Antibiotic resistance in the environment: a contribution from metagenomic studies

Johan Bengtsson-Palme

Department of Infectious Diseases
Institute of Biomedicine
Sahlgrenska Academy at the University of Gothenburg

UNIVERSITY OF GOTHENBURG

Gothenburg 2016

Cover illustration: Johan Bengtsson-Palme
Photo of the author: Annette Palme

Antibiotic resistance in the environment:
a contribution from metagenomic studies

johan.bengtsson-palme@microbiology.se
http://microbiology.se

ISBN 978-91-628-9758-1 (print)
978-91-628-9759-8 (PDF)
http://hdl.handle.net/2077/41843

Printed in Gothenburg, Sweden 2016
Ineko AB

Till min familj,
och de som kommer efter mig

Abstract

Antibiotic resistance accounts for hundreds of thousands of deaths annually, and its projected increase has made the WHO recognize it as a major global health threat. In the last decade, evidence has mounted suggesting that the environment plays an important role in the progression of resistance. The external environment acts as a source of resistance genes for human pathogens, but is also an important dissemination route allowing the spread of resistant bacteria between different environments and human populations. In this thesis, large-scale DNA sequencing techniques are used to gain a better understanding of the risks associated with environmental antibiotic resistance. A key task in this process is the quantification of the number of antibiotic resistance genes in different environments using metagenomics. However, equally important is to put this information into a larger perspective, by including, for example, taxonomic data, concentrations of antibiotics present, and the genomic contexts of identified resistance genes. This thesis presents a software tool – Metaxa2 – for improved taxonomic analysis of shotgun metagenomic data, which is shown to give more accurate taxonomic classifications of short read data than other tools (Paper I). It also provides theoretically predicted no-effect concentrations for 111 antibiotics (Paper II), and experimentally determined minimal selective concentrations for tetracycline (Paper III). Furthermore, resistance genes are quantified in two environments suggested to pose selective conditions for resistance: sewage treatment plants (Paper IV) and a lake exposed to waste from pharmaceutical production (Paper V). There was no clear evidence for selection of antibiotic resistance genes in sewage treatment plants, however other factors such as oxygen availability seem to have much stronger effects on these microbial communities, which may mask small selective effects of antibiotics and other co-selective agents. In contrast, in the lake subjected to industrial pharmaceutical pollution, resistance genes and mobile genetic elements were both diverse and abundant. Finally, Paper VI shows that travel contributes to the spread of resistance genes against several different classes of antibiotics between countries with higher resistance rates and Sweden. In Paper IV–VI, the genetic contexts of resistance genes were assessed through metagenomic assembly, showing how different resistance genes are linked to each other in different environments. Through these means, the thesis contributes knowledge about risk settings for development and transmission of antibiotic resistance genes, which can be used to guide risk assessment and management schemes to delay or reduce clinical resistance development.

Populärvetenskaplig sammanfattning

Antibiotika är fantastiska läkemedel som har gjort det möjligt att enkelt bota sjukdomar som tidigare ofta ledde till döden. Sedan Alexander Fleming upptäckte penicillinet har miljontals människoliv räddats med hjälp av vad som tidigare ofta kallades ”mirakelmediciner”. Vi använder dock inte bara antibiotika för att bota sjukdomar, utan även kirurgiska ingrepp, cancervård och vård av för tidigt födda barn är ofta direkt beroende av fungerande antibiotika. Det är med andra ord svårt att tänka sig hur den moderna sjukvården skulle vara utformad utan effektiva antibiotika.

Tyvärr har de senaste 25 åren inneburit att allt fler bakterier överlever behandling med antibiotika, så kallad *antibiotikaresistens*. Särskilt oroande är att många bakterier idag kan motstå flera olika typer av antibiotika och att denna utveckling verkar gå allt snabbare. Resistensutvecklingen är 2000-talets hälsokris och antibiotikaresistenta bakterier beräknas orsaka hundratusentals dödsfall varje år. WHO har kallat situationen en av de största utmaningarna för hela den moderna sjukvården. En stor del av denna utveckling beror på att bakterier har förvärvat nya gener som ger upphov till resistens. De kan göra detta eftersom många bakterier har förmåga att utbyta gener med varandra, särskilt under stress. Många, kanske till och med de flesta, av dessa nya gener har sitt ursprung i ofarliga bakterier som lever i miljön. Till exempel så har jordbakterier och bakterier som orsakar sjukdomar hos människor i vissa fall exakt samma resistensgener – trots att deras övriga gener uppvisar mycket begränsade likheter. Man har också hittat resistensgener i jordprover från 30 000 år gammal permafrost, tillsammans med DNA från mammutar. Detta talar för att miljön har en viktig roll i både spridning och utveckling av antibiotikaresistens, samt att resistensgener från miljön i värsta fall kan dyka upp i sjukdomsbakterier som då inte längre går att behandla. Vi vet dock fortfarande väldigt lite om exakt hur dessa processer går till och vilka miljöer som utgör särskilt stora risker för att resistens ska spridas till sjukdomsbakterier. Vi vet inte heller om de halter av antibiotika som påträffas i till exempel kommunala reningsverk kan driva på utvecklingen av resistenta bakterier, eftersom de halter av antibiotika som krävs för att ge resistenta bakterier en konkurrensfördel inte är kända.

Flera av studierna i den här avhandlingen använder storskalig sekvensering av DNA från bakteriesamhällen i olika miljöer, så kallad *metagenomik*, för att bättre förstå riskerna med antibiotikaresistens i miljön. För att bättre sätta resultaten i ett sammanhang har vi också undersökt vilka bakteriearter som finns i de olika miljöerna, samt vilka koncentrationer av antibiotika som kan förväntas ge resistenta bakterier i miljön en konkurrensfördel. För att kunna göra detta har vi inom ramen för avhandlingen behövt utveckla nya verktyg och referensverk och avhandlingens första del handlar om dessa.

I det första delarbetet presenteras ett nytt datorprogram – Metaxa2 – för att analysera vilka arter som finns i ett mikrobiellt samhälle baserat på sekvensering av blandat DNA från alla arter i ett prov. Vi visar att Metaxa2 överlag är bättre än

andra liknande datorprogram på att korrekt klassificera arter baserat på den typ av data som ofta generas i metagenomik-studier.

