Faculty of Bioscience Engineering

Academic year 2014-2015

Assessing pathogen invasion based on community evenness and metabolic similarity

Wai Kit Tsang

Promotors: Prof. Dr. ir. Nico Boon, Prof. Dr. Willem Waegeman

Tutors: ir. Michiel Stock, Dr. Ramiro Vilchez-Vargas

Thesis submitted in fulfilment of the requirements for the degree of

Master of Science in Bioscience Engineering: Environmental Technology

Faculty of Bioscience Engineering

Academic year 2014-2015

Assessing pathogen invasion based on community evenness and metabolic similarity

Wai Kit Tsang

Promotors: Prof. Dr. ir. Nico Boon, Prof. Dr. Willem Waegeman

Tutors: ir. Michiel Stock, Dr. Ramiro Vilchez-Vargas

Thesis submitted in fulfilment of the requirements for the degree of

Master of Science in Bioscience Engineering: Environmental Technology

Deze pagina is niet beschikbaar omdat ze persoonsgegevens bevat.
Universiteitsbibliotheek Gent, 2021.

This page is not available because it contains personal information.
Ghent University, Library, 2021.

Preface

This topic would never have come to fruition without the help of other people. First and foremost, let me express my gratitude to Prof. Boon, Prof. Waegeman, Michiel and Ramiro for figuring out a topic between both research groups. Several discussions were necessary to delineate the subject and set up a somewhat reasonable framework. Special thanks to Michiel and Ramiro who put in the effort to sit together and converted vague ideas into doable experiments, Ramiro for finding the subject of pathogen invasion and Michiel for solving the issues with the experimental design and providing serious guidance throughout the second semester. I would also like to thank Elham for guiding me with the initial experiments and keeping me company during the flow cytometer measurements, a machine that I would have undoubtedly cursed a lot more when it malfunctioned at unexpected moments. Whenever the experiments did not seem to progress, Prof. Boon was always available for a meeting, many thanks! Prof. Peter Vandamme I would like to thank for pitching the idea of freezing all the synthetic communities beforehand and making my lab life a lot easier, Kim Heylen, thanks, for helping out with the pathogen selection.

Of course, I cannot forget the LabMET-ians roaming around the labs: Benjamin for pointing out some intricacies of the flow cytometer, Jana for showing how to revive pathogen, Tim for sequencing the samples (respect for such labour intensive work). I also learned a lot about efficiency from FM and would like to thank him for showing a quicker way to make pathogen antibiotic resistant. Some people who (unknowingly perhaps) also contributed to my thesis are Aisling who referred me to fundamental literature about diversity, Prof. Thas who provided me with the formula to calculate the Gini-coefficient, Massimo for explaining the rationale behind the Gini-coefficient, Kim for explaining some stuff about databases, Annelies for counting colonies, the members of the EPIC-cluster for valuable feedback. In general I would like to thank LabMET and many others whom I forgot to mention for a pleasant work environment.

Even though a great part of the preface is devoted to LabMET, I would also like to thank

Prof. Waegeman and Michiel for inspiring me and being truly data-driven. Finally, last but not least, thanks to all the awesome people I met throughout my studies and Kim Desplenter in particular for hanging around with me during my master years.

Contents

List of Figures

List of Tables

List of abbreviations

ACK	Acknowledgement
AIC	Akaike Information Criterion
BCCM	Belgian Co-ordinated Collections of Micro-organisms
BIC	Bayesian Information Criterion
Co	Community organization
CCA	Canonical Correlation Analysis
DNA	Deoxyribonucleic acid
FL	Fluorescent Light
FSC	Forward-Scattered light
GUI	Graphical User Interface
LB	Lysogeny Broth
LF	Longpass Filter
LMG	Laboratory Microbiology at the Faculty of Sciences Ghent University
MRM	Microbial Resource Management
PL	Pareto-Lorenz curve
mRNA	messenger RNA
rRNA	ribosomal RNA
RNA	Ribonucleic acid
SG	SYBR GREEN
SSC	Side-Scattered light
WWTP	Waste Water Treatment Plant

Abstract

Synthetic microbial ecology concerns itself with unravelling the principles and mechanisms in ecological theory by assembling synthetic communities with a certain rationale. Here the effect of community evenness on pathogen invasion has been explored. 194 communities have been assembled consisting of ten bacteria, in which ranging over all communities the evenness is optimally varied. Invasion experiments were conducted with four pathogens, but the relation between evenness and invasion was inconclusive. Similarity-based diversity has been evaluated and seems a promising tool in analysing ecosystems. Finally, a basic recommender engine has been built to find suitable bacteria for an ecosystem, which can be adjusted to generate suggestions for synthetic communities.

Samenvatting

Synthetische microbiële ecologie heeft als doel het ontrafelen van ecologische principes en mechanismen door synthetische gemeenschappen samen te stellen met een bepaald rationale. Het effect van 'community evenness' op pathogeeninvasie werd hier onderzocht. 194 bacteriële gemeenschappen werden samengesteld van elks tien bacteriën, waarbij over de verschillende gemeenschappen de 'evenness' optimaal werd gevarieerd. Invasie-experimenten werden uitgevoerd met vier pathogenen. Het verband tussen 'evenness' en de pathogeeninvasie kon echter niet eenduidig aangetoond worden. Een similariteitsgevoelige diversiteitsindex werd hier geëvalueerd en lijkt een veelbelovend hulpmiddel te zijn voor de analyse van ecosystemen. Tenslotte werd er nog een eenvoudige 'recommender engine' geïmplementeerd die geschikte bacteriën kan aanraden voor een microbiële gemeenschap, en die, mits enkele aanpassingen, in staat is om ecosystemen te suggereren.

Chapter 1

Introduction

The presence of pathogens in microbial ecosystems is not always desirable and could be detrimental to the functionality of a biological ecosystem. Where in the past antibiotics were a simple solution to treat pathogens, nowadays such an approach is perhaps not always ideal, because of increasing levels of antibiotic resistance (Reacher et al., 2000; Phillips et al., 2004; Kemper, 2008). Microbial Resource Management (MRM) argues that microbial ecosystems can be managed according to ecological principles and that manipulation of a biological community can increase resistance to biological invasion (De Roy et al., 2013). Synthetic ecology probes deeper into the principles of ecological theory by setting up experiments from a bottom-up approach where bacteria in pure cultures are assembled into a synthetic community (De Roy et al., 2014). Hypotheses can be tested optimally by creating experiments that are maximally adapted to the research questions at hand, giving better insights in ecological theory.

The role of community evenness is analysed and, more specifically, its influence on pathogen invasion. Evenness can be understood as the relative abundance distribution of different species in a community; a community in which species are equally abundant has high evenness, while a community in which species differ widely in abundance has low evenness (Smith and Wilson, 1996).

In a first part of the thesis, lab experiments were performed in which synthetic communties were created with varying evenness and frozen in stock. After that the different communities were invaded with one pathogen each time. The data gathered from the experiments was analysed to assess whether community evenness has a significant influence on pathogens. In a second part, a fairly recent similarity-based diversity index (Leinster and Cobbold, 2012) was tested to check whether it can provide any value to the debate surrounding diversity indices (Jurasinski and Koch, 2011) by applying the index in the analysis of pathogen invasion. Finally, the similarity data was used to explore the possibilities for building a recommender engine in synthetic ecology.

Chapter 2

Indices for describing ecosystems

2.1 Pathogen invasion

Pathogens are recognized as microorganisms capable of causing disease in a host. In the early 20^{th} century, pathogens were attributed microbial functions that could injure host tissues (Smith, 1913; Zinsser, 1915). Although the susceptibility of the host was acknowledged, the focal point in disease research rested primarily on the pathogens' microbial properties (Hoeprich, 1989; Lipsitch and Moxon, 1997; Watson and Brandly, 1949). The role of the environment in disease tranmission, however, is not to be underestimated, as the occurence of diseases can strongly be related to microbial ecology (Casadevall and Pirofski, 1999). The development of type 1 diabetes in infants, for example, is preceded by a decrease in microbial gut diversity (Kostic et al., 2015). In natural soil ecosystems, survival of the invader is known to be negatively correlated with species diversity in soil microbiota (van Elsas et al., 2012).

From an ecological perspective, invasion is defined by two primary steps, namely the transport to a new location (colonisation) and the local increase in abundance of the invasive species (persistence) (Shea and Chesson, 2002). The introduction of invasive pathogenic microorganisms can easily disrupt the functionality of the community (Van der Putten et al., 2007). Not all communities, however, are so easily invaded, as they show different levels of resistance to exogenous species. The community structure is suggested to play a fundamental role in the resistance to invasive species (Dillon et al., 2005).

2.2 Community structure

In order to understand the concept of diversity and its link to pathogen invasion, the most frequent terminology is introduced with regards to community structure. Species richness and evenness are considered key factors in the success of pathogen invasion (Hillebrand and Matthiessen, 2009). Whenever an index describing a community – evenness, diversity or diversity profile – is discussed, each time a fully censused community is considered of S species, with proportions

denoted by $p_1, ..., p_S$ so that

$$p_i > 0 \text{ and } \sum_{i=1}^{S} p_i = 1.$$

The proportions are written as $\mathbf{p}= (p_1, ..., p_S)$.

2.2.1 Species richness

Species richness commonly denotes the absolute number of species in a community (Wittebolle et al., 2009). Often the species richness can be determined straightforward by counting the number of species in a community, and is otherwise estimated (Gotelli and Colwell, 2011). In terms of invasion, the higher the species richness, the less susceptible a community is to invasion (Hooper et al., 2005; Symstad, 2000). Functionally diverse communities are arguably more robust in dealing with invasive species (Jousset et al., 2011; Eisenhauer et al., 2012). Three underlying drivers support this idea: the higher the community's species richness, the higher the probability a species is present with properties that negatively impact the invader, also known as the 'sampling effect' (Huston, 1997; Fargione and Tilman, 2005). In addition, more species in a community result in more possibilities for teamwork, leading to positive interactions. Cooperation to inhibit the invasion by allochtonous microorganisms is also known as the 'facilitation effect' (Hodgson et al., 2002; Bruno et al., 2003). Finally, species may prefer different food resources and living conditions. A more species diverse community leaves less niche opportunities remaining for newcomers. This phenomenon is described as the 'competition effect' (Stachowicz et al., 1999; Naeem et al., 2000; Hodgson et al., 2002; Costello et al., 2012; Shea and Chesson, 2002).

Despite the acceptance of these three mechanisms, the effect of species richness is not that straightforward to observe. In surface water and wastewater processing systems as well as in microbial soil ecosystems, a higher resistance to invasion has been observed in microbial communities with a higher species richness (Van Nevel et al., 2013a; Cook et al., 2006; van Elsas et al., 2012). Other studies report no relation between species richness and invasion resistance (Meiners et al., 2004; Peart and Foin, 1985), while some argue that more species rich communities are more susceptible to invasion (Dunstan and Johnson, 2004; Robinson et al., 1995; Jiang and Morin, 2004). Open systems are for example more prone to invasion because of the high turnover rate (Stachowicz et al., 2002). In very large systems, local patterns can provide opportunities for invasion due to species loss at small scale (Levine, 2000). Possibilities that invaders can work together with native members should also be considered (Stachowicz et al., 2002; Bell et al., 2005).

2.2.2 Evenness

In addition to species richness, another important factor determining community structure is evenness, which basically describes the relative abundance species in a community. When

communities have a low evenness, only one or few species dominate the community. When the abundance distribution of species, however, is more equal, communities have a higher evenness. Uneven communities are considered more fragile, since only a minority of species is dominating the community, leaving more niches available for invasion (Hillebrand et al., 2008). De Roy et al. (2013) showed that species evenness and the invasion probability are negatively correlated. As species richness was deliberately kept constant in her experiments, communities with lower evenness were shown to be more susceptible to invasion than those with higher evenness. The negative effect of evenness on invasion can be attributed to decreased niche availability. Wilsey and Polley (2002) reported similar observations.

Smith and Wilson (1996) reviewed all popular evenness indices and described fifteen desirable criteria for the concept evenness, four of which were considered essential:

1. Evenness should be invariant under replication: it should not change when a dataset is replicated so that each of the species gives rise to n new species of the same absolute abundance as the original one
2. Evenness should decrease when abundance is transferred from a less abundant species to a more abundant one
3. Evenness should decrease when a very rare species is added to the population
4. Evenness should be invariant to scale, so that it depends on the proportional (not absolute) species abundances

In addition several desirable features are mentioned in the literature:

5. The index is maximal when species abundances are equal
6. The maximum value of the index is 1.0
7. The index is minimal, for any number of species, when the species abundances are as unequal as possible
8. Unrealistically uneven communities should not be necessary before the index value is low (arbitrarily defined as 0.05)
9. The minimum value of the index is 0
10. The minimum is attainable with any number of species
11. The index should show a value in the middle of the scale for values we intuitively consider intermediate (arbitrarily defined as 0.25 to 0.75)
12. The index should respond in a reasonable way to a series of communities that changes in evenness (using a series proposed by (Alatalo, 1981)) - “reasonable” taken to mean it should describe a concave function, i.e. the derivative function f’ of a continuous evenness curve f is monotonically decreasing over its domain

13. The index should be symmetric with respect to minor and abundant species (i.e. a community with several abundant species and one minor species should have the same evenness value as one with several minor species and one abundant one)

14. Skewed distributions should give a lower value

The concept of evenness is quantified by various indices. Pielou's evenness is one of the more frequently used measures, which satisfies the four essential requirements. This index is also better known as the Shannon-Wiener index, but rescaled between 0 and 1 (Pielou, 1969). The Shannon-Wiener index is calculated as:

$$H' = -\sum_{i=1}^{S} p_i \ln p_i. \tag{2.1}$$

Pielou's evenness is then calculated as:

$$\begin{aligned} J' &= \frac{H'}{H'_{max}} \\ &= -\sum_{i=1}^{S} p_i \ln p_i / \ln S. \end{aligned} \tag{2.2}$$

To get a better intuition of this concept, an element of the summation $-p_i \ln p_i$ can be plotted within its domain $p_i \in [0, 1]$ see Figure 2.1.

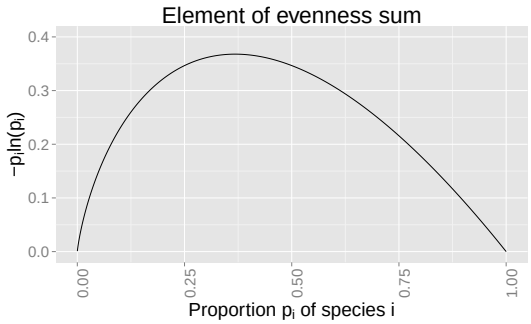


Figure 2.1: Plot of an element of evenness' summation $-p_i \ln p_i$.

In Figure 2.1 it can be seen that whenever species i takes over the culture, thus dominating the community, that $-p_i \ln p_i$ decreases. (When talking about 'dominating', the author means that one species occurs 10 to 1000 times more.) When a limited cluster of species dominates, the remaining proportions of other species are negligible in the Shannon-Wiener Index so that the evenness decreases and a community is regarded as uneven. Contrary, when a bacterial culture contains species that are relatively equal in abundance to eachother, none of the elements in the summation are negligible and the Shannon diversity index reaches its maximum. Some examples of community distributions with low to high evenness can be found in Figure 2.2. In cultures

with low evenness typically one bacteria has taken over the community, whereas in case of high evenness all species are relatively equal in abundance.

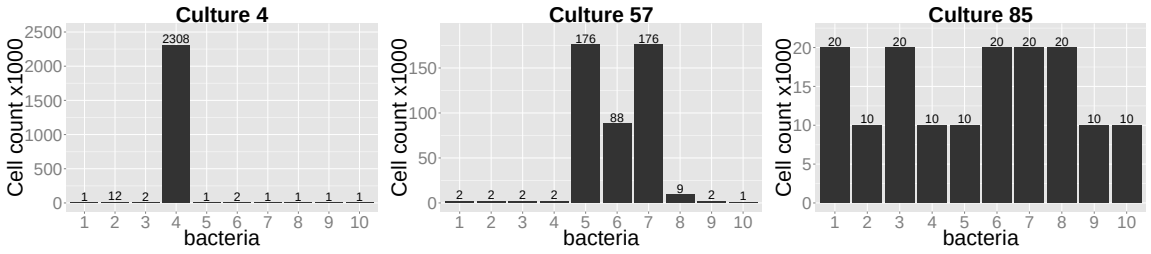


Figure 2.2: Bacterial cultures with low evenness in the left plot, medium evennes plotted in the center and high evenness on the right plot. On the x-axis are ten bacteria. The y-axis describes the total cell count for each bacteria in each culture.

Albeit not the most popular measure in ecology, the Gini-coefficient can also be used to quantify evenness. This measure is graphically represented in Lorenz-curves (Rousseau et al., 1999). The Gini-coefficient quantifies a community deviation from the most even community as the surface between the perfect evenness diagonal and the Lorenz curve. The more uneven a community is, the larger the surface between the perfect evenness diagonal and the Lorenz curve. Note how Pielou's evenness describes a perfectly even community with a score of one, while the Gini-coefficient has a score of zero, as the Lorenz-curve collapses on the diagonal.

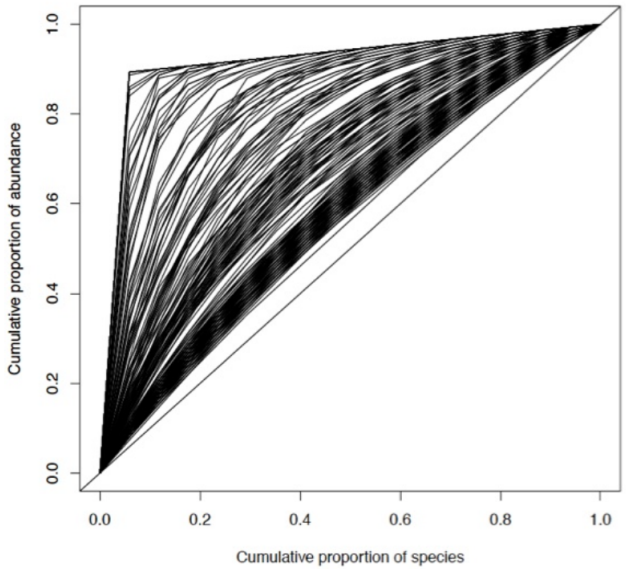


Figure 2.3: Examples of Lorenz curves (De Roy, 2014).

Pielou and Gini are more or less opposites, but there are more evenness indices to be found in the literature (Tuomisto, 2012). Choosing a specific evenness definition for a dataset, however, may not be so transparent, since each definition has its own subtleties (Beisel et al., 2003). There is no general agreement on which index is most fitting for ecological problems. Some prefer Pielou's definition (Zhang et al., 2012; Pommier et al., 2010), others opt for Gini's coefficient (Wittebolle et al., 2009), while still others choose one of the other numerous eveness indices (Tuomisto, 2012).

2.2.3 Species Diversity

The origins of diversity

Whereas species richness generally denotes the total number of species in a community (Wittebolle et al., 2009) and evenness in general the relative abundance of species in community ecology (Smith and Wilson, 1996), species diversity incorporates both richness and evenness (Pielou, 1977). When comparing ecosystem biodiversities, species diversity is often used as an aid (Hill, 1973). Despite repeated efforts to come to terms with the terminology (Ricotta, 2005), applying the concept of species diversity is not always transparent, since each group of authors insists on a different usage (Jurasinski and Koch, 2011). The history of species diversity starts with the terms *alpha, beta* and *gamma* diversity coined by Whittaker (1960) to describe the vegetation of the Siskiyou mountains. The three new diversities serve as a means to describe vegetation diversity on different levels: (i) *alpha-diversity* "richness in species of a particular stand or community", i.e. the species diversity within *one* place (ii) *beta-diversity* "extent of change of community composition", or in other words the change or difference between two different communities and (iii) *gamma-diversity*, the "species diversity of a number of community samples", which can be understood as a combination of *alpha-diversity* and *beta-diversity* (Whittaker, 1960, 1972).

A simple example will show how these three diversities relate to each other and how they can be applied. *Alpha-diversity* in its most strict meaning denotes the species diversity within a sample, which is also commonly called a site, a sampling unit or a plot depending on the field of research. *Gamma-diversity* describes, similarly to *alpha-diversity*, the species diversity within one area, but at a higher aggregational level combining several samples within one area (Figure 2.4). *Alpha-* and *gamma-diversity* thus basically describe the species within one area, but on a different scale, the former within one sample and the latter within an investigation area consisting of different samples (Whittaker, 1960). Concretely, there exist two measures to describe *alpha-* and *gamma-diversity*. It can be simply defined as the sampled or estimated species richness, based on samples or species accumulation curves (Gotelli and Colwell, 2001). Alternatively, it can describe the species-abundance distribution, for example indices of Simpson (1949), Shannon and Weaver (1949) or Fisher's alpha (Fisher et al., 1943).

In general, *beta-diversity* is used to describe the overall difference between two units (Condit et al., 2002; Koleff et al., 2003; Chave, 2004; Chust et al., 2006). Originally, Whittaker (1960) defined the relation of *beta-diversity* multiplicatively in terms of *alpha-* and *gamma-diversity* as:

$$\beta = \frac{\gamma}{\alpha}. \tag{2.3}$$

The inverse of β as in formula (2.3) describes the proportion of species richness found in an average sample, serving indirectly as a similarity measure between species compositions. When all samples share the same species compositions, β becomes 1 and if each species only appears in one single unit β approaches $1/S$. Since its conception, several new β-definitions have been introduced, such as additive partitioning (Veech et al., 2002), resemblance by (dis)similarity

coefficients (Jaccard, 1901; Sørensen, 1948), slope of the distance decay relationship (Whittaker, 1960), sum of squares of a species matrix (Legendre et al., 2005) or gradient length in ordination space (Whittaker, 1956). That is to say, there are more than 30 explanations for β, the one more obscure than the other (Tuomisto, 2010b), and this could be easily seen as the source of confusion in the current debate. For clarity and simplicity's sake, the multiplicative formulation (2.3) for β will therefore be used for the remainder of the chapter.

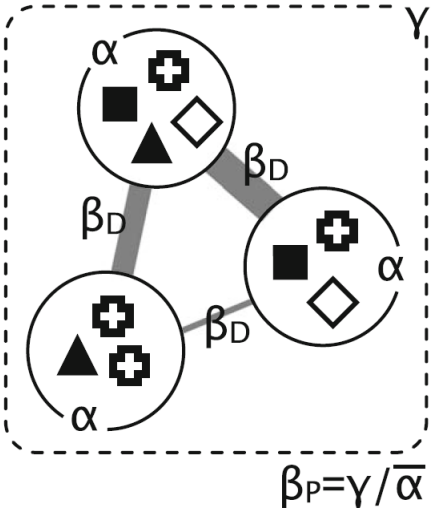


Figure 2.4: Example of *alpha, beta* and *gamma-diversity*. The circles are vegetation samples with different species, each depicted by a symbol. The *gamma-diversity* is the entire set of species over all the samples. *Alpha-diversity* is here simply shown as species richness in a sample. Illustrating β-diversity is not that easy due to its multifaceted nature. Differentation diversity (β_D) is shown here as pairwise similarity between samples (the broader the gray bar connecting two samples, the higher the similarity). The proportional diversity (β_P) is the relation between the two scales of investigation so that $\beta_P \neq \beta_D$ (Jurasinski et al., 2009).

The partitioning theorem

How do *alpha-, beta-* and *gamma*-diversity exactly relate to the aforementioned species richness and evenness? The multiplicative law by Whittaker (1960) states that any species diversity can be partitioned into two seperate components:

$$\gamma = \alpha \cdot \beta. \tag{2.4}$$

Here, species diversity, or in Whittaker's terminology the γ-diversity, can be split into two components: species richness and evenness (Pielou, 1977). Species richness is considered to be the α-diversity, and evenness the β-diversity or the difference across the different samples. The vocabulary of the Greek diversities enables a discussion concerning the difficulties in the selection of species richness and evenness on a higher level, as certain properties for species diversity are also relevant for evenness. Concretely, one of the fundamental and necessary pillars in the diversity debate states that species evenness should be unaffected by species richness (Heip, 1974). Jost (2007) recognised, however, that the multiplicative law (2.4) does not directly hold

for all species diversities at the individual level. Hill (1973) already defined a solution for the lack of this universal multiplicative property in each species diversity with the so-called Hill-number or "true diversity":

$$ {}^{q}D(\mathbf{p}) = \Big(\sum_{i=1}^{S} p_i^q \Big)^{1/(1-q)}, \tag{2.5} $$

where q is the order of the diversity index and $q \neq 1$; for all indices that are functions of $\sum_{i=1}^{S} p_i^q$, the true diversity depends only on the value of q and the species frequencies p_i, and *not on the functional form of the index*. In other words, all the most common diversity indices can be converted to one of the Hill diversity indices with order q. A couple of examples could elucidate its flexibility. Independent whether one uses Simpson concentration, inverse Simpson-concentration or the Gini-Simpson index, all give the same diversity:

$$ {}^{2}D = 1/\Big(\sum_{i=1}^{S} p_i^2 \Big), \tag{2.6} $$

where the superscript 2 on the diversity indicates that this is a diversity of order 2 (Jost, 2006). When the diversity has an order zero (q=0), it is completely insensitive to species frequencies and is reduced to species richness. A special case exists where Eq. (2.5) is undefined when q=1 but its limit can be calculated and equals

$$ {}^{1}D = \exp\Big(-\sum_{i=1}^{S} p_i \ln p_i \Big) = \exp(\mathrm{H}), \tag{2.7} $$

which can be recognized as the exponential of Shannon entropy. The proof for this particular case can be found in Appendix A.1. An overview of how common diversity indices can be converted to true diversities can be found in Table 2.1.

Table 2.1: Conversion of common indices to true diversities (Jost, 2007).

Index H	**Diversity in terms of** H	**Diversity in terms of** p_i
Species Richness $H = \sum_{i=1}^{S} p_i^0$	H	$\sum_{i=1}^{S} p_i^0$
Shannon Entropy $H = -\sum_{i=1}^{S} p_i \ln p_i$	$\exp(H)$	$\exp(-\sum_{i=1}^{S} p_i \ln p_i)$
Simpson concentration $H = \sum_{i=1}^{S} p_i^2$	$1/H$	$1/\sum_{i=1}^{S} p_i^2$
Gini-Simpson index $H = 1 - \sum_{i=1}^{S} p_i^2$	$1/(1-H)$	$1/\sum_{i=1}^{S} p_i^2$
HCDT entropies $H = (1 - \sum_{i=1}^{S} p_i^q)/(q-1)$	$[(1-(q-1)H)]^{1/(1-q)}$	$(\sum_{i=1}^{S} p_i^q)^{1/(1-q)}$
Renyi entropies $H = (-\ln \sum_{i=1}^{S} p_i^q)/(q-1)$	$\exp(H)$	$(\sum_{i=1}^{S} p_i^q)^{1/(1-q)}$

Using the property that allows conversion of common diversity indices into "true diversity" (2.5) Jost (2007) formulated the partitioning theorem which states that if a diversity index or compositional complexity consists of two independent components, then the numbers equivalent of the measure must equal the product of the numbers equivalent of the first component and the numbers equivalent of the second component. Let D be a number equivalent of true diversity, H_{tot} a diversity index, which can be decomposed into H_A and H_B so that the partitioning theorem can be written as:

$$D(H_{tot}) = D(H_A) \cdot D(H_B). \tag{2.8}$$

With Formula (2.8), true diversity can be decomposed in a product of independent components. Writing out these decompositions, the formulas have different expressions (Table 2.2). To convert a common diversity H to its true diversity, Table 2.1 can be consulted. Examples of the working principle to derive these formulas can be found in Appendix A.1.

Diversity cannot be decomposed into independent components

Jost (2010), though, pointed out that decomposition of diversity into independent richness and evenness components is mathematically impossible. Consider the true diversity Eq. (2.5), where the limit of $q \to 1$ and the Hill number boils down to the exponent of the Shannon index (2.1):

$${}^1D = \exp\Big(-\sum_{i=1}^{S} p_i \ln p_i \Big) = \exp(H_{Sh}) \tag{2.9}$$

The exponent of the Shannon index can be rewritten as the product of species richness S and an undetermined evenness factor $EF_{0,1}$:

$$e^{H_{Sh}} = S \cdot EF_{0,1} \tag{2.10}$$

When decomposing the exponent of Shannon entropy $e^{H_{Sh}}$, however, the evenness factor $EF_{0,1}$ becomes restrained by the species richness S as shown in Eq. (2.11). Since the minimum value of $e^{H_{Sh}}$ is 1 and its maximum value the number of species S, the evenness factor $EF_{0,1}$ will have a value between $1/S$ and 1.

$$EF_{0,1} = \frac{e^{H_{Sh}}}{S} = \frac{{}^1D(H_{Sh})}{S} \tag{2.11}$$

In consequence, decomposition of diversity into richness and evenness leads to a violation of the partitioning theorem, which states in this particular case that *if* components S and $e^{H_{Sh}}$

Table 2.2: Examples of the partitioning theorem.

Diversity index	**formula**
Species Richness	$H_A \cdot H_B = H_{tot}$
Shannon Entropy	$H_A + H_B = H_{tot}$
Exponential of Shannon Entropy	$H_A \cdot H_B = H_{tot}$
Gini-Simpson index	$H_A + H_B - (H_A \cdot H_B) = H_{tot}$
Simpson concentration	$H_A \cdot H_B = H_{tot}$
HCDT entropies	$H_A + H_B - (q-1) \cdot (H_A) \cdot (H_B) = H_{tot}$
Renyi entropies	$H_A + H_B = H_{tot}$

are independent, that neither factor puts a constraint on each other (Jost, 2007). So instead of partitioning species diversity, Jost (2010) proposes to partition richness:

$$\text{Richness} = \text{Diversity} \times \text{Unevenness}, \tag{2.12}$$

where Diversity equals equation (2.5) so that equation (2.12) can be rearranged as:

$$\text{Unevenness} = \text{Richness}/\text{Diversity} = 1/\text{Evenness} \tag{2.13}$$

so that the inverse still becomes:

$$\text{Evenness} = \text{Diversity}/\text{Richness} \tag{2.14}$$

Although equations (2.12 - 2.14) are logically equivalent, the starting point for the derivation of evenness meets the requirement of numerical independence for its components. For each step in the derivation, it can then be checked that the theoretical requirements are still met. Popular evenness measures are therefore more suited to analyse a community, as they satisfy the condition of numerical independence.

