resource concentrations are displayed over the course of 1 day. Next, after 400 days, all three independent simulations of WT07 are shown to have evolved to anticipate as a single lineage with two metabolic modes.

2.2.6 Individual anticipation by tuning and trimming the gene regulatory network

The previous section illustrates how multiple lineages can coexist because the predictable serial transfer protocol produces temporal niches. However, many of our WTs do not show any tendency to differentiate like this, and instead always adapt to the serial transfer protocol as a single lineage (**Figure 2.6D**). In order to better understand this, we will now look at the intracellular dynamics of WT07, and how it changes when adapting to the protocol. WT07 is one of the more “clever” WTs with a relatively complex GRN, and displays strong responses in gene expression when exposed to fluctuations. In **Figure 2.6B** we show that WT07 consistently

adapts to the protocol by switching between two modes of metabolism, where importer proteins are primed and ready at the beginning of the cycle, and exporter proteins and anabolic enzymes are suppressed during stationary phase. Despite some differences in the structure of the evolved GRNs, the protein allocation patterns are virtually indistinguishable across the three replicate evolutionary experiments. Interestingly, although no parallel changes were observed in the kinetic parameters of proteins, we do observe the parallel loss of a energy-sensing transcription factor as well as increased sensitivity of the TF that senses the external resource C. In other words, even though all mutations are equally likely, evolution apparently happened mostly through loss, and tuning and trimming of the GRN. Modulation between two metabolic modes allows this single lineage to switch between log and stationary phase, occupying both temporal niches. Indeed, a second lineage never appeared for this WT (Figure **Figure 2.6B** and Table S2).

2.2.7 Individual and collective solutions have similar macro-level observables

We have illustrated how all of our evolutionary experiments result in two modes, one with high yield, and another with high growth rates and lower yield. We have also shown how populations could or could not diversify into two strains, and how certain populations used regulated gene expression to adapt to all growth phases by itself. The four different combinations of collectives vs individual and regulating vs. non-regulating solutions, and their daily yield, is shown in **Figure**

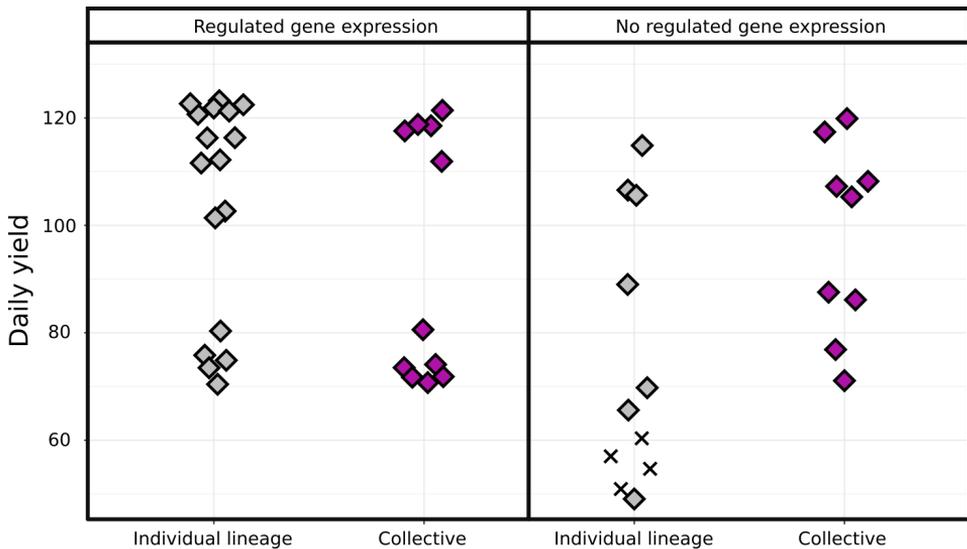


Figure 2.7: Individual and collective solutions have similar macro-level observables
The daily yield for all the evolved populations is shown, for groups of individual / collective solutions with and without regulated gene expression. Colours and symbols are identical to previous figures (grey=no coexistence, purple=coexistence). Only the non-regulating, individual lineages perform significantly worse than any of the other groups (performing all 6 Wilcoxon rank-sum tests with $\alpha = 0.05$)

2.7. As can be seen, all these combinations anticipate the serial transfer protocol using either the high yield or high growth rate strategy, and achieve similar values. The non-regulating individual solutions however clearly perform more poorly, as these populations lack the ability to fill both temporal niches (note that gene discoveries are disabled during the serial transfer experiment, so gene regulation cannot evolve *de novo*). Also note that, although the regulating WTs could fill both temporal niches by themselves, this does not prevent balanced polymorphisms from forming repeatedly. These results show that either a collective solution and/or gene regulation is required to be well-adapted to a serial transfer protocol, and that which solution is used is not observable on the overall macro-level.

2.3 Discussion

In this study we have taken a serendipitous approach to study how microbes adapt to a serial transfer protocol, and to what extent this is determined by their evolutionary history. The Virtual Microbe modelling framework serves this goal by building biology from the bottom up, *i.e.* implementing basic biological features and their interactions. We observe that regardless of their evolutionary history, all WTs learn to anticipate the regularity of the serial transfer protocol by evolving a fine-tuned balance between high growth rate and yield. Long-term survival without nutrients, which is now masked from natural selection, always deteriorates after prolonged exposure to such a protocol. Furthermore, this anticipation is done in two distinct ways. The high yield mode makes sure that the cells are ready to divide as soon as transferred to a fresh medium, whereas the high growth rate mode maximally exploits the medium but results in a poor performance during the stationary phase. We next show that WTs have similar trajectories towards a growth versus yield trade-off, but may subsequently diverge along it. Polymorphisms within populations are frequently observed, which can happen by means of cross-feeding interactions, resource specialisation, or by means of growth vs. yield specialisation. We furthermore find that these evolved collectives are dependent on one another, as both lineages perform better in the presence of the other. Finally, we show that regulated gene expression allows for an individual lineage to fill both temporal niches by itself, but that populations without regulated gene expression can still be well adapted to the protocol by diverging into two strains. In general, our results are robust to details in the serial transfer protocol, such as using only a single resource, or varying the interval between transfers (see Table S2). The anticipation effects therefore appear to be generic features of microbes exposed to prolonged evolution in a serial transfer protocol.

How do our results map onto experimental evolution in the lab? *E. coli* REL606 has been subjected to a daily serial transfer protocol for over 30 years ($\sim 70,000$ generations) in the LTEE. Many of our observations are very similar to the LTEE, such as the improved growth rate and cell sizes during the log phase (Lenski and Travisano, 1994), the (quasi-)stable dynamics of coexisting lineages (Good et al., 2017), and “leapfrogging” dynamics (*e.g.* **Figure 2.5A-B**) where an abundant lineage is overtaken by another lineage before rising to fixation (Papadopoulos et al., 1999; Sniegowski and Gerrish, 2010). The comparison with respect to the growth rates, yield, and the anticipation effects discussed in this work, is however less straightforward. We have observed how all our WTs quickly evolve to be maximally efficient given our artificial chemistry, and only subsequently diverge along the apparent growth versus yield trade-off (see **Figure 2.6**). For *E. coli*, growth and yield have continued to improve so far, and although a trade-off has been observed *within* the populations (Novak et al., 2006), no growth versus yield trade-off between the replicate populations has been observed yet. Nevertheless, we propose that anticipation of periodic environmental change, and a growth versus yield trade-off, provides testable hypotheses for the LTEE and similar experimental studies.

More similarities with empirical studies are found in the surprising number of experiments that result in balanced polymorphisms. A repeatedly observed mechanism for such a polymorphism is cross-feeding (Turner et al., 2015; Consuegra et al., 2017;

2

Rozen et al., 2005, 2009), where modeling has shown that this adaptive diversification involves character displacement and strong niche construction (Großkopf et al., 2016), and furthermore strongly depend on the regularity of a serial transfer protocol (Rocabert et al., 2017). We however also found balanced polymorphisms that did not include cross-feeding, involving one lineage with high growth rates during log phase and a slower growing lineage which performs better in stationary phase. Similar mechanisms of coexistence has been observed in respiratory and fermenting strains of *Saccharomyces cerevisiae* in chemostat (Wortel et al., 2016), and single nucleotide mapping has furthermore revealed the existence of this trade-off (Li et al., 2019). These results are directly related to r/K selection theory (Pianka, 1970), which describes an inherent conflict between the quantity and quality of ones offspring. Indeed, these dynamics have been shown to lead to two species stably coexisting in microbial populations (Turner et al., 1996; Smith, 2011; Manhart et al., 2018). Manhart & Shakhnovich (Manhart and Shakhnovich, 2018) furthermore show that an unlimited number of species can theoretically coexist within a serial transfer protocol, occupying any niche on a trade-off continuum. Here we show that these dynamics can emerge from a more complex eco-evolutionary setting. However, our results suggest that the trade-off between growth and yield is not continuous, as intermediate solutions rarely evolve. This is caused by the fact that as soon as the volume-at-transfer for our digital microbes is smaller than the division volume (*i.e.* something else than the main nutrient becomes limiting for division), a cell may as well exploit its resources fully.

Experimental evolution of *Pseudomonas fluorescens* has shown that different evolutionary paths can lead to the same phenotypic adaptations in a new environment (Rainey and Travisano, 1998; Lind et al., 2015). On the other hand, many studies have also suggested that adaptation can often entail mutations in the same genes (Ratcliff et al., 2015; Damkiær et al., 2013). In our experiments, prior adaptations can in some cases strongly shape the way subsequent evolution plays out, but these evolutionary constraints can strongly differ between WTs (**Suppl. Figure S2.6**). Furthermore, these data show that these evolutionary constraints may or may not diminish after prolonged evolution. There is a lot of variation on the predictability during the serial transfer experiment, revealing that evolutionary constraints by means of historical contingencies, are themselves the result of contingencies.

A factor that has been hypothesised to strongly impact the predictability and evolvability of biological systems are their GRNs (Crombach and Hogeweg, 2008; Raman and Wagner, 2011; Pelosi et al., 2006; Wagner, 2011), where for example global transcription factors could serve as mutational targets with large-scale phenotypic effects (Hindré et al., 2012). While our results (**Figure 2.6B**) clearly show an example where similar mutations result in similar adaptive changes, other regulating WTs showed much less predictability. For example, WT #09 is another strong regulating WT, but showed different outcomes with respect to diversification and regulation in all 3 cases. In other words, while the GRN appears to add knobs and buttons for evolution to push, other mechanisms are clearly available to adapt and be fit in a serial transfer protocol. One such mechanism could be ‘metabolic regulation’, which has recently been shown to be able to achieve very high levels of robustness without leading to a loss in adaptive degrees of freedom (Millard et al., 2017). Because all the kinetic parameters of enzymes (K_m , v_{max} , etc.) in the

Virtual Microbes are freely evolvable, it is likely that this metabolic regulation of homeostasis plays a very important role in Virtual Microbes. This could furthermore explain why the differences in evolvability between regulating and non-regulating populations were smaller than we initially expected. We have indeed observed that, for certain WTs, a change in metabolism could bypass regulated protein expression by means of kinetic neofunctionalisation of importer proteins, that evolved to be sensitive to different concentrations. Although such a solution does waste more building blocks on the continuous production of importer proteins, it is also much more responsive to environmental changes. It is possible that subtle differences like this explain, for example, why two of our WTs were much more sensitive to extinction by over-exploiting the medium than others. Furthermore, although the phenotypes that are reachable can be limited by prior evolution (Blount et al., 2012), the trajectories of evolution may be much less predictable on the long-term (Bajić et al., 2018). The role of metabolic regulation, and how this interplays with the repeatability and timescales of evolution, is a promising endeavour for future studies.

2.3.1 Who is anticipating what?

Our experiments reveal how populations of microbes can evolve to anticipate the regularity of a serial transfer protocol, trusting that new resources will be delivered on time. The concept of microbial populations anticipating predictable changes is frequently observed in nature (Mitchell et al., 2009; Wang et al., 2015; Mitchell et al., 2009), and is supported by theoretical models (Mitchell and Pilpel, 2011; Tagkopoulos et al., 2008). This form of anticipation however typically entails an environmental cue, where a preceding unrelated signal is used to anticipate environmental changes, usually followed by individuals taking some form of action. Without the necessity of such a cue, we show that anticipation can readily emerge in many different ways from an eco-evolutionary process. Although our form of anticipation is more passive, where not an individual but the system as a whole has temporal dynamics that accurately fit the protocol, this does not necessarily exclude individual-based anticipation. Like WT#07, most of the evolved regulating populations actually did not evolve to down-regulate their resource importers during the stationary phase, despite having repeatedly evolved to down-regulate other catabolic and anabolic enzymes (illustrated in **Figure 2.6B**). Since no more resource is available, and building blocks are consumed in order to keep expressing these importer proteins, this clearly does not have a positive impact during the late stationary phase. One can wonder why these individuals seem to keep the engine running. Whereas bet-hedging strategies have been shown to be a way to deal with irregular environmental changes (Veening et al., 2008b; Balaban et al., 2004; Solopova et al., 2014; Veening et al., 2008a; Siegal, 2015; New et al., 2014), this passive form of anticipation can be a way deal with regular, predictable changes in the environment. Furthermore, this could potentially be the first step towards active anticipation by means of a circadian rhythm, such as the sunflower heliotropism (Atamian et al., 2016) and the diurnal migration of life in lakes and oceans (Luell, Consalvey et al., 2004; Gaten et al., 2008).

2.3.2 Moving towards an eco-evolutionary understanding

The dynamics of Virtual Microbes expose that even a simple serial transfer protocol entails much more than sequentially evolving higher and higher growth rates. Instead, adaptation is an eco-evolutionary process that strongly depends on prior evolution, timescales, the presence of other competitors and mutants, and transient fitness effects. Although we found that competition experiments generally favoured the evolved population over the ancestral WTs, there were exceptions to this rule. It is therefore possible that the ancestral WTs perform better in such an experiment, but that this does not describe the stable eco-evolutionary attractor. Indeed, survival of the fittest is an eco-evolutionary process where any emerging lineage interacts with other lineages (or with other mutants) through changes in the environment, often resulting in a collective, community-based solution rather than the winner of all pair-wise interactions (Manhart and Shakhnovich, 2018). Furthermore, faster growth becomes less and less important as populations become better adapted to the serial transfer protocol, perhaps making the aforementioned interactions between lineages increasingly relevant. Other recent studies have recently elucidated the importance of eco-evolutionary dynamics (Govaert et al., 2018; Manhart and Shakhnovich, 2018), and how this can readily give rise to coexistence of multiple strains which could not have formed from a classical adaptive dynamics perspective (Vetsigian, 2017; Kotil and Vetsigian, 2018). Indeed, metagenomics have revealed much more diversity in the LTEE than previously anticipated (Good et al., 2017). Shifting focus from competition experiments towards the ever-changing selection pressures that emerge from the eco-evolutionary dynamics and interactions, will make the field of experimental evolution harder, but more intriguing, to study.

2.4 Conclusions

We have studied how *in silico* WTs of Virtual Microbes adapt to a serial transfer protocol like that of the LTEE. The LTEE has shown a sustained increase in competitive fitness, and intensive research displays how the evolved clones are still improving their growth rates with respect to their ancestor as to this day (Wiser and Lenski, 2015; Wiser et al., 2013; Lenski et al., 2015). Our experiments have generated a novel hypothesis that microbes in a serial transfer protocol will eventually evolve to anticipate the regular resource interval, and can do so by evolving either a high growth rate mode, or a high yield mode. Both these modes can be achieved by a single individual lineage, or by a collective of two strains which both have their own temporal niche. Taken together, our results reveal important insights into the dynamics and relevant selection pressures in experimental evolution, advancing our understanding of the eco-evolutionary dynamics of microbes.

List of abbreviations

- LTEE: Long Term Evolution Experiment (first published by R Lenski, 1991)
- WT: wild type (plural: WTs)
- TF: Transcription Factor (plural: TFs)
- GRN: Gene Regulatory Network (plural: GRNs)

2.5 Methods

A full description of the model and underlying equations is available online (bitbucket.org/thocu/virtual-microbes and <https://virtualmicrobes.readthedocs.io>). Here we summarize the sections of these documents that are relevant to this study.

2.5.1 Finding generic patterns of evolution

Experimental evolution is, of course, done on organisms that have evolved for a long time under a wide variety of conditions. These studied organisms all have their own evolutionary history, and differences in how they deal with starvation, stress, changes in resource etc. With Virtual Microbes we are able to evolve a *de novo* set of “wild types” (WTs), adapted to live in such severely fluctuating resource conditions. We can then explore how these WTs adapt to experimental evolution, and find generic patterns of evolution. To find generic patterns without being biased towards specific solutions, the biology of Virtual Microbes build-up from many levels with many degrees of freedom. One downside of this strategy can be that it can be hard for readers to understand all the underlying assumptions and algorithm and that many simulations result in a slightly different anecdote. However, we encourage the reader to read this work as though reading about ‘real’ biological evolution, where the experiments reveal new generic patterns and generate new hypotheses. With or without an understanding of the mechanistic details, relatively simple multilevel models can capture the eco-evolutionary dynamics of microbes, allowing us to study what happens, what else emerges from these dynamics “for free”, and equally important: what needs further explanation?

2.5.2 Model overview

Virtual Microbes metabolise, grow and divide on a spatial grid (**Figure 2.1C**). Here, we use two parallel 40×40 grids with wrapped boundary conditions. One grid contains the Virtual Microbes and empty grid-points, and the other describes the local environment in which the Virtual Microbes live. This environmental layer holds influxed metabolites, waste products of Virtual Microbes, and spilled metabolites from lysing cells (**Figure 2.1B**). In order to express proteins, grow, and maintain their cell size, Virtual Microbes must synthesize predefined metabolite(s), which we call building blocks. These building blocks are not directly provided, but must be synthesized by the Virtual Microbes by expressing the right proteins, allowing them to pump metabolites into the cell, and convert metabolites into one another (**Figure 2.1A**). The expression of these proteins depends on genes on genomes that undergo a wide variety of possible mutations upon reproduction (**Table 6.1**). Genomes are circular lists of genes, each with their own unique properties (*e.g.* K_m , V_{max} for enzymes, K_{ligand} and binding motif for TFs). The level of expression is unique for each gene, and is determined by its evolvable basal transcription rate and how this rate is modulated by transcription factors. When an enzyme or transporter gene is expressed, that specific reaction will take place within the cell that carries that gene. Note however that in the complete metabolic universe, many more possible reactions exist. The genome of an evolved Virtual Microbes will typically only use a subset of all the possible reactions. Genes to catalyse new reactions and novel TFs

Table 2.1: Types of mutations and their probabilities in WT evolution and serial transfer protocol (STP)

Mutation	Description	Prob (WT evolution)	Prob (STP)
Duplication	A stretch of 1 or more genes is duplicated in tandem	0.005	0.0015
Deletion	A stretch of 1 or more genes is deleted	0.005	0.0015
Inversion	A stretch of 1 or more genes is inverted in order	0.005	0.0015
Translocation	A stretch of 1 or more genes is moved to a random location	0.005	0.0015
(stretch length)	Geometrically distributed with $p = 0.3$	-	-
Gene discovery	Per time-step probability of discovering a new (randomly parameterised) gene.	0.0002	(disabled)
HGT	Per time-step probability of copying a gene from a cell closeby	0.002	(disabled)
Point mutation	Per gene per generation probability of modifying a single parameter of a gene (promoter strength, Michaelis Menten constants)	0.005	0.0015
Regulatory mutation	Per gene per generation probability of (partially) modifying the upstream binary operator sequence of a gene	0.005	0.0015

can be discovered through rare events. Which genes end up being selected for is not explicitly defined, but the result of a birth and death process. Birth depends on the availability of empty space and resources to synthesize new building blocks, whereas death depends on the ability to survive under a variety of different conditions and the potential accumulation (and avoidance) of toxicity. The resulting survival of the fittest (referred to as “competitive fitness” by Fragata *et al.*, 2018) is an emergent phenomenon of eco-evolutionary dynamics (Fragata *et al.*, 2018).

Metabolic universe The metabolic universe in Virtual Microbes is an automatically generated (or user defined) set of metabolites and reactions between them. The simple metabolic universe used in this study was automatically generated by a simple algorithm that defines 4 classes of molecules, how they can be converted into one another by means of 6 reactions, how fast they degrade, diffuse over the membranes, *etc.* (see Table 2.4).

The metabolism is simulated on the grid in terms of Ordinary Differential Equations (ODEs) using the GNU Scientific Library in Cython. These ODEs include the influx of molecules into the system, transport or diffusion across the membrane, intracellular metabolism (including expression and decay of proteins), biomass production, cell volume, the build-up of toxicity, *etc.* Diffusion between grid points was implemented as a simple local diffusion process, and is interleaved with the ODEs for efficiency. The number of simulations was limited to 16 WTs and 16×3

Table 2.2: Gene level mutations and the boundary conditions

Parameter	Gene Types	Value range in simulation
Promoter Strength	Enzyme, Transporter, TF	[0.001, 10]
$K_{substrate}$	Enzyme, Transporter	[0.001, 10]
K_{energy}	Transporter	[0.001, 10]
K_{ligand}	TF	[0.001, 10]
$K_{operator}$	TF	[0.001, 10]
v_{max}	Enzyme, Transporter	[0.001, 10]
effect-bound	TF	[0.001, 10]
effect-apo	TF	[0.001, 10]
ligand	TF	A, B, C, or e
exporting	Transporter	[True,False]
sense-external	TF	[True,False]
binding-motif	TF	bit flip at random position
operator-sequence	Enzyme, Transporter, TF	bit flip at random position

Table 2.3: Grid setup and environmental forcing in WT evolution and serial transfer protocol (STP)

Option (WT evolution)	Description	Value or range
Maximum population size	As defined by the size of the grid (40x40)	4900
Sub-grids	The grid is sub-divided into n grids where fluctuations are independent	4
Fluctuation frequency	Probability (per time step) of 1 metabolite (A or C) changes in influx in one of the sub-grids	0.01
Fluctuation range	New influx of metabolite is sampled uniformly from range	$[10^{-5}, 10^{-1}]$
Extracellular metabolite outflux	Rate at which metabolites outside of cells wash out	0.01
Option (serial transfer protocol)	Description	Value
Maximum population size	As defined by the size of the grid (70x70)	4900
Number of cells serially transferred	A (near) tenfold dilution of cells	500
Time steps of cycle	This represents, for example, the “24 hour” serial transfer protocol of the LTEE	50 (AUT)
[A] at beginning of cycle	Amount of resource A given at the beginning of the cycle	1.25
[C] at beginning of cycle	Am amount of resource C given at the beginning of the cycle	1.25
Extracellular metabolite outflux	Assuming metabolites can no longer wash out of the system	0.0

Table 2.4: A priori defined metabolites and reactions in artificial chemistry

Metabolite	Mass	Class	Degradation rate	Diffusion rate	Toxicity level
A	4	Resource	0.01	0.02	0.2
B	5	Building block	0.1	0.0015	0.2
C	6	Resource	0.01	0.015	0.2
e	1	Energy carrier	0.5	0.0015	0.2
Potential reactions (6)					
1C → 1B + 1e	1C → 1A + 2e	1A + 1B → 1C	2A → 1C	2A → 1B	1B → 1A + 1D

“lab” experiments due to computational feasibility. Statistics in this study only report effect sizes, as p-values are irrelevant in simulated studies (White et al., 2014).

Transmembrane transport For all molecules, transporters exist that import or export molecules across the cell membrane. Michaelis-Menten kinetics determine the transmembrane transportation with rate v :

$$v = v_{max\mathcal{T}} \cdot [\mathcal{T}] \cdot \frac{[S] \cdot [e]}{([S] + K_S) \cdot ([e] + K_e)}$$

where $[\mathcal{T}]$ is the concentration of the transporter protein, $[S]$ is the concentration of substrate transported, and $[e]$ is the concentration of available energy carrier metabolites. K_S and K_E are the Michaelis-Menten constants for the substrate and energy carrier respectfully. Depending on the direction of transport (importing or exporting) $[S]$ is either the external or the internal concentration of the substrate. Note that for any gene on the genome of a Virtual Microbe, $v_{max\mathcal{T}}$, K_S and K_E are all freely evolvable parameters.

Metabolism Similar to the transport, metabolic rates are catalysed by proteins by Michaelis-Menten kinetics with rate v :

$$v = v_{max\mathcal{E}} \cdot [\mathcal{E}] \cdot \frac{\prod_{R \in \mathcal{R}} [R]}{\prod_{R \in \mathcal{R}} ([R] + K_R)}$$

where $[\mathcal{E}]$ is the concentration of the enzyme catalysing the reaction, \mathcal{R} the set of all reactant metabolites, and K_R and $v_{max\mathcal{E}}$ are evolvable kinetic parameters of enzyme \mathcal{E} .

Biomass production Virtual microbes convert building block B to a biomass product P , which is consumed for cell growth and maintenance $Growth(B)$ and protein production $Prod(B)$, and determines strength with which individuals compete to reproduce. Biomass is next converted to cell volume with a fixed rate, and used for protein expression depending on the demands by the evolved genome. In other words, high rates of expression demand more biomass product for proteins, leaving less biomass product to invest in cell volume or maintenance (see cell volume growth). In total, the rate of change of P then becomes

$$\frac{dP}{dt} = \text{Production}(B) - \text{Growth}(B) - \text{Proteinexpression}(B) - \text{dilution} - \text{degradation}$$

where B is the concentration of building block metabolites. Production is a linear conversion of B into P , whereas growth, protein expression, and dilution depend on the dynamics of the cell. Biomass product is then consumed by cellular growth and protein expression which are a function of the building block concentration, is diluted proportional to the changes in cell volume, and degradation is fixed.

Consumption for protein expression is summed over all genes:

$$\sum_{i=1}^{N_{genes}} Pr_i \cdot Reg_i$$

where Pr_i is the basal expression rate of gene i , either up or down-regulated if transcription factors are bound to its operator sequence Reg_i (see transcriptional regulation).

Cell volume growth We assume that cell volumes approach a maximum cell size $MaxV$, and that there is a continuous turnover d of the cell volume at steady state, ensuring the necessity to keep on metabolising even if there is no possibility to reproduce (*i.e.* if the grid points are all full). Volume then changes as

$$\frac{dV}{dt} = g \cdot V \cdot \frac{1 - V}{MaxV} - d \cdot V$$

Transcriptional regulation The rates at which genes are expressed is a function of the basal expression rate of the gene and the concentrations of binding TFs and their molecular ligands. The intrinsic basal expression rate of a gene is encoded by a strength parameter in a gene's promoter region. This basal expression rate can be modulated by TFs that bind to an operator sequence associated with the gene. Binding sites and TF binding motifs are modelled as bit-strings and matching depends on a certain fraction of sequence complementarity. If a minimum complementarity is chosen < 1 a match may occur anywhere within the full length of the operator binding sequence and the TF binding motif. The maximum fraction of complementarity achieved between matching sequences linearly scales the strength with which a TF binds the target gene. In addition to binding strength following from sequence complementarity, TFs encode an intrinsic binding affinity for promoters K_b , representing the structural stability of the TF-DNA binding complex.

TFs can, themselves, be bound to small ligand molecules with binding affinity K_l , altering the regulatory effect they exert on downstream genes. These effects are encoded by parameters eff_{bound} and eff_{apo} for the ligand-bound and ligand-free state of the TF, respectively, and evolve independently. Ligand binding to TFs is assumed to be a fast process, relative to enzymatic and transcription-translation dynamics, and modeled at quasi steady state. We determine the fraction of TF that is not bound by any of its ligands L :

$$W_{apo} = \prod_{l \in L} \left(1 - \frac{[l]}{[l] + K_l}\right)$$

The fraction of time that a TF τ in a particular state σ (bound or apo) is bound to a particular operator o :

$$V_o = \frac{[\tau_\sigma] \cdot c_{\tau o} \cdot K_{b_\tau}}{1 + \sum_{\sigma \in \mathcal{S}} \sum_{\tau \in \mathcal{T}} [\tau_\sigma] \cdot c_{\tau o} \cdot K_{b_\tau}}$$

depends on the inherent binding affinity K_{b_τ} as well as the sequence complementarity score $c_{\tau o}$ between the TF binding motif and the operator sequence (Neyfakh et al., 2006). The binding polynomial in the denominator is the partition function of all TFs \mathcal{T} in any of the states \mathcal{S} that can bind the operator. Note that small declines in the concentration of free TFs due to binding to operators are neglected.

Now, the operator mediated regulation function for any gene is given by

$$Reg = \sum V_i \cdot E_i$$

with V_i the fraction of time that the operator is either unbound or bound by a TF in either ligand bound or unbound state and E_i the regulatory effect of that state (1 if unbound or eff_{bound} or eff_{apo} when bound by a ligand bound or ligand free TF, respectively). Finally, protein concentrations $[\mathcal{P}]$ are governed by the function:

$$\frac{d[\mathcal{P}]}{dt} = Pr \cdot Reg \cdot degr \cdot [\mathcal{P}]$$

where Pr is the evolvable parameter *promoter strength* and $degr$ a fixed protein degradation rate which is not evolvable.

Toxicity and death Virtual Microbe death is a stochastic process depending on a basal death rate, which is potentially increased when internal metabolite concentrations reach a toxic threshold. A cumulative toxic effect is computed over the current life time τ of a microbe as

$$e_{tox} = \sum_{m \in M} \int_{t=0}^{\tau} f(m, t) dt$$

for all internal molecules M , with

$$f(m, t) = \max(0, \frac{[m]_t - tox_m}{tox_m})$$

the toxic effect function for the concentration of molecule m at time t with toxicity threshold tox_m . This toxic effect increases the death rate d of microbes starting at the intrinsic death rate r

$$d = \frac{e_{tox}}{s + e_{tox}} \cdot (1 - r) + r$$

where s scales the toxic effect. Virtual Microbes that survive after an update cycle retain the toxic level they accumulated so far. Apart from toxicity and stochastic

death, cells can also starve. When insufficient biomass product is available to keep up the slowly decaying volume of the cell, the cells decrease in volume. If the cell volume drops below a *minimally viable volume*, this cell is automatically removed from the grid.

Reproduction When an empty grid point is available, the 8 (or less) neighbouring competitors get to compete to reproduce into the grid point. During the ‘*in silico* serial transfer protocol’ (see below), all cells are continuously mixed, so 8 (or less) random competitors are sampled. When cells compete for reproduction, the cells are ranked according to cell size. The “winner” is then drawn from a roulette wheel with weights proportional to this ranking. Upon reproduction, cell volume is divided equally between parent and offspring, and the genome is copied with mutations (see below). Molecule and protein concentrations remaining constant. Toxic effects built up during the parent’s lifetime do not carry over to offspring.

Genome and mutations The genome is a circular list of explicit genes and their promoter region, organised like “pearls on a string”. Genes can be enzymes, transporters, or transcription factors. At birth, the genome is subject to various types of mutations. Large mutations include duplications, deletions, inversions, and translocations of stretches of genes (see **Table 6.1**). At the single gene level, point mutations allow all evolvable parameters to mutate individually (see **Table 2.2**). Horizontal gene transfer can occur on every time step. Innovations are an abstraction of “HGT from an external (off-grid) source”, and allow randomly parameterised genes to be discovered at any given moment with a low probability.

2.5.3 Experimental setup

Metabolic network and wild type evolution We use a very simple metabolic network with 2 resource metabolites, 1 building block metabolite, and an energy carrier (**Figure 2.2A**). We initialised 16 minimally viable Virtual Microbes, and evolved them for ~10.000-15.000 generations in fluctuating resource conditions by applying random fluctuations of the influx rates for the A and the C resource. Because the rate of influx for the two resource metabolites fluctuates between very high (10^{-1}) and very low values (10^{-5}), conditions can be very poor, very rich, and/or potentially toxic. To avoid total extinction, we subdivided the 40×40 grid into four 20×20 subspaces, in which these fluctuations are independent (see **Figure 2.2B**). Note however that these subspaces do not impede diffusion and reproduction, but merely define the rate at which resources flux into different positions on the grid. In this study, the microbes do not migrate during their lifetime. These conditions, summarized in **Table 2.3**, aim to simulate natural resource fluctuations, evolving what we call “wild types” (WTs) of Virtual Microbes. (see supplement S1)

The initial population consists of cells that have 3 enzymes, 3 pumps, and 5 transcription factors. All these proteins are randomly parameterized, meaning that these proteins are unlikely to have good binding affinities and catalytic rates. The amount of building block required to grow and produce protein is therefore very minimal in the early stages of evolution, and is increased up to a fixed level as the Virtual Microbes become more productive over time.

***In silico* serial transfer protocol** We mimic a serial transfer protocol by taking our evolved WTs and – instead of fluctuating the resource conditions – periodically supplying a strong pulse of both the A- and the C-resource. While WTs are evolved in a spatial setting where resources flux in and out of the system, we here mix all cells and resources continuously and fully close the system, meaning no metabolites wash

in or out of the system during the daily cycle. To apply strong bottlenecks while at the same time allowing for sufficient growth, we increased the size of the grid from 40×40 to 70×70 . We then dilute population approximately tenfold, transferring 500 cells to the next cycle. Horizontal gene transfer between cells was disabled to represent the modified (asexual) *E. coli* REL606 clone that is used in the LTEE (Lenski et al., 1991). Finally, as the strong bottlenecks cause more genetic drift in our small populations than in the WT evolution, we found it necessary to dial back the mutation rates for the evolution of WTs to 30% to avoid over-exploiting mutants from appearing too easily (see **Table 6.1**). Other parameters of the serial transfer protocol are listed in **Table 2.3**.

Growth rate and yield measurements Yield was approximated by taking sum of all cell volumes. We measured yield both within a single serial transfer cycle (“daily yield”), and as the extended yield when we tested for long-term survival. As all WTs had slightly temporal growth rate dynamics, we estimated the growth rates as the average building block production during the first half of the protocol.

Characterising coexistence Using the neutral lineage markers (also see Supplementary **Suppl. Figure S2.7**), we manually characterised coexistence by looking at the dynamics of neutral lineage markers. When two neutral markers had relatively stable frequencies as visualised in **Figure 2.5B-C** for at least 10,000 time steps (approximately 100 generations), it was scored as coexistence. Sometimes coexistence did not last until the end of the simulation, which we refer to as quasi-stable coexistence.

Further configuration of Virtual Microbes Apart from the parameters within the confines of this article (Table 1-4), we have used the default settings for Virtual Microbes release 0.1.4, with the configuration files provided in Supplementary Section 2. Further details on the model and parametrisation are available online <https://bitbucket.org/thocu/virtual-microbes>

2.6 Supplementary Material

S1 - Evolution of Virtual Microbes wild types

Convergent and divergent evolution in Virtual Microbe wild types

In the evolution of our WT_s we observed strong convergence as well as divergence in the metabolic and gene regulatory networks that evolved. Because the evolved populations consist of a rich mix of different genotypes, we here describe the WT_s by profiling the gene repertoires and GRNs at the end of the simulation (~10.000 generations). For this, we took 20 (maximally unrelated) individuals from the evolved populations and determined the consensus metabolism (**Figure 2.1A**). While there is some diversity in the metabolic networks across WT_s, the shared gene repertoire constitutes a metabolic network that forms a metabolic cycle complemented with resource importers and an exporter for the C metabolite (**Figure 2.1B**). We observed that the discovery of both the metabolic cycle as well as the exporter favour survival, as it coincides with an increase in population size and a decrease in the number of generations per time step (**Suppl. Figure S2.4**). Note that in Virtual Microbes survival is improved by avoiding toxic effects of high metabolite concentrations and by only investing in growth when conditions are favourable for growth. The latter can be done via gene regulatory networks that respond to the quality of the environment, but we also found forms of metabolic regulation where microbes accurately fine-tuned kinetic parameters to automatically maintain homeostasis.

Although the shared gene repertoire from **Suppl. Figure S2.1B** does not contain transcription factors (TFs), all of the 16 WT_s have at least one type of TF. These TFs can constitutively repress or activate certain genes, or can respond to environmental conditions by binding to a ligand molecule. The latter response depends on the kinetic properties of the TFs and the properties of the genes which they regulate, all of which are evolvable (see Table S1). To get a better overview of how the WT_s respond to environmental stimuli we therefore chose to directly measure the gene expression levels in a variety of different resource concentrations (displayed for 6 WT_s in **Suppl. Figure S2.2**). On the level of these GRNs, and their sensitivity the environment, we clearly see signs of strong divergent evolution. Note however, that the *effect* on the importer and exporter proteins seems very similar between WT_s with different networks, showing that similar responses can be encoded by different GRNs. Finally, as seen in these graphs, some WT_s have no response to environmental stimuli. We found that these non-regulating WT_s are equally “fit”, in that they have the same rates of building block production and death rates (see **Suppl. Figure S2.3**). However, the majority (11/16) WT_s evolved clear regulatory mechanisms.

In short, during the *de novo* evolution of Virtual Microbe WT_s, some evolved features seem highly predictable. Namely, all have evolved the metabolic cycle, all express both resource importer proteins, and all but one WT have a C-exporter. On the other hand, regulatory mechanisms and some of the secondary reactions display considerable diversity. Note that this divergence cannot be explained by differences in initial conditions or fluctuations in resource concentrations, because the WT_s only differ with respect to the mutations that have happened in their evolutionary history. However, as shown in the main text, these differences have a profound effect on further evolution.

Table S1 - Important parameters for TFs for an environmental response

Property TF	Description
Expression TF	The TF itself needs to be sufficiently expressed
Binding motif	The binary binding motif (10 bits) must have a sufficient match to the operator sequences of genes (50 bits) in order to affect their expression
K_{ligand}	If the binding constant to the ligand K_{ligand} is not in range of the observed concentration of metabolites, the TF will always have the same (up or down) regulatory effect, regardless of the environment or internal concentrations.
Effect of ligand	The ligand-bound and ligand-unbound regulatory effects of TFs need to be different to effectively change expression of genes given any environmental stimulus

Table S2 - Anticipation and polymorphisms are also observed when changing in the serial transfer protocol For different transfer times, dilutions, and resources concentrations, seven WTs (11-16, and WT07 from **Figure 2.6** from the main text) have been tested for the anticipation effect and polymorphisms. Note that anticipation is not tested by prolonging the cycle (like in the main text), but by comparing the patterns in cell cycle dynamics with those from Figure 3 in the main text. If a clear decrease in cell volume was observed at the end of the cycle, it was scored as anticipation. Polymorphisms are scored as defined in the methods.

Shorter transfer time (25), 800 cycles		
WT	Anticipation (Yes/No)	Polymorphism (Yes,No,Quasi)
11	Y	N
12	Y	N
13	Y	Q
14	Y	N
15	Y	Q
16	Y	Q
07	Y	N

Higher transfer time (75), ~250 cycles		
WT	Anticipation (Yes/No)	Polymorphism (Yes,No,Quasi)
11	Y	N
12	Y	N
13	Y	Y
14	Y	N
15	Y	N
16	Y	Q
07	Y	N

S2 - Virtual Microbe configuration

The evolution of these WTs was done with Virtual Microbes version 0.1.4 which is publicly available as a Python package. Complete documentation on the methods is publicly available on <http://bitbucket.org/thocu/virtualmicrobes>. Full details on the configuration are attached to the publication on the website of BMC Evolutionary Biology.

S3 - Supplementary Figures

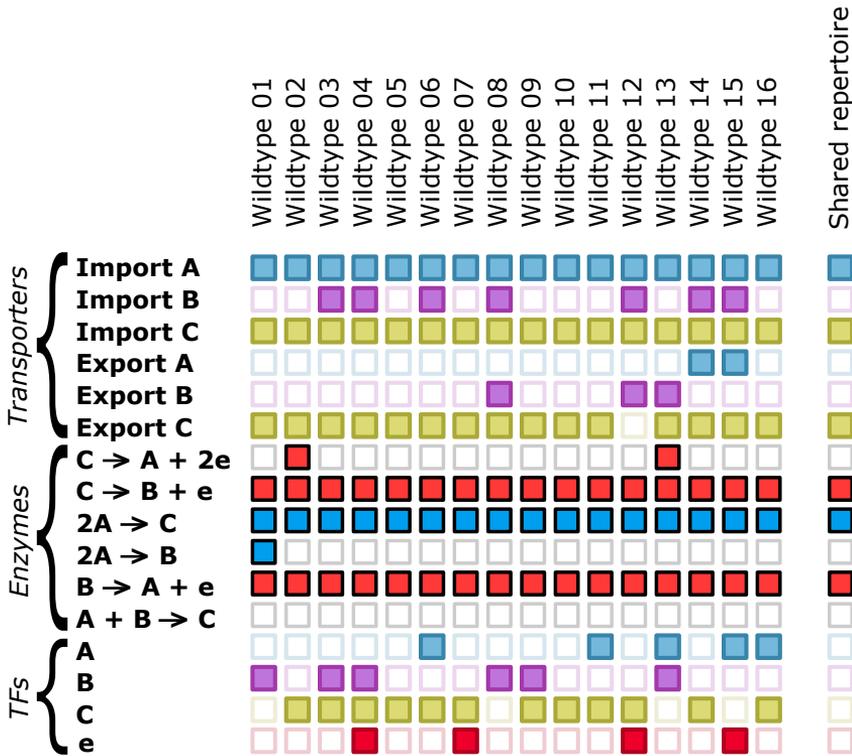


Figure S2.1: The evolved gene repertoires for all 16 WT The gene repertoires of WT (20 maximally unrelated individuals) is displayed for all 16 replicate simulations after 1.000.000 time steps. Rows represent the different types of proteins (transporters, enzymes and TFs), and the columns the gene repertoires. Note that the presence of a gene does not imply it is functional, since properties such as K_s and V_{max} might be poorly parameterised. Genes found in less than 4 cells or genes with low concentrations (*i.e.* low expression) were omitted. The separate column depicts the genes found in at least 90% of WT.

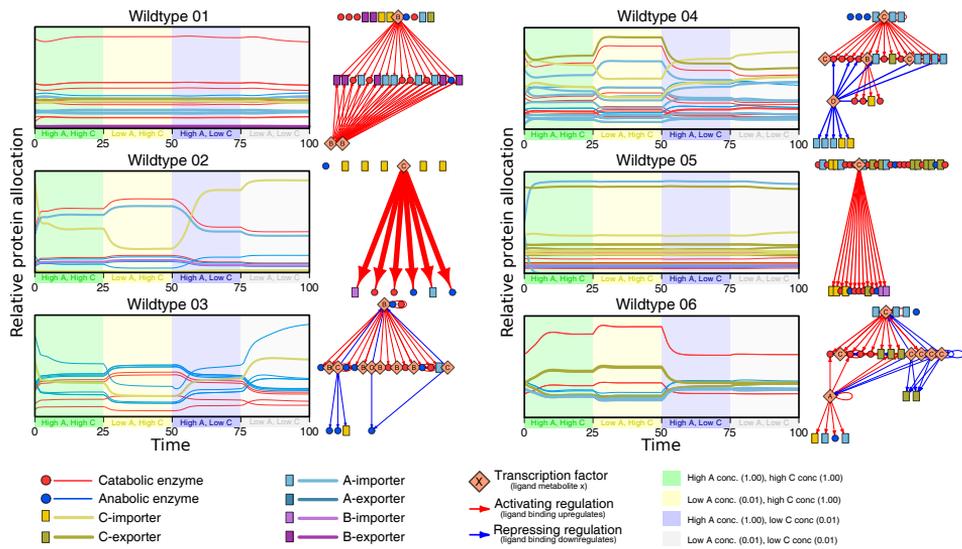


Figure S2.2: WTs have great diversity of transcriptional regulation, and not all respond to changing resource concentrations The graphs show how the protein allocation shifts to imposed changes in resource concentration of the environment. Single clones from the WTs were taken and resource concentrations of the A- and C-resource were varied from low (0.01) to high (1.0). The gene regulatory network responsible for these changes is displayed next to each graph. The colours for different enzymes are as displayed in the legend. Thicker arrows in the gene regulatory network represent higher expression levels of the transcription factors. Genes with very low expression levels were omitted for clarity.

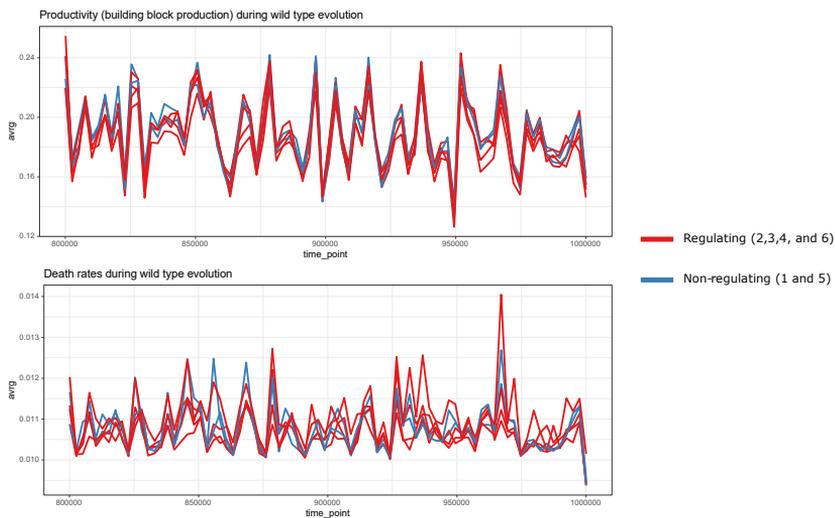


Figure S2.3: Different WTs have similar building block production rates and death rates. For 6 WTs, the last 2000 generation of WT evolution did not show any differences in “fitness”, which is approximated by the rates of building block production and death. Furthermore, regulating WTs showed no difference compared to non-regulating WTs.

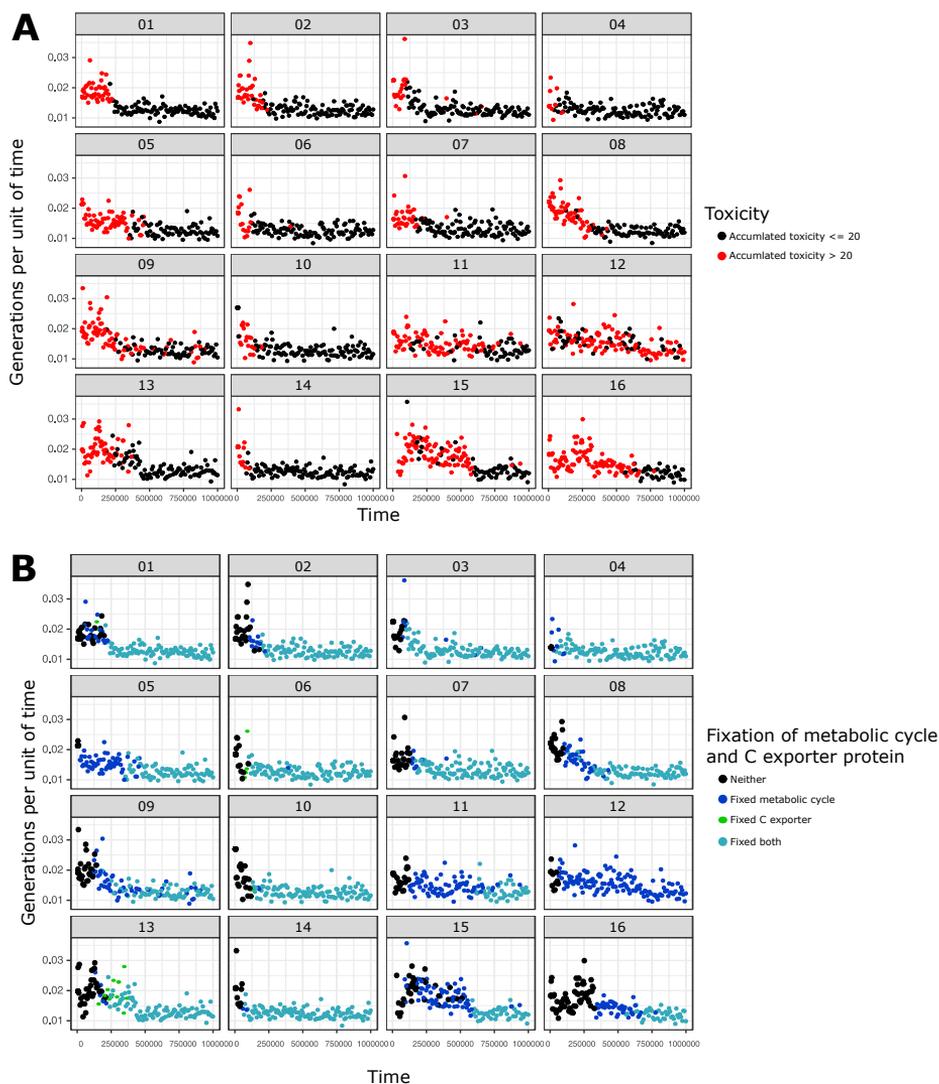


Figure S2.4: Decrease in number of generations per time step for WTs coincides with a decrease in toxicity and/or fixation of a metabolic cycle Decrease in number of generations per time step coincides with a decrease in toxicity accumulation and/or fixation of a metabolic cycle. Every dot represents an average over a 100 generations of simulation. For panel A, red dots represent a toxicity level above which death rate is increased at least threefold (toxicity > 20), and black dots represent the lower toxic levels (toxicity ≤ 20). In panel B, the cyan dots represent the fixation of both the C-exporter and the metabolic cycle, green dots the fixation of only the C-exporter, blue dots the fixation of the metabolic cycle, and black dots the fixation of neither of these features.

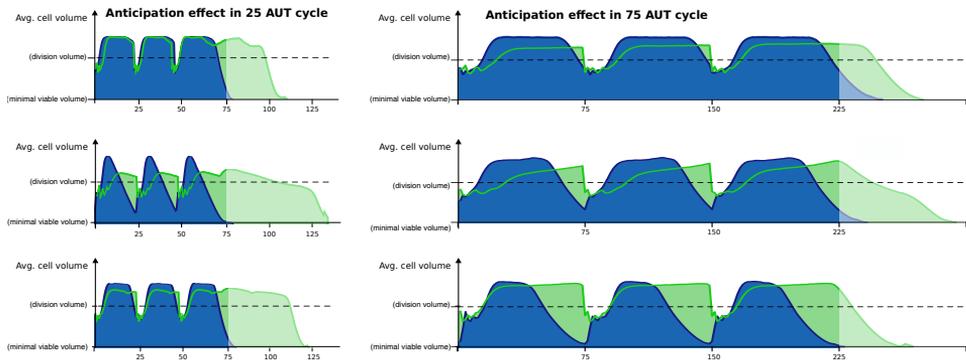


Figure S2.5: The anticipation effect is present after evolving in shorter (left) and longer (right) time intervals.

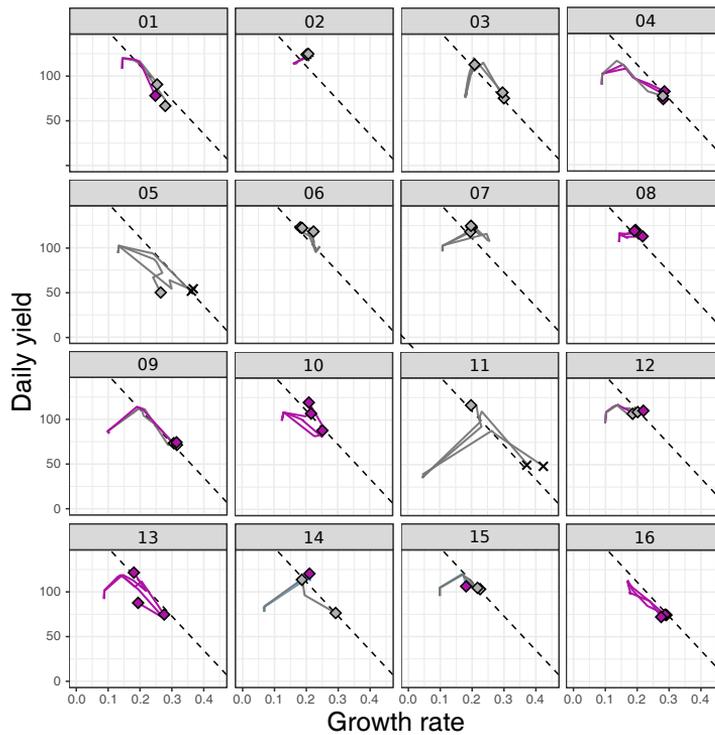


Figure S2.6: Three replicate serial transfer experiments for all 16 WT. This image shows all the data from the examples from Figure 2.4.

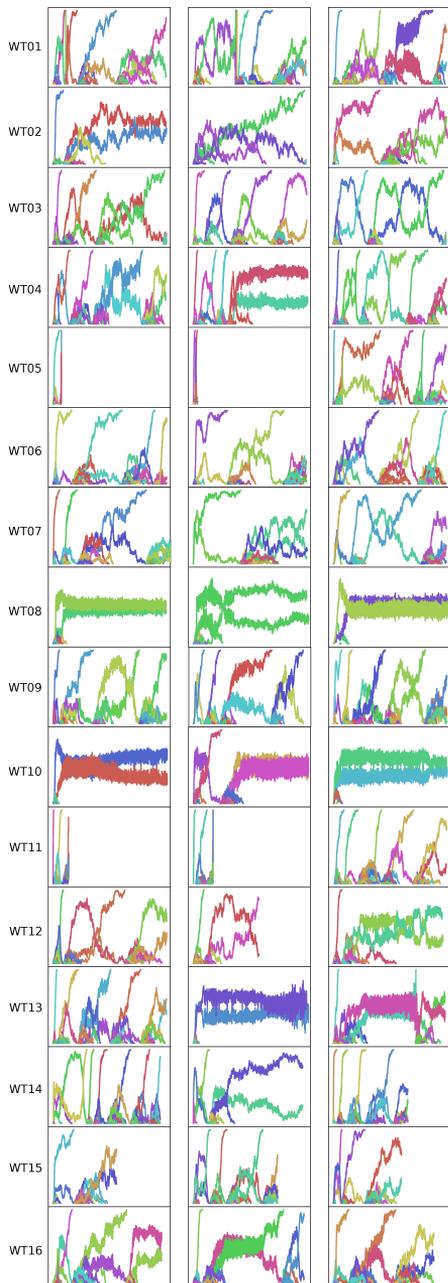


Figure S2.7: Dynamics of lineage markers in all simulations The neutral lineage markers for all 3 replicates of all WTs is shown, from which the coexistence was inferred (see methods).

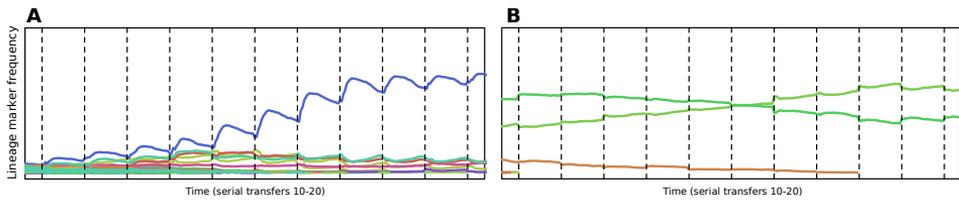


Figure S2.8: Dynamics of different invading mutants A) Neutral lineage markers linked to mutations which cause high growth rates but poor survival increase in frequency early, but are losing in frequency during stationary phase. B) Neutral lineage markers linked to mutations that have higher growth rates without trading off against survival show no such temporal fitness effect.

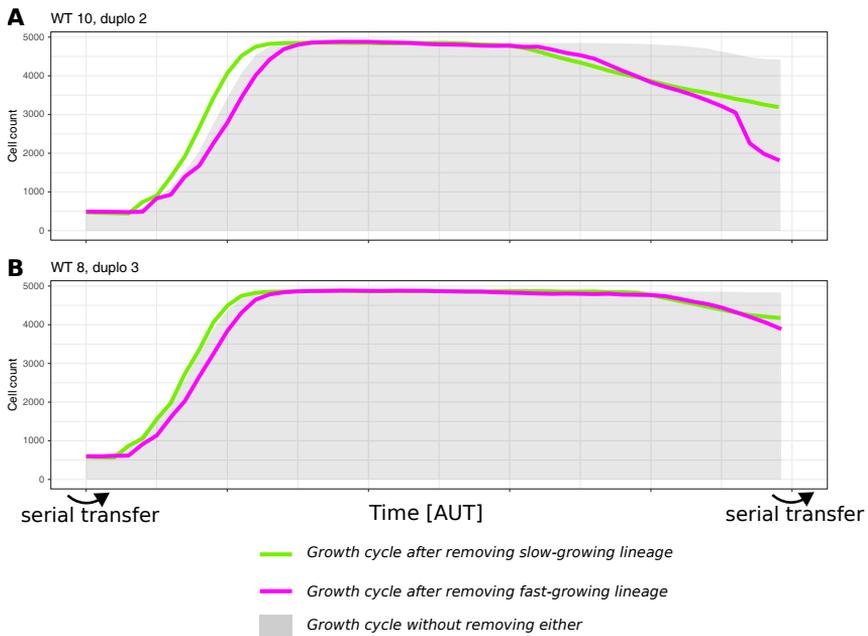


Figure S2.9: Lineages based on resource specialisation (top) and growth vs. yield specialisation (bottom) perform better in the presence of the other In the upper panel, two resource specialists over-exploit their resource in the absence of the other. In the bottom panel, without evident resource specialisation, both lineages have increased toxicity (given that this happens through enzymatic changes, this is likely due to changes in fluxes), which accumulates to significant levels near the end of the cycle. This figure thus shows that lineages can grow dependent on one another, as they are part of each-others environment.

“FALSE VIEWS, IF SUPPORTED BY SOME EVIDENCE,
DO LITTLE HARM, FOR EVERYONE TAKES A SALUTARY
PLEASURE IN PROVING THEIR FALSENESS.”

—Charles Darwin

Chapter 3

Contingent evolution of alternative metabolic network topologies determines whether cross-feeding evolves

JEROEN MELJER^{1,2}, BRAM VAN DIJK^{1,2}, AND PAULIEN HOGEWEG¹

1 UTRECHT UNIVERSITY, UTRECHT, THE NETHERLANDS

2 CONTRIBUTED EQUALLY TO THIS STUDY

UNDER REVISION AT NATURE COMMUNICATIONS BIOLOGY

Abstract

Metabolic exchange is widespread in natural microbial communities and an important driver of ecosystem structure and diversity, yet it remains unclear what determines whether microbes evolve division of labor or maintain metabolic autonomy. Here we use a mechanistic model to study how metabolic strategies evolve in a constant, one resource environment, when metabolic networks are allowed to freely evolve. Initially identical ancestral communities of digital organisms follow different evolutionary trajectories, as some communities are dominated by a single, autonomous lineage, while others are formed by stably coexisting lineages that cross-feed on essential building blocks. Our results show how without presupposed cellular trade-offs or external drivers such as temporal niches, diverse metabolic strategies spontaneously emerge from the interplay between ecology and metabolic constraints that arise during the evolution of metabolic networks. Thus, in the long term, whether microbes remain autonomous or evolve division of labour is an evolutionary contingency.

3.1 Introduction

Natural microbial communities are typically complex, composed of many different taxonomic groups that stably coexist. The complexity of these communities might partly reflect the complexity of their environment, which can generate and maintain diversity by allowing specialisation on different niches (Lenski et al., 1991; Rainey and Travisano, 1998; Rozen and Lenski, 2000; Le Gac et al., 2008; Rozen et al., 2009; Le Gac et al., 2012; Herron and Doebeli, 2013; Kinnersley et al., 2014; Rocabert et al., 2017; Behringer et al., 2018; Dal Co et al., 2019b). Experimental evolution has shown how initially clonal populations can adaptively diversify into stably coexisting ecotypes, each specialized on pre-existing niches defined by available nutrients (Herron and Doebeli, 2013), spatial structure (Rainey and Travisano, 1998; Dal Co et al., 2019b), temporal variability such as the feast and famine cycles in serial transfer (Lenski et al., 1991; Rozen and Lenski, 2000; Le Gac et al., 2008; Rozen et al., 2009; Le Gac et al., 2012; Kinnersley et al., 2014; Good et al., 2017), or combinations thereof (Behringer et al., 2018). However, even in constant, unstructured environments with a single limiting carbon source metabolic diversification routinely evolves (Helling et al., 1987; Rosenzweig et al., 1994; Treves et al., 1998; Kinnersley et al., 2014). Here, new niches are constructed by microbes themselves, as metabolic byproducts released by one become the growth substrate for another, allowing stable coexistence mediated by metabolic interactions. For example, initially clonal *Escherichia coli* populations grown in a glucose-limited chemostat genetically diversify into a lineage that rapidly (but inefficiently) grows on the provided resource, and lineages that specialize in using the overflow acetate produced by the first lineage (Rosenzweig et al., 1994; Kinnersley et al., 2009, 2014). Niche construction can thus lead to stable coexistence even with a single resource, which according to the competitive exclusion principle (Gause, 1934) would support only a single species.