I det andra delarbetet beräknar vi teoretiskt vilka halter av antibiotika som riskerar driva på utveckling av resistens mot antibiotika i komplexa bakteriesamhällen (minsta selektiva koncentrationer). Vi antar här att de halter som driver utveckling av resistens alltid är lika stora eller mindre än de halter som dödar bakterier eller hindrar deras tillväxt. Genom att utgå från kliniskt tillgänglig information om hur känsliga ett mycket stort antal olika bakteriestammar är för olika antibiotika har vi sedan uppskattat gränsvärden för 111 olika antibiotika. Dessa gränsvärden bör inte överskridas om man vill undvika utveckling av antibiotikaresistens. För antibiotikumet tetracyklin uppskattades denna minsta selektiva koncentration till 1 µg/L. I det tredje delarbetet följer vi upp denna studie för just tetracyklin genom en mängd experiment där bakterier får tillväxa i akvarier med olika höga halter av tetracyklin. Slutsatsen av denna studie är att just 1 µg/L verkade vara en rimlig uppskattning av den minsta halt som kan driva på resistensutveckling.

Med denna kunskap undersöker vi sedan två andra miljöer som potentiellt kan bidra till ökad resistensutveckling: svenska avloppsreningsverk och en indisk sjö förorenad med avfall från produktion av läkemedel, bland annat antibiotika. I svenska reningsverk hittade vi koncentrationer av ett par antibiotika som eventuellt kan vara tillräckligt höga för att bidra till resistensutveckling. Vi kunde dock inte se några tydliga tecken på att en sådan utveckling faktiskt ägt rum i reningsverken, och inte heller några tydliga bevis för att andra substanser som antibakteriella biocider och metaller skulle orsaka utveckling av antibiotika-resistens i denna miljö. Det var dock tydligt att andra faktorer, som t.ex. syretillgång, påverkar bakterierna mycket mer än vad halterna av antibiotika, metaller och biocider gör. Därför kan det finnas effekter som vi inte kan upptäcka med metagenomik, eftersom metoden är alltför grovkornig. I den indiska sjön, som undersöktes i det femte delarbetet, såg vi däremot tydliga effekter på förekomsten av resistensgener, samt på de gener som bidrar till att flytta resistensgener mellan olika bakterier. Detta pekar på att utsläpp av antibiotika från antibiotikaproduktion kan vara en viktig drivkraft i de processer som orsakar resistensutveckling i miljön.

Slutligen har vi undersökt hur resande påverkar hur resistensgener och resistenta bakterier sprids över jorden tillsammans med de bakterier som normalt lever i tarmen. Vi studerade här avföringsprov från 35 svenska studenter som rest till Indien eller Centralafrika och fann att resistensgener var vanligare i tarmen efter resan än de var före. Dessutom ökade förekomsten av de gener som bidrar till att flytta DNA mellan bakterier. Detta tyder på att det räcker att vistas i en miljö med en värre resistenssituation än i Sverige för att samla på sig resistenta bakterier. Eftersom dessa resistenta bakterier kan spridas utan att vi själva blir sjuka och märker av dem, kan de snabbt förflytta sig mellan olika världsdelar och resande utgör därmed en viktig spridningsväg för resistens över jorden.

Mycket tyder på att det oftast är en väldigt liten nackdel för bakterier att bära på gener som ger resistens mot antibiotika, och så fort det finns antibiotika närvarande utgör dessa gener en stor fördel. Det är därför viktigt att undvika att skapa miljöer där bakterier kan utveckla och sprida resistensgener. Det är också centralt att försöka stänga de vägar som resistenta bakterier från miljön kan ta för att hamna i människor och byta gener med människans tarmbakterier. Antalet olika resistensgener som påträffas i miljön är stort och det är därför troligt att det fortfarande finns massor av okända resistensgener i miljön som kan hamna i sjukdomsbakterier. Att identifiera dessa och så långt som möjligt försöka förhindra att de sprids är en enorm utmaning, men extremt angeläget för att fördröja utvecklingen av resistenta sjukdomsbakterier. Metagenomik utgör bara en liten pusselbit i denna process, men kan ändå bidra med viktig information för att t.ex. identifiera vilka miljöer som utgör särskilda risksituationer. Den här avhandlingen bidrar till denna kunskapsbas genom att utveckla verktyg för analys av resistensgener och deras sammanhang i metagenom, genom att undersöka tre särskilt viktiga miljöer där resistensgener eventuellt kan utvecklas och spridas, samt genom att föreslå gränsvärden för utsläpp av antibiotika till miljön.

List of Papers

I. **Metaxa2: improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data**
Bengtsson-Palme J, Hartmann M, Eriksson KM, Pal C, Thorell K, Larsson DGJ, Nilsson RH
Molecular Ecology Resources 15, 6, 1403–1414 (2015)

II. **Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation**
Bengtsson-Palme J, Larsson DGJ
Environment International 86, 140–149 (2016)

III. **Minimal selective concentrations of tetracycline in complex aquatic bacterial biofilms**
Lundström SV, Östman M, Bengtsson-Palme J, Rutgersson C, Thoudal M, Sircar T, Blanck H, Eriksson KM, Tysklind M, Flach C-F, Larsson DGJ
Science of the Total Environment 553, 587–595 (2016)

IV. **Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics**
Bengtsson-Palme J, Hammarén R, Pal C, Östman M, Björlenius B, Flach C-F, Fick J, Kristiansson E, Tysklind M, Larsson DGJ
Manuscript

V. **Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India**
Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E, Larsson DGJ
Frontiers in Microbiology 5, 648 (2014)

VI. **The human gut microbiome as a transporter of antibiotic resistance genes between continents**
Bengtsson-Palme J, Angelin M, Huss M, Kjellqvist S, Kristiansson E, Palmgren H, Larsson DGJ, Johansson A
Antimicrobial Agents and Chemotherapy 59, 10, 6551–6560 (2015)

VII. **Antibiotic resistance genes in the environment: prioritizing risks**
Bengtsson-Palme J, Larsson DGJ
Nature Reviews Microbiology 13, 369 (2015)

See **http://microbiology.se/publications/**
for additional publications by the author.