2.2.4 Similarity based diversity

Hill (1973) already searched for a continuum of possible diversity measures. Building further on the so-called Hill numbers (i.e. the effective numbers), Jost (2010) derived several useful evenness and inequality measures from species richness and diversity Eq. (2.5). Depending on the order q of the diversity ${}^qD$, Eq. (2.5) gains different meanings. While Tuomisto (2010a) establishes the possibility for consistency in relating different ecosystems, Leinster and Cobbold (2012) tie all the metrics together into a convenient tool by proposing a graphical representation of diversity profiles.

The proposed diversity profile takes two inputs: (i) relative abundance data describing the proportions $\mathbf{p}$ in which they are present and (ii) similarity data $\mathbf{Z}$, for each pair of species, a number Z_{ij} specifying how similar they are. The similarities between species are in consequence encoded in an $S \times S$ matrix $\mathbf{Z} = (Z_{ij})$, with Z_{ij} measuring the similarity between the ith and jth species. It is assumed that $0 \leq Z_{ij} \leq 1$, with 0 indicating total dissimilarity and 1 indicating identical species so that $Z_{ii} = \mathbf{I}$ is assumed and distinct species have nothing in common. Although most simple examples of similarity matrices are symmetric ($Z_{ij} = Z_{ji}$), symmetry is not part of the definition of a similarity matrix. Given these two inputs, the true diversity for *every* q is calculated, and plotted against q.

Leinster and Cobbold's similarity-based diversity is defined as:

$$ {}^{q}D^{\mathbf{Z}}(\mathbf{p}) = \Big(\sum p_i (\mathbf{Zp})_i^{q-1} \Big)^{\frac{1}{1-q}}, \tag{2.15} $$

where q is the community's order and $q \neq 1$ (but its limit $q \to 1$ exists) so that

$$ (\mathbf{Zp})_i = \sum_{i=1}^{S} Z_{ij} p_j. $$

Instead of a single measure to indicate a community's diversity, a family of measures can be used: ${}^{q}D^{\mathbf{Z}}(\mathbf{p})$, for each value of the sensitivity parameter q (Table 2.3).

The *diversity profile* of a community is the graph of ${}^{q}D^{\mathbf{Z}}(\mathbf{p})$ against q. The sensitivity parameter q can be understood as the *in*sensitivity to rare species. The higher the q parameter becomes, the less weight rare species have (Leinster and Cobbold, 2012). An example of such diversity profiles is shown in Figure 2.5, where the "naive" identity matrix is used, i.e. all species are maximally dissimilar from each other.

2.3 Incorporating species similarity

A 'naive' similarity matrix $\mathbf{Z} = \mathbf{I}$ assumes that distinct species have nothing in common. Depending on the availability and relevance of information, various similarity matrices can be used. Different similarity matrices lead to different diversities, allowing for a very flexible analysis depending on the characteristic of importance. Very crude taxonomic similarity matrices can be created (Warwick et al., 1995). Shimatani (2001) calculated not only conventional indices without incorporating species relatedness, but used available taxonomic information so that a better interpretation was achieved of the effects of thinning on contrasting vegetation structures.

Hughes et al. (2008) revealed significant effects in literature of genetic diversity on ecological processes such as primary productivity, population recovery from disturbance, interspecific

Table 2.3: Overview for varying orders of diversity (Leinster and Cobbold, 2012).

Notes: For the the three entries marked "compare," see the Appendix. Key to abbreviations: CCJ, Chao-Chiu-Jost; HSG, Hurlbert-Smith-Grassle; AKB, Allen-Kon-Bar-Yam. References (in the sequence shown in the table, left to right, then top to bottom): Faith (1992), Rao (1982*a*, *b*), Chao et al. (2010), Hill (1973), Simpson (1949), Berger and Parker (1970), Hurlbert (1971), Smith and Grassle (1977), Ricotta and Szeidl (2006), Allen et al. (2009), Patil and Taillie (1982), Tsallis (1988), Gini (1912).
† This quantity is undefined.

	Sensitive to rare species	⟷		Insensitive to rare species	
Diversity measure	$q = 0$	$q = 1$	$q = 2$	$q = \infty$	Remarks
Diversity ${}^qD^{\mathbf{Z}}$	compare Faith		$\frac{1}{1 - \text{Rao}}$		compare CCJ
Naive diversity ${}^qD = {}^qD^{\mathbf{I}}$ (Hill numbers)	species richness	exp(Shannon)	inverse Simpson concentration	$\frac{1}{\text{Berger-Parker}}$	${}^2D, \ldots, {}^mD$ give HSG
Entropy ${}^qH^{\mathbf{Z}}$ (Ricotta-Szeidl)		compare AKB	Rao's quadratic entropy	†	
Naive entropy ${}^qH = {}^qH^{\mathbf{I}}$ (Patil-Taillie-Tsallis)	species richness minus 1	Shannon entropy	Gini-Simpson	†	

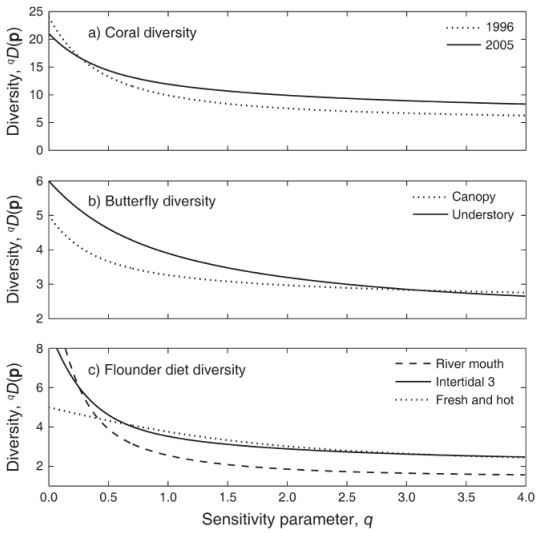


Figure 2.5: Examples of diversity profiles using naive similarity matrix $\mathbf{Z} = \mathbf{I}$ (Leinster and Cobbold, 2012). (a) Ecosystem of coral cover on reef from 1996 to 2005. Diversity profiles cross so that it is not possible to unambiguously say which community is more diverse (Riegl et al., 2009). The species richness ($q = 0$) decreased in 2005, but the community became more even for $q > 1$. (b) Butterfly diversity is more diverse overall in the understory except for high q where Canopy dominates the curve, i.e. when dominance in the community plays a more important role than the presence of a couple of rare butterflies (DeVries et al., 1997). (c) Diversity profiles for diet flounders where one pair-profile crosses twice. Which community is more diverse depends on the sensitivity parameter q (Wirjoatmodjo, 1980). More info can be gained by using different similarity matrices $\mathbf{Z}$.

competition, community structure, and fluxes of energy and nutrients. Genetic information can be easily implemented in similarity matrices. Petchey and Gaston (2006) reviewed how functional diversity can explain and predict the impact of organisms on ecosystems and thereby provide a mechanistic link between the two. The choice of functional traits with which organisms are distinguished and how the diversity of that trait information is summarized into a measure of functional diversity determines the outcome of the similarity matrix.

Construction of such similarity matrices does not need to be difficult. Simply associating data to

each species regarding its characteristics and calculating the corresponding similarity coefficients Z_{ij} in terms of some notion of difference between the associated data is sufficient for usage in diversity profiles (Leinster and Cobbold, 2012).

2.4 Understanding microbial communities

When crossing the bridge from microbial ecology to practical experiments, synthetic microbial communities can play an important role in the development of a framework. By intelligently devising synthetic microbial consortia, the behaviour of individual microbial populations has been programmed for useful projects over the last years. Synthetic microbial ecology has proven its practical value in several fields over the last couple of years, in furthering co-fermentation, for example (Chen, 2011), in biofuel exctraction (Pandhal and Noirel, 2014) or in waste water treatment plants (Ho et al., 2014). One aspect, though, where its importance cannot be underestimated is in developing understanding in microbial ecology.

Prosser et al. (2007) suggest that new insights and techniques should be guided by a theory-driven approach. Deliberately selected experiments with synthetic microbial communities could aid in the verification of hypotheses, thus accelerating theory generation. By carefully making an experimental design, utilizing specific methods, and assembling microbial communities, it is possible to exploit the full potential of microbial ecology.

Contrary to synthetic consortia, natural microbial communities easily contain more than thousand species (Curtis et al., 2002), which makes it experimentally a very challenging to nearly impossible task to determine which species are key drivers in a community. In addition, natural ecosystems are influenced by different effects depending on their surroundings. In a typical open system, like a waste water treatment plant, random colonisation effects strongly dominate (Ofiteru et al., 2010), while in more closed systems a more stable system occurs after a certain period (Costello et al., 2012). These factors can limit explorations in microbial theory. By artifically assembling a number of selected species in initially well-defined media, key features of a natural community can be retained (Großkopf and Soyer, 2014). With a significantly decreased number of species, complexity is reduced and controllability increased, so that synthetic communities are often preferred over complex communities to investigate ecological theories (De Roy et al., 2014).

The tools developed to understand the workings of microbial communities originate from a desire to control them, as more and more global challenges were fully grasped such as the production of green house gases (Crowley, 2000) and the supply of drinking water (Shearer, 2005). The need to manage microbial resources in order to steer and control mixed cultures and communities was quickly recognized and the term Microbial Resource Management (MRM) was coined (Verstraete et al., 2007). This concept tackles fundamental questions such as ‘what species are present’, ‘who is doing what with whom’ and ‘how abundant is each species?’ These questions easily blend into a methodology to gain a deeper understanding of microbial communities. Going beyond basic interactions between individual cells (Little et al., 2008; Klitgord and Segrè, 2010), Marzorati et al. (2008) initially defined microbial community parameters in the framework of Microbial

Resource Management as tools to steer communities.

2.5 Research Objectives

The goal of this thesis is researching pathogen invasion in function of community evenness and similarity among species. The set-up for this thesis consists of two parts. A first part will deal with a set-up in the lab, where synthetic communities are created in function of evenness and afterwards invaded with pathogens. The quantified pathogen growth in the communities serves as a means to analyse the effects of evenness on the pathogen invasion. In a second part, pathogen invasion will be further explored by looking into databases for additional information and combining these into the analysis with the diversity profiles. Different similarity measures will be checked for their effect.

Chapter 3

Material and Methods

3.1 Experimental Set-up

Before entering the lab, an experimental design is proposed to answer the research questions as clearly and efficiently as possible. Such a design identifies which specific questions are at stake, and, more importantly, deals with the analysis of data generated from the experiments. Sources of variability outside the scope of the research question are acknowledged and reduced as much as possible. Conditions will be adapted to increase the validity of the experiment. In addition, causal inferences can be drawn between dependent and independent variables and possible confounding factors can be removed preemptively.

Here, the design consists of two perspectives, namely a microbiological aspect involving statistical analysis and a part about predictive modelling. From a microbiological point of view, the objective is to test whether an ecological community's evenness significantly influences the invasion potential of pathogens. For the modelling part, the data of the pathogen invasion and the assmebled communities in the experiments will be assessed with similarity matrices based on information about the respective bacteria.

Both perspectives demand different conditions for their optimal design. In case of the predictive model, samples have to cover sufficient points of the feature space so as to reflect the diversity of the cocultures, while from a microbiological angle the samples have to be selected in such a manner that a significant difference can be tested between pathogen invasion in communities with low and high evenness. As the statistical conditions for the microbial test are more strict than those for predictive models, the concept of evenness will be explained, and afterwards the requirements for training a predictive model will be discussed.

As shown by Jost (2010) Pielou's evenness Eq. (2.2) possesses all four desirable traits and it can be decomposed into independent components. As such the synthetic communities will be set-up in function of the evenness. Only after the results for the experiments are obtained will the similarity matrices be included to assess their impact on the analysis.

The microbiological question now is whether the success rate of pathogen invasion will depend on this concept of evenness. To clarify the subject, the null-hypothesis can be formulated unambiguously as:

$$H_0 : \mathcal{P}(\text{Invasion} \cap \text{evenness} = \text{High}) = \mathcal{P}(\text{Invasion} \cap \text{evenness} = \text{Low}). \tag{3.1}$$

To isolate the variable evenness, the cultures are created with a fixed number of species, which is in our case ten bacteria, so that no confounding occurs between the number of species and the evenness. The variance introduced in the cultures are defined by each bacteria's total cell count. In addition to the fixed number of species per culture several other variables have to be controlled. The cell count has to be uncorrelated with evenness for bacterial communities over the entire domain, otherwise confounding might occur between these two variables. With regards to practical execution of the experiment, pipetting complexity also has to be limited in order to reduce the error rate. Changing of pipetting volumes has to be limited and the number of concentrations which are to be made is constrained. Since the cell count is quantified by using flow cytometry, the bacteria used are refreshed 24 hours prior to the quantification, which places an upper bound on the maximum starting concentration. As the cultures to be prepared are frozen and diluted 100 times afterwards, the initial concentrations also have a lower limit. Furthermore, the distribution of bacterial cultures has to answer questions from a statistical sound perspective, as well be used for predictive modelling. A two factor design would be ideal for testing the difference in pathogen invasion between high and low evenness communities, while for the latter randomly distributed samples would be optimal. The final result tries to take into account all mentioned pitfalls.

Two initial pipetting volumes 400 and 800 µL were chosen for the experiment with four concentrations $10^7, 10^6, 10^5$ and 10^4 cells/mL. Opting for only two possible volumes per bacteria controls the number of volume switches per culture. One million bacterial cultures were simulated from these solutions. 10^4 cells/mL approaches the lower limit of workable concentrations. After 100x dilution, 100 cells/mL hovers on the edge of quantifiable cell count. For each community evenness and cell count ('cellcount') were calculated. The cell count is plotted against the evenness in Figure 3.1.

From Figure 3.1 there appears to be no constant cell count that covers the entire evenness domain. Since the data points are not randomly scattered over the plot, sampling from this initial pool gives way to a correlated set-up where distinguishing between evenness and cell count will be difficult. To avoid such confounding dangers, choosing for a block design is a feasible way to decorrelate these two variables. A degree of randomness still has to be applied. Microbially, the priority lies in checking for a difference in pathogen invasion between communities with high and low evennness, which ideally can be tested with a two factor design. Aiming to use the data additionally for predictive modelling, though, will require more random data. A compromise between all these considerations is a combination of stratified and randomly selected cultures. For the experiments, a total of 200 data points struck a good balance between the practical

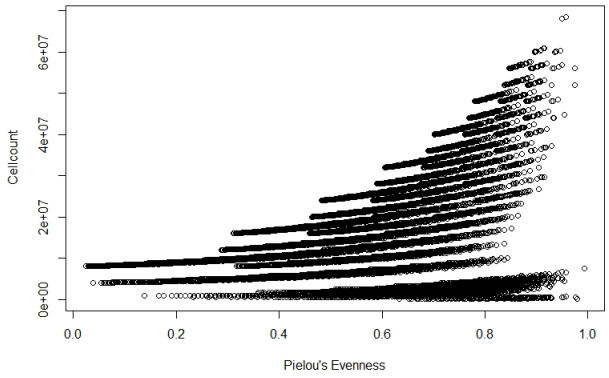


Figure 3.1: One million cultures on cell count vs evenness plot.

workability and the data requirements. Half of the data points were selected according to a certain stratification while the remaining fraction was chosen randomly. The stratification boils down to 40 cultures with low evenness, 40 cultures with high evenness and 20 ones with medium evenness. These partitions were distributed between high and low cell count so as not to find a correlation with evenness. 40 cultures were selected with high cell count, 40 with low cell count and 20 with a medium cell count. The idea is visualized in Figure 3.2 where the final design equals a combination of 100 cultures sampled from a certain stratification and 94 cultures randomly chosen.

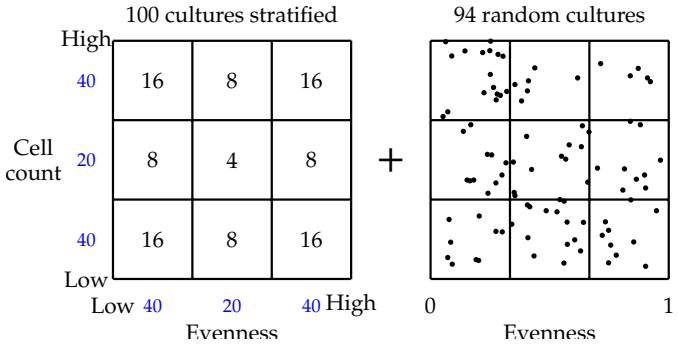


Figure 3.2: 100 cultures stratified and 94 random samples.

A subset in the plot has to be found where cell count is decorrelated from the evenness. All duplicate samples are removed, yielding Figure 3.3. The volumes and cell count are normalized to 1500 µL for the 2 mL cryovials. Figure 3.4 shows that the subset of the plot with cell count lower than 1.5 million total cells, contains data points in each quadrant. Slicing the pool of data points to this subset will decrease the number of concentrations that need to be prepared as the cell count goes down. After this the data points can be selected according to the design in Figure 3.2. In Figure 3.5 the evenness of the randomly generated cultures describes a bell-shaped curve. Figure 3.6 shows how the bacterial cultures are distributed in terms of evenness. Around an evenness of 0.5, a bell-shaped distribution can be found, while at the highest and the lowest evenness cultures are increased. In Figure 3.7 it can be checked whether the variable evenness is sufficiently decorrelated from the total cell count. This final distribution nicely combines a two factor design and a Gaussian evenness distribution, while maintaining a fairly orthogonal design with a correlation of -0.2402 between the total cell count and evenness.

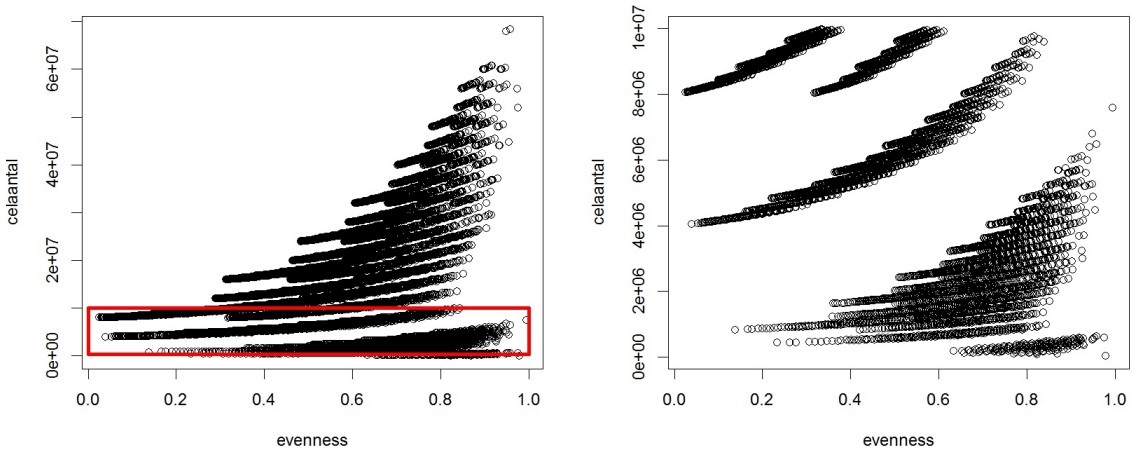


Figure 3.3: One million cultures are simulated in the left plot. The red box indicates from which region the data is sampled. The right plot shows a close-up of the red rectangle in the left plot.

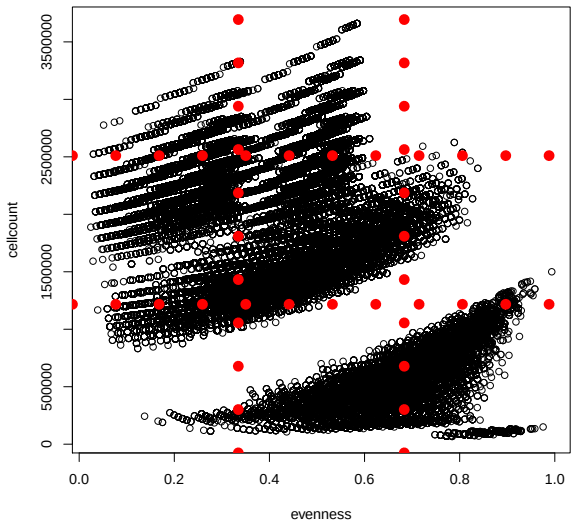


Figure 3.4: The points are represented in most of the quadrants where cell count is lower than 1.5 million. The dotted red lines represent the division in the quadrants.

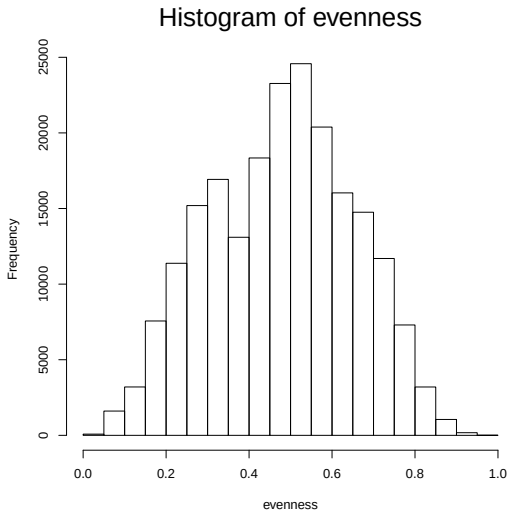


Figure 3.5: The evenness distribution that is going to be sampled.

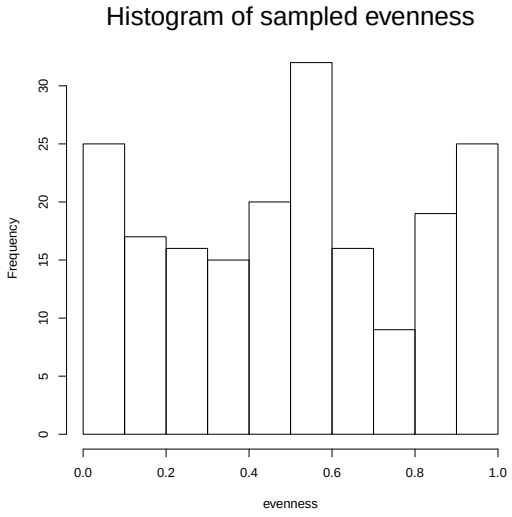


Figure 3.6: The evenness distribution with 100 cultures stratified and 94 random.

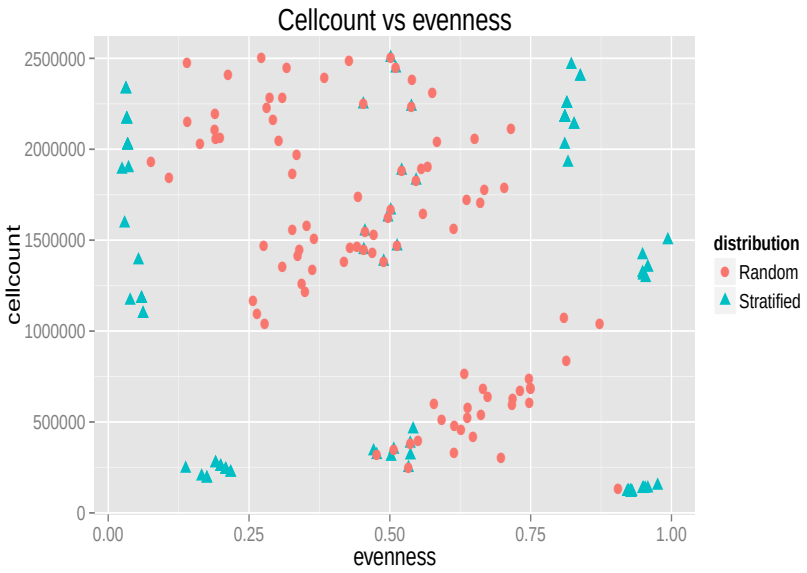


Figure 3.7: The experimental design's final distribution with 100 cultures stratified and 94 random.

3.2 Selection of the bacteria

For assembling the synthetic communities, the bacteria were chosen based on preliminary research. Elham Ehsani (personal communication, August 2014) quantified the growth curves via flow cytometry from an initial pool of 110 bacterial isolates from different envrionments such as soils, underground, water, and sand filters in bioreactors. In order to maximize the potential for a stable synthetic community, ten bacteria shown in Table 3.1 were selected based on similarity between growth curves. Since the bacteria are repeatedly quantified via flow cytometry, the selection procedure also takes into accounts the ease of separability between the noise and the cells during the gating of the FL1-FL3 graphs discussed in Section 3.3.3. For the creation of the communities, the same bacteria are each time used to fix the species richness, i.e. the number of species present in a community (Marzorati et al., 2008). The abundance of each species, however, varies from community to community, as described in Section 3.1. To prepare the communities with the correct cell count, the concentration of cells has to be determined in the pure cultures, a dilution series is to be determined and each community is created by carefully pipetting the correct volumes into a recipient.

Table 3.1: The ten bacteria chosen for creating the synthetic communities. The organisms were selected based on similarity in growth curves via flow cytometry. 16S rRNA sequences ensured that neither of the ten bacteria have the same sequence of nucleotides in the 16S rRNA-region.

Selected bacteria
Pseudomonas sp.
Bacillus sp.
Serratia sp.
Burkholderia Cepacia
Paracoccus sp.
Enterococcus sp.
Agrobacterium sp.
Rhizobium Duejenense
Delftia sp.
Aeromonas sp.

To determine the concentration of cells in each pure sample, the bacteria are refreshed each time 24 hours beforehand in LB Broth medium and incubated in dark conditions at T = 28 °C.

3.3 Flow Cytometry

3.3.1 Principle

In order to quantify the cell concentration in the pure samples, the flow cytometer is used, an optical instrument based on lasers and detectors. A liquid sample gets pumped through the machine and goes through the flow cell via a capillary. The capillary is extremely thin, so that any particles go one by one through the flow cell. As such flow cytometry can be applied to quantify bacteria (see Figure 3.8). Particles are detected based on Forward-Scattered light (FSC) and Side-Scattered light (SSC), the properties of which are related to the cell content (See Figure 3.9). In the flow cell, they pass across one (or multiple) laser beams. The scattered light is collected at different wavelengths, revealing information on different cell properties. When a laser beam collides with a cell, the characteristics and the wavelength of the scattered light is determined by the cell membrane and the cell surface granular structures. Cells are stained with fluorescent stains (fluorochromes). If the molecules of these stains get excitated by a specific wavelength of the laser, they emit light, which can be detected by the appropriate detector (BD Biosciences, 2000).

For high-throughput samping the BD Accuri^{TM} C6 flowcytometer (BD Biosciences) autosampler was used with the accompanying BD Accuri^{TM} C6 software. A 488 nm argon ion laser is used for excitation. Photomultiplier tubes, photodiode dectors detect green and red fluorescen with resp. 530/28 nm (FL1) and a 670/LP longpass filter (LF3). All samples are collected on a logarithmic scale. Milli-Q water serves as mantle liquid. The samples are measured on instrumental settings with a standard filter set (Table 3.2) (Van Nevel et al., 2013b).

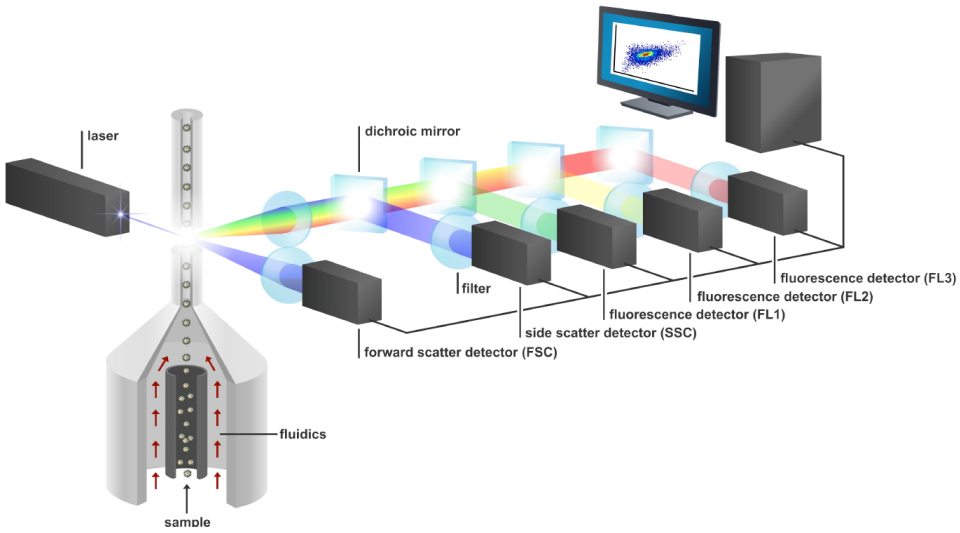


Figure 3.8: Working principle of a flowcytometer. A liquid sample is pumped through a flow cell via a capillary to the centre of a laser. The light collides with the particles causing scattering of the light, which is received by different detectors (De Roy, 2014).

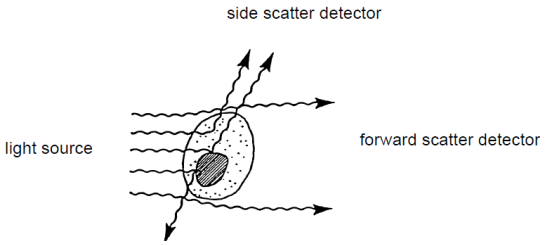


Figure 3.9: Light scattering properties of cells in flow cytometry. Forward-scattered light is proportional to particle size and Side-scattered light is proportional to cell granularity or internal complexity (BD , 2000).

Table 3.2: The flow cytometer settings during autosampling.