Several mechanisms have been proposed that might both explain the evolution of cross-feeding in simple, constant environments as well as the prevalence of cross-feeding in natural communities (Rosenthal et al., 2011; Kwong et al., 2014; Pande et al., 2014; Zelezniak et al., 2015; Garcia et al., 2015; Blasche et al., 2019) (see (D'Souza and Kost, 2016) for a review). For example, cellular or metabolic trade offs might favour metabolic specialisation (Johnson et al., 2012; Großkopf et al., 2016), and division of labour might increase productivity of a community (Thommes et al., 2019). Alternatively, the “Black Queen Hypothesis” (BQH) holds that gene loss can be adaptive for functions that are costly, leaky and essential—such as the production of essential building blocks that end up in the environment through diffusion or lysis—provided other community members compensate for the lost function (Morris et al., 2012; D'Souza et al., 2014; Pande and Kost, 2017). By removing two metabolic genes in *E. coli*, a recent study showed that in a synthetic community with engineered dependencies for amino acids, strains grow up to 20% faster compared to the autonomous wild type, suggesting positive selection for the loss of biosynthetic genes (Pande et al., 2014). Cross-feeding might thus be caused by a gene loss ratchet in a community of initially autonomous microbes that, driven by escaping the costs of bearing biosynthesis genes, evolve complementary metabolic networks and become dependent (Dekel and Alon, 2005; Tsoi et al., 2018).

Theoretical studies on the evolution of metabolic dependency typically pre-define a

limited number of metabolic strategies (Stump et al., 2018b), take cellular trade-offs for granted (Pfeiffer et al., 2001; Gudelj et al., 2007; Großkopf et al., 2016; Josephides and Swain, 2017; Stump et al., 2018b; San Roman and Wagner, 2018), or explicitly assume external forcing such as seasonality (Rocabert et al., 2017; van Dijk et al., 2019), and investigate how other conditions favour autonomous or cross-feeding strategies or prevent the evolution of metabolic free-loaders (Stump et al., 2018a). For example, by constructing genome-scale metabolic models and limiting the number of reactions that can be performed by a single individual, under certain conditions multiple cross-feeding strategies together can outperform an autonomous strategy (Thommes et al., 2019). However, these trade-offs are not always necessarily a inescapable physical reality that must be faced, but are themselves evolved cellular properties (Roller et al., 2016), and little attention is given to the evolution *of* trade-offs compared to the evolution *on* trade-offs. In addition, these models tend to assume ecological and evolutionary timescales are separated - the fate of each individual mutation is determined by playing out ecological dynamics before the next mutation arrives. We know from experimental studies (Kinnersley et al., 2019; Good et al., 2017; Levy et al., 2015; D’Souza and Kost, 2016; Ba et al., 2019; Cvijović et al., 2018) that microbial communities do not operate in such a mutation-limited domain, and instead, multiple mutations continuously emerge and cause complex evolutionary dynamics and contingency. Theoretical work (Van Der Laan and Hogeweg, 1995; Kotil and Vetsigian, 2018) demonstrated that under such true “eco-evolutionary” conditions, evolutionary dynamics qualitatively change and result in more complex and diverse ecosystems.

Here, we investigate how different metabolic strategies can evolve with a bottom-up, mechanistic model of microbial eco-evolutionary dynamics that explicitly accounts for the high diversity in communities, and without assuming cellular trade-offs or external drivers for metabolic division of labour. We show that initially identical populations –when propagated under identical conditions– can reach two qualitatively different eco-evolutionary attractors: a community of cross-feeding metabolic specialists, or alternatively, a community formed by a single lineage of microbes that are metabolically autonomous, producing all metabolites they need. Which type of community evolves is dependent on, and can be predicted from a “frozen metabolic accident”: the topology of the metabolic network that typically fixes much earlier in evolution. Differences between these topologies appear to be neutral across populations at time of fixation, but have far reaching repercussions for subsequent evolution.

3.2 Results

Model overview

In the model, we explicitly incorporate a “chemical universe”, metabolism, cell growth and division, genome evolution, and a two-dimensional spatial environment that all co-evolve (**Figure 3.1**). We do not predefine fitness (such as a target genotype, or biomass reaction to be optimized) but set basic rules for cell growth, reproduction, and death. This means whether a mutation is beneficial, neutral, or deleterious depends on local environmental conditions and interactions, cellular state, and genomic background, all of which are shaped by prior evolution in the model. As a consequence, metabolic or ecological strategies are not predefined but emerge during evolutionary simulations, as microbes evolve and explore the possibilities of the chemical universe and reshape their local environment by metabolite uptake and exchange. This approach allows us to *de novo* create microbial communities with their own evolutionary histories, and study them with access to a perfect digital fossil record. For a detailed model description and parameters, see Methods.

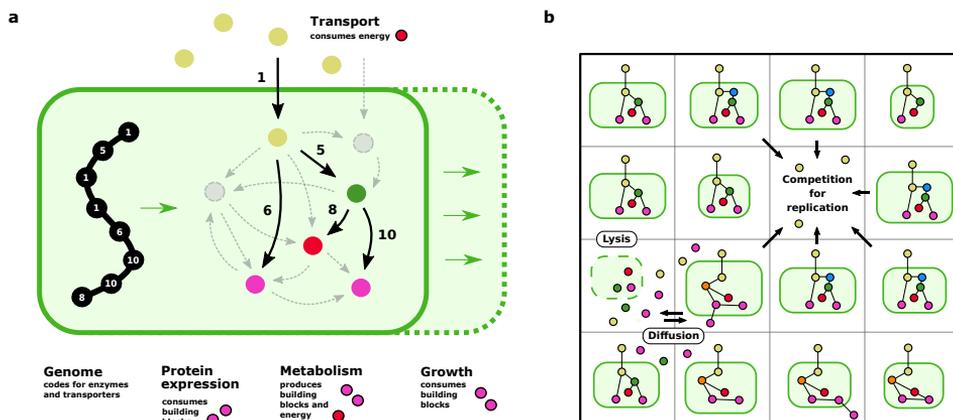


Figure 3.1: Model of microbial eco-evolutionary dynamics. **a.** Genes on a linear genome code for specific metabolic enzymes that catalyze individual reactions of the metabolic network. To express proteins and grow, microbes require two non-substitutable building block metabolites B_1 and B_2 (pink) that do not natively occur in their environment, but can be metabolized from the single provided resource R (yellow) by expressing the right metabolic pathways. Active transport of metabolites across the cell membrane requires an energy metabolite E (red). The genome of a single microbe typically covers a small subset of the complete “chemical universe” of 59 reactions (see **Table 3.2**). **b.** Microbes compete for space and metabolites on a 45×45 lattice. They can reproduce in an adjacent empty space if they meet the minimal division cell size. Here, microbes NE and W of the empty space are too small to reproduce. Upon replication, genomes can mutate through gene duplication and deletion, discovery of new genes, and point mutations that can change the expression rate and kinetic parameters of individual genes. New genes can also be acquired via horizontal gene transfer from nearby microbes. Active transport of metabolites and lysis changes the composition of a microbes’ local environment.

Microbes compete for a single resource molecule R and limited space on a 2d grid, reproducing locally into empty neighboring sites (**Figure 3.1a-b**). They require two essential (non-substitutable) “building block” metabolites B_1 , B_2 for cell growth and expressing proteins that perform metabolic functions. Building

blocks do not natively occur in the environment, but can be synthesized from the provided resource by expressing relevant metabolic proteins. In addition to building blocks, microbes require energy metabolite E to operate transporter proteins to pump metabolites (such as the provided resource) in and out of the cell. The chemical universe available to evolution to meet these metabolic demands consists of a predefined set of 9 metabolites (R , B_1 , B_2 , M_{1-5} , and E) connected by 59 reactions (43 conversion reactions and 16 transport reactions, see **Supplementary Table 3.2**), which contains many redundant pathways that provides many degrees of freedom to form functional metabolic networks. Proteins catalyze individual reactions (*e.g.* $1R \rightarrow 1M_1 + 5E$), that can be combined to form metabolic pathways. They are coded on the microbes' genome, which typically covers only a small subset of all reactions. When cells reproduce the genome can mutate, allowing the metabolic networks to evolve by tuning the rates of individual reactions through point mutations (basal expression rate and kinetic parameters of the enzyme) and gene copy number (gene deletion, duplication). New pathways can be formed by discovering new genes, or through horizontal gene transfer from nearby cells.

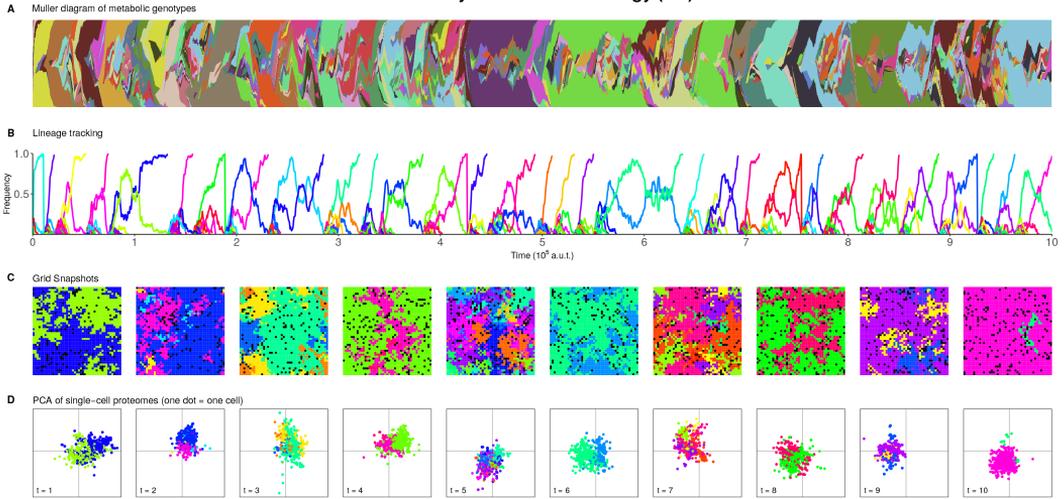
Cells can reproduce in a neighbouring empty site if they meet a minimal division size (**Figure 3.1b**), with competition biased towards larger cells when multiple cells are eligible. Cell death is modeled as a stochastic process, with a basal death rate that is potentially elevated when internal metabolites reach toxic concentrations. Cell lysis releases all internal metabolites into the local environment, which then locally diffuse and become available for other nearby microbes to take up. In this way, microbes change the metabolite composition of their local environment through active transport, passive diffusion across the cell membrane, and cell death (see fig **Figure 3.1b**). Motivated by experimental work (Dal Co et al., 2020, 2019b) that shows that micro-scale gradients quickly establish and influence microbial metabolism, we first consider evolution in a spatially structured environment with limited diffusion (mimicking biofilm conditions), and subsequently investigate evolution simulating a well-mixed medium.

We constructed an initial population by generating 2025 minimally viable microbes with randomly parameterized genes to import the provided resource R and synthesize both required building blocks. We then evolved 60 identical copies of this population in parallel under the exact same conditions for 10^6 time steps ($\sim 4 \cdot 10^5$ generations), while fluxing in food metabolite R at a constant rate at all grid points. Using this model, we examine whether “ecosystem based” metabolic strategies evolve, *i.e.* cross-feeding species with complementary metabolic networks, or “individual based” strategies in the form of autonomous microbes that produce all required building blocks.

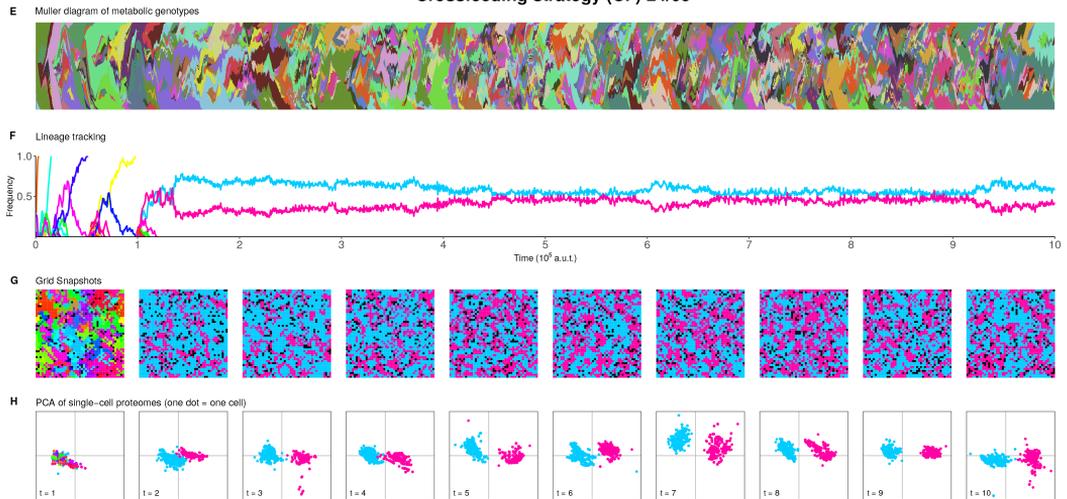
3.2.1 Diverse metabolic strategies evolve on a single resource

We investigated the evolution of metabolic strategies with a mechanistic model, first focusing on the effect of contingency with a parallel evolution experiment. The ancestral community consists of microbes with metabolic networks composed of a food importer and randomly selected genes to produce both building blocks, all of which have randomly sampled kinetic parameters and expression rates. During the

Metabolically autonomous strategy (AU) 21/60



Crossfeeding strategy (CF) 24/60



Switching between autonomous/crossfeeding strategies (SW) 15/60

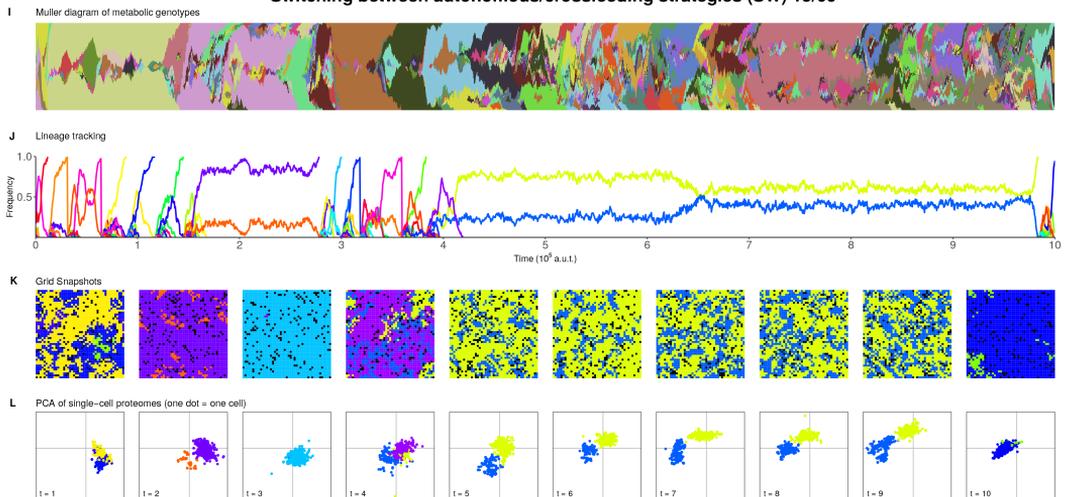


Figure 3.2: Emergence of diverse metabolic strategies **a-d** Example of population dominated by a single autonomous lineage. **a.** Muller plot showing relative frequencies of different metabolic genotypes. Clades of microbes with different metabolic genotypes (colours) continuously evolve and results in complex evolutionary dynamics of competition and leapfrogging throughout the experiment. **b.** Tracking ancestral relationships with renewing lineage markers shows a continued turnover in markers, indicating that at any point during the simulation all microbes have a recent common ancestor. **c.** Snapshots of spatial environment. **d.** Principal component analysis of single-cell proteomes shows that in these communities all microbes express similar proteins. **e-h** Example of a population that diversifies in two lineages that cross-feed on essential building blocks. **g.** Lineages form an interleaved pattern in the spatial environment (see Supplementary Fig. 6 and Supplementary Movie 1). Single-cell proteomes (h) show that these lineages express different metabolic enzymes. **i-l** Example of a population that switches between autonomous and cross-feeding strategies. Lineage markers are redistributed when a single marker fixes in the whole population. PCAs coloured for lineage markers, and composed per simulation on relative single-cell protein expressions, see methods for details.

simulations, point mutations fix that tune fluxes through specific reactions, and metabolic networks are extended with reactions that are dedicated to producing energy – which allows increased food uptake – and reactions that process byproducts for more energy and/or building blocks. Furthermore, importers are recruited to recycle building blocks that accumulate in the environment through cell lysis. Thus eventually, all populations evolve efficient, closed metabolic networks that make use of all produced metabolites.

However, in all replicate simulations, different metabolic genotypes continuously arise and compete for dominance within the population (**Figure 3.2**). The dynamics are complex, and characterized by clonal interference and frequent hitchhiking, leapfrogging, and horizontal gene transfer. To condense these complex dynamics, we use lineage tracking and analyse the cell proteomes of these lineages. This reveals that some of these heterogeneous populations are dominated by a single lineage that performs all the metabolic functions outlined above by itself (from here on referred to as the autonomous strategy, AU), while other populations diversify in two complimentary, cross-feeding lineages (CF) that specialize in producing one, and importing the other building block (**Figure 3.2a-b**). These lineages in CF populations coexist for tens of thousands of generations, typically until the end of the simulation, but in some simulations single or repeated switching between AU and CF strategies occurs (SW populations, see **Figure 3.2c**).

To investigate the nature of the cross-feeding interaction we examined whether lineages could survive in absence of the other. Specifically, at different time points after the two lineages emerged, we tested metabolic dependencies in the standing diversity of CF populations by removing all microbes from one lineage, and preventing further mutations to occur in the remaining lineage (**Figure 3.3a-c**). We find that both lineages nearly always (98.8% of cases tested) grow faster in the presence of their partner (**Figure 3.3d-e**, **Suppl. Figure S3.2**), supplementing their own metabolism with building blocks produced by the other lineage. Generally, cross-feeding communities consist of a major lineage that produces both building blocks and can survive by itself, and a minor lineage that is obligately dependent on the major lineage for one of the building blocks and goes extinct when the major lineage is removed (**Figure 3.3c**). These dependencies are not constant during the simulation, but can increase, decrease, completely switch direction, and

change to fully co-dependent, as reflected by large changes in the size of population bottlenecks of surviving lineages that follow lineage removal (**Figure 3.3c**), and changes in the relative frequencies of CF lineages in the main experiment (**Figure 3.2b**, **Suppl. Figure S3.1**)

At any given time during a simulation, mutants with the opposing strategy can be found within a community, but even though simulations last tens of thousands of generations, communities only switch strategy a couple of times and most communities (45/60) do not switch at all (**Figure 3.2**, **Suppl. Figure S3.1**). For example, in cross-feeding populations autonomous mutants can easily evolve via horizontal gene transfer between lineages with complementary metabolic networks. When a cross-feeding lineage is removed, such mutants in the remaining lineage can successfully take over and found a new, completely autonomous population, but their growth rates are higher in the context of the original cross-feeding population, where they exploit the environment created by the whole ecosystem (**Figure 3.3b-d**, **Suppl. Figure S3.3**). This explains why they cannot replace the resident CF community in the main evolutionary experiment. Similarly, in AU simulations, gene loss produces mutants that specialize on producing or importing only one building block, but these are only transiently present and fail to invade in the AU population that imports and produces both building blocks. Apparently, AU and CF are stable eco-evolutionary attractors, with only rare occurrences of populations switching between them. Since all simulations started from the same ancestral community, this shows evolutionary contingency determines what kind of community evolves. Before we further consider the consequences for predicting evolution, we first need to understand exactly what causes which strategy evolves.

3.2.2 The evolution of cross-feeding is not explained by protein cost

The evolution of cross-feeding in our simulations is characterized by loss of genes for building block synthesis and/or transporters, and results in smaller genomes for CF compared to AU strategies. As we assume an explicit cost for expressing proteins, and producing building blocks is an “inescapable public good” as they are required for growth and released into the environment when cells die, evolution of cross-feeding could thus be driven by Black Queen dynamics (Morris et al., 2012; D’Souza et al., 2014; Pande and Kost, 2017). To test this, we study the effect of protein expression costs on the evolution of metabolic strategies. Surprisingly, the emergence of AU, CF and SW strategies is robust to increasing or decreasing the costs of proteins expression an order of magnitude (**Suppl. Figure S3.4**). Although some of the dynamics change (for example, lower expression costs allow larger genomes to evolve, and increased expression costs causes the evolutionary dynamics to slow down), all strategies evolve under all conditions, and are stable eco-evolutionary attractors. Thus, even though in our model the production of building blocks acts as a public good, and protein expression has an explicit cost that can be reduced by gene loss, the evolution of cross-feeding is not driven by gene loss to escape this cost.

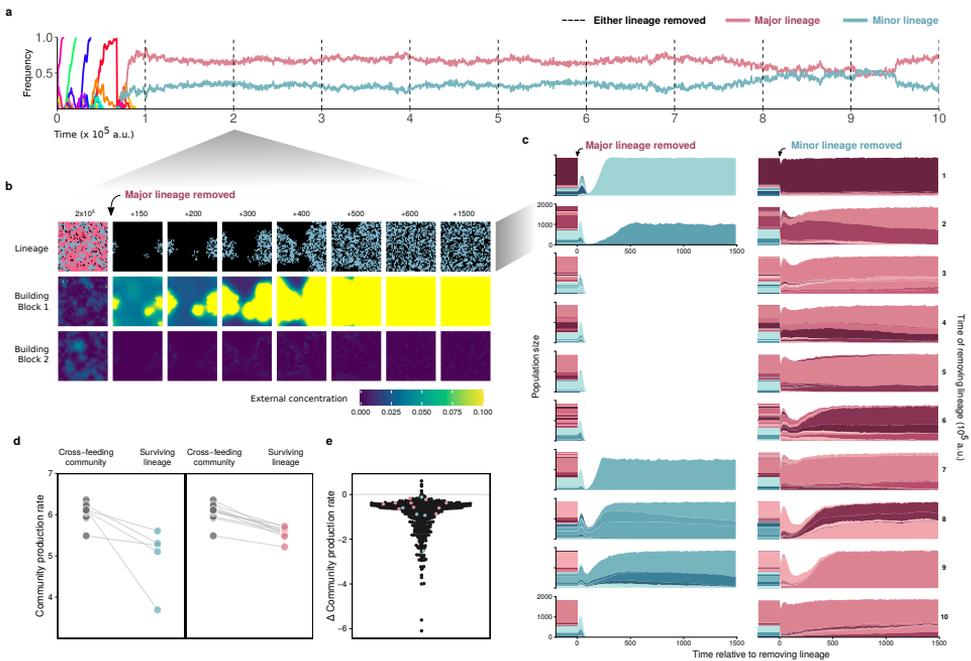


Figure 3.3: Metabolic dependencies in cross-feeding communities. We tested metabolic dependencies in 29 selected populations by removing either lineage at different time points after cross-feeding evolved, and without allowing further mutations to occur. **a-d.** Example of 2×10 tests of metabolic dependency in replicate population 23. Times indicated with dashed lines in (a). **b.** When removing the major lineage (pink) at $t = 2 \times 10^5$, most microbes of the remaining lineage (blue) die out. However, a rare mutant is able to grow by itself, though it cannot import building block 1 and does not reach a high abundance. **c.** Outcome of removing lineages for all time points in (a), with different metabolic genotypes within each lineage indicated with shades of the lineage colour. Typically, the minor lineage goes extinct or contains only few mutants that survive in isolation, reflecting obligate dependency on the major lineage. In contrast microbes in the major lineage can mostly survive without the minor lineage. These dependencies are not constant over evolutionary time as metabolic genotypes that dominate within each lineage change. Note that directly following lineage removal, all remaining lineages can initially quickly grow on the limited store of building blocks that were produced by partner lineage and are still present in the environment. **d.** Community production rates before and 1500 time steps after lineage removal. All surviving minor and major lineages have higher growth rates in the context of the original cross-feeding population. **e.** Difference in community growth rate for surviving lineages in 484 tests of metabolic dependency in 29 populations. 407 out of 412 (98.8%) tested lineages that survive have reduced growth rates in isolation. Surviving lineages from a-d are highlighted.

3.2.3 Trade-offs emerge during the evolution of metabolic networks

To look for signatures for CF and AU strategies, we further investigated the diversity of metabolic networks that evolved. First, we clustered populations based on metabolic gene frequencies at the end of the simulation (see **Figure 3.4a**). This shows that all populations share a core set of 5 genes for the uptake of the food resource, and production and uptake of both building blocks. For certain replicates, the genes for production and uptake of building blocks are only present in a subset of the population, reflecting how the cross-feeding communities have a distributed metabolic network. Note that populations strongly differ in which reaction is recruited to produce energy, and how byproducts from this reaction are further metabolized. Typically a single dedicated energy reaction fixes in a population. Although clustering is dominated by individual energy generating reactions which clusters AU and CF populations with a few exceptions (**Figure 3.4a**), no single gene acts as a signature for AU or CF community type.

3

Next, we consider how these gene frequencies change during the complete evolutionary trajectories by PCA (**Figure 3.5, Suppl. Figure S3.5**). We find that the major component separates CF and AU communities, and that SW populations move along this axis accordingly when switching strategy. Moreover, based on the energy reactions recruited by each strategy, the metabolic networks can be classified in two different topologies that associate exclusively with either strategy: networks that degrade the food metabolite for energy (*i.e.* $R \rightarrow \text{energy} + \text{byproduct}$) for CF communities, and networks that degrade either building block for energy (*i.e.* $B1$ or $B2 \rightarrow \text{energy} + \text{byproduct}$) for AU communities (**Figure 3.4b**). Interestingly, SW communities are composed of microbes with metabolic networks that have a hybrid topology (*i.e.* they degrade building blocks as well as food for energy), or are rare cases of simulations where one topology invades and replaces the other (**Figure 3.4, Figure 3.6**).

Why do these topologies determine metabolic strategies? The amount of energy available to a microbe is limited, and as a consequence, importing more of one metabolite trades off with importing others, depending on what metabolite is used as an energy precursor. If microbes create energy by degrading the food resource, taking up other metabolites such as building blocks lowers the cell's energy budget. The amount of additional metabolites that can be imported is thus constrained for metabolic networks with this topology. As building blocks are produced and accumulate in the environment, this creates two niches (*i.e.* one for each building block) that can be exploited by different lineages. In contrast, when building blocks are degraded for energy, importing them increases the energy budget and does not trade off with importing the food resource, allowing individual microbes to import both building blocks as well as retain their competitive ability for the food resource. Furthermore, as only one of the two building blocks is used for energy in AU communities, a cross-feeding scenario with this topology would be inherently asymmetrical and unstable, as one lineage would be dependent on the other for both energy and building blocks, while the other lineage only requires the complementary building block.

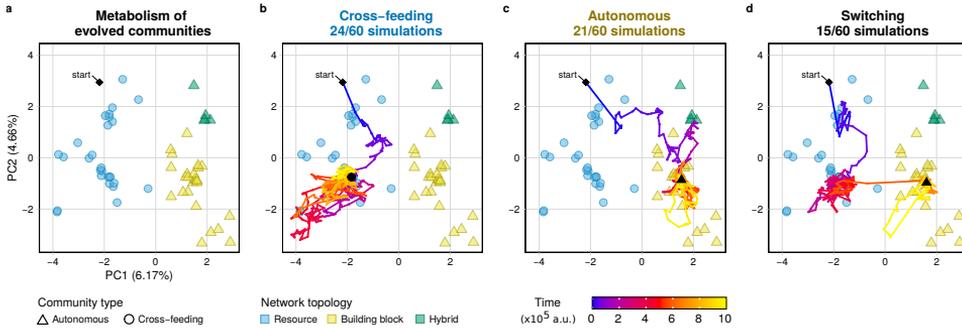


Figure 3.5: Evolutionary trajectories towards community attractors **a.** PCA of gene frequencies of 59 metabolic genes in 58 communities over the whole duration of the experiment. One dot represents one community. For clarity only the initial community and final time point of the simulations are shown. This separates communities by strategy along the first component, and reveals that topology of the evolved metabolic network determines metabolic strategy of the community: networks with reactions that degrade resource R for energy cross-feed on building blocks, whereas networks with reactions that degrade building block B_1 or B_2 for energy remain metabolically autonomous and consume all building blocks from the environment. **b-d.** Evolutionary trajectory showing all time points in the PCA for **b** a community that evolves cross-feeding (community 9), **c** metabolic autonomy (community 4), and **d** switches between strategies (community 18). For visualisation purposes outlier populations 59 and 47 were omitted from this analysis (see Supplementary Fig. 5 for analysis including these outliers)

when we stop mixing, only populations with the associated topology quickly diversify in two cross-feeding lineages, signifying that it is the interplay between environmental structure and evolved metabolic constraints that drives cross-feeding.

Finally, reasoning that long-term coexistence might result in increased robustness of the cross-feeding interaction, we tested the ecological and evolutionary stability of cross-feeding communities from the original experiment by transfer to a well-mixed medium. Specifically, we subjected 7 randomly chosen CF populations to well-mixed conditions at varying time steps after metabolic exchange evolved, while either allowing or preventing further mutations to occur. In all ‘ecology-only’ tests (*i.e.* without mutation), cross-feeding is stably maintained and population size and community productivity increases. The increased productivity makes intuitive sense, as less productive mutants are outcompeted and cannot re-appear due to lack of mutations. Moreover, under unmixed conditions local reproduction and metabolite diffusion limit the interface between both lineages and therefore reduce efficient exchange of building blocks.

In contrast, we find that all the evolutionary simulations under mixed conditions are quickly taken over by autonomous mutants. Strikingly, the resulting AU communities have smaller population sizes and productivity than their ancestral CF community. This shows that while spatial structure puts an upper limit to the efficiency of cross-feeding, it also protects against autonomous metabolic strategies. Consistent with previous results (Stump et al., 2018b; Pande et al., 2016), we find that spatial structure is needed to evolve and maintain metabolic cross-feeding, but also find that whether cross-feeding evolves or not depends on constraints of previous metabolic adaptations. As microbes evolve to produce more energy from either the resource or one of the building blocks, importing one metabolite trades of

with importing others. We find that shape of this trade-off is an evolved property of the metabolic network and the local environmental niches they construct.

3.2.5 Metabolic strategies are an evolutionary contingency

Our results demonstrate that the topology of the evolved metabolic network, combined with spatial structure, determines whether metabolic exchange evolves or not. Which topology evolves in a population is arbitrary, and often establishes early on. For simulations where the cost of protein expression is increased, the topology often fixes up to tens of thousands of generations before metabolic networks ‘mature’ by making use of all building blocks that accumulate in the environment. What eco-evolutionary strategy will evolve when microbes finally evolve to tap into that source can be predicted from the evolved topology (see **Figure 3.6b**), realizing a fate already cemented earlier in its evolutionary history. However, it is interesting to note that exact prediction is limited by several factors. Firstly, while evolution of the topology of the metabolic network is typically “founder controlled”, where the energy reaction that establishes itself first in the community quickly accumulates more beneficial mutations and is never outcompeted by other energy reaction genes that are discovered later on, in rare cases mutants with a different energy type can invade and replace the original population, changing community fate (**Figure 3.6a**). Secondly, microbes that have a hybrid metabolism can switch between strategies as they evolve and different reactions dominate the metabolic network, finally remaining in either strategy if one type of energy reaction is lost (**Figure 3.6a**).

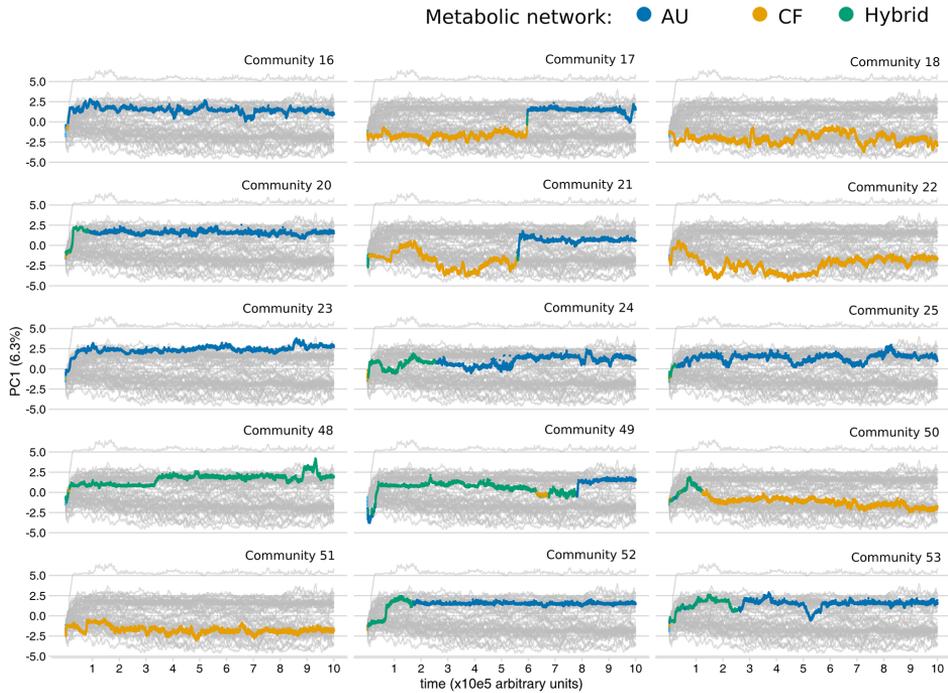
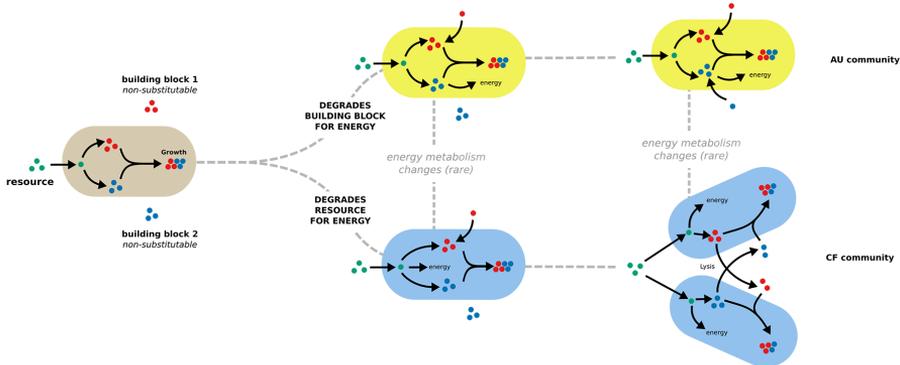
a Examples of evolutionary trajectories and topology of metabolic network

b Cartoon of evolutionary dynamics and ecosystem fate


Figure 3.6: Prior metabolic adaptations constrain future ecological roles. **a.** Evolutionary trajectories of example communities (first component from PCA in Fig. 3.5 v.s. time) towards cross-feeding (negative y-value) or autonomous (positive y-value) strategy, coloured for topology of the metabolic network. Grey lines indicate trajectories of all other communities. Changes in dominant network topology cause a switch in community strategy. **b.** Cartoon of evolutionary trajectories. Earlier metabolic adaptations that fix in the initial population dictate final eco-evolutionary attractor, but are an evolutionary contingency. However, prediction is limited because the duration of each depicted stage is unpredictable, and cases where mutants with alternate network topology invade and replace the population (dashed arrow in **b**, population 305 and 402 in **a**) are possible.

3.3 Discussion

Previous experimental studies have shown repeated evolution of cross-feeding (Rosenzweig et al., 1994; Treves et al., 1998; Pfeiffer and Bonhoeffer, 2004; Rozen et al., 2009; D’Souza and Kost, 2016), and the emergence of cross-feeding is typically interpreted as driven by physiological constraints. By explicitly incorporating constraints such as trade-offs in resource uptake or production (Großkopf et al., 2016), or a limit in the maximum number of reactions within a single organism (Thommes et al., 2019), theoretical approaches demonstrate how these constraints can force metabolic specialization and drive the evolution of cross-feeding. Here, rather than pre-defining such physiological constraints in our model, we demonstrate how metabolic constraints can themselves evolve, and whether or not they are driving cross-feeding is a historic contingency.

Next to metabolic constraints, other factors important for the evolution of metabolic exchange have been previously identified. For example, in batch culture repeated cycles of feast and starvation generate temporal niches that explain stable coexistence of cross-feeding ecotypes (Rozen et al., 2009; Le Gac et al., 2012; Großkopf et al., 2016; Rocabert et al., 2017; van Dijk et al., 2019). In these studies cross-feeding is unidirectional, *i.e.* one lineage feeds on metabolic byproducts of the other. In contrast, we here show bidirectional cross-feeding of two non-substitutable building blocks, and in a constant environment. This matches recent experimental observations of the evolution of cross-feeding on essential nutrients (D’Souza and Kost, 2016), and stable coexistence mediated by uptake of debris from lysed cells of a coexisting lineage (Rozen et al., 2009).

Another suggested driver for the evolution of cross-feeding is that the collective exploitation of resources can be more efficient or productive than individual-based solutions (Rosenthal et al., 2011; Alessi et al., 2018; Thommes et al., 2019; Yang et al., 2020). In our simulations, autonomous mutants in CF communities are indeed less productive than the resident cross-feeders (**Figure 3.3** and **Suppl. Figure S3.2 and S3.3**). However, communities that evolve a completely AU strategy are equally or more productive than CF communities, showing that productivity alone may not be sufficient to explain widespread metabolic dependencies observed in nature. Thus, we should realize that community productivity itself is an evolved property of the evolved metabolism, and that division of labour is not necessarily more productive, as productivity depends on evolved metabolic constraints. Metabolic dependencies can also arise due to non-adaptive gene loss caused by genetic drift in small populations (Kuo et al., 2009). In our model, this seems not to be the case as cross-feeding has a clear adaptive benefit (**Suppl. Figure S3.2**), and lost functions are easily regained by HGT.

Recent experimental work showed that initially clonal microbial communities quickly become genetically heterogeneous (Ba et al., 2019; Good et al., 2017), and that in the resulting complex evolutionary dynamics, the fate of individual mutations does not only depend on that specific mutation’s fitness but mainly on what new mutations accumulate in that background (Ba et al., 2019). Likewise, we find that when no further mutations are allowed to occur, AU mutants already present in CF population never fixate in the well-mixed selection regime, but when they can acquire more mutations always take over and dominate evolutionary outcome. In other

words, the outcome of evolution (*i.e.* the type of community that evolves) depends critically on interlocking evolutionary and ecological timescales, recapitulating earlier theoretical work (Van Der Laan and Hogeweg, 1995; Kotil and Vetsigian, 2018).

It is interesting to note that in each cross-feeding strain more mutants gain partial dominance than is the case in autonomous lineages (**Figure 3.2**). The amount of mutations in synthetic cross-feeding versus autonomous strains was recently studied in (Lilja and Johnson, 2019). Contrasting with our results, they found that although, under strong selection, the sum of the mutations in the two cross-feeding strains was larger and more varied than in the autonomous strain, the separate cross-feeding strains acquired less mutations than the autonomous strain. We conclude that the speed and trajectory of evolutionary dynamics in relation to co-evolving, interdependent lineages under various selection regimes is a promising avenue of future computational, theoretical and experimental research.

3

For computational feasibility and simplicity we here considered a simple metabolism with only two essential building blocks, while in reality many metabolites such as carbon sources, nitrogen sources, amino acids, nucleotides, and vitamins are required for growth, and also exchanged in microbial communities (D'Souza et al., 2018). It would be interesting to extend the model metabolism with more building blocks and investigate whether the here reported distinct community types (AU and CF) are maintained, or whether communities with intermediate levels of metabolic division of labour evolve. Interestingly, this dichotomy in community types matches a recent paper that investigated the metabolism of natural communities composed of up to 40 species. They reported that natural communities in a wide range of environments are either highly cooperative, with smaller genomes and diverse metabolic dependencies, or highly competitive, and dominated by species with larger genomes and overlapping nutritional needs (Machado et al., 2020). This suggests the dichotomy we here find is a general property of microbial communities. Furthermore, it would be interesting to investigate whether the rapid switching between CF and AU community types we find, also occurs in natural communities. It is exciting to see that patterns emerging in the eco-evolutionary dynamics of our multi-level Virtual Microbe model are now detectable through metagenomic analysis of natural microbial communities. We look forward to future cross-feeding of the two approaches to enhance our understanding of the microbial world.

3.4 Methods

3.4.1 Classification of metabolic types and construction of Muller diagrams

Populations were highly heterogeneous throughout evolutionary simulations, with a single or a couple genotypically identical individuals present at any time. We found it useful for interpretation and visualization to classify microbes based on presence and absence of all 59 metabolic genes on their genome, ignoring gene copy number, promoter strength and kinetic parameters. At simulation time we tracked gain and loss mutations that changed this metabolic genotype, and collected abundances and phylogenetic relationships of the clades. For creating Muller diagrams we subsampled the data every 500 time steps, and used a cutoff of 5% of maximum population size to remove metabolic type that never became abundant. We follow the convention to stack every new clade emerging from the center of its parent genotype in chronological order of appearance.

3.4.2 Lineage tracking with renewing markers

To allow tracking of ecological and evolutionary dynamics we used a technique similar to barcoding methods used in experimental evolution (Blundell and Levy, 2014). We label each microbe in the initial population with a unique neutral marker that it passes on to their offspring. If, at any point during the simulation, all living microbes have the same marker - indicating that they share a common ancestor that was alive when markers were distributed- markers are renewed by giving each microbe a new unique marker. Note that these markers only track lineage dynamics and do not correspond directly to the metabolic genotypes shown in the Muller plots. Markers do not reflect genetic or phenotypic identity nor provide high-resolution insight in the thousands of independent adaptive and nonadaptive mutations competing within a population.

3.4.3 Analysis of within-population diversity (PCA of single-cell proteomics)

To investigate different metabolic strategies of individual microbes within a community, we measured differential investments across all metabolic reactions by normalizing protein concentrations of each metabolic protein to the total protein concentration in the cell. We collected protein data from all cells for time points $(1, 2, \dots, 10) \times 10^5$, and then performed a separate PCA for each simulation.

3.4.4 Testing metabolic dependency

To assess metabolic dependency of microbes in cross-feeding populations, we removed all microbes in one lineage (including internal metabolites) and set mutation rates to zero. Following removal we continued the simulation for 2000 time steps or population extinction. For all CF populations in the main experiment we tested both lineages at every multiple of 10^5 time steps after cross-feeding evolved. For Fig. 3.3 we coloured the diversity of metabolic genotypes within each lineage in different shades of the lineage colour. Production rates of microbes were measured

at the time step preceding just lineage removal, and, if populations survived, after 2000 time steps following removal.

3.4.5 Analysis of between-population diversity (PCA of metagenomics)

We performed a single principal component analysis on the gene frequencies in all replicate populations for the whole duration of experiment. For clarity of visualisation we omitted two outliers (population 47 and 59) from this analysis, which did not impact the result (see Supplementary Figure 3.5 for an analysis including these outliers).

3.4.6 Full model description

Virtual Microbes was previously described in (van Dijk et al., 2019), and extends an earlier model (Cuypers and Hogeweg, 2012, 2014; Cuypers et al., 2017). It is a highly configurable framework designed to study emergent, eco-evolutionary microbial dynamics under a wide range of possible conditions. For clarity we here only describe the configuration and features used in the present study. The full documentation, installation instructions and source code are available at <https://bitbucket.org/thocu/virtualmicrobes>.

Overview Virtual Microbes is an agent-based model where each individual microbe occupies a grid point on a 2d plane. Population dynamics (*i.e.* competition, reproduction and death) play out every time step. Within a time step, cellular processes (cellular growth, gene expression, metabolic reactions including uptake, excretion, and diffusion across the cell membrane, metabolite and protein decay) and environmental metabolite dynamics (*i.e.* food influx, degradation, and diffusion between grid points) are updated using ordinary differential equations (ODEs).

Chemical universe The chemical universe is the *a priori* defined set of all molecules and possible reactions between them. We procedurally generated an artificial biochemistry consisting of 9 molecules (a single resource R , two designated building block molecules B_1 , B_2 , energy carrier E , and 4 intermediate metabolites M_{1-5}) and 59 metabolic reactions. This provides evolution considerable degrees of freedom for forming metabolic networks to produce the required building blocks and energy from provided resource. See **Supplemental Table 3.2** for the full set of reactions, and the online repository for a config file used to generate the biochemistry.

Transport For all metabolites, transporters exist that import or export the metabolite across the cell membrane. Transporters \mathcal{T} catalyse the transport of substrate S over the cell membrane by consuming energy metabolite E . Transport rate v is then given by Michaelis-Menten kinetics:

$$v = v_{max_{\mathcal{T}}} \cdot \mathcal{T} \cdot \frac{S \cdot E}{(S + K_S) \cdot (E + K_E)} \quad (3.1)$$

where K_S and K_E are the binding affinity constants of respectively the substrate and energy molecule, to the transporter. Depending on the direction of transport

(importing or exporting) S is either the external or the internal concentration of the substrate. $v_{max\mathcal{T}}$, K_S and K_E are all evolvable parameters of the focal transporter gene.

Metabolism Metabolic enzymes catalyze reactions of the general form



converting reactant metabolites $\{R_i, \dots\}$ to products $\{P_j, \dots\}$. The rate of catalysis v is calculated with Michaelis-Menten kinetics as follows:

$$v = v_{max\mathcal{E}} \cdot \mathcal{E} \cdot \frac{\prod_{R \in \mathcal{R}} R}{\prod_{R \in \mathcal{R}} (R + K_R)}$$

where \mathcal{E} is the concentration of the enzyme catalysing the reaction, \mathcal{R} the set of all reactant metabolites, and K_R and $v_{max\mathcal{E}}$ are evolvable kinetic parameters of enzyme \mathcal{E} .

Degradation and dilution of molecules Concentrations of all molecules (*i.e.* proteins, product, and metabolites) are adjusted according to the change in cell volume $dVol/dt$. In addition all molecules degrade with a molecule-specific rate *degradation*.

Production budget Virtual Microbes convert building blocks B_1 and B_2 to a production budget \mathcal{B} with rate *production*, favoring homeostasis of building block concentrations with:

$$production = B_1 \cdot B_2 \cdot \frac{1}{1 + (|B_1 - \frac{B_1+B_2}{2}| + |B_2 - \frac{B_1+B_2}{2}|)}$$

This budget is spent for increasing cell size $cost_{growth}$, and producing proteins by gene expression $cost_{expr}$. Note that to maintain sufficient selection pressure to select for adaptive mutations as building block synthesis rates of populations increase during evolutionary simulations, we scale the amount of budget spent with $\mathcal{B}_{scaling}$:

$$\mathcal{B}_{scaling} = \frac{\mathcal{B}}{\mathcal{B} + \mathcal{B}_{pop}}$$

where \mathcal{B}_{pop} is the time averaged population production value. The concentration of \mathcal{B} then changes with a rate:

$$\frac{d\mathcal{B}}{dt} = production - cost_{growth} - cost_{expr} - degradation - dilution$$

When cells grow, \mathcal{B} is consumed proportional to the rate of growth

$$cost_{growth} = Vol_{growth} \cdot \mathcal{B}_{scaling} \cdot \mathcal{B}$$

Total consumption of product for gene expression is summed over all genes:

$$cost_{expr} = \sum_{i=1}^{N_{genes}} Pr_i \cdot \mathcal{B}_{scaling} \cdot \mathcal{B}$$

Cell volume growth We assume that if Virtual Microbes grow without being able to divide due to lack of space they approach a maximum cell size Vol_{max} with rate g . We implement a cost for maintaining cell volume with a small shrinking rate s . Volume changes with a rate

$$\frac{dVol}{dt} = Vol_{growth} - s \cdot Vol$$

Where Vol_{growth} is given by

$$Vol_{growth} = g \cdot Vol \cdot \frac{1 - Vol}{Vol_{max}} \cdot \mathcal{B}_{scaling}$$

Protein expression Protein concentrations \mathcal{P} for any given gene are governed by the function:

$$\frac{d\mathcal{P}}{dt} = Pr \cdot \mathcal{B}_{scaling} - degradation - dilution$$

where Pr is the per gene evolvable parameter promoter strength. In addition, Virtual Microbes features explicit transcription factor proteins that can modify protein transcription. TFs were not relevant for this study and not recruited during evolutionary simulations, and are here omitted for clarity. See the full documentation for a description of TFs.

Toxicity and death Microbe death is a stochastic process depending on an intrinsic death rate $r = 0.03$, which is potentially increased when internal metabolite concentrations reach a toxic threshold. This cumulative toxic effect e_{tox} is computed over all internal metabolites M and the current life time τ of a microbe as

$$e_{tox} = \sum_{m \in M} \int_{t=0}^{\tau} f(m, t) dt$$

with the toxic effect function $f(m, t)$ for the concentration of metabolite m at time t with metabolite-specific toxicity threshold tox_m

$$f(m, t) = \max(0, \frac{m_t - tox_m}{tox_m})$$

This toxic effect increases the death rate d of microbes starting at the intrinsic death rate r

$$d = \frac{e_{tox}}{s + e_{tox}} \cdot (1 - r) + r$$

where s scales the toxic effect. Microbes that survive after an update cycle retain the toxic level they accumulated so far. If cell size drops below a minimal cell size threshold, cells are automatically marked for death.

Reproduction Microbes grow and reproduce on a 45×45 grid, competing for metabolites and empty space. A microbe can reproduce into an empty grid point in the Moore neighborhood if the minimal cell size is met. When multiple microbes are eligible for reproduction in a single empty space, reproduction chance depends

on the microbes' production value P , with a dynamically scaled chance for a “no reproduction” event occurring. Cell volume is divided equally between parent and offspring, *i.e.* molecule and protein concentrations remain constant. Toxic effects built up during the parent's lifetime do not carry over to offspring.

Genome and mutations The genome is a linear organisation of genes coding for enzymes, transporters and transcription factors. When a microbe successfully reproduces it divides and the offspring inherits a copy of the parent's genome, which can then be subject to various types of mutations. Stretches of genes can be duplicated, deleted, inverted, or translocated to another position on the genome (**Table 6.1**). At the single gene level, all evolvable parameters can mutate individually (**Table 6.1**) Horizontal gene transfer can occur on every time step. Innovations are modeled as HGT from an external (off-grid) source, and can occur every time step.

Initialization and experimental setup We constructed an initial population of 2025 “minimally viable” Virtual Microbes by generating genomes containing one importer gene for the food resource and 5 random enzymes, with the constraint that the resulting metabolic network was at least capable of producing the necessary building blocks and energy from the provided resource. Genes were randomly parameterized and ordered on the genome. 69 identical copies of this population were propagated in parallel for $1 \cdot 10^6$ time steps corresponding to $\sim 4 \cdot 10^5$ generations (or until they reached their final eco-evolutionary attractor, which took up to $5.34 \cdot 10^6$ time steps for some simulations) under the exact same conditions, differing only in the mutation seed.

Mutation	Description	Probability
Duplication	A stretch of 1 or more genes is duplicated in tandem	0.001
Deletion	A stretch of 1 or more genes is deleted	0.001
Inversion	A stretch of 1 or more genes is inverted in order	0.001
Translocation	A stretch of 1 or more genes is moved to a random location on the genome	0.001
Stretch length	Geometrically distributed with $p = 0.3$	-
Gene discovery	Per time-step probability of discovering a new (randomly parameterized) gene.	0.0001
Horizontal Gene Transfer	Per time-step probability of copying a gene from a cell in a neighboring site	0.0005
Point mutation	Per gene per generation probability of modifying a single parameter of a gene (promoter strength, Michaelis Menten constants)	0.02
Parameter	Gene Types	Value range
Promoter Strength	Enzyme, Transporter	[0.01, 8]
$K_{substrate}$	Enzyme, Transporter	[0.01, 8]
K_{energy}	Transporter	[0.01, 8]
V_{max}	Enzyme, Transporter	[0.01, 8]
Exporting	Transporter	[True, False]

Table 3.1: Types of mutations and their probabilities

3.4.7 Code availability

Full documentation, installation instructions, and source code for Virtual Microbes are available at <https://bitbucket.org/thocu/virtualmicrobes>. Configuration files to re-run the evolutionary simulations and scripts to analyze the data will be made available at publication.

Acknowledgements

This work was supported by the EvoEvo project (European Commission 7th Framework Programme (FP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies, ICT-610427). J.M. was supported by the NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, www.nwo.nl) grant 022.005.023, and the Utrecht University BINF fund for theoretical research in bioinformatics. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Virtual Microbes was written by Thomas D. Cuyper and Bram van Dijk.

Author contributions

J.M. wrote the manuscript with input from B.v.D. and P.H.; J.M., B.v.D. and P.H. contributed equally to the experimental design, analysis and interpretation of results.

3.5 Supplementary Material

Metabolite properties						
Label	Class	Influx rate	Degr rate	Toxicity	Energy	Diff rate
R	Provided	0.002	0.0003	0.077	9	0.011
B2	BB	0	0.0100	0.103	8	0.012
B1	BB	0	0.0100	0.058	8	0.012
E	Energy	0	0.1000	0.065	1	0.050
M1	-	0	0.0006	0.105	6	0.014
M2	-	0	0.0003	0.080	4	0.019
M3	-	0	0.0006	0.079	7	0.013
M4	-	0	0.0014	0.158	6	0.014
M5	-	0	0.0008	0.047	5	0.016

Reactions			
R → B2 + E	B2 → M1 + 2 E	M2 + M3 → B1	2 M2 → M4
R → B1 + E	B2 → M2 + 4 E	M2 + M3 → B2	
R → M1 + 3 E	B2 → M3 + E	M2 + M5 → M4	M3 + M5 → R
R → M2 + 5 E	B2 → M4 + 2 E	M2 + M4 → M3	M3 → M5 + 2 E
R → M4 + 3 E	B2 → M5 + 3 E	M2 + M5 → M3	M3 → M4 + E
R → M5 + 4 E		M2 + M4 → B2	
R → M3 + 2 E	M1 + M4 → R	M2 + M5 → B2	M4 → M2 + 2 E
	M1 + M2 → M3	M2 + M5 → R	M4 + M5 → B2
B1 → M1 + 2 E	M1 → M2 + 2 E	M2 + M5 → B1	M4 + M5 → R
B1 → M2 + 4 E	2 M2 → M5	M2 + M5 → M1	M4 + M5 → B1
B1 → M4 + 2 E	2 M2 → M1	2 M2 → B2	
B1 → M5 + 3 E	2 M1 → R	2 M2 → B1	M5 → M2 + E

Table 3.2: Chemical universe Using Virtual Microbes we generated a biochemistry comprised of 9 metabolites and 59 reactions, with one provided resource (R), two non-substitutable building blocks (BBs) and a single energy molecule (E). See Code availability for config files used for generating this biochemistry.

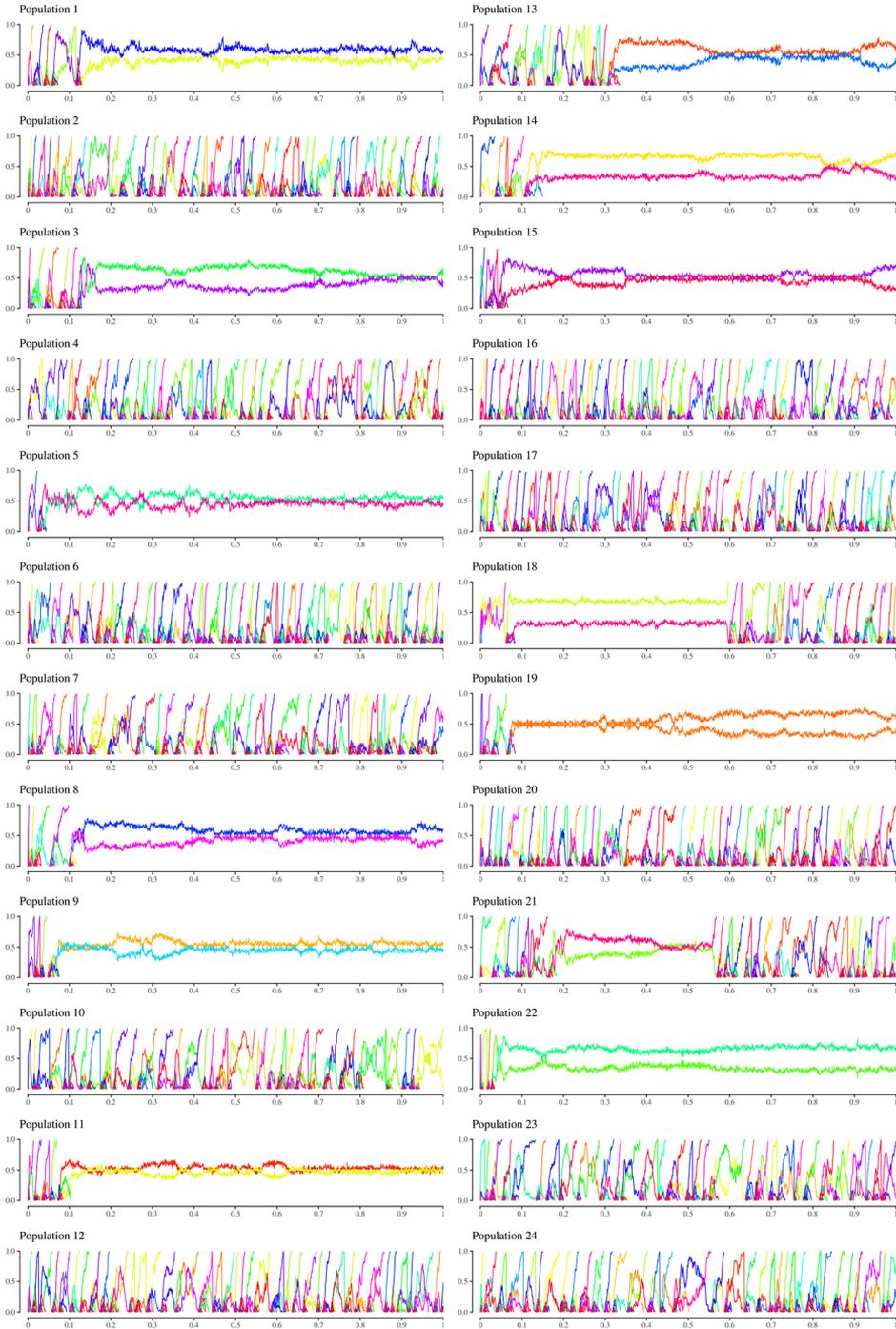


Figure S3.1: Neutral lineage markers for all populations of the main evolutionary experiment, displayed like in fig.2 of main text..

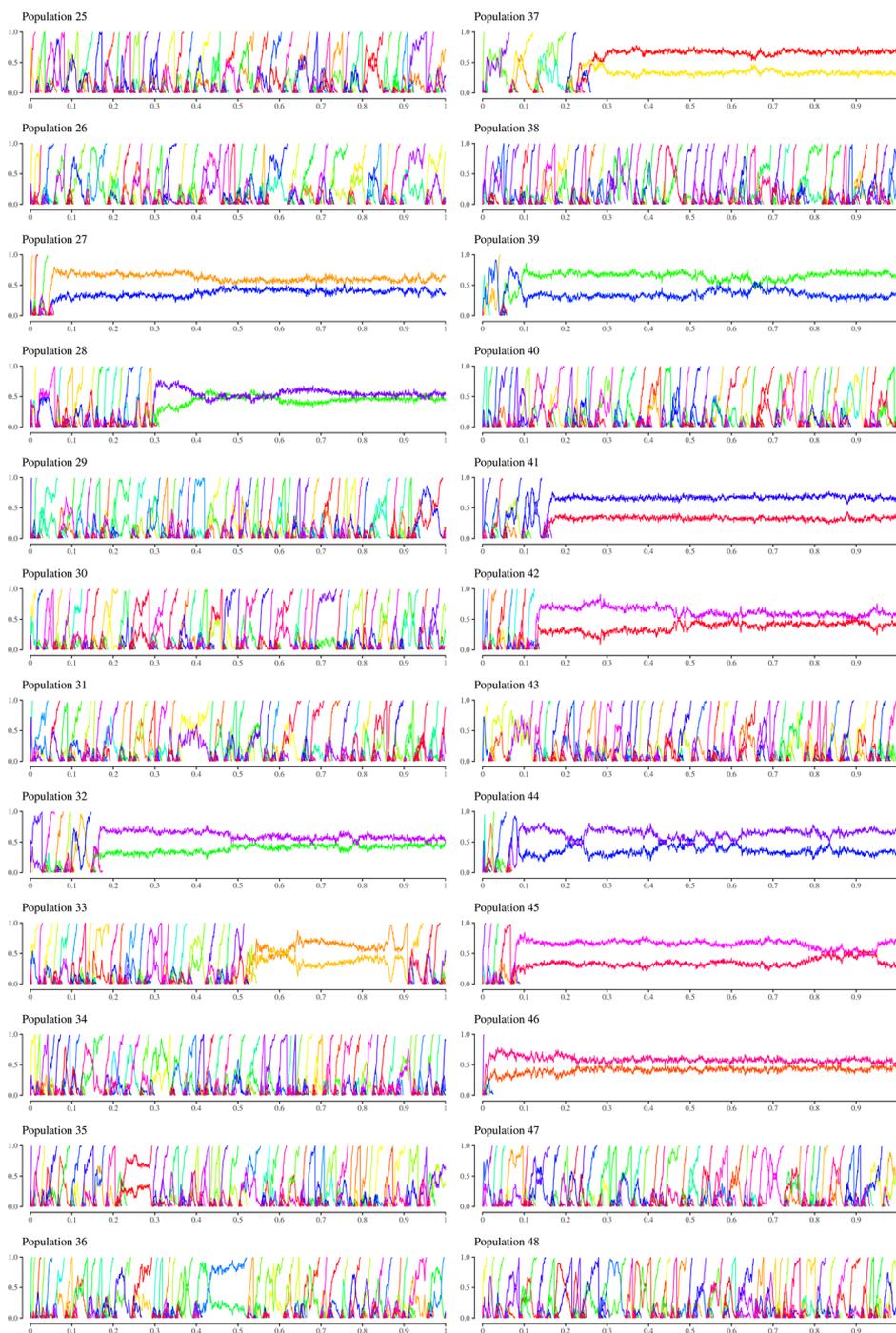


Figure S3.1: (continued).