Contents

Introduction

The importance of antibiotics

In the twentieth century, the ability to treat bacterial infections was revolutionized by a novel category of drugs – antibiotics, defined as any small molecule that antagonizes growth of microbes (Clardy *et al.* 2009). This age of "wonder drugs" begun with the introduction of sulfonamide in 1910, although at the time its mechanism of action was still unclear (Zaffiri *et al.* 2012). However, the real transformation of healthcare triggered by antibiotics came with Alexander Fleming's discovery of penicillin (Fleming 1929), and its introduction as a human antibiotic in 1941 (Chain *et al.* 1940), with mass production a few years later (Zaffiri *et al.* 2012). Since then, around two scores of antibiotics classes with different modes of actions have been introduced to the market, along with a large variety of derivatives within each class (Coates *et al.* 2011). The vast majority of classes were introduced in the 1950-ies and 60-ies, and for a long time antibiotics made diseases such as tuberculosis, pneumonia, gonorrhea, and puerperal sepsis easily treatable. However, virtually no novel classes of antibiotics have become available for treatment in the last fifteen years (Bush 2012), indicating a stagnation of the development of new therapeutic options. Today, antibiotics are widely used to treat bacterial infections, but are also integral as treatment and prophylaxis in surgery, as well as for cancer, neonatal and elderly care. Furthermore, antibiotics are widely used in agriculture for livestock, although to varying degree in different regions of the world (Hollis & Ahmed 2013; Hellman *et al.* 2014). It is difficult to imagine how modern healthcare would function without antibiotics, and along with hygiene and vaccines, antibiotics clearly represent one of the most important steps forward in the treatment of infectious diseases.

Emergence of antibiotic resistance

Already in the 1940-ies when penicillin was first used as an antibiotic, enzymes that could render bacteria resistant against it were described (Abraham & Chain 1940). This discovery foreshadowed a development we have since seen for every new class of antibiotics introduced, regardless of whether it has been derived from natural products, or has been a completely novel, chemically synthesized compound – only the time between introduction and resistance emergence has varied (Schmieder & Edwards 2012). The prevalence of antibiotic resistance among clinically relevant bacteria has steadily increased with antibiotics usage (Pendleton *et al.* 2013; Wattal & Goel 2014). In addition, pathogens are increasingly resistant to several different antibiotics – so called multidrug resistance – further complicating treatment strategies (Alekshun & Levy 2007; Nikaido 2009; Oliveira *et al.* 2015). Perhaps most alarming is the dramatic increase of resistance towards what is viewed as last-resort antibiotics: carbapenems, vancomycin, and piperacillin/tazobactam combinations (Laxminarayan 2014; European Centre for Disease Prevention and Control 2013).

The rapid surge of resistance among pathogens has been fueled by the ability of bacteria to share genes with each other through a process called horizontal gene transfer (HGT) (Thomas & Nielsen 2005). Resistance genes against antibiotics can through these gene transfer processes move between bacterial cells and species on mobile genetic elements (MGEs) such as plasmids and integrons, which can be shared as needed (Stokes & Gillings 2011). This also allows for resistance genes against several different compounds to be collected on the same MGE and move together, giving rise to transferrable multidrug resistance. Furthermore, once these genes are situated together on the same plasmid, treatment with one antibiotic will select for resistance against not only the antibiotic used, but also all other compounds for which resistance genes are present on the same MGE, so called co-resistance. Once resistance has emerged on an MGE, spread among pathogens can be quick, as shown in the case of the NDM-1 carbapenemase. The NDM-1 gene codes for an enzyme that can catalyze cleavage of most forms of beta-lactam antibiotics including carbapenems, and first appeared in a Swedish patient hospitalized in India in 2007 (Yong *et al.* 2009). The gene has subsequently been found to be widespread in the Indian environment (Walsh *et al.* 2011), and is nowadays – less than ten years later – detected in clinical isolates worldwide (Wilson & Chen 2012; Johnson & Woodford 2013). Developments like this have urged the WHO to consider antibiotic resistance as a global challenge so serious that it threatens the fundamental achievement of modern medicine (WHO 2014). Antibiotic resistance has been attributed to annual costs of at least 1.5 billion euros in Europe alone (Norrby *et al.* 2009) and has been estimated to account for 700,000 deaths every year (Review on Antimicrobial Resistance 2014). The problem is set to get worse over time, as bacteria seem to be more resistant rather than less and antibiotics usage is not in decline (Laxminarayan 2014). Recently, the last class of antibiotics where resistance was limited to individual bacterial strains – polymyxins – was faced with a resistance gene able to spread between bacteria through horizontal gene transfer (Liu *et al.* 2016). This means that for each class of antibiotics in use, corresponding resistance now exists on MGEs. Judging from the lessons learned from NDM-1, the *mcr-1* gene providing resistance to polymyxin – such as colistin – may be posed for similar development, perhaps signifying the start of a post-antibiotic era (Kåhrström 2013; WHO 2014).

The role of the environment

It is clear that human use of antibiotics, including overuse and misuse, is a large driver behind the global resistance development. However, evidence is mounting that resistance genes we see in pathogens today did not initially appear in the clinical setting, but have their origins in the environment (Martinez 2008; Wellington *et al.* 2013). The external environment hosts a large diversity of resistance genes, many of which have never been seen in human-associated bacteria (Allen *et al.* 2009; Lang *et al.* 2010; Martiny *et al.* 2011; Munck *et al.* 2015). This should not come as a surprise, since many of the compounds we use as antibiotics are derived from environmental microorganisms. Thus, antibiotics have been part of microbial ecosystems for much

longer than they have been in clinical use, and many resistance genes may have evolved as countermeasures against antibiotics or as protection mechanisms to withstand the antibiotic by the producers themselves. Along the same lines, resistance genes are essentially omnipresent, having been detected even in pristine environments such as glaciers (Segawa *et al.* 2012). Furthermore, resistance genes similar to those found in human pathogens today have been discovered in 30,000 years old permafrost samples (D'Costa *et al.* 2011), and soil bacteria harbors resistance genes identical to those found in pathogens – including their flanking regions (Forsberg *et al.* 2012). Taken together, it seems most likely that the environment constitutes a source of novel resistance determinants to human-associated bacteria (Wright 2010; Finley *et al.* 2013).