Instrument settings	
FL1: 533/30 (green)	blue laser (488 nm)
FL3: 670 LP (far red)	blue laser (488 nm)
Run limits	time: 30 seconds
Fluidics	medium speed
Primary threshold	FL1: 500
Secondary threshold	/
Wash settings	None
Agitate plate	None
Measuring speed	Fast
Measurement volume	25 µL

3.3.2 Staining with SYBR Green

The total cellcount can be determined by SYBR Green (SG) staining with flow cytometry. SYBR Green stains every cell and is therefore suited for a total cell count. However, no distinction can be made between intact and damaged cells. SG enters all cells, binds to the DNA and gets excited by the blue 488 nm laser. The stained cells typically show similar green (FL1) and red (FL3) fluorescence and are detected in a FL1-FL3 graph. Since only the living cells are of interest for creating the communities, even after refreshing the bacteria 24 hours beforehand, using flow cytometry serves only as a rough estimation of the total cellcount for each bacteria in a community. Before invading the communities the exact abundance is to be estimated with a more accurate method.

The staining of the pure cultures was executed in 96-well plates with a final volume of 200 µL. The different reagents were added with a multichannel pipet. In preparing the samples these are diluted 1000 and 5000 times by making a dilution series of 10, 100 and 1000. For a 10 times dilution while staining, 178 µL of physiological solution (0.85 % NaCl), 20 µL of diluted sample and 2 µL Sybr Green are added together. After preparing the 1000 and 5000 times dilutions in four fold, these are incubated for 13 min. at T = 37 °C, after which they are quantified in the flow cytometer (De Roy et al., 2012).

3.3.3 Data interpretation: flow cytometry gating

In case of SYBR Green staining, the two parameters of interest are FL1 (green fluorescence) and FL3 (red fluorescence). The most important step during the post-processing of flow cytometry data is defining the regions in which cells or particles with different characteristics are located. In case of SYBR Green staining these regions are: (i) noise (Organics, inorganics and electric noise) and (ii) cells.

Background is generally located below a fluorescence intensity of 1000 in the fluorescence channels FL1 and FL3. This background cluster likely contains organics e.g. viruses and free DNA. Another fluorescent cluster, situated on the diagonal, can consist of inorganic salts, containing autofluorescence. Stained cells will typically show a slightly higher green fluorescence compared to red fluorescence and are situated on the right hand side of the noise. Cells will have a green fluorescence FL1 of at least 1500-2000 fluorescence intensity.

3.4 Assembling the synthetic communities

The communities are to be invaded at each new invasion experiment with one pathogen. To assemble one community, the pure cultures are quantified each time, the correct dilutions are prepared, the communities are added together in the correct volumes and only then can a pathogen be introduced. Quantifying ten bacteria plus one pathogen with the flow cytometer each day can easily take up three hours due to four replicates. Preparing the correct dilutions can take up another hour, and pipetting the communities to invade another three hours. In total

repeating this procedure everyday would be quite laboursome. In order to reduce the workload, all the 194 synthetic communities are assembled in advance and frozen at a temperature of -80°C.

Storing of bacterial cultures for long-term preservation in the -80°C freezer is enabled by using cryoprotectants (glycerol). This is a chemical compound that limits damage to biological tissues in freezing conditions. Without protection, cells will rupture when they freeze as a result of expanding water ruining tissue samples. A 40% glycerol stock is prepared by adding 25 ml of glycerol (80%) and 25 ml demineralized H_2O in a glass bottle (100 mL) (using a 1:2 dilution) and autoclaving at 121°C for 20 min.

The 194 synthetic communities are labelled from 1 to 194, where the first 100 communities are stratified to optimally test the null-hypothesis of Eq. (3.1), while communities 101 to 194's composition are randomly drawn from simulations. The correct volumes for each bacteria are pipetted straight into the cryovials, after which 375 µL of glycerol stock is added. The cryovial is capped, vortexed briefly and stored in the -80°C freezer (Deweirdt, 2013). The entire workflow for one day is shown in Figure 3.10.

Since each new invasion of pathogen in the synthetic communities requires the stock to be thawed, the bacteria in the synthetic community can easily suffer from this repeated thawing-freezing cycle (Kerckhof et al., 2014). All communities are therefore retrieved from the stock, and pipetted into 96 well plates so that a stock is distributed over different plates. 50 µL of each community is pipetted into a normal 96-well plate, so that each plate is organised with different bacterial communities. Two lay-outs used in the experiments are shown in Figure 3.11, the remaining lay-outs can be found in Figure F.1. For each community there is one replica. The order of the replica's are not randomized to minimize the pipetting errors. If cross-contamination occurs, communities with equally low or high evenness have a higher chance of affecting each other within their own stratification for communities 1 to 100. For the remaining communities 101 to 194 these were drawn from a simulated random distribution so that cross-contamination can affect groups with either evenness. After pipetting 50 µL of each community into its corresponding well, the plates are sealed with transparent film and stored in the freezer at -80 °C. The organisation of the plates is shown in Table 3.3.

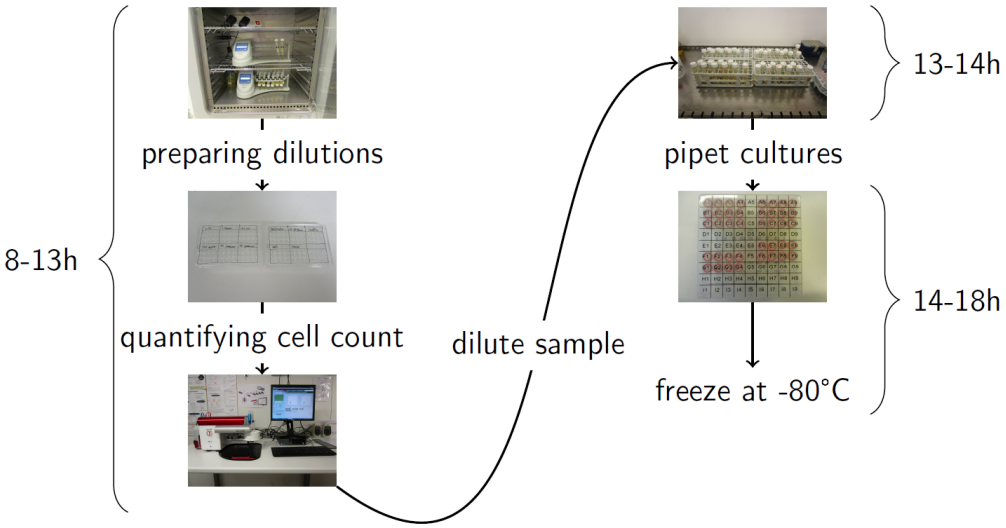
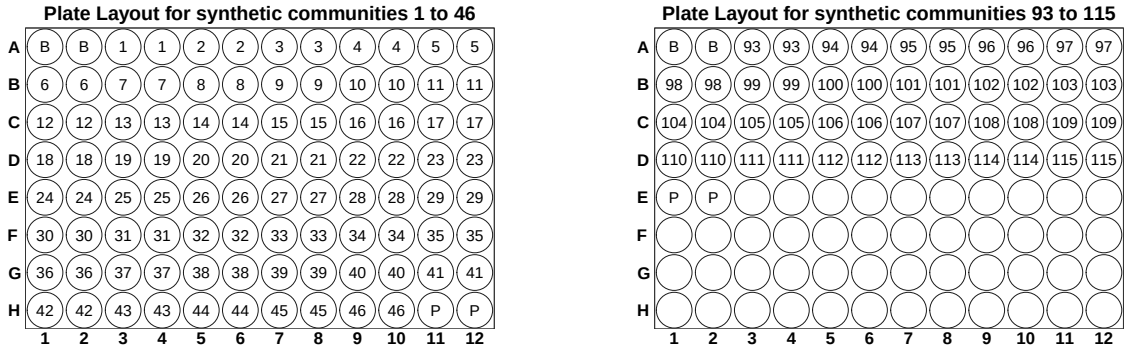


Figure 3.10: Daily flowchart for freezing the communities. A regular workday starts at 8'o clock in the morning. The concentration of bacteria in the pure cultures are determined via flow cytometry in the morning. In the afternoon, the bacteria are diluted to the different concentrations, from which the glycerol stocks are prepared.



(a) Communities 1 to 46. **(b)** Communities 93 to 115.

Figure 3.11: Lay-out for normal 96 well plates. Each well contains 50µL of the resp. community. The letter B are left empty, as they are intended as blanks. The letter P is left empty as a control for the pathogen, without any synthetic commnity. The numbers signify the resp. community number. Figures are made using R-package *ggplot2bdc* (Connelly, 2015)

Table 3.3: Organisation of the synthetic communities over the 96 well plates.

Plate	Communities
1	1 to 46
2	47 to 92
3	93 to 115
4	116 to 161
5	162 to 194

3.5 Determining antibiotic resistance in the synthetic communities

The growth of the pathogen invaded in the communities is quantified by plating in LB Agar with antibiotics. Distinguishing between pathogen and bacteria from the synthetic community is possible by making the pathogen resistant to antibiotics, while the bacteria in the synthetic community remain susceptible. To determine the concentration at which the bacteria in the synthetic community are resistant, they are tested on a gradient of antibiotics. In molecular biotechnology, it is a common selection strategy to use antibiotics in the medium to select for transconjugant organisms that contain a specific gene cassette with a "resistance marker". Although other strategies (heavy metals, temperature, etc.) exist the use of the antibiotics is most commonplace. Certainly when an exogenous gene cassette is located on a plasmid, the use of antibiotics is required to ensure selective pressure for a strain to retain the plasmid. Nonetheless, caution is necessary in applying and discarding of antibiotics as environmental exposure can lead to resistant environmental strains that could be pathogenic.

3.5.1 Streptomycin

Streptomycin binds to the small 16S rRNA of the 30S subunit of the bacterial ribosome, interfering with the binding of formyl-methionyl-tRNA to the 30S subunit. This leads to codon misreading, eventual inhibition of protein synthesis and ultimately death of microbial cells through mechanisms that are still not understood. Speculation on this mechanism indicates that the binding of the molecule to the 30S subunit interferes with 50S subunit association with the mRNA strand. This results in an unstable ribosomal-mRNA complex, leading to a frameshift mutation and defective protein synthesis; leading to cell death (Engelberg and Artman, 1964). A stock solution of streptomycin is prepared by dissolving 1 g of streptomycin sulphate salt (Sigma Aldrich) in 10 mL distilled water, filtered and stored at -20°C.

3.5.2 Spectinomycin

Spectinomycin acts against gram-negative and gram-positive bacteria. This antibiotic acts by inhibiting protein synthesis and elongation by binding to the bacterial 30S ribosomal subunit and interfering with peptidyl tRNA translocation, leading to disorders of ostomotic stability and lysis of microbial cells. A stock solution of streptomycin is prepared by dissolving 1 g of spectinomycin sulphate salt (Sigma Aldrich) in 10 mL distilled water, filtered and stored at -20°C.

3.5.3 Experiment

The bacteria in the synthetic community are grown in LB broth at 28 °C in dark conditions and continously rotated in an agitator. 48 hours after refreshing the bacteria in LB Broth, 150 *micro*L of each pure culture is combined in a cryovial and vortexed. Plates of LB Broth with agar (Sigma Aldrich) are prepared on a gradient of 0, 100, 200, 400, 600 and 800 mg/L with

Streptomycin. The solution of antibiotic is only added after sufficient cooldown (T < 60°C) to prevent deactiviation. 100 µL of the mixed cultures are pipetted on the agar plates, triangulated, sealed with parafilm and stored in dark conditions at 28°C. After 48 hours, the plates are checked for colony formation. Similar experiment is repeated for the antibiotic Spectinomycin. The lowest concentration where no colonies form is the lower limit of antibiotic resistance that the pathogen have to survive in order to distinguish growth between the pathogen and the members of the synthetic community.

3.6 Pathogen selection

To decide which pathogen are eligible for invading the synthetic communities, a pathogen list was requested from BCCM/LMG Bacteria Collection (Laboratory of Microbiology Ghent). All biosafety level 3 pathogen are removed from the list as well as the ones that have a higher risk to infect humans are also removed (Table E.1). Bacteria are selected from the same phylgenetic background (Table E.2), namely from the Class Gammaproteobacteria (Table 3.4). After subjecting the pathogen to antibiotic treatment the selection can be further narrowed down, depending on the results.

Table 3.4: Pathogen selection for invasion in a synthetic community.

10 selected bacteria	**Strain Number Bel**
Edwardsiella tarda	LMG 2793
Edwardsiella ictaluri	LMG 7860
Enterobacter cloacae subsp.cloacae(Enterobacter cloacae)	LMG 2783
Escherichia fergusonii	LMG 7866
Klebsiella oxytoca	LMG 3055
Proteus mirabilis	LMG 2954
Proteus inconstans_Providencia alcalifaciens/stuartii	LMG 7878
Providencia alcalifaciens_Proteus inconstans	LMG 7878
Raoultella planticola	LMG 3056
Raoultella terrigena	LMG 3203
Serratia marcescens subsp.marcescens	LMG 2792
Shigella flexneri	LMG 10472 & LMG 21935
Shigella sonnei	LMG 10473
Moraxella osloensis	LMG 1043
Moraxella(Subgenus Moraxella) bovis	LMG 986
Actinobacillus equuli subsp.equuli	LMG 3736
Actinobacillus equuli subsp.equuli	LMG 3736

3.7 Making pathogen Strepto- and spectinomycin resistant

3.7.1 Experiment: Treatment on agar plates

The selected pathogen in Table 3.4 are subjected to a strepto- and a spectinomycin treatment. Initially, the pathogen are stored in LB broth in 10 mL polypropylene tube in the incubator at 28 °C in dark conditions while continuously being rotated. Similar to the antibiotics treatment of the bacteria of the synthetic community, an antibiotics gradient is prepared of 0, 100, 200, 400, 600 and 800 mg/L in agar plates. 100 µL of each pathogen is pipetted on the plates, triangulated, sealed with parafilm and stored in the incubator at 28 °C in dark conditions. After four days, they are checked for colony formation. A colony is picked from the plate with the highest antibiotic concentration, resuspended in LB broth, and replated on the plates with higher concentrations, after which they are stored in the incubator. This procedure is repeated until the pathogen are more antibiotic resistant than the synthetic community.

3.7.2 Experiment: Treatment in LB broth

If the pathogens in experiment Section 3.7.1 do not survive a concentration higher than 100 mg/L, an alternative approach can be used. Instead of plating the pathogen, they are suspended in LB broth with low concentrations of antibiotics. A concentration gradient of 0, 10, 20, 50, 100 and 200 mg/L is prepared in LB broth. When bacterial growth is visually confirmed in the 10 mL polypropylene tubes, 100 µL of LB broth containing pathogen can be pipetted into tubes with higher concentrations of antibiotics. These steps are repeated until the pathogen reach a higher resistance than the synthetic community.

3.8 Determining invasion concentration of pathogen

3.8.1 Experiment

Similar to the synthetic communities, the pathogen are brought to the same metabolic state for the final experiment of invading the synthetic communities and are thus frozen in -80 °C. The pathogen are inoculated in LB broth in dark conditions at 28 °C during 24 hours, after which they are quantified in the flow cytometer. Concentrations of 10, 100 and 1000 cells/mL in glycerol are prepared in a cryovial in a 100 times upconcentration and frozen at -80 °C. After 48 hours the pathogen are retrieved from the stock, together with a sample of the synthetic communities at -80 °C. 15 µL of pathogen stock and 15 µL of synthetic community are pipetted in a 96-deep well plate and diluted with 1470 µL of LB broth. This procedure is repeated for 4 communities with duplicates. A well is filled with synthetic communities serving as a blank, in order to test for contamination in the plates. The 96 deep-well plate is sealed with a film and placed in an agitator at 380 rotations/minute in dark conditions at 28 °C during 48 hours. Agar plates are prepared with the corresponding antibiotics concentration to which the pathogen are resistant. For each of the 3 invasion concentrations 10, 100 and 1000 cells/mL of the pathogens,

a 10 to 10^5 dilution series is prepared in physiological solution (0.85 % NaCl). 10 µL of each diluted community invaded with pathogen is dropped on agar plates in order to optimize for the droplet method (Herigstad et al., 2001). The plates are sealed with parafilm and stored in dark conditions at 28 °C. After 48 hours the colonies are counted. The invasion concentration and dilution factor at which the colonies are countable is selected for the ensuing experiments. The invasion concentration is kept the same for all other pathogen, while the dilution factor before plating via the droplet method has to be determined for every new pathogen.

3.8.2 Experiment: Freezing pathogen in glycerol stock

After determining the correct invasion concentration, each pathogen is first quantified in the flow cytometer, after which it is frozen in a glycerol stock.

3.9 Invading communities

All previous experiments are preparatory steps for the actual gist of pathogen invasion in the synthetic communities, the results of which enable to test the null-hypothesis in Eq. (3.1). At this stage of the experiment, all ingredients are ready for starting the invasion in the synthetic communities. All 194 synthetic communities are prepared in a glycerol stock in normal 96 well plates and stored in the freezer at -80 °C (cf. Section 3.4). These can be readily pulled out of the stock and invaded with pathogen. The antibiotics resistance of the synthetic community has been determined (cf. Section 3.5) and the pathogen are resistant to antibiotics at a higher concentration than the synthetic communities' (cf. Section 3.7). A pathogen stock has been prepared in glycerol so that they are in the same metabolic state as the synthetic community (cf. Section 3.8).

3.9.1 Experiment

One 96 well-plate containing synthetic communities and two cryovials stock of the same pathogen are pulled out of the -80° freezer, and thawed during one hour. Meanwhile each well of a 96 deep-well plate is filled with 1470 µL LB broth. After one hour the synthetic communities in the normal 96 well plates are spun in the centrifuge during 10 minutes at 2204 g. The transparant film is removed and 15 µL of the communities are pipetted into the respective wells using a multichannel pipet, since the lay-out is exactly the same for the deepwell plates as the communities in the stock (cf. Figure 3.11 and F.1). 15 µL of the pathogen stock is inoculated in the 96 deep-well plates. Now that the 96-deep well plates contain the synthetic communities as well as the pathogen, the deep-well plate is sealed with a film and stored in the incubator in dark conditions at 28 °C while being agitated at 380 rpm.

After 48 hours in the incubator, the 96 deep well plates can be plated. Each well is diluted until the correct factor is reached via a dilution series in physiological solution (0.85 % NaCl) (cf. Section 3.8). A 10 times dilution is executed by pipetting 180 µL of physiological solution in

normal 96 well plates. 20 µL of the deep-well plates is pipetted into each corresponding well. This 10× dilution procedure is repeated until the target dilution concentration is reached.

The workload is decreased by making use of a multichannel pipet. To accelerate the speed for the invading experiments, a range of possible dilutions can be tested and all those dilutions are plated so that the right concentration does not have to be determined in advance. From the diluted communities invaded with a pathogen, 10 µL can be pipetted onto an agar plate with the tested antibiotic concentration (cf. Section 3.7) via the droplet method (Herigstad et al., 2001). In order to facilitate the plating, square plates are used which are divided into smaller squares by drawing a grid at the bottom of the well. These are indexed for easy recognition. In each square, there is enough space for a droplet of a diluted community with pathogen. After drying the droplets, the square plates are sealed with parafilm and stored at 28 °C in dark conditions in the incubator. After 48 hours, or less when needed, the colonies can be counted. For each pathogen a cycle is required of 5 total plates to invade all communities from 1 to 194. The order of the plates are shown in Table 3.3.

All the plating lay-outs can be found in appendix G.

After 48 hours in the incubator, the 96 deep well plates can be plated. Each well is diluted until the correct factor is reached via a dilution series in physiological solution (0.85 % NaCl) (Section 3.8). A 10 times dilution is executed by pipetting 180 µL of physiological solution in normal 96 well plates. 20 µL of the deep-well plates is pipetted into each corresponding well. The workload is decreased by making use of a multichannel pipet. To accelerate the speed for the invading experiments, a range of possible dilutions can be tested and all those dilutions are plated so that the right concentration does not have to be determined in advance. From the diluted communities invaded with a pathogen, 10 µL can be pipetted onto an agar plate with the tested antibiotic concentration (Section 3.7) via the droplet method (Herigstad et al., 2001). In order to facilitate the plating, square plates are used which are divided into smaller squares by drawing a grid at the bottom of the well. These are indexed for easy recognition. In each square, there is enough space for a droplet of a diluted community with pathogen. After drying the droplets, the square plates are sealed with parafilm and stored at 28 °C in dark conditions in the incubator. After 48 hours, or less when needed, the colonies can be counted. For each pathogen a cycle is required of 5 total plates to invade all communities from 1 to 194. The order of the plates are shown in Table 3.3. The workflow for invading with one pathogen can be found in D.1. The suggested workflow becomes interesting after two days, when it is possible to start a second set of invasion experiments with a new pathogen. The schedule is set up in such a manner that it allows for the invasion to happen between the plating experiments. A flowchart of the workflow is shown in Figure 3.12.

1. Start invasion in 96 deep-well plate. Revive pathogen and synthetic community from −80°C freezer and thaw for approx. 60 min.

2. Spin synthetic community 10 min. in centrifuge at 2204g and remove transparent film. Pipet 1470 µL LB broth in each well. Pipet 15 µL of each synthetic community in well, and invade with 15 µL of pathogen.

3. After 48h in the incubator at 28 °C, make dilution series for plating in physiological solution.

4. Plate from the dilution series via the droplet method. After drying store plates in the incubator at 28 °C.

5. After 48h count the pathogen colonies formed in the plates for each pathogen after two day invasion in communities.

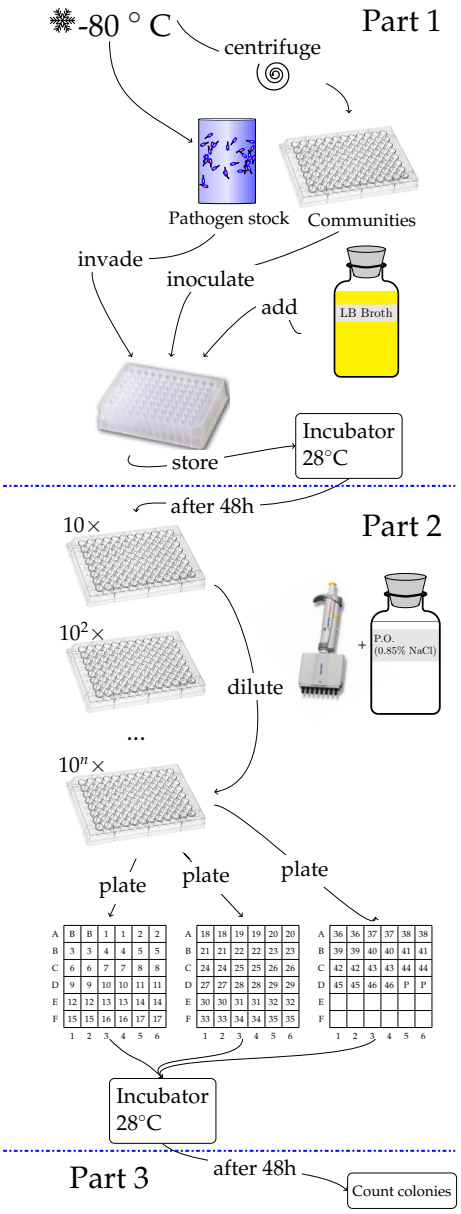


Figure 3.12: Flowchart for invasion of synthetic communities.

3.10 Similarity Measures

For the incorporation of similarity measures in the analysis two methods were used, namely Canonical Correlation Analysis (CCA) and Jaccard Index.

3.10.1 Canonical Correlation Analysis

Canonical Correlation Analysis (CCA) was introduced by Hotelling (1936), and is a technique often used in statistics and data analysis. CCA analyses the relation between a pair of datasets, i.e. matrices, and finds the direction of maximal correlation between the pair of matrices. In terms of linear algebra, CCA quantifies the similarities between two subspaces, namely the columns of a pair of matrices. Geometrically, CCA calculates the cosine of the *principle* angles

between the two subspaces (Avron et al., 2014). Different computational methods are used in practice to accelerate the calculations for CCA (Golub and Zha, 1995), but the one most frequently used and algorithmically simple to understand approach is shown here (Malacarne, 2014).

Definition

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{m \times l}$ be data matrices, and assume that $p = \text{rank}(\mathbf{A}) \geq \text{rank}(\mathbf{B}) = q$. The objective of CCA is to find the linear combinations

$$\mathbf{x}^T\mathbf{A} \text{ and } \mathbf{y}^T\mathbf{B}$$

such that these maximize the correlations

$$\rho = \text{Corr}(\mathbf{x}^T\mathbf{A}, \mathbf{y}^T\mathbf{B}), \tag{3.2}$$

where the correlation coefficients are defined as

$$\text{Corr}(\mathbf{A}, \mathbf{B}) = \frac{\text{cov}(\mathbf{A}, \mathbf{B})}{\sqrt{\text{var}(\mathbf{A}) \cdot \text{var}(\mathbf{Y})}}, \tag{3.3}$$

with

$$\text{cov}(\mathbf{A}, \mathbf{B}) = \sum_{i=1}^{N} \frac{(a_i - \mu_A)(b_i - \mu_B)}{N}. \tag{3.4}$$

Alternatively formulated, the canonical correlations $\rho_1(\mathbf{A}, \mathbf{B}) \geq \rho_2(\mathbf{A}, \mathbf{B}) \geq \ldots \geq \rho_q(\mathbf{A}, \mathbf{B})$ for the matrix pair $(\mathbf{A}, \mathbf{B})$ are defined recursively by the following formula (i = 1, ..., q):

$$\rho_i(\mathbf{A}, \mathbf{B}) = \max_{\mathbf{x} \in \mathbb{A}_i, \mathbf{y} \in \mathbb{B}_i} \rho(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) =: \rho(\mathbf{A}\mathbf{x}_i, \mathbf{B}\mathbf{y}_i)$$

The variance-covariance matrices $\mathbf{S_{AA}}$, $\mathbf{S_{BB}}$ of $\mathbf{A}$ and $\mathbf{B}$ and $\mathbf{S_{AB}}$ represent the covariance matrix of $\mathbf{A}$ and $\mathbf{B}$, and $\mathbf{S_{BA}}$ is its transpose. By making use of properties:

$$\text{Var}(\mathbf{U}\mathbf{A} + c) = \mathbf{U}\text{Var}(\mathbf{A})\mathbf{U}^T, \tag{3.5}$$

$$\text{Cov}(\mathbf{U}\mathbf{A}, \mathbf{V}\mathbf{B}) = \mathbf{U}\text{Cov}(\mathbf{A}, \mathbf{B})\mathbf{V}^T, \tag{3.6}$$

where $\mathbf{U}$ and $\mathbf{V}$ are conformable and c is a constant, Eq. (3.3) can be rewritten as:

$$\mathrm{Corr}(\mathbf{X}^{\mathrm{T}}\mathbf{A}, \mathbf{Y}^{\mathrm{T}}\mathbf{B}) = \frac{\mathbf{X}^{\mathrm{T}}\mathbf{S_{AB}}\mathbf{Y}}{(\mathbf{X}^{T}\mathbf{S_{AA}}\mathbf{X})^{1/2}(\mathbf{Y}^{T}\mathbf{S_{BB}}\mathbf{Y})^{1/2}} \tag{3.7}$$

where

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pk} \end{pmatrix} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$$

$$\mathbf{Y} = \begin{pmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{p1} & \cdots & y_{pk} \end{pmatrix} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)$$

$k = \mathrm{rank}(\mathbf{S_{AB}}) = \mathrm{rank}(\mathbf{S_{BA}})$.

Solution procedure

From Eq. (3.3) to (3.6):

$$\mathbf{S_{AA}}^{-1}\mathbf{S_{AB}}\mathbf{S_{BB}}^{-1}\mathbf{S_{BA}}\mathbf{X} = \lambda^2\mathbf{X}, \tag{3.8}$$

where λ is the largest eigenvalue for the characteristic equations. The eigenvectors $\mathbf{x}$ and $\mathbf{y}$ can be found via substitution, but here the mean of the squared eigenvalues are used indicate the overall similarity.

Interpretation of CCA

The most difficult part of CCA is perhaps its interpretation. The similarity between two data matrices $\mathbf{A}$ and $\mathbf{B}$ is determined by applying CCA which results in finding (i) the *principle* angles such that the canonical correlations are maximal and (ii) the canonical correlations. The largest value of the canonical correlations is often selected as a measure of similarity between two data matrices. This approach assumes that most of the information about the relation between two data matrices is captured by the maximum of the canonical correlations. In large dimensional spaces, however, it is possible that the largest canonical correlation does not give a representative view of the overall canonical correlations, and thus the overall similarity. Therefore, here it is chosen to use the trace of the canonical correlations as a measure of similarity, which also corrects for the lack of manual curation in the SEED-database.

3.10.2 Jaccard index: compound, reaction or location-based

The Jaccard index is a measure of similarity between two sets and is defined as the size of the intersection divided by the union of the sets. Let both vectors $\mathbf{x}$ and $\mathbf{y}$ be a set of compounds, reactions or locations, then Jaccard's index (Jaccard, 1901) is defined as:

$$J(\mathbf{x}, \mathbf{y}) = \frac{|\mathbf{x} \cap \mathbf{y}|}{|\mathbf{x} \cup \mathbf{y}|} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{x}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} - \mathbf{x}^T\mathbf{y}} \tag{3.9}$$

so that

$$0 \leq J(\mathbf{x}, \mathbf{y}) \leq 1.$$

The more similar two sample sets are, the closer Jaccard's index $J(\mathbf{x}, \mathbf{y})$ is to one. The locations for each bacteria are downloaded from Chaffron et al. (2010). Each organism is linked with one or more locations from an article. The organism names are matched with the first part of their full name with all the reaction models. The locations are attributed to each organism based on absence or presence (0 or 1) without taking into account the frequency of occurence for each organism.