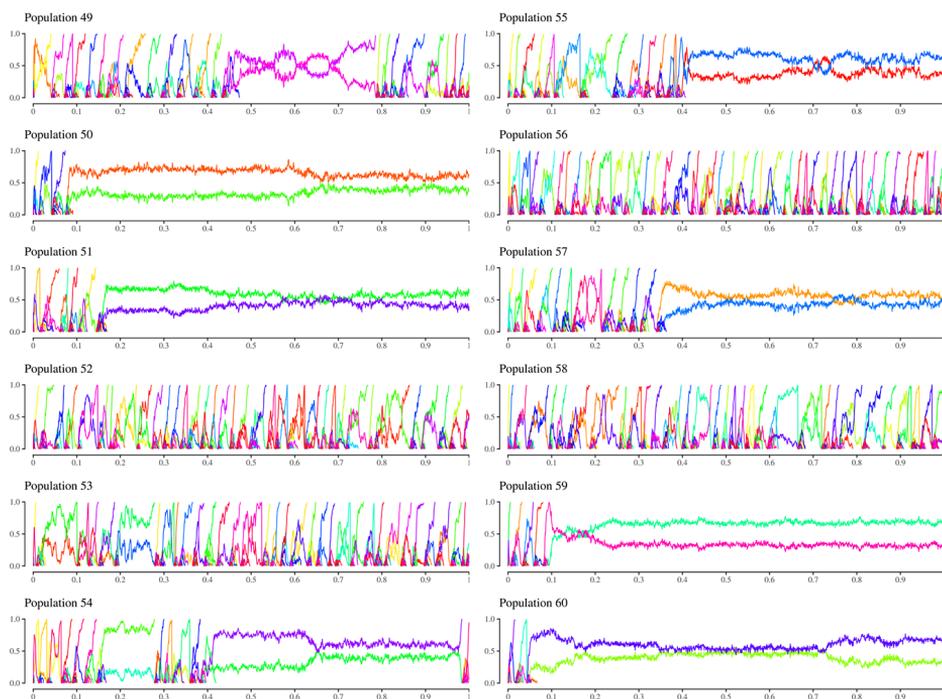
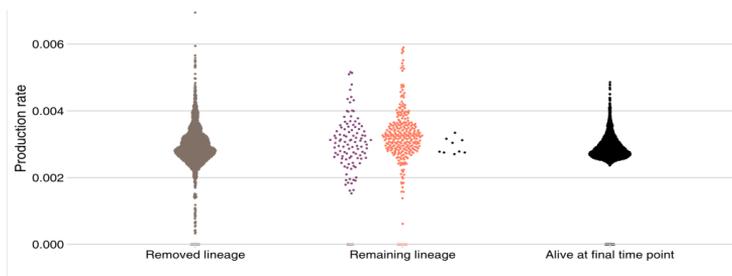


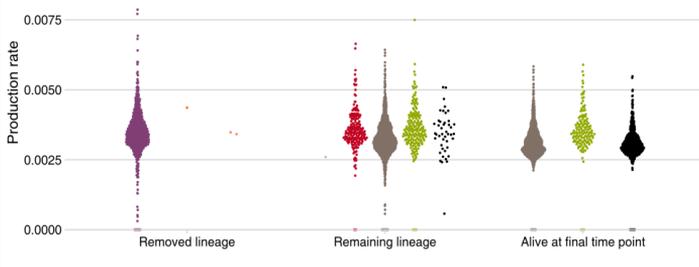
Figure S3.1: (continued).

Productivity of different metabolic types before and after removal

Lineage B remains at t=1



Lineage A remains at t=7



Lineage B remains at t=8

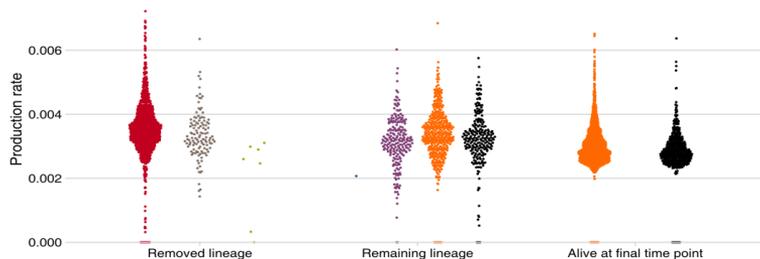


Figure S 3.3: Individual production rates in cross-feeding populations. Individual building block production rates (one dot = one microbe) for three cases in the lineage removal experiment shown in the main text. In the context of the CF population (left, middle), autonomous mutants (in black) that import and produce both building blocks have equal or higher production rates compared to other, co-occurring metabolic types. In contrast, their production rate in the new populations after one lineage is removed (right) is lower compared to the growth rates in the original CF population. This explains why these AU mutants cannot replace the CF population during the main experiment.

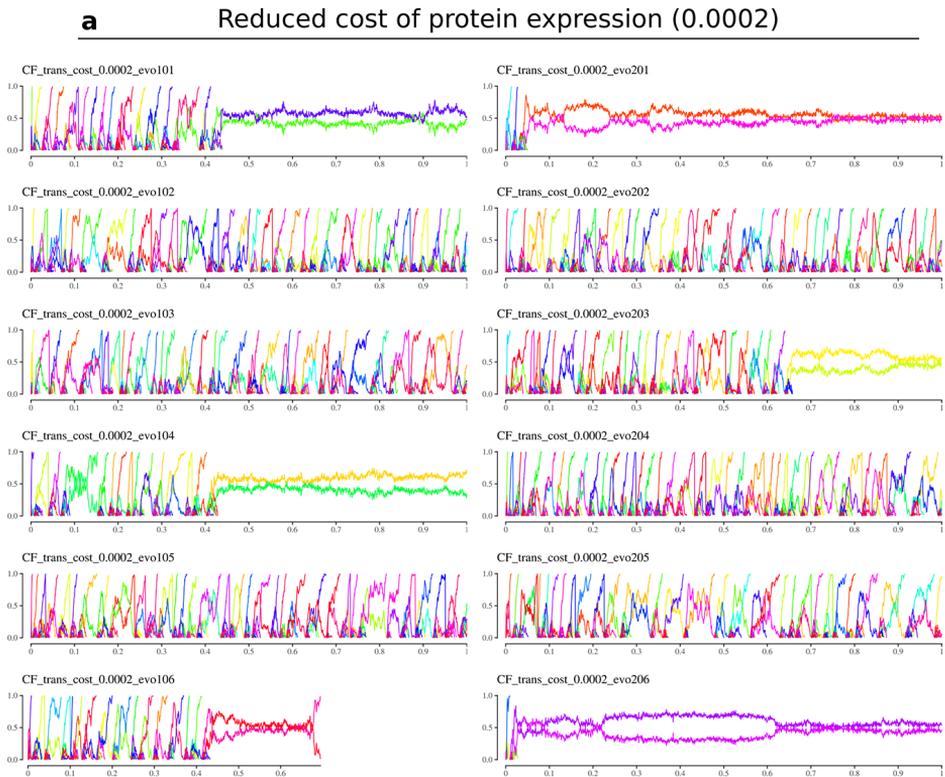


Figure S3.4: Neutral lineage markers for evolutionary simulations with reduced or increased protein expression costs To investigate the effect of protein expression costs in our simulations, we lowered or increased the cost of expressing proteins tenfold, and repeated the experiment for 23 populations. This again resulted in both cross-feeding, autonomous, and switching populations, which reliably associated with the topology of the evolved metabolic networks (see main text for details). Increasing expression costs slowed down evolutionary dynamics, and two simulations (marked with an asterisk) fixed the topology of the metabolic network but did not reach their final eco-evolutionary attractor within the duration of the simulation.

b Increased cost of protein expression (0.02)

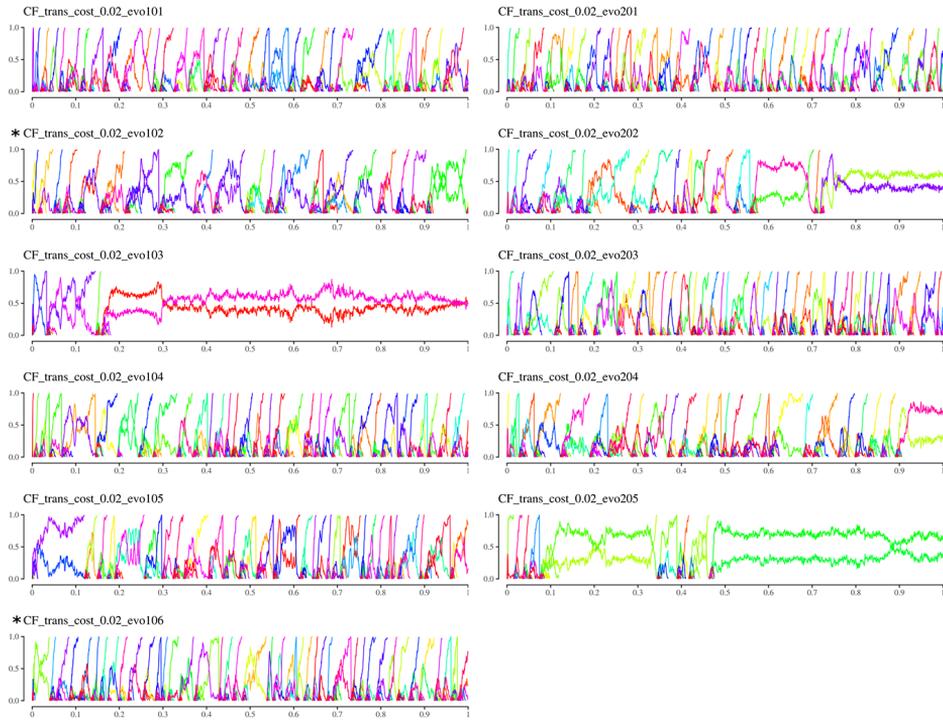


Figure S3.4: continued

Supplementary Figure 5.
PCA of metabolic networks, including
2 outlier communities not shown in Fig. 5.a
of the main text

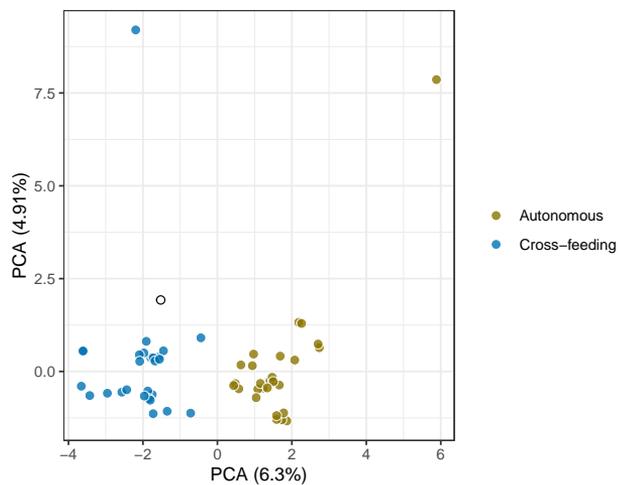
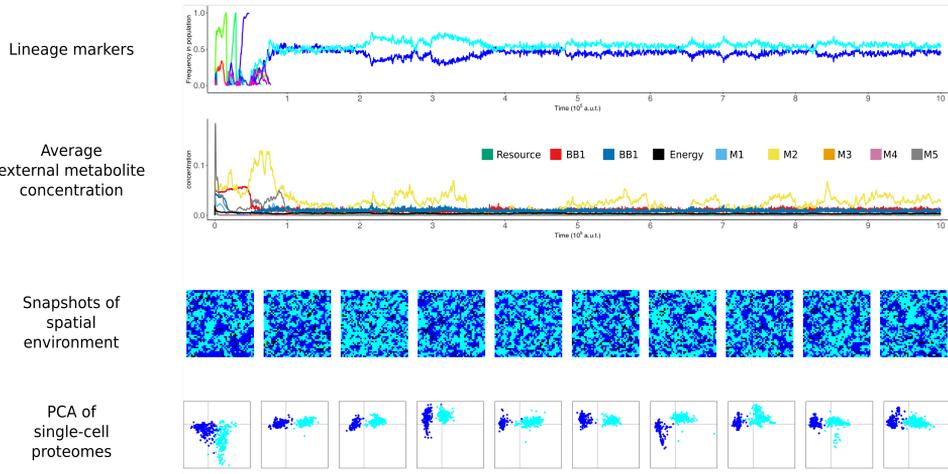


Figure S3.5: PCA of evolutionary trajectories of metabolic networks. PCA of gene frequencies in the main evolutionary experiment, including two outlier populations. See main text for details.

a. Population 9 (CF)



b. Population 33 (SW)

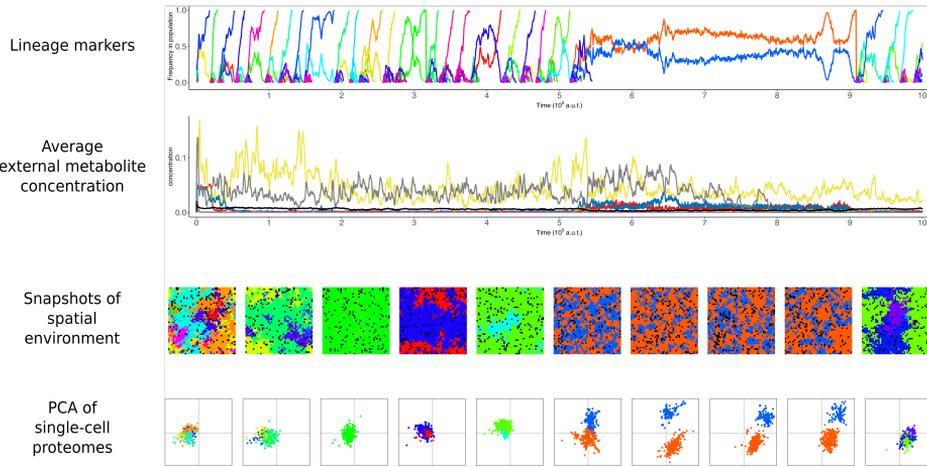


Figure S3.6: Neutral lineage markers, external metabolites, snapshots of the spatial environment, and PCA of microbial proteomes for two communities.

Supplementary Figure 7: Cross-feeding lineages self-organise into a spatially interleaved community

As discussed in the main text, the efficiency of cross-feeding was limited in spatially structured populations. Due to local reproduction, cells continuously find themselves surrounded by their own conspecifics which do not (sufficiently) provide the necessary building blocks. As we also observed how the evolved cross-feeding lineages would form interleaved patterns in space, we thus asked the question: are two cross-feeding lineages more spatially dispersed than autonomous cells? To investigate this, we isolated two typical cross-feeding lineages from one of our evolutionary simulations, and one metabolically autonomous lineage from another simulation. When the system was initialised with the two cross-feeding genotypes spatially separated (see **Supplementary Fig. 3.7a**), the cells quickly self-organised into an interleaved mixture again (also see **Supplementary Movie 1**). We then did the same experiment with a single, metabolically autonomous genotype, which was coloured with two neutral markers. Interestingly, in the timespan where the cross-feeding lineages had formed a mixed community, these metabolically autonomous cells remain spatially separated (see **Supplementary Fig. 3.7b**). Moreover, when these metabolically autonomous cells were initialised into a perfectly mixed state, the resulting pattern after some time was *less* dispersed than that of the two cross-feeding lineages. From this we conclude that cross-feeding communities are ‘over-dispersed’ in space.

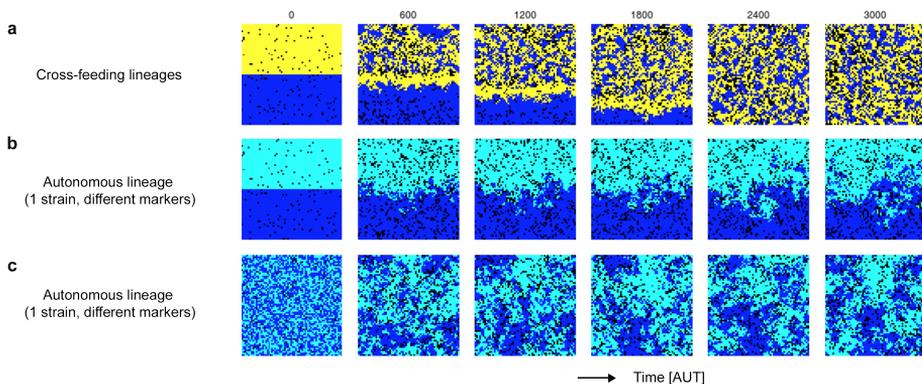
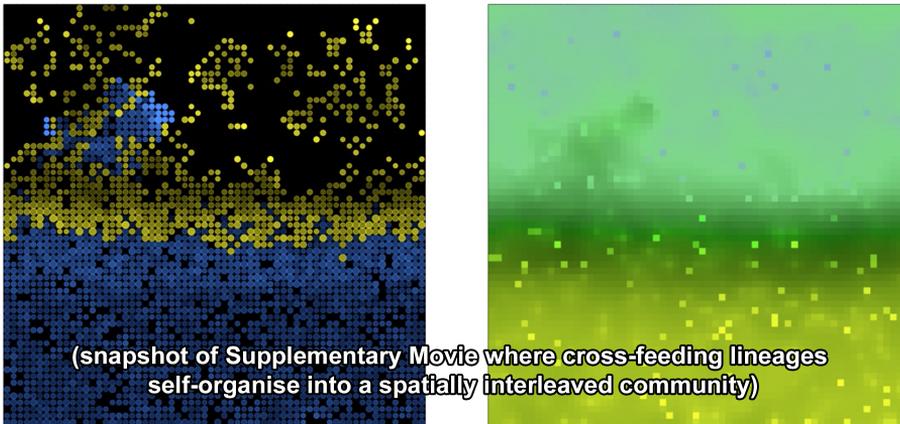


Figure S3.7: Cross-feeding lineages self-organise into a spatially interleaved community. **a.** Two genotypes isolated from a cross-feeding community quickly self-organise into an interleaved mixture. **b.** One metabolically autonomous genotype, coloured with two different neutral markers, does not yield an interleaved pattern. **c.** When one metabolically autonomous genotype is initialised perfectly mixed, it quickly results in a spatial distribution that is less mixed than the two cross-feeding lineages.

Description of Supplementary Movie I

Supplementary Movie I is a detailed illustration of how cross-feeding lineages self-organise into spatially interleaved communities. The top two rows are initialised with cells of one lineage, while the bottom rows are filled with cells from the other lineage (coloured yellow and blue respectively, see snapshot below). On the left-hand side, cells are depicted as circles. The size of the circle indicates the cell volume, and the brightness indicates the budget production rate. The right-hand side of the figure indicates which building blocks are present in the local environment, using a gradient from yellow to blue. A green colour thus indicates the building blocks are approximately equally abundant.

Time: 000319



3

By inspecting this movie we can detect the following details:

1. Cells are growing fastest (are brighter) at the interface between the lineages
2. The yellow type is more dependent on blue than *vice versa*, growing mainly at the interface between lineages, and barely surviving on their own.
3. Blue cells expand into the empty space where yellow has failed to thrive, consuming the abundance of necessary building blocks left behind by yellow.
4. In response, yellow cells can grow and divide again, eventually resulting in an interleaved pattern of cells.

Note that this movie is merely an illustrative example of how cross-feeding cells can intermix, and the details on the inter-dependencies may differ across evolved communities.

“CONSCIOUSNESS IS THE ONE THING IN THIS UNIVERSE
THAT CANNOT BE AN ILLUSION.”

—Sam Harris

Chapter 4

In silico gene-level evolution explains microbial population diversity through differential gene mobility

BRAM VAN DIJK¹ & PAULIEN HOGEWEG¹

¹ UTRECHT UNIVERSITY, UTRECHT, THE NETHERLANDS

GENOME BIOLOGY AND EVOLUTION VOLUME 8, ISSUE 1, JANUARY 2016, PAGES 176–188

Abstract

Microbial communities can show astonishing ecological and phylogenetic diversity. What is the role of pervasive horizontal gene transfer (HGT) in shaping this diversity in the presence of clonally expanding “killer strains”? Does HGT of antibiotic production and resistance genes erase phylogenetic structure? To answer these questions, we study a spatial eco-evolutionary model of prokaryotes, inspired by recent findings on antagonistic interactions in *Vibrionaceae* populations. We find that toxin genes evolve to be highly mobile, whereas resistance genes minimise mobility. This differential gene mobility is a requirement to maintain a diverse and dynamic ecosystem. The resistance gene repertoire acts as a core genome that corresponds to the phylogeny of cells, whereas toxin genes do not follow this phylogeny and have a patchy distribution. We also show that interstrain HGT makes the emergent phylogenetic structure robust to selective sweeps. Finally, in this evolved ecosystem we observe antagonistic interactions between, rather than within, spatially structured subpopulations, as has been previously observed for prokaryotes in soils and oceans. In contrast to ascribing the diversification and evolution of microbial communities to clonal dynamics, we show that multilevel evolution can elegantly explain the observed phylogenetic structure and ecosystem diversity.

4.1 Introduction

Recent metagenomics studies have led to the realisation that the commonly observed phenotypic diversity within microbial ecosystems is just the tip of the iceberg. High-throughput sequencing has revealed an immense variation at the level of DNA sequences (Ochman et al., 2000) and ecological adaptation (Preheim et al., 2011), even in narrow taxonomic groups. How this diversity comes about, and moreover how it is maintained, is still an open question.

An important source of diversity in prokaryotes is the remarkable flexibility of their genome content (Kolstø, 1997; Snel et al., 2002; Koonin and Wolf, 2008). This flexibility is not equal for all parts of the genome, as some genes are more dispensable than others. The well-conserved core genes generally encode basal cellular functions, and are complemented with a variable and more niche-specific set of accessory genes. The core genome tends to have a phylogeny that is consistent with that of the species. On the other hand, the accessory genome has a phylogeny that differs from this vertical inheritance pattern (e.g., Shapiro et al. 2012), showing that horizontal gene transfer (HGT) is an important process in shaping the flexibility and diversity of prokaryotes (Fraser et al., 2009). Indeed, recent results have shown that the rate at which HGT happens is up to 2 orders of magnitude higher than per-gene point mutations (Puigbò et al., 2014). These results indicate that the prokaryotes evolve mainly through gene content evolution, which can change in relatively short timescales.

Populations of bacteria do not only display high diversity in gene repertoires but also at the phylogenetic level, as was recently shown for Vibrionaceae (Preheim et al., 2011) and *Streptomyces* sp. (Vetsigian et al., 2011). Strikingly, this phylogenetic diversity is maintained despite the presence of “superkiller” strains (Vetsigian et al., 2011; Cordero et al., 2012b). Vetsigian *et al.* (2011) explained this superkiller paradox in a model assuming high rates of *de novo* gene gain (e.g., HGT from a distant source), which creates an evolutionary arms race. However, the resulting clonal diversity and short evolutionary life span of lineages appear to fall short in explaining the observed phylogenetic diversity in natural populations. Surprisingly, both studies also indicate that killing happens less frequently between related individuals, that is, close kin are spared from killing. The frequency of killing is low up to a certain genetic threshold, after which it sharply increases. This observation once again suggests high levels of phylogenetic structure. The combination of high phenotypic diversity and high phylogenetic diversity suggests that adaptive genes sweep the population on their own (Shapiro et al., 2012), rather than resulting in genome-wide, clonal sweeps. Furthermore, as *de novo* gene discovery fails to reproduce these patterns, internal HGT of locally present DNA is a likely candidate for the flexibility and diversity of prokaryotic populations.

In *Vibrio ordalii*, a toxin production gene was identified that is part of a mobile genetic element (MGE) (Cordero et al., 2012b). Surprisingly, this toxin-production gene was not linked to the corresponding resistance factor, and can thus only transfer to individuals that already have the resistance factor in their genetic background. The resistance of *V. ordalii* was indeed shown to be ancestral, and other resistant-only individuals show no trace of the MGE. These results indicate that the resistance gene is part of the core genome, whereas the toxin gene is

a mobile accessory gene which is frequently transferred within and potentially between populations of Vibrionaceae. This genome structure appears to prevent clonal dynamics, and impacts the observed diversity and interaction dynamics by means of gene-level selection.

We here ask the question how local HGT impacts population diversity and the differential gene mobility as observed for the toxin and resistance genes in *Vibrio*'s. Furthermore, we wish to understand what selection pressures underlie this process, and how it shapes core-, accessory-, and pan-genomes. We investigate these questions by using a minimal model of microbial evolution, where cells undergo HGT of locally present DNA. The model includes three different levels: genes, cells, and spatial patterns. Cells compete locally for space by means of antagonistic interactions of toxin and resistance genes. Cells stochastically take up genes from the local extracellular DNA (eDNA) pool. Genes encode their own mobility parameter, which represents the likelihood of integration into the genome after uptake. Antagonistic interactions take place between spatially patterned subpopulations. We focus on the evolution of gene mobility, and how this influences diversity and population structure. The aim is to reveal the intrinsic evolutionary pressure of gene mobility, abstracting away from the mechanisms underlying this process, such as uptake sequences (Mell and Redfield, 2014) or illegitimate recombination (de Vries and Wackernagel, 2002). Our minimal model shows how HGT and the resulting interplay of gene- and cell-level selection can explain the observed ecosystem diversity and the different phylogenetic patterns of core and accessory genomes of natural prokaryotic populations.

4.2 Methods

4.2.1 General Description

We model a simple *in silico* microbial population that is subject to local HGT of toxin and resistance genes (for an overview, see **Figure 4.1**). A spatial grid is implemented on which a single species is competing for reproduction space through local interactions by means of 20 sets of predefined toxin- and resistance-genes. We assume that the costs for producing toxins or resistance factors are equal, and focus our observations on the evolution of gene mobility. The gene mobility is a gene-specific, evolvable feature which determines the chance of successful integration after uptake (*i.e.*, abstracting away from mechanisms such as flanking sequence homology, MGEs, etc.). The genes available for uptake originate from dying individuals, leaving their genetic information behind. This pool of eDNA undergoes diffusion and decay, whereas individuals in the proximity of this genetic material can stochastically take up genes.

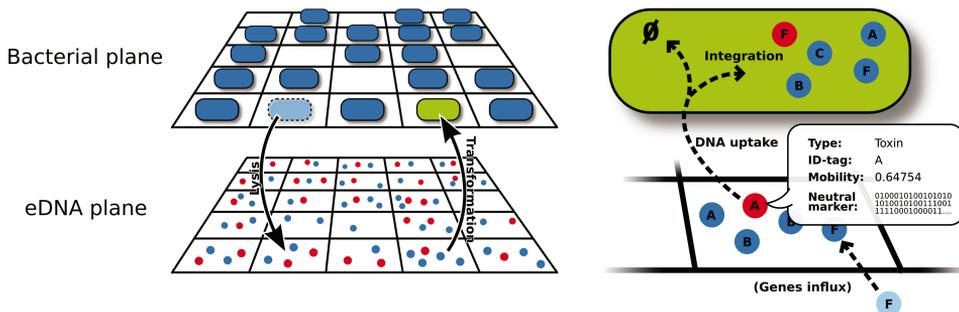


Figure 4.1: Graphical representation of the modelled microbial ecosystem. Microbe-like individuals compete on a grid, with fitness-proportional reproduction (R_i and B_i , equations 4.2 and 4.1) and stochastic death (D_i , equation 4.3). Fitness is decreased for every gene in the genome of an individual linearly (parameter c). D_i is increased for each toxin produced in the Moore (8) neighbourhood for which the individual does not produce resistance factors (parameter t). Dying individuals leave their genes behind (blue and red circles for resistance and toxin genes, respectively), creating an eDNA pool that is diffused using Toffoli–Margolus Diffusion (Toffoli and Margolus, 1987). Genes can then be taken up with a fixed probability (u), and get integrated into the genome with a probability that is evolved per gene (m).

4.2.2 Population Dynamics

Our model is a stochastic cellular automaton, where grid points represent reproduction space. They are either empty or contain one individual bacterium. Every time step, all grid points are synchronously updated based on their own state, and the state of adjacent grid points in the Moore Neighbourhood (eight adjacent sites). If the grid point is empty, local neighbours can compete for reproduction (R_i) into the empty spot with chances that are dependent on their relative birth rate (B_i) with respect to the birth rate of the other competitors (see **Equation 4.1** and **Equation 4.2**). B_i is linearly decreasing for the total number of toxin- and resistance genes produced by that individual (G_{t+r}), with a fixed per-gene cost (c). It is also possible that none of the individuals reproduces, for example, when there

are very few or unfit individuals surrounding the empty grid point (this chance is scaled by parameter e).

Individual bacteria die with a constant probability (d), increased by the toxicity parameter (t) for every direct neighbour that is capable of killing the focal cell (**Equation 4.3**). Note that killing is self-inclusive, meaning that individuals need both the toxin gene and the resistance gene to survive.

$$B_i = \max\left(0, 1 - c \cdot \sum_{j=1}^{n_{genes}} (G_{ij})\right), \quad (4.1)$$

$$R_i = \frac{B_i}{\sum_j B_j + e}, \quad (4.2)$$

$$D_i = \min\left(1, d + t \cdot \sum_{j=1}^{n_{neigh}} \sum_{k=1}^{n_{genes}} (G_{i,k(res)} == 0 \ \&\& \ G_{j,k(tox)} > 0)\right). \quad (4.3)$$

4.2.3 Gene-Level Dynamics

The genomes of bacteria are implemented as simple collections of toxin- and resistance genes, where the order of genes is not taken into account (*i.e.*, excluding the possibility of physical coupling of genes). In this model, 20 different pairs of toxin- and resistance genes are taken into account (a total of 40 unique genes). The genes have an identity-tag (letters $A - T$ in **Figure 4.1**), a mobility parameter (m), and a neutral bit-string marker used for phylogenetic reconstruction of gene trees (see details on phylogenetic reconstruction below). Upon reproduction, an individual copies its collection of genes allowing for mutations on both the mobility parameter (m), and bit flip mutations on the bit-string marker. There is also a fixed per-gene probability of loss (l). *De novo* gene discovery and gene duplication do not happen as a result of replicating the genome for reproduction. However, both gene duplications and gene discoveries can happen through a simplified form of HGT in the model.

HGT, and the resulting gene-level dynamics, is included by modelling a simple process of uptake of eDNA (as shown in **Figure 4.1**). Individuals that die leave their genes behind, which diffuse and decay with a fixed rate. Genes that reside on the same spot as any bacterium have a fixed chance (u) to be taken up by the bacterium. Subsequently, the gene can be integrated into the genome of the individual with a chance $P(m)$ which is an evolvable property of that toxin or resistance gene. If that cell already had a copy of that gene, it pays the cost without getting any ecological benefit. Thus, in this model, having multiple copies of a gene is always detrimental to the direct fitness, and can only be advantageous in an evolutionary context (*e.g.*, redundancy when one of the copies is lost). If the gene is not integrated into the genome after uptake, it is degraded ($P(1 - m) \rightarrow \emptyset$).

For example, if toxin A in **Figure 4.1** (coloured red) gets taken up by the green cell, it gets incorporated into the genome with a chance (in the figure indicated as $\sim 65\%$), which is unique for that instance of gene A. Because of mutations, many other instances of toxin gene A are present with different values for m , resulting in

a process of natural selection on the gene-level. Note that in this example, the cell indeed has the resistance gene for compound A, meaning it could give a selective advantage to the cell if it would get integrated into the genome, indirectly also benefiting the toxin gene.

Finally, every time step a fixed number of random toxin and resistance genes flux into the eDNA-pool (f). These genes flux in at a random position. They can have a novel and unique identity, but the mobility is sampled from the existing ecosystem, under the assumption that there is influx from a relatively distant, but otherwise similar ecosystem. This influx parameter f is used to study the effect of closed ecosystems (low f), where the majority of genes comes from the population itself, and gradually more open ecosystems (higher f).

4.2.4 Visualising Fitness Landscapes

Fitness landscapes are generated to understand when the loss/gain of a gene is beneficial and gives an immediate benefit for a cell. Here we define fitness as $(B_i - D_i)$, and calculate the average difference in fitness with respect to the eight adjacent neighbors, before and after the mutation. The difference between before and after (E_i) gives the immediate benefit that a gain/loss of a gene would give the focal cell (**Equation 4.4**). Note that long-term, indirect effects (e.g., toxin gain enabling to locally overgrow more than the direct neighbors) are not included in this measure. We also measure E_i with eight randomly chosen individuals, which helps to understand the importance of local interactions in the ecosystem:

$$E_i = \frac{\sum_{j=1}^n ((B_i^* - D_i^*) - (B_j - D_j^*))}{8} - \frac{\sum_{j=1}^n ((B_i - D_i) - (B_j - D_j))}{8} \quad (4.4)$$

where n is the number of living surrounding neighbouring cells (max 8), and B_i^* , D_i^* , and D_j^* are the modified birth/death rates after the gain/loss of the gene (B_j cannot be affected).

4.2.5 Strain Diversity and Phylogenetic Reconstruction

In this model, we distinguish between phenotypic strain diversity and phylogenetic diversity. Strain diversity is measured at the phenotypic level, by the presence or absence of resistance and toxin-production. In other words, strains with multiple (redundant) copies of a gene do not contribute to strain diversity. To exclude transient mutants, only strains consisting of more than 20 individuals are included.

In addition to strain diversity, phylogenetic trees are constructed to study the phylogenetic diversity and the balance between vertical and horizontal transfer of genes. We added neutral bit-string markers to both the cells and to each of the genes. Note however that these markers are strictly for observational purposes, and have no meaning other than to reconstruct the evolutionary history. All individuals carry a neutral marker that is only inherited vertically, but the marker for genes can both be inherited vertically and horizontally (through the transformation-like process described above). Upon inheritance, the 1.000-bit-long binary strings mutate with a

chance of 10^{-6} per bit. Distance matrices are made of the populations of genes and individuals, allowing us to make species trees, gene trees, and most importantly, to map closely related genes onto the species tree. All trees were generated in R version 3.0.2 using `hclust` with an average similarity method.

4.2.6 Parameters used in the model

Unless stated otherwise, the parameters for the model are given by Table 4.1.

Table 4.1: Default Parameters for the Model

Parameter	Value
Grid size	250×250
Base death (d)	0.05 per time step
Toxicity (t)	0.3 per toxin
Gene cost (c)	0.05 per gene
Gene loss (l)	0.0005 per gene per replication
Rate of uptake (u)	0.01 per cell per gene
Initial HGT rates	0.1
Gene influx No. (f)	3 per time step
Decay genes	0.1
eDNA diffusion No.	3 per time step
Initial gene mobility	0.1
Mobility mutation rate	0.01 per replication
Mobility mutation size	Uniform ± 0.05
Neutral marker mutation rate	10^{-6} per bit

4.3 Results

We first focus on some key results of the model under parameter conditions given by **Table 4.1**. Both gene costs and loss were set as low as possible without resulting in the loss of too many toxin genes through the spread of universal resistance. In larger grids even lower values can be used without toxin genes being lost, but we chose a relatively small grid (250×250) for computational efficiency. In the section “Evolution of the Fitness Landscape”, we discuss how too low a gene loss/cost eventually leads to the loss of all or most genes.

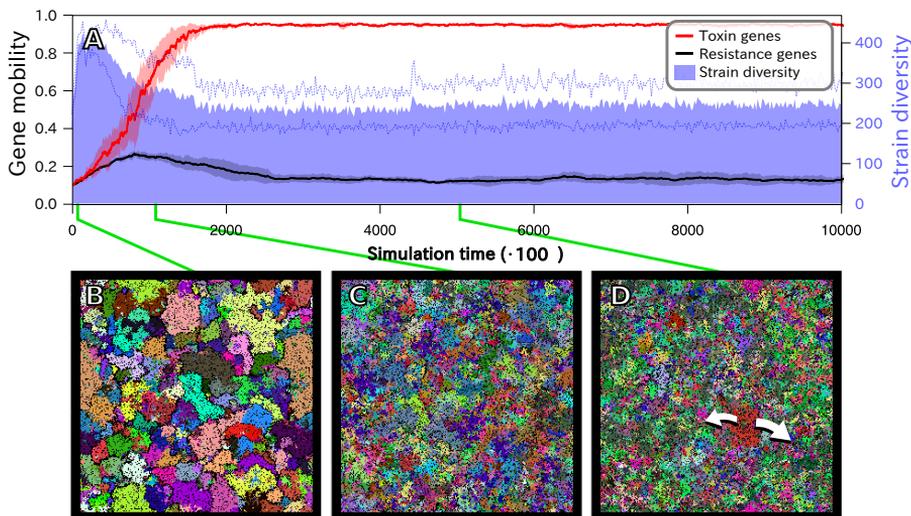


Figure 4.2: **A)** Evolution toward differential gene mobility. $T_{mobility}$ (red line), $R_{mobility}$ (black line), and strain diversity (blue area) are plotted against time. Dotted lines and shaded background depict the standard deviation of 5 independent simulations. Only strains that were prevalent (>20 individuals) were accounted for the strain diversity. **B-D)** Snapshots of the field-dynamics, where black depicts empty cells, and all other colors are unique strains. The white arrows in panel D depict how the red strain locally overgrows sensitive individuals.

4.3.1 Evolution of differential gene mobility

Individuals are initialised by randomly assigning relatively immobile toxin and resistance genes ($m \sim 0.1$) to a subset of the population, yielding a population where individual strains mutually kill one another (see Supplementary Movie S1). During this initial phase, the average mobility of both toxin and resistance genes, respectively called $T_{mobility}$ and $R_{mobility}$ henceforward, is increasing (**Figure 4.2A**). The flux of genes favours the spread of resistance, and after some time, the pattern of mutual killing (**Figure 4.2B**) virtually disappears (**Figure 4.2C**). From this point, the mobility of resistance genes ($R_{mobility}$) no longer increases, but minimises to approximately 0.10, whereas $T_{mobility}$ evolves to a high value (~ 0.95). Surprisingly, diversity is maintained even after long-term evolution, despite the high levels of resistance. In this equilibrium phase, long periods of static coexistence are interspersed with localised killing (**Figure 4.2D** and Supplementary Movie S1). We next wished to understand how the divergence of $T_{mobility}$ and $R_{mobility}$ comes

about, and how this process maintains a diverse population. **Figure 4.3A** displays how the average genome composition of individuals changes over time. First, there is a strong increase in individual resistance (grey area), as was already evident by the aforementioned disappearing of mutual killing. The benefit of producing toxins should thus be decreasing, as individuals become resistant to a variety of toxins. Indeed, the decline in the number of toxin genes per individual (red area) reflects that producing a toxin is, on average, costly. Some toxin genes are completely lost from the population, as can be seen in the declining number of toxins in the pan genome (dark blue line). However, toxins genes that evolve higher mobility than their corresponding resistance factor appear to be maintained (**Figure 4.3B**, light grey lines), even after long-term evolution. Note that toxin and resistance genes with less diverged gene mobility, eventually go extinct (**Figure 4.3B**, red lines). We conclude that toxin genes can survive prolonged periods of unrewarding toxin production, by shifting selection toward the gene-level. The persistence of these genes in the population is rewarded when a strain that has lost resistance to the toxin is locally overgrown (as shown in **Figure 4.2D**). These results explain how only toxin and resistance genes with strongly differential gene mobility are maintained after long-term evolution, which in turn helps to uphold strain diversity.

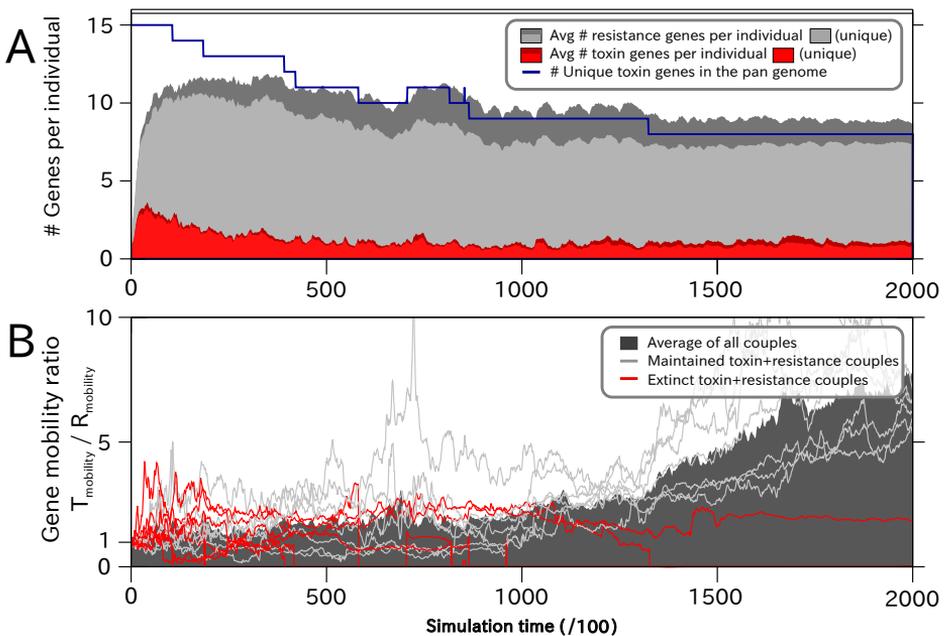


Figure 4.3: High levels of resistance promote the evolution of differential gene mobility. A) Average number of different types of toxin genes (red) and resistance genes (grey) per individual. The darker shadow displays the total of non-unique genes of both types. The total number of toxins in the pan genome is depicted by the blue line. B) The ratio between $T_{mobility}$ and $R_{mobility}$ for all corresponding toxin and resistance genes is displayed, where maintained couples are displayed in grey, and couples that go extinct before 1×10^6 simulation steps are displayed in red. The average gene mobility ratio is displayed by the dark grey area.

4.3.2 Differential gene mobility is required for phenotypic diversity

Spatial patterns allow for many strains to coexist, because there is no all-against-all competition. Indeed, mixing either the population or the eDNA results in the extinction of most or all toxin and resistance genes. Furthermore, the mixing of eDNA results in “selfish” DNA which replicates at the expense of their hosts (**Suppl. Figure S4.1**). As the influx of new genes (f) is low (**Table 4.1**), maintaining diversity depends mostly on maintaining standing variation. Does the observed diversity in our simulations so far depend on the evolution of the differential gene mobility, or is the HGT on its own sufficient to explain the maintained it? To answer this question, simulations are done where gene mobility was fixed, and not allowed to evolve. **Figure 4.4** shows that diversity always decreases to very low values when $T_{mobility}$ and $R_{mobility}$ are fixed to the same value, even when $T_{mobility}$ and $R_{mobility}$ are both equally high (green line). The diversity of the previous simulations (blue area) is evidently much higher. When the mobility of toxin and resistance genes is diverged from the start of the simulation (grey area), the population retains very high diversity within the time span of the simulation, showing that differential gene mobility indeed helps in preventing the decline of diversity. Concluding, in our closed ecosystem with low influx, diversity can only be maintained through differential gene mobility.

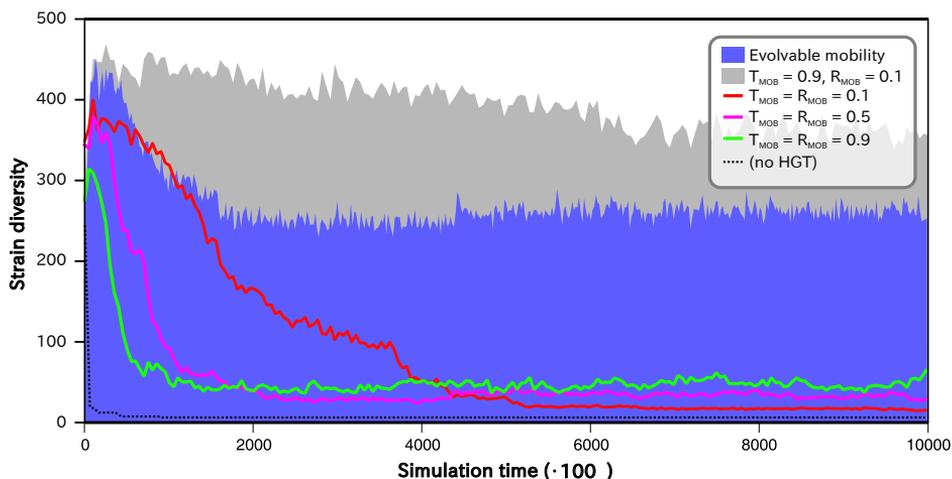


Figure 4.4: Differential gene mobility is required to maintain phenotypic diversity. Average strain diversity is depicted over evolutionary time, for different scenarios where $T_{mobility}$ and $R_{mobility}$ do not evolve (coloured lines). For comparison, the diversity from the main simulations is also displayed (blue area). The grey area depicts simulations where toxins and resistance genes have differential gene mobility (0.9 and 0.1, respectively) that does not evolve.

4.3.3 Evolution of the fitness landscape

Why does $R_{mobility}$ remain low, if higher mobility could help prevent local sweeps? Why are there no all-resistant individuals, out-competing all toxin producing individuals? To further understand the evolved genome composition, and the

emergence of differential gene mobility, we calculated the fitness effects for gaining and losing genes, both early and late in the evolution. **Figure 4.5A** shows that gaining a resistance gene is almost always detrimental for the cell. This peak at -0.05 reflects the per-gene costs (c parameter) the host pays for a resistance gene, which yields no direct benefit. In contrast, figure **Figure 4.5B** shows that the effect of gaining a toxin gene is much more varied, and although it is often detrimental because of self-killing or the cost of the gene, it can be very advantageous as well. Interestingly, long-term evolution (black bars) increases both neutrality and the beneficial effects of gaining a toxin gene. A fairly weak initial bias towards gaining toxin genes leads to evolution of higher $T_{mobility}$, which is in turn maintained when $R_{mobility}$ evolves towards lower values.

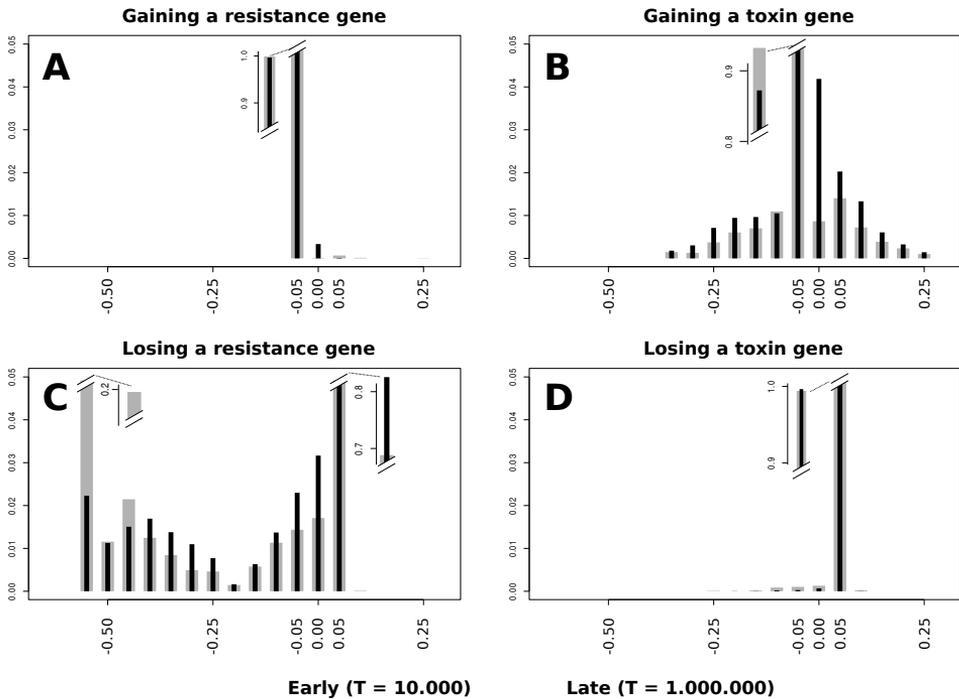


Figure 4.5: Evolution of the fitness landscape of gaining or losing toxin and resistance genes. The distribution of fitness ($B_i - D_i$) advantages with respect to local neighbours for gaining/losing genes. Blue bars indicate an early time point where $T_{mobility}$ and $R_{mobility}$ have not diverged yet. Black bars indicate the end of the simulation, where $T_{mobility}$ and $R_{mobility}$ have been diverged for a long time. Each type of mutation (A - D) was repeated 1.000 times for every individual in the population. All error bars were very small, and omitted for clarity.

Loss of genes (*i.e.*, a non-evolvable parameter) shows an opposite trend. Losing resistance genes has a varied effect on fitness, and is frequently deleterious (**Figure 4.5C**), whereas losing toxin genes is almost always slightly beneficial (**Figure 4.5D**). Together these fitness effects lead to the observed population with high levels of resistance and much lower levels of toxin production (**Figure 4.4**), and the evolution of differential gene mobility (**Figure 4.2**). Furthermore, the fitness landscapes show that long-term evolution leads to significantly higher neutrality, reflecting the long-term evolution to higher robustness known for static fitness landscapes (Van Nimwegen et al., 1999; Wagner, 2005). In contrast to these classical

studies, the evolution of the fitness landscape is not only due to the evolution of genome composition but also due to the evolved population structure and local pattern formation. To quantify the role of local pattern formation, **Suppl. Figure S4.2** displays the fitness landscapes (analogous to **Figure 4.5**) for the evolved population structure and genome composition, excluding the influence of local pattern by choosing random interactors instead of neighbours. Most strikingly, the beneficial effects of gaining toxin genes are not present, showing that gaining a toxin gene hardly ever yields a “global” advantage. Spatial pattern formation is indeed essential for the outcome of the model. If spatial pattern formation is prevented, diversity collapses in the very short term (data not shown). We conclude that the functional differences between toxin and resistance genes, and therewith the differences in fitness landscapes, play an important role in the evolution of differential gene mobility, and its consequences.

Finally, the fitness landscape also clarifies why too low per gene costs/loss leads to the loss of most genes. We can infer from **Figure 4.5B** that the advantageous effects of gaining a toxin are very rare. When gene cost or gene loss is too low, multi-resistant cells readily emerge. This would make the advantageous effects of gaining toxin genes even more rare. Therefore, even very mobile toxin genes are lost from the population, which in the long run leads to the loss of the resistance genes, which are now merely costly. Indeed, **Suppl. Figure S4.3** shows that when gene cost or gene loss is low, the strain diversity is much lower. Even so, differential gene mobility still evolves. Increasing gene cost or gene loss generally yields clonal red-queen dynamics, and does not result in the evolution of differential gene mobility.

4

4.3.4 Core resistance repertoires and accessory toxins

After $T_{mobility}$ and $R_{mobility}$ have diverged to nearly a 10-fold difference, we next focus on how toxin and resistance genes are distributed among the various strains. For this, we have tagged both individuals and genes with a neutral bit-string marker (see Materials and Methods). Note that this marker is only used for observational purposes, and has no meaning otherwise. The markers for individuals can only be inherited vertically, which we use to generate a reference “species”-tree. Next, the phylogeny of genes is reconstructing using their own neutral markers. When the effective rate of HGT is high, closely related genes will appear in different clades in the reference tree. As shown in **Figure 4.6A**, the evolved differential gene mobility is reflected in how corresponding toxin and resistance genes are mapped onto the phylogenetic tree. Closely related clusters of a typical resistance gene (depicted as different colours in the inner circle) appear together in the tree, whereas the gene for the corresponding toxin appears in many different clades (outer circle, also see **Suppl. Figure S4.4** for a representation of all genes). To quantify this effect for all genes and in more replicate simulations, we use a Robinson–Foulds metric (Robinson and Foulds, 1981), as a measure for tree-distance (**Figure 4.6B**). Clearly, resistance genes are a better representation of the species tree than toxin genes are. The evolved differential gene mobility leads to persistent and core-like resistance repertoires with more transient accessory-like toxin genes.

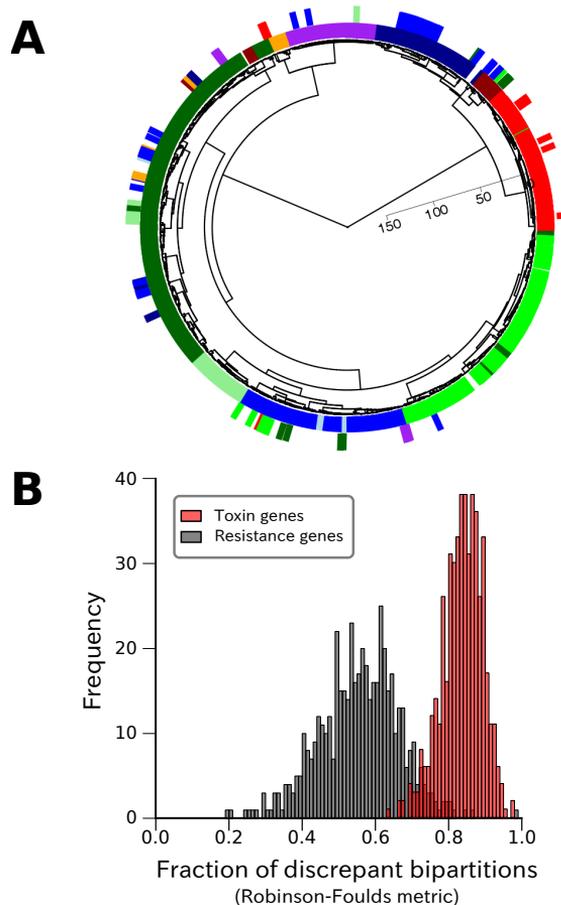


Figure 4.6: Phylogenetic reconstruction shows core-like resistance with a patchy distribution of toxin genes. **A)** For a typical corresponding toxin and resistance gene, closely related genes (different colours) are mapped onto the phylogenetic tree of 500 randomly sampled individuals from a typical simulation. The inner circle depicts how the phylogeny of the resistance gene closely resembles the phylogenetic tree. The corresponding toxin in the outer circle displays that the toxin gene has a much more scrambled distribution. Although this effect exists on many levels (i.e., does not strictly depend on how many gene clusters are made), this example uses ten gene clusters. **B)** Frequency histograms of the fraction of discrepant bipartitions, a Robinson–Foulds metric (Robinson and Foulds 1981), predicted by both gene-types for a total of ten simulations. The distribution for resistance genes and toxin genes has significantly different means ($P \ll 10^{-14}$, Welch’s t-test).

4.3.5 Strains are cohesive units of interaction

The sparing of close kin as described by Vetsigian *et al.* (2011) and Cordero *et al.* (2012) is a feature that requires a non-clonal, phylogenetic structure. The clonal populations in the model by Vetsigian *et al.* (2011) indeed fail to reproduce this observed pattern, as the likelihood of killing increases even for very small genetic distances across strains. To study this feature for our simulations, **Figure 4.7** displays how the cumulative probability of killing relates to the genetic distance between the target and receiver. Killing is not observed between highly related

individuals, and higher frequencies of killing increase only for much greater genetic distances. As seen in the phylogenetic tree plotted along the same axis in **Figure 4.7**, the level of taxonomic structure below this threshold reflects the core-like resistance repertoires of strains.

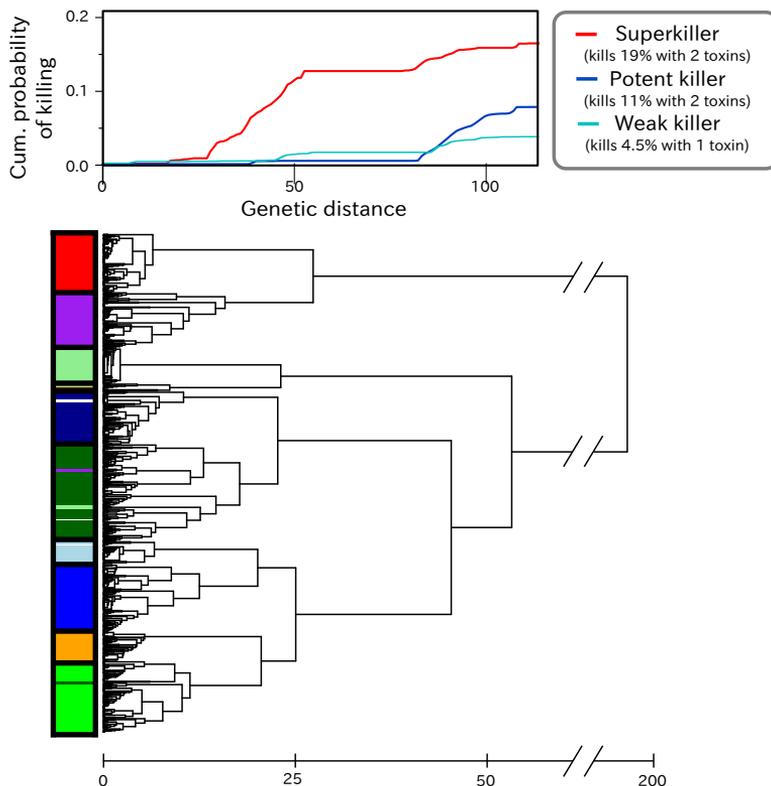


Figure 4.7: Strains spare close kin from killing In an all-against-all killing assay, we compute the cumulative probability that the focal individual can kill a target individual given the genetic distance between them. The phylogenetic tree of 500 randomly sampled individuals is displayed along the same axis, clustered into ten clades based on a core-like resistance gene (different colours). These ten clades correspond to the genetic distance where all profiles stop sparing close kin (~ 40).

Similar to the all-against-all assay of *Vibrionaceae* sp. by Cordero *et al.* (2012), strains in our model are also cohesive populations, which act as emergent units of interaction. Killing happens more likely between clades than within, in accordance to the high phylogenetic structure and the core-like behaviour of resistance repertoires. Spatial structure plays an important role in maintaining this pattern. Strains that lose resistance are less likely to survive if the corresponding toxin is produced by neighbouring (*i.e.*, more related) strains. The surprising interactions observed in all-against-all assays (*e.g.*, super-killers) are not the interactions that determine the eco-evolutionary outcome of our model.

4.3.6 Diversity is robust to selective sweeps

Because of the small influx of novel genes, the discovery of new toxins is very rare. This is because a strain would firstly have to acquire resistance to a toxin which

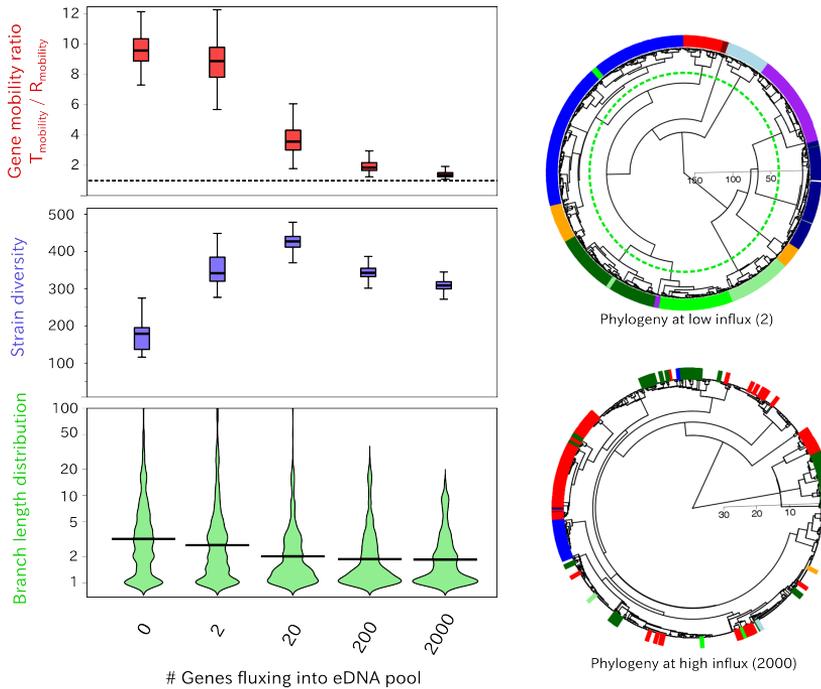


Figure 4.8: Differential gene mobility evolves only in a closed ecosystem. The effect of increasing influx on the gene mobility ratio (red), the strain diversity (blue), and the branch length distribution (green). Two phylogenetic trees display the difference in the population structure between a typical ecosystem with low influx and a typical ecosystem with high in flux. Clusters of an arbitrary resistance gene (different colours) are mapped to the trees to visualise the presence/absence of core genomes. The dashed green line displays how a single core-genome at low influx has a distance to MRCA comparable to the entire tree of the ecosystem with high influx.

is not produced by the population, and subsequently discover the corresponding toxin gene. Nevertheless, it happens in three of the ten replicated experiments. We therefore ask the question whether such an event would erase all existing phenotypic diversity and phylogenetic structure. Surprisingly, diversity increases in all three cases, despite the fact that these events are followed by what looks like a selective sweep (Supplementary Movie S2 and **Suppl. Figure S4.5**). The distance to the most recent common ancestor and most of the underlying population structure is maintained even after the sweep, as sufficient HGT happens between the sweeping strain and other pre-existing strains. We varied the diffusion of eDNA as a proxy for inter-strain gene flow and subjected these populations to similar selective sweeps. When diffusion of eDNA is very low, HGT happens more frequently within strains than between them. When we subjected populations with low eDNA diffusion to a selective sweep, two out of four sweeps lead to a phylogenetic bottleneck, which never happens with higher rates of eDNA diffusion (see **Suppl. Figure S4.6**). This shows that when the inter-strain gene flow is sufficiently faster than a selective sweep, the genes sweep the population on their own, allowing not only strain diversity but also phylogenetic diversity to persist.