The environment plays an important role in at least two parts of resistance development: as a source of resistance genes to pathogens, and for the dissemination of resistant bacteria, including human pathogens. As described above, the environment can function as a resistance gene pool for pathogens. In this context, it can contribute arenas with sufficient selection pressure to promote recruitment of novel resistance determinants, but the same settings can also aid in rearrangement of existing resistance factors. The latter scenario may effectively create MGEs carrying multiple resistance genes (co-resistance), more efficient resistance gene chimeras, or mobilize genes that were previously bound to chromosomes. Since estimates have pointed to the existence of a staggering thousand billion billion billion (10^{30}) bacterial cells on earth (Kallmeyer *et al.* 2012), such rearrangement events are likely to happen continuously. However, most of these do not occur in settings where a selection pressure for maintaining novel genetic rearrangement exists, and they are consequently not fixated in the bacterial population. The second role of the environment is as a dissemination route for resistant bacteria, such as pathogens travelling between hosts. In this latter context, environments such as sewage treatment plants and agriculture are likely to be important for the spread of resistance (Pruden *et al.* 2006; 2013; Review on Antimicrobial Resistance 2015).

Assessing risks related to environmental antibiotic resistance

To assess the risks associated with environmental antibiotic resistance, the magnitude of the contribution of the environment needs to be quantified (Pruden *et al.* 2013; Ashbolt *et al.* 2013; Berendonk *et al.* 2015). Unfortunately, important information required to perform such a quantification of risks is currently lacking. There are several important knowledge gaps that need to be overcome in order to enable proper risk assessment of environmental antibiotic resistance (Table 1). With regards to the emergence of novel resistance determinants, the understanding of the environments where they appear in contexts that enable transfer to human pathogens is limited. It has been suggested that particular "hot-spot" environments, such as those subjected to pharmaceutical pollution or sewage discharges, as well as aquaculture and agriculture, could be potential environments for resistance emergence (Ashbolt *et al.* 2013; Berendonk *et al.* 2015). However, it remains unclear if these environments are

Table 1. Selected knowledge gaps hindering assessment of risks associated with environmental antibiotic resistance

Open question	Some suggestions
Where do horizontally transferrable resistance determinants emerge?	Polluted environments, sewage treatment plants, aquaculture, agriculture (Ashbolt *et al.* 2013; Berendonk *et al.* 2015)
What concentrations of antibiotics and other toxicants are selective for resistance?	Determination and predictions of minimal selective concentrations for antibiotics (Tello *et al.* 2012; Gullberg *et al.* 2011; 2014; Paper II)
Which environments have the potential to drive resistance selection in bacterial communities?	Likely: humans and animals given antibiotics, industrially polluted environments, aquaculture Possible: sewage, sewage treatment plants, waste disposal (Ashbolt *et al.* 2013; Larsson 2014a)
What roles do mobile genetic elements play in resistance development?	Transfer of resistance between bacteria, mobilization of chromosomal resistance genes, rearrangement of existing resistance determinants (Stokes & Gillings 2011)
What concentrations of antibiotics and other toxicants induce horizontal gene transfer?	Sub-inhibitory concentrations of antibiotics (Beaber *et al.* 2004; Prudhomme *et al.* 2006), few minimal concentrations determined (Jutkina *et al.* 2016)
What are the dissemination routes for resistance genes to human pathogens?	Water bodies (Lupo *et al.* 2012; Pruden 2014), agriculture and food trade (Rolain 2013; European Food Safety Authority & European Centre for Disease Prevention and Control 2013)
Which dissemination routes from selective environments connect to environments with human pathogens?	Water bodies and agriculture have large potential
How can risks associated with known and novel resistance genes be weighed against each other?	Viewpoints vary (Martinez *et al.* 2015; Berendonk *et al.* 2015; Paper VII)

actually where such novel resistance factors emerge, or if they are barely selected for in these settings. For a resistance determinant to be fixated in a bacterial population rather than being lost due to other competitive factors, a selection pressure favoring maintenance of the resistance gene is likely to be the most important factor. However, knowledge of selective concentrations of antibiotics is lacking, particularly in complex communities (Ågerstrand *et al.* 2015), although these concentrations are likely to be below the concentrations completely inhibiting growth (Gullberg *et al.* 2011; Andersson & Hughes 2012). Furthermore, other agents than antibiotics, such as metals and antibacterial biocides, may indirectly contribute to selection for resistance determinants via co-selection (Baker-Austin *et al.* 2006; Wales & Davies 2015) but at what concentrations and in which settings is not known. This makes it complicated to address which environments that actually have selective potential. That said, in some instances selection pressures are are evident, since concentrations of antibiotics well-above the minimal inhibitory concentrations (MICs) for many bacterial pathogens

have been measured, e.g. in sediments polluted by discharges from pharmaceutical manufacturing (Larsson *et al.* 2007; Fick *et al.* 2009).

Mobilization of novel resistance determinants is aided by the induction of horizontal gene transfer. Exactly which roles different MGEs play in the emergence of resistance is not clear. Likely, integrons and transposons greatly contribute to the mobilization of chromosomal genes to plasmids that can spread through bacterial communities (Poirel *et al.* 2009; van Hoek *et al.* 2011; Stokes & Gillings 2011; Il'ina 2012). However, research on when transposases and integrases are induced, when horizontal transfer of plasmids occurs, and the dependence of these processes on the concentrations of antibiotics and other toxicants is still in its infancy (Marcinek *et al.* 1998; Nagel *et al.* 2011). It is known that the transfer of genetic material between bacteria increases upon exposure to sub-inhibitory levels of antibiotics (Prudhomme *et al.* 2006; López & Blázquez 2009; Johnson *et al.* 2015), an effect that has been at least partially attributed to the bacterial SOS response (Beaber *et al.* 2004; Guerin *et al.* 2009), which in turn is dependent on toxicant concentrations (Dörr *et al.* 2009; Torres-Barceló *et al.* 2015). That said, the lowest concentrations that cause these effects remain unknown (Paper II).