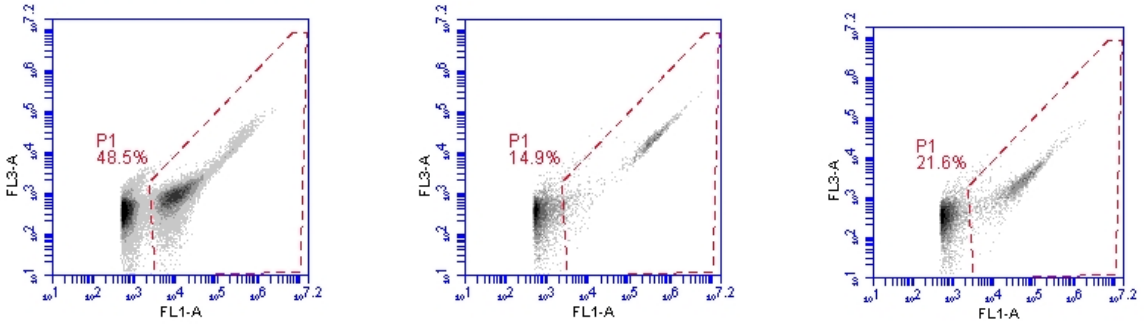
Chapter 4

Results

4.1 Lab results

4.1.1 Quantifying bacteria and freezing at -80 °C

The ten selected bacteria (Cf. Table 3.1) for the synthetic communities are quantified via flow cytometry and are refreshed 24 hours beforehand. The gating is done each time manually by E. Ehsani based on results from her previous experiments (E. Ehsani, Pers. Comm.).



(a) *Burkholderia* 1000 times diluted. **(b)** *Enterococcus* 5000 times diluted. **(c)** *Delftia* Sp. 1000 times diluted

Figure 4.1: The dots within the dashed red line are the cells of the bacteria in question. Except for similarity in growth curves, the bacteria were selected on the basis of seperability from the background. Each bacteria is measured in a 1000× and 5000× dilution. Based on the ratio of the quantified cell count in the dilutions, the correct concentration was determined.

After the correct concentration is obtained via Equation (4.1), stock solutions of $10^7\times, 10^6\times, 10^5\times$ and $10^4\times$ dilution are prepared for the ten bacteria; from these solutions the glycerol stocks are prepared in cryovials using the charts in appendix C. Each day, approximately between 10 and 18 communities are frozen as stock in cryoboxes with five replicates for each community. Preparation of a stock allows that the different synthetic communities do not have to be prepared each time at the beginning of a new invasion experiment. Repeating the quantification via flow

cytometry and assembling procedure of the synthetic communities for each invasion experiment can easily take up to 8 hours. With a stock, starting an invasion experiment consists of pulling the communities out of the freezer and thawing them for one hour.

$$\text{Concentration} = \text{cell count per 25 µL} \times \text{dilution factor} \times 25 \times 10^3 \frac{\mu L}{mL}, \tag{4.1}$$

4.1.2 Preparing 96-well plate stocks

When the stocks for the synthetic communities are retrieved from the freezer, and refrozen after invasion, they were not able to survive these cycli of thawing and freezing. The synthetic communities are therefore retrieved from the freezer, thawed and 50 µL of stock is pipetted in 96-well plates according to a predetermined lay-out (cf. Appendix F). This procedure enables an acceleration of the workflow, as the 96-well plates can be pulled from the freezer without affecting the remaining stock.

4.1.3 Testing the streptomycin resistance of the synthetic communities

To quantify the pathogen growth selectively, pathogens should have a higher antibiotic resistance than those of the bacteria in the synthetic community. The antibiotic resistance of the bacteria in the synthetic community are tested (Table 3.1). The bacteria in the synthetic community are not able to grow in streptomycin concentrations higher than 400 mg/L (cf. Table 4.1). Pathogen that can survive a streptomycin concentration of 400 mg/L are therefore suitable for invasion experiments, since they can be quantified afterwards.

4.1.4 Making the pathogen streptomycin resistant

Initially, a set of pathogen is subjected to a streptomycin treatment, but not all pathogens were able to resist so easily (Table 4.2).

4.1.5 Testing the spectinomycin resistance of synthetic community

Since not all pathogens were able to survive at a streptomycin concentration higher than 400 mg/L, the synthetic community was also tested for spectinomycin resistance. The pathogens were tested whether they could gain a higher resistance than a concentration of 50 mg/L spectinomycin (cf. Table 4.3).

4.1.6 Making the pathogen spectinomycin resistant

Pathogens were treated on a spectinomycin gradient to see whether they are suitable for invasion experiments. Pathogens that were already streptomycin resistant to a certain concentration were

Table 4.1: Synthetic communities are tested on streptomycin resistance after revival from the freezer for 24 hours. These are cultured on LB Agar afterwards and stored in dark conditions at 28 °C.

	Control	**100**	**200**	**400**	**600**	**800**	µg/ mL Streptomycin
Synthetic community	✓	✓	✓	✓	✗	✗	

Table 4.2: Pathogens are cultured on LB Agar and stored in dark conditions at 28 °C. The pathogens here are tested on a resistance to streptomycin.

	Pathogen name	**Control**	**100**	**200**	**400**	**600**	**800**	µg/µL	Enterobacter
LMG19137	*Staphylococcus*	✓	✓	✓	✓	✓	✗		No
LMG1232	*Bordetella Bronchiseptica*	✓	✓	✓	✓	✓	✓		No
LMG7866	*Escherichia Fergusonii*	✓	✓	✓	✓	✓	✓		Yes
LMG1043	*Moraxella osloensis*	✓	✓	✓	✓	✗	✗		No
LMG8064	*Moraxella bovis*	✗	✗	✗	✗	✗	✗		No
LMG10473	*Shigella Sonnei*	✓	✗	✗	✗	✗	✗		Yes
LMG8064	*Staphylococcus*	✓	✗	✗	✗	✗	✗		No
LMG3056	*Raoultella planticola*	✓	✗	✗	✗	✗	✗		Yes
LMG3055	*Klebsiella oxytoca*	✓	✗	✗	✗	✗	✗		Yes
LMG2783	*Enterobacter cloacae subsp.cloacae*	✓	✓	✓	✓	✓	✗		Yes
LMG8337	*Chryseobacterium indologenes*	✓	✓	✓	✓	✓	✓		No
LMG3736	*Actinobacillus equuli subsp.equuli*	✓	✗	✗	✗	✗	✗		No
LMG3203	*Raoultella terrigena*	✓	✓	✓	✓	✓	✓		Yes
LMG2792	*Serratia marcescens subsp.marcescens*	✓	✗	✗	✗	✗	✗		Yes
LMG7878	*Proteus Inconstans*	✓	✓	✗	✗	✗	✗		No
LMG2954	*Proteus mirabilis*	✓	✓	✗	✗	✗	✗		Yes
LMG2793	*Edwardsiella tarda*	✓	✗	✗	✗	✗	✗		Yes
LMG10472	*Shigella flexneri*	✗	✗	✗	✗	✗	✗		Yes

subjected to an additional spectinomcyin treatment.

4.1.7 Testing the dilution factors for plating and invading

Pathogen LMG7866 is refreshed 24 hours beforehand after which it is quantified via flow cytometry. Pathogen concentrations of 10, 100 and 1000 cells/mL are prepared and invaded in 4 different communities with low, medium low, medium high and high evenness. The synthetic communities are thawn from the freezer, invaded with three intensity levels and stored in the incubator at 28°C for 48 hours, after which they are plated on the concentration of antibiotics where pathogen LMG7866 can survive, namely 600 mg/L streptomycin and 100 mg/L spectinomycin. The colonies can be counted at a $10^5\times$ dilution (cf. Table 4.5). Two types of colonies could be distinguished, but these were confirmed to be both pathogen LMG7866 after 16S-sequencing. A pathogen concentration of 100 cells/mL was chosen for the invasion experiments and for the quantification after invasion a $10^5\times$ dilution was chosen. For each pathogen, however, it is necessary to determine this again during a first invasion experiment.

Table 4.3: Synthetic communities are tested on streptomycin resistance after revival from the freezer for 24 hours. These are cultured on LB Agar afterwards and stored in dark conditions at 28 °C.

	Control	**10**	**20**	**50**	**100**	**200**	µg/ µL Spectinomycin
Synthetic community	✓	✓	✓	✓	✗	✗	

Table 4.4: Pathogens are cultured on LB Agar and stored in dark conditions at 28 °C. Contaminated pathogens are removed.

	Pathogen name	**Control**	**10**	**20**	**50**	**100**	**200**	mg/L	Streptomycin resist.
LMG8337	*Chrysobacterium Indologenes*	✓	✓	✓	✓	✓	✓		No
LMG7866	*Escherichia Fergusonii*	✓	✓	✓	✓	✓	✗		600 mg/L
LMG3056	*Raoultella planticola*	✓	✓	✓	✓	✓	✓		contaminated
LMG3055	*Klebsiella oxytoca*	✓	✓	✓	✓	✓	✓		contaminated
LMG2783	*Enterobacter cloacae subsp.cloacae*	✓	✓	✓	✓	✓	✓		600 mg/L
LMG3736	*Actinobacillus equuli subsp.equuli*	✓	✓	✓	✓	✗	✗		No
LMG3203	*Raoultella terrigena*	✓	✓	✓	✗	✗	✗		No
LMG2792	*Serratia marcescens subsp.marcescens*	✓	✓	✓	✓	✗	✗		No
LMG1232	*Bordetella bronchiseptica*	✓	✓	✓	✓	✓	✓		No
LMG10472	*Shigella flexneri*	✓	✓	✓	✓	✗	✗		No
LMG2793	*Edwardsiella tarda*	✓	✓	✓	✗	✗	✗		No
LMG2954	*Proteus mirabilis*	✓	✓	✓	✓	✓	✓		No
LMG7878	*Proteus Inconstans*	✓	✓	✓	✓	✓	✗		No
LMG3736	*Actinobacillus equuli subsp.equuli*	✓	✓	✓	✗	✗	✗		No

4.1.8 Storing pathogen in the freezer and testing whether they are able to survive after thawing

Only a selection of Enterobacter is quantified in the flowcytometer and frozen at -80°C, because of time constraints. The pathogen are inoculated in LB broth 24 hours prior to quantification via flow cytometry. After quantification these are diluted to a concentration of 100 cells/mL in a glycerol stock and stored in the -80°C.

4.1.9 Invading with pathogen and plating

The pathogen are pulled from the freezer, together with the synthetic communities and invaded in 96-deep well plates with LB Broth. After 48 hours the invaded communities are plated on LB Agar with concentration of streptomycin and/or spectinomycin (cf. Table 4.7) in square Petri plates (cf. Appendix G). The square Petri plates are stored in the incubator at a temperature of 28°C and counted after 48 hours. Some are counted already after 24 hours, because of their growth speed.

Table 4.5: The counts were annotated with + indicate cluster or a number indicating the colonies followed by Y(ellow) or W(hite) to denote its colour. These two types of colonies were distinguished when plating pathogen LMG7866, but after 16S-sequencing it was confirmed that both were from LMG7866.

Invasion Concentration	**Dilution factor for plating on LB Agar**							
LMG7866	**1×**	**10×**	**100×**	**1000×**	**10000× A**	**10000× B**	**100000× A**	**100000× B**
Low 10	+	+	+	+	+	+	10Y /7W	10Y /6W
Low 100	+	+	+	+	+	+	10Y	6Y/4W
Low 1000	+	+	+	+	+	+	9Y/4W	13Y/4W
Medium Low 10	+	+	+	+	15Y/5W	15Y/8W	2Y	1W
Medium Low 100	+	+	+	+	+	+	5Y/2W	5Y/1W
Medium Low 1000	+	+	+	+	+	24Y/8W	3Y/1W	7Y/4W
Medium High 10	+	+	+	+	10Y/8W	15Y/7W	2Y	1Y
Medium High 100	+	+	+	+	18Y/6W	19Y/9W	4Y/1W	3Y/1W
Medium High 1000	+	+	+	+	+	+	13Y	11Y/5W
High 10	+	+	+	+	+	+	20Y/18W	18Y/15W
High 100	+	+	+	+	+	+	16Y/12W	18Y/13W
High 1000	+	+	+	+	+	+	19Y/7W	17Y/14W

Table 4.6: Pathogens are cultured on LB Broth and stored in dark conditions at 28 °C.

	Pathogen name	**Frozen**	**Streptomycin resistance (mg/L)**	**Streptomycin resistance (mg/L)**
LMG7866	*Escherichia Fergusonii*	Yes	600	100
LMG2783	*Enterobacter cloacae subsp.cloacae*	Yes	600	200
LMG7878	*Proteus Inconstans*	Yes	0	200
LMG3203	*Raoultella terrigena*	Yes	0	200
LMG2954	*Proteus mirabilis*	Yes	0	200

4.1.10 Description of the data

The dataset contains ten abundances for each bacteria in the synthetic community, the total cell count for each community and Pielou's evenness index.

4.2 Results

4.2.1 Statistical analysis

Initially Poisson regression models are fitted to test for the potential effects of Pielou's evenness and the initial total cell count of the 194 communities on the number of colonies formed by the respective pathogen. The count data of the colonies make the Poisson regression method suitable for this problem. The statistical analysis was performed with the R-program. To get a first exploration of the data, the count data is visualised for each pathogen in Figure 4.2. The distribution of the pathogen colonies can be further explored via histograms (Figure 4.3).

An initial Poisson regression model is fitted seperately for each of the four pathogen to explore

Table 4.7: Pathogens are cultured on LB Broth and stored in dark conditions at 28 °C.

	Pathogen name	**Frozen**	**Streptomycin resistance (mg/L)**	**Streptomycin resistance (mg/L)**
LMG7866	*Escherichia Fergusonii*	Yes	600	100
LMG2783	*Enterobacter cloacae subsp.cloacae*	Yes	600	200
LMG7878	*Proteus Inconstans*	Yes	0	200
LMG3203	*Raoultella terrigena*	Yes	0	200
LMG2954	*Proteus mirabilis*	Yes	0	200

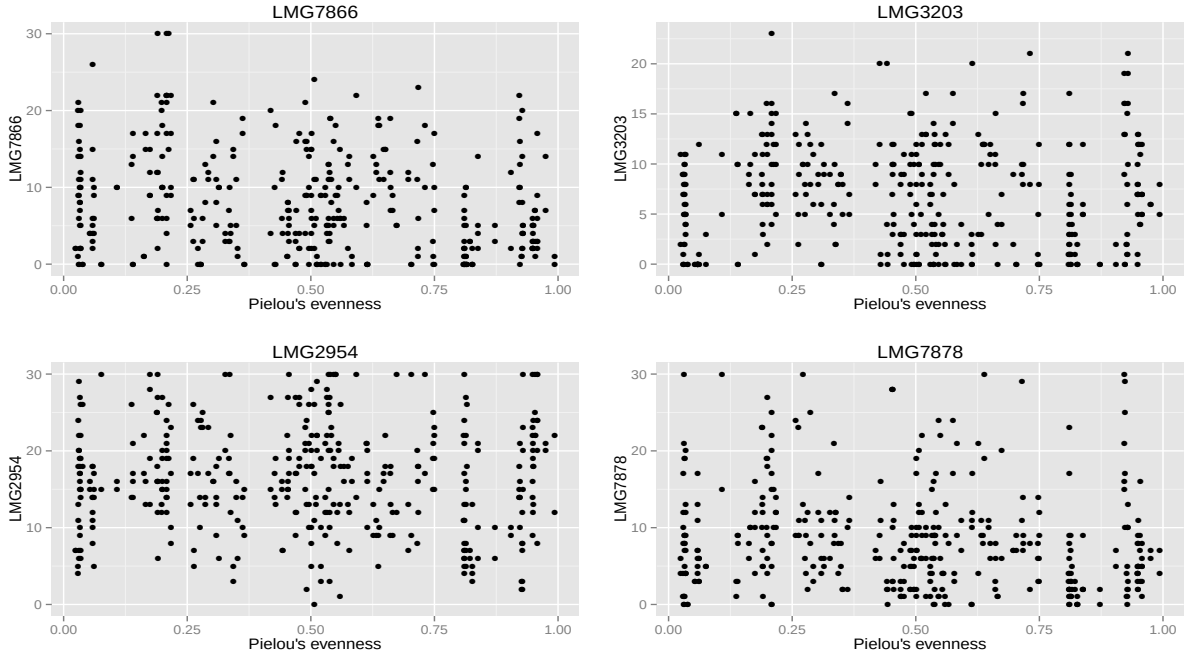


Figure 4.2: Each pathogen is plotted in function of evenness. The relation between the colonies counted and the evenness is visually difficult to distinguish, as the totall cell count might interfere as a confounding factor.

whether the interaction factor is significant between the initial total cell count in the community and the evenness. It is assumed that the response variable Y has a Poisson distribution, i.e. $y_i \sim \text{Poisson}(\mu_i)$ for $i = 1, \ldots, N$ where $N = 388$, because of 2×194 observations per pathogen. The expected count of y_i is $E(Y) = \mu_i$, so that μ can be understood as the mean number of quantified pathogen colonies after 48 hours, but the number of observed colonies also shifts around its mean μ with standard deviation $\sqrt{\mu}$ since the Poisson distribution assumes $Var(y) = \mu$. The Poisson model is defined as:

$$\begin{aligned} \log \mu = \; & \beta_0 + \beta_1 \times evenness + \beta_2 \times \log_{10}(cellcount) + \\ & \beta_3 \times evenness \times \log_{10}(cellcount) + \epsilon, \end{aligned} \tag{4.2}$$

where μ can be retrieved by taking the exponent of both sides in Eq. (4.2).

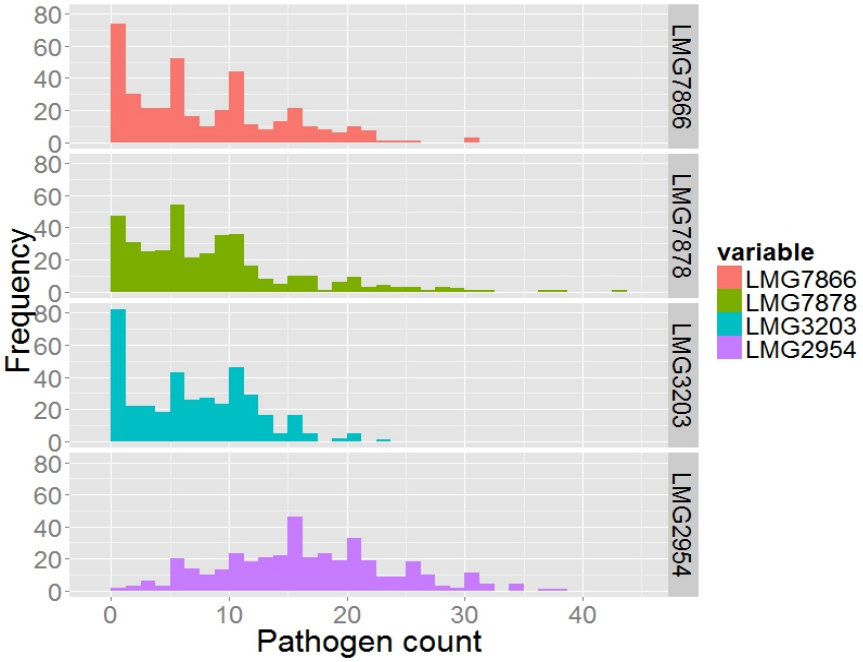


Figure 4.3: For each pathogen a histogram of the counted colonies are plotted after invasion. For pathogen LMG7866 and LMG3203 there appears to be a lot of zeroes, indicating that there might be 'overdispersion', i.e. the distribution being skewed towards one side as a result of a greater variability.

In Table 4.8 the models for pathogen LMG7866, LMG3203 and LMG2954 are not showing any significant interaction factors ($p > 0.05$). For pathogen LMG7878 the interaction between the initial total cell count and the evennes cannot be removed as its p-value is < 0.05. The model can be updated by removing the insignificant interaction factors resulting in model (4.3).

$$\log \lambda = \beta_0 + \beta_1 \times evenness + \beta_2 \times \log_{10}(cellcount) + \epsilon \tag{4.3}$$

For pathogen LMG7866, LMG3203 and LMG2954 the results of the final model are shown in Table 4.9. The parameter for Pielou's evenness are significant for pathogen LMG7866 and LMG3203. The normality of the residuals and the homoscedastisity are checked and confirmed (cf. Appendix). On the basis of these Poisson models the quantified pathogen counts are checked for 'overdispersion' via a regression-based overdispersion test (Cameron and Trivedi, 1990). Such a test basically checks whether the assumption of conditional variance-mean equality holds ($E[Y] = Var(Y)$) against the alternative hypothesis that the conditional variance, however, varies quadratically $Var[Y_i|X_i] = \mu_i + \alpha\mu_i^2$, thus accounting for the overdispersion (Greene, 2008). In consequence if $\alpha > 0$ and significant, then the distribution is overdispersed and a negative binomial regression is more favourable. The tests will be performed for LMG7866, LMG2954 and LMG3203 as the data for LMG7878 is too noisy.

In case of overdispersion a negative binomial regression can account for overdispersion by varying the conditional variance $Var[Y_i|X_i]$ parameter over the domain. First the interaction-effects were

Table 4.8: Poisson Regression is fitted for each pathogen seperately including parameters for Pielou's evenness, total initial cell count and an interaction factor. In the cases of pathogen LMG7866, LMG2954 and LMG3203 the interaction factors are not significant and are to be removed from the model. For LMG7878, however, the interaction factor is significant.

	LMG7866	p-value	LMG2954	p-value	LMG3203	p-value	LMG7878	p-value
(Intercept)	4.68	$< \mathbf{2E\text{-}16}^{***}$	3.93	$< \mathbf{2E\text{-}16}^{***}$	3.53	$\mathbf{3.29E\text{-}10}^{***}$	2.56	$\mathbf{1.87E\text{-}7}^{***}$
	(0.47)		(0.37)		(0.56)		(0.49)	
evenness	0.15	0.845	−0.70	0.2	−0.48	0.56	2.35	$\mathbf{0.0013}^{**}$
	(0.74)		(0.54)		(0.82)		(0.73)	
log10(cellcount)	−0.38	$\mathbf{1.56E\text{-}6}^{***}$	−0.18	$\mathbf{0.0027}^{**}$	−0.26	$\mathbf{0.0046}^{**}$	−0.03	0.70
	(0.08)		(0.06)		(0.09)		(0.08)	
evenness:log10(cellcount)	−0.17	0.185	0.10	0.28	0.07	0.63	−0.49	$\mathbf{7.6E\text{-}5}^{***}$
	(0.13)		(0.09)		(0.14)		(0.12)	
AIC	3375.16		2995.03		3016.24		3307.94	
BIC	3391.00		3010.88		3032.09		3323.78	
Log Likelihood	-1683.58		-1493.52		-1504.12		-1649.97	
Deviance	2104.75		1234.04		1787.87		1925.17	
Num. obs.	388		388		388		388	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

checked, but these were not significant. After removing the interaction factors, the statistical analysis of the negative binomial regression was performed. In Table 4.11, it can be seen that only for pathogen LMG7866 all coefficients are significant. The assumptions of homoscedasticity are checked in the appendix.

For pathogen LMG7866 and LMG2954, the cellcount has a significant influence on the number of pathogen colonies. The higher the initial total cells in the synthetic communities before invasion, the less the number of pathogen colonies counted after 48 hours of invasion. For pathogen LMG3203 the initial total cell count in the synthetic community has no significant influence. The evenness parameter is only significant for pathogen LMG7866, indicating that more even communities will have less pathogen colonies after 48 hours of invasion from LMG7866.

4.3 Similarity Measures

4.3.1 Data retrieval and preprocessing

In order to determine the pairwise similarity measures between bacteria, information is to be retrieved for each bacteria. For several bacteria, it has been determined which reactions occur in their metabolic pathway. This metabolic information can be retrieved from databases for several bacteria to calculate similarity measures. Here, the reactions for each bacteria available on the SEED-server, together with the corresponding names, are retrieved via the Network-Based SEED API (Henry et al., 2010). All reaction models that do not have a full name in addition to their code name are filtered so that the final collection contains 2991 reaction models. The metabolic data are converted to stoichiometric matrices for computing pairwise similarity measures between

Table 4.9: The interaction factors for pathogen LMG7866, LMG2954 and LMG3203 are removed from the initial models. The parameters for Pielou's evenness and the total initial cell count are significant in the case of pathogen LMG7866 and LMG2954. The evenness parameter has a p-value higher than 0.05, however, for LMG3203, while the total initial cell count remains significant.

	LMG7866	*p*-value	LMG2954	*p*-value	LMG3203	*p*-value
(Intercept)	5.19	$< \mathbf{2E\text{-}16}^{***}$	3.59	$< \mathbf{2E\text{-}16}^{***}$	3.30	$< \mathbf{2E\text{-}16}^{***}$
	(0.26)		(0.19)		(0.29)	
evenness	−0.83	$< \mathbf{2E\text{-}16}^{***}$	−0.11	$\mathbf{0.017}^{*}$	−0.09	0.20
	(0.07)		(0.05)		(0.07)	
log10(cellcount)	−0.46	$< \mathbf{2E\text{-}16}^{***}$	−0.12	$\mathbf{3.6E\text{-}5}^{***}$	−0.22	$\mathbf{9.9E\text{-}7}^{***}$
	(0.04)		(0.03)		(0.05)	
AIC	3374.91		2994.22		3014.47	
BIC	3386.79		3006.10		3026.35	
Log Likelihood	-1684.45		-1494.11		-1504.24	
Deviance	2106.50		1235.23		1788.10	
Num. obs.	388		388		388	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4.10: Results overdispersion test for pathogen LMG7866, LMG2954 and LMG3203. If $\alpha > 0$ and its $p-$value significant, then the distribution is overdispersed and a negative binomial regression could be more suitable.

	LMG7866	*p*-value	LMG2954	*p*-value	LMG3203	*p*-value
α	0.468	$< \mathbf{2.2E\text{-}16}^{***}$	2.0.128	$< \mathbf{2.2E\text{-}16}^{***}$	0.409	$< \mathbf{2.2E\text{-}16}^{***}$

alternative hypothesis: true dispersion is greater than 1

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

each organism. When the Gibbs free energy ΔG is larger than 1000 kcal/mol, this is set to zero and its standard deviation σ_d to 10 kcal/mol. If the Gibbs free energy ΔG is unavailable, this is set to 50 kcal/mol and its standard deviation σ_d to 1 kcal/mol. Let each stoichiometric matrix $\mathbf{S}$ be defined as a $m \times p$ matrix where the p columns are reactions and the m rows are the compounds. The S_{ij} values indicate the number of metabolites consumed (< 0) and produced (> 0) for each reaction.

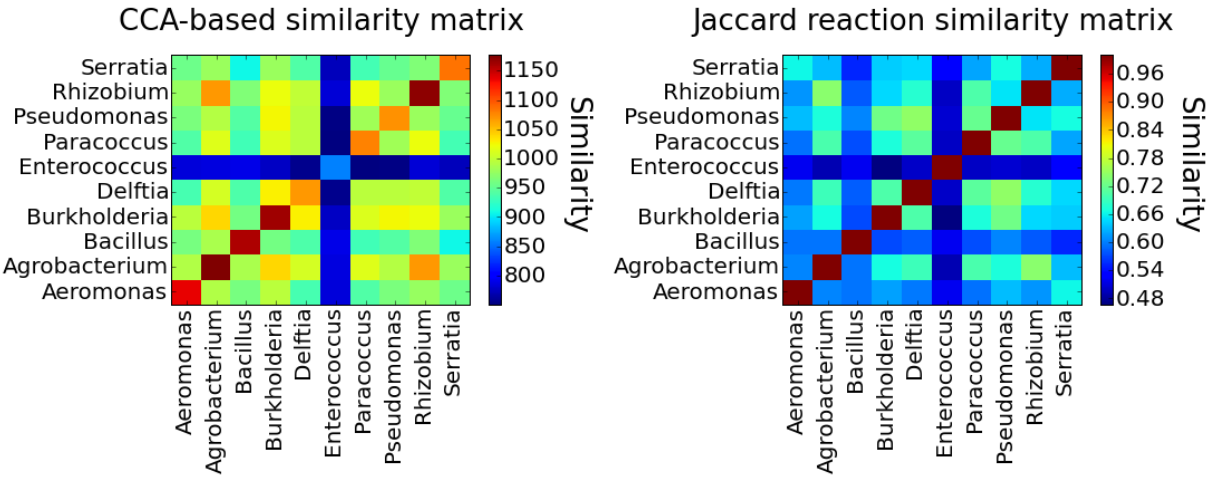
4.3.2 Similarity matrices

Similarity measures can be calculated pairwise between each bacteria using CCA Eq. (3.8) and the Jaccard index Eq. (3.9). A higher value for a similarity measure indicates that the two bacteria are more similar to each other, while a lower value indicates that they are not very similar to each other. The similarity information between each bacteria can be represented in a matrix. The calculated similarity matrices for the ten bacteria from the synthetic communities can be found in Figure 4.4. For the CCA-matrix Fig 4.4a the trace of the canonical correlations were calculated to account for the noise in the SEED-database.

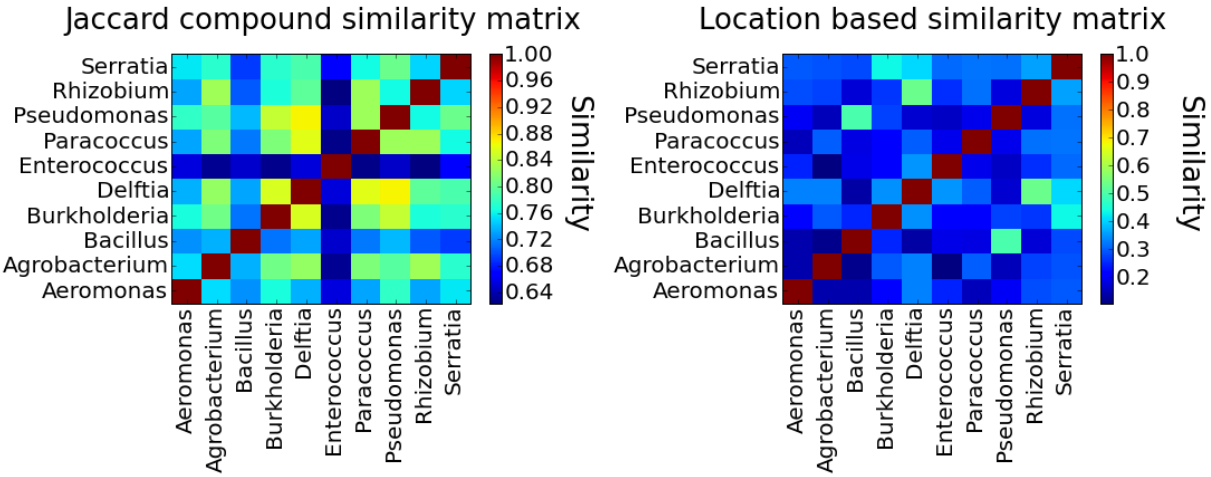
Table 4.11: Output for the negative binomial models. The interaction factors are removed as they were not significant. Only for pathogen LMG7866 all parameters are significant. The evenness coefficient is negative as is the logarithmically transformed cell count. For the other two pathogen neither the evenness nor the logarithm of the cell count is significant.