4.3.7 Differential gene mobility evolves only in a closed ecosystem

The results discussed so far are examples from a relatively closed ecosystem, that is, with low influx of genes. Next, we focus on gradually more open ecosystems by increasing the external influx of genes into the eDNA plane (**Figure 4.8**). For more open ecosystems, the difference between $T_{mobility}$ and $R_{mobility}$ eventually disappears, where both $T_{mobility}$ and $R_{mobility}$ evolve towards high values. Although this increasing influx mostly has a positive effect on the strain diversity, the branch length distribution shows that this scenario describes an evolutionary arms race where novel clones continuously replace one another (see Supplementary Movie S3. Indeed, the phylogenetic trees show that at high influx phylogenetic diversity is much lower than at lower influx. Also note that the core-like resistance repertoire as discussed above no longer exists at this parameter-range. Differential gene mobility and the resulting core and accessory genomes are most strongly observed at low influx, that is, when most eDNA originates from the local population itself.

4.4 Discussion

We showed that differential gene mobility evolves in a simple model where HGT occurs by incorporation of locally present eDNA. Toxin genes evolve high mobility, whereas resistance genes are mostly vertically inherited. This differential gene mobility is furthermore required to maintain a diverse population, as diversity collapses when mobility is equal for toxin and resistance genes. Moreover, for parameter values which do not evolve differential gene mobility (**Suppl. Figure S4.3**), the phylogenetic diversity as observed *in vivo* does not occur. The differential gene mobility appears to generate the within-species population structure and interaction dynamics as observed *in vivo* (Vetsigian et al., 2011; Cordero et al., 2012b; Cui et al., 2015). Our model shows that many observations on microbial ecosystems can be explained by local HGT and the resulting interplay between gene- and cell-level evolution.

4.4.1 Closed ecosystems with high diversity

Interspecies HGT challenges our concept of species (Zhaxybayeva and Doolittle, 2011). On the other hand, within-species HGT on a more local scale is now challenging our expectations on the eco-evolutionary dynamics of prokaryotes (Shapiro et al., 2012; Overballe-Petersen and Willerslev, 2014; Takeuchi et al., 2014). We have shown in that the dynamics of a closed ecosystem, that is, with HGT through a local pool of eDNA, is very different from open ecosystems where new genes are continuously discovered. The closed ecosystem is very congruent with the observation of core-like resistomes in natural ecosystems (Forsberg et al., 2014). Furthermore, despite the closed and relatively small ecosystem, adaptations such as toxin production can be present because of gene-level selection. This shows that high levels of diversity and the presence of “super killers” (Vetsigian et al., 2011; Cordero et al., 2012b) do not necessarily contradict one another. Moreover, the diversity and dynamics of the closed ecosystem do not depend on high rates of *de novo* gene discovery, as in many other modelling approaches (Vetsigian et al., 2011; Cui et al., 2015), but on local HGT.

Although observations in our model are strikingly similar to certain *in vivo* observations, bacteria in agricultural soil (Heuer et al., 2011) or hospital wastewater (Al Naiemi et al., 2005) might have the dynamics of an open ecosystem (*e.g.*, as a result of anthropogenic influences). We have shown that increasing the external influx of genes results in very rapid dynamics of clonal expansions driven by the continuous discovery of novel genes (**Figure 4.8** and Supplementary Movie S3). Although superficially this ongoing *de novo* gene discovery results in higher strain diversity, we show that repeated genome-wide selective sweeps actually negatively impact phylogenetic diversity. In other words, our results suggest that, counterintuitively, genetic diversity is highest when the majority of available adaptations originates from the population itself.

Earlier research has also challenged the idea diversity is reduced by genome-wide selective sweeps of beneficial adaptations. Shapiro et al. (2012) propose that HGT can actually result in gene-specific selective sweeps, and how the diversity of the population is not necessarily purged by this process. Indeed, Takeuchi et al. (2015) have studied a model of gene-specific sweeps of a generic beneficial adaptation through negative frequency-dependent selection. Here, we have also shown gene-specific sweeps, but now in a context in which the benefit of a gene depends on the local neighbourhood and the genetic background of the host cell. As such, newly discovered toxin and resistance genes spread through the population on their own, maintaining phylogenetic diversity (**Suppl. Figure S4.5**).

The evolutionary impact of HGT, and evolution of the other HGT-related processes, can be very different depending on the scale, the donor and recipient species, and the environment. Together with recent studies, our results show that microbial evolution is not driven by clonal dynamics, and that even small, closed niches can be expected to harbour a lot of genetic and phenotypic diversity.

4.4.2 Rarely beneficial genes evolve high mobility

The question of how and when HGT could be favourable is in most studies focused on the potential benefits for the cell (Martinez, 2009; Heuer et al., 2011). However, in our model we take into account both gene- and the cell-level selection, and show that it is actually the least beneficial genes, toxin genes, that evolve high mobility. As shown in **Figure 4.3**, as highly resistant individuals evolve, only toxin genes with mobility higher than their corresponding resistance gene are able to survive. We interpreted the low mobility of resistance in closed ecosystems as reduction of fitness cost of redundant genes, as individuals in the vicinity are already resistant. We have indeed shown that taking up resistance genes is rarely beneficial (**Figure 4.5**). Congruent with these interpretations is the fact that high diffusion of eDNA leads to increased gene mobility for both types of genes (**Suppl. Figure S4.1**), once again revealing how locality impacts the process of HGT. Despite the sparse presence of toxin producing individuals, resistance is sufficiently selected on the cell-level, whereas toxin genes need to evolve high gene mobility to survive. We tested the notion that less beneficial genes evolve higher gene mobility in a simple model, showing that there is indeed a significant anti-correlation between gene mobility and the frequency of selection ($P \ll 10^{-3}$,

$R^2 = 0.4611$, see Supplementary Material for implementation). As can be seen from **Figure 4.7**, the process described above results in core-like resistance repertoires with a patchy distribution of toxin genes. A similar pattern is observed in core- and accessory genomes in vivo as accessory genes are only occasionally useful (*i.e.*, pathogenicity or secondary metabolism, (Hacker and Carniel, 2001); (Norman et al., 2009)). This pattern could be explained by the dominant role of loss of genes (Wolf and Koonin, 2013), but this does not explain the discrepancy between the phylogeny of genes and cells. We show that HGT plays an important role in shaping this observed genome mosaicism (Zhaxybayeva et al., 2004).

4.4.3 Who wants to take up eDNA?

In our model, we assumed that the uptake of eDNA by cells is fixed. Although not all routes of HGT can be prevented by the cell, many bacterial species are known to have evolved regulation of DNA-uptake (Seitz and Blokesch, 2013). This suggests that uptake of eDNA is only occasionally useful. Indeed, when the uptake of eDNA is made evolvable in our model, cells avoid uptake under many conditions except high gene-loss (data not shown). What conditions make the uptake of eDNA favourable? Although DNA can serve as a nutrient (Finkel and Kolter, 2001), which might lead to transformation as a side-effect (Macfadyen et al., 2001), more direct benefits of natural competence are also possible, and poorly understood. Contrasting the regulated uptake of eDNA that is observed for most bacteria, the human pathogens of *Neisseria* sp. and *Helicobacter pylori* both constitutively take up eDNA (Seitz and Blokesch, 2013). The strong environmental pressure by the immune system might directly select for diversity, potentially explaining the constitutive uptake of eDNA. Indeed, earlier studies have shown that the uptake of eDNA can prevent the loss of genomic information (Vogan and Higgs, 2011; Johnston et al., 2013; Takeuchi et al., 2014). Indeed, by generating and maintaining diversity HGT could contribute to the evolvability of prokaryotes (Hindré et al., 2012), shaping the standing variation on which natural selection acts. In other words, an important entry point for future research would be the evolution of natural competence in relation to the environmental variability, and the signals used by the cells to induce this process.

Genes in our model are not only taken up with a fixed rate but also strictly independent of each other. In other words, two genes cannot get transferred together in a single event, as we assume no genetic linkage between genes. Although such decoupled behaviour is observed in the Vibrionaceae discussed by Cordero et al., other research commonly observes both genes together on a plasmid or MGEs (Alonso et al., 2000; Van Melderren and De Bast, 2009). Linking genes can circumvent some of the risks of HGT, and could even feedback on the earlier questions regarding the evolution of natural transformation. What conditions favour independent toxin and resistance genes, and when does the alternative solution of, for example, toxin-antitoxin systems evolve? To investigate this, the model can be extended to include the evolution of genome structure, where HGT of fragmented DNA or plasmids could include multiple genes in a single event. Answering the aforementioned questions will further increase our understanding on multilevel evolution and the selection pressures that acts on these different biological levels.

We conclude that HGT and the resulting interplay between gene-level and cell-level

selection can structure genomes and populations of microbial communities as observed *in vivo*.

Acknowledgements

The authors thank the editor and referees for their critical view, resulting in improvement of both the content and style of the manuscript. This work was supported by the European Commission 7th Framework Programme (FPFP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427).

4.5 Supplementary Material

Explanation of parameter choice

The high phenotypic diversity of bacteria has been previously explained with an evolutionary arms race, *i.e.* dynamics of continuous clonal expansions (Vetsigian et al., 2011). However, results from Preheim *et al.* (2011) and Cordero *et al.* (2012) show both phenotypic diversity and ‘deep’ phylogeny. Therefore, we chose parameters to observe both high diversity and long-lived lineages. For computational reasons, the size of the grid was kept small. For this small ecosystem, both costs and loss were set as low as was possible, but high enough to prevent too many paralogous genes from emerging within a strain. When either cost or loss is set too low, all-resistant phenotypes are readily formed, and only a few genes are retained in the population on the long run (also see **Suppl. Figure S4.3**).

Note that differential gene mobility happens in a broader range of parameters than the ones used in the main article (*e.g.* low gene loss/cost). However, either the strain diversity or the pairwise genetic distances are much lower in these scenarios. At high gene loss/cost, clonality increases and the differential gene mobility is not observed.

Lastly, we chose the influx of external adaptations (f) to be small. This way, we focus on local HGT within a single population. Higher influx-rates, and the clonal dynamics they entail, are discussed in the paragraph ‘Differential gene mobility evolves only in a closed ecosystem’.

4

Supplementary model

To confirm our interpretation that rarely beneficial genes evolve higher gene mobility we made a second model based on the model described in “In silico gene-level evolution explains microbial population diversity through differential gene mobility”. Here, toxins are not produced by the individuals themselves, but are present in the environment in different frequencies. Other model parameters are equivalent to the model from the main figures. We started the simulation with a single all-resistant individual and simulated 2,000,000 time steps. In general, resistance genes that are needed too infrequently got lost, rarely beneficial resistant genes are maintained through high gene mobility, and frequently beneficial resistance genes evolve lower gene mobility. In other words, there is an anti-correlation between gene mobility and the frequency of selection ($P \ll 10^{-3}$, $R^2 = 0.4611$).

Supplementary Videos

Supplementary videos are available on the website of Genome Biology and Evolution: <https://doi.org/10.1093/gbe/evv255>.

Supplementary Figures

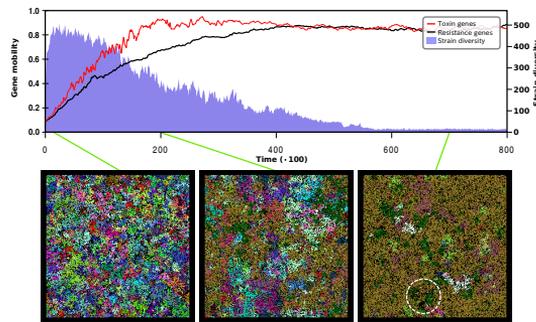


Figure S4.1: Mixing eDNA yields “selfish” DNA and very low diversity

The importance of locality in the process of HGT in our model is displayed by fully mixing the eDNA every simulation step. This way, HGT happens just as likely between neighbours than between two random individuals. In this scenario, gene mobility for both toxin and resistance genes becomes high. Furthermore, cells can become highly unfit by carrying lots of redundant genes, and can even die out locally, as can be seen from the encircled area in the bottom right. Such “selfish elements” are not observed in our regular simulations as highly mobile genes would decrease the fitness of the local population, and would subsequently be overgrown by fitter strains.

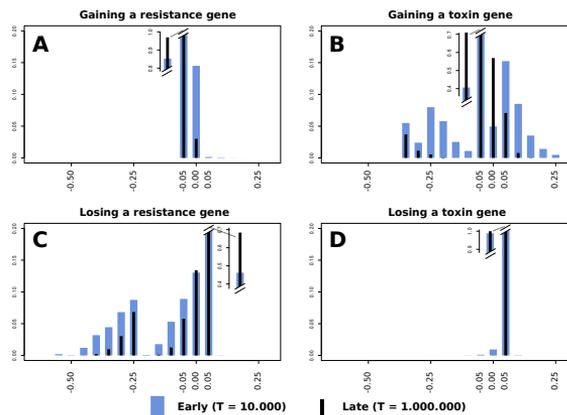


Figure S4.2: Fitness on a local scale differs markedly from fitness on a global scale.

Analogous to **Figure 4.5**, the distribution of fitness ($B_i - D_i$) advantages with respect to random interactors (instead of neighbours) from the population when gaining / losing random genes. Blue bars indicate an early time point where $T_{mobility}$ and $R_{mobility}$ have not diverged yet. Black bars indicate the end of the simulation, where $T_{mobility}$ and $R_{mobility}$ have been diverged for a long time. Each type of mutation (A-D) was repeated 1000 times for every individual in the population. All error bars were very small, and omitted for clarity. We see that the gain of toxin genes is never beneficial on a global scale, but was often beneficial on the local scale from **Figure 4.5**. Also the strong detrimental effect of losing resistance genes is absent. This can be understood when we consider an individual within a toxin production strain, for which losing resistance is extremely deleterious with respect to local competitors (*i.e.* self-killing and killing by neighbours), but much less deleterious when we consider random competitors from outside the toxin producing strain. When comparing with **Figure 4.5**, we conclude that the emergent pattern formation is important for determining the fate of genes.

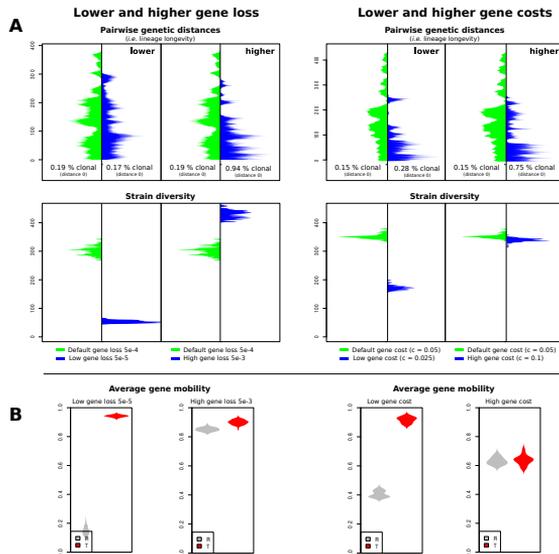


Figure S4.3: Parameter values that differ strongly from the main model do not show the combination of high phenotypic and phylogenetic diversity as observed in vivo. A) Parameters used in the main model (green) show higher strain diversity and/or higher pairwise genetic distances (*i.e.* ‘deep’ phylogeny) than when gene loss/cost are much lower/higher (blue). B) Differential gene mobility for higher and lower gene loss/cost.

4

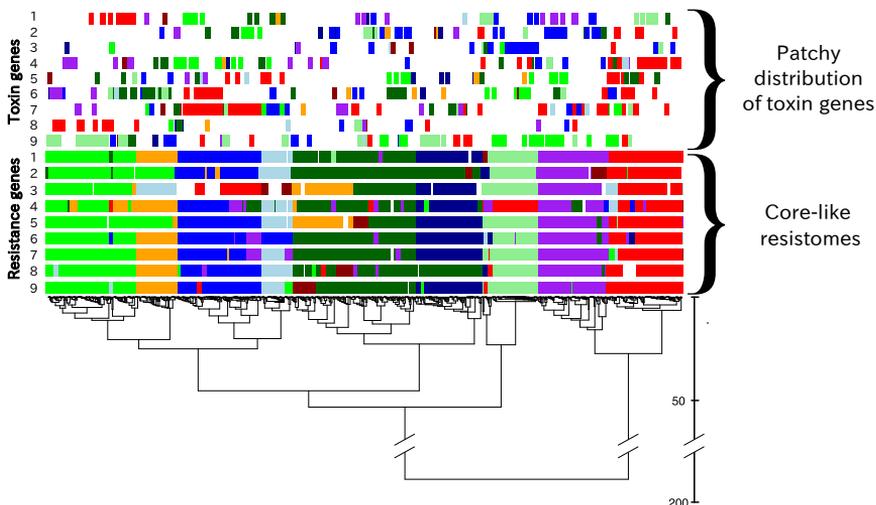


Figure S4.4: Long-term evolution structures core-like resistance repertoires with a patchy distribution of toxin genes.

To give a more complete overview of the phylogenetic structure displayed in Figure 6, we map closely related genes (colours) found in 500 individuals to the phylogenetic tree (for more details, see methods section ‘Strain diversity and phylogenetic reconstruction’).

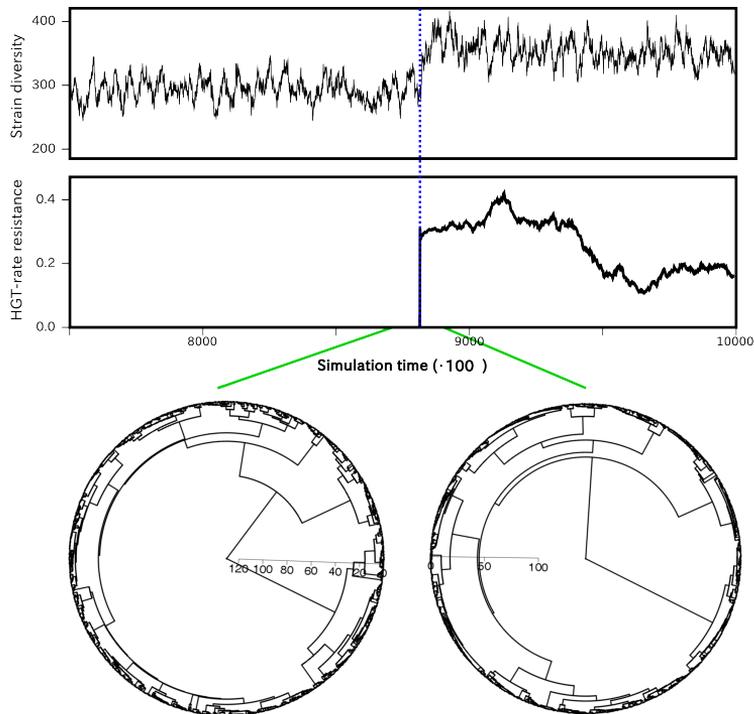
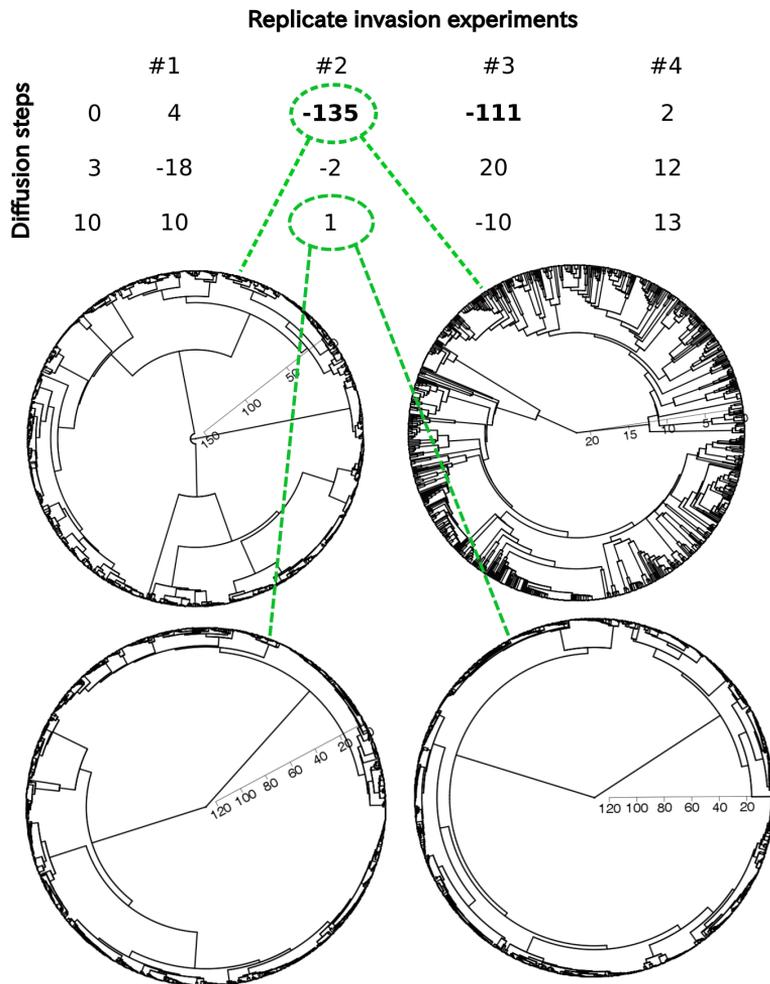


Figure S4.5: Diversity and lineages are robust to selective sweeps

When a new toxin+resistance couple is discovered by an individual (indicated by the blue vertical line), this does not lead to a genome-wide selective sweep. As strain diversity increases, also the global phylogenetic structure is maintained. Similar to the resistance genes at initialisation, the mobility of the new resistance gene first increases as it spreads, after which it decreases again.



4

Figure S 4.6: With sufficient eDNA diffusion, phylogenetic diversity is robust to genome-wide selective sweeps.

To mimic a selective sweep, evolved ecosystems are subjected to the discovery of a new toxin gene. This is done in simple invasion experiments, where a single individual is given a new toxin gene and the corresponding resistance gene. This experiment can lead to genome-wide selective sweep, or alternatively, the gene can sweep the population on its own. The experiment is repeated four times for different levels of eDNA diffusion. To show what happens, we measure the change in coalescence distance before and after the sweep (displayed in the table in the upper part of the figure). The effects are also visualised in the tree before (left) and after (right) the sweep, in the lower part of figure. When eDNA diffusion is low, 2 out of 4 replicate experiments yield a major loss in coalescence distance (-135 and -111). In these cases, the population has gone through a bottleneck, meaning that the invading genome has become the common ancestor of all individuals. This bottleneck is also apparent when comparing the tree before (left) and after the sweep (right). When eDNA diffusion is higher, the phylogenetic diversity is maintained.

“EVERYBODY DOES HAVE A BOOK IN THEM,
BUT IN MOST CASES THAT’S WHERE IT SHOULD STAY.”

—Christopher Hitchens

Chapter 5

Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements

BRAM VAN DIJK¹, HILJE M. DOEKES¹, NOBUTO TAKEUCHI², PAULIEN HOGEWEG¹

¹ UTRECHT UNIVERSITY, UTRECHT, THE NETHERLANDS

² UNIVERSITY OF AUCKLAND, AUCKLAND, NEW ZEALAND

UNDER REVISION AT eLIFE

Abstract

Horizontal gene transfer (HGT) and gene loss result in rapid changes in the gene content of bacteria. While HGT aids bacteria to adapt to new environments, it also carries risks such as selfish genetic elements (SGEs). Here, we use modelling to study how HGT of slightly beneficial genes impacts growth rates of bacterial populations, and if bacterial collectives can evolve to take up DNA despite selfish elements. We find four classes of slightly beneficial genes: indispensable, enrichable, rescuable, and unrescuable genes. Rescuable genes — genes with small fitness benefits that are lost from the population without HGT — can be collectively retained by a community that engages in costly HGT. While this ‘gene-sharing’ cannot evolve in well-mixed cultures, it does evolve in a spatial population like a biofilm. Despite enabling infection by harmful SGEs, the uptake of foreign DNA is evolutionarily maintained by the hosts, explaining the coexistence of bacteria and SGEs.

5.1 Introduction

Horizontal Gene Transfer (HGT), the transmission of genetic material between unrelated individuals, is a major factor driving prokaryotic evolution (Ochman et al., 2000; Doolittle and Zhaxybayeva, 2009; Vogan and Higgs, 2011). Recent estimates of the rate of HGT in closely related bacteria are staggeringly high (Iranzo et al., 2019; Sakoparnig et al., 2019), with HGT possibly even outpacing gradual sequence evolution (Hao and Golding, 2006; Puigbò et al., 2014; Vos et al., 2015). Combining this with the fact that prokaryotes adapt mostly through rapid gene loss (Kuo and Ochman, 2009; Morris et al., 2012), bacterial adaptation appears to be mainly driven by changes in gene content (Snel et al., 2002; Treangen and Rocha, 2011; Nowell et al., 2014). Rather than waiting for rare beneficial mutations to arise, taking up tried-and-true genes from a shared ‘mobile gene pool’ allows bacteria to adapt quickly to different ecological opportunities (Jain et al., 2003; Wiedenbeck and Cohan, 2011; Casacuberta and González, 2013; Mell and Redfield, 2014; Niehus et al., 2015; Lopatkin et al., 2016). Indeed, many bacterial species show patterns consistent with this rapid turn-over of genes, where strains from a single niche contain a relatively small set of genes, while the set of genes found by sampling strains from various niches (*i.e.* the pan-genome) is much richer (Welch et al., 2002; Lefébure and Stanhope, 2007; Touchon et al., 2009; Kim et al., 2015). Hence, genes appear to be rapidly lost from any individual lineage, but are retained in a much larger gene pool through HGT.

When considering the effects of HGT on gene content, it is important to note that HGT can add novel genes to the genome. This might happen through plasmid transfer (*i.e.* conjugation), but also through recombination. For example, when bacteria take up DNA from their environment (*i.e.* transformation), genes inserted between two homologous regions may be integrated into the genome. The DNA that has been taken up may also carry mobile genetic elements (MGEs), which can integrate into the genome without a requirement for sequence similarity. Such processes have been coined “additive HGT” (Thomas and Nielsen, 2005; Choi et al., 2012; Soucy et al., 2015), which is distinct from “replacing HGT” because of its ability to *copy* genes from one individual to another. On the one hand, DNA uptake may be beneficial for the cells, as it allows adaptations to new environments (Casacuberta and González, 2013; Mell and Redfield, 2014; Lopatkin et al., 2016) or the recovery of lost genes (Vogan and Higgs, 2011; Takeuchi et al., 2014). On the other hand, DNA uptake also poses a risk in the form of Selfish Genetic Elements (SGEs) such as transposons, and may furthermore cause chromosome disruptions and cytotoxicity (Baltrus, 2013). Finally, the translocation of DNA and the required state of competence generally comes at energetic and metabolic costs for the cell in terms of expressing and operating DNA uptake machinery (Ephrussi-Taylor and Freed, 1964; Bergé et al., 2017; Villa et al., 2019). Hence, while picking up genes can be very beneficial for bacteria when adapting to a new environment, taking up foreign DNA is also a costly and highly risky endeavour (Vogan and Higgs, 2011; Baltrus, 2013). Given these disadvantages, is the uptake of DNA ever adaptive for bacteria when the environment does *not* change? Can HGT be considered an evolved trait of bacteria, or is it only a side-effect of other unrelated processes like infection by SGEs or DNA repair (Redfield, 2001)?

To address these questions, we here present and analyse a model of a bacterial

population undergoing HGT of a single gene. We consider HGT through the uptake of genes from a shared pool of DNA, *i.e.* bacterial transformation. We assume that HGT is a costly process for the host cells, and that these costs are proportional to the rate of uptake. Such costs may reflect energetic costs for translocating DNA, growth impediments during the state of natural competence, or the various risks of incorporating foreign DNA into the genome. We show that this form of HGT has a positive impact on population growth rates by recovering slightly beneficial genes, which are hard to maintain in the population through selection alone. Based on whether or not the genes are lost from the population without HGT, and whether HGT can improve the population growth rate, we find that genes fall into one of five gene classes: (i) *indispensable genes*, that are never lost from the population, and for which HGT is therefore unnecessary and deleterious, (ii) *enrichable genes*, that are not lost from the population, but enriching the genes via HGT can nevertheless improve growth rates, (iii) *rescuable genes*, which are lost from the population without HGT, but can be rescued by HGT which improves population growth rates, (iv) *unrescuable genes* which are also lost from the population without HGT, but recovering them with HGT does not improve growth rates, and (v) *selfish genetic elements*, which confer a fitness penalty but can persist through HGT. For enrichable and rescuable genes, where HGT can increase population growth rates, we also investigate if HGT, *i.e.* the ability of cells to take up DNA, can evolve *de novo*. While the bacteria readily evolve to use HGT for enrichable genes, having sufficient donor cells to interact with, evolving HGT to ‘rescue’ rescuable genes faces a problem: HGT is needed for the gene to persist in the population, but sufficient donor cells are required to make HGT adaptive. This paradox is however resolved in a spatially structured population like a biofilm, as even a minority of donor cells can be locally abundant, giving rise to a localised ‘gene-sharing’ community that eventually overgrows the whole population. Finally, in this spatial eco-evolutionary context, HGT is evolutionarily maintained even when exploited by harmful genetic parasites, resulting in stable coexistence of bacteria and SGEs. Our model provides important insights and search images for how slightly beneficial genes may spread, or fail to spread, in an evolving microbial population.

5.2 Results

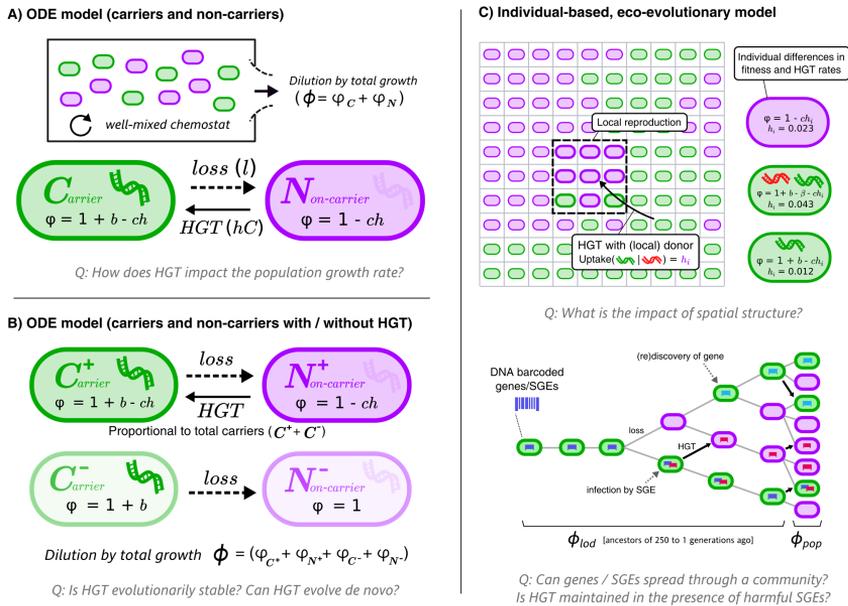


Figure 5.1: Graphical overviews of the different models: This study uses a series of models with gradually increasing complexity. The first two models (**A+B**) are composed of Ordinary Differential Equations (ODEs), and the third model (**C**) is an individual-based model (IBM). The models describe a population of bacterial cells which either carry a beneficial gene (carriers, C) or do not carry the genes (non-carriers, N), both of which engage in costly HGT (*e.g.* take up DNA or enter a state of competence). The cells are competing for a limited resource, where the intrinsic growth is 1, b is the growth rate advantage (or disadvantage) for carrying the gene, l is the rate at which the gene is lost, h is the rate of HGT (*i.e.* uptake / competence), c is the cost of HGT, φ is the growth rate of sub-populations / individual cells, and ϕ represents the total growth rate. In the IBM, each cell has an individual rate of HGT (arbitrary values are shown in the cartoon), which we use to study the *de novo* evolution of HGT. The IBM also makes a distinction between the average growth rate of the population (ϕ_{pop}) and the average growth rate of the line of descent (ϕ_{log} , previous 250 generations of cells). In the IBM, both beneficial genes (with benefit b , green) and harmful selfish genetic elements (SGEs, red) with a fitness penalty β are taken into account. Genes and SGEs are tagged with a unique barcode when they flux in, which are inherited upon reproduction or transfer. Parameters c , h and l are assumed to be positive.

Horizontal gene transfer (HGT) is a key component of bacterial evolution, which in concert with gene loss can result in rapid changes in gene content. While HGT can aid bacteria to adapt to new environments, it also carries risks since bacteria may pick up selfish genetic elements (SGEs). Here, we use modelling to study how bacterial growth rates are affected by HGT by means of the costly uptake of DNA from the environment. We study the impact of HGT of slightly beneficial genes, when bacteria can evolve to engage in HGT (*i.e.* take up DNA) to improve their growth rates, and when HGT is evolutionarily maintained by the hosts despite harmful SGEs. We first find that we can distinguish between

four classes of slightly beneficial genes: indispensable, enrichable, rescuable, and unrescuable genes. Rescuable genes — genes that confer small fitness benefits and are lost from the population in the absence of HGT — can be collectively retained by a bacterial community that engages in HGT. While this ‘gene-sharing’ cannot evolve in well-mixed cultures, it does evolve in a spatially structured population such as a biofilm. Although HGT does enable infection by harmful SGEs, HGT is nevertheless evolutionarily maintained by the hosts, explaining the stable coexistence and co-evolution of bacteria and SGEs as observed in nature.

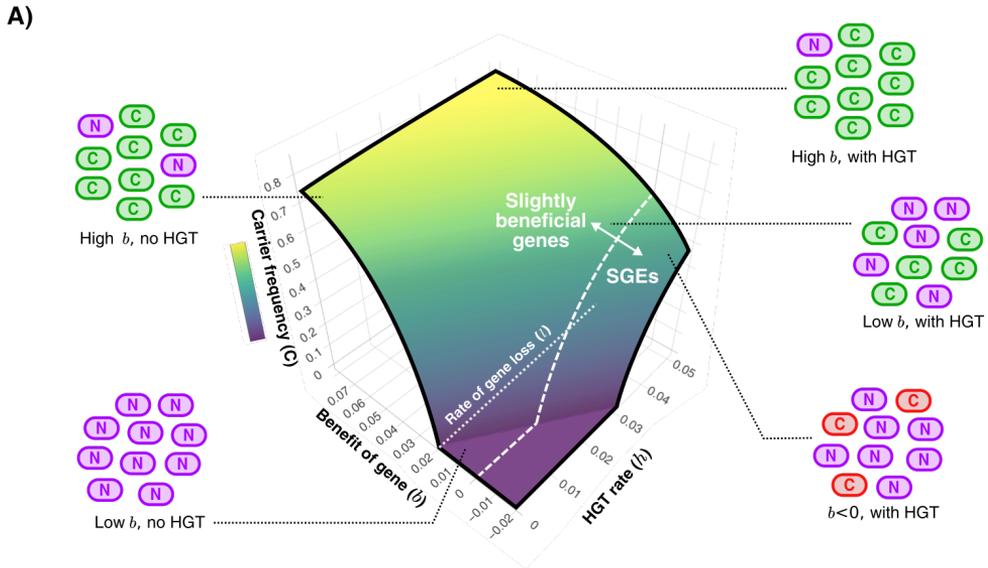
Starting with the simplest model depicted in **Figure 5.1A**, we first illustrate how the steady-state frequency of carrier cells depends on the benefit of the gene (b) and the rate of HGT (h). **Figure 5.2A** shows that if the gene is sufficiently beneficial, most of the population will consist of carrier cells with or without HGT. Despite being continuously lost, these genes are beneficial enough to readily persist in the population through selection. An increased rate of HGT results in only marginally more carrier cells. For genes with a much smaller benefit, HGT can have a large impact on the frequency of carrier cells in the population. In fact, if the benefit is very small ($b < l$, white dotted line), carriers do not survive in the absence of HGT at all, but can occur in fairly high frequencies with sufficient HGT. Note however that the mere survival of carriers with beneficial genes does not imply a positive impact on the population growth rate, as the model assumes HGT comes at a cost. Actually, at sufficiently high rates of HGT, carrier cells with costly genes ($b < 0$) can also persist in the population, which by definition is deleterious for growth. These costly genes could either be genes that are expressed but not useful in the current environment, or Selfish Genetic Elements (SGEs). Throughout this study, we consider genes with $b < 0$ to be SGEs.

5.2.1 Slightly beneficial genes fall into distinct gene classes

To better understand the impact of HGT, we next study how HGT impacts the population growth rate (ϕ). The population growth rate in steady state is given by **Equation 5.1** displayed below (see full derivation in Supplementary Section 1). The function is comprised of two parts; one where the population consists only of non-carriers (if $h \leq l - b$), and one where carriers survive and the gene persists within the population (if $h > l - b$). When the gene persists, an optimal growth rate is found at $h_{opt} = \sqrt{bl/c} - b$. (see Supplementary Material).

$$\phi^*(h) = \begin{cases} 1 - ch & \text{if } h \leq (l - b) \text{ (gene cannot persist)} \\ 1 - ch + b - \frac{bl}{b+h} & \text{if } h > (l - b) \text{ (gene persists).} \end{cases} \quad (5.1)$$

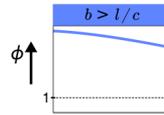
By analysing **Equation 5.1**, we find that we can distinguish distinct classes of genes depending on (i) whether HGT is required for the gene to persist within the population, and (ii) whether HGT is beneficial for the population growth rate (**Figure 5.2B**). When genes are highly beneficial ($b > l/c$), HGT is not required for the gene to persist, and HGT does not improve the population growth rate. In other words, although transferring these *indispensable genes* yields a small increase in the number of carrier cells, this does not outweigh the costs of HGT. When considering lower values of b , HGT is still not required for the gene to persist within the population, but transferring these *enrichable genes* is nevertheless beneficial



B)

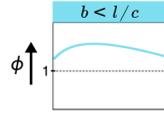
Indispensable genes
 These genes are readily maintained. Transferring these genes with costly HGT always diminishes the growth rate of the population.

HGT is...	
REQUIRED	BENEFICIAL
No	No



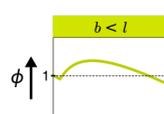
Enrichable genes
 These genes are maintained without HGT, but HGT can improve the population growth rate.

REQUIRED	BENEFICIAL
No	Yes



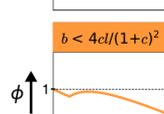
Rescuable genes
 These genes are lost from the population without HGT. If the rate of HGT is high enough, the genes persist and the population growth rate is improved.

REQUIRED	BENEFICIAL
Yes	Yes



Unrescuable genes
 These genes are lost from the population without HGT. Despite being maintained at intermediate rates of HGT, the population grows fastest without HGT.

REQUIRED	BENEFICIAL
Yes	No



Selfish genetic elements
 Selfish genetic elements (SGEs) can only persist at very high rates of HGT. When they do, the population growth rate is diminished.

REQUIRED	BENEFICIAL
Yes	No

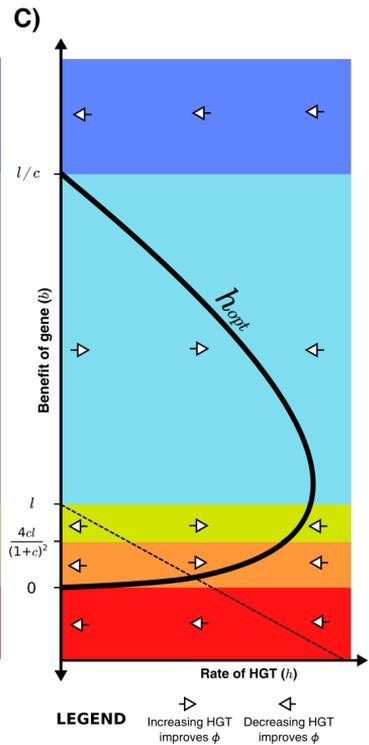
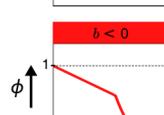


Figure 5.2: HGT can help genes persist in the population, resulting in distinct gene classes of slightly beneficial genes **A)** For the model shown in **Figure 5.1A**, the frequency of carrier cells is shown in a 3D surface plot for different values of b and h . This function is derived in the Supplementary Material and given by Equation 5.10, and here drawn for $l = 0.02$ and $c = 0.2$. The white dashed line ($b = 0$) gives the boundary between slightly beneficial genes and selfish genetic elements (SGEs). Cartoons illustrate how, for a very beneficial gene (high b), HGT leads only to a mild increase in carrier cells, how HGT has a large impact when the gene brings a smaller fitness effect (low b), and how SGEs can also persist with high HGT rates ($b < 0$). **B)** Different classes of slightly beneficial genes can be distinguished based on (i) if HGT is *required* for the gene to persist within the population and (ii) if HGT is *beneficial* for population growth rates. The graphs on the right-hand side show, for each of these classes, how an increasing rate of HGT (x-axis) influences the population growth ϕ (y-axis). **C)** A bifurcation diagram shows how the population growth rate is either improved or diminished by HGT for different values of the rate of HGT (h , x-axis) and the benefit parameter (b , y-axis). The HGT rate that optimises population growth rates ($h_{opt} = \sqrt{bl/c} - b$) is depicted by the thick black curve. The dashed line is given by $h = l - b$, above which the genes are able to persist in the population. Finally, white arrows depict whether $\delta\phi/\delta h$ is positive or negative, indicating how more/less HGT changes the population growth rate.

for population growth rates. For even lower benefit ($b < l$), HGT is a necessity for the gene to persist within the population, but the population growth rate can be improved by means of intermediate rates of HGT. We call these genes *rescuable genes*. If we consider genes with even smaller fitness effects ($b < 4cl/(1+c)^2$), HGT is still required for the survival of these genes, but the population growth rates are highest in the absence of HGT. Thus, despite being defined as a beneficial gene ($b > 0$), transferring these *unrescuable genes* is not beneficial. Finally, we can consider SGEs, genes with a negative effect on fitness ($b < 0$). These genetic parasites can only persist in the population at very high rates of HGT, but are of course never beneficial for the population growth rate. **Figure 5.2C** shows a bifurcation diagram that summarises how increasing or decreasing rates of HGT impact the population growth rate for these different classes.

5.2.2 HGT is an evolutionarily stable strategy, but cannot evolve to ‘rescue’ rescuable genes

By analysing the simple model of cells undergoing HGT, we have found 5 distinct gene classes. For two of these classes, namely enrichable and rescuable genes, moderate rates of HGT improve the population growth rates. We next study (i) whether HGT of enrichable and rescuable genes is an evolutionarily stable strategy, and (ii) if bacteria can evolve this strategy *de novo*. To answer these questions, we consider two competing species: one that takes up DNA, and one that does not (HGT^+ and HGT^- respectively, see **Figure 5.1B**). With this model, we have studied the evolution of DNA uptake by means of adaptive dynamics (Metz et al., 1995). If HGT^- cannot invade HGT^+ , we call HGT an evolutionarily stable strategy, and if HGT^+ can invade HGT^- we call HGT evolvable.

We found that HGT is an evolutionarily stable strategy for both enrichable and

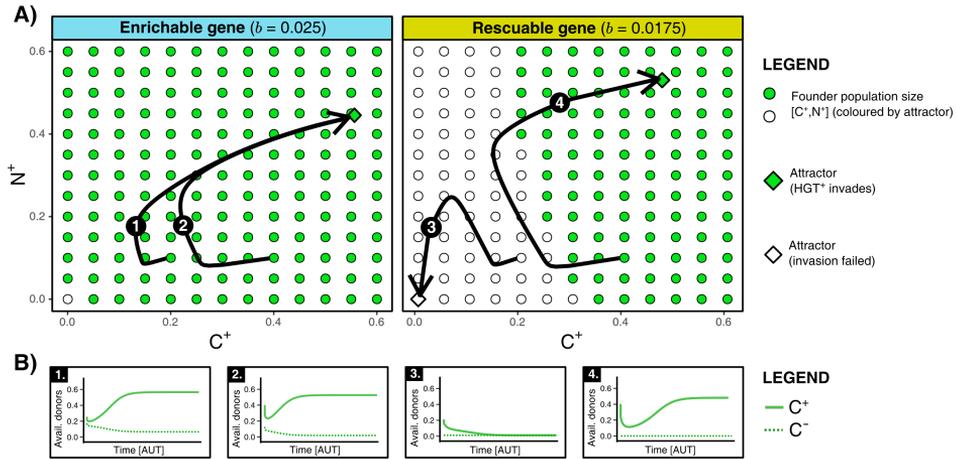


Figure 5.3: HGT is an evolutionarily stable strategy, but is evolutionarily inaccessible for rescuable genes due to a lack of gene-carrying donor cells. **A)** For an enrichable and a rescuable gene ($b = 0.025$ and $b = 0.0175$ respectively), a 2D projection of the 4D state space is shown. For various founder sizes (combinations of carriers, C^+ , and non-carriers, N^+), the result of invasion of HGT^+ (that has the optimal rate of HGT, $h = h_{opt}$) into HGT^- ($h = 0$) is shown. C^+/N^+ -combination that successfully invade are annotated as green dots, while failed invasions are coloured white. HGT^+ always successfully invades for enrichable genes. For a rescuable gene, low founder population sizes of C^+ and N^+ (white dots) fail to invade, whereas they can invade at higher population sizes (green dots). Black arrows (1-4) show the trajectories starting from two founder population sizes. **B)** for the four trajectories from panel A, the graphs show the temporal dynamics of gene-carrying donor cells.

rescuable genes, but that HGT is evolvable only for enrichable genes (see Supplementary Material for full analysis). In other words, invasion of HGT^+ -mutants is not possible with respect to rescuable genes. Even when we assume that the invading HGT^+ -mutant has the optimal rate of HGT, it cannot invade into a population of HGT^- cells in steady state. These results were confirmed by numerical analysis, which indeed shows that HGT^+ only invades when the founding population size of HGT^+ (C^+/N^+) is relatively large (see **Figure 5.3A**). This failure to reach the alternative (fitter) evolutionary attractor is caused by positive frequency-dependent selection (known as the Allee effect). Invading mutants, *i.e.* a small population of HGT^+ cells, contain few carrier cells to act as donors for HGT. Moreover, since the resident population of HGT^- is also not able to retain the rescuable genes, the resident population can also not serve as a donor (see **Figure 5.3B**). As such, the costs of HGT for an invading HGT^+ -mutant do not outweigh the potential benefits. In summary, while HGT is an evolutionarily stable strategy, cells cannot evolve HGT to ‘rescue’ rescuable genes.

5.2.3 Spatial structure hinders the maintenance of genes, making HGT adaptive for a wider range of genes

So far, we have studied a well-mixed population of cells that undergoes all-against-all competition, and found that HGT is advantageous for slightly beneficial genes that (i) are not too beneficial, as these genes readily persist within the population without

HGT, and (ii) are beneficial enough to compensate for the costly HGT. Next, we study the same dynamics of carrier and non-carrier cells in a spatially explicit, eco-evolutionary context. We do this by implementing an individual-based model (IBM), where bacterial cells reside on a grid, interactions are local, and events like HGT and gene loss are implemented as stochastic processes (see Methods and **Figure 5.1C**). When the cells on this grid are sufficiently mixed each time step, the IBM should approximate the dynamics of the ODE model. However, when cellular mixing is minimal, the resulting spatially structured population is more analogous to that of a biofilm. What is the effect of this spatial structure?

We first analysed the IBM for a wide variety of values for b and h , and measured the average growth rates ϕ in the population. We can thus evaluate whether the

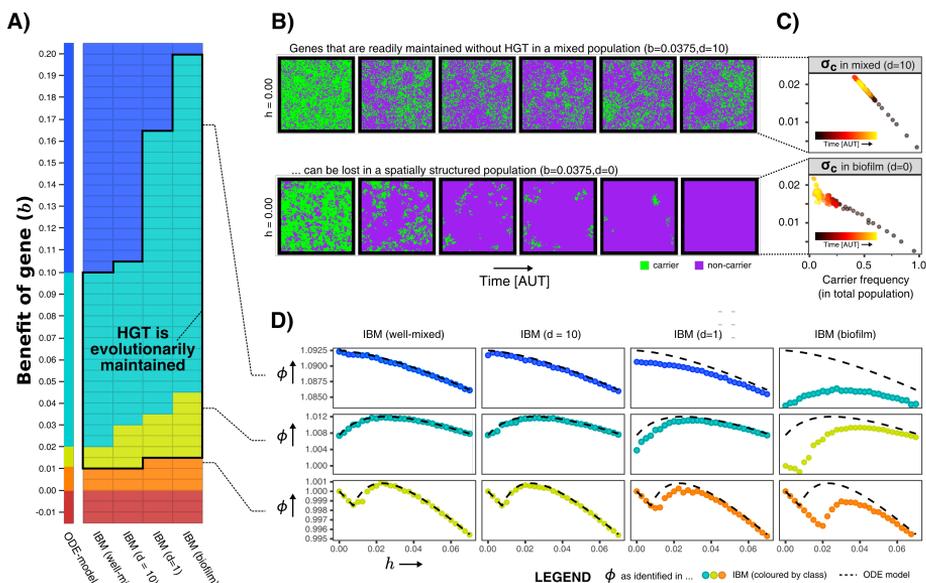


Figure 5.4: Spatial structure hinders the maintenance of slightly beneficial genes **A)** Each tile in this table represents a series of simulations in the individual-based model (IBM), where we first test which gene class (background colour) is found when sweeping over different HGT-rates (h -values), and next test whether HGT (*i.e.* DNA uptake) is evolutionarily maintained when starting with a population consisting of only carrier cells with $h = 0.05$ (shown with black outline). This was tested for the well-mixed IBM and the IBM with different levels of mixing (d). The continuum of gene classes from the ODE-model is presented for comparison. Colours are the same as in **Figure 5.2** (blue=indispensable, cyan=enrichable, yellow=rescuable, orange=unrescuable, red=selfish genetic element). **B)** Shown is the spatial grid of the IBM for two simulations with the same value of b , and no HGT. The gene readily persists in the mixed IBM (top panel, $d = 10$), while the gene does not persist in the spatially structured population (bottom panel, $d = 0$). **C)** For the simulations shown in B, we plot σ_c (see Methods) against the frequency of carrier cells. This value of σ_c indicates how much fitter carrier cells are than their local competitors, which decreases to 0 when all local competitors are also carriers. When comparing the top and bottom panel, this shows how clumping hinders the effective benefit of carrying a gene. **D)** For 3 rows from the table of A, it is illustrated how the effect of spatial clumping illustrated in B and C modifies the gene class found for specific b -values. The dashed line indicates the growth rates predicted by the ODE model.

mentioned gene classes (indispensable, enrichable, rescuable, unrescuable genes, and SGEs) are found under the same conditions as in the ODE model. **Figure 5.4A** shows that, when the IBM is well-mixed, the gene classes indeed occur at values of b identical to the ODE model. However, the gene classes shift to higher values of b when mixing is decreased, making the range of benefits in which genes are classified as enrichable and rescuable much broader. In these biofilm populations, HGT was indeed found to be evolutionarily stable for this wider range of fitness-effects (black outline), illustrating that it is not only the value of b , but also the ecological context in which a gene finds itself that determines whether or not HGT is adaptive.

What causes these gene classes to shift depending on this spatial context? How does an enrichable gene in the well-mixed system become rescuable in the spatially structured population, as though it is less beneficial? **Figure 5.4B** shows how this can be intuitively understood by taking into account how, with low mixing, individuals in a spatial system mostly compete with their own kind (*i.e.* progeny and conspecifics). Even when the majority of the population consists of non-carriers, carriers are still competing mostly with other carrier cells. Thus, the effective benefit of carrying the gene is lower in a biofilm, hence the gene becomes harder to maintain within the population. In **Figure 5.4C** is shown that, while carrier cells in well-mixed populations experience a competitive advantage of $\sim 2\%$ when carriers make up approximately half the population, carriers in a biofilm only reach a similar competitive advantage at very low carrier frequencies, *i.e.* when the carriers are almost extinct. At this point, the gene will readily be lost stochastically. The hampered ability of spatially structured populations to retain slightly beneficial genes, indeed changes how the population growth rate depends on the rate of HGT (**Figure 5.4D**).

5.2.4 Bacteria evolve DNA uptake for rescuable genes only in a spatially structured population

The results described in the previous section illustrate that HGT (*i.e.* the uptake of DNA) is an evolutionarily stable strategy for a much broader range of b -values (fitness effects of genes) in a spatially structured population than in a well-mixed culture. Many more genes are furthermore classified as rescuable in these spatially structured populations, meaning that they can only persist through HGT. We have concluded in a previous section that HGT cannot evolve to ‘rescue’ these rescuable genes in populations that are well-mixed, fully deterministic, and by only considering a single HGT^+ mutant type at a time. In the IBM on the other hand, the population is spatially structured, events are stochastic, and each individual cell has its own rate of DNA uptake. Can these different assumptions help to alleviate the Allee effect mediated by a lack of donor cells, which prevents the *de novo* evolution of HGT?

To answer the question posed above, we allowed the HGT-rate h (*i.e.* the costly uptake of DNA) of all individuals in the IBM to evolve (see Methods). When a non-carrier interacts with a (local) carrier, the h -value of this non-carrier (*i.e.* the recipient) determines the probability of accepting the gene. For simplicity, we will call individuals with an h -parameter greater than 0.02 HGT^+ , and the others

HGT^- . We start with a non-carrier population of HGT^- cells (with $h = 0.00$), simulate this population for some time (20,000 time steps), and then allow cells to sporadically discover rescuable genes. Since rescuable genes cannot persist without HGT, the fate of this gene depends on the ability of cells to engage in (local) HGT. Using this protocol, we investigate if the rescuable gene is able to spread through the evolution of HGT. We found that HGT never evolved for rescuable genes in well-mixed populations (**Figure 5.5A**), consistent with our prior results in the well-mixed ODE model. Thus, we can conclude that the level of stochasticity in the IBM is insufficient to overcome the aforementioned Allee effect caused by a lack of donor cells.

In the spatially structured population, HGT of rescuable genes *does* in fact emerge, therewith ‘rescuing’ the rescuable genes (**Figure 5.5B**). Interestingly however, we found that HGT did not always evolve immediately after the influx of rescuable genes started (yellow arrow), but nevertheless spread steadily once attained. To further elucidate the spread of genes, we barcoded each newly discovered gene with a unique ID, and visualised these on the spatial grid with different colours (**Figure 5.5C**). Initially, rescuable genes fail to invade, even though different barcodes may locally persist for a while (episode I). After some time however, one gene (green) manages to persist within a local community of transferring cells (episode II). This sets in motion a positive feedback mechanism, where the local abundance of the green gene alleviates the lack of donor cells, transforming nearby HGT^+ -mutants into carriers, and so on (also see Supplementary Movie). This emergent ‘gene-sharing’ community eventually overgrows the other cells, and the rescuable gene ultimately persists in up to $\sim 70\%$ of the population. After the influx of rescuable gene is stopped (episode III), the gene readily persists within the population, showing how this transferring community does not depend on the continuous influx of genes. In summary, HGT of rescuable genes through the uptake of DNA can only evolve if transfer happens within spatially localised sub-populations, and not under well-mixed conditions modelled by mass-action. Through a local ‘nucleation event’, communities can reach the alternative stable state that can maintain the rescuable gene. **Figure 5.5D** summarises the outcome of *de novo* HGT evolution for a broad range of genes (b -values) with different levels of mixing, revealing how the uptake of DNA evolves for many more genes in a spatially structured population. Moreover, while HGT of enrichable genes always evolved, HGT only evolved for rescuable genes in spatially structured populations. Finally, as expected from prior results, HGT never evolved for indispensable and unrescuable genes.

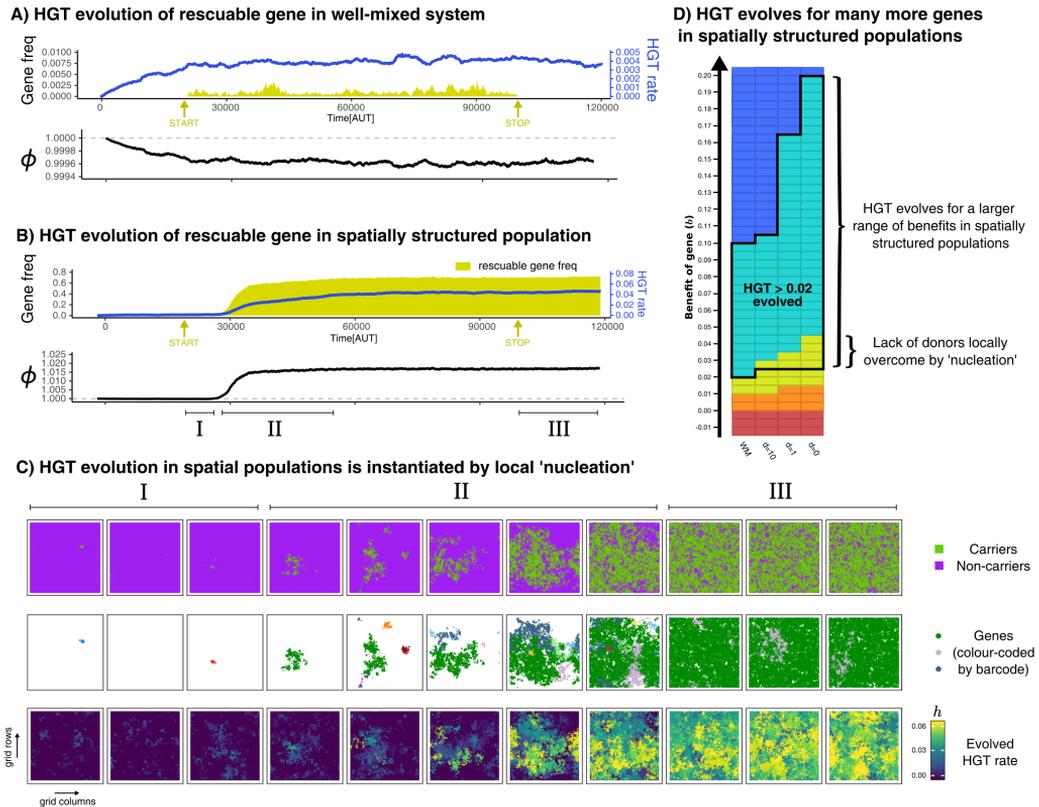


Figure 5.5: HGT of rescuable genes only evolves in spatially structured populations due to the emergence of 'gene-sharing' communities: A-B) Both panels show the frequency of a rescuable gene (yellow area) that is discovered with a very low probability ($5 \cdot 10^{-6}$ per time step), the mean evolved rate of DNA uptake within the population (blue line), and the growth rate of the population (black). Note that A and B have a different range in the y-axis for clarity. **C)** Different visualisation of the grid show, in the spatially structured populations, carrier cells with a rescuable gene (colour coded by the unique barcodes) spread after a local 'nucleation event'. A positive feedback loop follows, resulting in a 'gene-sharing' community which slowly overgrows the rest of the population. **D)** The outcome of *de novo* HGT evolution for the same combinations of fitness-effects and mixing as in Figure 5.4A. Parameters used: $h_{init} = 0.0$, $u = 5e-3$, $m = 0.05$, $l = 0.02$, $c = 0.1$, $f = 5 \cdot 10^{-6}$, $f_{start} = 20.000$, $f_{stop} = 100.000$, $n = 400$ (i.e. $N = 400^2$). For the well-mixed population, we chose a rescuable gene with the highest benefit from Figure 5.4A ($b = 0.0175$), and for the spatially structured population we used $b = 0.030$ (the average of the much broader range of rescuable genes). Both these genes require HGT to persist, and are thus rescuable within their own spatial context.