Another concern is the contribution of the environment to the dissemination of resistance genes and resistant bacteria (Pruden *et al.* 2013). To some extent, the environments that facilitate dissemination of human-associated resistant bacteria are the same as those enabling spread of non-resistant human pathogens. In this process, sewage, wastewater treatment plants, water bodies and food trade have been identified as important contributing factors (Fernando *et al.* 2010; Rolain 2013; Molton *et al.* 2013; European Food Safety Authority & European Centre for Disease Prevention and Control 2013; Pruden 2014). In addition, human travel is an important vehicle for transporting resistant bacteria around the world (Angelin *et al.* 2015), which means that once resistance emerges in a pathogen at some location, it can quickly gain global spread. These perspectives are important for limiting the spread of human-associated bacteria that have already acquired resistance. However, it is much less clear how harmless environmental bacteria carrying resistance genes disperse, and in which settings they have the possibility to interact with human-associated bacteria under conditions that would favor transfer of antibiotic resistance determinants. The dissemination routes that connect hot-spot environments for emergence and maintenance of resistance genes to humans and/or animals constitute propagation routes for resistance into the human population, and needs to be delineated. Rapid progress in DNA sequencing technology has opened up the possibility to study environmental antibiotic resistance on a large-scale using shotgun metagenomics (e.g. Kristiansson *et al.* 2011). However, the development of methods for metagenomic analysis is still in its early stages, and important tools for e.g. accurate taxonomic analysis are partially missing. Taken together, these obstacles makes it difficult to assess risks, and also to weight the risks associated with presence of known versus novel resistance factors in a given microbiome (Martinez *et al.* 2015; Paper VII).

The aims of this thesis

The overarching ambition of this thesis is to contribute knowledge towards the understanding of how the environment is involved in the emergence and transfer of antibiotic resistance genes. Specifically, the aims of this thesis are:

- To address the need for software that can reliably detect and extract rRNA fragments from shotgun metagenomic data, and accurately classify them to at least the genus level (Paper I)
- To broadly estimate theoretical minimal selective concentrations of antibiotics in complex microbial communities, providing guidance to regulatory efforts to prevent environmental resistance selection (Paper II)
- To experimentally determine the minimal selective concentration of tetracycline in complex microbial communities, using both genotypic and phenotypic endpoints (Paper III)
- To investigate if antibiotics exert a direct selection pressure for resistant bacteria in Swedish sewage treatment plants (Paper IV)
- To determine if antibiotics, biocides and/or metals could co-select for antibiotic resistance in sewage treatment plants (Paper IV)
- To understand how high concentrations of antibiotics resulting from pollution with pharmaceutical waste shape the resistome of environmental microbial communities (Paper V)
- To assess the context and potential mobility of resistance genes in polluted environments (Paper V)
- To investigate the extent to which resistance genes are carried within the gut microbiome of visitors to geographical regions with higher prevalence of resistant bacteria at their return to Sweden (Paper VI)

Through these specific investigations, the thesis contributes knowledge towards the identification of environments that have potential to present selective conditions for antibiotic resistance to bacterial communities. The thesis also aims to shed light on the role of horizontal gene transfer in environmental resistance development, and seeks to verify suggested dissemination routes for resistance genes. Finally, the ultimate objective of the thesis is to synthesize this knowledge to enable better risk assessment of environmental antibiotic resistance (Paper VII).

Metagenomics in antibiotic resistance research

Studying the environmental resistome

Resistance patterns among bacteria have traditionally been studied using culturing on media selecting for resistant colonies. This method has the advantage of showing phenotypic resistance directly and allows connection of physiological features to genetic information using PCR or genome sequencing. It also provides for isolation of resistance plasmids, which can give unambiguous insights into co-resistance patterns and the degree of transferability of resistance genes (see e.g. Flach *et al.* 2015). The isolate culturing approach works well for the study of many resistant pathogens, which can relatively easily be cultivated under laboratory conditions. However, the vast majority of microorganisms in nature cannot be cultivated, at least not by standard methods (Amann *et al.* 1995). This limits the possible scope of this method and thereby veils much of the diversity of species and resistance factors, particularly in environmental communities. For this reason, culture-independent methods to study resistance genes in environmental samples have been developed. A commonly applied approach to quantify resistance gene abundances is quantitative real-time PCR (qPCR; Heid *et al.* 1996). In this method, the abundance of an investigated resistance gene is quantified relative to, e.g., the abundance of 16S rRNA genes or the total volume of the sample. Quantitative PCR has in this way been used to study resistance gene abundances in, for example, soil (Knapp *et al.* 2011), aquaculture (Tamminen *et al.* 2011; Muziasari *et al.* 2014), sewage treatment plants (Gao *et al.* 2012; Laht *et al.* 2014), and areas polluted by pharmaceutical pollution (Rutgersson *et al.* 2014). Furthermore, large-scale qPCR arrays allowing the study of hundreds of resistance gene variants in parallel have been developed and applied to study the resistomes of swine farms (Zhu *et al.* 2013). However, even in the latter case, qPCR is restricted to a fixed number of resistance genes, for which sequences must be known to enable the construction of PCR primers. Thus, while qPCR is highly sensitive and can detect resistance genes at very low abundances, it remains a somewhat limited and largely non-explorative approach.

To facilitate the study of previously undescribed genes and proteins in uncultivable organisms, metagenomics was developed (Handelsman *et al.* 1998). The term "metagenome" refers to the collection of genomes from all organisms in a given environment (or sample), and initially their genetic content was studied by fragmenting the total DNA from an entire community into shorter pieces, which were then inserted into cultivable bacteria. The recipient strains were grown on plates selective for the function of interest. For the study of antibiotic resistance, selective plates containing antibiotics were used. Recipient strains surviving on these plates had their inserted sequences from the metagenome sequenced and further characterized. Using this technique, which subsequently has been named *functional metagenomics*, novel resistance determinants have been uncovered from soil (Allen *et al.* 2009; Lang *et al.* 2010; Torres-Cortés *et al.* 2011; Udikovic-Kolic *et al.* 2014), permafrost (Perron *et al.*

2015b), sea water (Hatosy & Martiny 2015), cow manure (Wichmann *et al.* 2014), birds (Martiny *et al.* 2011), sewage sludge (Munck *et al.* 2015), and the human gut (Sommer *et al.* 2009). Functional metagenomics has taught us that that there is an enormous diversity of resistance genes that we have not yet encountered in human pathogens, even in the human gut (Sommer *et al.* 2010; Moore *et al.* 2011). Still, there are important limitations of functional metagenomics that calls for the use of alternative approaches. First, it is highly time-consuming to perform the experiments needed for a single screen at a sufficiently large scale. Second, since resistance genes are not that common in most environments, very large numbers of DNA recipients often need to be screened to detect a single resistance gene to a given antibiotic. Third, for a resistance gene to actually confer phenotypic resistance, the entire gene (or at least most of it) must be captured inside the DNA fragment inserted, as it will otherwise not remain functional. Furthermore, the gene must also be compatible with the cultivable host, both in terms of functionality and gene expression. Finally, even though the number of resistant recipients can be counted and compared between samples, this does only provides a rough measure of the resistance gene abundances in the studied communities, making functional metagenomics less suitable for quantitative resistance gene screening.