	LMG7866	*p*-value	LMG2954	*p*-value	LMG3203	*p*-value
(Intercept)	5.47^{***}	$< \mathbf{3.6E\text{-}14}^{***}$	3.60^{***}	$< \mathbf{2E\text{-}16}^{***}$	3.23^{***}	$\mathbf{5.4E\text{-}6}^{***}$
	(0.72)		(0.35)		(0.71)	
evenness	-0.95^{***}	$< \mathbf{3.4E\text{-}8}^{***}$	-0.11	0.188	-0.09	0.609
	(0.17)		(0.08)		(0.17)	
log10(cellcount)	-0.50^{***}	$< \mathbf{1.6E\text{-}5}^{***}$	-0.13^{*}	$\mathbf{0.023}^{*}$	-0.21	0.061
	(0.12)		(0.06)		(0.11)	
AIC	2390.49		2630.76		2310.08	
BIC	2406.34		2646.60		2325.92	
Log Likelihood	-1191.25		-1311.38		-1151.04	
Deviance	464.29		416.84		472.86	
Num. obs.	388		388		388	

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$



(a) Similarity matrix based on Canonical Correlation. **(b)** Similarity matrix based on reactions via Jaccard.

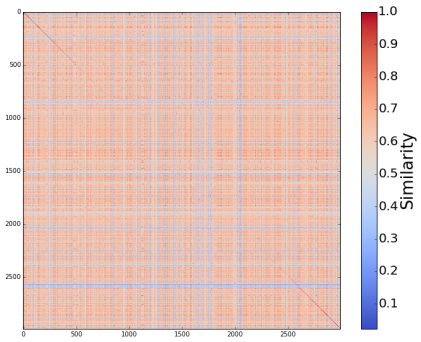


(c) Similarity matrix based on compounds via Jaccard.

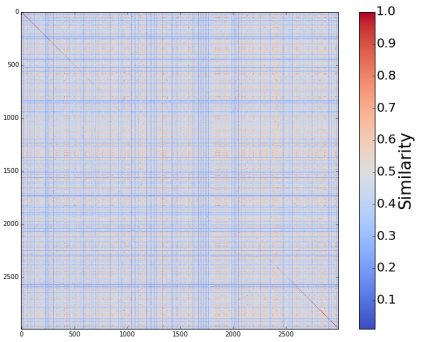
(d) Similarity matrix based on location.

Figure 4.4: Similarity matrices are shown here for the ten bacteria in the synthetic community based on different similarity measures. For the CCA-matrix in **(a)** the values are based on the overall correlations to account for the noise in the data.

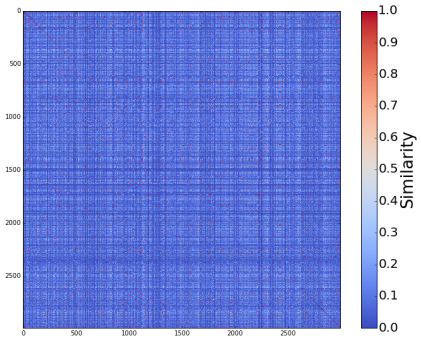
The similarity indices were calculated pairwise for the 2991 bacteria pulled from the database and stored in 2991 × 2991 matrices (Figure 4.5). The similarity matrix based on CCA, however, was computationally too intensive to compute and is shown to be less relevant later on. While the similarity measures for the ten bacteria can be visually distinguished in the similarity matrices (Figure 4.4), looking up the similarities for the 2991 bacteria by visually consulting the plots is not possible in very large similarity matrices. Interpretation of very large similarity matrix is therefore not trivial.



(a) Similarity matrix based on the compounds via Jaccard's similarity metric.



(b) Similarity matrix based on reactions via Jaccard's similarity metric.

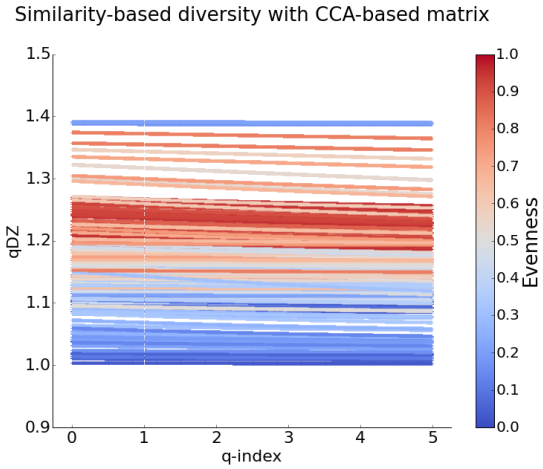


(c) Similarity matrix based on the location via Jaccard's similarity metric.

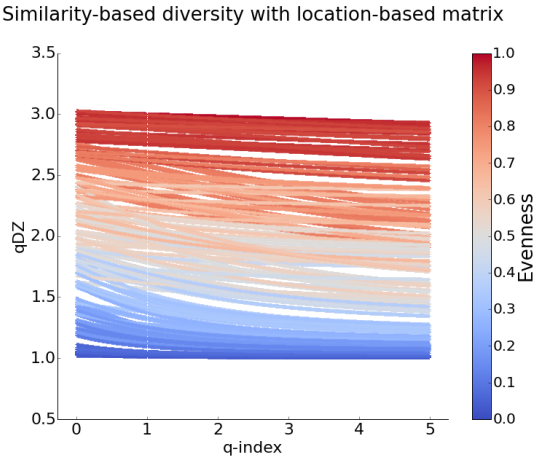
Figure 4.5: Similarity matrices are shown here for all 2991 bacteria based on different information but calculated via the Jaccard similarity metric. For similarity matrix **(a)** the values are much closer to one, indicating a high overall similarity, while this is lower for **(b)** and lowest overall for similarity matrix **(c)**.

4.3.3 Evaluation of the similarity-based diversity index

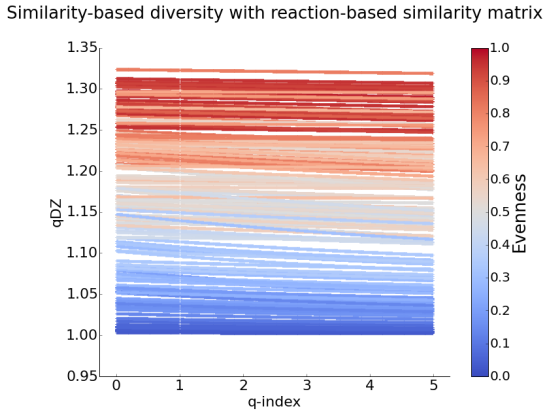
The similarity matrices are evaluated on the basis of the similarity-based diversity index in Eq. (2.15). The diversity profiles for the synthetic communities can be plotted in function of the diversity order q. The communities with a higher evenness are shown to be more diverse on the basis of the similarity-based diversity index (Figure 4.6). The influence of the initial diversity on the final invasion by pathogen LMG7866 is not so obvious to distinguish (Figure 4.7). Only the q-profiles for pathogen LMG7866 are plotted, since this invader generated the least noisy output.



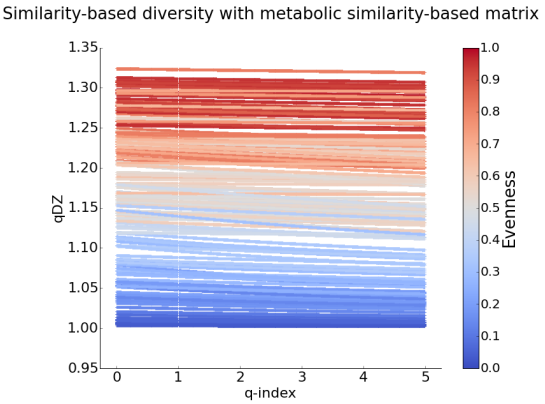
(a) Diversity profiles with similarity matrix based on CCA for 194 communities.



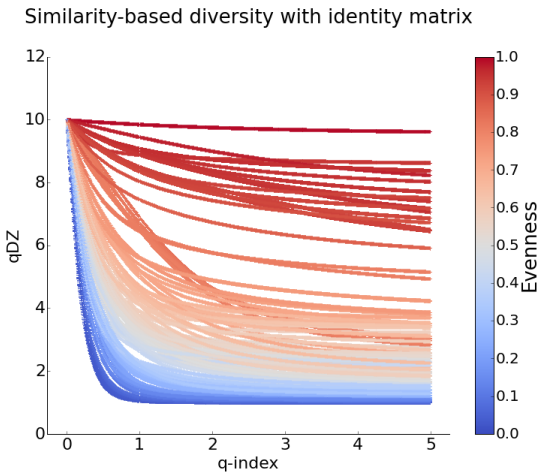
(b) Diversity profiles with similarity matrix based on location for 194 communities.



(c) Diversity profiles with similarity matrix based on reactions for 194 communities.

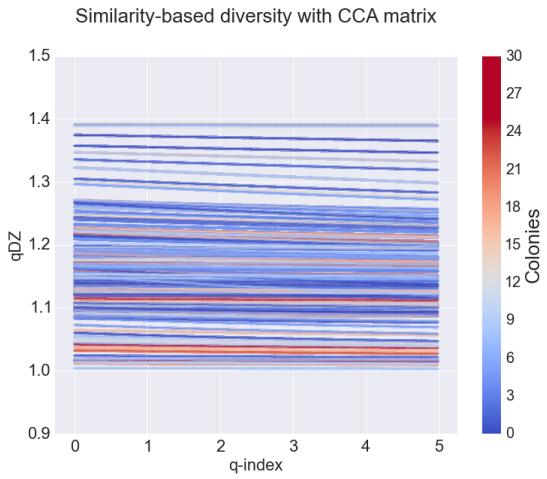


(d) Diversity profiles with similarity matrix based on compounds for 194 communities.

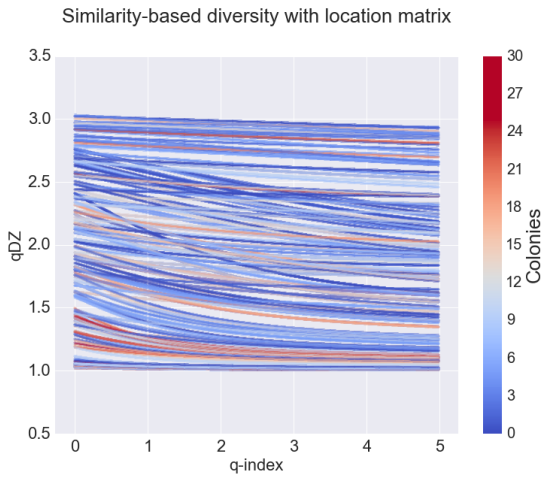


(e) Diversity profiles with identity matrix for 194 communities.

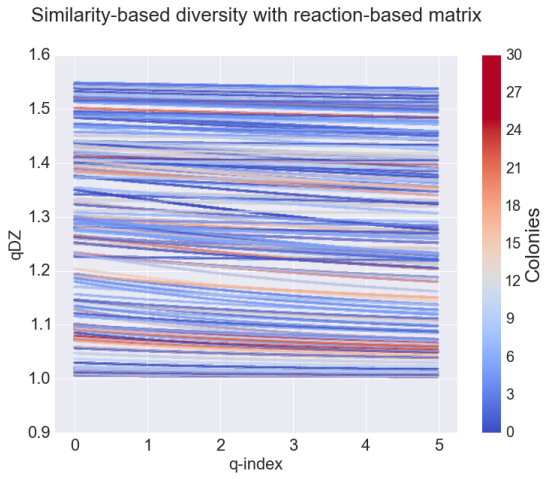
Figure 4.6: The diversity profiles plotted for all synthetic communities based on different similarity matrices. The graphical q-profiles are coloured according to the evenness for each community. Communities with a higher evenness are shown to be more diverse as they score higher on the similarity-based diversity.



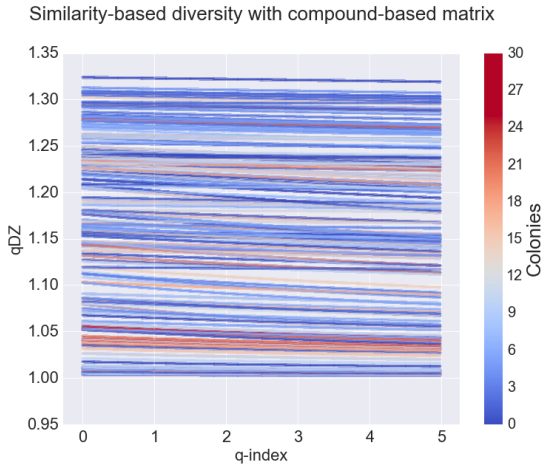
(a) Diversity profiles with similarity matrix based on CCA for 194 communities.



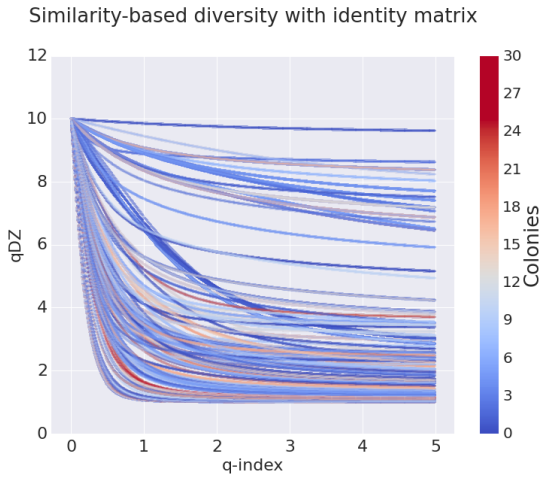
(b) Diversity profiles with similarity matrix based on location for 194 communities.



(c) Diversity profiles with similarity matrix based on reactions for 194 communities.



(d) Diversity profiles with similarity matrix based on compounds for 194 communities.



(e) Diversity profiles based on identity matrix for 194 communities.

Figure 4.7: The diversity profiles plotted for all synthetic communities based on different similarity matrices. The graphical q-profiles are coloured according to the invasion by pathogen LMG7866 for each community. Communities with a higher invasion rate are shown to be initially less diverse, although the pattern is difficult to distinguish, because of the overdispersion in the data (Fig. 4.3).

Instead of Pielou's evenness (Eq. (2.2)) the q-index is evaluated by calculating the residuals for each statistical model using the similarity-based diversity (Eq. (2.15)) based on four similarity matrices, namely CCA and Jaccard with information about the locations, reactions and compounds. The performance is significantly lower for the statistical models that quantify the invasion of pathogen LMG7866 when the q-index is based on the identity matrix. Overall there is a lower residual for the statistical models around a diversity order of q = 1 (Fig. 4.8). The residuals for CCA are higher over the entire range than for the other indices. Similarity matrices using location, reactions and the identity matrix score lower than the residuals of the negative binomial model.

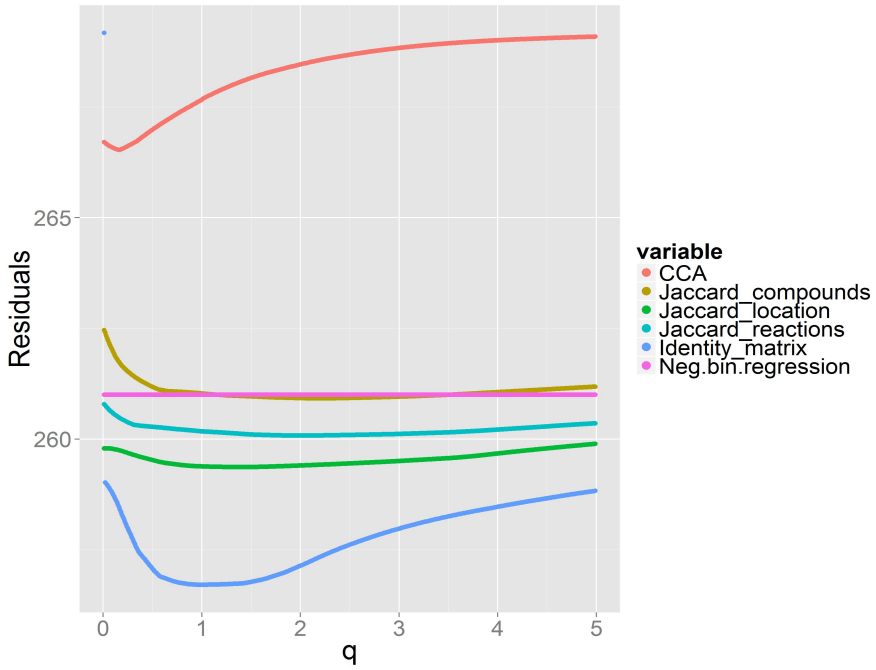
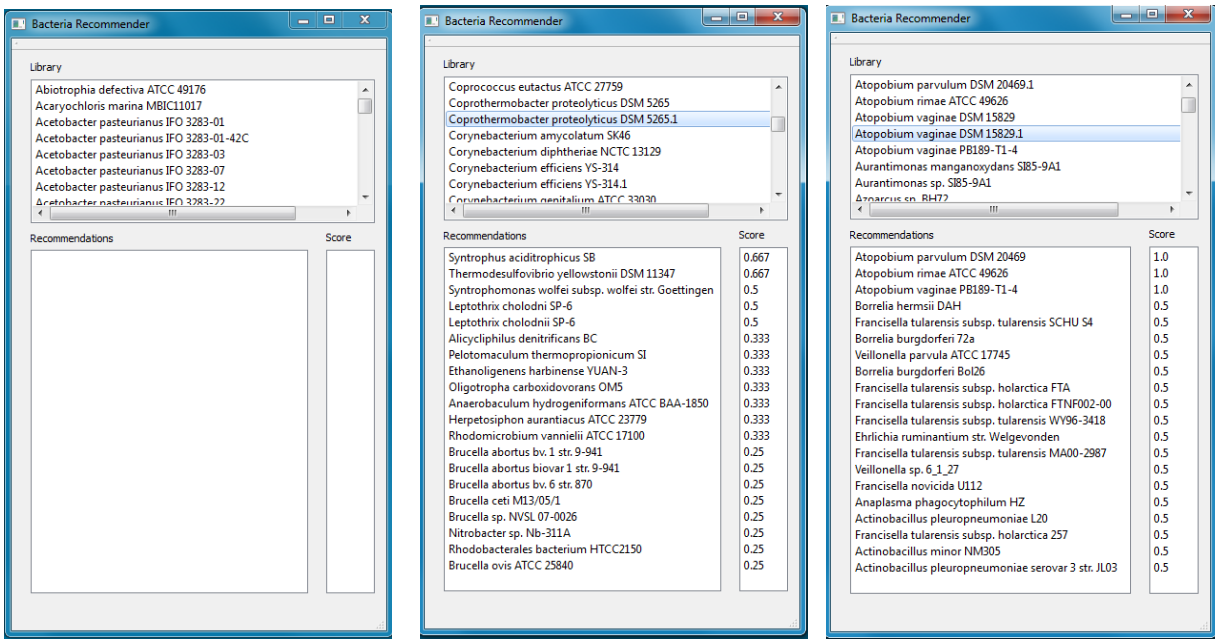


Figure 4.8: The residuals – i.e. the difference between the observed value and the estimated function value – for each q index is plotted to determine the best performance for the models that predict the colonies of pathogen LMG7866 after invasion. For a q in the neighbourhood of one all indices have a lower residual score. The lower the residual score the better the performance of the q in the statistical models, since the estimated function values deviate less from the observed values.

4.4 Building a recommender engine

A typical extension of similarity measures is its application in recommendation engines. When devising synthetic microbial communities, the selection procedure of the bacteria for the synthetic communities can be labour intensive. Numerous experiments have to be conducted before finding suitable matches between bacteria. The purpose of building a recommender engine exists in suggesting bacteria that are likely to interact with each other in a useful manner. The similarity metrics can be calculated between different items. Based on the pairwise similarity scores a recommendation can be made for an item that is quite similar. Since the metabolic data of the bacteria are already retrieved from the SEED-database (Henry et al., 2010) as well as the locations from Chaffron et al. (2010), the similarity metrics can easily be calculated between

each bacteria and the resulting values can be stored in a large matrix. By looking into each row of such a matrix for the highest scores, recommendations can be made for very similar bacteria. A primitive recommender engine for bacteria is built here using the similarity matrix based on location (Fig. 4.9). Such a memory-based method typically uses an entire database to generate predictions (Su and Khoshgoftaar, 2009).



(a) Recommender program for bacteria after start-up.

(b) Recommender program when you select *Coprothermobacter*.

(c) Recommender program when you select *Atopobium*.

Figure 4.9: Recommender engine based on the location similarity matrix. In the topbox a bacteria in the library can be selected. In the left bottom list the best 20 recommendations will appear. In the right bottom box the corresponding similarity score based on location are shown for the recommended bacteria.

Chapter 5

Discussion

5.1 Evenness: key to pathogen invasion

Community evenness is shown to be a key factor in the success of pathogen invasion in microbial communities (Dillon et al., 2005). The functional stability and the resistance to environmental stress of an ecosystem with low evenness is shown to be different from communities with high evenness (Wittebolle et al., 2009). Uneven comunities can be considered more fragile, since only one or few species are dominating the community, leaving more niches available for invasion (Hillebrand et al., 2008). Experimental evidence suggests indeed that invasion probability and species evenness are negatively correlated (De Roy et al., 2013).

Although the literature indicates that diversity indices are relevant for the functionality and the resistance in community structures (van Elsas et al., 2012; Shade et al., 2012; Rogers et al., 2013), Jost (2010) pointed out via the partitioning theorem Eq. (2.8) that decomposition of diversity into independent richness and evenness components is mathematically impossible. The requirement for numerical independence of richness and evenness plays an important role in the analysis of a community structure, as both components should reflect behaviour of the real world and not be constrained by mathematical artefacts. Diversity indices as such are a measure for biodiversity, but not necessarily 'diversity' itself (Jost, 2006). A possible solution to the problems surrounding the debate of diversity indices is suggested by using Pielou's evenness in Eq. (2.2) to describe the community structure. Despite the fact that Pielou's evenness is often rejected, because of its dependence on species richness, it is actually an excellent measure for community structure, as it also encapsulates the component of species richness within an ecosystem, as well as the species distribution (Jost, 2010).

A null hypothesis Eq. (3.1) has been formulated to test whether the probability for invasion is the same in communities with high evenness as in communities with low evenness. To optimally test for this question, an experimental design has been set-up where 194 synthetic communities with ten bacteria are created via simulation, taking into account the number of workable dilution concentrations, the total cellcount and the entire range of evenness. When

measuring environmental samples, it is difficult to explore the influence of factors such as species richness, evenness and diversity on other phenomena, as they are easily confounded with random noise found in complex environmental samples, possibly leading to contradictory conclusions (Wilsey and Polley, 2002; Emery and Gross, 2007). Creating synthetic communities allows for controlling the composition and conditions for an ecosystem, enabling a more clean way to explore microbial hypotheses and theories (De Roy et al., 2014). After preparing the synthetic communities and storing them in the freezer, four pathogen have been selected for invasion.

The counted colonies were initially analysed with a Poisson regression model, but the distribution of the reponse variable showed overdispersion, suggesting that such a model may not be the correct approach. Where Poisson regression assumes that $E[X] = Var(Y)$, a negative binomial model introduces a dispersion parameter, which allows the conditional variance $Var(Y_i|X_i)$ to vary quadratically over the domain (Hilbe, 2011). Pathogen LMG7878 was rejected as too noisy to analyse since there was confounding between the total cellcount in the community and the evenness. After analysing the interactions between evenness and cellcount with the negative binomial models, the interaction factors were rejected as insignificant. In the final models without interaction effects, the three remaining pathogen showed different results. For pathogen LMG7866 and LMG2954 the parameter of log10(cellcount) was significant and negative. When a community with a very high initial cellcount is invaded, the invaders are competing with a lot of cells for resources and are less likely to grow succesfully (Hibbing et al., 2010; Griffin et al., 2004). The parameter of total cellcount is not significant for pathogen LMG3203, indicating that independent of the total cellcount the invasion probability is the same.

In the experimental design, the total cellcount over all the synthetic communities has been kept as constant as possible to single out the influence of the evenness parameter. If the total cellcount would be of no significance for the invasion probability, as is the case for pathogen LMG3203, it can be attributed to this decision to keep the total cellcount as constant as possible, since it would be only a confounding factor for interpreting the influence of evenness. The parameter of evenness is only significant for pathogen LMG7866. This parameter is negative, indicating that a very even community has a lower probability of being dominated by newly introduced cells of pathogen LMG7866. The model's assumption of homoscedasticity has been checked in the Appendix. However, since for each of the four pathogen the significance of the parameters is different, the defined null-hypothesis Eq. (3.1) cannot be rejected. No conclusion is possible here about the influence of evenness on pathogen invasion overall.

Several explanations are possible for these results. The technique used for preparing the synthetic communities involves preparation in glycerol stock and freezing them at -80° C. Kerckhof et al. (2014) analysed the functionality and the community structure of mixed microbial systems when they are treated with a cryopreservant protocol and stored in the freezer at -80° C. After thawing the mixed microbial communities, the relative abundance of the communities can alter so that an extra control is needed to check for the exact abundance in each synthetic community. Initially it was planned for the synthetic communities to be sent to Illumina sequencing allowing for the correction of the abundances in the synthetic communities (Bentley et al., 2008), but because of

time constraints this was not possible anymore. Another factor playing a role is the pipetting complexity, which can be sensitive to human errors. Creating a lot of communities can help remediate average out the errors, but here also the errors could possibly be solved with the Illumina sequencing to determine which bacteria are exactly present in each community. A final note that is taken into account is the fact that there is a lot of variance in biological data, which possibly can be remediated by collecting more data.

Although the influence of evenness is not evident from all the experiments, only four pathogen are used for invasion. Finding relevant results about the role of evenness on pathogen invasion would require a lot more pathogen to be invaded in the communities. Here a lot of time was invested in preparing the stock of the synthetic communities and making the pathogen resistant so there was not a lot of time left for invasion experiments. The inconclusive nature of these results indicates that more data could possibly help in clarifying the picture. Invasion experiments with an entire batch of pathogen would enable a better understanding of the relation of evenness on pathogen invasion and allow for more general conclusions.

5.2 Similarity-based diversity: a different perspective

A possible way to incorporate more information in the analysis is by using a similarity-based diversity index. One possible approach is suggested by Leinster and Cobbold (2012), who propose a graphical representation of the diversity with q-profiles using Eq. (2.15). Depending on the similarity between bacteria in a community, different interactions are possible, for example competition, symbiosis or commensalism (West et al., 2007; Faust and Raes, 2012). Using a simple measure such as evenness does not take into account the diverse nature of interaction and communication between bacteria (Little et al., 2008; Klitgord and Segrè, 2010), which can have an influence on an ecosystem's resistance to pathogen invasion (Manefield and Turner, 2002).

An advantage of a similarity sensitive measure is its initial inspiration by the Hill numbers, which converts the established diversity indices to effective numbers (Hill, 1973). Raw diversity indices attempt to quantify significant differences in diversity between communities, but are difficult to interpret. The values are not limited between zero and one, as is the case for evenness, so that identifying communities with high and low diversity is not straightforward (Pallmann et al., 2012). The Hill numbers on the contrary allow for insight in the biology of an ecosystem by conveying different information depending on the order q of the diversity (Jost, 2006). Similarity-based diversity inherits these properties of having an intuitive interpretation, i.e. being an 'effective' number, from the Hill-numbers.

The similarity-based diversity by Leinster and Cobbold (2012) takes two inputs, namely the relative abundance data and similarity data. In the experimental design, the relative abundance data has already been determined and crudely quantified via flow cytometry. The similarity data can be obtained from different sources, such as phylogenetic trees, occurence in environmental samples or functionality. Here metabolic information was obtained from the SEED-server via the Network-Based SEED API (Henry et al., 2010). A second source of information was used

for determining the locations from Chaffron et al. (2010) who grouped publicly available 16S ribosomal RNA sequences and searched for co-occurences in environments.

The metabolic information was used to calculate pairwise similarity measures between the bacteria. Overall canonical correlations were determined between the synthetic community's ten bacteria based on their stoichiometric matrices, as well as the Jaccard indices for each bacteria's set of reactions, compounds and most frequent locations. The similarity metrics were stored in similarity matrices. To illustrate the effect of evenness on the diversity in the q-profiles, all synthetic communities are plotted and coloured according to evenness (Figure 4.6). It can be seen that communities with higher evenness score higher on the similarity-based diversity ${}^{\mathbf{q}}\mathbf{D}^{\mathbf{Z}}$ over the entire range of q's, while the communities with lower evenness score lower. All the invaded synthetic communities are also plotted in Figure 4.7 and coloured according to the counted colonies after invasion.

Only plots are shown for pathogen LMG7866, since primarily the data for this pathogen were least noisy. It can be seen in Figure 4.7 with some difficulty that the red curves are slightly less diverse as they have a lower diversity over the entire q-domain. The pattern, however, is not so obvious as in Figure 4.6. For q-profiles based on the identity matrix (Figure 4.6e), at q = 0 the graphs converge at a diversity of ten, which basically indicates the species richness. At q = 1 the diversity profiles diverge, and the similarity-based diversity boils down to its meaning of evenness. When q becomes larger than two, more weight is given to the similarity matrix. By using diversity profiles the interpretation of diversity in a community is facilitated.

To evaluate the relevance of diversity profiles, the similarity-based diversity can be used in the statistical analysis. The negative binomial model can be calculated each time where the similarity-based diversity indices are used instead of Pielou's evenness. For each q, the residuals of the corresponding statical model can be plotted to determine the optimal q for analysis and to see whether it adds extra value to the analysis (Figure 4.8). The optimal q can be found around the region where q = 1. By using similarity matrices the residuals of the statistical models are slightly decreased. Although the scale of improvement is quite marginal, the synthetic communities consist of only ten bacteria.