5.2.5 HGT is evolutionarily maintained in the presence of harmful SGEs

We have shown that HGT can be adaptive and evolvable for bacteria in order to enrich or rescue slightly beneficial genes. We next investigated if HGT can be maintained under the pressure of harmful SGEs, genetic parasites that spread through horizontal transfer. Our earlier analysis has shown that, when the rate of HGT cannot evolve, SGEs (genes with $b < 0$) can persist within the population as long as $h > l - b$. However, when the rate of HGT is allowed to evolve, bacteria may lower their rate of DNA uptake to avoid these genetic parasites. Therefore, we next investigate whether bacteria in the IBM will maintain their ability to take up foreign DNA in the presence of SGEs. For this, we consider a population that evolved to engage in HGT of a rescuable gene ($b=0.03$), and expose this population to a low influx of SGEs which confer a fitness penalty (β). We study if these SGEs, despite their fitness penalty, can persist within this bacterial population, and if HGT is evolutionarily maintained by the hosts. **Figure 5.6A** shows that, when the fitness penalty of the SGEs is small relative to the benefit of the rescuable gene (hereafter called “weak SGEs”, $\beta = 0.01$), these genetic parasites quickly rise to very high frequencies within the population. Although the host cells gradually evolve lower HGT rates in response (from $h \pm 0.05$ it stabilises around $h \pm 0.04$, also see **Suppl. Figure S5.5C**), HGT, the rescuable gene, and the SGEs are evolutionarily maintained. When the influx of SGEs is stopped, the cells (and their beneficial gene) stably coexists with these genetic parasites.

Strikingly, if we introduce SGEs whose fitness penalty is greater than the benefit of the gene (“strong SGEs”, $\beta = 0.04$), we also observe the coexistence of cells, rescuable genes, and SGEs. By looking at the initial invasion dynamics (**Figure 5.6B**), we can see that these strong SGEs cannot rise to very high frequencies. As the hosts evolve lower rates of DNA uptake, these genetic parasites are pushed to very low frequencies. However, the reduced threat of genetic parasites causes the host cells to once again increase their rates of DNA uptake, leading to a secondary outbreak of SGEs (**Figure 5.6B**, from $T=300,000$ onwards). It is interesting to note that, while the population growth rates (ϕ_{pop}) clearly decrease due to this second infection, the growth rates along the line-of-descent (ϕ_{lod} , see methods) remains largely unaffected. Thus, while a sub-set of the population has been infected, individuals in this infected strain will not be amongst the long-term ancestors. Counter-intuitively, strong SGEs only have a minor impact on bacterial growth rates, as they are only retained at very low frequencies. Weaker SGEs instead impose a significant burden on the population by rising to much higher frequencies. Indeed, when SGEs are very costly, the population purges them entirely by evolving lower rates of DNA uptake (also see Supplementary material, **Suppl. Figure S5.5C**). Finally, note how stopping the influx of SGEs does not impact the long-term coexistence of cells, beneficial genes, and these strong SGEs (**Figure 5.6C** and **Suppl. Figure S5.5C**).

To better understand the co-evolutionary process between SGEs and bacteria engaging in HGT of rescuable genes, **Figure 5.6D** shows long-term dynamics of barcoded SGEs in this spatial system. Although a diverse set of SGEs are initially discovered in parallel (coloured by their unique barcode), eventually only a single barcode remains after the influx of SGEs is stopped. Moreover, it can also be seen how SGEs are either locally abundant, or entirely absent. Thus, spatially

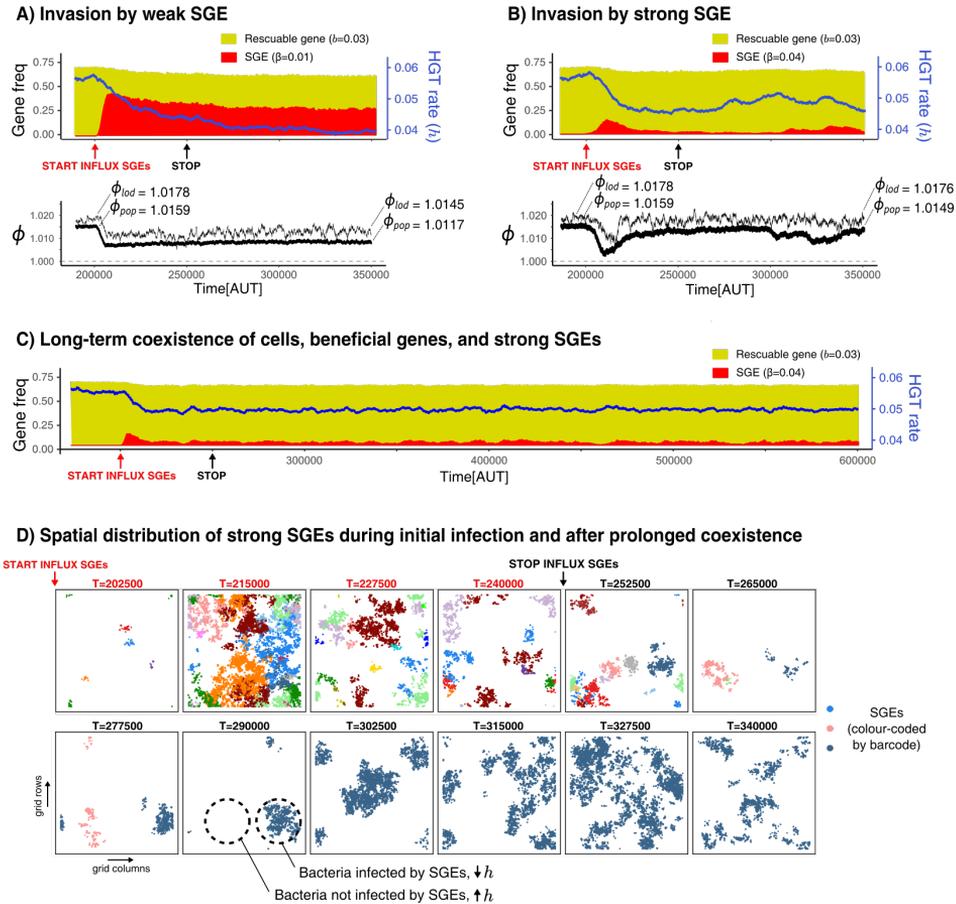


Figure 5.6: Selfish Genetic Elements (SGEs) can invade and stably coexist with their host cell: A-B Temporal dynamics for a population that has evolved to rescue a slightly beneficial gene ($b = 0.03$) invaded by a weak SGE (A, $\beta = 0.01$) and a strong SGE (B, $\beta = 0.04$) respectively. The blue line indicates the rate of HGT (i.e. DNA uptake) as evolved by the host cells. In the bottom graphs, the thick black line is the average growth rate of the population (ϕ_{pop}), and the thin black line is ϕ_{lod} , the average growth rate along the line of descent (250-1 generations ago). The ϕ_{pop} and ϕ_{lod} that are annotated with the dashed lines are the average of the first/final 200 generations. **C)** This panel shows the long-term coexistence of cells, beneficial genes, and a strong SGEs ($\beta = 0.04$). **D)** Panels show the spatial distribution of SGEs (coloured by their unique barcodes). The top row shows this during the invasion (open ecosystem) and the bottom row shows this during prolonged coexistence (closed ecosystem). Note that the empty sites (white) only indicate the absence of SGEs, not of bacterial cells, which are instead present in every grid point. Parameters used: h -parameters and frequency of carriers as evolved from Figure 5.5, $u = 5e-3$, $m = 0.05$, $l = 0.02$, $c = 0.1$, $i = 1e-5$, $i_{start} = 200,000$, $i_{stop} = 250,000$, $n = 400$ (i.e. $N=400^2$).

separated strains of bacteria experience opposing selection pressures for taking up DNA. Lower rates of uptake are favoured in the presence of these strong SGEs, but higher rates of uptake are favoured when these genetic parasites have (locally) died out. Indeed, this heterogeneity of SGEs is crucial for the strong SGEs to persist, as well-mixed populations can only retain weaker SGEs (see **Suppl. Figure S5.5**). Interestingly, we also found that strong SGEs failed to persist when HGT was too localised (*e.g.* only between neighbouring cells), as the SGEs then could not escape to a new pool of hosts that have high rates of DNA uptake (**Suppl. Figure S5.6**). We conclude that, in a spatially structured population, strong SGEs can stably coexist in a bacterial population which maintains its ability to engage in HGT to ‘rescue’ rescuable genes.

5.3 Discussion

We have studied the balance between the advantages and disadvantages of HGT by modelling a simple bacterial population undergoing the uptake of genes from a shared DNA pool. Our analysis shows that we can categorise slightly beneficial genes based on whether genes are lost from the population without HGT, and whether HGT of these genes can improve the population growth rate. This results in five distinct gene classes: (i) *indispensable genes*, that readily persist within the population and for which HGT is therefore always deleterious, (ii) *enrichable genes* which are not lost from the population without HGT, but moderate rates of HGT are adaptive, (iii) *rescuable genes* which are lost from the population without HGT, but can be rescued by HGT which improves population growth rates, and (iv) *unrescuable genes*, that are also lost from the population without HGT, but recovering them with HGT does not improve population growth rates, and (v) *selfish genetic elements*, genes that confer a fitness penalty, but can persist within the population with HGT. We further investigated if HGT of these respective gene classes is an evolutionarily stable strategy, and if the bacterial cells can evolve to engage in HGT of these genes *de novo*. We found that horizontal transfer of enrichable and rescuable genes is indeed a evolutionarily stable strategy, but can only evolve from scratch for enrichable genes. The evolution of HGT to ‘rescue’ a rescuable gene faces a problem under well-mixed conditions: HGT is required for the gene to persist, but sufficient carriers of the gene are necessary to evolve HGT. By modelling this process in a spatially structured population, we show that HGT can nevertheless evolve for these rescuable genes. As carriers of the gene can be locally abundant, emergent communities form that locally retain the gene via HGT, therewith slowly outgrowing other individuals. Finally we show that once stable transferring communities have evolved, selfish genetic elements (SGEs) can stably coexist with the bacterial population and the beneficial genes. In spite of these genetic parasites exploiting the host’s ability to transfer, HGT is evolutionarily maintained, providing a doorway to the co-evolutionary process between bacteria and SGEs.

Here, we have used a simple model which ignores many of the complexities of natural bacterial populations, *e.g.* by assuming a constant environment and studying the impact of HGT one gene at a time. In reality, microbial ecosystems experience many changing selection pressures. Besides fluctuations in the available resources, traits like antimicrobial resistance, toxin production, and cooperative behaviour may confer large fitness benefits only under specific ecological circumstances (Riley and Wertz, 2002; Cordero et al., 2012a; Vogwill and MacLean, 2015; Gerardin et al., 2016; Hehemann et al., 2016; Dimitriu et al., 2019). Moreover, when considering the impact of HGT, we should also consider its implications for the genome as a whole, rather than only considering single genes. To take into account the impact of HGT on the whole genome, it is important to consider the possibility of differential gene mobility, where genes associated with MGEs may transfer much more frequently than other genes (Rankin et al., 2011). However, we should then also take into account selection pressures that allow MGEs to have both mutualistic and parasitic relationships with their host (Harrison et al., 2015). Although our extremely simple model ignores all the above intricacies, it retrospectively shines light on dynamics observed in our earlier eco-evolutionary modelling of antagonistic interactions (van

Dijk and Hogeweg, 2015). In this multilevel model, which included changing selection pressures and the co-evolution of cells and mutualistic/parasitic genes, we showed that rarely beneficial toxin genes evolved to transfer much more frequently than their corresponding resistance factors. Despite the complex eco-evolutionary dynamics in this model, we can in hindsight understand how toxin genes evolved to correspond to ‘rescuable’ genes, and the resistance genes evolved to be ‘indispensable’. However, the rescuable toxin genes in this model experienced alternating selection pressures, only sporadically being highly beneficial for their hosts. Therefore, it remained unclear whether the ability of HGT to rescue genes required such alternating selection pressures, which was suggested to be important in earlier studies (Bergstrom et al., 2000; Rankin et al., 2011). Here, our simple model has revealed that bacteria may also benefit from preferentially mobilising genes that are, constantly, slightly beneficial.

We have studied the impact of HGT by considering bacterial transformation as the mechanism by which genes transfer, *i.e.* by assuming (i) engaging in HGT is continuously costly and not only when it is successful, (ii) all carrier cells can act as a donor regardless of their ability to take up DNA themselves, and (iii) the ability for HGT is determined by the recipient cell. However, note that similar gene classes and comparable dynamics are observed when modifying these assumptions in our model (see Supplementary material, **Suppl. Figure S5.1**, **Suppl. Figure S5.2**, and **Suppl. Figure S5.3**), and that some of our key results may therefore hold for different mechanisms of HGT like conjugation. Indeed, there are some interesting parallels with earlier modelling work on plasmid persistence by Bergstrom et al. (2000). On the one hand, this study shows that if genes carried on plasmids are sufficiently beneficial, HGT is not necessary, and plasmid-bearing cells are outcompeted by plasmid-free cells that have integrated the beneficial trait into the bacterial chromosome. These dynamics resemble those of the indispensable genes in our model, where transfer is not adaptive as the trait readily persists without HGT. On the other hand, Bergstrom et al. show that plasmids that are intermittently favoured can be retained by shuttling them back and forth between strains or species, reminiscent of the dynamics we have revealed for rescuable genes. However, there are of course some very important differences between conjugation and transformation, where plasmids for example confers other burdens on the host (San Millan and Craig maclean, 2019), cells require direct contact to conjugate, and the F-factor necessary for conjugation can itself reside either on a plasmid or on the bacterial chromosome. A promising direction for future modelling therefore is to more directly compare these two distinct mechanisms of HGT, so that we may one day learn to predict which genes will spread in an evolving microbial population.

Besides investigating the impact of HGT for a range of different fitness effects, we have also shown how spatial structure is a key component for the emergence of HGT of rescuable genes. Both conjugation and transformation have indeed been observed to occur more frequently in biofilms than in well-mixed cultures (Madsen et al., 2012), and plasmids have furthermore been shown to be more persistent in biofilms (Stalder et al., 2020). On the other hand, spatial structure can slow down adaptation in asexual populations because individuals are mostly competing with their related conspecifics (Gordo and Campos, 2006; Habets et al., 2007; Chacón and Harcombe, 2019). Relatedness has indeed been shown to be an

important factor in stabilising HGT, for example of cooperative traits (Mc Ginty et al., 2011). Our model shows that, also without explicitly taking cooperation into account, HGT can only evolve in ‘gene-sharing’ communities which emerge in by local reproduction in spatially structured populations. Thus, not only are relatedness and spatial structure necessarily intertwined, they are crucial for the rare ‘nucleation events’ that initiates evolution towards increased rates of HGT. Intriguingly, similar nucleation events have been observed in origin of life studies (Wu and Higgs, 2012) and models of microbial antagonistic interactions (Kotil and Vetsigian, 2018). These types of emergent evolutionary transitions highlight how studying evolution under well-mixed conditions, and one mutant at a time, can be highly misleading. Studying biological systems in a spatial context will help us to better understand which eco-evolutionary outcomes are accessible, and maintainable, by evolution.

5.3.1 Horizontal Gene Transfer: rescue or catastrophe?

In nature, HGT can happen through a variety of mechanisms that each have their own potential advantages and disadvantages for the host cell (Vogan and Higgs, 2011; Baltrus, 2013). Bacteria do not always have full control over the rates at which HGT happens, especially when considering it as a side-effects of other processes (Redfield, 2001). However, it remains an intriguing question under which specific circumstances bacteria benefit from HGT, whether it is a side-effect or not. By abstracting away from the different mechanisms of HGT, and what it means for a gene to be “beneficial”, we have revealed the conditions under which HGT is an adaptive trait for the host cells. In a similar spirit, earlier modelling by Vogan & Higgs has shown that HGT can be adaptive with respect to genes that are frequently lost (Vogan and Higgs, 2011). However, in their work, natural selection eventually favoured improved replication accuracy, therewith decreasing the advantage of HGT. Other models have shown that HGT is beneficial to mitigate the effects of Mullers Ratchet (Muller, 1964) by decreasing assortment load (Takeuchi et al., 2014; Vig-Milkovics et al., 2019), analogous to the impact of sex and recombination on the balance between drift and selection (Lynch et al., 1995; Schultz and Lynch, 1997; Lynch et al., 2016; Vos et al., 2019). Our work complements these aforementioned studies by showing that, however low the rate of gene loss may be, there may always be a class of slightly beneficial traits for which HGT is adaptive and evolvable. Although genes with such small fitness effects are very hard to detect experimentally (Bataillon, 2000; Wisner and Lenski, 2015), our model is a proof of principle that HGT may play a key role in preventing the loss of these genes, which may explain the differential rates of HGT as observed in the data (Nogueira et al., 2009; Rankin et al., 2011; Madsen et al., 2012; Novick and Doolittle, 2020). With the upswing and improvement of experimental techniques like Hi-C metagenomics (Beitel et al., 2014; Burton et al., 2014) and DNA barcoding (Blundell and Levy, 2014; Ba et al., 2019), we will soon have more insights into the eco-evolutionary dynamics of small-effect mutations (Li et al., 2018; Lerner et al., 2019) and accessory genes (Quistad et al., 2019; Yaffe and Relman, 2020), and we may learn when HGT can come to rescue a microbial population, and when it may be nothing more than a catastrophe.

5.4 Methods

5.4.1 General overview

In this work, we study the dynamics of bacteria undergoing HGT of slightly beneficial genes and Selfish Genetic Elements (SGEs). We do this by modelling the same processes with gradually increasing complexity, starting from simple Ordinary Differential Equations (ODEs), and then evaluating the same dynamics in an Individual-based Model (IBM). A graphical representation of these models is found in the main text (**Figure 5.1**). The models consider the competition between cells of two types: carrier cells (C) that carry a gene, and non-carrier cells (N). When carrier cells contain a beneficial gene (*i.e.* it is a beneficial trait), they grow faster than the non-carrier cells (N). However, carriers may lose this beneficial gene with a fixed rate l . Both cell types take up DNA with rate h , which comes with a cost c . This cost is equal for both cell types, reflecting for example the costs of expressing and operating the DNA uptake machinery, other metabolic burdens of the physiological state of natural competence, or the risks associated with taking up foreign DNA. Proportional to the density of available carrier cells, non-carriers can be transformed back into a carrier cell by means of “additive” HGT. Both models use a chemostat assumption, where cells wash out at a rate proportional to the rate of growth, ensuring a constant population size in steady state.

5.4.2 ODE model(s)

By modelling the dynamics described above by means of ODEs, we assume a well-mixed population of cells that compete according to all-against-all dynamics (*i.e.* mass-action). Our equations describing the density of carrier (C) and non-carrier (N) cells are given in **Equation 5.2**, where b is the benefit of the carried gene (or burden if $b < 0$), l is the rate of gene loss, h is the rate at which cells engage in HGT, c is the continuous cost for engaging in HGT, and HGT transforms a non-carrier into a carrier when they interact (hCN). This cost for HGT (c) is equal for both cell types, meaning that whatever the costs may entail, we assume they are continuously paid. Finally, the total amount of growth (ϕ) is subtracted from both populations, meaning that the population density in steady state is always 1.

$$\begin{aligned}
 \frac{dC}{dt} &= \underbrace{(1 - ch + b)C}_{\text{reproduction of } C} - \underbrace{lC}_{\text{gene loss}} + \underbrace{hCN}_{\text{HGT}} - \underbrace{\phi C}_{\text{chemostat}} \\
 \frac{dN}{dt} &= \underbrace{(1 - ch)N}_{\text{reproduction of } N} + \underbrace{lC}_{\text{gene loss}} - \underbrace{hCN}_{\text{HGT}} - \underbrace{\phi N}_{\text{chemostat}} \\
 \phi &= \underbrace{(1 - ch + b)C}_{\text{total growth of } C} + \underbrace{(1 - ch)N}_{\text{total growth of } N}
 \end{aligned} \tag{5.2}$$

$C + N = 1$ (constant population size, ensured by chemostat assumption.)

From the above model, we derived how the population growth rate (ϕ) depends on both b and h (see **Equation 5.1** in the main text), which shows the conditions under which HGT improves the total growth rate of the population. To analyse

whether or not HGT could evolve, we extended the two-variable ODE model above (of cells with the same h) to a four-variable ODE model (of two species with a different h , see **Figure 5.1B** and **Equation 5.3** below). We use this extension to study whether or not a species with HGT (C^+ and N^+ , $h > 0$) could invade upon a species without HGT (C^- and N^- , $h = 0$), and *vice versa* (see Supplementary material for full analysis). Finally, we also extended the ODE model to study the impact on growth rates for cells that engage in HGT of both a beneficial gene and a Selfish Genetic Element (SGE), which can be found in the Supplementary Material.

$$\begin{aligned}
 \frac{dC^-}{dt} &= (1 + b)C^- - lC^- - \phi C^- \\
 \frac{dN^-}{dt} &= N^- + lC^- - \phi N^- \\
 \frac{dC^+}{dt} &= (1 + b - ch)C^+ - lC^+ + hN^+(C^- + C^+) - \phi C^+ \\
 \frac{dN^+}{dt} &= (1 - ch)N^+ + lC^+ - hN^+(C^- + C^+) - \phi N^+ \\
 \phi &= (1 + b)C^- + N^- + (1 + b - ch)C^+ + (1 - ch)N^+
 \end{aligned} \tag{5.3}$$

5.4.3 Individual-based model

The individual-based model (IBM) describes the same dynamics as the ODE models, but differs in some important aspects. Firstly, individuals are discrete entities that live on a 2D grid, and reproduce locally. This allows us to study the model with and without spatial pattern formation by modifying the rate at which cells mix. When mixing is disabled or very limited, a spatially structured population like that of a biofilm will form, while an increased amount of cellular mixing will approximate a well-mixed culture. Under well-mixed conditions, individuals will interact with random individuals in the population (approximating the all-against-all dynamics of the ODEs), while individuals will interact mostly with their conspecifics in case of the biofilm. We explicitly define a competition range (focal cell plus its 8 neighbouring grid points) and a HGT range (all cells within distance t) which determine smaller samples of the total population with which individuals can interact. Each individual (potentially) has its own h -parameter, allowing us to study the evolution of HGT in an eco-evolutionary context (see implementation of mutations below). As we primarily focus on the question if cells benefit from taking up genes from their environment or other cells, we assume that the h -parameter of the recipient cell determines the probability of HGT. While we tested if HGT could evolve *de novo* by studying the invasion criterion for HGT^+ (cells that can take up DNA) in the ODE model, the we study the *de novo* evolution of HGT in the IBM by continuously introducing carriers at a low rate. With a rate f , genes with benefit b^* are (re)discovered, allowing us to study how and if newly discovered genes / selfish elements spread through the population. Finally, note that processes such as gene loss, HGT, and competition are no longer deterministic like in the ODEs, but implemented as events that can stochastically happen at each simulated

time step. To ensure the chance-events in the IBM (reproduction, HGT, gene loss) accurately represent the rates as used in the ODE, all probabilities were multiplied by a small constant $\Delta T=0.1$.

Updating grid points: All grid points (i, j) in the IBM contain a single cell which can be a carrier or non-carrier ($b_{i,j} = b$ for carriers, $b_{i,j} = 0$ for non-carrier), which can carry a SGE ($\beta_{i,j} = \beta$ for SGE infected cells, $\beta_{i,j} = 0$ for uninfected cells), and have an individual HGT-parameter $h_{i,j}$. At each time step, local reproduction happens in each grid point i, j by drawing a random individual from the Moore (9 cells) neighbourhood and letting it reproduce with a probability proportional to its growth-rate $\varphi_{i,j}$:

$$\varphi_{i,j} = 1 + b_{i,j} - \beta_{i,j} - ch_{i,j} \quad (5.4)$$

When reproduction happens, the winner cell replicates and replaces the cell in grid point i, j . This newborn cell is an exact copy of the mother cell. Next, all cells are also updated to include the processes of stochastic gene loss with rate l , HGT with rate $h_{i,j}$, and gene/SGE discoveries based on the influx-rate f . Finally, with a small probability u , the HGT rate of any individual can mutate, where a cell uniformly samples a new parameter between $h_{i,j} - m$ and $h_{i,j} + m$.

IBM growth rates: With respect to growth rates, the simulated IBM model does not only track the average growth rate of all cells in the population (ϕ_{pop}), but also tracks the growth rate of the line of descent that gave rise to the current population (ϕ_{lod}). While ϕ_{pop} is comparable to ϕ in the ODE-model, ϕ_{lod} gives us insights into how the long-term ancestors are impacted by HGT. We also measure the average competitive advantage of carrier cells (σ_c) by calculating, for each carrier cell, how much higher φ is than that of the 8 neighbouring individuals. When all competitors of a also carrier cells, this value is 0. When all competitors are non-carriers, this value approaches b . σ_c is the average of these values.

IBM barcoding: We tag all influxed genes and SGEs with a unique identifier, allowing us to visualise how genes / SGEs spread through the population (analogous to DNA barcoding (Blundell and Levy, 2014; Levy et al., 2015; Ba et al., 2019; Lerner et al., 2019)). These barcodes also allow us to investigate whether or not these genes are continuously rediscovered, or form long lineages of genes that persist within the population.

Parameters used:

Throughout most of this study, the gene loss l was set to 0.02 and the cost for HGT was set to $c = 0.2$. In general, our results do not depend on the absolute value of these two parameters. Instead, our results depends on the relationship between these two parameters, meaning that for lower (arguably more realistic) rates of gene loss, the relevant parameter range will still be extensive if the costs for HGT are also lower. If the costs are very high, the range where HGT is adaptive (*i.e.* enrichable and rescuable genes) becomes much more narrow, but is nevertheless retained (*i.e.* the gene classes discussed in **Figure 5.2** simply shift to lower values of b). As these costs may entail a wide variety of different burdens on the cell (operating and expressing DNA pumps, diverting resources to the physiological state of natural competence, the burden of expressing redundant gene copies, chromosome disruptions, cytotoxicity,

and the risks of SGEs), we do not know quantitative information to judge the relevant value for this cost parameter. Instead, we argue that our model instead gives conceptual clarifications and demonstrations of novel possibilities. Parameters such as the benefit (b), the HGT-rate (h), the amount of mixing (d), and the HGT distance (t) have been extensively swept, as discussed in the main text / Supplementary Material. In these cases, the used parameters are given in the captions of the relevant figures. When comparing the IBM with the ODE models (*e.g.* occurrence of gene classes), evolution of h was disabled ($u = 0.0$). For the *de novo* evolution of HGT, the initial population consisted only of *non-carrier* cells, but genes fluxed in at a low rate ($f = 5 \cdot 10^{-6}$), while the initial level of HGT ($h = 0.0$) was allowed to evolve with $u = 5 \cdot 10^{-5}$ with a uniform step size of $m = 0.05$. Finally, when testing whether HGT could be maintained, no influx of genes was present ($f = 0.0$), but the initial population consisted of carrier-cells that already engage in HGT ($h = 0.05$, see supplementary material for evolved rates of HGT). All experiments in the IBM with SGEs were done with slightly lower costs ($c = 0.1$), to compensate for the extra costs imposed by these genetic parasites. All the important parameters of our models are summarised in **Table 5.1**.

Table 5.1: Description of parameters used in the models

Parameter (general)	Description
Gene loss (l)	Rate at which carrier cells lose the beneficial gene
HGT rate (h)	Rate at which non-carriers are transformed into carriers (when interacting with carrier cells)
Benefit of gene (b)	Growth rate benefit for carrier cells (or penalty for negative b)
Costs of HGT (c)	Growth rate penalty for the rate of HGT
Parameter (IBM only)	Description
Grid size (n)	The simulation is done on a square grid of $n \times n$ cells
Mixing rate (d)	Every time step, the grid is mixed d times using the Margolus Diffusion algorithm (Toffoli and Margolus, 1987). Alternatively, the population was <i>well-mixed</i> by assigning new positions at random every time step.
Competition range (s)	Sub-population of $s \times s$ cells surrounding focal grid point that compete for reproduction
HGT distance (t)	Sub-population of $d \times d$ cells surrounding focal grid point from which a random potential donor is sampled for HGT
Influx genes (f)	A small probability for any cell to discover a gene <i>de novo</i>
Benefit of influxed gene (b^*)	Growth rate benefit for carrier cells (or penalty for negative b)
Influx SGEs (i)	A small probability for any cell to be infected by an SGE <i>de novo</i>
Fitness penalty of SGE (β)	The fitness penalty imposed by the SGE
Mutation rate (μ)	Chance of mutating the evolvable HGT-rate
Mutation step (m)	Uniform step size of mutations

5.4.4 Software used

The analytical model was numerically analysed using `grind.R` by R.J. de Boer (<http://tbb.bio.uu.nl/rdb>), a R script that uses the `deSolve` R-package (Soetaert et al., 2010). The simulated model was implemented in `Cash` (Cellular Automaton simulated hardware) version 2.1, an free and easy-to-use library to make simple spatially explicit simulations (originally created by R.J. de Boer & A.D. Staritsk, further developed by Nobuto Takeuchi and Bram van Dijk). Visualisation of both models was done in R using `ggplot` (Wickham, 2016) and `plotly` (Inc., 2015). Simulations were run in Linux Ubuntu 16.04 LTS using GNU `parallel`(Job).

Both the R-scripts for ODE analysis and the IBM code implemented in C, are available online https://github.com/bramvandijk88/HGT_Genes_And_SGEs.

Acknowledgements

This work was supported by the European Commission 7th Framework Programme (FPFP7-ICT-2013.9.6 FET Proactive: Evolving Living Technologies) EvoEvo project (ICT-610427) and the Human Frontier Science Program (RGY0072/2015).

5.5 Supplementary Material

This supplementary material includes the mathematical derivations of the results discussed in the main text and some extra insights and figures. The source code material to reproduce the numerical simulations we have done (both in the main text and in this supplementary material), is available online (https://github.com/bramvandijk88/HGT_Genes_And_SGEs).

Part I: Mathematical analyses

Equilibria and population growth rate of a single population

As described in the main text, we consider a population of cells that either carry or do not carry a gene. The dynamics of the density of carriers (C) and non-carriers (N) are described by:

$$\frac{dC}{dt} = \underbrace{(1 - ch + b)C}_{\text{reproduction of } C} - \underbrace{lC}_{\text{gene loss}} + \underbrace{hCN}_{\text{HGT}} - \underbrace{\phi C}_{\text{chemostat}} \quad (5.5)$$

$$\frac{dN}{dt} = \underbrace{(1 - ch)N}_{\text{reproduction of } N} + \underbrace{lC}_{\text{gene loss}} - \underbrace{hCN}_{\text{HGT}} - \underbrace{\phi N}_{\text{chemostat}} \quad (5.6)$$

$$\phi = \underbrace{(1 - ch + b)C}_{\text{total growth of } C} + \underbrace{(1 - ch)N}_{\text{total growth of } N} \quad (5.7)$$

$$C + N = 1 \text{ (constant population size, ensured by chemostat assumption.)} \quad (5.8)$$

Equilibria and their stability

The equilibria of Eq 5.5–5.8 are found by solving $\frac{dC}{dt} = \frac{dN}{dt} = 0$.

$$\text{Let } \frac{dC}{dt} = (1 + b - ch)C - lC + hNC - \phi C = 0.$$

Then either $C = 0$,

or $\phi = 1 + b - ch - l + hN$ and

$$\begin{aligned} 1 + b - ch - l + hN &= (1 + b - ch)C + (1 - ch)N \\ \iff 1 + b - ch - l + h(1 - C) &= (1 + b - ch)C + (1 - ch)(1 - C) \\ \iff 1 + b - ch - l + h &= (1 + b - ch + h - 1 + ch)C + 1 - ch \\ \iff b - l + h &= (b + h)C \\ \iff C &= \frac{b - l + h}{b + h} = 1 - \frac{l}{b + h} \end{aligned}$$

Using $C + N = 1$, we find that the system has two equilibria:

$$\text{equilibrium (i): } C^* = 0, \quad N^* = 1, \quad (5.9)$$

$$\text{equilibrium (ii): } C^* = 1 - \frac{l}{b + h}, \quad N^* = \frac{l}{b + h} \quad (5.10)$$

Next, we study under what conditions the gene can persist in the population described by Eq 5.5–5.8. Note that this is equivalent to asking when equilibrium (i)

is unstable, *i.e.*, when the carrying cells (C) can invade on a resident population of non-carrying cells (N) at carrying capacity. When the system is in equilibrium (i), $C^* = 0$, $N^* = 1$, and $\phi^* = (1 - ch)$. The dynamics of the carrying cells can then be approximated by

$$\frac{dC}{dt} \approx (1 + b - ch - l + hN^* - \phi^*)C = (1 + b - ch - l + h - (1 - ch))C = (b + h - l)C,$$

and the carrying cells can invade iff $\frac{dC}{dt} > 0$, *i.e.*, iff

$$b + h - l > 0. \quad (5.11)$$

From Eq 5.11 we can conclude that genes which yield a sufficient growth rate benefit to overcome the loss rate ($b > l$) do not need HGT in order to persist in a population. Slightly beneficial genes, however, only persist when $h > (l - b)$. HGT, serving as a plausible “back-mutation”, prevents the eventual loss of such a gene from the population.

Population growth rate ϕ in steady state as a function of HGT rate h

Even though we have shown above that some genes can only persist in a population at sufficiently high rates of HGT, the survival of these genes does not necessarily imply that HGT also improves the actual growth rate of the population under these conditions, as the model also assumes a cost for higher rates of HGT. To gain better insight into when HGT improves the steady state growth rate, we will next consider how the population growth rate ϕ depends on h .

The population growth rate in steady state, ϕ^* , is given by:

$$\phi^*(h) = (1 + b - ch)C^* + (1 - ch)N^* \quad (5.12)$$

$$= \begin{cases} 1 - ch & \text{if } h \leq (l - b) \text{ (gene cannot persist);} \\ 1 - ch + b - \frac{bl}{b+h} & \text{if } h > (l - b) \text{ (gene persists).} \end{cases} \quad (5.13)$$

To determine the effect of the rate of HGT, h , on the steady state population growth rate ϕ^* , we differentiate Eq 5.13 with respect to h :

$$\frac{\partial \phi^*}{\partial h} = \begin{cases} -c & \text{if } h \leq (l - b); \\ -c + \frac{bl}{(b+h)^2} & \text{if } h > (l - b). \end{cases} \quad (5.14)$$

As long as $h < (l - b)$, $\frac{\partial \phi^*}{\partial h} = -c < 0$ and an increase in HGT rate h will decrease the population growth rate at steady state $\phi^*(h)$. For, $h > (l - b)$, the population growth rate ϕ^* might however have a local optimum, which we can find by setting $\frac{\partial \phi^*}{\partial h}$ to 0:

$$\begin{aligned} \frac{bl}{(b+h)^2} - c &= 0 \\ \iff (b+h)^2 &= \frac{bl}{c} \\ \text{from which we can solve} \\ h_{\text{opt}} &= \sqrt{\frac{bl}{c}} - b \end{aligned} \quad (5.15)$$

Note that this optimum is only obtained in the function $\phi^*(h)$ if $h_{\text{opt}} > (l - b)$:

$$\sqrt{\frac{bl}{c}} - b > l - b \quad (5.16)$$

$$\iff \frac{bl}{c} > l^2 \quad (5.17)$$

$$\iff b > lc. \quad (5.18)$$

(This is the same condition found when solving $\frac{\partial \phi^*}{\partial h} > 0$ at $h = (l - b)$)

Furthermore, since h is the rate of HGT, we are only interested in positive values of h . $h_{\text{opt}} > 0$ iff

$$\sqrt{\frac{bl}{c}} > b \quad (5.19)$$

$$b < \frac{l}{c}. \quad (5.20)$$

Under the conditions of Eq 5.18 and 5.20, the second derivative of ϕ^* to h is

$$\frac{\partial^2 \phi^*}{\partial h^2} = \frac{-2bl}{(b + h)^3},$$

which is negative if the parameters b and l are ≥ 0 . Hence, when $\phi^*(h)$ has an optimum for a positive HGT rate h_{opt} , this local optimum is a maximum. The growth rate in this local maximum is larger than the growth rate at $h = 0$, $\phi^*(0) = 1$, iff

$$\phi^*(h_{\text{opt}}) = 1 - ch_{\text{opt}} + b - \frac{bl}{b + h_{\text{opt}}} > 1 \quad (5.21)$$

$$\iff 1 + cb - \sqrt{bcl} + b - \frac{bl}{\sqrt{\frac{bl}{c}}} > 1 \quad (5.22)$$

$$\iff b(1 + c) - 2\sqrt{bcl} > 0 \quad (5.23)$$

$$\iff b > \frac{4lc}{(1 + c)^2}. \quad (5.24)$$

Summarising, the population growth rate at equilibrium, ϕ^* , decreases linearly with the risks ch when $h < (l - b)$ due to the costs of HGT (see Eq 5.13). Under these conditions, the growth rate does not depend on b because the gene cannot persist in the population. When $h > (l - b)$, the gene does persist within the population, resulting in an extra term $b - \frac{bl}{b+h}$ in the growth rate $\phi^*(h)$. This extra term approaches a maximal benefit of b for high values of h . The burden of HGT ch will however eventually outweigh this benefit for increasing rates of HGT. A (local) optimal rate of HGT can found at $h_{\text{opt}} = \sqrt{bl/c} - b$, as long as $b > lc$. This optimal HGT rate is greater than 1, meaning that HGT improves the population growth rate at steady state, if the genes have a minimal benefit (see Eq 5.24). However, when the benefit is too large ($b > l/c$), the optimal HGT rate becomes $h_{\text{opt}} < 0$. As negative values for HGT are biologically unsound, HGT never improves the population growth rate in steady state for genes with such a high fitness benefit.

Following these derivations, genes can be divided in different classes based on the value of the fitness benefit b and the consequent effect of HGT on the population growth rate at steady state (see main text and **Figure 5.2**):

Selfish Genetic Elements (SGEs) ($b < 0$) Carrying the gene confers a fitness cost. Increasing HGT-rates only lower the equilibrium population growth rate ϕ^* .

Unrescuable genes ($b < l$ and $b < \frac{4lc}{(1+c)^2}$) Genes confer a small fitness benefit, but this benefit is too small to overcome gene loss. Furthermore, no positive HGT rate h improves the population growth rate $\phi^*(h)$ over the population growth rate in the absence of HGT ($\phi^*(0) = 1$).

Rescuable genes ($\frac{4lc}{(1+c)^2} < b < l$) Genes confer a small fitness benefit and cannot persist in a population in the absence of HGT, but can be rescued by a sufficiently high HGT rate ($h > (l - b)$). For some HGT rate $h_{\text{opt}} > 0$ the equilibrium growth rate $\phi^*(h) > 1$, indicating that HGT can improve the growth rate of the population.

Enrichable genes ($l < b < l/c$) Genes confer a sufficient fitness benefit to persist in a population in the absence of HGT. HGT can however improve the equilibrium population growth rate $\phi^*(h_{\text{opt}})$.

Indispensable genes ($b > l/c$) Genes confer a large fitness benefit and can persist in a population in the absence of HGT. HGT furthermore does not improve the equilibrium population growth rate.

Evolutionary stability of HGT^+ and HGT^- populations

To study whether HGT is an evolvable trait, we will consider 1) if HGT can evolve *de novo*, and 2) if HGT can be evolutionarily maintained. For this, we extended the two-variable model of one species to a four-variable model of two species: a HGT^+ -species that engages in HGT, and a HGT^- -species that does not (**Suppl. Figure S5.1B, Equation 5.25-5.28**). We analysed under what conditions the HGT^+ -species can invade an equilibrium of the HGT^- -species, and *vice versa*. We found that HGT can only evolve for an enrichable gene, but is evolutionarily maintained for both enrichable and rescuable genes. The following paragraphs will elaborate on how these results are derived:

Consider a HGT^+ -species (C^+ , N^+) and a HGT^- -species (C^- , N^-) that differ in their HGT rate h , but are identical otherwise. The dynamics of the density of cells carrying and not carrying the gene of the two species can be described by the following equations:

$$\frac{dC^-}{dt} = (1 + b)C^- - lC^- - \phi C^- \quad (5.25)$$

$$\frac{dN^-}{dt} = N^- + lC^- - \phi N^- \quad (5.26)$$

$$\frac{dC^+}{dt} = (1 + b - ch)C^+ - lC^+ + hN^+(C^- + C^+) - \phi C^+ \quad (5.27)$$

$$\frac{dN^+}{dt} = (1 - ch)N^+ + lC^+ - hN^+(C^- + C^+) - \phi N^+ \quad (5.28)$$

$$\phi = (1 + b)C^- + N^- + (1 + b - ch)C^+ + (1 - ch)N^+ \quad (5.29)$$

$$C^- + N^- + C^+ + N^+ = 1. \quad (5.30)$$

Note that we include horizontal gene transfer from HGT^- -cells carrying the gene to HGT^+ -cells that do not yet carry the gene. In other words, we consider a situation in which the propensity of HGT is determined by the recipient cell, and not by the donor. This is inspired by for instance the process of transformation, in which the recipient cell “decides” whether or not it takes up extracellular DNA.

If HGT is evolvable *de novo*, the HGT^+ species should be able to invade a HGT^- population in steady state. In other words, the equilibrium state $(C^-, N^-, C^+, N^+) = (\hat{C}^-, \hat{N}^-, 0, 0)$ should be unstable.

Around the equilibrium $(\hat{C}^-, \hat{N}^-, 0, 0)$, the dynamics of the HGT^+ -species are linearly approximated by

$$\begin{pmatrix} \frac{dC^+}{dt} \\ \frac{dN^+}{dt} \end{pmatrix} \approx \mathbf{J} \begin{pmatrix} C^+ \\ N^+ \end{pmatrix},$$

where

$$\mathbf{J} = \begin{pmatrix} 1 + b - ch - l - \hat{\phi} & h\hat{C}^- \\ l & 1 - ch - h\hat{C}^- - \hat{\phi} \end{pmatrix}.$$

The HGT^+ -species can invade iff the dominant eigenvalue of \mathbf{J} is positive.

Note that the equilibrium densities of \widehat{C}^- and \widehat{N}^- depend on b and l . As derived in the previous section,

$$\text{if } b \leq l, \quad \widehat{C}^- = 0 \quad \text{and} \quad \widehat{N}^- = 1, \quad \text{while} \quad (5.31)$$

$$\text{if } b > l, \quad \widehat{C}^- = 1 - \frac{l}{b} \quad \text{and} \quad \widehat{N}^- = \frac{l}{b}. \quad (5.32)$$

We will consider both possibilities separately.

In the case of unrescuable and rescuable genes ($0 < b \leq l$), the equilibrium densities of \widehat{C}^- and \widehat{N}^- are given by Eq 5.31. Then, $\widehat{\phi} = 1$ and the Jacobian matrix

$$\mathbf{J} = \begin{pmatrix} b - ch - l & 0 \\ l & -ch \end{pmatrix}.$$

The eigenvalues of \mathbf{J} are $\lambda_1 = b - ch - l$ and $\lambda_2 = -ch$. The second eigenvalue $\lambda_2 < 0$ as long as HGT comes at some cost $c > 0$ (the HGT-rate h of a HGT⁺-species is always positive). At the same time, λ_1 is also negative because we consider genes with a small benefit, $0 < b \leq l$. Hence, we conclude that for unrescuable and more importantly for rescuable genes, an HGT⁺-species cannot invade on a HGT⁻-population at equilibrium, and HGT can hence never evolve *de novo*.

In the case of enrichable and indispensable genes ($b > l$), the equilibrium densities of \widehat{C}^- and \widehat{N}^- are given by Eq 5.32. Now, $\widehat{\phi} = (1 + b)(1 - \frac{l}{b}) + \frac{l}{b} = 1 + b - l$, and the Jacobian matrix

$$\mathbf{J} = \begin{pmatrix} -ch & h(1 - \frac{l}{b}) \\ l & l - b - ch - h(1 - \frac{l}{b}) \end{pmatrix}.$$

The eigenvalues of \mathbf{J} should now be solved from

$$(-ch - \lambda)(l - b - ch - h(1 - \frac{l}{b}) - \lambda) - lh(1 - \frac{l}{b}) = 0 \quad (5.33)$$

$$\iff \lambda^2 - \lambda(l - b - 2ch - h(1 - \frac{l}{b})) + (bch - lch + c^2h^2 + ch^2(1 - \frac{l}{b}) - lh(1 - \frac{l}{b})) = 0. \quad (5.34)$$

Let

$$\beta = l - b - 2ch - h(1 - \frac{l}{b}), \quad \text{and} \quad (5.35)$$

$$\gamma = bch - lch + c^2h^2 + ch^2(1 - \frac{l}{b}) - lh(1 - \frac{l}{b}). \quad (5.36)$$

Then, the eigenvalues of \mathbf{J} are equal to $\lambda_{1,2} = \frac{1}{2}(\beta \pm \sqrt{\beta^2 - 4\gamma})$. Remember that we are interested in the sign of the dominant eigenvalue. If the eigenvalues are complex ($\beta^2 < 4\gamma$), the real part of the eigenvalues $\text{Re}(\lambda_{1,2}) > 0$ iff $\beta > 0$. If the eigenvalues are real, the dominant eigenvalue is $\lambda_1 = \frac{1}{2}(\beta + \sqrt{\beta^2 - 4\gamma})$, and $\lambda_1 > 0$ iff $\beta > 0$ or $\sqrt{\beta^2 - 4\gamma} > \beta \iff \gamma < 0$.

First, consider the possibility $\beta > 0$. Then we should have

$$l - b - 2ch - h\left(1 - \frac{l}{b}\right) > 0 \quad (5.37)$$

$$\iff l - b > h\left(2c + \left(1 - \frac{l}{b}\right)\right). \quad (5.38)$$

This is however a contradiction, since we here deal with genes for which $b > l$ and hence $l - b < 0$, but $\widehat{C}^- = 1 - \frac{l}{b} > 0$, $c > 0$ and $h > 0$. Hence, β is always negative and the dominant eigenvalue is positive only if $\gamma < 0$. From $\gamma < 0$, we find

$$bch - lch + c^2h^2 + ch^2\left(1 - \frac{l}{b}\right) - lh\left(1 - \frac{l}{b}\right) < 0 \quad (5.39)$$

$$\iff c(b - l + ch) + (ch - l)\left(1 - \frac{l}{b}\right) < 0 \quad (5.40)$$

Trying to solve Eq 5.40 for any value of h would yield a complicated condition on the value of b . However, we can further simplify Eq 5.40 by asking if a HGT⁺-species with a very small (but positive) HGT-rate could invade. For $h = \epsilon \approx 0$, Eq 5.40 reduces to

$$c(b - l) - l\left(1 - \frac{l}{b}\right) < 0, \quad (5.41)$$

from which we can solve

$$c(b - l) - l\left(1 - \frac{l}{b}\right) < 0 \quad (5.42)$$

$$\iff cb^2 - l(c + 1)b + l^2 < 0 \quad (5.43)$$

$$\iff (cb - l)(b - l) < 0. \quad (5.44)$$

Since we consider enrichable and indispensable genes, with $b > l$, condition 5.44 can only be true if $cb < l \iff b < l/c$, which is exactly the condition that separates enrichable from indispensable genes. Hence, we conclude that for enrichable genes ($l < b < l/c$), a HGT⁺-species with a small but positive HGT-rate can always invade on a HGT⁻-population at equilibrium, and that HGT can hence evolve *de novo*.

So far, we have determined under what conditions a HGT⁻-population is evolutionarily stable. We can however ask the same for a HGT⁺-population. In other words, even though it may not be reached by gradual evolution, can HGT be *maintained*? To answer this question, we next consider the evolutionary stability of the HGT⁺-equilibrium: $(C^-, N^-, C^+, N^+) = (0, 0, \tilde{C}^+, \tilde{N}^+)$.

Again, the densities of C^+ - and N^+ -cells at equilibrium depend on the values of b , l and h (see Eq 5.9–5.10 in the previous section):

$$\text{if } b \leq l - h, \quad \tilde{C}^+ = 0 \quad \text{and} \quad \tilde{N}^+ = 1, \quad \text{while} \quad (5.45)$$

$$\text{if } b > l - h, \quad \tilde{C}^+ = 1 - \frac{l}{b + h} \quad \text{and} \quad \tilde{N}^+ = \frac{l}{b + h}. \quad (5.46)$$

If $b \leq l - h$, the gene does not persist in the population and HGT hence does not confer any benefit, while still imposing a cost on the N^+ -cells. Under these conditions, the N^- -cells, that do not carry the cost of HGT, will always be able to invade.

For the more interesting case in which the gene does persist in a HGT⁺-population (Eq 5.46), we now linearise the dynamics of the HGT⁻-species around the equilibrium:

$$\begin{pmatrix} \frac{dC^-}{dt} \\ \frac{dN^-}{dt} \end{pmatrix} = \mathbf{J} \begin{pmatrix} C^- \\ N^- \end{pmatrix}$$

$$\text{with } \mathbf{J} = \begin{pmatrix} (1+b) - l - \tilde{\phi} & 0 \\ l & 1 - \tilde{\phi} \end{pmatrix}$$

$$\text{and } \tilde{\phi} = (1+b-ch)\left(1 - \frac{l}{b+h}\right) + (1-ch)\frac{l}{b+h} = (1-ch) + b\left(1 - \frac{l}{b+h}\right).$$

Again, the HGT⁻-species can invade if the dominant eigenvalue of \mathbf{J} is positive, and hence the HGT⁺-species of equilibrium is evolutionarily stable if both eigenvalues are negative. The eigenvalues of \mathbf{J} are $\lambda_1 = 1 + b - l - \tilde{\phi}$ and $\lambda_2 = 1 - \tilde{\phi}$.

For the first eigenvalue, we find

$$\lambda_1 < 0 \quad (5.47)$$

$$\iff 0 > 1 + b - l - \tilde{\phi} \quad (5.48)$$

$$\iff 0 > 1 + b - l - (1 - ch) - b\left(1 - \frac{l}{b+h}\right) \quad (5.49)$$

$$\iff 0 > bl + ch(b+h) - l(b+h) \quad (5.50)$$

$$\iff lh > ch(b+h) \quad (5.51)$$

$$\iff l > c(b+h) \quad (5.52)$$

$$\iff c < \frac{l}{b+h} \quad (5.53)$$

Hence, this first eigenvalue is negative as long as the costs of HGT are not too large. For the second eigenvalue, we find:

$$\lambda_2 < 0 \quad (5.54)$$

$$\iff 0 > 1 - \tilde{\phi} \quad (5.55)$$

$$\iff 0 > 1 - (1 - ch) - b\left(1 - \frac{l}{b+h}\right) \quad (5.56)$$

$$\iff 0 > ch - b\left(1 - \frac{l}{b+h}\right) \quad (5.57)$$

$$\iff ch < b\left(1 - \frac{l}{b+h}\right) \quad (5.58)$$

$$\iff c < \frac{b\left(1 - \frac{l}{b+h}\right)}{h}. \quad (5.59)$$

Remember that we considered a HGT⁺-population in which the gene can persist, *i.e.*, $b+h > l$. Hence $\frac{l}{b+h} < 1$ and the right hand side in Eq 5.59 is positive. Hence, we can again conclude that there are some non-zero costs for which λ_2 is negative. Combining the results in Eq 5.53 and 5.59, we see that for some costs, HGT can be maintained. For rescuable genes with costs that satisfy conditions 5.53 and 5.59, there is an Allee effect with respect to HGT: HGT can be evolutionarily maintained,

but it cannot evolve *de novo*. This result can be intuitively understood. Small (invading) HGT⁺-populations pay the continuous costs for HGT, but hardly ever interact with their conspecifics, and hence the positive fitness effects of maintaining the slightly beneficial gene are too small to overcome the costs for HGT. Higher fitness can only be achieved when the population size is large enough, such that the benefits conferred by HGT outweigh its costs. The presence of an Allee effect was confirmed by numerically integrating Eq 5.25–5.28 for different initial conditions. We then indeed see that the system converges to different equilibria depending on the initial frequency of HGT⁺-cells (see **Figure 5.3**).

Part II: Supplementary results and figures

Similar gene classes exist with different assumptions for the cost of HGT

In the main text, we assume that the costs for HGT scale proportional with the rate of HGT. Moreover, we assume that the costs are always present, *i.e.* also when no carriers are present to interact with. The assumption is that both carriers and non-carriers can take up DNA, but do so to no avail when no carrier DNA is present. Below, we show what the effect is on population growth rates when these assumptions are varied. Specifically, we consider a scenario with constant costs, and a scenario where the costs are only paid when carriers are present from which to take up DNA.

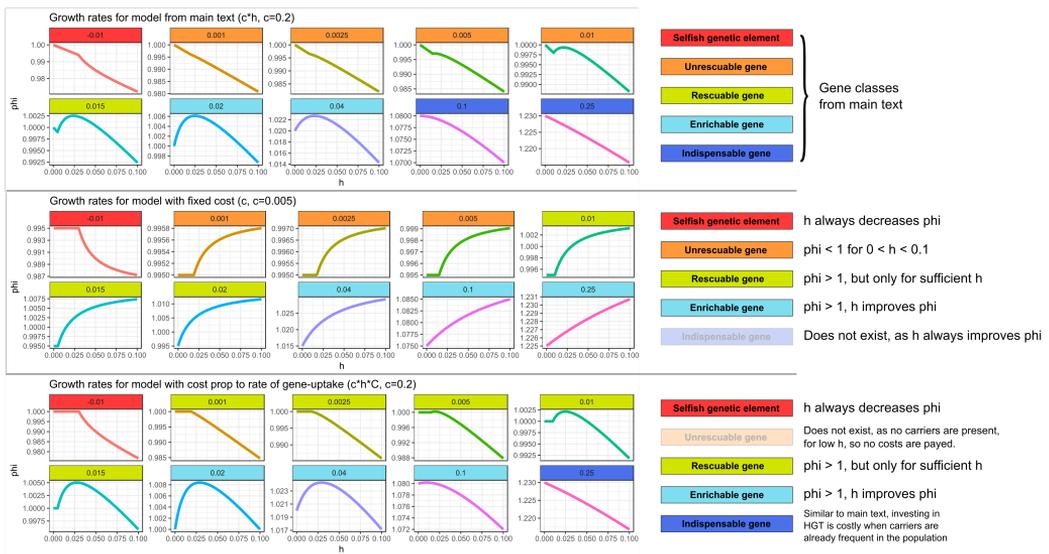


Figure S5.1: Similar gene classes exist with different assumptions for the cost of HGT

In the top panel (with costs as in the main text), we can observe all the gene classes discussed in the main text. In the middle panel (with fixed costs), HGT always improves the growth rates for genes with $b > 0$. However, for low b , the growth rate is lower than 1 even for very high h . Therefore, it is not adaptive to invest in costly HGT (the maximal growth rate is 1.0 with $h = 0$). Thus, for this range of h 's, we observe a similar 'unrescuable' gene class as the one discussed in the main text. When growth rates can be improved within this range of h 's, the genes can be compared to the rescuable genes or enrichable genes, depending on whether or not HGT is required for the gene to persist. Indispensable genes do not exist under this regime, as HGT always increases growth rates when $b > 0$. We argue however that a constant cost with extremely high rates of HGT is fairly unrealistic, as every HGT-event carries inherent risks, for example by chromosome disruptions, cytotoxicity, and the integration of SGEs. In the bottom panel, we investigate a scenario where costs are only present when actually taking up DNA from carrier cells (*i.e.* instead of $c \cdot h$, the costs are $c \cdot h \cdot C$). Under this scenario, most gene classes from the main text exist. Unrescuable genes however do not exist, as there are no cost if no carriers are present. Similarly, the Allee effect as discussed in the main text is also not present, as this Allee effect requires costs in the absence of donor cells.

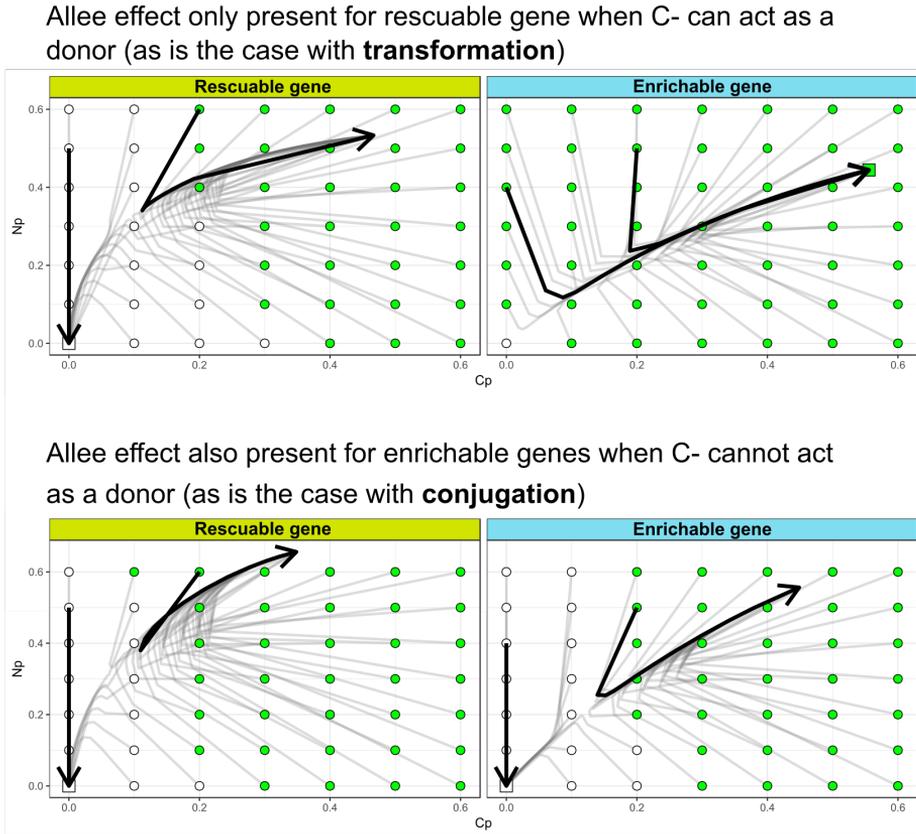


Figure S5.2: The Allee effect for *de novo* evolution of HGT is also present for enrichable genes when C^- (a carrier that cannot engage in HGT) cannot act as a donor.

Comparing our results with conjugation-like scenarios

In the main text, we discussed how HGT may not be able to evolve for rescuable genes due to the Allee effect: HGT is only adaptive if the gene is present, but the gene can only be present with HGT. Here, we modified our model to represent a scenario more similar to bacterial conjugation, by assuming that C^- (*i.e.* a carrier without the ability for HGT) cannot act as a donor. Interestingly, under these assumptions we found that the enrichable gene class also shows the Allee effect (see image below). This is intuitive, as in this scenario the HGT^+ strains lack donors to interact with, regardless of the gene already being maintained by the resident HGT^- strain. Thus, similar to the results discussed in the main text, HGT^+ cannot invade. Indeed, we found that this conjugation-like scenario has different dynamics for the evolution of HGT, where both rescuable and enrichable genes require the ‘nucleation events’ presented in the main text.

Interestingly, when studying the emergence of a ‘gene-sharing’ community (using a slow rate of gene-influx as in the main text), we found that it most readily evolved when HGT happened at intermediate physical distances. After 50.000 time steps, both the conjugation-like (C^- cannot act as donor) and the transformation-like (C^- can act as donor) mechanisms of HGT show that only a subset of the populations managed to evolve HGT for low physical distances (see figure below). The details of this figure of course depend on b , as slightly more beneficial genes may spread easier. It is however interesting to

consider how the ‘gene-sharing’ discussed in the main text, works best when the genes can be picked up from afar, ensuring that the clumping of carrier cells (see **Figure 5.5**) is alleviated. This distance-effect is of course not applicable to conjugation, as physical contact is required between cells in order to conjugate. Finally, note that other than these change in the evolutionary accessibility of HGT, the other results discussed in the main text remain qualitatively similar.

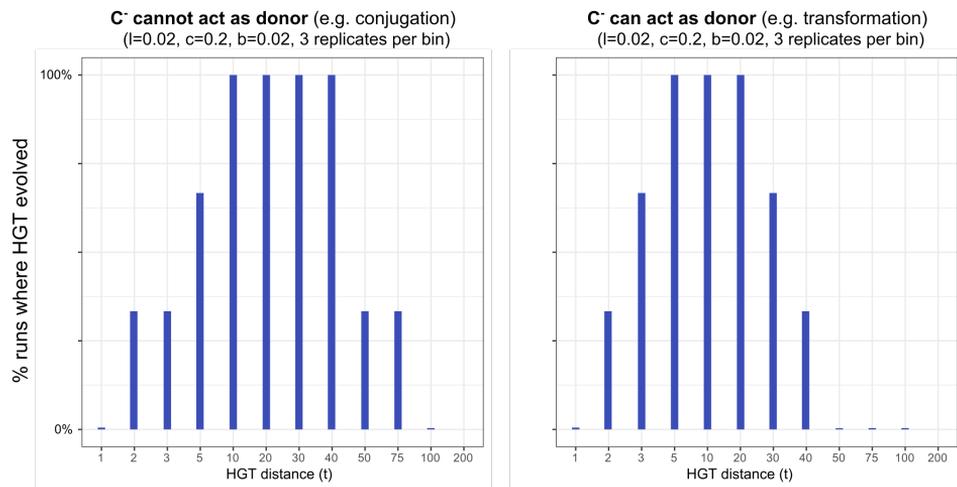
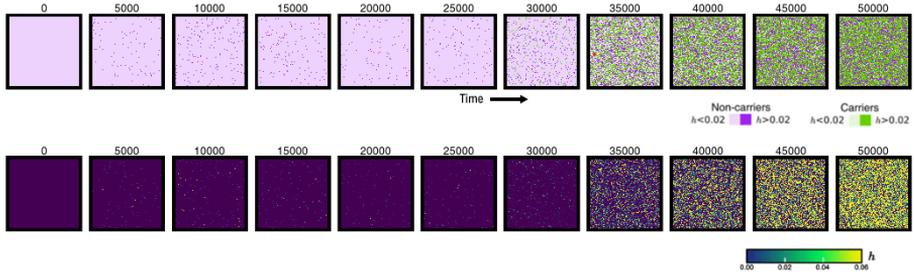


Figure S5.3: HGT evolves most readily when transfer happens between donor / recipient pairs of intermediate distances. On the left-hand side, we consider this for a conjugation-like scenario (C^- , a carrier that cannot engage in HGT, cannot act as a donor), while on the right-hand side we consider this for a transformation-like scenario (C^- , a carrier that cannot engage in HGT, *can* act as a donor).