The drawbacks of functional metagenomics suggest that a more convenient method to study the metagenomes of different communities is needed. Luckily, an alternative methodology exists, enabled by rapidly declining costs of DNA sequencing throughout the last decade (Metzker 2010; Hayden 2013; Heather & Chain 2016). In this approach, the total metagenomic DNA of a community is randomly fragmented and sequenced by high-throughput DNA sequencing, often referred to as *shotgun metagenomics* (Wooley *et al.* 2010). The resulting DNA fragments can be analyzed using similarity searches to sequence databases, or assembled into longer stretches of DNA, allowing for the reconstruction of complete genes from the relatively short read fragments. However, although shotgun metagenomics is less limited to particular predetermined target genes than qPCR, it still essentially requires that the obtained genes, or close variants of them, are present in a reference database to enable assignment of them to a (predicted) resistance phenotype. That said, since sequence data can be stored and re-used later, shotgun metagenomics allows for retrospective analysis of resistance genes identified after the initial study has been completed (see e.g. Forslund *et al.* 2013). Furthermore, using homology-based methods novel resistance genes can be unraveled which may then be confirmed in the laboratory, as has been done for the *qnr* fluoroquinolone resistance genes (Boulund *et al.* 2012; Flach *et al.* 2013). Shotgun metagenomics has been applied to quantify the abundances of many resistance genes in parallel, for example in environments subjected to pharmaceutical pollution (Kristiansson *et al.* 2011), sewage treatment plants (Yang *et al.* 2013; 2014), sea water (Port *et al.* 2012), tap water (Shi *et al.* 2013), and the human gut (Forslund *et al.* 2013; Hu *et al.* 2013). However, in terms of measuring specific gene abundances, metagenomics is less sensible than qPCR, particularly when only a couple of million reads are generated per sample. In this

respect, Illumina sequencing was a major step forward compared to pyrosequencing, simply due to the lower costs associated with each read. Limited sequencing depth affects the sensitivity to estimate both the abundances and diversity of resistance genes in the sample, which will be discussed in a later section.

One major advantage of shotgun metagenomics compared to qPCR and functional metagenomics is the ability to detect changes in taxonomic composition and other functional genes, for example those involved in horizontal gene transfer. This can provide clues about whether the resistance genes detected have potential to move between bacterial cells or not. Furthermore, through metagenomic assembly it is sometimes possible to uncover co-resistance patterns, or even completely novel resistance plasmids (Kristiansson *et al.* 2011). In this thesis, the main method for studying the resistance patterns of microbial communities has been shotgun metagenomics (Papers III–VI). In addition, culturing approaches and/or qPCR have been applied to complement the metagenomic data in Papers III and IV, and culturing followed by whole-genome sequencing was used in Paper VI.

Obtaining sequence data from microbial communities

As a first step of any metagenomics analysis, DNA must be extracted from the community. This is usually done using standard DNA extraction kits. However, as environmental samples comprise a large diversity of different bacteria and also may contain contaminants of different kinds, this process is not always straightforward. In addition, although sequencing protocols nowadays require less than a µg of DNA, amplification of the DNA may be needed to obtain sufficient quantity or concentration. It is important to understand that the extraction protocols and amplification strategies (if used) can bias gene frequencies, as not all bacterial species are affected equally by the reagents used. Bias has been shown to result from differences between DNA extraction kits (Knauth *et al.* 2013; McCarthy *et al.* 2015), storage of samples (Choo *et al.* 2015; McCarthy *et al.* 2015), DNA amplification kits (Pinard *et al.* 2006), as well as due to biological variation of, for example, GC-content (Dohm *et al.* 2008). All these factors contribute noise to the samples already before the sequencing is taking place. However, different sequencing techniques also produce different results, partially because of differences in sequenced length for each fragment, but also due to the different methodologies used to determine the nucleotides (Glenn 2011). In this thesis, Illumina sequencing has been employed exclusively, so in this respect samples should be comparable. However, since different extraction kits have been used (and in the case of Paper V also DNA amplification), there might be biases between studies and sample types, and thus cross-study comparisons should be interpreted with some caution. Although the exact details have varied somewhat between studies, the sequence data has, before any other analyses have been performed, been filtered with respect to sequencing adapters and low-quality reads. In Paper V, PETKit (Bengtsson-Palme 2012) was used for read trimming and filtering, but in Papers III, IV & VI, this was replaced by a software called Trim Galore! (Babraham Bioinformatics 2012). Trim Galore! is faster, offers a

higher degree of flexibility, and can remove remnants of the Illumina sequence adapters from the data in a single step, and was therefore preferred over PETKit in the later studies. All analyses of sequence data in all studies of this thesis are based on the quality-filtered reads obtained after this filtering step.