Similarity matrices can possibly add more value to an analysis when more bacteria are present in a community, since more possibilities are available for complex interactions which the evenness criterion cannot encapsulate (Little et al., 2008). An additional advantage with similarity-sensitive diversities is that communities with different species richness can be compared more easily via diversity profiles. Similarity-sensitive diversities are 'effective' numbers with a biological and intuitive meaning so that comparison between ecosystems is possible from a different environment. In this case, however, it was not possible to determine the exact role of similarity-based diversities on pathogen invasion, despite the fact that extra information was used. The quality of 'dry lab' analysis can only go so far as the quality of the 'wet lab' data, a problem which is encountered in other labs as well (Prinz et al., 2011). Overestimating the power of a sample can easily lead to spurious and irreproducible conclusions (Button et al., 2013). Without overinterpreting the results, it can be said that similarity-sensitive diversity (Leinster and Cobbold, 2012) can be

helpful, but still more data and experiments are needed to confirm this observation. A typical approach to solve the lack of data is to incorporate a-priori information about the relation between evenness and pathogen invasion found in the literature via Bayesian analysis. The Bayesian framework allows the user to introduce biological information with an informative prior so that the posterior belief includes information from the analysis on the data, as well as insights from other experiments (Kruschke, 2010).

5.3 Recommender engine: bacteria assemble

Similarity metrics are often used in recommender systems. Collaborative filtering techniques make use of preferences expressed by a group of users and can make recommendations for other users (Su and Khoshgoftaar, 2009). Here similarity data based on locations indicate whether a bacteria co-occurs with another bacteria. The preprocessing and selection of the bacteria for the synthetic communities involved a labour intensive procedure of quantifying growth curves via flow cytometry, sequencing of 16S rRNA of bacteria and post processing of data (E. Ehsani, personal communication). Here the ten bacteria were selected on the basis of similarity of growth curves and differences in 16S rRNA nucleotids. Pairwise similarities between bacteria can be determined by gathering and processing information from articles and databases. These similarity metrics could possibly be exploited to narrow down the possible experiments that were initially conducted where 110 bacterial isolates were quantified (E. Ehsani personal communication).

Here a similarity matrix has been constructed between 2991 bacteria, which is an example of a memory-based collaborative filtering technique (Su and Khoshgoftaar, 2009). All available data from the filtered bacteria were used to analyse the relationships between different bacteria, which can be used to generate recommendations (Sarwar et al., 2001). The relationships between bacteria were quantified via the Jaccard-similarity metric Eq. (3.9). The recommender process basically exists of finding the k most similar bacteria (nearest neighbors) after computing the similarities, then aggregate the neighbors to get the top-N most frequent bacteria as recommendation (Su and Khoshgoftaar, 2009). In the primitive recommender system (Fig. 4.9), twenty bacteria were selected with the highest scoring similarities for each of the 2991 bacteria. The Graphical User Interface simply allows the user to click in the library to look up the bacteria that are most similar to the current selection.

Synthetic microbial ecology mainly concerns itself with ecosystems that are created by a bottom-up approach where two or more defined microbial populations are assembled in a well-characterised and controlled environment (De Roy et al., 2014). Useful applications of this idea can be found in bio-augmentation, where a selection of bacteria with favourable characteristics is grown in media, after which they are introduced in a contaminated soils to accelerate the degradation of contaminants (de Lorenzo, 2008). Although the pre-conditioning phase of a community should not be underestimated (Pagaling et al., 2014), the strategy of synthetic microbial ecology can be beneficial from a theoretical point of view, which has already been demonstrated by successes in biotechnology (Pandhal and Noirel, 2014).

A necessary part, however, that could be subject to improvement is the selection procedure for finding the most favourable interactions, which might be the bottleneck in terms of man hours in synthetic ecology. Such experiments to find suitable matches between bacteria have already been performed by Stock et al. (2013), who managed to predict co-culture responses based on other interactions of the same heterotrophs and methanotrophs. The primitive recommender system built here is only limited to generating pairwise recommendations between bacteria within the similarity matrix. It is, however, not a far cry to extend the possibilities of a similarity matrix to generate synthetic ecosystems with suitable characteristics and interactions. Xu et al. (2012) already found that user preferences for certain items can depend on the subgroup to which the user belongs. By identifying such subgroups on the basis of interesting traits, it might be possible to suggest novel co-occurence networks for bacteria in synthetic ecology.

Chapter 6

Conclusion and perspectives

Microbial management holds a lot of potential for steering microbial ecosystems (Verstraete et al., 2007). A crucial link in the manipulation of microbial communities is the necessity to understand the mechanics and principles behind ecological theory (De Roy et al., 2014). Here, the influence of community structure on pathogen invasion has been examined. 194 synthetic communities with varying evenness have been assembled according to the principles of synthetic ecology (Großkopf and Soyer, 2014). All 194 communities were invaded separately each time with four pathogen and the resulting colonies were counted; no significant link between evenness and counted pathogen colonies were present and the results were different for all four pathogen.

Although the results were difficult to interpret, it was recognised that the dataset was still limited and that variance in biological data is a factor that can play a significant role. To remediate such issues, more invasion experiments should be conducted so that more data and information can be obtained. In order to cope with the scarce data, bacterial information was pulled from the SEED-database (Henry et al., 2010) and from Chaffron et al. (2010) to determine the pairwise similarities. By incorporating the similarity-based information into diversity indices, these could be evaluated in the analysis. The performance of the similarity-sensitive diversities in the statical models showed only a slight improvement of the residuals. Its full potential, however, could be tested on more complex ecosystems, where more species are present and more subtle interactions are possible.

The primitve recommender engine that was implemented could aid when assembling synthetic communities. In particular it can be useful for decreasing the workload in the selection procedure. A possible extension to a recommender engine could be the implementation to generate co-occurence networks by determining sub-groups in the similarity matrix (Xu et al., 2012).

Finally, a lot of preparatory work has been done to create stocks of 194 synthetic communities. Still remaining in the freezer are four cryovials of replica's for each synthetic community. Due to time constraints only invasion experiments could be executed with four pathogen, but follow-up on the communities in the freezer is possible.

Bibliography

Alatalo, R. V. Problems in the measurement of evenness in ecology. *Oikos*, pages 199–204, 1981.

Avron, H., Boutsidis, C., Toledo, S., and Zouzias, A. Efficient dimensionality reduction for canonical correlation analysis. *SIAM Journal on Scientific Computing*, 36(5):S111–S131, 2014.

Introduction to Flow Cytometry: A Learning Guide. BD Biosciences, 2000.

Beisel, J.-N., Usseglio-Polatera, P., Bachmann, V., and Moreteau, J.-C. A comparative analysis of evenness index sensitivity. *International review of hydrobiology*, 88(1):3–15, 2003.

Bell, T., Newman, J., Silverman, B. W., Turner, S. L., and Lilley, A. K. The contribution of species richness and composition to bacterial services. *Nature*, 436:1157–1160, 2005.

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., et al. Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218):53–59, 2008.

Bruno, J. F., Stachowicz, J. J., and Bertness, M. D. Inclusion of facilitation into ecological theory. *Trends in Ecology & Evolution*, 18(3):119–125, 2003.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., and Munafò, M. R. Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5):365–376, 2013.

Cameron, A. C. and Trivedi, P. K. Regression-based tests for overdispersion in the poisson model. *Journal of econometrics*, 46(3):347–364, 1990.

Casadevall, A. and Pirofski, L.-A. Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity. *Infection and Immunity*, 67(8):3703–3713, 1999.

Chaffron, S., Rehrauer, H., Pernthaler, J., and von Mering, C. A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome research*, 20(7): 947–959, 2010.

Chave, J. Neutral theory and community ecology. *Ecology letters*, 7(3):241–253, 2004.

Chen, Y. Development and application of co-culture for ethanol production by co-fermentation of glucose and xylose: a systematic review. *Journal of Industrial Microbiology & Biotechnology*, 38(5):581–597, 2011.

Chust, G., Chave, J., Condit, R., Aguilar, S., Lao, S., and Pérez, R. Determinants and spatial modeling of tree β-diversity in a tropical forest landscape in panama. *Journal of Vegetation Science*, 17(1):83–92, 2006.

Condit, R., Pitman, N., Leigh, E. G., Chave, J., Terborgh, J., Foster, R. B., Núnez, P., Aguilar, S., Valencia, R., Villa, G., et al. Beta-diversity in tropical forest trees. *Science*, 295(5555): 666–669, 2002.

Connelly, B. D. *ggplot2bdc: Themes and other functions for working with ggplot2*, 2015. R package version 0.2.0.

Cook, K., Garland, J., Layton, A., Dionisi, H., Levine, L., and Sayler, G. Effect of microbial species richness on community stability and community function in a model plant-based wastewater processing system. *Microbial Ecology*, 52(4):725–737, 2006.

Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J., and Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science*, 336:1255–1262, 2012.

Crowley, T. J. Causes of climate change over the past 1000 years. *Science*, 289:270–277, 2000.

Curtis, T. P., Sloan, W. T., and Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 10494–10499, 2002.

de Lorenzo, V. Systems biology approaches to bioremediation. *Current opinion in biotechnology*, 19(6):579–589, 2008.

De Roy, K. *Microbial Resource Management: Introducing New Tools and Ecological Theories*. PhD thesis, Ghent University, Belgium, 2014.

De Roy, K., Clement, L., Thas, O., Wang, Y., and Boon, N. Flow cytometry for fast microbial community fingerprinting. *Water Research*, 46(3):907–919, 2012.

De Roy, K., Marzorati, M., Negroni, A., Thas, O., Balloi, A., Fava, F., Verstraete, W., Daffonchio, D., and Boon, N. Environmental conditions and community evenness determine the outcome of biological invasion. *Nature Communications*, 4:1383, 2013.

De Roy, K., Marzorati, M., Van den Abbeele, P., Van de Wiele, T., and Boon, N. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environmental Microbiology*, 16(6):1472–81, June 2014.

DeVries, P. J., Murray, D., and Lande, R. Species diversity in vertical, horizontal, and temporal dimensions of a fruit-feeding butterfly community in an ecuadorian rainforest. *Biological journal of the Linnean Society*, 62(3):343–364, 1997.

Deweirdt, R. *Dietary Fat and the Human Gut Microbiome*. PhD thesis, Ghent University, 2013.

Dillon, R., Vennard, C., Buckling, A., and Charnley, A. Diversity of locust gut bacteria protects against pathogen invasion. *Ecology Letters*, 8(12):1291–1298, 2005.

Dunstan, P. K. and Johnson, C. R. Invasion rates increase with species richness in a marine epibenthic community by two mechanisms. *Oecologia*, 138(2):285–292, 2004.

Eisenhauer, N., Scheu, S., and Jousset, A. Bacterial diversity stabilizes community productivity. *PLoS One*, 7(3):e34517, 2012.

Emery, S. M. and Gross, K. L. Dominant species identity, not community evenness, regulates invasion in experimental grassland plant communities. *Ecology*, 88(4):954–964, 2007.

Engelberg, H. and Artman, M. Studies on streptomycin-dependent bacteria: effect of streptomycin on protein synthesis by streptomycin-sensitive, streptomycin-resistant and streptomycin-dependent, mutants of escherichia coli. *Biochimica et Biophysica Acta (BBA)-Specialized Section on Nucleic Acids and Related Subjects*, 80(2):256–268, 1964.

Fargione, J. E. and Tilman, D. Diversity decreases invasion via both sampling and complementarity effects. *Ecology Letters*, 8(6):604–611, 2005.

Faust, K. and Raes, J. Microbial interactions: from networks to models. *Nature reviews. Microbiology*, 10(8):538–50, August 2012.

Fisher, R. A., Corbet, A. S., and Williams, C. B. The relation between the number of species and the number of individuals in a random sample of an animal population. *The Journal of Animal Ecology*, pages 42–58, 1943.

Golub, G. H. and Zha, H. *The canonical correlations of matrix pairs and their numerical computation*. Springer, 1995.

Gotelli, N. J. and Colwell, R. K. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters*, 4(4):379–391, 2001.

Gotelli, N. J. and Colwell, R. K. Estimating species richness. *Biological diversity: frontiers in measurement and assessment*, 12:39–54, 2011.

Greene, W. Functional forms for the negative binomial model for count data. *Economics Letters*, 99(3):585–590, 2008.

Griffin, A. S., West, S. A., and Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature*, 430(7003):1024–1027, 2004.

Großkopf, T. and Soyer, O. Synthetic microbial communities. *Current Opinion in Microbiology*, 18:72–77, 2014.

Heip, C. A new index measuring evenness. *Journal of the Marine Biological Association of the United Kingdom*, 54(03):555–557, 1974.

Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B., and Stevens, R. L. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, 28(9):977–982, 2010.

Herigstad, B., Hamilton, M., and Heersink, J. How to optimize the drop plate method for enumerating bacteria. *Journal of Microbiological Methods*, 44(2):121–129, 2001.

Hibbing, M. E., Fuqua, C., Parsek, M. R., and Peterson, S. B. Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews. Microbiology*, 8(1):15–25, January 2010.

Hilbe, J. *Negative binomial regression*. Cambridge University Press, 2011.

Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology*, 54(2): 427–432, 1973.

Hillebrand, H. and Matthiessen, B. Biodiversity in a complex world: consolidation and progress in functional biodiversity research. *Ecology letters*, 12(12):1405–1419, 2009.

Hillebrand, H., Bennett, D. M., and Cadotte, M. W. Consequences of dominance: a review of evenness effects on local and regional ecosystem processes. *Ecology*, 89(6):1510–1520, 2008.

Ho, A., de Roy, K., Thas, O., De Neve, J., Hoefman, S., Vandamme, P., Heylen, K., and Boon, N. The more, the merrier: heterotroph richness stimulates methanotrophic activity. *The ISME journal*, 8:1945–1948, 2014.

Hodgson, D. J., Rainey, P. B., and Buckling, A. Mechanisms linking diversity, productivity and invasibility in experimental bacterial communities. *Proceedings of the Royal Society of London B: Biological Sciences*, 269(1506):2277–2283, 2002.

Hoeprich, P. Host-parasite relationships and the pathogenesis of infectious disease. *Infectious diseases, PD Hoeprich and M. C. Jordan (eds.). Lippincott, Philadelphia, Pennsylvania*, pages 41–53, 1989.

Hooper, D. U., Chapin Iii, F., Ewel, J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J., Lodge, D., Loreau, M., Naeem, S., et al. Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecological Monographs*, 75(1):3–35, 2005.

Hotelling, H. Relations between two sets of variates. *Biometrika*, pages 321–377, 1936.

Hughes, A. R., Inouye, B. D., Johnson, M. T., Underwood, N., and Vellend, M. Ecological consequences of genetic diversity. *Ecology Letters*, 11(6):609–623, 2008.

Huston, M. A. Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia*, 110(4):449–460, 1997.

Jaccard, P. Etude comparative de la distribution florale d'une portion des alpes et du jura. *Bull Soc Vaudoise Sci Nat*, 37:547–579, 1901.

Jiang, L. and Morin, P. J. Productivity gradients cause positive diversity–invasibility relationships in microbial communities. *Ecology Letters*, 7(11):1047–1057, 2004.

Jost, L. Entropy and diversity. *Oikos*, 113(2):363–375, 2006.

Jost, L. Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10): 2427–2439, 2007.

Jost, L. The relation between evenness and diversity. *Diversity*, 2(2):207–232, 2010.

Jousset, A., Schulz, W., Scheu, S., and Eisenhauer, N. Intraspecific genotypic richness and relatedness predict the invasibility of microbial communities. *The ISME journal*, 5(7):1108–1114, 2011.

Jurasinski, G. and Koch, M. Commentary: do we have a consistent terminology for species diversity? we are on the way. *Oecologia*, 167(4):893–902, 2011.

Jurasinski, G., Retzer, V., and Beierkuhnlein, C. Inventory, differentiation, and proportional diversity: a consistent terminology for quantifying species diversity. *Oecologia*, 159(1):15–26, 2009.

Kemper, N. Veterinary antibiotics in the aquatic and terrestrial environment. *Ecological indicators*, 8(1):1–13, 2008.

Kerckhof, F.-M., Courtens, E. N., Geirnaert, A., Hoefman, S., Ho, A., Vilchez-Vargas, R., Pieper, D. H., Jauregui, R., Vlaeminck, S. E., Van de Wiele, T., et al. Optimized cryopreservation of mixed microbial communities for conserved functionality and diversity. *PloS one*, 9(6):e99517, 2014.

Klitgord, N. and Segrè, D. Environments that induce synthetic microbial ecosystems. *PLoS Computational Biology*, 6, 2010.

Koleff, P., Lennon, J. J., and Gaston, K. J. Are there latitudinal gradients in species turnover? *Global Ecology and Biogeography*, 12(6):483–498, 2003.

Kostic, A. D., Gevers, D., Siljander, H., Vatanen, T., , Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al. The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. *Cell Host & Microbe*, 17(2):260–273, 2015.

Kruschke, J. *Doing Bayesian Data Analysis: A Tutorial Introduction with R*. Elsevier Science, 2010. ISBN 9780123814869.

Legendre, P., Borcard, D., and Peres-Neto, P. R. Analyzing beta diversity: partitioning the spatial variation of community composition data. *Ecological Monographs*, 75(4):435–450, 2005.

Leinster, T. and Cobbold, C. A. Measuring diversity: the importance of species similarity. *Ecology*, 93(3):477–489, 2012.

Levine, J. M. Species diversity and biological invasions: relating local process to community pattern. *Science*, 288(5467):852–854, 2000.

Lipsitch, M. and Moxon, E. R. Virulence and transmissibility of pathogens: what is the relationship? *Trends in Microbiology*, 5(1):31–37, 1997.

Little, A. E. F., Robinson, C. J., Peterson, S. B., Raffa, K. F., and Handelsman, J. Rules of engagement: interspecies interactions that regulate microbial communities. *Annual Review of Microbiology*, 62:375–401, January 2008.

Malacarne, R. L. Canonical correlation analysis. *Mathematica Journal*, 16, 2014.

Manefield, M. and Turner, S. L. Quorum sensing in context: out of molecular biology and into microbial ecology. *Microbiology*, 148(12):3762–3764, 2002.

Marzorati, M., Wittebolle, L., Boon, N., Daffonchio, D., and Verstraete, W. How to get more out of molecular fingerprints: Practical tools for microbial ecology. *Environmental Microbiology*, 10:1571–1581, 2008.

Meiners, S. J., Cadenasso, M. L., and Pickett, S. T. Beyond biodiversity: individualistic controls of invasion in a self-assembled community. *Ecology Letters*, 7(2):121–126, 2004.

Naeem, S., Knops, J. M., Tilman, D., Howe, K. M., Kennedy, T., and Gale, S. Plant diversity increases resistance to invasion in the absence of covarying extrinsic factors. *Oikos*, 91(1): 97–108, 2000.

Ofiteru, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., and Sloan, W. T. Combined niche and neutral effects in a microbial wastewater treatment community. *Proceedings of the National Academy of Sciences of the United States of America*, 107:15345–15350, 2010.

Pagaling, E., Strathdee, F., Spears, B. M., Cates, M. E., Allen, R. J., and Free, A. Community history affects the predictability of microbial ecosystem development. *The ISME journal*, 8(1): 19–30, 2014.

Pallmann, P., Schaarschmidt, F., Hothorn, L. A., Fischer, C., Nacke, H., Priesnitz, K. U., and Schork, N. J. Assessing group differences in biodiversity by simultaneously testing a user-defined selection of diversity indices. *Molecular ecology resources*, 12(6):1068–1078, 2012.

Pandhal, J. and Noirel, J. Synthetic microbial ecosystems for biotechnology. *Biotechnology letters*, 36(6):1141–51, 2014.

Peart, D. R. and Foin, T. C. Analysis and prediction of population and community change: a grassland case study. In *The population Structure of Vegetation*, pages 313–339. Springer, 1985.

Petchey, O. L. and Gaston, K. J. Functional diversity: back to basics and looking forward. *Ecology Letters*, 9(6):741–758, 2006.

Phillips, I., Casewell, M., Cox, T., De Groot, B., Friis, C., Jones, R., Nightingale, C., Preston, R., and Waddell, J. Does the use of antibiotics in food animals pose a risk to human health? a critical review of published data. *Journal of Antimicrobial Chemotherapy*, 53(1):28–52, 2004.

Pielou, E. *An introduction to Mathematical Ecology*. Wiley-Interscience, 1969.

Pielou, E. *Mathematical Ecology*. Wiley-Interscience publication. John Wiley & Sons, 1977. ISBN 9780471019930.

Pommier, T., Neal, P. R., Gasol, J. M., Coll, M., Acinas, S. G., and Pedrós-Alió, C. Spatial patterns of bacterial richness and evenness in the nw mediterranean sea explored by pyrosequencing of the 16s rrna. 2010.

Prinz, F., Schlange, T., and Asadullah, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature reviews Drug discovery*, 10(9):712–712, 2011.

Prosser, J. I., Bohannan, B. J. M., Curtis, T. P., Ellis, R. J., Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. J., Osborn, a. M., Solan, M., van der Gast, C. J., and Young, J. P. W. The role of ecological theory in microbial ecology. *Nature Reviews. Microbiology*, 5(5):384–92, May 2007.

Reacher, M. H., Shah, A., Livermore, D. M., Wale, M. C., Graham, C., Johnson, A. P., Heine, H., Monnickendam, M. A., Barker, K. F., James, D., et al. Bacteraemia and antibiotic resistance of its pathogens reported in england and wales between 1990 and 1998: trend analysis. *Bmj*, 320(7229):213–216, 2000.

Ricotta, C. Through the jungle of biological diversity. *Acta Biotheoretica*, 53(1):29–38, 2005.

Riegl, B., Purkis, S. J., Keck, J., and Rowlands, G. Monitored and modeled coral population dynamics and the refuge concept. *Marine Pollution Bulletin*, 58(1):24–38, 2009.

Robinson, G. R., Quinn, J. F., and Stanton, M. L. Invasibility of experimental habitat islands in a california winter annual grassland. *Ecology*, 76(3):786–794, 1995.

Rogers, G. B., Hoffman, L. R., Carroll, M. P., and Bruce, K. D. Interpreting infective microbiota: the importance of an ecological perspective. *Trends in microbiology*, 21(6):271–276, 2013.

Rousseau, R., Van Hecke, P., Nijssen, D., and Bogaert, J. The relationship between diversity profiles, evenness and species richness based on partial ordering. *Environmental and Ecological Statistics*, 6(2):211–223, 1999.

Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Bürgmann, H., Huber, D. H., Langenheder, S., Lennon, J. T., Martiny, J. B. H., Matulich, K. L., Schmidt, T. M., and Handelsman, J. Fundamentals of microbial community resistance and resilience. *Frontiers in microbiology*, 3(December):417, January 2012.

Shannon, C. E. and Weaver, W. *The Mathematical Theory of Communication.* Univ of Illinois Press, 1949.

Shea, K. and Chesson, P. Community ecology theory as a framework for biological invasions. *Trends in Ecology & Evolution*, 17(4):170–176, 2002.

Shearer, A. W. Whether the weather: Comments on 'An abrupt climate change scenario and its implications for United States national security'. *Futures*, 37:445–463, 2005.

Shimatani, K. On the measurement of species diversity incorporating species differences. *Oikos*, 93(1):135–147, 2001.

Simpson, E. H. Measurement of diversity. *Nature*, 1949.

Smith, B. and Wilson, J. B. A consumer's guide to evenness indices. *Oikos*, pages 70–82, 1996.

Smith, T. An attempt to interpret present-day uses of vaccines. *Journal of the American Medical Association*, 60(21):1591–1599, 1913.

Sørensen, T. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content*. Biologiske Skrifter. Biologiske Skrifter, 1948.

Stachowicz, J. J., Whitlatch, R. B., and Osman, R. W. Species diversity and invasion resistance in a marine ecosystem. *Science*, 286(5444):1577–1579, 1999.

Stachowicz, J. J., Fried, H., Osman, R. W., and Whitlatch, R. B. Biodiversity, invasion resistance, and marine ecosystem function: reconciling pattern and process. *Ecology*, 83(9):2575–2590, 2002.

Stock, M., Hoefman, S., Kerckhof, F.-M., Boon, N., De Vos, P., De Baets, B., Heylen, K., and Waegeman, W. Exploration and prediction of interactions between methanotrophs and heterotrophs. *Research in microbiology*, 164(10):1045–54, December 2013.

Su, X. and Khoshgoftaar, T. M. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009(Section 3):1–19, 2009.

Symstad, A. J. A test of the effects of functional group richness and composition on grassland invasibility. *Ecology*, 81(1):99–109, 2000.

Tuomisto, H. A consistent terminology for quantifying species diversity? yes, it does exist. *Oecologia*, 164(4):853–860, 2010a.

Tuomisto, H. A diversity of beta diversities: straightening up a concept gone awry. part 2. quantifying beta diversity and related phenomena. *Ecography*, 33(1):23–45, 2010b.

Tuomisto, H. An updated consumer's guide to evenness and related indices. *Oikos*, 121:1203–1218, 2012.

Van der Putten, W. H., Klironomos, J. N., and Wardle, D. A. Microbial ecology of biological invasions. *The ISME Journal*, 1(1):28–37, 2007.

van Elsas, J. D., Chiurazzi, M., Mallon, C. A., Elhottova, D., Kristufek, V., and Salles, J. F. a. Microbial diversity determines the invasion of soil by a bacterial pathogen. *Proceedings of the National Academy of Sciences of the United States of America*, 109(4):1159–64, 2012.

Van Nevel, S., De Roy, K., and Boon, N. Bacterial invasion potential in water is determined by nutrient availability and the indigenous community. *FEMS Microbiology Ecology*, 85(3): 593–603, 2013a.

Van Nevel, S., Koetzsch, S., Weilenmann, H.-U., Boon, N., and Hammes, F. Routine bacterial analysis with automated flow cytometry. *Journal of Microbiological Methods*, 94(2):73–76, 2013b.

Veech, J. A., Summerville, K. S., Crist, T. O., and Gering, J. C. The additive partitioning of species diversity: recent revival of an old idea. *Oikos*, 99(1):3–9, 2002.

Verstraete, W., Wittebolle, L., Heylen, K., Vanparys, B., de Vos, P., van de Wiele, T., and Boon, N. Microbial Resource Management: The road to go for environmental biotechnology. *Engineering in Life Sciences*, 7:117–126, 2007.

Warwick, R., Clarke, K., et al. New'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Marine Ecology Progress Series*, 129(1):301–305, 1995.

Watson, D. and Brandly, C. Virulence and pathogenicity. *Annual Reviews in Microbiology*, 3(1): 195–220, 1949.

West, S. a., Griffin, a. S., and Gardner, a. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology*, 20(2):415–32, March 2007.

Whittaker, R. H. Evolution and measurement of species diversity. *Taxon*, pages 213–251, 1972.

Whittaker, R. H. Vegetation of the great smoky mountains. *Ecological Monographs*, 26(1):1–80, 1956.

Whittaker, R. H. Vegetation of the siskiyou mountains, oregon and california. *Ecological monographs*, 30(3):279–338, 1960.

Wilsey, B. J. and Polley, H. W. Reductions in grassland species evenness increase dicot seedling invasion and spittle bug infestation. *Ecology Letters*, 5(5):676–684, 2002.

Wirjoatmodjo, S. *Growth, food and movement of flounder, Platichthys flesus (L.) in an estuary.* PhD thesis, The New University of Ulster, Coleraine, Northern Ireland, UK, 1980.

Wittebolle, L., Marzorati, M., Clement, L., Balloi, A., Daffonchio, D., Heylen, K., De Vos, P., Verstraete, W., and Boon, N. Initial community evenness favours functionality under selective stress. *Nature*, 458:623–626, 2009.

Xu, B., Bu, J., Chen, C., and Cai, D. An exploration of improving collaborative recommender systems via user-item subgroups. In *Proceedings of the 21st international conference on World Wide Web*, pages 21–30. ACM, 2012.

Zhang, Y., Chen, H. Y., and Reich, P. B. Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *Journal of Ecology*, 100(3):742–749, 2012.

Zinsser, H. Scientific books: Infection and resistance. *Science*, 41:686, 1915.

Appendices

Appendix A

Proof and derivations

A.1 Diversity of order $q = 1$

This proof can be directly found in Hill (1973), but is written here for the reader's convenience.

A.1.1 Proposition

Let $p_1, p_2, ..., p_S$ be positive numbers such that

$$\sum p_i = 1$$

and let the Hill diversity be

$${}^qD = (\sum p_i^q)^{1/(1-q)}.$$

${}^qD$ is continuous with derivatives of all orders at $q = 1$, and

$${}^1D = \lim_{q \to 1} {}^qD = \exp(-\sum p_i \ln p_i). \tag{A.1}$$

A.1.2 Proof

Three standard results of mathematical analysis are used here.

(a) for small values of x, $\exp(x) \approx 1 + x$.

(b) for small values of x, $\ln|1 + x| \approx x$.

(c) Let $A(x)$ and $B(x)$ be functions of x which can be expanded as power series in a neighbourhood of $x = 0$, and let $A(x)$ and $B(x)$ be 0 at $x = 0$. If the ratio $A(x)/B(x)$ is continuous at $x = 0$, then it also has derivatives of all orders, and it can be expanded as a power series.

Claim

$$\lim_{q\to 1}\big(\sum p_i^q\big)^{1/(1-q)} = \exp\big(-\sum p_i \ln p_i\big) \tag{A.2}$$

Proof
Substitute $q = 1 + b$ in Eq. (A.2) and take the natural logarithm of both sides so that

$$\lim_{b\to 0}\frac{1}{b}\ln\big(\sum p_i^{1+b}\big) = \sum p_i \ln p_i \tag{A.3}$$

The left hand side of (A.3) can be rewritten

$$\begin{aligned}
&\lim_{b\to 0}\frac{1}{b}\ln\big(\sum p_i \cdot p_i b\big)\\
&\lim_{b\to 0}\frac{1}{b}\ln\big(\sum p_i \exp(b\ln p_i)\big)\\
&\lim_{b\to 0}\frac{1}{b}\ln\big(\sum p_i + b\sum p_i(\ln p_i)\big) \qquad \text{via (a)}
\end{aligned}$$

Since $\sum p_i = 1$ and using result (b) the left hand side of equation (A.3) can be seen to be equal to its right side.