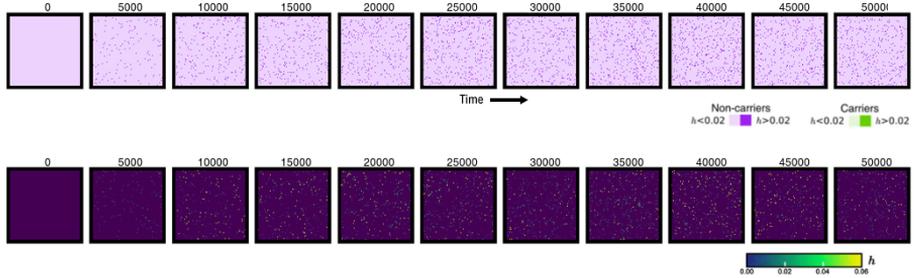
In the well-mixed IBM, HGT only evolves for enrichable genes

In the main text we discussed that HGT cannot evolve for genes that cannot persist without HGT. For these genes, a lack of donor cells does not allow mutants that engage in HGT to get a significant fitness benefit, even when they actually *do* carry the beneficial gene. To get over this so-called Allee effect, a large number of gene-carrying individuals has to simultaneously start engaging in HGT. We have also shown that, in the spatially structure populations, HGT *does* evolve for genes that could not persist without HGT, as it is more likely that the lack of donor cells is, at least locally, overcome. This supplementary figure summarises this result, by showing that, even though HGT does evolve for enrichable genes under well-mixed conditions, it indeed fails to evolve for rescuable genes.

A) HGT evolution for an enrichable gene ($b=0.03$, well-mixed)



B) HGT evolution for a rescuable gene ($b=0.0175$, well-mixed)



C) HGT evolution for a rescuable gene ($b=0.03$, no mixing)

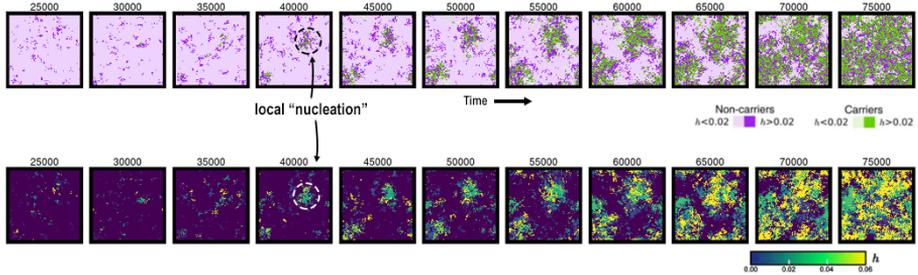
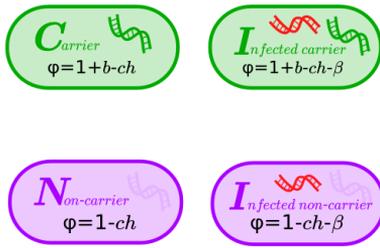


Figure S5.4: HGT evolution in IBM under various conditions

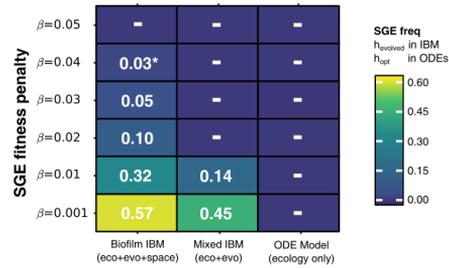
Maintaining weak and strong Selfish Genetic Elements

In the main text we have discussed how SGEs can coexist along-side their hosts and slightly beneficial genes, even when the their fitness-penalty is greater than the benefit of the gene. However, this was only observed in the spatially structured model, as illustrated in the figure below.

A) Cartoon of the 4 cell types



B) Persistence of SGEs



* in small populations as used here, co-evolution between the hosts HGT-rate and SGEs may cause extinction. Larger grids as shown in the main text are more resilient to this.

C) SGEs with intermediate fitness penalty impose the highest cost on the host

(simulated in the spatially structured populations of 200x200 cells)

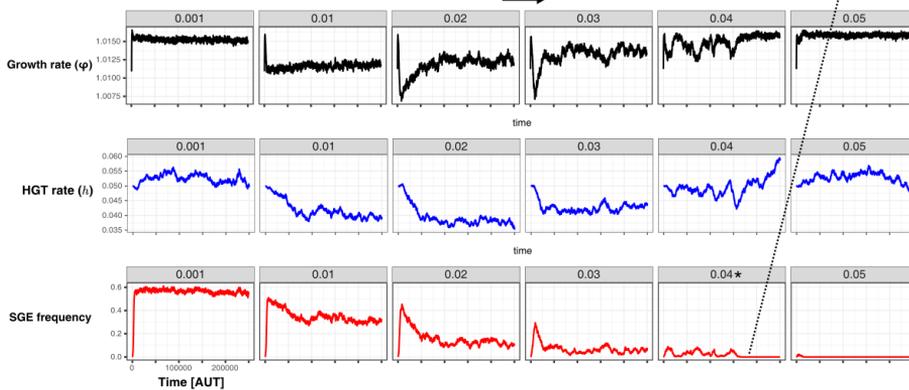


Figure S5.5: Persistence of SGEs in various implementations of our model. A shows a cartoon of the cell types, between which competition was modeled in a various ways. In B we show for these different implementations how many SGEs persist within the populations for SGEs with different penalties. For the IBM, we simulated for 250.000 time steps and calculated the average SGE-frequency in the final 100 generations. For the ODE model, we chose the optimal rate of HGT (h_{opt}), and numerically integrated the equilibrium concentrations of infected cells. Finally, C shows the temporal dynamics of the growth rate (ϕ), HGT-rate (h), and the SGE frequency, in the spatially structured simulations. As this parameter sweep had slightly smaller populations sizes as used in the main text, the strong SGE could eventually go extinct (this is annotated with an asterisk).

Equations for Suppl. Figure S5.5A

$$\begin{aligned}
\frac{dC}{dt} &= \gamma(1 + b - ch)C - lC + lD + h(NC + 0.5ND - CP - CD) - \phi C \\
\frac{dN}{dt} &= \gamma(1 - ch)N + lC + lP - hN(C + P + D) - \phi N \\
\frac{dP}{dt} &= \gamma(1 - \beta - ch)P + lD - lP + h(NP + ND/2 - CP - PD) - \phi P \\
\frac{dD}{dt} &= \gamma(1 + b - \beta - ch)D - lD/2 + h(CP + CD + PC + PD) - \phi D \\
\phi &= \gamma((1 + b - ch)C + (1 - ch)N + (1 - \beta - ch)P + (1 + b - \beta - ch)D)
\end{aligned} \tag{5.60}$$

Strong SGEs fail to spread / persist in the population at low HGT-distances

In the main text we have discussed how we found that strong SGEs (genetic parasites with a greater penalty than the beneficial gene) could nevertheless stably coexist with an evolving population of cells. However, this persistence of SGEs relies on their ability to escape to new susceptible hosts who have not experienced SGEs for some time (and therefore have evolved elevated HGT rates). In this supplementary figure, it is indeed seen how the distance influences the spread / persistence of SGEs. If the distance between donor and acceptor is very local ($d = 1$), SGEs cannot spread even while they are still fluxing in (top row). For an intermediate HGT-distance ($1 < d < 10$), the SGEs persist for a bit as long as they flux in, but die out when influx is stopped (middle row). For larger HGT distances ($d > 10$), we found that SGEs can persist even after the influx was stopped.

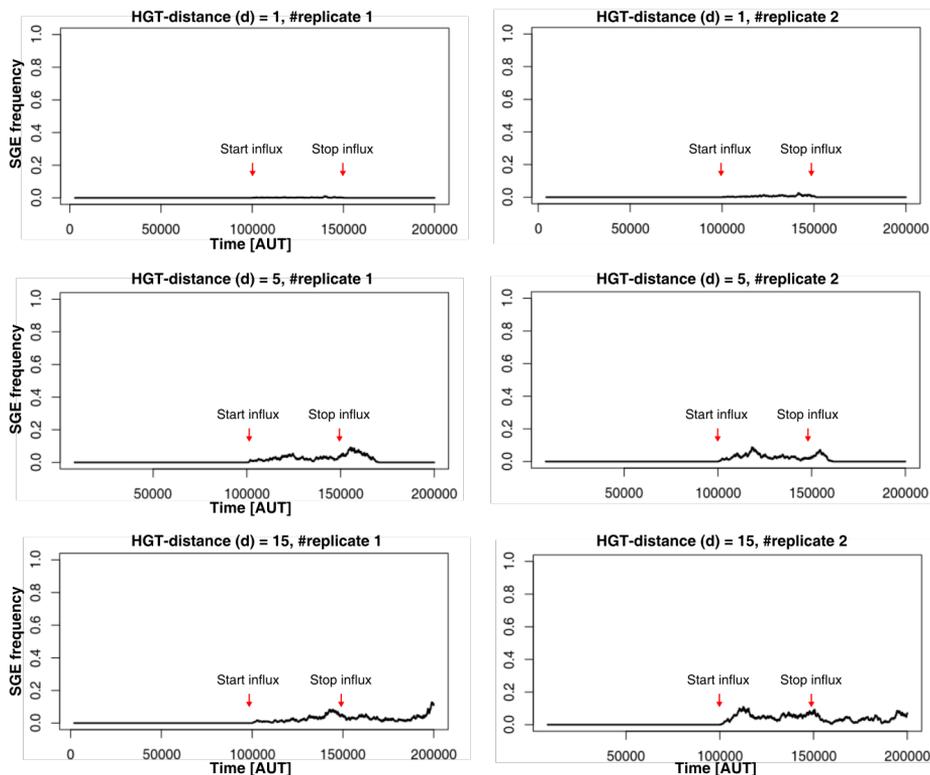


Figure S5.6: SGEs persistence in open and closed ecosystems. (parameters used as in main text)

Supplementary Movie - Gene-sharing ‘nucleation events’ and coexistence with SGEs

This annotated supplementary movie illustrates how gene-sharing of rescuable genes emerges through a ‘nucleation’-event, allowing local communities eventually overgrow all other cells. Where a local sub-community initially transitions to the alternative HGT^+ state, eventually the whole population will be taken over. Similar mechanisms have been observed in origin of life studies Wu and Higgs (2012) and microbial community transitions Kotil and Vetsigian (2018). We also show here how SGEs (here with $b = 0.01$) are able to infect, and stably coexist, with this gene-transferring community.

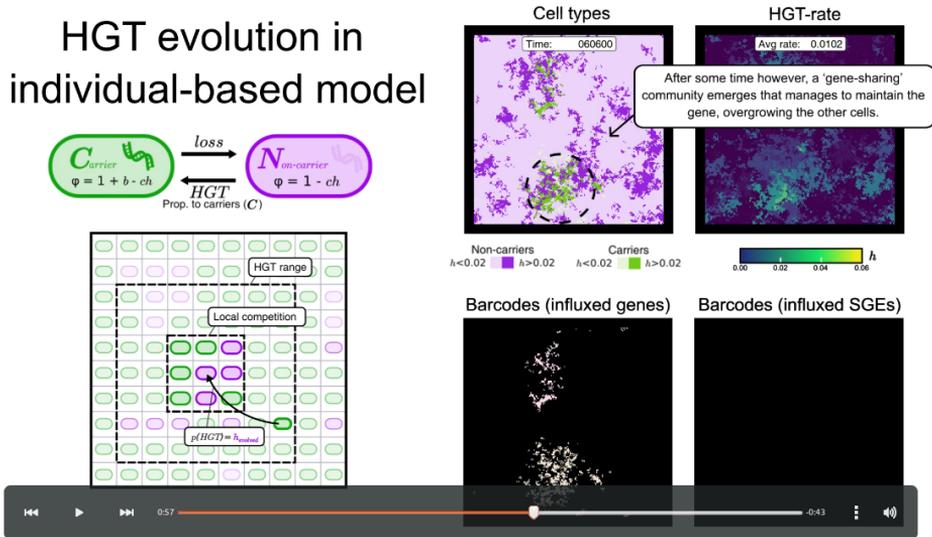


Figure S5.7: A snapshot from the supplementary movie. (<https://youtu.be/cpJh-CeFPm4>)

Table 5.2: Values of h_{opt} that are attained when testing if HGT could be maintained. Starting from $h = 0.05$, we ran the model with evolving h to test if HGT would be maintained. Positions in this table correspond to Figure 4A in the main text. If $h_{evolved}$ persisted at values > 0.001 after 500.000 time steps, we call HGT maintainable. Otherwise, the value in this table is 0.

(b)	IBM (well-mixed)	IBM (d=10)	IBM (d=1)	IBM (d=0)
0.2025	0	0	0	0
0.1975	0	0	0	0.0035
0.1925	0	0	0	0.009
0.1875	0	0	0	0.016
0.1825	0	0	0	0.017
0.1775	0	0	0	0.022
0.1725	0	0	0	0.024
0.1675	0	0	0	0.028
0.1625	0	0	0.003	0.033
0.1575	0	0	0.003	0.033
0.1525	0	0	0.0035	0.035
0.1475	0	0	0.004	0.035
0.1425	0	0	0.004	0.035
0.1375	0	0	0.006	0.035
0.1325	0	0	0.008	0.035
0.1275	0	0	0.01	0.036
0.1225	0	0	0.011	0.036
0.1175	0	0	0.014	0.036
0.1125	0	0	0.016	0.036
0.1075	0	0	0.018	0.036
0.1025	0	0.002	0.02	0.036
0.0975	0.002	0.004	0.021	0.036
0.0925	0.004	0.006	0.022	0.037
0.0875	0.006	0.006	0.023	0.037
0.0825	0.008	0.008	0.023	0.037
0.0775	0.01	0.01	0.025	0.037
0.0725	0.012	0.012	0.026	0.037
0.0675	0.014	0.014	0.027	0.038
0.0625	0.016	0.017	0.028	0.038
0.0575	0.018	0.018	0.028	0.038
0.0525	0.02	0.021	0.028	0.038
0.0475	0.021	0.021	0.029	0.038
0.0425	0.023	0.023	0.029	0.039
0.0375	0.024	0.025	0.03	0.041
0.0325	0.025	0.025	0.03	0.041
0.0275	0.025	0.026	0.031	0.041
0.0225	0.025	0.025	0.032	0.041
0.0175	0.024	0.025	0.029	0.037
0.0125	0.023	0.024	0	0
0.0075	0	0	0	0
0.0025	0	0	0	0
-0.0025	0	0	0	0
-0.0075	0	0	0	0
-0.0125	0	0	0	0

“YOU CAN’T SEE WHICH GRASS IS GREENER.
CHANCES ARE IT’S NEITHER.
EITHER WAY IT’S EASIER TO SEE THE DIFFERENCE,
WHEN YOU’RE SITTING ON THE FENCE.”

—Tim Minchin

Chapter 6

The evolution of alternative genome architectures through Horizontal Gene Transfer and Gene Duplications

BRAM VAN DIJK¹ & PAULIEN HOGEWEG¹

¹ UTRECHT UNIVERSITY, UTRECHT, THE NETHERLANDS

IN PREPARATION

Abstract

Staying fit in light of ongoing mutations is a continuous challenge for evolving microbial populations. Where finite populations can accumulate slightly deleterious mutations, even large populations can only sustain a limited amount of genetic information. Earlier work has shown the importance of Horizontal Gene Transfer in prokaryotes, maintaining genetic information by retrieving unbroken genes from the mobile gene pool. However, recent other work has shown that in eukaryotes, gene duplications can provide “backups” with which to restore broken genes. Here we contrast HGT and gene duplication by studying the long-term evolution of genomes under either of these two mechanisms of gene gain. We find that genomes within these mutational regimes evolve very distinct genome architectures in terms of their genome size dynamics and the basal expression rates of genes. We find that both genome architectures avoid genome deterioration, but by very different means.

6.1 Introduction

Staying fit in light of ongoing mutations is a continuous challenge for evolving microbial populations. Even when a microorganism is well-adapted to a particular niche, natural selection is an imperfect process where genetic drift can cause the accumulation of deleterious mutations. When mutations are only slightly deleterious, the selection coefficient between mutants and mutation-free individuals is insufficient to prevent finite populations from entirely losing the unmutated genes from the population, a process that is known as Mullers Ratchet (Muller, 1964; Pamilo et al., 1987). Similarly, there is an upper limit to the accumulation of beneficial mutations due to a concept known as “the information threshold” (Eigen, 1971; Takeuchi et al., 2005). The latter concept is an entirely deterministic process, and does not depend on the stochastic nature of mutations or finite population sizes. Thus, even in a best-case scenario, evolving organisms are always at risk of their genomes deteriorating.

Several mechanisms to prevent the deterioration of genomes have recently come to light. In bacteria, deleterious mutations and gene loss are often offset by extensive gene gain via horizontal gene transfer (HGT) (Hao and Golding, 2006; Puigbò et al., 2014; Vos et al., 2015; Nowack et al., 2016; Sakoparnig et al., 2019). While HGT is well-known as a mechanism to gain access to new genes (*i.e.* adapt to new environments), it can also repair broken genes by means of homologous recombination. With a mobile gene pool acting as a shared genomic resource (Fullmer et al., 2015), bacteria may benefit from HGT when gene loss or inactivating mutations are high (Vogan and Higgs, 2011; Takeuchi et al., 2015), which is analogous to how sex impacts the balance between drift and selection (Lynch et al., 1995; Schultz and Lynch, 1997; Lynch et al., 2016; Vos et al., 2019). Besides the possibility of *recombining* genes however, HGT also has the ability to add new genes to the genome, a process that has been coined “additive HGT” (Thomas and Nielsen, 2005; Choi et al., 2012; Soucy et al., 2015), which our previous work has shown to play an important role in retaining genes that are only slightly beneficial (van Dijk et al., 2020).

An alternative solution for gaining genes prevalent in eukaryotes is gene duplication. Besides the role of gene duplications in driving important evolutionary innovations (van Dam et al., 2011; Berke and Snel, 2015; Tromer et al., 2019), it may also play an important role in avoiding genome deterioration. By having multiple copies of a gene with a specific function, one copy may serve as a backup for when the other gene is lost or inactivated (Gu et al., 2003). After inactivation, the preserved copy may once again be duplicated, therewith effectively “repairing” the damage. This importance of this mechanism was recently highlighted by Colizzi and Hogeweg (2019), showing that increased duplication and deletions indeed purges broken genes in light of transcription induced mutations of ribosomal RNA genes. Their work highlights how dynamics of gene duplication and deletion provides a solution to prevent genome deterioration in the absence of the homologous recombination. However, it is important to consider that the aforementioned additive HGT, *i.e.* the uptake of genes from other cells or the environment, can act very similar to a gene duplication, as genomes may (or may not) pick up genes they already carry. How does the mechanism of duplications and deletions by Colizzi & Hogeweg relate

to these additive effects of HGT, where genes are picked up from a shared mobile gene pool, rather than from one's own genome?

Here, we contrast the impact of “regular” gene duplications with additive HGT on the long-term genome evolution of simple microbe-like organisms. We do so by exposing a pre-evolved “Virtual Microbe” from a previous study (Meijer et al. 2019, Chapter 3) to two different mutational regimes, one regime with only gene duplications and no additive HGT, and another regime with only additive HGT and no gene duplications. In both scenarios, the total number of expected mutations (deletions, duplications/HGT, and smaller point mutations) are kept the same, the only difference being the origin of duplicated genes. We show that these different regimes result in distinct genome architectures, where the genomes evolved with duplications are larger and show greater fluctuations in size than the genomes evolved with HGT. Finally, we implement a minimal model that can reproduce this distinction between duplications and additive HGT, showing how both are alternative mechanisms to the same end: staying fit in light of ongoing mutations.

6.2 Results

Throughout this study we contrast two different mutational regimes, both describing evolution with a rapid turnover of genes. While the populations in these regimes have identical rates of gene gain, individuals within half of the populations can only gain genes by duplicating their own genes, whereas the individuals in the other half of the populations can only gain genes by taking up genes from other individuals in the population. We will refer to these populations as DUP-populations (for those that gain genes through duplications), and HGT-populations henceforth. We will first explore these mutational regimes in Virtual Microbes, a bottom-up model of the eco-evolutionary dynamics of microbes, after which we present a minimal model that can capture some of the key insights.

6.2.1 Exploring the effects of gene duplications and additive HGT in Virtual Microbes

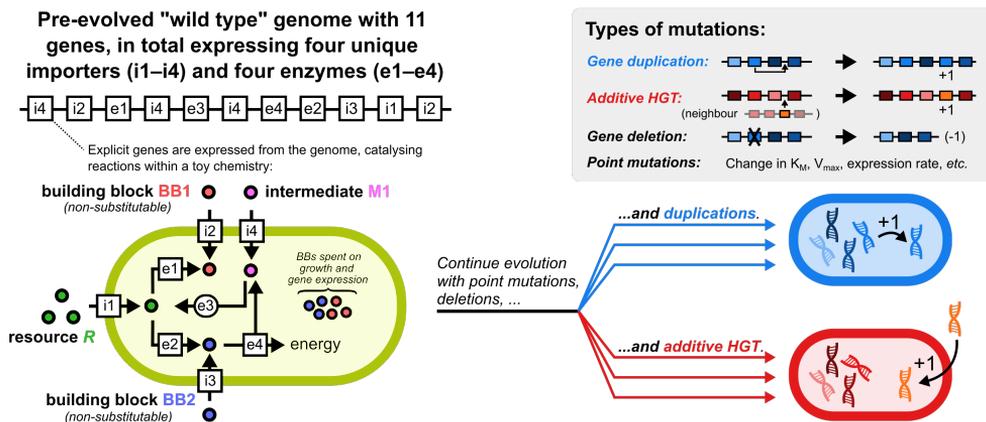


Figure 6.1: Experimental setup: increasing the turnover rate of the gene content of a pre-evolved digital organisms to study the resulting genome dynamics. We expose a pre-evolved Virtual Microbe (*i.e.* a “wild type” with a unique evolutionary history, Chapter 3) to a changing mutational regime. This wild type was evolved under moderate rates of gene deletion, duplication, and HGT, which we increased approximately 10-fold (see Methods for details). In one regime (coloured blue), genomes can only expand in genome size due to gene duplications, whereas in another regime they can only expand by HGT (*i.e.* additively taking up genes from their conspecifics). Other mutations, namely gene deletions and point mutations that change gene parameters, are identical across these regimes. During this experiment, *de novo* discovery of genes is prohibited so that we can focus our analysis on how genomic information is maintained under these new mutational regimes.

We start with a pre-evolved Virtual Microbe “wild type” from an earlier study (Meijer et al., 2019), which did *not* evolve with a rapid turnover of genes (see Methods for more details). Briefly, this wild type expresses four unique importer genes (i1i4), four enzymatic proteins (e1e4), and has a total genome size of 11 due to carrying multiple copies of genes encoding for i2 and i4 (see **Figure 6.1**). Except for using either gene duplications or HGT, all mutations are identical across populations, and include gene deletions (occurring at the same rate as duplications or HGT), changes in Michaelis-Menten constants (*i.e.* K_m , v_{max} , etc.), and changes

in the basal expression rates of individual genes. Note that Virtual Microbes have no explicit penalty for the size of their genome, but excessive gene expression will waste valuable building blocks that could be invested in cell growth.

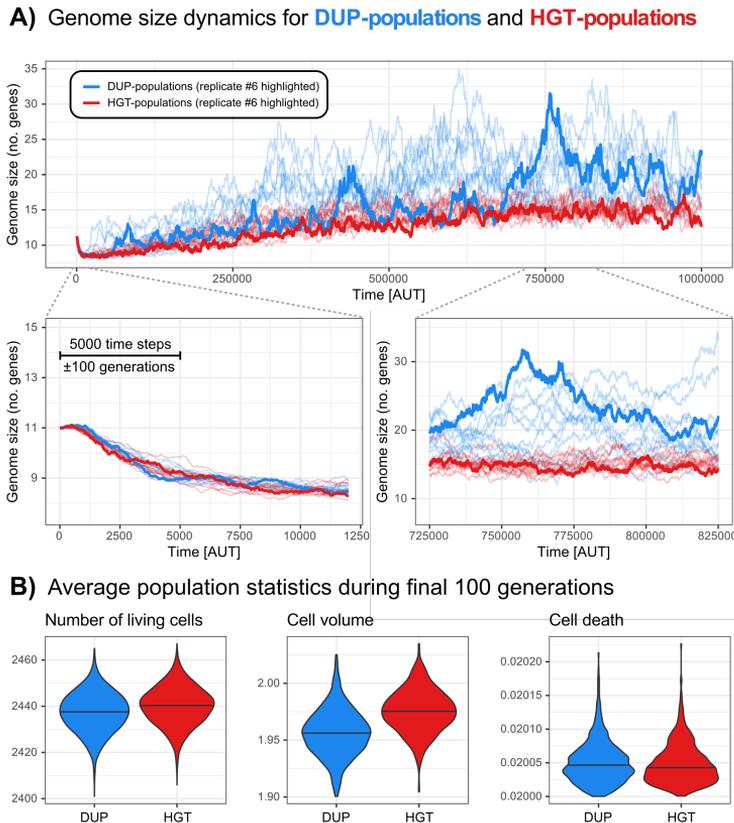


Figure 6.2: DUP-populations and HGT-populations show distinct genome size dynamics after long-term evolution. Panel A shows for 10 DUP-populations (blue) and 10 HGT-populations (red) how the average genome size of the population changes over time. One replicate population (6) is highlighted for illustrative purposes. Panel B shows how some relevant population statistics during the final 100 generations, indicating how the HGT-populations outperform the DUP-populations.

6.2.2 Gene duplication and gene transfer yield distinct genome size dynamics

After being exposed to a 10-fold increase in the rates of gene duplication and deletion, all populations undergo a substantial decrease in genome size during the first 100-200 generations (see **Figure 6.2A**). While this initial decline in genome size happens very rapidly, a long-term evolutionary process results in steady increases in the genome size for all populations. However, the DUP-populations and HGT-populations clearly show distinct patterns. While the HGT-populations stabilise around a genome size of ± 15 , and show only minor fluctuations in genome

size, the genomes in the DUP-populations show transient genome inflations to up to 30–35 genes. Given that de novo gene discovery is disabled (*i.e.* the content of the pan genome is fixed), these striking fluctuations in genome size are only due to copy number variations in the pre-existing gene content. During the final 100 generations, HGT-populations consist of more living cells, have larger cell volumes, and similar death rates **Figure 6.2B**. In other words, from the same amount of resources, HGT-populations produce more (living) biomass.

6.2.3 Initially brittle genomes become robust to null-mutations in the long term

Next we investigated if the individual genomes in the DUP- and HGT-populations showed different patterns in the evolution of their gene content. Note that, although de novo gene discovery was prohibited, differential gene loss can still result in differences in gene content between individuals within and across populations. We analysed the fraction of the population that carried specific gene types (*i.e.* are able to perform a particular metabolic reaction). **Figure 6.3** shows the changes

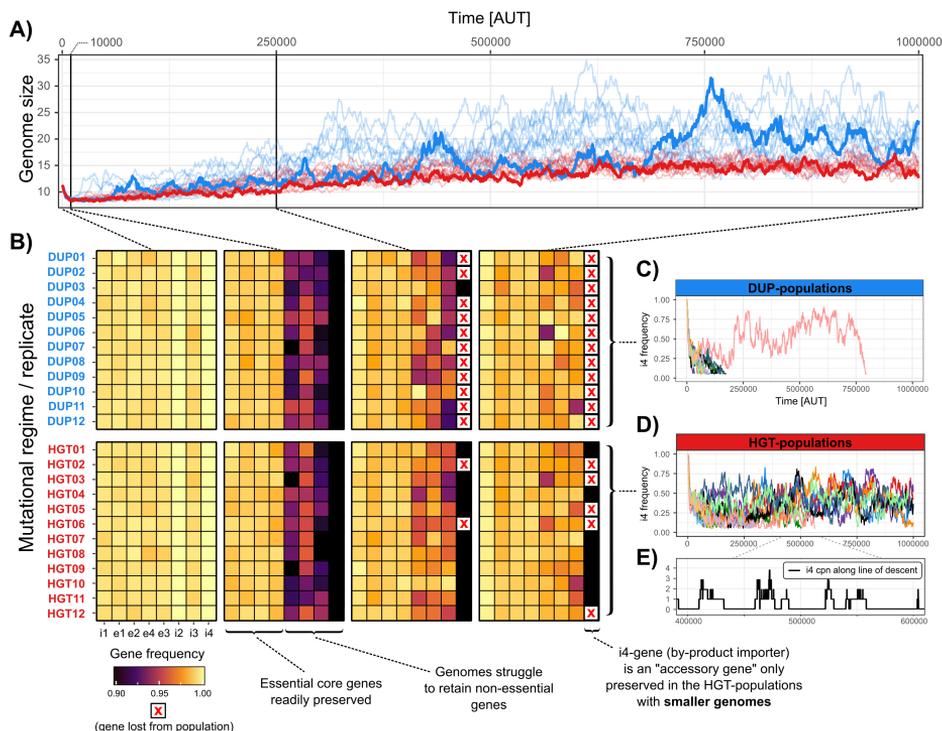


Figure 6.3: Changes in gene frequencies in the DUP- and HGT-populations reflect how initially brittle genomes improve in mutational robustness in the long term. A) Genome size dynamics (similar to **Figure 6.2A**) shown for reference. B) Gene frequencies in the populations after 50, 10,000, 250,000, and 1,000,000 time steps in new mutational regimes. C) The frequency of the i4-gene over time for in the DUP-populations. D) The frequency of the i4-gene in the HGT-populations, accompanied by the copy number of this gene along a line-of-descent.

in these frequencies in the populations over time. As this experiment was started with a clonal population, all genes are still present in high frequencies early in the simulation (**Figure 6.3B**, first two panels). After 10.000 time steps however, many genes show decreased frequencies (**Figure 6.3B**, second panel). Although genes for essential core processes (*e.g.* importing resources, producing the two essential building blocks, and producing energy) are readily preserved in the genomes, the genomes struggle to retain non-essential genes (*e.g.* recycling and reabsorbing spilled by-products). One gene in particular appears hard to maintain (*i4*) and drops to frequencies well below the illustrated 90% cut-off. **Figure 6.3C+D** show how this gene is in fact only retained in some of the HGT-populations, lingering at frequencies of approximately 50% by continuously being lost and rediscovered along the line of descent **Figure 6.3E**. Surprisingly however, the genomes from HGT-populations that have lost the *i4* gene are still more productive (produce more biomass) than the genomes from the DUP-populations, indicating that the differences in fitness are not only due to this extra accessory gene (**Suppl. Figure S6.2**). For the other non-essential genes, **Figure 6.3B** shows how both the DUP- and HGT-populations evolve to be more robust to null-mutations in the long-term.

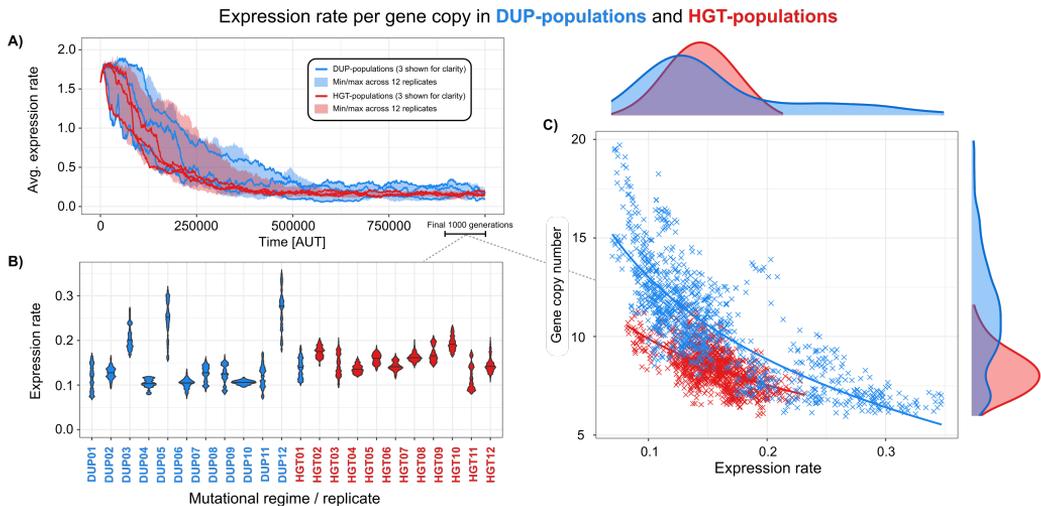


Figure 6.4: To avoid over-expression, the average individual expression rate of genes gradually evolves to lower values. A) The average expression rate of genes in the population is shown over time, where the area shows the minimal and maximal value across all 12 replicates, and 3 replicates are highlighted for illustrative purposes. B) For all replicate simulations, the expression rates as sampled during the final 1000 generations are shown in a violin plot. C) Each point in the graph is the average genome size in a population versus the average expression rate of its genes, as sampled during the final 1000 generations. Blue symbols are drawn on top of red symbols. The regression lines are fitted logarithmic curves. The histograms along the axes depict the corresponding densities of gene copy numbers and expression rates.

6.2.4 The evolution of expression rates underpins different genome size dynamics

From the results above we can conclude that there is no direct relationship between genome size and the number of unique genes, as the smaller genomes from the HGT-populations are surprisingly able to retain more different types of genes. Thus, to better understand the differences in genome size, we further investigated what entailed the variations in gene copy numbers. First, note that having many gene copies (*i.e.* a large genome) does not impose a direct burden on the cells, but that over-expression of genes may waste valuable building blocks. However, not performing certain metabolic functions at all may be much more costly, explaining why genomes in this regime of high gene turn-over require multiple copies to act as a “backup”. Concurrent with the inflation of the genome, **Figure 6.4A** shows that the average expression rate per gene declines to much lower levels than that of the initial ancestor. Although in equilibrium both the DUP- and HGT-populations have similar average expression of ± 0.15 (**Figure 6.4B**), the DUP-populations show much greater fluctuations, reminiscent of the above-mentioned fluctuations in genome size. From **Figure 6.4C** it can indeed be seen how the DUP-populations and HGT-populations show very distinct patterns in terms of how the expression rate scales with gene copy number. Where the DUP-populations appears to have either many lowly expressed genes or few highly expressed genes, the genomes from the HGT-populations stably maintain relatively moderate copy numbers with intermediate expression rates. Based on these observations, we hypothesised that the dynamics of having backup genes on the one hand, and the costs of over-expression on the other hand, are sufficient to describe the observed differences between DUP- and HGT-populations.

6.2.5 A minimal model of the evolution of gene copy number and expression rate

So far, we have studied how in the complex eco-evolutionary setting of Virtual Microbes, the processes of HGT and gene duplications yield fundamentally different genome architectures in terms of genome size and the individual expression rates of genes. The bottom-up design and open-ended nature of Virtual Microbes allowed the model to reveal an unforeseen pattern of stable and fluctuating genome sizes, and surprise us with the relatively counter-intuitive insights that the smaller genomes from the HGT-populations contain more different types of genes. Note that apart from the basal expression rates, the individual genes showed no clear differences in other functional parameters (e.g. K_m or V_{max}). We therefore hypothesised that these observations could be primarily understood in terms of copy number variations on the one hand, and evolving rates of gene expression on the other hand. To test this hypothesis, we composed a minimal model that can indeed yield similar observations (**Figure 6.5**, see Methods for details).

Briefly, the model describes the evolution of genomes on a simple fitness landscape. The genomes only express a single type of gene that needs to be expressed sufficiently, but not more so (*i.e.* over-expression yields a fitness penalty). Like in the Virtual Microbes model, the copy numbers of this gene can vary through gene deletions

and duplication / HGT, and the expression rate of individual genes can mutate upon reproduction. A genome carrying a single gene that is expressed sufficiently has the same *direct* fitness as a genome carrying multiple copies that are together similarly expressed.

Indeed, the simple model confirms that the genome size dynamics as observed in **Figure 6.2A** can be understood in terms of the evolution of expression rates **Figure 6.6**. Where DUP-populations have large (fluctuating) genomes with lower gene expression, HGT-populations have small genomes with high(er) gene expression rates. Surprisingly however, where we have shown how the Virtual Microbes evolved with HGT were more productive than those evolved with duplications, the two alternative mutational regimes yield identical fitness in this minimal model. The Virtual Microbes evolved with HGT thus appear to exploit the many more degrees of freedom that are present in an eco-evolutionary model like Virtual Microbes.

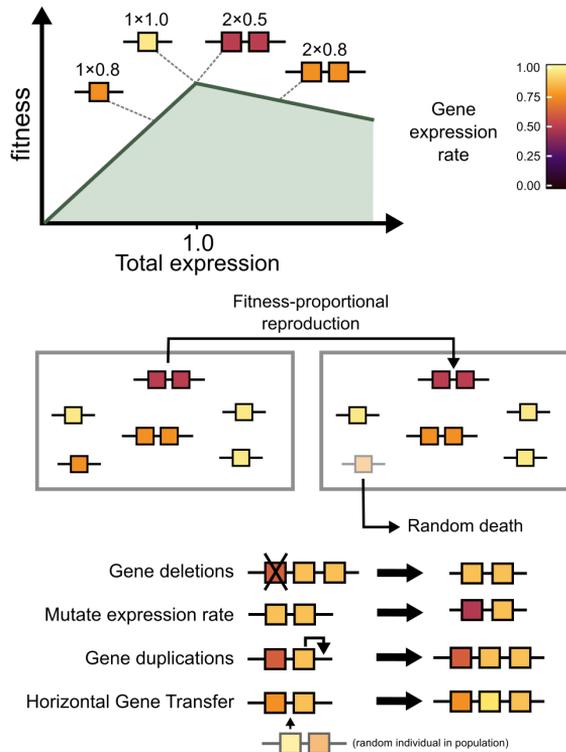


Figure 6.5: A minimal model describing genome evolution on a simple fitness landscape. To study whether copy number variations and individually evolving expression rates of genes were sufficient to explain the patterns as observed, we implemented a minimal model that includes only these variables. Where the Virtual Microbe model had 8 different gene types that performed various functions, this minimal model only considers a single type the total expression of which can give a total fitness of 1.0. Individuals expression two genes with expression 0.5 are as fit as individuals expressing a single gene with expression 1.0, or four genes with expression 0.25, etc. (see Methods for full implementation)

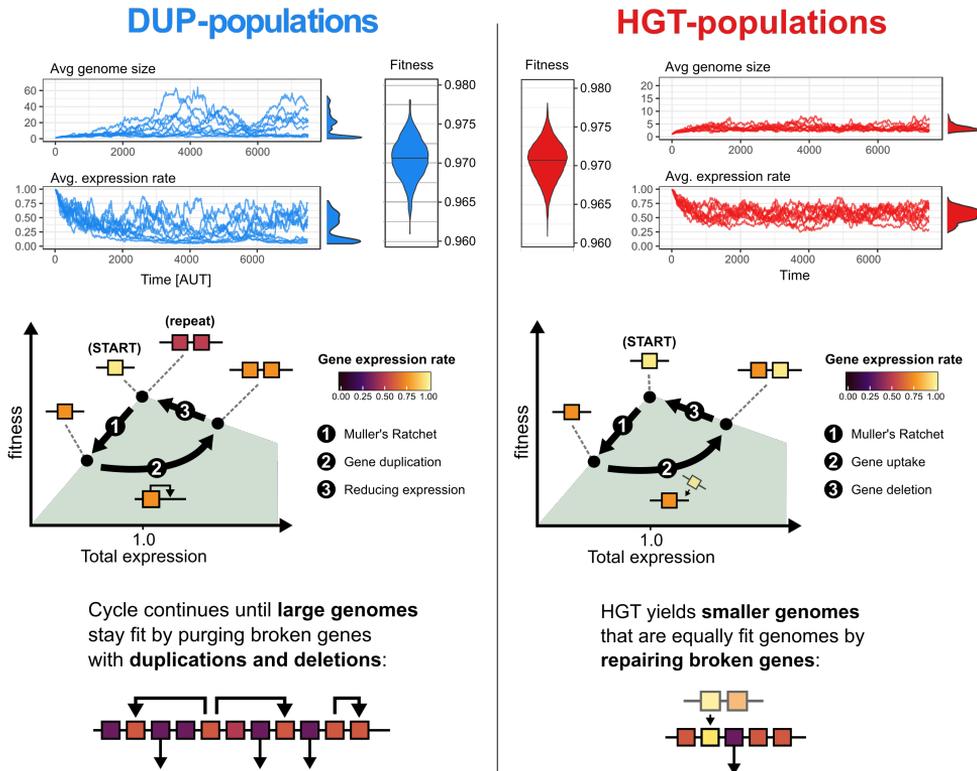


Figure 6.6: Gene duplications and gene uptake yield alternative genome architectures, but both avoid genome deterioration. Using the minimal model from **Figure 6.5**, the observed genome size dynamics of Virtual Microbes can be explained in terms of copy number variations and individual expression rates of genes. The graphs at the top include 10 replicate simulations for each mutational regime to illustrate the consistency of the pattern.

6.3 Discussion

In this study, we have explored and contrasted the differential effects of intragenomic gene duplications and additive HGT. We do so by evolving Virtual Microbes in one of two extreme mutational regimes: one where only duplications are allowed, and one where genes can only be gained by taking up genes from other individuals. We find that even in a simple, constant environment, these two mutational regimes yield very different dynamics in terms of genome size. Genomes evolved with only gene duplications are larger, and furthermore show much greater fluctuations in genome size than genomes evolved with only HGT. Nevertheless, genomes evolved with HGT retain more unique types of genes, being able to continuously rescue an accessory gene that is lost from genomes without HGT. Further, we show that this ability to retrieve genes can also explain why genomes evolved with HGT are smaller, being able to repair the damage of slightly deleterious mutations in the expression rates of genes. The large and fluctuating genomes evolved with gene duplications on the other hand cannot easily repair this damage, and resort to making extra gene copies in order to retain sufficient gene expression. Finally, the minimal model has also shown that, although both mechanisms result in very distinct signatures in their genome dynamics, they stay equally fit in light of ongoing mutations.

Similar mechanisms to avoid genome deterioration to the ones revealed by this study have been previously reported. One such mechanism is to avoid the accumulation of deleterious mutations by recombining with unbroken genes from the gene pool (Takeuchi et al., 2014). This mechanism was however described in the absence of copy number variations, where each allele only occurred once in each genome. By assuming HGT is additive rather than replacing (see Choi et al. 2012; Soucy et al. 2015), we here studied how this ability of HGT to “rescue” genes from the gene pool (van Dijk et al., 2020) interplays with genome architecture. Recently modelling work has shown that — in the absence of HGT — increased rates of duplications and deletions can indeed purge broken gene clusters from evolving yeast genomes (Colizzi and Hogeweg, 2019). Our results complement these earlier studies by directly comparing these alternative mechanisms, and showing that these mechanisms can emerge in an eco-evolutionary model with many additional degrees of freedom.

6.3.1 Mutations as drivers of genome size evolution

It has been known for decades that, even within a single bacterial species, genomes can vary substantially in size (Herdman, 1985). If large genomes pose a direct energetic burden on the host, relying on HGT to (re)gain access to genes in the pangenome may allow bacteria to adapt to various ecological opportunities whilst still having a small genome. Our model however does not include such an explicit cost on the size of the genome, but nevertheless shows that a mutational regime dominated by HGT results in the evolution of compact genome architectures. In other words, we see that engaging in HGT results in small genomes, rather than small genomes resulting in the necessity to use HGT. Moreover note that the much larger genomes evolved in the absence of HGT can attain the same fitness, reaffirming the absence of direct costs on genome size. Indeed, other modelling

studies that do not assume a direct cost on genome size have revealed how mutations can be important drivers of the expansion or shrinkage of genomes. For example, genome size dynamics may be dominated by the interplay between neutrality and adaptability (Cuypers and Hogeweg, 2012), or may be driven by inherent biases towards shrinking even when duplications appear to outpace deletions (Fischer et al., 2014). More recently, Rutten *et al.* have shown that very specific genome architectures may evolve to reshape the distribution of fitness effects (Rutten et al., 2019), where hypermutators evolve to compress their coding sequences but inflate the non-coding parts of their genome. The large genome results in more duplications and deletions, and increases the selection coefficient between individuals and their mutated offspring. Thus, although dealing with ongoing mutations by evolving an even bigger genome appears initially counter-intuitive, the evolution of specific genome architectures can help to resolve this.

6.4 Conclusion

We have revealed how initially identical genomes show striking differences in their genome size dynamics depending solely on the origin of duplicated genes. Genomes that rely on “regular” gene duplications purge broken genes by large genome inflation and subsequent streamlining, while genomes that rely on HGT remain small by picking up genes from a shared gene pool. We conclude that, where staying fit in light of ongoing mutations is a continuous challenge for evolving organisms, genomes will organise themselves to avoid the deterioration of their genes.

6.5 Methods

6.5.1 Overview of the Virtual Microbes model

A full description of the model and underlying equations is available online (bitbucket.org/thocu/virtual-microbes and <https://virtualmicrobes.readthedocs.io>). Here we summarise the sections of these documents that are relevant to this study.

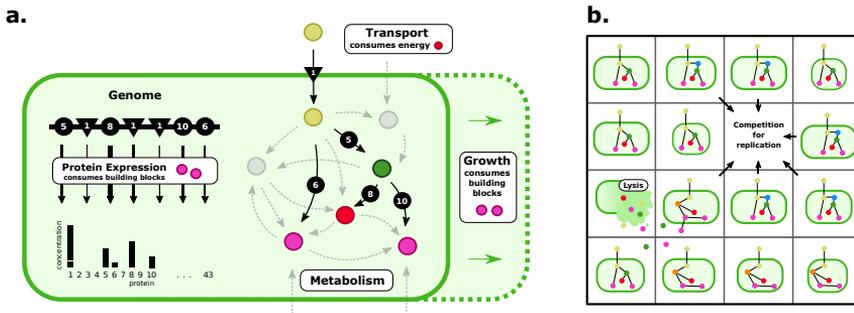


Figure 6.7: Virtual Microbes overview

a. Genes on a linear genome code for specific metabolic enzymes that catalyse individual reactions within an “artificial biochemistry”. To grow and express proteins microbes require two non-substitutable building block metabolites BB_1 and BB_2 , (purple) that do not natively occur in their environment, but can be metabolized from the single provided resource R (yellow) by expressing functional metabolic networks. Active transport of metabolites across the cell membrane requires an energy metabolite E (red). **b.** Microbes compete for space and metabolites on a 50×50 lattice. If they reach a minimal division cell size, they can reproduce in an empty space in eight adjacent sites – here, microbes NE and W of the empty site are too small to reproduce. Upon replication, genomes can mutate through gene duplication, gene deletion, and horizontal gene transfer. Individual genes can also mutate their expression rates and kinetic parameters. Active transport of metabolites and lysis changes the composition of a microbes’ local environment.

Virtual Microbes metabolise, grow and divide on a spatial grid (Figure 6.7). Here, we used two parallel 50×50 grids with wrapped boundary conditions. One grid contains the Virtual Microbes and empty grid-points, and the other describes the local environment in which the Virtual Microbes live. This environmental layer holds the influxed resource R and other spilled metabolites such as building blocks and by-products. In order to express proteins, grow, and maintain their cell volume, Virtual Microbes must use predefined metabolite(s), which we call building blocks. These building blocks are not directly provided, but must be synthesised by the Virtual Microbes by expressing the right proteins, allowing them to pump metabolites into the cell, and convert metabolites into one another (Figure 2.1A). The expression of these proteins depends on genes on a circular genome that changes in content due to gene deletions, gene duplications, and/or uptake of genes from neighbouring conspecifics. Smaller “point mutations” can change parameters of genes (*e.g.* the K_m , V_{max} , or the expression rate parameter of an individual gene). When an enzyme or transporter gene is expressed, that specific reaction will take place within the cell that carries that gene. Which genomes are most fit is not explicitly defined, but the result of a birth and death process, where birth depends on the availability of empty space and the synthesis of new building blocks, whereas

Table 6.1: Implemented changes in mutation rates

Mutation	Description	Rate (this study)	Rate (wild type)
Point mutation	Per gene per generation probability of modifying a single parameter of a gene (expression rate or Michaelis Menten constants)	0.01 / 0.01	0.02
Deletion	Per gene probability of deletion	0.01 / 0.01	0.001
Duplication	Per gene probability of duplication	0.01 / 0.00	0.001
HGT	Upon birth, per gene probability of picking up a gene from an individual closeby	0.00 / 0.01	$\pm 0.0016^*$

* In the evolutionary history of the wild type, the uptake of genes (HGT) was implemented as a probability per time step, and did not depend on genome size. Here, to have a fair comparison between the effects of duplications and additive HGT, we implemented the uptake of genes to also occur on a per-gene basis. The effective rate of additive HGT for the ancestor was still much lower than in the new regime (± 0.0016 per gene).

a minimal death is implemented to free up new reproduction space. While new-born cells are born with a minimal rate of cell death, overzealous metabolic activity can result in the accumulation of by-products or other metabolites that increases death rates. The resulting survival of the fittest (referred to as “competitive fitness” by Fragata et al. 2018) is determined by many potential variables, and is an emergent phenomenon of the eco-evolutionary dynamics.

The metabolism described above is simulated on the grid in terms of Ordinary Differential Equations (ODEs). These ODEs include the influx of molecules into the system, transport or diffusion across the membrane, intracellular metabolism (including expression and decay of proteins), biomass production, cell volume, the build-up of toxicity, *etc.*, and are fully described in previous work (van Dijk et al., 2019).

Implemented changes in mutation rates

In the main text, we take a pre-evolved Virtual Microbe from an earlier study (Meijer et al., 2019) and subject it to two different mutational regimes. Populations in both regimes undergo gene deletions and point mutations that change the parameters of genes (see **Table 6.1**). One set of replicate populations (called DUP-populations in the main text) can only use intragenomic gene duplications to increase gene copy numbers. The other set of replicate populations (called HGT-populations) can only take up genes from their conspecifics, and do not undergo intragenomic gene duplications.

Minimal model on a simple fitness landscape

To study whether copy number variations and individually evolving expression rates of genes were sufficient to explain the patterns as observed, we implemented a minimal model that includes only these variables. Where the Virtual Microbe model had 8 different gene types that performed various functions, this minimal model only takes into account a single type of gene. An individual is defined as a vector of expression rates:

$$\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad (6.1)$$

Giving an individual i a total expression E_i of:

$$E_i = \sum_{j=1}^n e_j \quad (6.2)$$

The total gene expression linearly increase the individual's fitness F_i , with a maximum fitness benefit of 1. Expression also gives the individuals a small fitness cost ($c = 0.01$, see Equation 6.3), resulting in the fitness function:

$$F_i = \min(1.0, E_i) - E_i \cdot c \quad (6.3)$$

where the maximum fitness is $1 - c$.

Every time step, N individuals evolve according to a simple Moran process (also see cartoon in main text **Figure 6.5**). First, an individual is sampled proportional to its fitness. Then, this individual reproduces, replacing a random individual in the population. Upon reproduction there is a per-gene probability of gene deletion, gene duplication, gene uptake, and mutations in the expression-rate parameter. The new expression-rate parameter is a value between the current value, plus or minus a small uniformly sampled step size of 0.1. Assuming breaking genes is easier than fixing them, 90% of mutation on the expression rate lower expression, whereas 10% increase the expression rate. This model was implemented in Python 2.7, and is publicly available: <https://github.com/bramvandijk88/HGTvsDUP>.

6.6 Supplementary Material

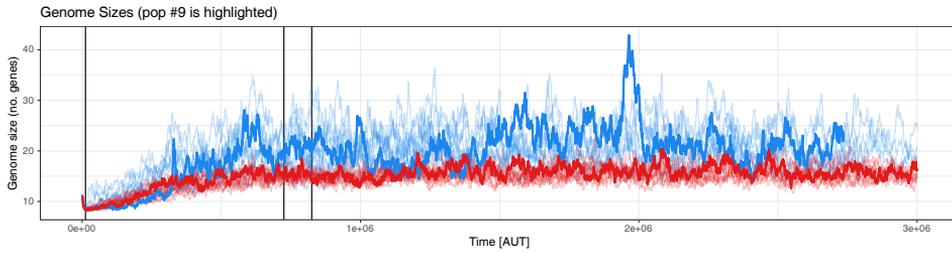


Figure S 6.1: Distinct patterns of genome size dynamics for DUP- and HGT-populations persist indefinitely. For 10 DUP-populations (blue) and 10 HGT-populations (red), this graph shows how the average genome size of the populations changes over time, extended until up to 3×10^6 timesteps ($\sim 60,000$ generations). One replicate population (9) is highlighted for illustrative purposes.

HGT-populations are more productive (more living biomass) even when the *i4*-gene is not retained

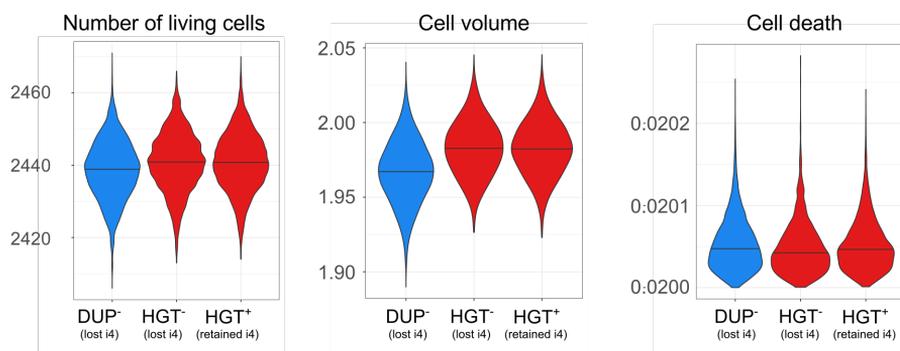


Figure 6.2: HGT-populations produce more biomass even when the *i4*-gene is not retained. In the main text it is shown how HGT-populations are more productive (produce more biomass) than the DUP-populations, but we also show how they are able to retain an accessory gene. However, even the populations that have not retained this gene are more productive than the DUP-populations.

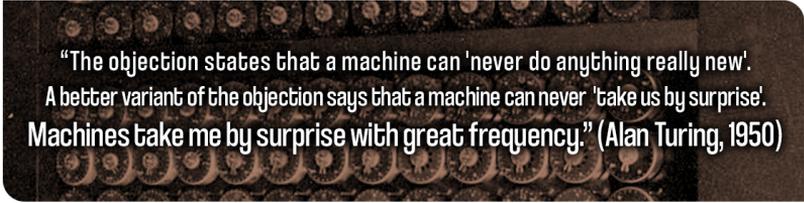
“MACHINES TAKE ME BY SURPRISE WITH GREAT FREQUENCY.”

—Alan Turing

Chapter 7

Discussion

Throughout this thesis we investigated the ecological and evolutionary dynamics of life at the micro scale. We explored the influence and emergence of different levels of selection, and how the interplay between these levels impacts the diversity of microbial populations. We have also studied the extent to which evolution is predictable, and how this predictability is impacted by prior evolution. More into detail, we have studied how genes and genomes are shaped by the process of Horizontal Gene Transfer (HGT). All this was done, as the title of this thesis suggests, “on a Virtual Grain of Sand”, *i.e.* using computational modelling. Our main objective was to gain insight into the biological evolution of microorganisms, and doing so by using computational models often led to unforeseen insights. One may wonder how it is that we found ourselves so frequently ‘taken by surprise’ by the models we ourselves designed. Note however that the ability of computation to yield surprising behaviour was already discussed by Alan Turing in the 1950s:



“The objection states that a machine can ‘never do anything really new’.
A better variant of the objection says that a machine can never ‘take us by surprise’.
Machines take me by surprise with great frequency.” (Alan Turing, 1950)

Thus, besides using this final chapter to discuss open questions and speculate on future directions, I also wish to address the extraordinary ability of computational models to yield unforeseen insights. More specifically, I argue that the ability of models to surprise us is best taken advantage of by building multilevel models from the bottom up. But before doing so, let us first recapitulate some of the main insights that were gained in this thesis.

7.1 Chapter reviews

In **Chapter 2** we set out to study the predictability of evolution by exposing a set of pre-evolved Virtual “wild types” to a simple serial transfer protocol. We found that independent of their evolutionary history, all wild types learn to anticipate the predictable resource cycle by evolving very fine-tuned growth dynamics. Despite the repeatability of this anticipatory behaviour however, our experiments also revealed the great many underlying mechanisms by which this outcome could be attained.

Some replicate experiments resulted in single lineages that regulate their gene expression to be adapted to both feast and starvation, while other experiments resulted in multiple lineages that specialise on these respective niches. Other evolutionary outcomes were less elegant, simply consisting of a single lineage that exploits as much resource as is possible, making sure to survive until the next pulse of resources. When comparing these mechanisms across the different wild types, we found some wild types repeatedly used the same mechanism in a very predictable fashion, while others evolved a different mechanism every time the experiment was repeated. Thus, where we set out to answer whether evolution is predictable or not, we surprisingly found that the predictability of evolution can *itself* be unpredictable.

Chapter 3 continued on the theme of the predictability of evolution, but this time in a spatially structured population. In this experiment, although only a single resource was externally provided, metabolites accumulated in the local environment through efflux and cell death. This allows for metabolic exchange of important building blocks and by-products. We found that the evolved populations would either contain two interdependent cross-feeding species, or a single species that is entirely self-sufficient. The difference between these outcomes was strongly determined by prior evolution in the energy metabolism of the populations, which was itself determined by chance. In other words, where we could not predict precisely which population would evolve cross-feeding early in the evolutionary experiment, our predictions became very accurate once certain metabolic signatures had evolved. Similar to Chapter 2, we again find that predictability is an evolved property.

Where an important theme of the first two chapters was the predictability of evolution, they also repeatedly showed how microbial diversity can emerge. For example, Chapter 2 showed species adapted to temporal niches of feast and starvation, whereas in Chapter 3 species repeatedly emerged that evolved to be metabolically dependent on one another. Instead of the emergence of diversity, however, **Chapter 4** revealed a new mechanism to *maintain* microbial diversity. More specifically, we studied the importance of gene-level selection and horizontal gene transfer (HGT) on maintaining a rich set of toxin production genes and their corresponding resistance factors. We showed how, when genes are given the opportunity to evolve their individual rates of horizontal transfer, the model always converged to a phylogenetically diverse population of hyper-immune strains with very frequent horizontal transfer of toxin genes. Surprisingly, while one would expect that such frequent transfer of toxin genes would abolish diversity, the resulting population would retain many strains that diversified a long time ago. Moreover, this “deep” phylogenetic diversity was not observed when toxin genes were not allowed to transfer more frequently than their corresponding resistance factors. In other words, in this particular case, evolution self-organised to maintain high diversity. Where earlier modelling by Vetsigian et al. (2011) explained the diversity of antagonistic interactions by implementing the continuous discovery of novel genes, the resulting Red Queen dynamics contradict the deep phylogenetic relationships that are observed in natural populations. Instead, our model shows how a highly diverse set of toxin and resistance genes can be maintained even in a small bacterial population, without disrupting the rich diversity of strains and their phylogenetic relationships. Taken together, this reconciles two observations in the literature that initially seem contradictory: rampant HGT and deep phylogenetic diversity.

In the previous chapter we have studied the impact of HGT by taking the perspective of the genes. An important result of this study was that the toxin genes, which were on average only slightly beneficial, could only persist by evolving high mobility. However, it is also important to consider how gene mobility is also (partially) determined by the host, for example when bacteria take up DNA from the environment. Therefore, in **Chapter 5** we set out to better understand HGT from the perspective of the cell. We study the impact of HGT on bacterial growth rates, and subsequently investigating whether or not the cellular uptake of DNA could evolve *de novo*. First, using a simple analytical model, we show that HGT indeed allows for slightly beneficial genes (*i.e.* like the toxin genes) to persist within a bacterial population, and that this can indeed improve the growth rate of a bacterial population. Next, we found that the *de novo* evolution of the ability of cells to engage in HGT faced an interesting paradox: certain genes were unable to persist without HGT, but HGT was only adaptive for cells in the presence of the gene. By studying these dynamics in a spatially structured population however, this apparent paradox was resolved as local “gene-sharing” communities emerge. A second surprising observation in the spatial model was that bacteria could coexist even with highly deleterious selfish genetic elements (SGEs), retaining their ability to take up DNA in spite of these genetic parasites. In well-mixed populations, the global threat of such deleterious SGEs would result in strong selection against DNA uptake. Thus, only in a spatial context can we understand the long-term coexistence of SGEs and bacteria that engage in HGT through the uptake of DNA.