Detecting and quantifying resistance genes in metagenomes

Gaining insights into the resistance gene content of a microbial community from sequence data requires the ability to detect resistance genes among sequence fragments derived from a multitude of different genes. This is achieved through similarity searches, employing the principle that genes that share homology often perform similar functions. This principle is at the heart of bioinformatic methods, but depending on the questions asked, its usefulness differs. Often, changes of only a few amino acid residues in a protein can alter substrate preferences (Smooker *et al.* 2000; Johnson *et al.* 2001), binding sites (Glaser *et al.* 2005; Dabrazhynetskaya *et al.* 2009) or the overall functions (Atkinson & Babbitt 2009; Bianchi & Díez-Sampedro 2010) of certain proteins. Therefore, the validity of the assumption that a read matching to a protein in a reference database comes from a gene encoding a protein with the same function is dependent on *how* similar the read is to the reference sequence. This means that the choice of method for assigning function to metagenomic reads depends on which stringency one aims for. In the case of mobilized resistance genes, their sequences show limited variation once they have appeared on MGEs (Pal *et al.* 2014). Chromosomal resistance genes (and other chromosomal genes as well) tend to have a lesser degree of conservation between species, and it is therefore harder to detect non-mobile resistance genes with certainty than mobilized ones. Because of the inherent dependency on sequence similarity, selecting an appropriate sequence identity cutoff for calling a matching read a resistance gene becomes crucial (Martinez *et al.* 2015). At the same time, reads come with a certain degree of sequencing errors, and there might be slight differences between resistance genes that do have the same function. Therefore, one wants to allow to a certain degree of mismatches between the read and the reference sequence – the question is: how large can this difference be if stringency is to be maintained? The answer to that question depends on how similar resistance genes known to carry out the same function are. However, the percent identity of functionally verified resistance genes within the same group varies substantially (Figure 1). The average sequence identity between sequences associated with the same gene name and function differs between 68% and completely identical, and the lowest identity between two sequences with the same gene name can be as low as 52.8% (the *vanSG* vancomycin resistance gene). However, applying a universal cutoff of 50% sequence identity would produce an immense number of false positive hits. Using the CTX-M beta-lactamase as an example, performing a BLAST search (Altschul *et al.* 1997) against the NCBI protein database (NCBI Resource Coordinators 2015) with the CTX-M sequences as queries yields more than 2000 matches at a 50% identity cutoff (requiring 30 matching amino acids, corresponding to the length of a typical Illumina read). Many of these

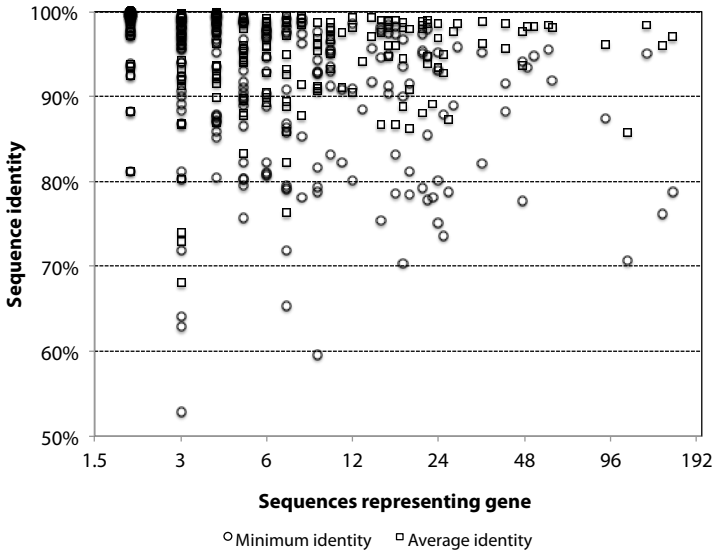


Figure 1. Sequence identity between variants assigned to the same resistance gene group in the Resqu database. Sequences were aligned using MAFFT (Katoh & Toh 2008) and pairwise identities were calculated as the number of identical amino acids in corresponding positions, discarding gaps in one or both of the sequences.

sequences belong to other classes of beta-lactamases, indicating that this cutoff would not be feasible.

Indeed, there is not foolproof approach to make sure that a read comes from a functional resistance gene. Even if 100% identical to a resistance gene, the read only represents a part of the gene sequence, and the gene the read is derived from may, for example, be truncated and thus non-functional. However, as seen in the example with CTX-M, it is important that the cutoffs are not set too low to retain stringency. Thus, requiring sequence identity of 80-95% is probably warranted. Furthermore, the larger the datasets grow, the more computing resources are required to process them. Read mapping allowing for a large number of mismatches is computationally much more expensive than searching for high-identity matches. Thus, the choice of cutoff value becomes a tradeoff between speed, sensitivity and stringency. In this thesis, the Vmatch software (Kurtz 2010) has been used for matching reads to reference databases of resistance genes. Vmatch utilizes suffix trees, which are extremely efficient data structures for matching reads with high identity to reference data. Generally, a cutoff of two amino acid mismatches per read has been used, corresponding to a percent identity of 90-94%, depending on the read length. To avoid missing known mobile resistance genes, the database therefore includes all confirmed variants of each gene, meaning that a read matching to any of these variants has been counted as a resistance gene fragment.

To quantify resistance gene abundances, the reads mapped to resistance gene variants have been summed for each resistance gene type (i.e. individual gene names).

This yields a raw number of reads associated with each resistance gene type in every sample. To avoid overestimating the abundance of long genes (which will recruit more matching reads simply because there are more amino acids to align to), each count has been divided by the length of the reference gene. Furthermore, since samples are sequenced to varying depth, i.e. the total number of reads generated differ, and may not contain similar proportions of bacteria, the length-normalized counts have been further divided by the number of bacterial 16S rRNA sequences in each sample, and finally divided by the length of the 16S rRNA gene. The end product is a number that represents the number of reads matching to a resistance gene per bacterial 16S rRNA. These numbers are more comparable between samples, and can also to some degree be compared to values from qPCR studies normalized in a similar way.

It should be noted that read mapping against a reference database is not suitable for detecting novel resistance genes, for reasons outlined earlier. To successfully predict novel resistance determinants not yet present in the database computationally, prior knowledge of the specific gene type is, in principle, required. Through modeling of conserved motifs, discovery of novel resistance proteins is possible (Boulund *et al.* 2012), but without very specific models for the genes studied the risks of over- and under-prediction are high. Instead, functional screening for novel resistance genes is more likely to arrive at useful results (Allen *et al.* 2009; Sommer *et al.* 2009; Munck *et al.* 2015), since computational predictions nevertheless need to be tested in the laboratory to have their function verified (Flach *et al.* 2013).