The continuity of the derivatives follows from (c) and if $A(x)$ is well behaved, so also is $\exp(-A(x))$. Q.E.D.

A.2 Examples of partitioning theorem

For true diversities it holds that they can be partitioned multiplicatively. To convert a common diversity H to its true diversity, Table 2.1 can be consulted for its conversion formula.

A.2.1 Species Richness

$$H_{tot} = H_A \cdot H_B$$

A.2.2 Shannon entropy

$$\begin{aligned}
\exp(H_{tot}) &= \exp(H_A)\cdot\exp(H_B)\\
\exp(H_{tot}) &= \exp(H_A + H_B)\\
H_{tot} &= H_A + H_B
\end{aligned}$$

A.2.3 Gini-Simpson

$$\begin{aligned}
\frac{1}{1-H_{tot}} &= \frac{1}{1-H_A} \cdot \frac{1}{1-H_B} \\
1-H_{tot} &= (1-H_A) \cdot (1-H_B) \\
1-H_{tot} &= 1-H_A-H_B+H_AH_B \\
H_{tot} &= H_A+H_B-H_AH_B
\end{aligned}$$

Appendix B

Code

All code for the experimental design and the statistical analysis can be found at `https://github.ugent.be/watsang/master-thesis`. All checks for statistical assumptions can be found there as well.

Appendix C

Tables pipetting

Table C.1: The ten bacteria chosen for creating the synthetic communities. The organisms were selected based on similarity in growth curves. The abbreviation is the name used in the pipetting table.

Selected bacteria	**Abbreviation**
Pseudomonas sp. (10 TYP)	10 TYP
Bacillus sp. (1 Bacillus)	1 Bacillus
Serratia sp. (14.3 ISO1)	14.3ISO1
Burkholderia Cepacia (Burkholderia)	25.2 Burkho
Paracoccus sp. (42 Paracoccus)	42 Para
Enterococcus sp. (59 Enterococcus)	59 Entero
Agrobacterium sp. (Beijirinckia)	Beijirinckia
Rhizobium Duejenense (63 Rhizobium)	63 Rhizi
Delftia Sp.	Delftia
Aeromonas sp. (K62)	K62 Aero

	10 TYP	1 Bacillus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
1	188 x10^7	188 x10^4	94 x10^4	188 x10^4	188 x10^4	94 x10^4	188 x10^4	94 x10^4	94 x10^4	188 x10^4
2	79 x10^4	158 x10^4	158 x10^7	158 x10^4	158 x10^4	158 x10^4	158 x10^4	158 x10^4	158 x10^4	158 x10^4
3	115 x10^4	115 x10^5	115 x10^4	231 x10^7	115 x10^4	231 x10^4	115 x10^4	115 x10^4	115 x10^4	231 x10^4
4	115 x10^4	115 x10^5	231 x10^4	231 x10^7	115 x10^4	231 x10^4	115 x10^4	115 x10^4	115 x10^4	115 x10^4
5	115 x10^5	231 x10^7	115 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^4	115 x10^4	231 x10^4	115 x10^4
6	107 x10^4	107 x10^5	107 x10^4	107 x10^4	214 x10^7	214 x10^4	214 x10^4	107 x10^4	107 x10^4	214 x10^4
7	107 x10^5	107 x10^4	214 x10^7	214 x10^4	214 x10^4	214 x10^4	107 x10^4	107 x10^4	107 x10^4	107 x10^4
8	107 x10^4	214 x10^4	214 x10^4	107 x10^4	107 x10^4	107 x10^4	107 x10^5	107 x10^4	214 x10^4	214 x10^7
9	107 x10^4	107 x10^4	107 x10^4	214 x10^4	214 x10^4	214 x10^4	214 x10^7	107 x10^4	107 x10^4	107 x10^5
10	107 x10^4	107 x10^4	107 x10^4	214 x10^4	214 x10^4	214 x10^4	107 x10^4	107 x10^4	214 x10^7	107 x10^5
11	100 x10^5	100 x10^4	100 x10^4	200 x10^4	100 x10^4	200 x10^4	200 x10^4	200 x10^4	200 x10^7	100 x10^4
12	200 x10^4	200 x10^4	200 x10^4	200 x10^4	100 x10^4	100 x10^4	100 x10^4	100 x10^4	200 x10^7	100 x10^5
13	200 x10^4	200 x10^7	100 x10^4	100 x10^4	200 x10^4	100 x10^5	100 x10^4	200 x10^4	200 x10^4	100 x10^4
14	200 x10^4	200 x10^4	100 x10^4	100 x10^4	200 x10^4	100 x10^5	200 x10^4	200 x10^7	100 x10^4	100 x10^4
15	100 x10^4	200 x10^4	100 x10^4	100 x10^4	200 x10^4	200 x10^7	100 x10^5	100 x10^4	200 x10^4	200 x10^4
16	188 x10^4	94 x10^5	94 x10^4	188 x10^4	188 x10^4	188 x10^4	94 x10^4	188 x10^7	188 x10^4	94 x10^4
17	115 x10^4	231 x10^4	231 x10^4	115 x10^4	115 x10^7	115 x10^4	115 x10^4	115 x10^4	115 x10^4	231 x10^4
18	136 x10^4	136 x10^4	136 x10^4	136 x10^7	136 x10^4	136 x10^4	136 x10^4	136 x10^4	273 x10^4	136 x10^5
19	115 x10^4	231 x10^4	115 x10^4	115 x10^7	231 x10^4	115 x10^4	115 x10^4	115 x10^5	115 x10^4	231 x10^4
20	115 x10^4	231 x10^4	115 x10^4	115 x10^4	231 x10^4	115 x10^4	231 x10^4	115 x10^5	115 x10^7	115 x10^4
21	115 x10^7	115 x10^4	231 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^5
22	231 x10^4	115 x10^4	115 x10^4	231 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^7	115 x10^5	115 x10^4
23	107 x10^4	214 x10^4	214 x10^4	214 x10^4	107 x10^5	107 x10^4	214 x10^4	107 x10^4	107 x10^4	107 x10^7
24	107 x10^4	214 x10^4	107 x10^5	107 x10^7	214 x10^4	107 x10^4	214 x10^4	107 x10^4	214 x10^4	107 x10^4
25	115 x10^4	115 x10^4	231 x10^4	231 x10^4	231 x10^6	115 x10^4	115 x10^4	115 x10^4	115 x10^4	115 x10^4

Figure C.1: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Badllus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
26	188 x10^4	94 x10^4	94 x10^4	94 x10^4	188 x10^4	188 x10^4	94 x10^4	188 x10^4	188 x10^6	188 x10^4
27	176 x10^6	176 x10^4	176 x10^4	176 x10^4	176 x10^4	176 x10^4	88 x10^4	88 x10^4	176 x10^4	88 x10^4
28	176 x10^4	176 x10^4	88 x10^4	176 x10^4	176 x10^4	176 x10^6	88 x10^4	176 x10^4	176 x10^4	88 x10^4
29	125 x10^4	125 x10^4	125 x10^5	125 x10^4	125 x10^4	125 x10^4	250 x10^4	125 x10^4	250 x10^6	125 x10^4
30	125 x10^4	125 x10^5	125 x10^4	250 x10^6	125 x10^4	250 x10^4	125 x10^4	125 x10^4	125 x10^4	125 x10^4
31	115 x10^4	231 x10^6	115 x10^4	115 x10^5	231 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^4	115 x10^4
32	115 x10^4	115 x10^4	115 x10^5	231 x10^4	115 x10^4	231 x10^6	115 x10^4	231 x10^4	115 x10^4	115 x10^4
33	115 x10^5	231 x10^6	231 x10^4	115 x10^4	115 x10^4	231 x10^4	115 x10^4	115 x10^4	115 x10^4	115 x10^4
34	214 x10^4	214 x10^6	107 x10^4	107 x10^4	107 x10^4	107 x10^4	214 x10^4	214 x10^4	107 x10^5	107 x10^4
35	214 x10^4	107 x10^4	107 x10^4	214 x10^4	107 x10^4	214 x10^6	214 x10^4	107 x10^4	107 x10^4	107 x10^5
36	107 x10^5	214 x10^4	214 x10^4	107 x10^4	107 x10^4	107 x10^4	214 x10^6	214 x10^4	107 x10^4	107 x10^4
37	107 x10^5	107 x10^4	214 x10^4	214 x10^6	107 x10^4	214 x10^4	107 x10^4	107 x10^4	107 x10^4	214 x10^4
38	107 x10^4	214 x10^4	214 x10^4	107 x10^4	214 x10^6	214 x10^4	107 x10^4	107 x10^4	107 x10^4	107 x10^5
39	200 x10^6	100 x10^4	100 x10^4	200 x10^4	200 x10^4	100 x10^4	200 x10^4	100 x10^4	100 x10^5	200 x10^4
40	100 x10^4	100 x10^5	100 x10^4	100 x10^4	200 x10^4	100 x10^4	200 x10^6	200 x10^4	200 x10^4	200 x10^4
41	107 x10^5	214 x10^5	107 x10^6	107 x10^5	107 x10^5	214 x10^6	214 x10^4	214 x10^6	107 x10^4	107 x10^7
42	125 x10^4	125 x10^4	125 x10^6	125 x10^6	125 x10^6	125 x10^6	250 x10^4	125 x10^7	250 x10^4	125 x10^6
43	214 x10^6	214 x10^6	107 x10^4	214 x10^6	107 x10^6	107 x10^4	107 x10^4	214 x10^4	107 x10^4	107 x10^7
44	188 x10^4	188 x10^5	94 x10^4	94 x10^7	94 x10^6	94 x10^7	188 x10^5	188 x10^5	188 x10^6	188 x10^5
45	214 x10^6	107 x10^6	214 x10^4	214 x10^4	107 x10^6	107 x10^5	107 x10^6	214 x10^4	107 x10^4	107 x10^7
46	107 x10^7	107 x10^5	107 x10^6	214 x10^5	107 x10^5	107 x10^6	107 x10^7	214 x10^5	214 x10^5	214 x10^4
47	107 x10^7	214 x10^5	107 x10^6	214 x10^4	214 x10^6	107 x10^5	214 x10^4	107 x10^4	107 x10^4	107 x10^7
48	200 x10^5	200 x10^5	100 x10^4	100 x10^4	200 x10^6	200 x10^4	100 x10^7	100 x10^4	100 x10^7	200 x10^4
49	188 x10^4	188 x10^4	188 x10^5	94 x10^5	188 x10^6	94 x10^6	188 x10^5	188 x10^5	94 x10^7	94 x10^6
50	100 x10^4	100 x10^6	100 x10^7	200 x10^5	200 x10^6	100 x10^6	200 x10^5	200 x10^4	200 x10^4	100 x10^4

Figure C.2: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Badllus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
51	188 x10^4	188 x10^6	188 x10^4	188 x10^5	94 x10^5	188 x10^6	94 x10^7	188 x10^5	94 x10^5	94 x10^6
52	214 x10^4	214 x10^6	107 x10^7	214 x10^4	107 x10^5	107 x10^6	107 x10^6	214 x10^5	107 x10^4	107 x10^5
53	167 x10^4	167 x10^5	167 x10^6	167 x10^5	167 x10^5	167 x10^4	167 x10^4	83 x10^5	83 x10^4	167 x10^5
54	94 x10^4	94 x10^5	94 x10^6	188 x10^5	188 x10^6	188 x10^4	188 x10^4	188 x10^4	188 x10^4	94 x10^4
55	200 x10^4	100 x10^4	200 x10^4	100 x10^6	100 x10^4	100 x10^5	200 x10^6	100 x10^5	200 x10^5	200 x10^4
56	107 x10^4	107 x10^4	107 x10^6	107 x10^5	214 x10^4	107 x10^5	214 x10^4	214 x10^5	214 x10^6	107 x10^5
57	176 x10^4	176 x10^4	176 x10^4	176 x10^4	176 x10^6	88 x10^6	176 x10^6	88 x10^5	176 x10^4	88 x10^4
58	176 x10^4	176 x10^5	176 x10^5	176 x10^4	88 x10^6	88 x10^4	176 x10^6	176 x10^4	88 x10^4	176 x10^4
59	200 x10^4	200 x10^6	100 x10^4	100 x10^4	200 x10^5	200 x10^4	100 x10^6	200 x10^4	100 x10^4	100 x10^5
60	88 x10^5	176 x10^5	176 x10^4	176 x10^5	176 x10^6	88 x10^6	176 x10^4	176 x10^4	176 x10^4	88 x10^4
61	100 x10^6	100 x10^6	200 x10^6	200 x10^6	100 x10^6	200 x10^6	200 x10^6	100 x10^7	200 x10^6	100 x10^6
62	200 x10^6	100 x10^6	100 x10^6	200 x10^6	100 x10^6	100 x10^6	200 x10^6	200 x10^6	200 x10^6	100 x10^7
63	176 x10^5	88 x10^7	176 x10^6	176 x10^6	176 x10^6	176 x10^6	176 x10^6	88 x10^6	176 x10^6	88 x10^6
64	176 x10^6	176 x10^6	88 x10^6	88 x10^7	176 x10^6	88 x10^6	176 x10^6	176 x10^5	176 x10^6	176 x10^6
65	214 x10^6	107 x10^6	107 x10^6	214 x10^6	107 x10^6	107 x10^6	107 x10^7	214 x10^6	214 x10^6	107 x10^6
66	158 x10^6	158 x10^6	158 x10^6	158 x10^6	158 x10^5	79 x10^7	158 x10^6	158 x10^6	158 x10^5	158 x10^6
67	188 x10^6	94 x10^6	94 x10^6	188 x10^6	188 x10^6	94 x10^4	188 x10^6	94 x10^7	188 x10^6	188 x10^6
68	94 x10^6	94 x10^7	188 x10^6	188 x10^6	188 x10^6	94 x10^6	188 x10^6	188 x10^6	94 x10^4	188 x10^6
69	188 x10^6	94 x10^7	188 x10^6	94 x10^6	94 x10^4	188 x10^6	188 x10^6	94 x10^6	188 x10^6	188 x10^6
70	188 x10^5	94 x10^6	188 x10^6	188 x10^6	188 x10^6	188 x10^6	188 x10^6	94 x10^6	94 x10^7	94 x10^6
71	188 x10^6	188 x10^6	188 x10^6	188 x10^6	94 x10^7	188 x10^6	188 x10^5	94 x10^6	94 x10^6	94 x10^6
72	188 x10^6	94 x10^6	188 x10^6	188 x10^5	188 x10^6	94 x10^7	188 x10^6	94 x10^6	188 x10^6	94 x10^6
73	94 x10^7	188 x10^6	94 x10^6	188 x10^6	94 x10^6	94 x10^6	188 x10^5	188 x10^6	188 x10^6	188 x10^6
74	94 x10^6	94 x10^6	188 x10^6	188 x10^6	188 x10^6	188 x10^6	188 x10^6	94 x10^6	94 x10^7	188 x10^5
75	188 x10^5	188 x10^6	188 x10^6	94 x10^6	188 x10^6	188 x10^6	94 x10^6	94 x10^6	188 x10^6	94 x10^7

Figure C.3: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Badllus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
76	167 x10^6	167 x10^6	167 x10^6	167 x10^6	83 x10^4	167 x10^6	167 x10^6	167 x10^6	83 x10^6	167 x10^6
77	200 x10^5	100 x10^6	200 x10^6	100 x10^6	100 x10^6	200 x10^6	100 x10^6	100 x10^6	200 x10^6	200 x10^6
78	200 x10^5	100 x10^5	200 x10^5	100 x10^5	100 x10^5	200 x10^5	200 x10^5	200 x10^5	100 x10^5	100 x10^5
79	83 x10^5	167 x10^5	167 x10^5	167 x10^5	167 x10^5	167 x10^5	167 x10^5	167 x10^5	167 x10^4	83 x10^5
80	176 x10^5	176 x10^5	176 x10^5	176 x10^5	88 x10^5	176 x10^4	88 x10^5	176 x10^5	176 x10^5	88 x10^5
81	188 x10^5	188 x10^5	188 x10^5	188 x10^4	188 x10^5	188 x10^5	94 x10^5	94 x10^5	94 x10^5	94 x10^5
82	188 x10^4	188 x10^5	94 x10^5	188 x10^5	94 x10^5	94 x10^5	188 x10^5	188 x10^5	94 x10^5	188 x10^5
83	100 x10^5	100 x10^5	200 x10^5	200 x10^5	100 x10^5	200 x10^5	200 x10^4	200 x10^5	100 x10^5	100 x10^5
84	125 x10^5	125 x10^5	125 x10^5	125 x10^5	125 x10^4	125 x10^5	125 x10^5	250 x10^4	250 x10^5	125 x10^5
85	125 x10^5	125 x10^5	125 x10^5	250 x10^4	125 x10^5	125 x10^5	125 x10^5	125 x10^4	250 x10^5	125 x10^5
86	107 x10^5	214 x10^5	107 x10^5	107 x10^5	214 x10^4	107 x10^5	107 x10^5	214 x10^5	107 x10^5	214 x10^4
87	107 x10^5	214 x10^4	214 x10^4	107 x10^5	107 x10^5	107 x10^5	214 x10^5	107 x10^5	214 x10^5	107 x10^5
88	107 x10^5	107 x10^5	214 x10^4	214 x10^4	214 x10^5	107 x10^5	214 x10^5	107 x10^5	107 x10^5	107 x10^5
89	83 x10^5	83 x10^5	167 x10^5	167 x10^5	167 x10^4	167 x10^5	167 x10^5	167 x10^5	167 x10^4	167 x10^5
90	176 x10^4	88 x10^5	176 x10^5	88 x10^5	176 x10^5	176 x10^5	88 x10^5	176 x10^5	176 x10^5	176 x10^4
91	176 x10^4	176 x10^5	176 x10^4	176 x10^5	88 x10^5	176 x10^5	176 x10^5	88 x10^5	176 x10^5	88 x10^5
92	94 x10^5	94 x10^5	188 x10^4	188 x10^5	188 x10^5	94 x10^5	188 x10^4	188 x10^5	94 x10^5	188 x10^5
93	94 x10^5	188 x10^4	188 x10^5	94 x10^5	188 x10^4	94 x10^5	94 x10^5	188 x10^5	188 x10^5	188 x10^5
94	107 x10^5	214 x10^5	107 x10^6	107 x10^5	107 x10^5	214 x10^6	214 x10^4	214 x10^6	107 x10^4	107 x10^7
95	100 x10^5	200 x10^4	100 x10^4	100 x10^5	200 x10^5	200 x10^5	100 x10^5	100 x10^6	200 x10^7	200 x10^5
96	94 x10^5	188 x10^4	94 x10^6	188 x10^5	188 x10^5	188 x10^4	94 x10^5	94 x10^7	188 x10^4	188 x10^4
97	200 x10^6	200 x10^6	100 x10^4	100 x10^7	200 x10^6	200 x10^6	100 x10^4	100 x10^6	200 x10^6	100 x10^5
98	94 x10^6	188 x10^6	94 x10^5	188 x10^4	94 x10^6	188 x10^4	94 x10^7	188 x10^6	188 x10^6	188 x10^5
99	200 x10^4	100 x10^7	100 x10^5	100 x10^6	100 x10^6	200 x10^4	200 x10^5	200 x10^4	100 x10^4	200 x10^5
100	200 x10^6	200 x10^4	200 x10^6	200 x10^4	100 x10^4	200 x10^5	100 x10^7	100 x10^5	100 x10^5	100 x10^5

Figure C.4: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Badillus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
101	176 x10^4	88 x10^5	88 x10^7	176 x10^4	176 x10^6	176 x10^6	176 x10^6	176 x10^5	88 x10^6	176 x10^6
102	188 x10^5	188 x10^4	94 x10^5	94 x10^6	188 x10^7	188 x10^5	188 x10^4	94 x10^5	94 x10^5	188 x10^5
103	214 x10^5	107 x10^5	107 x10^5	214 x10^5	107 x10^5	214 x10^5	214 x10^5	107 x10^5	107 x10^4	107 x10^4
104	188 x10^5	94 x10^4	188 x10^7	188 x10^4	188 x10^4	188 x10^5	94 x10^4	94 x10^5	94 x10^4	188 x10^4
105	231 x10^4	231 x10^6	115 x10^6	115 x10^4	115 x10^4	115 x10^4	115 x10^4	115 x10^7	231 x10^6	115 x10^4
106	200 x10^4	200 x10^4	100 x10^6	100 x10^6	200 x10^6	200 x10^5	100 x10^7	200 x10^6	100 x10^5	100 x10^5
107	167 x10^4	167 x10^5	167 x10^6	167 x10^5	167 x10^5	167 x10^4	167 x10^4	83 x10^5	83 x10^4	167 x10^5
108	188 x10^4	188 x10^5	94 x10^4	94 x10^5	188 x10^4	188 x10^4	94 x10^6	188 x10^7	94 x10^5	188 x10^5
109	100 x10^6	100 x10^6	200 x10^6	200 x10^6	200 x10^6	100 x10^6	100 x10^6	100 x10^4	200 x10^5	200 x10^5
110	94 x10^4	94 x10^5	94 x10^6	188 x10^5	188 x10^6	188 x10^4	188 x10^4	188 x10^4	188 x10^4	94 x10^4
111	125 x10^5	125 x10^6	125 x10^4	250 x10^5	250 x10^4	125 x10^5	125 x10^5	125 x10^5	125 x10^5	125 x10^7
112	188 x10^4	188 x10^5	94 x10^4	94 x10^4	94 x10^7	94 x10^5	188 x10^5	188 x10^6	188 x10^5	188 x10^5
113	200 x10^4	100 x10^4	200 x10^4	100 x10^6	100 x10^4	100 x10^5	200 x10^6	100 x10^5	200 x10^5	200 x10^4
114	88 x10^5	176 x10^6	176 x10^5	176 x10^5	176 x10^6	176 x10^5	176 x10^6	88 x10^5	176 x10^5	88 x10^5
115	176 x10^6	88 x10^4	88 x10^4	176 x10^5	176 x10^7	176 x10^4	176 x10^4	88 x10^4	176 x10^5	176 x10^6
116	107 x10^5	107 x10^4	214 x10^5	107 x10^6	107 x10^4	214 x10^6	214 x10^5	214 x10^4	107 x10^6	107 x10^6
117	188 x10^4	188 x10^7	188 x10^6	94 x10^5	188 x10^4	94 x10^5	94 x10^4	94 x10^5	188 x10^6	188 x10^4
118	214 x10^4	107 x10^4	214 x10^4	107 x10^5	214 x10^6	107 x10^6	107 x10^4	107 x10^6	214 x10^5	107 x10^5
119	200 x10^5	200 x10^6	200 x10^5	100 x10^4	200 x10^5	200 x10^5	100 x10^4	100 x10^6	100 x10^7	100 x10^4
120	107 x10^4	107 x10^6	214 x10^4	107 x10^6	214 x10^7	107 x10^5	107 x10^5	214 x10^4	214 x10^5	107 x10^4
121	94 x10^5	188 x10^4	94 x10^6	94 x10^5	188 x10^5	188 x10^4	188 x10^6	94 x10^5	188 x10^7	188 x10^5
122	200 x10^6	200 x10^6	200 x10^5	100 x10^5	200 x10^5	100 x10^4	100 x10^6	100 x10^4	200 x10^5	100 x10^6
123	176 x10^4	88 x10^4	176 x10^4	88 x10^5	176 x10^4	88 x10^5	176 x10^7	176 x10^5	176 x10^5	176 x10^5
124	188 x10^4	188 x10^4	188 x10^5	94 x10^5	188 x10^6	94 x10^6	188 x10^5	188 x10^5	94 x10^7	94 x10^6
125	115 x10^4	231 x10^5	231 x10^4	115 x10^7	231 x10^5	115 x10^4	115 x10^6	115 x10^4	115 x10^5	115 x10^7

Figure C.5: Volume in μL from concentrations where white = $10^7 \times diluted$, red = $10^6 \times diluted$, green = $10^5 \times diluted$, blue = $10^4 \times diluted$.

	10 TYP	1 Badllus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
126	125 x10^4	125 x10^4	125 x10^6	125 x10^6	125 x10^6	125 x10^6	250 x10^4	125 x10^7	250 x10^4	125 x10^6
127	188 x10^7	188 x10^4	188 x10^6	94 x10^4	94 x10^6	188 x10^4	188 x10^4	94 x10^5	188 x10^5	94 x10^6
128	176 x10^5	176 x10^5	176 x10^4	88 x10^6	176 x10^6	176 x10^6	176 x10^5	88 x10^6	176 x10^5	88 x10^4
129	214 x10^6	214 x10^6	107 x10^4	214 x10^6	107 x10^6	107 x10^4	107 x10^4	214 x10^4	107 x10^4	107 x10^7
130	115 x10^5	231 x10^7	231 x10^5	115 x10^4	115 x10^4	115 x10^6	115 x10^5	115 x10^4	231 x10^4	115 x10^4
131	188 x10^5	94 x10^5	188 x10^5	94 x10^5	188 x10^6	188 x10^6	188 x10^6	94 x10^5	94 x10^5	188 x10^4
132	100 x10^4	100 x10^6	100 x10^7	200 x10^5	200 x10^6	100 x10^6	200 x10^5	200 x10^4	200 x10^4	100 x10^4
133	88 x10^6	88 x10^5	88 x10^7	176 x10^5	176 x10^4	176 x10^4	176 x10^4	176 x10^4	176 x10^5	176 x10^5
134	150 x10^5	150 x10^7	150 x10^4	150 x10^6	150 x10^5	150 x10^5	150 x10^4	150 x10^5	150 x10^6	150 x10^4
135	188 x10^6	94 x10^4	94 x10^4	188 x10^4	188 x10^4	188 x10^5	188 x10^4	188 x10^7	94 x10^5	94 x10^5
136	188 x10^4	188 x10^5	94 x10^4	94 x10^7	94 x10^6	94 x10^7	188 x10^5	188 x10^5	188 x10^6	188 x10^5
137	158 x10^5	158 x10^4	158 x10^4	158 x10^5	158 x10^7	158 x10^6	158 x10^5	158 x10^6	79 x10^5	158 x10^5
138	100 x10^5	200 x10^5	200 x10^6	100 x10^5	100 x10^7	100 x10^5	200 x10^4	200 x10^6	100 x10^5	200 x10^4
139	107 x10^4	107 x10^4	214 x10^6	214 x10^4	107 x10^5	107 x10^4	107 x10^6	214 x10^4	107 x10^7	214 x10^4
140	125 x10^7	125 x10^4	125 x10^4	125 x10^4	125 x10^5	125 x10^5	250 x10^5	125 x10^6	125 x10^6	250 x10^5
141	200 x10^5	100 x10^6	200 x10^4	100 x10^4	100 x10^6	200 x10^4	100 x10^4	200 x10^4	200 x10^4	100 x10^6
142	231 x10^6	231 x10^6	115 x10^6	115 x10^5	115 x10^5	115 x10^4	231 x10^6	115 x10^5	115 x10^6	115 x10^6
143	188 x10^4	188 x10^6	188 x10^4	188 x10^5	94 x10^5	188 x10^6	94 x10^7	188 x10^5	94 x10^5	94 x10^6
144	88 x10^7	88 x10^6	176 x10^5	88 x10^6	176 x10^4	176 x10^6	176 x10^6	176 x10^6	176 x10^6	176 x10^4
145	188 x10^6	188 x10^4	188 x10^6	188 x10^5	94 x10^4	94 x10^6	94 x10^4	188 x10^4	94 x10^4	188 x10^5
146	100 x10^5	200 x10^5	200 x10^4	200 x10^6	200 x10^6	100 x10^6	100 x10^6	100 x10^4	100 x10^6	200 x10^4
147	214 x10^6	107 x10^6	214 x10^4	214 x10^4	107 x10^6	107 x10^5	107 x10^6	214 x10^4	107 x10^4	107 x10^7
148	115 x10^7	115 x10^4	231 x10^6	115 x10^5	115 x10^6	115 x10^5	231 x10^5	115 x10^6	115 x10^4	231 x10^6
149	107 x10^7	107 x10^5	107 x10^6	214 x10^5	107 x10^5	107 x10^6	107 x10^7	214 x10^5	214 x10^5	214 x10^4
150	115 x10^6	231 x10^6	115 x10^5	115 x10^6	231 x10^5	115 x10^5	231 x10^6	115 x10^5	115 x10^7	115 x10^4