In **Chapter 6** we contrasted the ability of HGT to “rescue” genes by taking up genes from the gene pool (*i.e.* Chapter 5) with gene duplications. Where HGT is especially common for prokaryotes, gaining genes by means of gene duplication is much more prevalent in eukaryotes. Through both mechanisms an extra gene is gained, but that the origin of that gene is of course very different. Without an initial intuition on the potentially important variables for comparing these two mechanisms of gene gain, we exposed a pre-evolved Virtual Microbe strain from Chapter 3 to two mutational regimes. In one mutational regime gene gain happened through additive HGT, whereas in the other regime gene gain happened through intragenomic gene duplications. A generic pattern that emerged was that the genomes exposed to gene duplications evolved to be much larger than genomes evolved with HGT. Moreover, the populations exposed to gene duplications fluctuated much more in size, *i.e.* periodically showing phases of extreme genome expansion. We found that the genome inflations could be explained by a reduction in individual expression rates of genes, which is subsequently compensated by duplicating genes. The populations under the HGT-regime however maintained much more stable expression rates, and therewith did not need to compensate by making extra gene copies. Thus, although populations in both mutational regimes converge to very different genome architectures, both adequately avoid the deterioration of their genomes.

7.2 Models and unforeseen insights

Models can be surprising in many different ways. In reviewing the chapters of this thesis, however, note how our models did not only surprise us in terms of the values of variables, but also gave us entirely new concepts to think about. Moreover, many

of these concepts were initially very counter-intuitive. For example, we especially observed high diversity in Chapter 4 in a “closed” ecosystem, *i.e.* when no new DNA was introduced. We were similarly surprised to find how microbes may learn to trust the hand that feeds (Chapter 2), how cross-feeding yields self-mixing populations that however behave very different than well-mixed systems (Chapter 3), that ‘gene-sharing’ can emerge in spatially structured populations (Chapter 5), and that different mechanisms of gene gain result in distinct genome architectures (Chapter 6). Below, I will argue that the ability of models to generate surprising insights is amplified when building a multilevel biological system from the bottom up. Such multilevel models have many degrees of freedom, and therewith many possible outcomes. Nevertheless, this thesis has shown how evolution tends to converge to only a limited set of outcomes, from which we can extract generic principles of evolutionary processes.

7.2.1 Building multilevel models from the bottom up

As was already touched upon in the introduction of this thesis, the mounting observational data from continuously advancing experimental techniques is giving us very detailed knowledge on the parts that make up biological systems, as well as their physiological details (**Figure 7.1**). Across many scales, we now understand a lot about the important components (genes, proteins, nutrients, cells, biofilms, *etc.*), as well as the structures and interactions that arise because of the interplay between these components (gene regulatory networks, metabolic networks, cell cycles, microbial communities, *etc.*). Due to this complexity and the many levels however, it is not always easy to interpret the underlying dynamics of these systems. In other words, what do these systems actually look like when the wheels are turning? This is of course where modelling plays an important role. However, many models in biology have been describing mostly the dynamics of these systems in the abstract, while the actual intricacies of the relevant biology remain on the background. For example, simple models may describe changes in variables such as population size, gene expression, or metabolic flux, but the actual structures (genomes, regulatory networks) and the many biological levels (genes, cells, collectives) are rarely considered explicitly. Of course, adding all the possible intricacies of life to a model is not analytically or computationally feasible, and the model may furthermore quickly become incomprehensible. Instead, a good model is somewhere in between simplicity and complexity, where it can be surprising on the one hand, and comprehensible on the other hand. However, with all the possible assumptions and details that could be added to a model, where do we begin?

7.2.2 Biological basic principles

Considering the problem of choosing what to include in a model, taking a bottom-up perspective can help. Where physicists commonly model from first principles (*i.e.* a basic assumptions that cannot be deduced any further), we could similarly contemplate on the “basic principles” of biology. Consider for example how processes like local reproduction, metabolism, and mutations, are simply part of the physical reality in which biological systems live. By implementing these basic principles, we can simply let the system “live its life”, where we may have very little prior

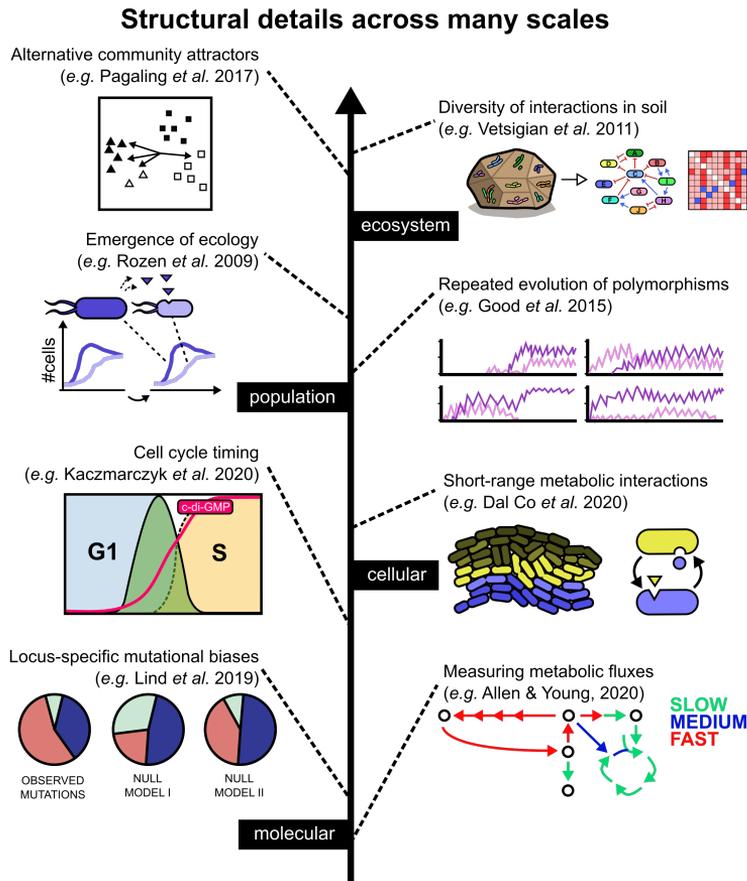


Figure 7.1: Mounting observational data on microbial metabolism, ecology, and evolution across many scales.

expectations of what the model *should* do. Indeed, digital evolution has often been shown to be surprisingly creative, and instead of the programmer debugging the model, the model oftentimes “debugs” the programmer (Lehman et al., 2018). With sufficient degrees of freedom, models can be an open-ended and unsupervised way to investigate what the system *will* do given a set of basic rules, generating new observables and concepts along the way.

7.2.3 The unique exploratory potential of complex models

Although a bottom-up model of biology will produce observables similar to fitness, population growth rates, and microbial communities, they are not explicitly modelled as such. A notable downside to this may be that a lot of complexity emerges from such a model, while a simpler TOP DOWN description could capture the behaviour just as well. Especially when combined with many degrees of freedom, a bottom-up model has many more parameters and variables than a simpler analytical model. What then, are the advantages of such a complex model?

When explaining existing observations, simple models are a powerful tool to investigate which mechanisms can explain the observation (and which mechanisms fail to explain it). Similarly, once having observed some interesting behaviour in a relatively complex model, simpler models may be applied to investigate further. Note however, given the initially surprising nature of some of the observed patterns, the simplest possible model may be a counterproductive starting point. For example, when we studied the effects of gene duplications and HGT in Chapter 6, we were initially surprised with the generic patterns of genome inflations and gene expression rates. As the evolution of gene expression rates is one of the many degrees of freedom in Virtual Microbes, it was one of the many variables that could have been important for these dynamics. By first allowing a model with many degrees of freedom to live its life, we were eventually able to capture the dynamics in a much simpler “model of our model”. In other words, seeing and experimenting with the behaviour in Virtual Microbes was the first essential step towards progress. Simply put, it is no use to search for diamonds in the dark without a torch: *“Complex models have the ability to shed light, and simple models search in that light.”* (Guillaume Beslon, personal communication). Daring to build multilevel bottom-up models, and the extra levels of complexity that may entail, will help us to better understand not only microbial evolution, but also biology writ large.

7.2.4 Structures, processes, and levels of selection

After having discussed how complex models can be used as an exploratory tool, I would finally like to touch upon another important advantage of building multilevel models from the bottom up. While many of the models presented in this thesis may look like typical individual-based models, note how they are not primarily focused on the individual at all. Instead, they typically also encapsulate many of the levels below the individual (*e.g.* genes and metabolism) and the levels above the individual (*e.g.* community structure and environmental interplay). A good example of this is the model of antagonistic interactions presented in Chapter 4, which includes gene- and cell-level selection, but also competition between spatially separated strains. Without focusing on the importance of each individual level, we can instead focus on what processes, structures, and patterns persist and are repeatedly re-produced. Indeed, by shifting focus from single units of selection towards evolution as a multilevel process (Hogeweg, 2012) that results in differential persistence of processes and structures (Hogeweg, 2007; Doolittle and Booth, 2017; Ibanez, 2020), we may better understand what outcomes are attainable, and maintainable, by evolution.

7.3 Outlook

Amongst other subjects, this thesis has primarily addressed the processes that make or break microbial diversity, the predictability of evolution, and the impact that HGT has on evolving microbial populations. In this final section, I would like to discuss some interesting open questions that relate to these topics, and suggest future directions to further advance the field of microbial ecology and evolution.

7.3.1 Embracing HGT in evolutionary theory

Interestingly, the effects of HGT have first been observed almost a century ago (Griffith, 1928), two-and-a-half decades before the structure of DNA was resolved (Watson et al., 1953). Despite this early discovery and the astounding rates at which we now know HGT happens in natural systems (Puigbò et al., 2014; Nowack et al., 2016; Sakoparnig et al., 2019; Tidjani et al., 2019), the process of HGT is still poorly embedded in much of evolutionary thinking. Much modelling still builds on the theory of population genetics (Fisher, 1930; Wright, 1931; Haldane, 1932), which besides rarely including dynamics of ongoing mutations, does not take into account the intricacies of HGT (**Figure 7.2**). Notably, HGT facilitates many processes that are not directly analogous to recombination or sex, such as shared gene pools, genes spreading independent of the host's cell cycle, and the introduction of new genes from external sources. In this thesis, we have seen how HGT can rescue slightly beneficial genes, how HGT helps toxin genes persist until rare ecological opportunities emerge, and how HGT helps to avoid genome deterioration in light of ongoing mutations. Moreover, our work has shown how HGT does not only impact the fitness of organisms, but it can also fundamentally change the outcome of evolutionary processes. I argue that by embracing HGT in our evolutionary theory, we can not only better understand microbial evolution, but may also be able to improve genetic algorithms (Tomko et al., 2013) and help with the future development of evolving technologies.

Evolution with HGT and ongoing mutations

(i.e. bacteria and mobile genetic elements)

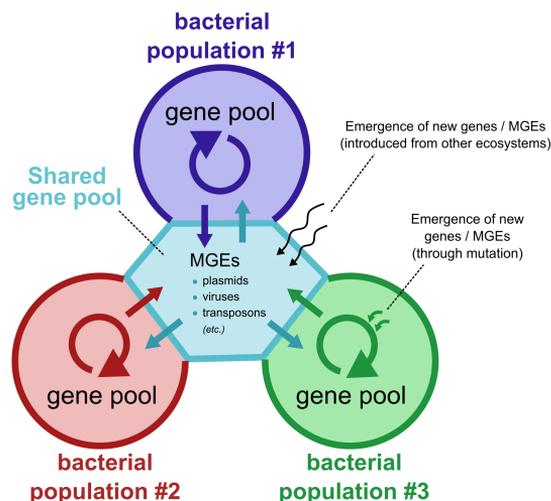


Figure 7.2: HGT opens up gene pools to share genes and mutations, both within and between populations / ecosystems Based on a figure from Nobuto Takeuchi's website <https://unidirectory.auckland.ac.nz/people/nobuto-takeuchi>

7.3.2 What is the impact of nonrandom mutations?

Evolution by means of natural selection is driven by variation and selection. While one could say that selection is by definition a nonrandom process, the processes that give rise to variation (*i.e.* mutations) are often considered to be random for all practical purposes. Although it is well-known that mutations are not random on the chemical level, it remains somewhat taboo to say that mutations can be directed to certain kinds of change. However, theoretical work has provided early proofs of concept that evolution can bias both (i) which genetic mutations are more likely to occur (Crombach and Hogeweg, 2007), or (ii) the phenotypes that are produced by random mutations (Crombach and Hogeweg, 2008), both resulting in quicker adaptation to two alternating environments. In other words, where first-order selection results in survival of the best adapted, second-order selection can give rise to lineages that produce specific genotypes or phenotypes through mutation. This so-called evolution of evolvability has also been shown to play an important role in Virtual Cells (Cuypers et al., 2017), where evolvability was shown to be an easily accessible and viable alternative to evolving phenotypic plasticity. Besides the possibility of improving the rates at which organisms adapt, the evolution of evolution (EvoEvo) can also have other implications for the long-term outcome of evolutionary processes. For example, computational modelling on the evolution of self-replicating RNA molecules has revealed the evolution of a “mutational division of labour”, where specific mutants help with the replication of their parents (Colizzi and Hogeweg, 2014). All the above theoretical studies reveal the importance of either (i) nonrandom mutations, or (ii) the production of nonrandom phenotypes through mutations.

From recent experiments, it has become increasingly clear how many mutations in biological organisms (point mutations, gene duplications and deletions, translocation, HGT) are themselves products of, and shaped by, evolution. For example, locus-specific mutational biases have been observed in *Escherichia coli* (Shewaramani et al., 2017), *Pseudomonas fluorescens* (Lind et al., 2019), and *Saccharomyces cerevisiae* (Zhu et al., 2017), and in some cases these biases are moreover shown to change depending on the environment (Maharjan and Ferenci, 2017; Liu and Zhang, 2019). In *Streptomyces coelicolor*, a specific mutational bias has even been shown to result in specialised mutants that produce costly antibiotics for the colony (Zhang et al., 2019). These experimental observations reveal the possible importance of both the production of non-random mutations, and the production of non-random phenotypes. However, we so far poorly understand whether these examples of mutational biases are idiosyncratic, or whether they are features that have themselves evolved. For example, is the “mutational division of labour” as observed in *Streptomyces* (Zhang et al., 2019) an example of second-order selection? The preceding theoretical work by Colizzi and Hogeweg (2014) demonstrates that, at the very least, this is possible. To further tackle questions like this, we need to better understand the mapping of genotype to phenotype (*i.e.* the GP-MAP), and how this GP-map can change by the evolution of specific genome architectures. Getting a grasp on this will be an exciting challenge for future theoretical and experimental work.

7.3.3 Is evolution predictable?

The crowning achievement of science is often considered to be its ability to predict, informing us on what will happen in the future, as well as what has happened in the past. Predicting the trajectory of evolution is however not a trivial task. Although experimental evolution is often highly repeatable (Flowers et al., 2009; Barrick and Lenski, 2013; Lind et al., 2015), the underlying mechanisms are not always completely understood. Nevertheless, we can now say with increasing confidence that *Escherichia coli* (strain REL606) will evolve a reduced lag-phase in a serial transfer protocol (Lenski et al., 1991), and that *Pseudomonas fluorescence* (strain SBW25) will evolve mat-forming morphotypes in broth microcosms (Rainey and Travisano, 1998). However, what would happen if swap these treatments, use different strains of *E. coli* or *P. fluorescence*, or use entirely different species of microbes? Ideally, if we really understand the underlying causes of how these species adapt to the experimental protocol, we should be able to predict evolution with some level of accuracy. However, the different evolutionary history of these species may constrain or potentiate very particular adaptations, and unless we disentangle how these constraints come about, and how we can recognise them beforehand, our predictions are likely to be of limited accuracy. In other words, where evolution is most likely to occur through “pathways of least resistance” (Schluter, 1996), this is convoluted by the fact that these pathways are themselves products of evolution. We have touched upon the impact of the evolutionary history of species on the predictability of subsequent evolution in Chapter 2, where we found that the degree of predictability was *itself* unpredictable. Some organisms always adapted to a serial transfer protocol in precisely the same way, whereas others evolved mechanistically different outcomes each time the experiment was repeated. Interestingly however, we also saw that on the population level, very generic behaviour emerged in the form “anticipating” the regularity of the resource cycle. This reveals that, besides the impact of prior evolution, the degree of predictability may vary with respect to genotypes, individuals phenotypes, or the evolved collective behaviours.

7.3.4 Concluding remarks

Clearly, there are many challenges in understanding the evolutionary processes that generate and maintain the astonishing diversity of microorganisms. Experimental evolution and bioinformatics have provided us with a treasure trove of data, which we used to design biologically interesting simplifications. This way, we have glimpsed at what life on a single grain of sand may look like. Besides the fact that the resulting “virtual laboratory” is a lot cheaper than an actual laboratory, it also has some other important advantages. Most notably, it ensures that we can study a system one variable at a time, allows us to add or remove important feedback mechanisms, and moreover has the extraordinary ability to surprise the biologist behind the machine. However, as appropriate for any scientific endeavour, many new questions will emerge along the way. In addition to the generation and analysis of new (meta)genomic data, I conclude that the best tool we have at our disposal to further understand microbial evolution is the laboratory, virtual or otherwise.

“THE DEEPEST SOLACE LIES IN UNDERSTANDING,
THIS ANCIENT UNSEEN STREAM,
A SHUDDER BEFORE THE BEAUTIFUL.”

—Tuomas Holopainen

Appendix

References

- Al Naiemi, N., B. Duim, P. H. Savelkoul, L. Spanjaard, E. De Jonge, A. Bart, C. M. Vandembroucke-Grauls, and M. D. de Jong
2005. Widespread transfer of resistance genes between bacterial species in an intensive care unit: implications for hospital epidemiology. *Journal of clinical microbiology*, 43(9):4862–4864.
- Albalat, R. and C. Cañestro
2016. Evolution by gene loss. *Nature Reviews Genetics*, 17(7):379.
- Alessi, A. M., S. M. Bird, N. C. Oates, Y. Li, A. A. Dowle, E. H. Novotny, E. R. deAzevedo, J. P. Bennett, I. Polikarpov, J. P. W. Young, et al.
2018. Defining functional diversity for lignocellulose degradation in a microbial community using multi-omics studies. *Biotechnology for biofuels*, 11(1):166.
- Allison, S. D.
2005. Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecology Letters*, 8(6):626–635.
- Alonso, G., G. Vilchez, and V. Rodríguez Lemoine
2000. How bacteria protect themselves against channel-forming colicins. *International Microbiology*, 3(2):81–88.
- Atamian, H. S., N. M. Creux, E. A. Brown, A. G. Garner, B. K. Blackman, and S. L. Harmer
2016. Circadian regulation of sunflower heliotropism, floral orientation, and pollinator visits. *Science*, 353(6299):587–590.
- Ba, A. N. N., I. Cvijović, J. I. R. Echenique, K. R. Lawrence, A. Rego-Costa, X. Liu, S. F. Levy, and M. M. Desai
2019. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, 575(7783):494–499.
- Bajić, D., J. C. Vila, Z. D. Blount, and A. Sánchez
2018. On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences*, 115(44):11286–11291.
- Balaban, N. Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler
2004. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625.
- Baltrus, D. A.
2013. Exploring the costs of horizontal gene transfer. *Trends in ecology & evolution*, 28(8):489–495.
- Barrick, J. E. and R. E. Lenski
2013. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12):827–839.
- Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim
2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268):1243–1247.
- Bataillon, T.
2000. Estimation of spontaneous genome-wide mutation rate parameters: whither beneficial mutations? *Heredity*, 84(5):497.
- Behringer, M. G., B. I. Choi, S. F. Miller, T. G. Doak, J. A. Karty, W. Guo, and M. Lynch
2018. *Escherichia coli* cultures maintain stable subpopulation structure during long-term evolution. *Proceedings of the National Academy of Sciences*, 115(20):E4642–E4650.
- Beitel, C. W., L. Froenicke, J. M. Lang, I. F. Korf, R. W. Micheltmore, J. A. Eisen, and A. E. Darling
2014. Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. *PeerJ*, 2:e415.
- Bell, G. A.
1997. Experimental evolution in *Chlamydomonas*. i. short-term selection in uniform and diverse environments. *Heredity*, 78(5):490–497.
- Bergé, M. J., C. Mercy, I. Mortier-Barrière, M. S. VanNieuwenhze, Y. V. Brun, C. Grangeasse, P. Polard, and N. Campo
2017. A programmed cell division delay preserves genome integrity during natural genetic transformation in *Streptococcus pneumoniae*. *Nature communications*, 8(1):1–13.
- Bergkessel, M., D. W. Basta, and D. K. Newman
2016. The physiology of growth arrest: uniting molecular and environmental microbiology. *Nat Rev Micro*, 14(9):549–562.

- Bergstrom, C. T., M. Lipsitch, and B. R. Levin
2000. Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics*, 155(4):1505–1519.
- Berke, L. and B. Snel
2015. The plant polycomb repressive complex 1 (prc1) existed in the ancestor of seed plants and has a complex duplication history. *BMC evolutionary biology*, 15(1):44.
- Beslon, G., S. F. Elena, P. Hogeweg, D. Schneider, and S. Stepney
2018. Evolving living technologies—insights from the evoevo project. In *International Symposium on Search Based Software Engineering*, Pp. 46–62. Springer.
- Biller, S. J., F. Schubotz, S. E. Roggensack, A. W. Thompson, R. E. Summons, and S. W. Chisholm
2014. Bacterial vesicles in marine ecosystems. *science*, 343(6167):183–186.
- Blasche, S., Y. Kim, R. Mars, E. Kafkia, M. Maansson, D. Machado, B. Teusink, J. Nielsen, V. Benes, R. Neves, et al.
2019. Emergence of stable coexistence in a complex microbial community through metabolic cooperation and spatio-temporal niche partitioning. *bioRxiv*, P. 541870.
- Blöchl, E., R. Rachel, S. Burggraf, D. Hafenbradl, H. W. Jannasch, and K. O. Stetter
1997. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 c. *Extremophiles*, 1(1):14–21.
- Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski
2012. Genomic analysis of a key innovation in an experimental *escherichia coli* population. *Nature*, 489(7417):513.
- Blundell, J. R. and S. F. Levy
2014. Beyond genome sequencing: lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer. *Genomics*, 104(6):417–430.
- Bruce, J., S. West, and A. Griffin
2017. Bacteriocins and the assembly of natural *pseudomonas fluorescens* populations. *Journal of evolutionary biology*, 30(2):352–360.
- Burton, J. N., I. Liachko, M. J. Dunham, and J. Shendure
2014. Species-level deconvolution of metagenome assemblies with hi-c-based contact probability maps. *G3: Genes, Genomes, Genetics*, 4(7):1339–1346.
- Casacuberta, E. and J. González
2013. The impact of transposable elements in environmental adaptation. *Molecular ecology*, 22(6):1503–1517.
- Chacón, J. M. and W. R. Harcombe
2019. Increasing growth rate slows adaptation when genotypes compete for diffusing resources. *Unpublished data, bioRxiv*, P. 616938.
- Chao, L. and B. R. Levin
1981. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences*, 78(10):6324–6328.
- Choi, S. C., M. D. Rasmussen, M. J. Hubisz, I. Gronau, M. J. Stanhope, and A. Siepel
2012. Replacing and additive horizontal gene transfer in streptococcus. *Molecular biology and evolution*, 29(11):3309–3320.
- Colizzi, E. S. and P. Hogeweg
2014. Evolution of functional diversification within quasispecies. *Genome biology and evolution*, 6(8):1990–2007.
- Colizzi, E. S. and P. Hogeweg
2016. High cost enhances cooperation through the interplay between evolution and self-organisation. *BMC evolutionary biology*, 16(1):31.
- Colizzi, E. S. and P. Hogeweg
2019. Transcriptional mutagenesis prevents ribosomal dna deterioration: The role of duplications and deletions. *Genome biology and evolution*, 11(11):3207–3217.
- Coman, C., B. Drugă, A. Hegedus, C. Sicora, and N. Dragoş
2013. Archaeal and bacterial diversity in two hot spring microbial mats from a geothermal region in romania. *Extremophiles*, 17(3):523–534.
- Consalvey, M., D. M. Paterson, and G. J. Underwood
2004. The ups and downs of life in a benthic biofilm: migration of benthic diatoms. *Diatom Research*, 19(2):181–202.

- Consuegra, J., J. Plucain, J. Gaffé, T. Hindré, and D. Schneider
2017. Genetic basis of exploiting ecological opportunity during the long-term diversification of a bacterial population. *Journal of molecular evolution*, 85(1-2):26–36.
- Cooper, T. F., D. E. Rozen, and R. E. Lenski
2003. Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 100(3):1072–1077.
- Cordero, O. X., L.-A. Ventouras, E. F. DeLong, and M. F. Polz
2012a. Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proceedings of the National Academy of Sciences*, 109(49):20059–20064.
- Cordero, O. X., H. Wildschutte, B. Kirkup, S. Proehl, L. Ngo, F. Hussain, F. Le Roux, T. Mincer, and M. F. Polz
2012b. Ecological populations of bacteria act as socially cohesive units of antibiotic production and resistance. *Science*, 337(6099):1228–1231.
- Costello, E. K., C. L. Lauber, M. Hamady, N. Fierer, J. I. Gordon, and R. Knight
2009. Bacterial community variation in human body habitats across space and time. *Science*, 326(5960):1694–1697.
- Crombach, A. and P. Hogeweg
2007. Chromosome rearrangements and the evolution of genome structuring and adaptability. *Molecular biology and evolution*, 24(5):1130–1139.
- Crombach, A. and P. Hogeweg
2008. Evolution of evolvability in gene regulatory networks. *PLoS computational biology*, 4(7):e1000112.
- Cui, Y., X. Yang, X. Didelot, C. Guo, D. Li, Y. Yan, Y. Zhang, Y. Yuan, H. Yang, J. Wang, et al.
2015. Epidemic clones, oceanic gene pools, and *eco-ld* in the free living marine pathogen *Vibrio parahaemolyticus*. *Molecular biology and evolution*, 32(6):1396–1410.
- Cuypers, T. D. and P. Hogeweg
2012. Virtual genomes in flux: an interplay of neutrality and adaptability explains genome expansion and streamlining. *Genome biology and evolution*, 4(3):212–229.
- Cuypers, T. D. and P. Hogeweg
2014. A synergism between adaptive effects and evolvability drives whole genome duplication to fixation. *PLoS computational biology*, 10(4).
- Cuypers, T. D., J. P. Rutten, and P. Hogeweg
2017. Evolution of evolvability and phenotypic plasticity in virtual cells. *BMC evolutionary biology*, 17(1):60.
- Cvijović, I., A. N. N. Ba, and M. M. Desai
2018. Experimental studies of evolutionary dynamics in microbes. *Trends in Genetics*, 34(9):693–703.
- Czárán, T. L., R. F. Hoekstra, and L. Pagie
2002. Chemical warfare between microbes promotes biodiversity. *Proceedings of the National Academy of Sciences*, 99(2):786–790.
- Dal Co, A., M. Ackermann, and S. van Vliet
2019a. Metabolic activity affects the response of single cells to a nutrient switch in structured populations. *Journal of the Royal Society Interface*, 16(156):20190182.
- Dal Co, A., S. Van Vliet, and M. Ackermann
2019b. Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *Philosophical Transactions of the Royal Society B*, 374(1786):20190080.
- Dal Co, A., S. van Vliet, D. J. Kiviet, S. Schlegel, and M. Ackermann
2020. Short-range interactions govern the dynamics and functions of microbial communities. *Nature Ecology & Evolution*, 4(3):366–375.
- Damkiær, S., L. Yang, S. Molin, and L. Jelsbak
2013. Evolutionary remodeling of global regulatory networks during long-term bacterial adaptation to human hosts. *Proceedings of the National Academy of Sciences*, 110(19):7766–7771.
- Darwin, C.
1859. On the origin of species. *published on*, 24.
- Datta, M. S., E. Sliwerska, J. Gore, M. F. Polz, and O. X. Cordero
2016. Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature communications*, 7(1):1–7.

- de Vries, J. and W. Wackernagel
2002. Integration of foreign dna during natural transformation of acinetobacter sp. by homology-facilitated illegitimate recombination. *Proceedings of the National Academy of Sciences*, 99(4):2094–2099.
- Dekel, E. and U. Alon
2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592.
- Dettman, J. R., C. Sirjusingh, L. M. Kohn, and J. B. Anderson
2007. Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature*, 447(7144):585.
- Dieckmann, U. and M. Doebeli
1999. On the origin of species by sympatric speciation. *Nature*, 400(6742):354.
- Dimitriu, T., F. Medaney, E. Amanatidou, J. Forsyth, R. J. Ellis, and B. Raymond
2019. Negative frequency dependent selection on plasmid carriage and low fitness costs maintain extended spectrum β -lactamases in escherichia coli. *Scientific Reports*, 9(1):1–7.
- Doolittle, W. F. and A. Booth
2017. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biology & Philosophy*, 32(1):5–24.
- Doolittle, W. F. and O. Zhaxybayeva
2009. On the origin of prokaryotic species. *Genome research*, 19(5):744–756.
- Draghi, J. and G. P. Wagner
2009. The evolutionary dynamics of evolvability in a gene network model. *Journal of evolutionary biology*, 22(3):599–611.
- D'Souza, G., S. Shitut, D. Preussger, G. Yousif, S. Waschina, and C. Kost
2018. Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Natural product reports*, 35(5):455–488.
- D'Souza, G., S. Waschina, S. Pande, K. Bohl, C. Kaleta, and C. Kost
2014. Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution*, 68(9):2559–2570.
- Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown, F. Rosenzweig, and D. Botstein
2002. Characteristic genome rearrangements in experimental evolution of saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 99(25):16144–16149.
- Dykes, G. A. and J. W. Hastings
1997. Selection and fitness in bacteriocin-producing bacteria. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1382):683–687.
- Dykhuisen, D. E.
1990. Experimental studies of natural selection in bacteria. *Annual Review of Ecology and Systematics*, 21(1):373–398.
- D'Souza, G. and C. Kost
2016. Experimental evolution of metabolic dependency in bacteria. *PLoS genetics*, 12(11).
- Eigen, M.
1971. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.
- Ephrussi-Taylor, H. and B. A. Freed
1964. Incorporation of thymidine and amino acids into deoxyribonucleic acid and acid-insoluble cell structures in pneumococcal cultures synchronized for competence to transform. *Journal of bacteriology*, 87(5):1211–1215.
- Feazel, L. M., L. K. Baumgartner, K. L. Peterson, D. N. Frank, J. K. Harris, and N. R. Pace
2009. Opportunistic pathogens enriched in showerhead biofilms. *Proceedings of the National Academy of Sciences*, 106(38):16393–16399.
- Finkel, S. E.
2006. Long-term survival during stationary phase: evolution and the gasp phenotype. *Nature Reviews Microbiology*, 4(2):113–120.
- Finkel, S. E. and R. Kolter
2001. Dna as a nutrient: novel role for bacterial competence gene homologs. *Journal of Bacteriology*, 183(21):6288–6293.

- Fischer, S., S. Bernard, G. Beslon, and C. Knibbe
2014. A model for genome size evolution. *Bulletin of mathematical biology*, 76(9):2249–2291.
- Fisher, R.
1930. The genetical theory of natural selection, 1st edn. clarendon.
- Flowers, J. M., Y. Hanzawa, M. C. Hall, R. C. Moore, and M. D. Purugganan
2009. Population genomics of the arabidopsis thaliana flowering time gene network. *Molecular biology and evolution*, 26(11):2475–2486.
- Forsberg, K. J., S. Patel, M. K. Gibson, C. L. Lauber, R. Knight, N. Fierer, and G. Dantas
2014. Bacterial phylogeny structures soil resistomes across habitats. *Nature*, 509(7502):612.
- Fragata, I., A. Blanckaert, M. A. D. Louro, D. A. Liberles, and C. Bank
2018. Evolution in the light of fitness landscape theory. *Trends in ecology & evolution*.
- Fraser, C., E. J. Alm, M. F. Polz, B. G. Spratt, and W. P. Hanage
2009. The bacterial species challenge: making sense of genetic and ecological diversity. *science*, 323(5915):741–746.
- Frazaõ, N., A. Sousa, M. Lässig, and I. Gordo
2019. Horizontal gene transfer overrides mutation in escherichia coli colonizing the mammalian gut. *Proceedings of the National Academy of Sciences*, 116(36):17906–17915.
- Frenkel, E. M., M. J. McDonald, J. D. Van Dyken, K. Kosheleva, G. I. Lang, and M. M. Desai
2015. Crowded growth leads to the spontaneous evolution of semistable coexistence in laboratory yeast populations. *Proceedings of the National Academy of Sciences*, 112(36):11306–11311.
- Fullmer, M. S., S. M. Soucy, and J. P. Gogarten
2015. The pan-genome as a shared genomic resource: mutual cheating, cooperation and the black queen hypothesis. *Frontiers in microbiology*, 6:728.
- Garcia, S. L., M. Buck, K. D. McMahon, H.-P. Grossart, A. Eiler, and F. Warnecke
2015. Auxotrophy and intrapopulation complementary in the ‘interactome’ of a cultivated freshwater model community. *Molecular ecology*, 24(17):4449–4459.
- Gaten, E., G. Tarling, H. Dowse, C. Kyriacou, and E. Rosato
2008. Is vertical migration in antarctic krill (euphausia superba) influenced by an underlying circadian rhythm? *Journal of genetics*, 87(5):473.
- Gause, G.
1934. The struggle for existence. ny.
- Gerardin, Y., M. Springer, and R. Kishony
2016. A competitive trade-off limits the selective advantage of increased antibiotic production. *Nature microbiology*, 1(12):16175.
- Germerodt, S., K. Bohl, A. Lück, S. Pande, A. Schröter, C. Kaleta, S. Schuster, and C. Kost
2016. Pervasive selection for cooperative cross-feeding in bacterial communities. *PLoS computational biology*, 12(6):e1004986.
- Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai
2017. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45.
- Gordo, I. and P. R. Campos
2006. Adaptive evolution in a spatially structured asexual population. *Genetica*, 127(1-3):217.
- Gorter, F. A., M. Manhart, and M. Ackermann
2019. Understanding the evolution of interspecies interactions in microbial communities. *bioRxiv*, P. 770156.
- Govaert, L., E. A. Fronhofer, S. Lion, C. Eizaguirre, D. Bonte, M. Egas, A. P. Hendry, A. D. B. Martins, C. J. Melián, J. A. Raeymaekers, et al.
2018. Eco-evolutionary feedbacks-theoretical models and perspectives. *arXiv preprint arXiv:1806.07633*.
- Granato, E. T., T. A. Meiller-Legrand, and K. R. Foster
2019. The evolution and ecology of bacterial warfare. *Current biology*, 29(11):R521–R537.
- Griffith, F.
1928. The significance of pneumococcal types. *Epidemiology & Infection*, 27(2):113–159.
- Großkopf, T., J. Consuegra, J. Gaffé, J. C. Willison, R. E. Lenski, O. S. Soyer, and D. Schneider
2016. Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC evolutionary biology*, 16(1):163.

- Gu, Z., L. M. Steinmetz, X. Gu, C. Scharfe, R. W. Davis, and W.-H. Li
2003. Role of duplicate genes in genetic robustness against null mutations. *Nature*, 421(6918):63–66.
- Gudelj, I., R. E. Beardmore, S. Arkin, and R. C. MacLean
2007. Constraints on microbial metabolism drive evolutionary diversification in homogeneous environments. *Journal of evolutionary biology*, 20(5):1882–1889.
- Habets, M. G., T. Czarán, R. F. Hoekstra, and J. A. G. De Visser
2007. Spatial structure inhibits the rate of invasion of beneficial mutations in asexual populations. *Proceedings of the Royal Society B: Biological Sciences*, 274(1622):2139–2143.
- Hacker, J. and E. Carniel
2001. Ecological fitness, genomic islands and bacterial pathogenicity: a darwinian view of the evolution of microbes. *EMBO reports*, 2(5):376–381.
- Haldane, J.
1932. The causes of evolution. new york and london.
- Hansen, S. K., P. B. Rainey, J. A. Haagensen, and S. Molin
2007. Evolution of species interactions in a biofilm community. *Nature*, 445(7127):533–536.
- Hao, W. and G. B. Golding
2006. The fate of laterally transferred genes: life in the fast lane to adaptation or death. *Genome research*, 16(5):636–643.
- Hardin, G.
1960. The competitive exclusion principle. *science*, 131(3409):1292–1297.
- Harrison, E., D. Guymer, A. J. Spiers, S. Paterson, and M. A. Brockhurst
2015. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. *Current Biology*, 25(15):2034–2039.
- Harrison, E., R. MacLean, V. Koufopanou, and A. Burt
2014. Sex drives intracellular conflict in yeast. *Journal of evolutionary biology*, 27(8):1757–1763.
- Hehemann, J.-H., P. Arevalo, M. S. Datta, X. Yu, C. H. Corzett, A. Henschel, S. P. Preheim, S. Timberlake, E. J. Alm, and M. F. Polz
2016. Adaptive radiation by waves of gene transfer leads to fine-scale resource partitioning in marine microbes. *Nature communications*, 7:12860.
- Helling, R. B., C. N. Vargas, and J. Adams
1987. Evolution of *escherichia coli* during growth in a constant environment. *Genetics*, 116(3):349–358.
- Herdman, M.
1985. The evolution of bacterial genomes. *The Evolution of Genome Size*, Pp. 37–68.
- Herrerías-Azcué, F., V. Pérez-Muñuzuri, and T. Galla
2018. Stirring does not make populations well mixed. *Scientific reports*, 8(1):1–9.
- Herron, M. D. and M. Doebeli
2013. Parallel evolutionary dynamics of adaptive diversification in *escherichia coli*. *PLoS biology*, 11(2).
- Heuer, H., H. Schmitt, and K. Smalla
2011. Antibiotic resistance gene spread due to manure application on agricultural fields. *Current opinion in microbiology*, 14(3):236–243.
- Higgins, D. and J. Dworkin
2012. Recent progress in *bacillus subtilis* sporulation. *FEMS microbiology reviews*, 36(1):131–148.
- Hindré, T., C. Knibbe, G. Beslon, and D. Schneider
2012. New insights into bacterial adaptation through in vivo and in silico experimental evolution. *Nature Reviews Microbiology*, 10(5):352.
- Hogeweg, P.
2007. From population dynamics to ecoinformatics: Ecosystems as multilevel information processing systems. *Ecological Informatics*, 2(2):103–111.
- Hogeweg, P.
2012. Toward a theory of multilevel evolution: long-term information integration shapes the mutational landscape and enhances evolvability. In *Evolutionary Systems Biology*, Pp. 195–224. Springer.

- Hogeweg, P. and B. Hesper
1991. Evolution as pattern processing: Todo as substrate for evolution. *From Animals to Animats*, Pp. 492–497.
- Holt, R. D.
2001. Species coexistence. *Encyclopedia of biodiversity*, 5:413–426.
- Ibanez, S.
2020. The evolution of ecosystem phenotypes. *Biological Theory*, Pp. 1–16.
- Inc., P. T.
2015. Collaborative data science.
- Iranzo, J., Y. I. Wolf, E. V. Koonin, and I. Sela
2019. Gene gain and loss push prokaryotes beyond the homologous recombination barrier and accelerate genome sequence divergence. *Nature Communications*, 10(1):1–10.
- Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake
2003. Horizontal gene transfer accelerates genome innovation and evolution. *Molecular biology and evolution*, 20(10):1598–1602.
- James, R., C. Kleanthous, and G. R. Moore
1996. The biology of e colicins: paradigms and paradoxes. *Microbiology*, 142(7):1569–1580.
- Job, Y. F. P.
. Gnu parallel: The command-line power tool.
- Johnson, D. R., F. Goldschmidt, E. E. Lilja, and M. Ackermann
2012. Metabolic specialization and the assembly of microbial communities. *The ISME journal*, 6(11):1985–1991.
- Johnston, C., B. Martin, C. Granadel, P. Polard, and J.-P. Claverys
2013. Programmed protection of foreign dna from restriction allows pathogenicity island exchange during pneumococcal transformation. *PLoS pathogens*, 9(2):e1003178.
- Jones, B. E., W. D. Grant, A. W. Duckworth, and G. G. Owenson
1998. Microbial diversity of soda lakes. *Extremophiles*, 2(3):191–200.
- Josephides, C. and P. S. Swain
2017. Predicting metabolic adaptation from networks of mutational paths. *Nature communications*, 8(1):1–15.
- Kim, J.-N., Y. Kim, Y. Jeong, J.-H. Roe, B.-G. Kim, and B.-K. Cho
2015. Comparative genomics reveals the core and accessory genomes of streptomycetes species. *J Microbiol Biotechnol*, 25(10):1599–605.
- Kinkel, L. L., D. C. Schlatter, K. Xiao, and A. D. Baines
2014. Sympatric inhibition and niche differentiation suggest alternative coevolutionary trajectories among streptomycetes. *The ISME journal*, 8(2):249–256.
- Kinnersley, M., K. Schwartz, J. Boswell, D.-D. Yang, G. Sherlock, and F. Rosenzweig
2019. Evolutionary dynamics of de novo mutations and mutant lineages arising in a simple, constant environment. *BioRxiv*, P. 540625.
- Kinnersley, M., J. Wenger, E. Kroll, J. Adams, G. Sherlock, and F. Rosenzweig
2014. Ex uno plures: clonal reinforcement drives evolution of a simple microbial community. *PLoS genetics*, 10(6).
- Kinnersley, M. A., W. E. Holben, and F. Rosenzweig
2009. E unibus plurum: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. *PLoS genetics*, 5(11).
- Kolstø, A.-B.
1997. Dynamic bacterial genome organization. *Molecular microbiology*, 24(2):241–248.
- Koonin, E. V. and Y. I. Wolf
2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic acids research*, 36(21):6688–6719.
- Korona, R., C. H. Nakatsu, L. J. Forney, and R. E. Lenski
1994. Evidence for multiple adaptive peaks from populations of bacteria evolving in a structured habitat. *Proceedings of the National Academy of Sciences*, 91(19):9037–9041.

- Kotil, S. E. and K. Vetsigian
2018. Emergence of evolutionary stable communities through eco-evolutionary tunneling. *bioRxiv*, P. 271015.
- Kuo, C.-H., N. A. Moran, and H. Ochman
2009. The consequences of genetic drift for bacterial genome complexity. *Genome research*, 19(8):1450–1454.
- Kuo, C.-H. and H. Ochman
2009. Deletional bias across the three domains of life. *Genome biology and evolution*, 1:145–152.
- Kwong, W. K., P. Engel, H. Koch, and N. A. Moran
2014. Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proceedings of the National Academy of Sciences*, 111(31):11509–11514.
- Laan, L., J. H. Koschwanez, and A. W. Murray
2015. Evolutionary adaptation after crippling cell polarization follows reproducible trajectories. *Elife*, 4:e09638.
- Lang, A. S., O. Zhaxybayeva, and J. T. Beatty
2012. Gene transfer agents: phage-like elements of genetic exchange. *Nature Reviews Microbiology*, 10(7):472–482.
- Le Gac, M., M. D. Brazas, M. Bertrand, J. G. Tyerman, C. C. Spencer, R. E. Hancock, and M. Doebeli
2008. Metabolic changes associated with adaptive diversification in *escherichia coli*. *Genetics*, 178(2):1049–1060.
- Le Gac, M., J. Plucain, T. Hindré, R. E. Lenski, and D. Schneider
2012. Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *escherichia coli*. *Proceedings of the National Academy of Sciences*, 109(24):9487–9492.
- Lefébure, T. and M. J. Stanhope
2007. Evolution of the core and pan-genome of *streptococcus*: positive selection, recombination, and genome composition. *Genome biology*, 8(5):R71.
- Lehman, J., J. Clune, D. Misevic, C. Adami, L. Altenberg, J. Beaulieu, P. J. Bentley, S. Bernard, G. Beslon, D. M. Bryson, et al.
2018. The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities. *arXiv preprint arXiv:1803.03453*.
- Lenski, R. E., M. R. Rose, S. C. Simpson, and S. C. Tadler
1991. Long-term experimental evolution in *escherichia coli*. i. adaptation and divergence during 2,000 generations. *The American Naturalist*, 138(6):1315–1341.
- Lenski, R. E. and M. Travisano
1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences*, 91(15):6808–6814.
- Lenski, R. E., M. J. Wiser, N. Ribbeck, Z. D. Blount, J. R. Nahum, J. J. Morris, L. Zaman, C. B. Turner, B. D. Wade, R. Maddamsetti, et al.
2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *escherichia coli*. *Proc. R. Soc. B*, 282(1821):20152292.
- Lepplae, R., D. Geeraerts, R. Hallez, J. Guglielmini, P. Dreze, and L. Van Melderen
2011. Diversity of bacterial type ii toxin–antitoxin systems: a comprehensive search and functional analysis of novel families. *Nucleic acids research*, 39(13):5513–5525.
- Lerner, J., M. Manhart, W. Jasinska, L. Gauthier, A. W. Serohijos, and S. Bershtein
2019. Chromosomal barcoding of *e. coli* populations reveals lineage diversity dynamics at high resolution. *bioRxiv*, P. 571505.
- Levy, S. F., J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock
2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181.
- Lewis, K.
2007. Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology*, 5(1):48–56.
- Li, F., M. L. Salit, and S. F. Levy
2018. Unbiased fitness estimation of pooled barcode or amplicon sequencing studies. *Cell systems*, 7(5):521–525.
- Li, Y., D. A. Petrov, and G. Sherlock
2019. Single nucleotide mapping of the locally accessible trait space in yeast reveals pareto fronts that constrain initial adaptation. *Unpublished data, bioRxiv*, P. 593947.

- Lili, L. N., N. F. Britton, and E. J. Feil
2007. The persistence of parasitic plasmids. *Genetics*, 177(1):399–405.
- Lilja, E. E. and D. R. Johnson
2019. Substrate cross-feeding affects the speed and trajectory of molecular evolution within a synthetic microbial assemblage. *BMC evolutionary biology*, 19(1):129.
- Lind, P. A., A. D. Farr, and P. B. Rainey
2015. Experimental evolution reveals hidden diversity in evolutionary pathways. *Elife*, 4:e07074.
- Lind, P. A., E. Libby, J. Herzog, and P. B. Rainey
2019. Predicting mutational routes to new adaptive phenotypes. *Elife*, 8:e38822.
- Liu, H. and J. Zhang
2019. Yeast spontaneous mutation rate and spectrum vary with environment. *Current Biology*, 29(10):1584–1591.
- Lopatkin, A. J., S. Huang, R. P. Smith, J. K. Srimani, T. A. Syssoeva, S. Bewick, D. K. Karig, and L. You
2016. Antibiotics as a selective driver for conjugation dynamics. *Nature microbiology*, 1(6):16044.
- Luell, S. M.
. The movements of fish and zooplankton at sunset in willand pond, new hampshire.
- Lukjancenko, O., T. M. Wassenaar, and D. W. Ussery
2010. Comparison of 61 sequenced escherichia coli genomes. *Microbial ecology*, 60(4):708–720.
- Lynch, M., M. S. Ackerman, J.-F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster
2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics*, 17(11):704.
- Lynch, M., J. Conery, and R. Bürger
1995. Mutational meltdowns in sexual populations. *Evolution*, 49(6):1067–1080.
- MacArthur, R.
1970. Species packing and competitive equilibrium for many species. *Theoretical population biology*, 1(1):1–11.
- Macfadyen, L. P., D. Chen, H. C. Vo, D. Liao, R. Sinotte, and R. J. Redfield
2001. Competence development by haemophilus influenzae is regulated by the availability of nucleic acid precursors. *Molecular microbiology*, 40(3):700–707.
- Machado, D., O. M. Maistrenko, S. Andrejev, Y. Kim, P. Bork, K. R. Patil, and K. R. Patil
2020. Polarization of microbial communities between competitive and cooperative metabolism. *bioRxiv*.
- Madsen, J. S., M. Burmølle, L. H. Hansen, and S. J. Sørensen
2012. The interconnection between biofilm formation and horizontal gene transfer. *FEMS Immunology & Medical Microbiology*, 65(2):183–195.
- Maharjan, R. P. and T. Ferenci
2017. A shifting mutational landscape in 6 nutritional states: Stress-induced mutagenesis as a series of distinct stress input–mutation output relationships. *PLoS biology*, 15(6):e2001477.
- Manhart, M., B. V. Adkar, and E. I. Shakhnovich
2018. Trade-offs between microbial growth phases lead to frequency-dependent and non-transitive selection. *Proc. R. Soc. B*, 285(1872):20172459.
- Manhart, M. and E. Shakhnovich
2018. Growth tradeoffs produce complex microbial communities on a single limiting resource. *bioRxiv*, P. 266569.
- Marchal, M., F. Goldschmidt, S. N. Derksen-Müller, S. Panke, M. Ackermann, and D. R. Johnson
2017. A passive mutualistic interaction promotes the evolution of spatial structure within microbial populations. *BMC evolutionary biology*, 17(1):106.
- Martinez, J. L.
2009. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proceedings of the Royal Society B: Biological Sciences*, 276(1667):2521–2530.
- Matic, I.
2016. Molecular mechanisms involved in the regulation of mutation rates in bacteria. *Periodicum biologorum*, 118(4):363–372.

- Mc Ginty, S. E., D. J. Rankin, and S. P. Brown
2011. Horizontal gene transfer and the evolution of bacterial cooperation. *Evolution: International Journal of Organic Evolution*, 65(1):21–32.
- McInerney, J. O., A. McNally, and M. J. O’connell
2017. Why prokaryotes have pangenomes. *Nature microbiology*, 2(4):1–5.
- Meijer, J., B. van Dijk, and P. Hogeweg
2019. Evolutionary contingencies drive emergent ecological interactions between microbes in a constant, one-niche environment (in preparation).
- Mell, J. C. and R. J. Redfield
2014. Natural competence and the evolution of dna uptake specificity. *Journal of bacteriology*, 196(8):1471–1483.
- Metz, J. A., S. A. Geritz, G. Meszéna, F. J. Jacobs, and J. S. Van Heerwaarden
1995. Adaptive dynamics: a geometrical study of the consequences of nearly faithful reproduction.
- Millard, P., K. Smallbone, and P. Mendes
2017. Metabolic regulation is sufficient for global and robust coordination of glucose uptake, catabolism, energy production and growth in escherichia coli. *PLoS computational biology*, 13(2):e1005396.
- Mitchell, A. and Y. Pilpel
2011. A mathematical model for adaptive prediction of environmental changes by microorganisms. *Proceedings of the National Academy of Sciences*, 108(17):7271–7276.
- Mitchell, A., G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel
2009. Adaptive prediction of environmental changes by microorganisms. *Nature*, 460(7252):220.
- Morris, J. J., R. E. Lenski, and E. R. Zinser
2012. The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio*, 3(2):e00036–12.
- Mozhayskiy, V. and I. Tagkopoulos
2012. Horizontal gene transfer dynamics and distribution of fitness effects during microbial in silico evolution. In *BMC bioinformatics*, volume 13, P. S13. Springer.
- Muller, H. J.
1964. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 1(1):2–9.
- Nadell, C. D., K. Drescher, and K. R. Foster
2016. Spatial structure, cooperation and competition in biofilms. *Nature Reviews Microbiology*, 14(9):589.
- Nannipieri, P., J. Ascher, M. Ceccherini, L. Landi, G. Pietramellara, and G. Renella
2003. Microbial diversity and soil functions. *European journal of soil science*, 54(4):655–670.
- New, A. M., B. Cerulus, S. K. Govers, G. Perez-Samper, B. Zhu, S. Boogmans, J. B. Xavier, and K. J. Verstrepen
2014. Different levels of catabolite repression optimize growth in stable and variable environments. *PLoS biology*, 12(1):e1001764.
- Neyfakh, A. A., N. N. Baranova, and L. J. Mizrokhi
2006. A system for studying evolution of life-like virtual organisms. *Biology Direct*, 1(1):23.
- Niehus, R., S. Mitri, A. G. Fletcher, and K. R. Foster
2015. Migration and horizontal gene transfer divide microbial genomes into multiple niches. *Nature communications*, 6:8924.
- Nogueira, T., D. J. Rankin, M. Touchon, F. Taddei, S. P. Brown, and E. P. Rocha
2009. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. *Current Biology*, 19(20):1683–1691.
- Norman, A., L. H. Hansen, and S. J. Sørensen
2009. Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1527):2275–2289.
- Novak, M., T. Pfeiffer, R. E. Lenski, U. Sauer, and S. Bonhoeffer
2006. Experimental tests for an evolutionary trade-off between growth rate and yield in e. coli. *The American Naturalist*, 168(2):242–251.

- Novick, A. and W. F. Doolittle
2020. Horizontal persistence and the complexity hypothesis. *Biology & Philosophy*, 35(1):2.
- Nowack, E. C., D. C. Price, D. Bhattacharya, A. Singer, M. Melkonian, and A. R. Grossman
2016. Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of paulinella chromatophora. *Proceedings of the National Academy of Sciences*, 113(43):12214–12219.
- Nowell, R. W., S. Green, B. E. Laue, and P. M. Sharp
2014. The extent of genome flux and its role in the differentiation of bacterial lineages. *Genome biology and evolution*, 6(6):1514–1529.
- Ochman, H., J. G. Lawrence, and E. A. Groisman
2000. Lateral gene transfer and the nature of bacterial innovation. *nature*, 405(6784):299.
- Overballe-Petersen, S. and E. Willerslev
2014. Horizontal transfer of short and degraded dna has evolutionary implications for microbes and eukaryotic sexual reproduction. *Bioessays*, 36(10):1005–1010.
- Pagie, L. and P. Hogeweg
1999. Colicin diversity: a result of eco-evolutionary dynamics. *Journal of Theoretical Biology*, 196(2):251–261.
- Pamilo, P., M. Nei, and W.-H. Li
1987. Accumulation of mutations in sexual and asexual populations. *Genetics Research*, 49(2):135–146.
- Pande, S., F. Kaftan, S. Lang, A. Svatoš, S. Germerodt, and C. Kost
2016. Privatization of cooperative benefits stabilizes mutualistic cross-feeding interactions in spatially structured environments. *The ISME journal*, 10(6):1413–1423.
- Pande, S. and C. Kost
2017. Bacterial unculturability and the formation of intercellular metabolic networks. *Trends in microbiology*, 25(5):349–361.
- Pande, S., H. Merker, K. Bohl, M. Reichelt, S. Schuster, L. F. De Figueiredo, C. Kaleta, and C. Kost
2014. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *The ISME journal*, 8(5):953–962.
- Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot
1999. Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences*, 96(7):3807–3812.
- Paterson, S., T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, et al.
2010. Antagonistic coevolution accelerates molecular evolution. *Nature*, 464(7286):275.
- Pedrós-Alió, C., M. Potvin, and C. Lovejoy
2015. Diversity of planktonic microorganisms in the arctic ocean. *Progress in oceanography*, 139:233–243.
- Pelosi, L., L. Kühn, D. Guetta, J. Garin, J. Geiselmann, R. E. Lenski, and D. Schneider
2006. Parallel changes in global protein profiles during long-term experimental evolution in escherichia coli. *Genetics*, 173(4):1851–1869.
- Pfeiffer, T. and S. Bonhoeffer
2004. Evolution of cross-feeding in microbial populations. *The American naturalist*, 163(6):E126–E135.
- Pfeiffer, T., S. Schuster, and S. Bonhoeffer
2001. Cooperation and competition in the evolution of atp-producing pathways. *Science*, 292(5516):504–507.
- Philippe, N., E. Crozat, R. E. Lenski, and D. Schneider
2007. Evolution of global regulatory networks during a long-term experiment with escherichia coli. *Bioessays*, 29(9):846–860.
- Pianka, E. R.
1970. On r-and k-selection. *The american naturalist*, 104(940):592–597.
- Piggot, P. J. and D. W. Hilbert
2004. Sporulation of bacillus subtilis. *Current opinion in microbiology*, 7(6):579–586.

- Plucain, J., T. Hindré, M. Le Gac, O. Tenaillon, S. Cruveiller, C. Médigue, N. Leiby, W. R. Harcombe, C. J. Marx, R. E. Lenski, et al.
2014. Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science*, P. 1242862.
- Preheim, S. P., S. Timberlake, and M. F. Polz
2011. Merging taxonomy with ecological population prediction in a case study of vibronaceae. *Appl. Environ. Microbiol.*, 77(20):7195–7206.
- Puigbò, P., A. E. Lobkovsky, D. M. Kristensen, Y. I. Wolf, and E. V. Koonin
2014. Genomes in turmoil: quantification of genome dynamics in prokaryote supergenomes. *BMC biology*, 12(1):66.
- Quistad, S. D., G. Doulier, and P. B. Rainey
2019. Experimental manipulation of selfish genetic elements links genes to microbial community function. *BioRxiv*, P. 608752.
- Rainey, P. B. and M. Travisano
1998. Adaptive radiation in a heterogeneous environment. *Nature*, 394(6688):69.
- Raman, K. and A. Wagner
2011. Evolvability and robustness in a complex signalling circuit. *Molecular BioSystems*, 7(4):1081–1092.
- Ranjard, L. and A. Richaume
2001. Quantitative and qualitative microscale distribution of bacteria in soil. *Research in microbiology*, 152(8):707–716.
- Rankin, D. J., E. P. Rocha, and S. P. Brown
2011. What traits are carried on mobile genetic elements, and why? *Heredity*, 106(1):1.
- Ratcliff, W. C., J. D. Fankhauser, D. W. Rogers, D. Greig, and M. Travisano
2015. Origins of multicellular evolvability in snowflake yeast. *Nature communications*, 6:6102.
- Redfield, R. J.
1988. Evolution of bacterial transformation: is sex with dead cells ever better than no sex at all? *Genetics*, 119(1):213–221.
- Redfield, R. J.
2001. Do bacteria have sex? *Nature Reviews Genetics*, 2(8):634.
- Reichenbach, T., M. Mobilia, and E. Frey
2007. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. *Nature*, 448(7157):1046.
- Riley, M. A.
1993. Molecular mechanisms of colicin evolution. *Molecular biology and evolution*, 10(6):1380–1395.
- Riley, M. A.
1998. Molecular mechanisms of bacteriocin evolution. *Annual review of genetics*, 32(1):255–278.
- Riley, M. A. and J. E. Wertz
2002. Bacteriocins: evolution, ecology, and application. *Annual Reviews in Microbiology*, 56(1):117–137.
- Ritz, K.
1998. Molecular approaches to the analysis of soil microbial community structure and function (environment and soil microorganisms, symposium). *Soil Microorganisms*, 52:7–21.
- Rocabert, C., C. Knibbe, J. Consuegra, D. Schneider, and G. Beslon
2017. Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification. *PLoS computational biology*, 13(3):e1005459.
- Roesch, L. F., R. R. Fulthorpe, A. Riva, G. Casella, A. K. Hadwin, A. D. Kent, S. H. Daroub, F. A. Camargo, W. G. Farmerie, and E. W. Triplett
2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, 1(4):283.
- Roller, B. R., S. F. Stoddard, and T. M. Schmidt
2016. Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nature microbiology*, 1(11):1–7.
- Rosenthal, A. Z., E. G. Matson, A. Eldar, and J. R. Leadbetter
2011. RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *The ISME journal*, 5(7):1133–1142.