Figure C.6: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Badillus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
151	100 x10^7	100 x10^4	200 x10^4	100 x10^4	200 x10^5	200 x10^5	200 x10^5	200 x10^4	100 x10^6	100 x10^4
152	125 x10^5	125 x10^4	250 x10^4	125 x10^6	125 x10^6	125 x10^4	125 x10^5	125 x10^6	125 x10^5	250 x10^4
153	100 x10^5	100 x10^7	200 x10^5	200 x10^5	100 x10^6	200 x10^6	200 x10^4	200 x10^5	100 x10^5	100 x10^7
154	273 x10^5	136 x10^6	136 x10^4	136 x10^4	136 x10^4	136 x10^4	136 x10^4	136 x10^5	136 x10^6	136 x10^6
155	100 x10^4	100 x10^5	100 x10^4	200 x10^6	100 x10^7	200 x10^4	200 x10^4	200 x10^4	200 x10^5	100 x10^6
156	100 x10^4	200 x10^6	200 x10^6	200 x10^5	200 x10^5	100 x10^5	200 x10^5	100 x10^6	100 x10^6	100 x10^5
157	188 x10^6	188 x10^6	94 x10^6	94 x10^5	188 x10^4	188 x10^5	188 x10^4	188 x10^5	94 x10^5	94 x10^5
158	107 x10^4	107 x10^4	107 x10^6	107 x10^5	214 x10^4	107 x10^5	214 x10^4	214 x10^5	214 x10^6	107 x10^5
159	214 x10^4	214 x10^6	107 x10^7	214 x10^4	107 x10^5	107 x10^6	107 x10^6	214 x10^5	107 x10^4	107 x10^5
160	214 x10^6	214 x10^6	107 x10^5	214 x10^4	214 x10^5	107 x10^4	107 x10^6	107 x10^4	107 x10^6	107 x10^4
161	214 x10^6	214 x10^6	107 x10^7	107 x10^6	107 x10^5	214 x10^6	214 x10^6	107 x10^5	107 x10^4	107 x10^4
162	214 x10^4	107 x10^7	107 x10^4	107 x10^4	214 x10^6	107 x10^4	214 x10^4	214 x10^6	107 x10^4	107 x10^4
163	176 x10^4	176 x10^6	88 x10^6	88 x10^5	176 x10^5	176 x10^5	176 x10^6	176 x10^5	176 x10^6	88 x10^7
164	107 x10^4	107 x10^7	214 x10^5	107 x10^4	107 x10^5	214 x10^6	107 x10^5	214 x10^5	107 x10^4	214 x10^4
165	200 x10^5	100 x10^6	100 x10^6	200 x10^6	100 x10^5	200 x10^4	200 x10^6	100 x10^6	200 x10^4	100 x10^6
166	100 x10^4	200 x10^4	200 x10^5	200 x10^7	100 x10^4	200 x10^5	200 x10^4	100 x10^4	100 x10^4	100 x10^6
167	176 x10^4	176 x10^4	88 x10^4	88 x10^5	176 x10^6	88 x10^7	176 x10^6	176 x10^6	176 x10^6	176 x10^6
168	100 x10^5	200 x10^6	100 x10^6	100 x10^5	100 x10^4	200 x10^4	200 x10^6	100 x10^7	200 x10^4	200 x10^4
169	100 x10^6	200 x10^4	100 x10^5	200 x10^6	200 x10^5	200 x10^4	100 x10^4	100 x10^5	100 x10^6	200 x10^7
170	231 x10^5	115 x10^5	115 x10^6	231 x10^4	115 x10^5	231 x10^6	115 x10^6	115 x10^5	115 x10^4	115 x10^4
171	250 x10^5	125 x10^4	125 x10^4	250 x10^4	125 x10^6	125 x10^4	125 x10^6	125 x10^5	125 x10^5	125 x10^7
172	115 x10^4	231 x10^4	115 x10^4	231 x10^5	115 x10^4	115 x10^6	115 x10^6	115 x10^7	115 x10^5	231 x10^5
173	94 x10^4	94 x10^7	94 x10^4	188 x10^6	188 x10^6	94 x10^6	188 x10^4	188 x10^4	188 x10^4	188 x10^5
174	231 x10^4	115 x10^5	115 x10^4	231 x10^5	115 x10^7	115 x10^7	115 x10^5	231 x10^5	115 x10^4	115 x10^5
175	200 x10^6	200 x10^5	100 x10^5	100 x10^7	200 x10^5	100 x10^4	200 x10^6	100 x10^5	100 x10^5	200 x10^6

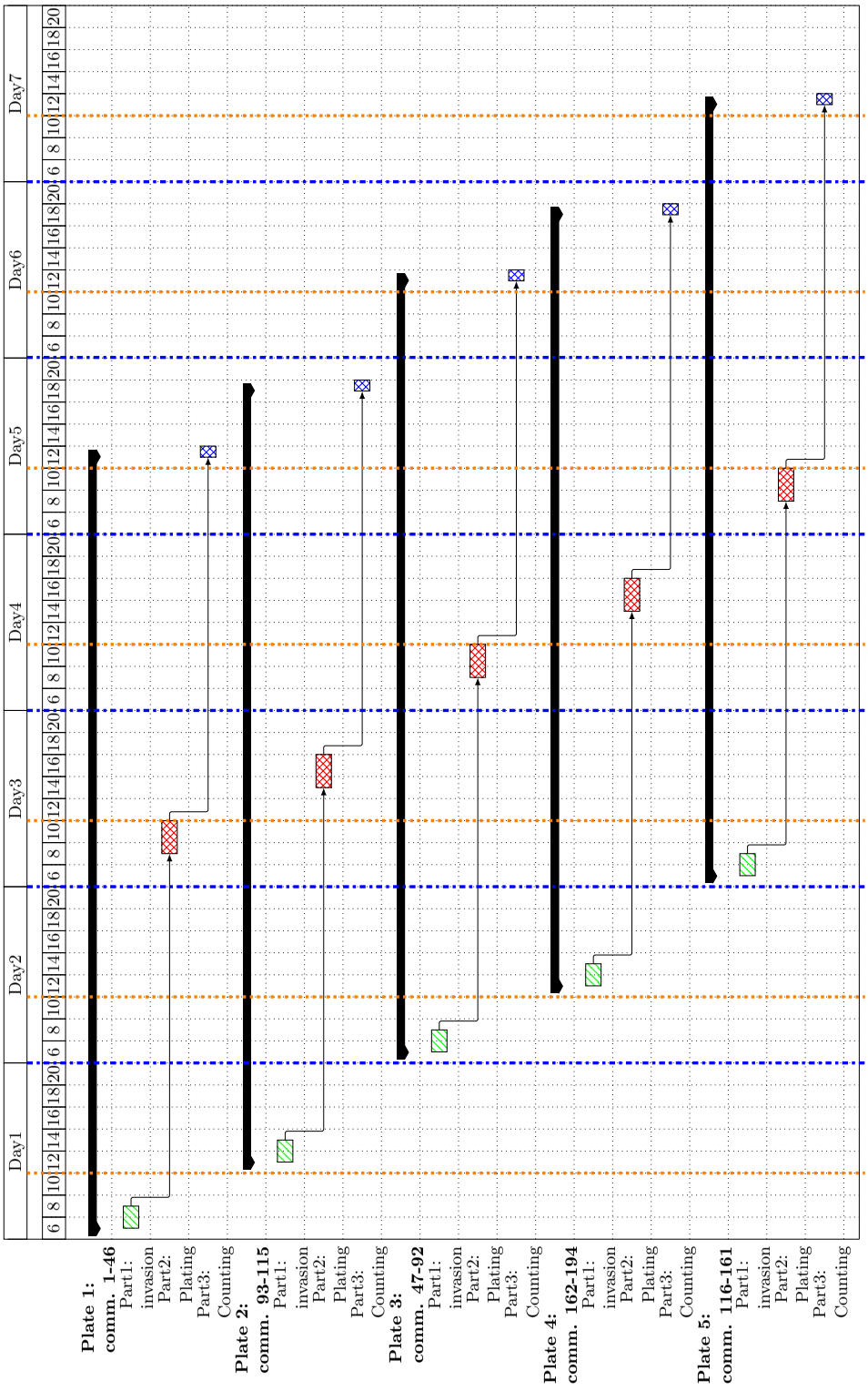
Figure C.7: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

	10 TYP	1 Bacillus	14.3 ISO1	25.2 Burkho	42 Para	59 Entero	Beijirinckia	63 Rhizi	Delftia	K62 Aeromo
176	214 x10^6	214 x10^6	107 x10^5	214 x10^4	214 x10^5	107 x10^4	107 x10^6	107 x10^4	107 x10^6	107 x10^4
177	214 x10^6	214 x10^6	107 x10^7	107 x10^6	107 x10^5	214 x10^6	214 x10^6	107 x10^5	107 x10^4	107 x10^4
178	214 x10^4	107 x10^7	107 x10^4	107 x10^4	214 x10^6	107 x10^4	214 x10^4	214 x10^6	107 x10^4	107 x10^4
179	176 x10^4	176 x10^6	88 x10^6	88 x10^5	176 x10^5	176 x10^5	176 x10^6	176 x10^5	176 x10^6	88 x10^7
180	107 x10^4	107 x10^7	214 x10^5	107 x10^4	107 x10^5	214 x10^6	107 x10^5	214 x10^5	107 x10^4	214 x10^4
181	200 x10^5	100 x10^6	100 x10^6	200 x10^6	100 x10^5	200 x10^4	200 x10^6	100 x10^6	200 x10^4	100 x10^6
182	100 x10^4	200 x10^4	200 x10^5	200 x10^7	100 x10^4	200 x10^5	200 x10^4	100 x10^4	100 x10^4	100 x10^6
183	176 x10^4	176 x10^4	88 x10^4	88 x10^5	176 x10^6	88 x10^7	176 x10^6	176 x10^6	176 x10^6	176 x10^6
184	100 x10^5	200 x10^6	100 x10^6	100 x10^5	100 x10^4	200 x10^4	200 x10^6	100 x10^7	200 x10^4	200 x10^4
185	100 x10^6	200 x10^4	100 x10^5	200 x10^6	200 x10^5	200 x10^4	100 x10^4	100 x10^5	100 x10^6	200 x10^7
186	231 x10^5	115 x10^5	115 x10^6	231 x10^4	115 x10^5	231 x10^6	115 x10^6	115 x10^5	115 x10^4	115 x10^4
187	250 x10^5	125 x10^4	125 x10^4	250 x10^4	125 x10^6	125 x10^4	125 x10^6	125 x10^5	125 x10^5	125 x10^7
188	115 x10^4	231 x10^4	115 x10^4	231 x10^5	115 x10^4	115 x10^6	115 x10^6	115 x10^7	115 x10^5	231 x10^5
189	94 x10^4	94 x10^7	94 x10^4	188 x10^6	188 x10^6	94 x10^6	188 x10^4	188 x10^4	188 x10^4	188 x10^5
190	231 x10^4	115 x10^5	115 x10^4	231 x10^5	115 x10^7	115 x10^7	115 x10^5	231 x10^5	115 x10^4	115 x10^5
191	200 x10^6	200 x10^5	100 x10^5	100 x10^7	200 x10^5	100 x10^4	200 x10^6	100 x10^5	100 x10^5	200 x10^6
192	100 x10^6	200 x10^5	200 x10^4	100 x10^6	100 x10^4	200 x10^4	100 x10^6	100 x10^6	200 x10^6	200 x10^5
193	214 x10^5	107 x10^6	214 x10^6	107 x10^7	214 x10^6	107 x10^4	107 x10^4	107 x10^4	214 x10^6	107 x10^4
194	200 x10^6	200 x10^5	200 x10^5	100 x10^5	200 x10^6	100 x10^4	100 x10^4	200 x10^4	100 x10^7	100 x10^5

Figure C.8: Volume in µL from concentrations where white = $10^7 \times$ *diluted*, red = $10^6 \times$ *diluted*, green = $10^5 \times$ *diluted*, blue = $10^4 \times$ *diluted*.

Appendix D

Workflow


Day1
Day2
Day3
Day4
Day5
Day6
Day7
6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20 6 8 10 12 14 16 18 20
Plate 1:
comm. 1-46
Part1: invasion
Part2: Plating
Part3: Counting
Plate 2:
comm. 93-115
Part1: invasion
Part2: Plating
Part3: Counting
Plate 3:
comm. 47-92
Part1: invasion
Part2: Plating
Part3: Counting
Plate 4:
comm. 162-194
Part1: invasion
Part2: Plating
Part3: Counting
Plate 5:
comm. 116-161
Part1: invasion
Part2: Plating
Part3: Counting

Appendix E

Tables

Table E.1: Pathogen list obtained from the Belgian Coordinated collections of micro-organisms after filtering on BSL and health risks (BCCM/LMG Bacteria Collection, Laboratory of Microbiology Ghent after filtering on BSL and health risks.

Bacterial species	Infect.	Strain No. Bel	Biological origin	Growth solid media	BSL
Corynebacterium urealyticum	_	LMG 19041	"patient, urine"	"Medium 14 or 151, 37°C"	2
Mycobacterium terrae	Y	LMG 10394	Sputum and gastric lavage specimen	"Medium 336, 37°C"	1
Mycobacterium peregrinum	Y	LMG 19256	bronchial aspiration of a child	"Medium 185, 28°C"	1
Mycobacterium smegmatis or Mycobacterium friedmannii	_	LMG 8190 & LMG 10267		"Medium 126, 37°C"	1 BIO/2 BIO
Nocardia asteroides	Y	LMG 4062		"Medium 78, 28°C"	2
Nocardia farcinica	Y	LMG 4079		"Medium 78, 37°C"	2
Rhodococcus equi (Corynebacterium equi)	_	LMG 7335 & LMG 18452		"Medium 1, 28°C"	2
Cellulosimicrobium cellulans	_	LMG 16121		"Medium 1 or 185, 28°C"	1
Elizabethkingia meningoseptica	_	LMG 12279	"premature infant, cerebrospinal fluid"	"Medium 1 or 185, 28°C"	2
Empedobacter brevis	_	LMG 4011	"human, bronchial secretions"	"Medium 1, 30°C"	1
Chryseobacterium indologenes (Flavobacteriumindologenes)	_	LMG 8337	"human, trachea at autopsy"	"Medium 1 or 185, 28°C"	2
Vagococcus fluvialis	_	LMG 9464	"chicken, faeces"	"Medium 217, 30°C"	1
Staphylococcus epidermidis	Y-	LMG 10273		"Medium 1, 37°C"	1
Staphylococcus schleiferi subsp.schleiferi	_	LMG 13347	jugular catheter	"Medium 1, 28°C"	2
Staphylococcus haemolyticus	Y-	LMG 13349	"Human, skin"	"Medium 1, 28°C"	2
Staphylococcus warneri	_	LMG 13354	"Human, skin"	"Medium 1, 28°C"	1
Staphylococcus caprae	_	LMG 19123	goat milk	"Medium 1, 37°C"	1
Staphylococcus schleiferi subsp.coagulans	_	LMG 19137 & LMG 22205	"dog, skin"	"Medium 185, 37°C"	2
Staphylococcus simulans	Y	LMG 20177	Oryza sativa	"Medium 1, 28°C"	Not Yet assigned
Staphylococcus xylosus	_	LMG 20180	Oryza sativa	"Medium 1, 37°C"	Not Yet assigned
Staphylococcus aureus subsp.anaerobius	Y-	LMG 22203	"young sheep, abscess"	"Medium 151, 28°C"	2
Staphylococcus aureus subsp.aureus	Y-	LMG 8064	"Human, lesion"	"Medium 1, 37°C"	2
Staphylococcus pseudintermedius	_	LMG 9079	"dog with paratoid salivary adentitis, blood culture"	"Medium 185, 28°C"	2
Lactococcus garvieae (Streptococcus garvieae)	_	LMG 8501	"Cow 830, mastitis sample"	"Medium 66, 30°C"	1

Continued on next page

Table E.1 – continued from previous page

Bacterial species	Infect.	Strain No. Bel	Biological origin	Growth solid media	BSL
Streptococcus uberis	_	LMG 9465		"Medium 22 or 66, 37°C"	1
Afipia felis	Y-	LMG 18884	"patient with cat scratch disease, lymph node tissue"	"Medium 115 or 14, 28°C"	2
Alcaligenes faecalis subsp.faecalis	_	LMG 1229		"Medium 1, 28°C"	Opportunistic pathogen
Bordetella bronchiseptica	_	LMG 1232	"dog, lung"	"Medium 1 or 185, 37°C"	2
Bordetella parapertussis	_	LMG 1816	patient suffering from whooping cough	"Medium 185, 37°C"	2
Bordetella hinzii	_	LMG 1872	sputum	"Medium 185, 37°C"	2
Oligella urethralis (Moraxella urethralis)	_	LMG 1015	Ear	"Medium 1 or 7, 28°C"	1
Burkholderia cepacia	_	LMG 1222	Allium cepa	"Medium 14, 28°C"	Opportunistic pathogen
Chromobacterium violaceum	Y-	LMG 1267	freshwater	"Medium 1, 25°C"	2
Aeromonas encheleia	_	LMG 13061	artesian well	"Medium 185, 30°C"	1
Aeromonas bestiarum	_	LMG 13444	Diseased fish	"Medium 185, 30°C"	1
Aeromonas eucrenophila	_	LMG 3774	"carp, ascites"	"Medium 185, 30°C"	1
Edwardsiella tarda	_	LMG 2793	"Human, faeces"	"Medium 1, 37°C"	2
Edwardsiella ictaluri	_	LMG 7860	"channel catfish (Ictalurus punctatus), enteric septicemia"	"Medium 22, 28°C"	2
Enterobacter aerogenes	_	LMG 2094	Sputum	"Medium 1, 30°C"	2
Enterobacter hormaechei	_	LMG 25826 & LMG 25827	"75-year-old male with metastatic lung cancer, blood"	"Medium 14, 28°C"	2
Enterobacter cloacae subsp.cloacae(Enterobacter cloacae)	Y-	LMG 2783	Cerebrospinal fluid	"Medium 1, 30°C"	2
Escherichia albertii	_	LMG 20972	"diarrhoeal child, stool"	"Medium 14, 28°C"	2
Escherichia fergusonii	_	LMG 7866	"1-year-old boy, faeces"	"Medium 1, 30°C"	2
Klebsiella oxytoca	Y-	LMG 3055	Pharyngeal tonsil	"Medium 1, 30°C"	2
Morganella morganii subsp.morgani	_	LMG 2097 & LMG 7874	"green-billed toucan (Ramphastos dicolorus), dysentery"	"Medium 1, 37°C"	2
Proteus morgani	Y	LMG 2097 & LMG 7874	"green-billed toucan (Ramphastos dicolorus), dysentery"	"Medium 1, 37°C"	2

Continued on next page

Table E.1 – continued from previous page

Bacterial species	Infect.	Strain No. Bel	Biological origin	Growth solid media	BSL
Plesiomonas shigelloides	_	LMG 4240		“Medium 1, 37°C”	2
Proteus mirabilis	Y-	LMG 2954		“Medium 1, 37°C”	2
Proteus inconstans_Providencia alcalifaciens/stuartii	_	LMG 7878	Faeces	“Medium 1, 37°C”	2
Providencia alcalifaciens_Proteus inconstans	_	LMG 7878	Faeces	“Medium 1, 37°C”	2
Raoultella planticola	Y-	LMG 3056	soil	“Medium 1, 30°C”	2
Raoultella terrigena	Y-	LMG 3203	used water	“Medium 1, 30°C”	1
Serratia marcescens subsp.marcescens	Y	LMG 2792	pond water	“Medium 1, 28°C”	Opportunistic pathogen
Shigella flexneri	_	LMG 10472 & LMG 21935		“Medium 1, 37°C”	2
Shigella sonnei	_	LMG 10473	“human, faeces”	“Medium 1, 37°C”	2
Moraxella osloensis	_	LMG 1043	Probably urine	“Medium 74 or 14, 33°C”	2
Moraxella catarrhalis	_	LMG 1133		“Medium 22 or 3, 33°C”	2
Moraxella(Subgenus Moraxella) bovis	Y-	LMG 986	“Cattle, pinkeye”	“Medium 3 or 7, 33°C”	1
Actinobacillus equuli subsp.equuli	_	LMG 3736	“Foal dead from sleepy-foal disease, blood”	“Medium 1 or 185, 30°C”	Not Yet assigned
Actinobacillus lignieresii	_	LMG 3737	“Bovine, lesions”	“Medium 4, 37°C”	Not Yet assigned
Pseudomonas aeruginosa	Y	LMG 1242		“Medium 2, 28°C”	2
“Stenotrophomonas maltophilia (Pseudomonas maltophilia, Xanthomonas maltophilia)”	_	LMG 957	“Human, blood culture”	“Medium 9 or 14, 28°C”	1

Table E.2: Pathogen list obtained from the Belgian Coordinated collections of micro-organisms after filtering on BSL and health risks (BCCM/LMG Bacteria Collection, Laboratory of Microbiology Ghent after filtering on BSL and health risks.

Bacterial species	**Phylum**	**Class**	**Order**	**Family**	**Genus**
Bacterial species	Phylum	Class	Order	Family	Genus
Corynebacterium urealyticum	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Corynebacteriaceae
Mycobacterium terrae	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Mycobacteriaceae
Mycobacterium peregrinum	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Mycobacteriaceae
Mycobacterium smegmatis or Mycobacterium friedmannii	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Mycobacteriaceae
Nocardia asteroides	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Nocardiaceae
Nocardia farcinica	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Nocardiaceae
Rhodococcus equi (Corynebacterium equi)	Actinobacteria	Actinobacteridae	Actinomycetales	Corynebacterineae	Nocardiaceae
Cellulosimicrobium cellulans	Actinobacteria	Actinobacteridae	Actinomycetales	Micrococcineae	Promicromonosporaceae
Elizabethkingia meningoseptica	Bacteroidetes / Chlorobi group	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Empedobacter brevis	Bacteroidetes / Chlorobigroup	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Chryseobacterium indologenes(Flavobacteriumindologenes)	Bacteroidetes / Chlorobigroup	Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae
Vagococcus fluvialis	Firmicutes	Bacilli	Lactobacillales	Enterococcaceae	Vagococcus
Staphylococcus epidermidis	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus schleiferi subsp.schleiferi	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus haemolyticus	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus warneri	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus caprae	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus schleiferi subsp.coagulans	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus simulans	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus xylosus	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus aureus subsp.anaerobius	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus aureus subsp.aureus	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Staphylococcus pseudintermedius	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	Staphylococcus
Lactococcus garvieae (Streptococcus garvieae)	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Lactococcus
Streptococcus uberis	Firmicutes	Bacilli	Lactobacillales	Streptococcaceae	Streptococcus
Afipia felis	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	Afipia

Continued on next page

Table E.2 – continued from previous page

Bacterial species	Phylum	Class	Order	Family	Genus
Alcaligenes faecalis subsp.faecalis	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Alcaligenes
Bordetella bronchiseptica	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Bordetella
Bordetella parapertussis	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Bordetella
Bordetella hinzii	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Bordetella
Oligella urethralis (Moraxella urethralis)	Proteobacteria	Betaproteobacteria	Burkholderiales	Alcaligenaceae	Oligella
Burkholderia cepacia	Proteobacteria	Betaproteobacteria	Burkholderiales	Burkholderiaceae	Burkholderia
Chromobacterium violaceum	Proteobacteria	Betaproteobacteria	Neisseriales	Chromobacteriaceae	Chromobacterium group
Aeromonas encheleia	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas
Aeromonas bestiarum	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas
Aeromonas eucrenophila	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas
Edwardsiella tarda	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Edwardsiella
Edwardsiella ictaluri	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Edwardsiella
Enterobacter aerogenes	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter
Enterobacter hormaechei	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter
Enterobacter cloacae subsp.cloacae(Enterobacter cloacae)	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Enterobacter
Escherichia albertii	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia
Escherichia fergusonii	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia
Klebsiella oxytoca	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Klebsiella
Morganella morganii subsp.morgani	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Morganella
Proteus morgani	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Morganella
Plesiomonas shigelloides	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Plesiomonas
Proteus mirabilis	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Proteus
Proteus inconstans____Providencia alcalifaciens/stuartii	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Proteus
Providencia alcalifaciens____Proteus inconstans	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Providencia
Raoultella planticola	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Raoultella
Raoultella terrigena	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Raoultella
Serratia marcescens subsp.marcescens	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Serratia
Shigella flexneri	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Shigella

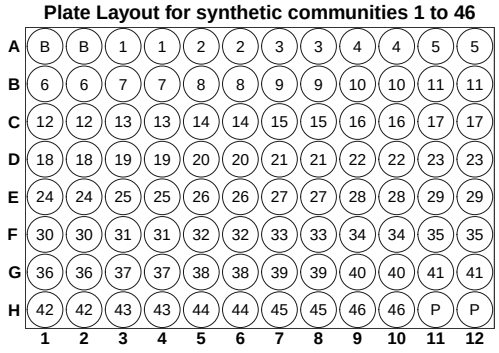
Continued on next page

Table E.2 – continued from previous page

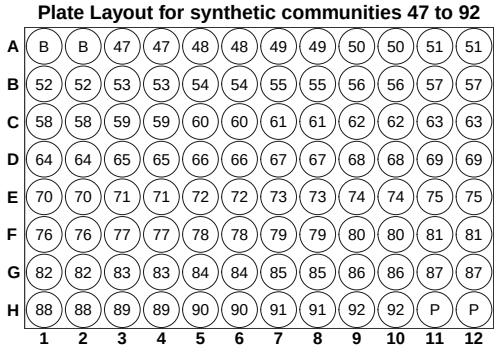
Bacterial species	Phylum	Class	Order	Family	Genus
Shigella sonnei	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Shigella
Moraxella osloensis	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella
Moraxella catarrhalis	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella
Moraxella(Subgenus Moraxella) bovis	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Moraxella
Actinobacillus equuli subsp.equuli	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Actinobacillus
Actinobacillus lignieresii	Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Actinobacillus
Pseudomonas aeruginosa	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas
"Stenotrophomonas maltophilia (Pseudomonas maltophilia, Xanthomonas maltophilia)"	Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas

Appendix F

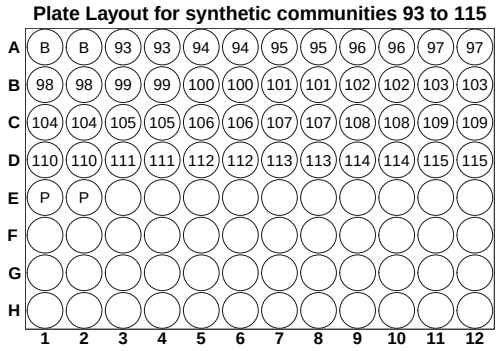
Configuration for the synthetic communities stock



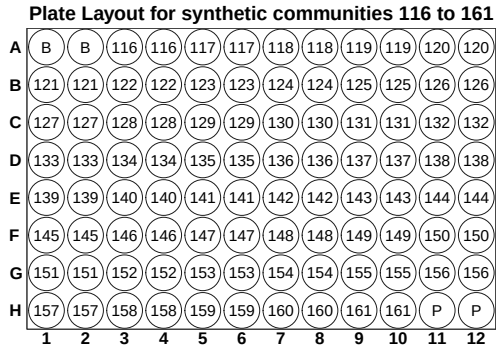
(a) Communities 1 to 46.



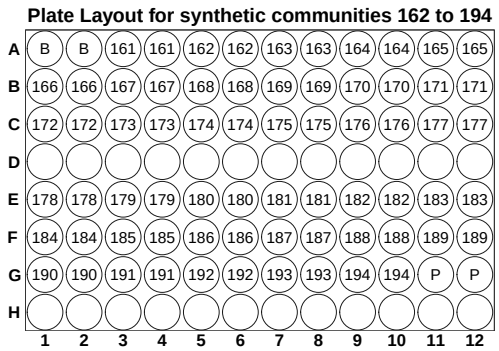
(b) Communities 47 to 92.



(c) Communities 93 to 115.



(d) Communities 116 to 161.



(e) Communities 162 to 194.

Figure F.1: Lay-out for normal 96 well plates. Each well contains 50µL of the resp. community. The letter B are left empty, as they are intended as blanks. The letter P is left empty as a control for the pathogen, without any synthetic commnity. Figures are made using R-package

Appendix G

Plating lay-out for synthetic communities

	1	2	3	4	5	6
A	B	B	1	1	2	2
B	3	3	4	4	5	5
C	6	6	7	7	8	8
D	9	9	10	10	11	11
E	12	12	13	13	14	14
F	15	15	16	16	17	17

(a) Communities 1 to 17.

	1	2	3	4	5	6
A	18	18	19	19	20	20
B	21	21	22	22	23	23
C	24	24	25	25	26	26
D	27	27	28	28	29	29
E	30	30	31	31	32	32
F	33	33	34	34	35	35

(b) Communities 18 to 35.

	1	2	3	4	5	6
A	36	36	37	37	38	38
B	39	39	40	40	41	41
C	42	42	43	43	44	44
D	45	45	46	46	P	P
E						
F						

(c) Communities 36 to 46.

Figure G.1: Plate lay-out for synthetic communities 1 to 46.

	1	2	3	4	5	6
A	B	B	47	47	48	48
B	49	49	50	50	51	51
C	52	52	53	53	54	54
D	55	55	56	56	57	57
E	58	58	59	59	60	60
F	61	61	62	62	63	63

(a) Communities 47 to 63.

	1	2	3	4	5	6
A	64	64	65	65	66	66
B	67	67	68	68	69	69
C	70	70	71	71	72	72
D	73	73	74	74	75	75
E	76	76	77	77	78	78
F	79	79	80	80	81	81

(b) Communities 64 to 81.

	1	2	3	4	5	6
A	82	82	83	83	84	84
B	85	85	86	86	87	87
C	88	88	89	89	90	90
D	91	91	92	92	P	P
E						
F						

(c) Communities 82 to 92.

Figure G.2: Plate lay-out for synthetic communities 47 to 92.

	1	2	3	4	5	6
A	B	B	93	93	94	94
B	95	95	96	96	97	97
C	98	98	99	99	100	100
D	101	101	102	102	103	103
E						
F						

(a) Communities 93 to 103.

	1	2	3	4	5	6
A	104	104	105	105	106	106
B	107	107	108	108	109	109
C	110	110	111	111	112	112
D	113	113	114	114	115	115
E	P	P				
F						

(b) Communities 104 to 115.

Figure G.3: Plate lay-out for synthetic communities 93 to 115.

	1	2	3	4	5	6
A	B	B	116	116	117	117
B	118	118	119	119	120	120
C	121	121	122	122	123	123
D	124	124	125	125	126	126
E	127	127	128	128	129	129
F	130	130	131	131	132	132

(a) Communities 116 to 132.

	1	2	3	4	5	6
A	133	133	134	134	135	135
B	136	136	137	137	138	138
C	139	139	140	140	141	141
D	142	142	143	143	144	144
E	145	145	146	146	147	147
F	148	148	149	149	150	150

(b) Communities 133 to 150.

	1	2	3	4	5	6
A	151	151	152	152	153	153
B	154	154	155	155	156	156
C	157	157	158	158	159	159
D	160	160	161	161	P	P
E						
F						

(c) Communities 151 to 161.

Figure G.4: Plate lay-out for synthetic communities 116 to 161.

	1	2	3	4	5	6
A	B	B	162	162	163	163
B	164	164	165	165	166	166
C	167	167	168	168	169	169
D	170	170	171	171	172	172
E	173	173	174	174	175	175
F	176	176	177	177	178	178

(a) Communities 162 to 178.

	1	2	3	4	5	6
A	179	179	180	180	181	181
B	182	182	183	183	184	184
C	185	185	186	186	187	187
D	188	188	189	189	190	190
E	191	191	192	192	193	193
F	194	194	P	P		

(b) Communities 179 to 194.

Figure G.5: Plate lay-out for synthetic communities 162 to 194.