- Rosenzweig, R. F., R. Sharp, D. S. Treves, and J. Adams
1994. Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics*, 137(4):903–917.
- Rozen, D. E. and R. E. Lenski
2000. Long-term experimental evolution in *Escherichia coli*. viii. dynamics of a balanced polymorphism. *The American Naturalist*, 155(1):24–35.
- Rozen, D. E., N. Philippe, J. Arjan de Visser, R. E. Lenski, and D. Schneider
2009. Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters*, 12(1):34–44.
- Rozen, D. E., D. Schneider, and R. E. Lenski
2005. Long-term experimental evolution in *Escherichia coli*. xiii. phylogenetic history of a balanced polymorphism. *Journal of Molecular Evolution*, 61(2):171–180.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, et al.
2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biology*, 5(3):e77.
- Rutten, J. P., P. Hogeweg, and G. Beslon
2019. Adapting the engine to the fuel: mutator populations can reduce the mutational load by reorganizing their genome structure. *BMC Evolutionary Biology*, 19(1):191.
- Sakoparnig, T., C. Field, and E. van Nimwegen
2019. Whole genome phylogenies reflect long-tailed distributions of recombination rates in many bacterial species. *bioRxiv*, P. 601914.
- Salverda, M. L., J. Koomen, B. Koopmanschap, M. P. Zwart, and J. A. G. de Visser
2017. Adaptive benefits from small mutation supplies in an antibiotic resistance enzyme. *Proceedings of the National Academy of Sciences*, P. 201712999.
- San Millan, A. and R. Craig Maclean
2019. Fitness costs of plasmids: a limit to plasmid transmission. *Microbial Transmission*, Pp. 65–79.
- San Roman, M. and A. Wagner
2018. An enormous potential for niche construction through bacterial cross-feeding in a homogeneous environment. *PLoS Computational Biology*, 14(7):e1006340.
- Sarubbi, E., K. Rudd, H. Xiao, K. Ikehara, M. Kalman, and M. Cashel
1989. Characterization of the *spot* gene of *Escherichia coli*. *Journal of Biological Chemistry*, 264(25):15074–15082.
- Schluter, D.
1996. Adaptive radiation along genetic lines of least resistance. *Evolution*, 50(5):1766–1774.
- Schultz, D., P. G. Wolyne, E. B. Jacob, and J. N. Onuchic
2009. Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*, 106(50):21027–21034.
- Schultz, S. T. and M. Lynch
1997. Mutation and extinction: the role of variable mutational effects, synergistic epistasis, beneficial mutations, and degree of outcrossing. *Evolution*, 51(5):1363–1371.
- Seitz, P. and M. Blokesch
2013. Cues and regulatory pathways involved in natural competence and transformation in pathogenic and environmental gram-negative bacteria. *FEMS Microbiology Reviews*, 37(3):336–363.
- Shah, D., Z. Zhang, A. B. Khodursky, N. Kaldalu, K. Kurg, and K. Lewis
2006. Persister: a distinct physiological state of *E. coli*. *BMC Microbiology*, 6(1):53.
- Shapiro, B. J., J. Friedman, O. X. Cordero, S. P. Preheim, S. C. Timberlake, G. Szabó, M. F. Polz, and E. J. Alm
2012. Population genomics of early events in the ecological differentiation of bacteria. *Science*, 336(6077):48–51.
- Shewaramani, S., T. J. Finn, S. C. Leahy, R. Kassen, P. B. Rainey, and C. D. Moon
2017. Anaerobically grown *Escherichia coli* has an enhanced mutation rate and distinct mutational spectra. *PLoS Genetics*, 13(1):e1006570.
- Siegal, M. L.
2015. Shifting sugars and shifting paradigms. *PLoS Biology*, 13(2):e1002068.

- Slater, F. R., M. J. Bailey, A. J. Tett, and S. L. Turner
2008. Progress towards understanding the fate of plasmids in bacterial communities. *FEMS Microbiology Ecology*, 66(1):3–13.
- Smith, H. L.
2011. Bacterial competition in serial transfer culture. *Mathematical biosciences*, 229(2):149–159.
- Snel, B., P. Bork, and M. A. Huynen
2002. Genomes in flux: the evolution of archaeal and proteobacterial gene content. *Genome research*, 12(1):17–25.
- Sniegowski, P. D. and P. J. Gerrish
2010. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1544):1255–1263.
- Soetaert, K., T. Petzoldt, and R. W. Setzer
2010. Solving differential equations in R: Package deSolve. *Journal of Statistical Software*, 33(9):1–25.
- Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl
2006. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proceedings of the National Academy of Sciences*, 103(32):12115–12120.
- Soler, N., M. Krupovic, E. Marguet, and P. Forterre
2015. Membrane vesicles in natural environments: a major challenge in viral ecology. *The ISME journal*, 9(4):793–796.
- Solopova, A., J. van Gestel, F. J. Weissing, H. Bachmann, B. Teusink, J. Kok, and O. P. Kuipers
2014. Bet-hedging during bacterial diauxic shift. *Proceedings of the National Academy of Sciences*, 111(20):7427–7432.
- Sommer, U.
1985. Comparison between steady state and non-steady state competition: experiments with natural phytoplankton. *Limnology and Oceanography*, 30(2):335–346.
- Sorokin, D. Y., T. Berben, E. D. Melton, L. Overmars, C. D. Vavourakis, and G. Muyzer
2014. Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles*, 18(5):791–809.
- Soucy, S. M., J. Huang, and J. P. Gogarten
2015. Horizontal gene transfer: building the web of life. *Nature Reviews Genetics*, 16(8):472–482.
- Stalder, T., B. Cornwell, J. Lacroix, B. Kohler, S. Dixon, H. Yano, B. Kerr, L. J. Forney, and E. M. Top
2020. Evolving populations in biofilms contain more persistent plasmids. *Molecular Biology and Evolution*.
- Staley, J. T. and J. J. Gosink
1999. Poles apart: biodiversity and biogeography of sea ice bacteria. *Annual reviews in Microbiology*, 53(1):189–215.
- Stanton, T. B.
2007. Prophage-like gene transfer agents—novel mechanisms of gene exchange for methanococcus, desulfovibrio, brachyspira, and rhodobacter species. *Anaerobe*, 13(2):43–49.
- Steven, B., R. L evell e, W. H. Pollard, and L. G. Whyte
2006. Microbial ecology and biodiversity in permafrost. *Extremophiles*, 10(4):259–267.
- Stump, S. M., E. C. Johnson, and C. A. Klausmeier
2018a. Local interactions and self-organized spatial patterns stabilize microbial cross-feeding against cheaters. *Journal of the Royal Society Interface*, 15(140):20170822.
- Stump, S. M., E. C. Johnson, Z. Sun, and C. A. Klausmeier
2018b. How spatial structure and neighbor uncertainty promote mutualists and weaken black queen effects. *Journal of theoretical biology*, 446:33–60.
- Tagkopoulos, I., Y.-C. Liu, and S. Tavazoie
2008. Predictive behavior within microbial genetic networks. *science*, 320(5881):1313–1317.
- Takeuchi, N., O. X. Cordero, E. V. Koonin, and K. Kaneko
2015. Gene-specific selective sweeps in bacteria and archaea caused by negative frequency-dependent selection. *BMC biology*, 13(1):20.

- Takeuchi, N. and P. Hogeweg
2009. On the degree of freedom in multilevel evolutionary models. *Proceedings: Levels of Selection and Individuality in Evolution: Conceptual Issues and the Role of Artificial Life Models*, P. 35.
- Takeuchi, N., K. Kaneko, and E. V. Koonin
2014. Horizontal gene transfer can rescue prokaryotes from muller's ratchet: benefit of dna from dead cells and population subdivision. *G3: Genes, Genomes, Genetics*, 4(2):325–339.
- Takeuchi, N., P. H. Poorthuis, and P. Hogeweg
2005. Phenotypic error threshold; additivity and epistasis in rna evolution. *BMC Evolutionary Biology*, 5(1):9.
- Tashiro, Y., Y. Hasegawa, M. Shintani, K. Takaki, M. Ohkuma, K. Kimbara, and H. Futamata
2017. Interaction of bacterial membrane vesicles with specific species and their potential for delivery to target cells. *Frontiers in microbiology*, 8:571.
- Thomas, C. M. and K. M. Nielsen
2005. Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature reviews microbiology*, 3(9):711.
- Thommes, M., T. Wang, Q. Zhao, I. C. Paschalidis, and D. Segrè
2019. Designing metabolic division of labor in microbial communities. *mSystems*, 4(2):e00263–18.
- Tidjani, A.-R., J.-N. Lorenzi, M. Toussaint, E. van Dijk, D. Naquin, O. Lespinet, C. Bontemps, and P. Leblond
2019. Massive gene flux drives genome diversity between sympatric streptomyces conspecifics. *mBio*, 10(5):e01533–19.
- Tilman, D.
1977. Resource competition between plankton algae: an experimental and theoretical approach. *Ecology*, 58(2):338–348.
- Tilman, D.
1982. *Resource competition and community structure*. Princeton university press.
- Toffoli, T. and N. Margolus
1987. Cellular automata machines: a new environment for modeling (scientific computation). In *Press Cambridge*, P. 259.
- Tomko, N., I. Harvey, and A. Philippides
2013. Unconstrain the population: The benefits of horizontal gene transfer in genetic algorithms. In *SmartData*, Pp. 117–127. Springer.
- Torsvik, V.
1994. Use of dna analysis to determine the diversity of microbial communities. *Beyond the biomass: compositional and functional analysis of soil microbial communities*, Pp. 39–48.
- Torsvik, V. and L. Øvreås
2002. Microbial diversity and function in soil: from genes to ecosystems. *Current opinion in microbiology*, 5(3):240–245.
- Touchon, M., C. Hoede, O. Tenaillon, V. Barbe, S. Baeriswyl, P. Bidet, E. Bingen, S. Bonacorsi, C. Bouchier, O. Bouvet, et al.
2009. Organised genome dynamics in the escherichia coli species results in highly diverse adaptive paths. *PLoS genetics*, 5(1):e1000344.
- Treangen, T. J. and E. P. Rocha
2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS genetics*, 7(1):e1001284.
- Treves, D. S., S. Manning, and J. Adams
1998. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of escherichia coli. *Molecular biology and evolution*, 15(7):789–797.
- Tromer, E. C., J. J. van Hooff, G. J. Kops, and B. Snel
2019. Mosaic origin of the eukaryotic kinetochore. *Proceedings of the National Academy of Sciences*, 116(26):12873–12882.
- Tsoi, R., F. Wu, C. Zhang, S. Bewick, D. Karig, and L. You
2018. Metabolic division of labor in microbial systems. *Proceedings of the National Academy of Sciences*, 115(10):2526–2531.
- Turner, C. B., Z. D. Blount, D. H. Mitchell, and R. E. Lenski
2015. Evolution and coexistence in response to a key innovation in a long-term evolution experiment with escherichia coli. *bioRxiv*, P. 020958.

- Turner, P. E., V. Souza, and R. E. Lenski
1996. Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology*, 77(7):2119–2129.
- van Dam, T. J., F. J. Zwartkruis, J. L. Bos, and B. Snel
2011. Evolution of the tor pathway. *Journal of Molecular Evolution*, 73(3-4):209–220.
- Van Der Laan, J. D. and P. Hogeweg
1995. Predator—prey coevolution: interactions across different timescales. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 259(1354):35–42.
- van Dijk, B. and P. Hogeweg
2015. In silico gene-level evolution explains microbial population diversity through differential gene mobility. *Genome biology and evolution*, 8(1):176–188.
- van Dijk, B., P. Hogeweg, H. Doekes, and N. Takeuchi
2020. Slightly beneficial genes are retained by evolving horizontal gene transfer despite selfish elements. *bioRxiv*.
- van Dijk, B., J. Meijer, T. D. Cuypers, and P. Hogeweg
2019. Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives. *BMC evolutionary biology*, 19(1):201.
- van Gestel, J. and R. Kolter
2019. When we stop thinking about microbes as cells. *Journal of molecular biology*.
- Van Ham, R. C., J. Kamerbeek, C. Palacios, C. Rausell, F. Abascal, U. Bastolla, J. M. Fernández, L. Jiménez, M. Postigo, F. J. Silva, et al.
2003. Reductive genome evolution in buchnera aphidicola. *Proceedings of the National Academy of Sciences*, 100(2):581–586.
- van Heerden, J. H., M. T. Wortel, F. J. Bruggeman, J. J. Heijnen, Y. J. Bollen, R. Planqué, J. Hulshof, T. G. O’Toole, S. A. Wahl, and B. Teusink
2014. Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. *Science*, 343(6174):1245114.
- Van Melderren, L. and M. S. De Bast
2009. Bacterial toxin–antitoxin systems: more than selfish entities? *PLoS genetics*, 5(3):e1000437.
- Van Nimwegen, E., J. P. Crutchfield, and M. Huynen
1999. Neutral evolution of mutational robustness. *Proceedings of the National Academy of Sciences*, 96(17):9716–9720.
- Van Vliet, S., M. Ackermann, et al.
2019. Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *BioRxiv*, P. 534149.
- van Vliet, S., D. J. Kiviet, S. Schlegel, M. Ackermann, et al.
2019. Short-range interactions govern cellular dynamics in microbial multi-genotype systems. *bioRxiv*, P. 530584.
- Veening, J.-W., W. K. Smits, and O. P. Kuipers
2008a. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.*, 62:193–210.
- Veening, J.-W., E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen
2008b. Bet-hedging and epigenetic inheritance in bacterial cell development. *Proceedings of the National Academy of Sciences*, 105(11):4393–4398.
- Vetsigian, K.
2017. Diverse modes of eco-evolutionary dynamics in communities of antibiotic-producing microorganisms. *Nature Ecology & Evolution*, 1:0189.
- Vetsigian, K., R. Jajoo, and R. Kishony
2011. Structure and evolution of streptomyces interaction networks in soil and in silico. *PLoS biology*, 9(10):e1001184.
- Vig-Milkovics, Z., I. Zachar, Á. Kun, A. Szilágyi, and E. Szathmáry
2019. Moderate sex between protocells can balance between a decrease in assortment load and an increase in parasite spread. *Journal of theoretical biology*, 462:304–310.
- Villa, T., L. Feijoo-Siota, A. Sánchez-Pérez, J. R. Rama, and C. Sieiro
2019. Horizontal gene transfer in bacteria, an overview of the mechanisms involved. In *Horizontal Gene Transfer*, Pp. 3–76. Springer.

- Vogan, A. A. and P. G. Higgs
2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. *Biology direct*, 6(1):1.
- Vogwill, T. and R. C. MacLean
2015. The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evolutionary applications*, 8(3):284–295.
- Vos, M., A. Buckling, and B. Kuijper
2019. Sexual selection in bacteria? *Trends in microbiology*.
- Vos, M., M. C. Hesselman, T. A. te Beek, M. W. van Passel, and A. Eyre-Walker
2015. Rates of lateral gene transfer in prokaryotes: high but why? *Trends in microbiology*, 23(10):598–605.
- Vos, M. and G. J. Velicer
2009. Social conflict in centimeter- and global-scale populations of the bacterium *myxococcus xanthus*. *Current Biology*, 19(20):1763–1767.
- Wagner, A.
2005. Robustness, evolvability, and neutrality. *FEBS letters*, 579(8):1772–1778.
- Wagner, A.
2011. The molecular origins of evolutionary innovations. *Trends in Genetics*, 27(10):397–410.
- Wang, J., E. Atolia, B. Hua, Y. Savir, R. Escalante-Chong, and M. Springer
2015. Natural variation in preparation for nutrient depletion reveals a cost–benefit tradeoff. *PLoS biology*, 13(1):e1002041.
- Watson, J. D., F. Crick, et al.
1953. A structure for deoxyribose nucleic acid.
- Welch, R. A., V. Burland, G. Plunkett, P. Redford, P. Roesch, D. Rasko, E. Buckles, S.-R. Liou, A. Boutin, J. Hackett, et al.
2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *escherichia coli*. *Proceedings of the National Academy of Sciences*, 99(26):17020–17024.
- Werren, J. H.
2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10863–10870.
- Werren, J. H., U. Nur, and C.-I. Wu
1988. Selfish genetic elements. *Trends in Ecology & Evolution*, 3(11):297–302.
- White, J. W., A. Rassweiler, J. F. Samhuri, A. C. Stier, and C. White
2014. Ecologists should not use statistical significance tests to interpret simulation model results. *Oikos*, 123(4):385–388.
- Wickham, H.
2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wiedenbeck, J. and F. M. Cohan
2011. Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS microbiology reviews*, 35(5):957–976.
- Wiser, M. J. and R. E. Lenski
2015. A comparison of methods to measure fitness in *escherichia coli*. *PLoS One*, 10(5):e0126210.
- Wiser, M. J., N. Ribeck, and R. E. Lenski
2013. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367.
- Wolf, Y. I. and E. V. Koonin
2013. Genome reduction as the dominant mode of evolution. *Bioessays*, 35(9):829–837.
- Wortel, M. T., E. Bosdriesz, B. Teusink, and F. J. Bruggeman
2016. Evolutionary pressures on microbial metabolic strategies in the chemostat. *Scientific reports*, 6:29503.
- Wright, S.
1931. Evolution in mendelian populations. *Genetics*, 16(2):97.
- Wu, M. and P. G. Higgs
2012. The origin of life is a spatially localized stochastic transition. *Biology direct*, 7(1):42.

- Yaffe, E. and D. A. Relman
2020. Tracking microbial evolution in the human gut using hi-c reveals extensive horizontal gene transfer, persistence and adaptation. *Nature Microbiology*, 5(2):343–353.
- Yang, D.-D., A. Alexander, M. Kinnersley, E. Cook, A. Caudy, A. Rosebrock, and F. Rosenzweig
2020. Fitness and productivity increase with ecotypic diversity among escherichia coli strains that coevolved in a simple, constant environment. *Applied and environmental microbiology*, 86(8).
- Young, I. M. and J. W. Crawford
2004. Interactions and self-organization in the soil-microbe complex. *Science*, 304(5677):1634–1637.
- Zelezniak, A., S. Andrejev, O. Ponomarova, D. R. Mende, P. Bork, and K. R. Patil
2015. Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences*, 112(20):6449–6454.
- Zhang, Z., F. de Barsy, M. Liem, A. Liakopoulos, Y. H. Choi, D. Claessen, and D. Rozen
2019. Antibiotic production is organized by a division of labour in streptomyces. *bioRxiv*, P. 560136.
- Zhaxybayeva, O. and W. F. Doolittle
2011. Lateral gene transfer. *Current Biology*, 21(7):R242–R246.
- Zhaxybayeva, O., P. Lapierre, and J. P. Gogarten
2004. Genome mosaicism and organismal lineages. *TRENDS in Genetics*, 20(5):254–260.
- Zhu, Y. O., G. Sherlock, and D. A. Petrov
2017. Extremely rare polymorphisms in saccharomyces cerevisiae allow inference of the mutational spectrum. *PLoS genetics*, 13(1).

Glossary

ACCESSORY GENES When sampling and sequencing multiple genomes (strains, species), accessory genes are the genes only found in a subset of all genomes.

ANTAGONISTIC INTERACTIONS Interactions between individuals or species where the presence of either participant impedes the growth or survival of the other. (Also see INTERFERENCE COMPETITION)

BALANCED POLYMORPHISM When multiple strains, species, or morphotypes co-occur in stable frequencies, often due to strong intraspecific competition.

BLACK QUEEN HYPOTHESIS The Black Queen hypothesis proposes a mechanism by which loss of important functions can be driven by selection. Organisms may adaptively lose costly genes whose function is performed by other members of the community if the goods provided by that function are publicly accessible. Such functions have been referred to as “leaky” functions.

BOTTOM UP In this thesis I use the term bottom up to describe a modelling design that uses “basic principles” that are axiomatic in biology, such as local reproduction, metabolic reaction networks, and mutations on structured genomes.

BUILDING BLOCKS Building blocks refer to the metabolites or compounds that a cell (*e.g.* a Virtual Microbe) must synthesise to grow and express proteins.

CLOSED PANGENOME A pangenome is called “closed” when it contains mostly core genes, with a limited number of accessory genes

COMPETITIVE EXCLUSION The competitive exclusion principles state that two species that compete for the exact same resources cannot stably coexist.

CONJUGATION A mechanism of horizontal gene transfer where two cells (a donor and recipient) make physical contact to exchange genetic elements.

CORE GENES When sampling and sequencing multiple genomes (strains, species), core genes are the genes found all genomes.

CROSS-FEEDING Cross-feeding in microbial ecology describes the phenomenon where species feed off the metabolic (by)products of another species.

ECO-EVOLUTIONARY DYNAMICS The process where ecological and evolutionary dynamics do not happen in isolation, but mutually influence one another as they play out on overlapping time scales.

EVOEVO EvoEvo is the study of how evolution itself evolves. Examples are how natural selection acts upon mutation rates, DNA repair, rates of horizontal gene transfer, the GP-MAP, and evolvability.

EVOLVABILITY The extent to which a system can undergo adaptive evolution. May be defined on the scale of individuals (the capacity of a genome to give rise to beneficial mutations) or the population (the ability of a population to generate and maintain adaptive genetic diversity).

EXPLOITATIVE COMPETITION Competition where one or more species directly compete for resources such as nutrients, space, light, or oxygen.

GENE TRANSFER AGENTS Phage-like particles that are employed by some bacteria to mediate horizontal gene transfer.

GENE-LEVEL SELECTION Gene-level selection refers to the selection pressures that act upon the spread of genetic elements regardless of its (potential) negative impact on the host. These selection pressures shape selfish genetic elements such as transposons and self-promoting plasmids.

GP-MAP The genotype–phenotype (GP) map is the manner in which changes in genotype (where information is stored) affect the phenotype (where selection acts).

HOMEOSTASIS Homeostasis is the ability of living systems to maintain sufficiently stable chemical and physical conditions.

HORIZONTAL GENE TRANSFER The process by which genes are transferred in any other way than vertical gene transfer (*i.e.* from parent to offspring).

INDIVIDUAL-BASED MODELS Individual-based models (IBMs) are systems with explicit individuals that may differ in for example their parameters, position in space, or local environmental conditions.

INTERFERENCE COMPETITION Competition where one or more species interfere with the reproduction or survival of others, for example by producing toxic compounds. (Also see ANTAGONISTIC INTERACTIONS).

MOBILE GENETIC ELEMENTS Genetic elements that can move around within and between genomes. An MGE that replicates itself at the cost of the host cell is often referred to as a SELFISH GENETIC ELEMENT.

MULLER’S RATCHET A process where bacterial populations irreversibly accumulate slightly deleterious mutations due to the effects of genetic drift in finite, asexual populations.

MUTUALISTIC INTERACTIONS Interactions between individuals or species where both participants have a net benefit.

OPEN PANGENOME A pangenome is called “open” when it contains relatively many accessory genes and few core genes

ORDINARY DIFFERENTIAL EQUATIONS Ordinary differential equations (ODEs) are differential equations with only one independent variable (usually time).

PANGENOME When sampling and sequencing multiple genomes (strains, species), the total collection of genes that is found is called the pangenome.

RED QUEEN DYNAMICS Red Queen dynamics are evolutionary dynamics where continuously changing selection force organisms to constantly adapt in order to persist.

SELFISH GENETIC ELEMENTS Genetic elements that can move around within and between genomes and have a negative impact on its host’s fitness. Examples are transposons, self-promoting plasmids, and bacteriophages.

TODO Instead of thinking about evolution as an optimisation process, Hogeweg & Hesper (1991) propose an alternative view for evolution that considers the simple rules that biological organisms have to adhere to (*i.e.* what there is to do), and how these rules gives rise to complex behavioural and environmental patterns.

TOP DOWN In this thesis I use the term top-down to describe models that are designed to match (fit) a certain observation.

TRANSDUCTION A mechanism of horizontal gene transfer where foreign DNA is introduced into a cell by a phage or another viral vector.

TRANSFORMATION A mechanism of horizontal gene transfer where a bacterial cell takes up and integrates extracellular DNA into its genome.

VERTICAL GENE TRANSFER The process by which genes are transferred from parent to offspring.

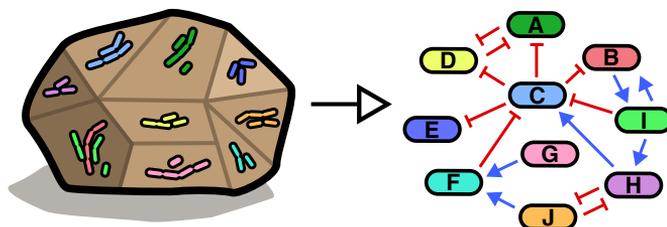
Begrippenlijst: De belangrijkste begrippen om de Nederlandse samenvatting van dit proefschrift beter te kunnen begrijpen.

Begrip	Omschrijving
Bioinformatica	Tegenwoordig wordt met bioinformatica vaak de analyse van biologische sequentiedata (DNA, RNA en eiwitten) bedoeld. De oorspronkelijke betekenis van het woord, door Ben Hesper en Paulien Hogeweg beschreven in 1978, was echter hoe biologische systemen <i>zelf</i> informatie verwerken.
Diversificatie	Het proces waardoor twee of meerdere soorten kunnen ontstaan uit één voorouderlijke soort.
DNA	DNA staat voor deoxyribonucleic acid (desoxyribonucleïnezuur), en is de informatiedrager van de cel. Op het DNA van de cel liggen genen (zie ook: genen) die voornamelijk coderen voor eiwitten (zie ook: eiwitten).
eDNA	Extracellulair DNA. Dit DNA bevindt zich niet in een levende cel, maar los in het milieu. Veel bacteriën en andere micro-organismen kunnen dit losse DNA opnemen (zie ook: horizontale genoverdracht), en de genen die zich op dit DNA bevinden zodoende tot expressie brengen (zie ook: genexpressie).
Eiwitten	Waar DNA de drager is van genetische informatie, zijn het de eiwitten (ook wel proteïnen genoemd) die uiteindelijk het werk doen. Eiwitten spelen ieder een specifieke rol om het overleven, groeien, en het delen van cellen te bewerkstelligen.
Eukaryoten	Organismen met celkern (zoals planten en dieren, maar ook bepaalde micro-organismen zoals bakkersgist)
Genen	Genen zijn discrete eenheden erfelijk materiaal (m.a.w. specifieke stukken DNA) die coderen voor bepaalde eindproducten (vaak een eiwit, zie ook: eiwitten).
Genduplicatie	Een specifieke mutatie (zie ook: mutatie) waarna een organisme twee (of meerdere) kopieën van hetzelfde gen draagt.
Genexpressie	Er wordt van genexpressie gesproken wanneer een gen dat zich op het DNA bevindt (zie ook: gen) daadwerkelijk wordt omgezet tot een eindproduct (bijv. een eiwit, zie ook: eiwitten). Genen in het genoom (zie ook: genoom) van een organisme komen niet altijd (even veel) tot expressie.
Genmobiliteit	De mate waarin een gen wordt overgedragen m.b.v. horizontale genoverdracht (zie ook: horizontale genoverdracht).
Genoom	Het genoom van een organisme is de volledige verzameling genetisch materiaal van een organisme. Een groot genoom kan veel (verschillende) genen bevatten, maar kan soms ook veel niet-coderende sequenties dragen (DNA wat niet voor genen codeert).
Horizontale genoverdracht	Wanneer genen op een andere wijze worden overgedragen dan van ouder op kind. Dit proces komt vaak voor bij micro-organismen.
Metabolisme	Dit proces wordt vaak ook wel stofwisseling genoemd. Metabolisme is het geheel van chemische processen dat in een cel plaatsvindt.
Model	Een model is een versimpeling van de werkelijkheid. Door modellen van een systeem te ontwerpen en bestuderen, verkrijgen we nieuwe inzichten over het gemodelleerde systeem.
Mutatie	Een kleine of grote verandering in het DNA van een organisme. Mutaties treden voornamelijk op tijdens celdeling, waardoor nakomelingen net iets anders zullen zijn dan hun ouder(s).
Prokaryoten	Organismen zonder celkern (voornamelijke micro-organismen)
Verticale genoverdracht	Wanneer genen worden overgedragen van ouder op kind.
Virtual Microbes	Een model dat is ontwikkeld tijdens mijn promotietraject. Virtual Microbes (virtuele micro-organismen) evolueren op basis van een artificiële chemie, en moeten leren om de juiste bouwstenen te kunnen produceren op basis van een gegeven voedselbron.
Wild-type	De verschijningsvorm waarin een bepaalde soort in de natuur voorkomt. Met het Virtual Microbes model hebben we vaak gebruik gemaakt van “virtuele” wild-types, door onze experimenten pas te beginnen nadat deze digitale organismen langdurig onder “natuurlijke” omstandigheden waren geëvolueerd.
Zelfzuchtige genen	Genen die d.m.v. horizontale genoverdracht in een populatie kunnen overleven, zonder nut (of zelfs met kosten) voor de gastheercel.

Samenvatting in het Nederlands

De diversiteit van soorten op een enkele zandkorrel

Overal in de natuur barst het van de diversiteit. Niet alleen zijn er eindeloos veel soorten zichtbaar met het blote oog, maar ook de allerkleinste milieus zitten vol met verschillende levensvormen. Zo zijn er met behulp van een microscoop bijvoorbeeld al heel veel verschillende micro-organismen te zien in een druppel slootwater. De diversiteit van micro-organismen is echter nog veel opmerkelijker gebleken in de afgelopen twintig jaar, omdat we naast microscopie nu ook precies kunnen reconstrueren welk genetisch materiaal (DNA) zich in een bepaald milieu bevindt. Met behulp van deze nieuwe methode is ontdekt dat in allerlei omgevingen veel meer soorten voorkomen dan aanvankelijk werd gedacht. Naast natuurlijke omgevingen zoals bodemgrond en water, leven er in bijvoorbeeld onze darmen en in douchebakken waanzinnig veel verschillende micro-organismen. Als we het echter hebben over de diversiteit binnen één enkel milieu, dan is vooral bodemgrond opmerkelijk divers. Zelfs op het kleinste korreltje bodemgrond, dat met een naald bemonsterd moet worden, leven tot wel honderd verschillende soorten micro-organismen. Bovendien kunnen deze verschillende soorten complexe interactienetwerken vormen, waar bijvoorbeeld de aanwezigheid van de ene soort de groei van een ander bevordert of verhindert (**Figuur 1**). Hoe kunnen al deze verschillende soorten bestaan op de schaal van een korrel zand?



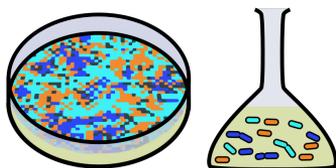
Figuur 1: Zelfs op een enkele korrel bodemgrond leven tientallen soorten micro-organismen. Deze micro-organismen vormen complexe interactie-netwerken, waarin de aanwezigheid van de ene soort de groei van de ander kan bevorderen (blauwe pijlen) of verhinderen (rode pijlen).

Computationele modellen als "virtueel laboratorium"

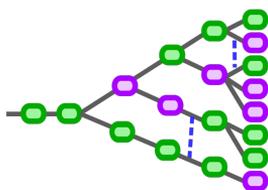
Charles Darwin heeft in 1859 in zijn boek "On the origin of species" al beschreven hoe de enorme diversiteit van levensvormen op aarde kan worden verklaard aan de hand van evolutie door natuurlijke selectie. Soorten die beter zijn aangepast aan hun omgeving blijven over, terwijl de andere soorten uitsterven. Zodoende zijn er heel veel soorten, simpelweg omdat er veel verschillende omgevingen zijn. Maar zijn er ook zoveel verschillende omgevingen op een enkele korrel zand?

Ruim anderhalve eeuw na Darwin's boek hebben we met behulp van experimenteel onderzoek naar de fysiologie en evolutie van micro-organismen, en de computationele analyse van genetisch materiaal (bioinformatica), al erg veel kennis vergaard over hoe het leven op de microschaal eruitziet. Daarnaast hebben ook theoretische benaderingen, zoals wiskundige of computationele modellen, bijgedragen aan onze kennis van het ontstaan en behouden van diversiteit. Veel van deze modellen gebruiken echter sterke versimpelingen van de daadwerkelijke details van het microbiële leven, bijvoorbeeld door microbiële groei te omschrijven met een eenvoudige

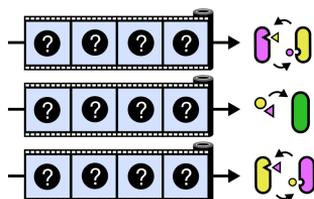
Computationele modellen kunnen dienen als een "virtueel laboratorium"



Kweken van micro-organismen door uit te platen of door goed te mixen.



Nauwkeurig bijhouden van de afstammingslijn en genoverdracht.



Herafspelen van het evolutieproces om de voorspelbaarheid hiervan te onderzoeken.

Figuur 2: Met behulp van een "virtueel laboratorium" kunnen we de ecologie en evolutie van microbiële systemen in actie zien, en zodoende de natuurlijke systemen beter leren begrijpen.

wiskundige vergelijking. Hoewel deze modellen erg goed de groei van microbiële populaties kunnen omschrijven, geven ze weinig inzicht in wat een micro-organisme eigenlijk moet *doen* om deze groei te bewerkstelligen.

Het doel van dit proefschrift is om modellen te ontwerpen waarin belangrijke basisprincipes van het leven op de microschaal voorop staan. Dit doen we door bijvoorbeeld chemische reacties zoals het produceren van eiwitten en andere bouwstenen, interacties tussen genen en de cel, en ruimtelijke populatiestructuur in de modellen te verwerken. Dergelijke computationele modellen omvatten, net als echte microbiële populaties, vaak vele schaalniveaus (van moleculen tot genen, genen tot cellen, etc.), die bovendien allemaal elkaar kunnen beïnvloeden. Met behulp van een dergelijk “virtueel laboratorium” (**Figuur 2**) kunnen we dan de ecologie en evolutie van micro-organismen in actie zien, en meer leren over wat er op kleine schaal in onze darmen en in de bodem gebeurt.

Samenvatting van de hoofdstukken

Naast een algemene introductie (**Hoofdstuk 1**) en een wetenschappelijke discussie (**Hoofdstuk 7**), bestaat dit proefschrift grofweg uit twee delen. In deel 1 (**Hoofdstuk 2 en 3**), ontwerpen en bestuderen we een model genaamd Virtual Microbes (virtuele micro-organismen), dat ik samen met Thomas Cuypers heb ontwikkeld tijdens mijn promotietraject. Dit model omvat veel belangrijke details van wat een micro-organisme "moet doen" om te kunnen overleven, groeien, en reproduceren. Zo moeten deze virtuele micro-organismen bijvoorbeeld bepaalde bouwstenen kunnen produceren door de juiste chemische reacties te gebruiken, maar moeten ze ook zorgen dat er geen toxische hoeveelheden restproducten opbouwen in hun cel. In dit eerste deel behandelen we hoe verschillende microbiële soorten kunnen vormen, en hoe goed we deze evolutie kunnen voorspellen. In het tweede deel (**Hoofdstuk 4, 5, en 6**) richten we ons specifiek op één proces. Naast het verkrijgen van genen van hun oudercel (zogenoemde verticale genoverdracht), kunnen bacteriën en andere microben ook genen opnemen van individuen uit dezelfde generatie. Dit proces heet horizontale genoverdracht. Omdat horizontale genoverdracht niet plaatsvindt in planten en dieren, is onze kennis over hoe dit evolutie beïnvloedt nog zeer beperkt. Deel 2 laat zien wat de gevolgen zijn van horizontale genoverdracht, en hoe het er voor kan zorgen dat er veel meer soorten micro-organismen behouden kunnen blijven in een klein ecosysteem.

In **Hoofdstuk 2** bestuderen we de bovengenoemde Virtual Microbes zich aanpassen aan omstandigheden zoals in een laboratorium. We doen dit door de omstandigheden na te bootsen van een bekend microbiële evolutie-experiment, dat in 1988 is gestart door Richard Lenski en nog tot de dag van vandaag in gang is. In dit experiment ondergaat de bacterie *Escherichia coli* elke 24 uur opnieuw een groeicyclus van exponentiële groei (zolang voedsel beschikbaar is) en stagnatie (wanneer het voedsel op is). Voordat dit experiment begon, heeft *E. coli* natuurlijk heel lang in de natuur geleefd, zowel in de bodem waar voedsel soms schaars is, als in de darmen van dieren, waar juist heel veel voedsel aanwezig is. Zodoende had het zogenoemde wild-type van *E. coli* allerlei aanpassingen om te kunnen omgaan met onvoorspelbare veranderingen in voedselvoorziening. Om de vergelijking met het

experimentele protocol te versterken, hebben we in het Virtual Microbes model eerst 16 verschillende “wild-types” geëvolueerd die, net als *E. coli*, hebben leren omgaan met lange en onvoorspelbare periodes van uithongering. Een belangrijk vraagstuk hierbij was: passen al deze 16 wild-types zich precies op dezelfde manier aan? Met andere woorden, hoe beïnvloedt de “natuurlijke” evolutie die aan het experiment vooraf is gegaan, toekomstige evolutie in het experimentele protocol? Onze resultaten laten zien dat voorgaande evolutie aan de ene kant veel invloed kan hebben op de precieze details van de uitkomst, maar dat er op grotere schaal vaak hetzelfde gebeurt. Zo leren alle 16 wild-types om de voorspelbare omstandigheden van het lab-protocol te vertrouwen, en de toevoer van nieuw voedsel nauwkeurig te anticiperen. Echter laten we ook zien dat deze anticipatie op veel verschillende manieren kan plaatsvinden. Zo ontstaan er regelmatig sub-soorten die specialiseren op verschillende groeifasen, waar de ene soort beter groeit zolang er veel voedsel beschikbaar is, en de andere soort beter overleeft als het voedsel op is. Echter zien we dat er ook één soort kan ontstaan die al deze groeifasen in zijn eentje benut. Daarnaast zijn er enkele wild-types die altijd dezelfde aanpassingen laten zien, terwijl andere wild-types iedere keer dat het experiment wordt herhaald totaal andere aanpassingen vertonen. Hoewel het voorspellen van de toekomst (en het verleden!) vaak een belangrijke mijlpaal is voor een wetenschappelijk vakgebied, laten onze resultaten zien dat de voorspelbaarheid van evolutie zelf onvoorspelbaar kan zijn.

Hoofdstuk 3 bouwt voort op de vraag of evolutie voorspelbaar is, maar nu onder meer natuurlijke omstandigheden. De Virtual Microbes evolueren in dit hoofdstuk in een meer complexe en ruimtelijk gestructureerde omgeving. We herhalen de evolutie voor één en dezelfde startpopulatie van micro-organismen, waarin de individuen zich moeten aanpassen om bepaalde bouwstenen te synthetiseren op basis van één voedselbron. Om dit doel te bewerkstelligen kunnen de micro-organismen vele alternatieve chemische reacties gebruiken, welke allemaal andere metabolieten (energiedragers, intermediaire producten, bijproducten, en eindproducten) op kunnen leveren. Door middel van opname en uitwisseling van deze metabolieten zullen de micro-organismen niet alleen kunnen groeien, maar ook hun lokale milieu beïnvloeden. Zo kunnen bijvoorbeeld belangrijke eindproducten ophopen in het milieu als deze niet volledig worden geconsumeerd door een cel. Als we dit evolutie-experiment vele keren herhalen, zien we dat de geëvolueerde populatie van micro-organismen of uit twee van elkaar afhankelijke soorten bestaat, óf bestaat uit één soort die zichzelf van alle benodigde bouwstenen voorziet. De twee afhankelijke soorten uit de eerste set populaties hebben een complementair metabolisme ontwikkeld, en beide soorten zijn daardoor van de andere soort afhankelijk om te kunnen blijven groeien en overleven. Omdat dergelijke afhankelijkheden ook in de natuur veel voorkomen, is het interessant om de vraag te stellen wat nou precies bepaalt of deze afhankelijkheid ontstaat. Door het metabolisme van de geëvolueerde soorten te bestuderen hebben we ontdekt dat het type uitkomst (twee afhankelijke soorten of een zelfvoorzienende soort) voornamelijk wordt bepaald door toevalligheden in hoe de micro-organismen zich hebben aangepast om energie te produceren. Deze aanpassing zelf wordt echter volledig bepaald door een kansproces, waardoor we dus niet ver van tevoren kunnen voorspellen welk van de twee uitkomsten zal evolueren. Desalniettemin, kunnen we nadat een specifiek energiemetabolisme is geëvolueerd, heel goed voorspellen welk van de twee uitkomsten zal optreden. Zeer vergelijkbaar met Hoofdstuk 2, zien we

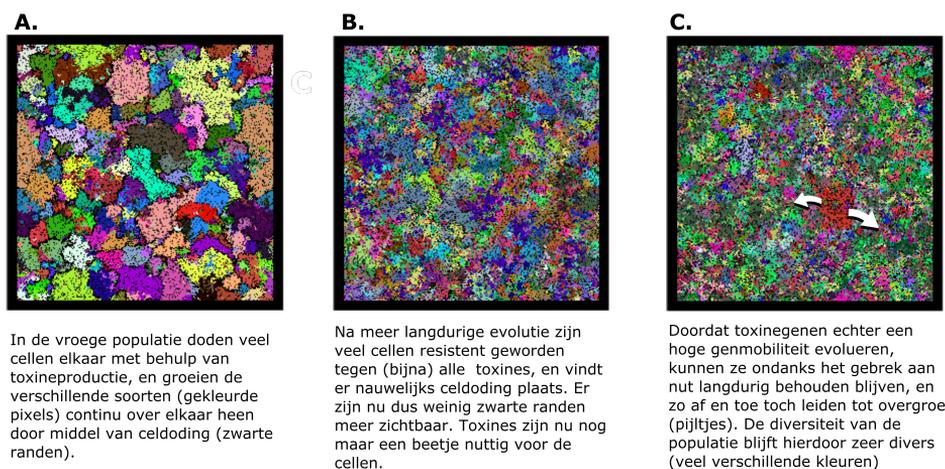
dus wederom dat de voorspelbaarheid van het evolutieproces zelf kan evolueren.

Zoals bovengenoemd, bestuderen we in het tweede deel van dit proefschrift het proces horizontale genoverdracht. Omdat door dit proces genen zich op een unieke manier van individu tot individu kunnen verplaatsen, heeft horizontale genoverdracht belangrijke consequenties voor wie of wat er precies geselecteerd wordt door het totale evolutionaire proces. Zo kunnen aan de ene kant micro-organismen belangrijke genen onderling delen, om zich zo sneller aan te kunnen passen aan een nieuwe omgeving waar nieuwe voedselbronnen of antibiotica aanwezig zijn. Aan de andere kant kunnen de genen zich echter ook zelfzuchtig gaan gedragen, en zichzelf gaan repliceren ten koste van hun gastheer. Een belangrijke vraag is dus: wanneer is horizontale genoverdracht nuttig voor de micro-organismen, en wanneer is het slechts nuttig voor de zelfzuchtige genen? Deze vraagstukken bestuderen we door expliciet het perspectief van genen, en dat van hun gastheer, te bestuderen. Ook bestuderen we hoe dit proces het behoud van diversiteit en genen kan beïnvloeden.

Van natuurlijke microbiële populaties weten we dat sommige genen veel gemakkelijker door middel van horizontale genoverdracht overgedragen kunnen worden dan andere genen. Een dergelijk verhoogde “genmobiliteit” kan bijvoorbeeld voorkomen als het gen zich op een zogenaamd mobiel genetisch element (MGE) bevindt. Maar precies hoeveel vaker zo’n mobiel genetisch element overgedragen wordt dan andere genen, en wat deze verschillen in genmobiliteit veroorzaakt, is nog niet helemaal duidelijk. Daarom geven we ons model in **Hoofdstuk 4** deze genmobiliteit geen vaste waarde, maar laten we ieder gen evolueren hoe gemakkelijk het zich van gastheer tot gastheer verplaatst. Dit doen we in een microbiële populatie waar de individuele gastheercellen verschillende toxine- en resistentiegenen kunnen produceren, waardoor er een complexe wapenwedloop ontstaat van toxineproductie, -resistentie, en -sensitiviteit (**Figuur 3A**). De individuen in de populatie raken deze genen soms kwijt, bijvoorbeeld doordat er een inactiverende mutatie optreedt. Het effect van een dergelijk genverlies is natuurlijk verschillend voor toxinegenen en resistentiegenen. Het verlies van een resistentiegen is alleen schadelijk als er (in de buurt) toxines geproduceerd worden, en het verlies van toxinegenen is alleen schadelijk als er in de buurt sensitieve cellen leven die je anders met behulp van deze toxines had kunnen verdrijven. Naast het verliezen van genen, kunnen cellen in dit model ook extracellulair DNA (eDNA) opnemen uit een eDNA-poel met daarin de toxine- en resistentiegenen van recent overleden cellen. Dit opnemen van genen uit de omgeving heet transformatie, en is een specifieke vorm van horizontale genoverdracht. Door middel van het opnemen van eDNA kunnen gastheercellen de bovengenoemde verloren genen weer terug krijgen, maar omdat de genen ook hun eigen genmobiliteit kunnen bepalen, kunnen er door deze opname van eDNA ook zelfzuchtige genen ontstaan. Vantevoren wisten we niet precies welke genen hun mobiliteit zouden gaan verhogen of verlagen, en of de zelfzuchtige genen een probleem voor hun gastheer zouden opleveren.

Onze resultaten laten zien dat, dankzij de opname van eDNA, alle micro-organismen eerst resistent worden tegen (bijna) alle toxines die in de populatie aanwezig zijn. Hierdoor zie je dat de eerder genoemde wapenwedloop een iets andere vorm krijgt, en er minder cellen door toxines worden gedood dan in de vroegere populaties (**Figuur 3B**). In deze nieuwe populatiesamenstelling, is het nut van toxine- en

resistentiegenen heel anders geworden. Resistentiegenen zijn regelmatig nuttig om te dragen, omdat er sporadisch individuen in je omgeving voorkomen die vreemde toxines kunnen produceren. Aan de andere kant zijn toxinegenen veel minder nuttig, omdat (bijna) iedere cel resistent is tegen alles. Echter treedt er eens in de zoveel tijd een “golf” van celdood plaats (**Figuur 3C**), omdat een subpopulatie cellen de resistentie tegen een zeldzaam toxine is verloren. Kortgezegd zijn de resistentiegenen dus zeer nuttig, en de toxinegenen slechts een beetje nuttig. Dit verschil in nut zorgt ook voor een groot verschil in het effect van het opname van deze genen uit de eDNA-poel. De opname van resistentiegenen is voor de gastheer vaak slechts een extra last om te dragen, omdat ze deze genen vaak helemaal niet waren kwijt zijn geraakt. Het opnemen van toxinegenen is ook lang niet altijd nuttig voor de cellen, omdat het merendeel van de populatie resistent is tegen bijna alle toxines. Echter is er sporadisch een enorm voordeel te behalen als een naburige stam cellen toch sensitief is geworden. Deze toxinegenen, die meestal alleen maar kosten met zich meebrengen maar sporadisch enorm voordelig zijn, evolueren een hoge genmobiliteit, en worden daardoor regelmatig gedeeld tussen cellen uit dezelfde generatie. Resistentiegenen daarentegen evolueren een minimale genmobiliteit, en worden dus voornamelijk van ouder- op dochtercel overgedragen. Deze resultaten geven dus inzicht in welke genen nuttig zijn om horizontaal over te dragen (hier de toxinegenen), en welke genen beter verticaal overgedragen kunnen worden. Ook vinden we dat in dit ruimtelijk gestructureerde model, zelfzuchtige genen eigenlijk geen groot probleem zijn voor de micro-organismen, omdat op plaatsen waar zelfzuchtige genen de kop opsteken een ongezonde stam ontstaat, die met de tijd door andere micro-organismen wordt overgroeid. Als we echter toestaan dat de genen zich ook tussen ver uit elkaar liggende individuen kunnen verplaatsen, verliest het systeem deze negatieve terugkoppeling en ontstaan er wel veel zelfzuchtige genen. Het probleem van zelfzuchtige genen lijkt dus vooral relevant voor goed gemengde systemen, zoals je die voornamelijk in het laboratorium vindt.



Figuur 3: Een computationeel model van micro-organismen die verschillende genen dragen om toxines (giftstoffen) te produceren, of om daartegen resistent te zijn. Het model illustreert hoe een groot aantal genen (ook genen met maar weinig nut) toch kunnen overleven in een relatief kleine populatie micro-organismen.

Hoofdstuk 5 vervolgt het onderzoek naar de effecten van horizontale genoverdracht, maar richt zich nu op het perspectief van de individuele cel: “wil” een cel eigenlijk wel DNA uit zijn omgeving opnemen, als dit eventueel met kosten en zelfzuchtige genen gepaard gaat? En houdt de verrassende conclusie dat de minst nuttige genen vaker worden overgedragen stand? Met een wiskundige analyse laten we zien dat horizontale genoverdracht inderdaad vooral nuttig is voor genen die slechts een beetje nuttig zijn voor hun gastheer. Sterker nog, een bepaalde klasse genen valt alleen “te redden” dankzij dit proces, en zouden anders niet behouden blijven in de populatie. Deze analyse toont echter ook aan dat de gastheren deze genen die “te redden” zijn, niet kunnen redden door de opname van DNA te evolueren. Dit probleem ontstaat door een paradox, waar het opnemen van DNA alleen nuttig is als het gen behouden is gebleven, maar het gen alleen behouden blijft als cellen DNA opnemen. Echter, wanneer we ditzelfde systeem in een ruimtelijk model bestuderen, zien we dat micro-organismen weldegelijk over deze drempel heen kunnen komen, door eerst lokaal de genen te delen met een klein collectief van cellen. Zelfs als we in dit scenario sterk zelfzuchtige genen aan de populatie toevoegen, blijven de gastheercellen DNA opnemen om de genen die “te redden” zijn, te redden.

In hoofdstuk 4 en 5 hebben we gezien dat horizontale genoverdracht als een belangrijk reparatiemechanisme kan optreden wanneer er genverlies optreedt. Een belangrijke vervolgvraag in **Hoofdstuk 6** is hoe deze reparatie zich verhoudt tot andere manieren om genverlies te voorkomen. Zo kunnen organismen bijvoorbeeld ook extra kopieën maken van genen, zodat ze altijd een backup hebben. Een dergelijke backup kan worden gemaakt door middel van genduplicatie, een mechanisme dat veel voorkomt bij eukaryote (kerndragende) organismen. Bacteriën en andere prokaryoten (organismen zonder celkern) gebruiken minder vaak genduplicaties, en doen vaker beroep op het bovengenoemde mechanisme van horizontale genoverdracht. Hoewel met behulp van beide mechanismen één extra gen aan een genoom wordt toegevoegd, is de oorsprong van dit gen heel anders. Een belangrijke onderzoeksvraag is dan wat de gevolgen zijn van deze twee verschillende mechanismen om (nieuwe) genen te verkrijgen. Om een beeld te krijgen van de mogelijke gevolgen, verkennen we het verschil tussen genduplicaties en horizontale genoverdracht eerst in het Virtual Microbe model uit hoofdstuk 2 en 3. We nemen een geëvolueerde microbe uit hoofdstuk 3, en vervolgen het evolutionaire proces lange tijd in populaties waarin óf alleen horizontale genoverdracht optreedt, óf alleen genduplicaties optreden. We zien dan dat organismen die horizontale genoverdracht gebruiken veel kleinere genomen evolueren, maar desondanks meer verschillende genen kunnen behouden. De organismen die geen horizontale genoverdracht gebruiken, maar in plaats daarvan alleen genduplicaties kunnen gebruiken evolueren een hogere, en sterk fluctuerende genomegrootte. Ook zien we dat deze veranderingen in genomegrootte gepaard gaan met het reduceren van de mate van eiwitexpressie per individueel gen. Om verder uit te diepen wat de minimale vereisten zijn voor deze patronen, ontwikkelen we vervolgens een mini-model. Dit mini-model reproduceert onze eerdere resultaten, en laat zien dat horizontale genoverdracht en genduplicaties twee alternatieve genomestructuren opleveren. In beide populaties voorkomen de genomen, op geheel andere wijze, dat ze op de lange termijn hun goede genen verliezen.

Slotwoord

In dit proefschrift hebben we onderzocht hoe de ecologie en evolutie van het allerkleinste leven er in actie uitziet. Zodoende hebben we de invloed van verschillende niveaus (bijv. moleculen, genen, cellen, en populaties) bestudeerd, en de wisselwerking tussen deze niveaus in kaart gebracht. We hebben gezien dat evolutie niet altijd even voorspelbaar is, en dat horizontale genoverdracht sterk kan beïnvloeden hoe evolutie te werk gaat. Dit hebben we allemaal bestudeerd, zoals de titel van dit proefschrift suggereert, “op een Virtuele Zandkorrel”, d.w.z. door middel van computationele modellen. Waar experimentele evolutie en bioinformatica ons hebben voorzien van een schatkist vol met data, hebben wij deze data gebruikt om biologisch interessante vereenvoudigingen te maken. Hoewel deze systemen zich soms erg ingewikkeld en counter-intuïtief gedragen, is dat tevens de kracht van deze modellen. De verrassende dynamiek die ontstaat leidt vaak tot nieuwe inzichten, die we nooit van tevoren hadden kunnen zien aankomen. Maar naast nieuwe inzichten leveren de experimenten in onze “virtuele laboratoria” ook veel nieuwe vragen op. Om deze vragen te beantwoorden, moeten we weer terug naar het – al dan niet virtuele – laboratorium.

Dankwoord

Laat ik bij het begin beginnen. Mam, Pap, jullie hebben mijn nieuwsgierigheid nooit ontmoedigd. Kinderen zitten vol met waarom-vragen, en hoewel jullie vaak een concreet antwoord voor deze vragen hadden, was “dat weet ik niet” misschien nog wel het meest leerzame antwoord. Soms raakte het geduld op, en kreeg ik antwoorden zoals “die blauwe plek is botrot” en “dit is de zelfvernietigingsknop van de Twingo”. Hoewel ik destijds wakker kon liggen van hoe dichtbij het stuur de zelfvernietigingsknop van de Volkswagen eigenlijk zat, ben ik blij met de nuchterheid die deze antwoorden me hebben bijgebracht. Ook mijn zussen Sanne en Noortje wil ik bedanken voor alle goede herinneringen, zoals je insmeren met modder om je vervolgens meteen weer schoon te douchen (een zeer belangrijke bezigheid natuurlijk), of drie uur lang naar hetzelfde liedje luisteren (één... negen!). Hierdoor herinner ik me soms nog maar al te goed hoe het was om een klein kind te zijn (in plaats van een groot kind). Over herinneringen aan mezelf als jonger persoon gesproken: Luuk, bedankt voor alle potjes Super Smash tijdens het afronden van mijn proefschrift. Zelfs met mijn favoriete beer-en-vogel duo kan ik maar zelden van je winnen.

Naast mijn directe familie, wil ik ook een aantal Cramwinckels bedanken. Hans en Jeannette, bedankt voor alle aanmoedigingen en hulp in het begin van mijn academische carrière. Ook wil ik Margot's broers Ewoud en Ivo bedanken. Toen ik nog twijfels had over of ik wel geschikt was voor de universiteit, heeft de vanzelfsprekendheid waarmee jullie met mij over complexe onderwerpen en wetenschap praatten, mij erg veel geholpen.

Af en toe moeten er gewoon liedjes over Kees gezongen worden. Margot, Sam, Rochelle, Lex, Lisette, Lindy, Niels, en Alex, bedankt voor alle keren dat we rond de tafel stonden en... dingen aten. Geen van ons is echt hetzelfde gebleven, maar eigenlijk gewoon stiekem wel. De komende jaren zal het feestje misschien iets vaker digitaal moeten, maar ook kijk ik uit naar het volgende potjes-/potvol-weekend.

In een groep van 220 Biologie studenten, heb ik een handvol studiegenoten ook als vrienden beschouwd. Rutger, Ino, Fabian, Koen, en Roderick, bedankt voor alle gekras en hoedjes in MIJN AANTEKENINGEN, de muziek, de kraaien, en de presentaties met out-of-bed look.

Next, I would like to thank the reading committee, adding at least five people to the list of people who thoroughly read my thesis. Martin Ackerman, Guillaume Beslon, Santiago Elena, Paul Higgs, and Paul Rainey, thank you for taking the time to read

my work. Much of your work has inspired the topics discussed throughout this thesis, and I hope I managed to inspire you back. Guillaume Beslon in particular, you have carried the weight of the EvoEvo project on your shoulders, and somehow still had time for encouraging words when I was ready to give up on Virtual Microbes. I'd also like to thank the other members of the EvoEvo-project for all the fruitful discussions.

I've been around TBB for far too long to never actually experience my final day there (bloody Coronavirus!). At the same time however, I don't see the finality of it. First of all, the lessons taught by (in order) Alexander Panvilov, Rob de Boer, Kirsten ten Tusscher, Can Kesmir, Paulien Hogeweg, and Berend Snel will be with me indefinitely. But more importantly, I am positive we will keep in touch. There are so many other TBB-people that I'd like to thank that I'd like to first cowardly thank "all of you". But more than a few people have done too much for me to not mention them by name.

First of all, there is of course Paulien. Paulien, you were my teacher, internship supervisor, and PhD promotor, and throughout the years you started to feel more and more like a friend. In the 8 years I have worked with you, I don't think we've ever had a single boring conversation. You have taught me to think critically, and to not be too eager to rely on the simplest possible explanation. But most of all, you (and through you, Ben) taught me that science is fun. Thank you both. Although I am thoroughly convinced that any individual elephant came about by random mutations, I am tempted to stick to the word "believe" when considering how both of your extremely non-random brains came about ;)

Next, it is impossible to overstate my gratitude to the following person. Jan Kees, you are single-handedly responsible for so many people actually achieving the end goal of a doctoral degree. Or in my case, a bachelor's + master's + a doctoral degree. I remember when I moved out of my parents house, and I asked my father for a shelf in the bathroom (yes, this is somehow about you, and not about my father). He told me that it would be difficult to put a shelf there. Five minutes later a shelf had mysteriously appeared in my bathroom. You did exactly the same thing with regard to computer problems, times a hundred. It happened exactly 0 times that you couldn't find a good solution to my problems, or at least better solution than I came up with. What's more, instead of complaining about the memory-intensive nightmare that was Virtual Microbes, you've always helped me find creative solutions to manage its complexity. It appears that "sacrifice" is truly in the DNA of the Amerongen family. Thank you. Thank you so much.

I also want to thank some people who are / use to be at TBB that helped or inspired me during my PhD. Otto Cordero, Paola Carrillo-Bustamante, Like Fokkens, Michael Seidl, Saikrishna Gadhamsetty (or Batman, who really knows), Bas Dutilh, and Nobuto Takeuchi. Nobuto, thank you for our emergent interactions. There's a few more at TBB that I cannot in clear conscience mention merely by enumeration.

Sandro and Renske, thank you for being my senpais. Besides helping me out with the science, we have had endless discussions about literally everything. If ever there was evidence that conversations are a powerful tool to make progress in pretty much

every domain, it was with you two. Sandro's Italian volume combines beautifully with Renske's eloquence, and I hope that there are many more conversations to be had.

Thomas, I don't know what is wrong with you, but of all the horrible great ideas you've had, I think Virtual Microbes was the most horrible great idea. Thank you for developing most of what is now Virtual Microbes, for your patience with my infinite requests for improvements, and for the endless discussions we've had with respect to "the scaling NON".

A few more TBB-associated names. Alessia, even though our time at TBB had a significant overlap, we somehow became friends right before you left. Luckily, we still enjoy grabbing an ice cream every now and then. Laurens, even your presence in the lab always felt like a feat of live performance. You are probably the jolliest mathematician I know. But then again, I don't know that many mathematicians. Juliane, thank you for our rides through the chicken park, and the constant reminder of what it is like to "have a plan". I hope you succeed with whatever it is you plan on doing. Lianne, although our student-supervisor relationship was cut short, I have really enjoyed our collaboration. Your hyperactive BS-detector was very useful for debugging both our code, and debugging me during our conversations. Ianthe, thank you for diving into the deep end with me concerning (sometimes painfully hard) philosophical topics. Lastly, Dieter, please keep the Paulien-should-write-a-book vibe going in my stead. And thank you for your thank you, it really brought a smile to my face.

Finally, I'd like to especially thank my paranymphs Hilje and Jeroen, who I respectively consider to be my academic big sister and (not so) little brother. Hilje, thank you for putting up with my ****. Jeroen, thank you for not putting up with my ****.

Als allerlaatste Margot. Dankjewel dat je er bent. Ik ben blij dat ik samen met jou, en niet alleen, in een proefschrift-corona-quarantaine zat. Nu is het tijd voor het volgende avontuur, waarin Engeland, Duitsland, Ierland, en heel veel andere landen flink befietsts en/of bewandeld gaan worden. Zin in.

Curriculum vitæ

Bram van Dijk was born on the 22nd of May (1988) in Apeldoorn, the Netherlands. After a vocational training in media design, he finished his entrance examination (*colloquium doctum*) to study Biology at Utrecht University in 2009. In 2012, he finished his Bachelor's degree with a thesis focused on the impact of horizontal gene transfer on microbial evolution, combining a literature study with theoretical modelling. Together with Paulien Hogeweg he continued to explore the evolution of microorganisms during his MSc track at Molecular and Cellular Life Sciences (Theoretical Biology track). He obtained his Master of Science degree in 2015, and started as a PhD candidate under supervision of Paulien Hogeweg in January 2015. The results of this PhD are described in this thesis.



List of Publications

2016

van Dijk, B., Hogeweg, P. (2016). In silico gene-level evolution explains microbial population diversity through differential gene mobility. *Genome biology and evolution*, 8(1), 176-188.

2019

van Dijk, B., Meijer, J., Cuypers, T. D., Hogeweg, P. (2019). Trusting the hand that feeds: microbes evolve to anticipate a serial transfer protocol as individuals or collectives. *BMC evolutionary biology*, 19(1), 201.

2020

van Dijk, B., Hogeweg, P., Doekes, H.M., Takeuchi, N. (2020). Slightly beneficial genes are retained by bacteria evolving DNA uptake despite selfish elements. *eLife* 9 (2020): e56801.

Meijer, J., **van Dijk. B.**, Hogeweg, P. (2020). Contingent evolution of alternative metabolic network topologies determines whether cross-feeding evolves. *Accepted at Nature Communications Biology*.

van Dijk, B., Hogeweg P. The evolution of alternative genome architectures through Horizontal Gene Transfer and Gene Duplications. *In preparation*, 2020.

“FINISHED, NOT PERFECT.”

—Jake Parker