Databases for resistance genes

In addition to the methodological aspects regarding gene quantification from metagenomic data, the choice of reference databases also has important implications for the quality of the information derived. Since annotation based on bioinformatic analysis of sequence similarity never will be more accurate than that of the reference sequences, selecting a reference database with high-quality annotation is crucial to arrive at relevant conclusions. Simply put, if the database only contains resistance genes against beta-lactams, you naturally cannot expect results to cover the full range of resistance genes in the sample, and the total resistance gene content in that environment will likely be grossly underestimated. On the other hand, if the database contains genes incorrectly predicted to have resistance functions, the resistance gene abundance of the sample will be overestimated. A number of databases containing antibiotic resistance gene information exist. An often used resource, particularly in the early papers using metagenomics to investigate antibiotic resistance, is the Antibiotic Resistance Genes Database (ARDB), established in 2008 (Liu & Pop 2009). However, a few problems exist with ARDB. Most prominently, its last update was in July 2009, meaning that any resistance gene discovered after that date is not included in the database (this includes e.g. the NDM-1 carbapenemase mentioned earlier). In addition, the ARDB does not make any difference between resistance genes with a confirmed resistance function and those predicted to confer resistance based on

homology. Thus, the database may contain sequences that in fact are not resistance genes. The ARDB has subsequently been structured by resistance types and had some obviously erroneous sequences removed (Yang *et al.* 2013), and this version of the database remains in use (e.g. in Ma *et al.* 2016). However, the basic problems of the database being outdated and that the majority of sequences do not have their functionality demonstrated prevails also in this version. The developers of ARDB instead recommend the use of the Comprehensive Antibiotic Resistance Database (CARD; McArthur *et al.* 2013). This database is still in active curation and is possibly the most comprehensive resource for antibiotic resistance gene information available. However, although CARD is based on thorough curation, it does not clearly separate experimentally verified and predicted entries. Furthermore, it is unclear if the genes in the database have been found on MGEs or only have been detected on chromosomes. That said, the use of a single reference sequence for every resistance gene increases the likelihood that each sequence has been confirmed to confer resistance in at least some species. Similar problems also haunts the ARG-ANNOT database (Gupta *et al.* 2014), although to a much larger extent. The ARG-ANNOT database employs what they refer to as "relaxed search criteria" to identify resistance genes, which in reality means that the database contains a multitude of sequences with poor annotation information, and that many entries are unlikely to be functional resistance genes. The value of ARG-ANNOT for identifying true resistance genes is thus limited. A more stringent approach to this has been taken by the ResFinder (Zankari *et al.* 2012) and Resqu (1928 Diagnostics 2012) databases. Both these databases only contain sequences of acquired antibiotic resistance genes present on MGEs. However, while ResFinder does not pose any experimental validation criteria for entries, Resqu also requires each gene to have been experimentally verified for inclusion in the database. That said, a drawback associated with Resqu is that it has not been updated since 2013, while ResFinder is still actively curated. In this thesis (Papers III-VI), we have used the Resqu database as reference, though in many cases we have also verified results against the ARDB and CARD databases.

In terms of resistance genes against other compounds that may act as co-selectors for antibiotic resistance, such as antibacterial biocides and metals, the available database options are more scarce, particularly for biocides. For metals, scattered efforts to create databases for particular metals exists, for example for arsenic (Cai *et al.* 2013) and copper (Li *et al.* 2014). However, in none of these cases actual verified function was required for inclusion, and sequences were instead included based on their annotation and similarity searches. Furthermore, there have been attempts to define broader sets of detoxification proteins (Bengtsson-Palme *et al.* 2014a), but such approaches are not well suited for annotating short-read metagenomic data. The lack of comprehensive databases for potential co-selective agents spurred our development of the BacMet database of resistance genes against antibacterial biocides and metals (Pal *et al.* 2014). This database contains information on experimentally verified resistance genes, as well as a separate part covering resistance genes predicted by similarity searches. Since this is to date the only comprehensive

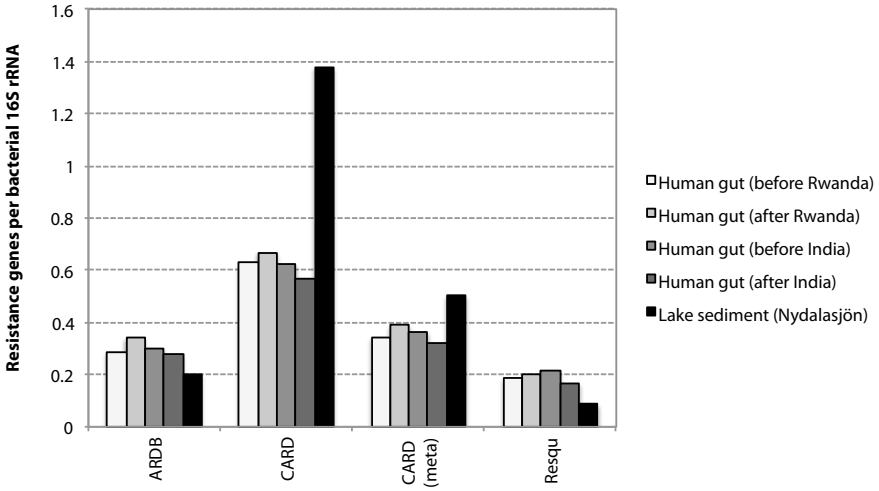


Figure 2. Differences in total resistance abundance reported by the same bioinformatic method using four different reference databases: ARDB, the full CARD database, the metagenomics-adapted version of CARD, and Resqu.

curated resource of biocide and metal resistance genes in bacteria, it has been used for the identification of such genes in this thesis (Paper IV).

How the database content affects results

Depending on the database used, reported resistance gene abundances may differ, despite that the same bioinformatics protocols are applied. For example, ARDB, CARD and Resqu report radically different numbers of resistance genes in the human gut and sediment from a Swedish lake (Figure 2). Resqu consistently reports the lowest numbers, likely since it only contains resistance genes with a verified resistance function that have been shown to be present on mobile genetic elements, and thus excludes many generic efflux pumps that may confer low-level antibiotic resistance. From a risk perspective, the mobile resistance genes are probably the most relevant to detect and quantify. Furthermore, many of the multidrug efflux pumps are relatively well conserved between variants having and not having capacity to export antibiotics (Martinez *et al.* 2015). Using the full CARD database consistently reports resistance gene counts two to three times higher than ARDB, while the version of CARD with target sequences removed reports roughly the same results as ARDB (although not for the lake sediments).

The reason why the full CARD database suggests much higher abundance of resistance genes is that in addition to genes that actually confer resistance thanks to their function, it also include target genes with mutations providing resistance. Genes containing such point mutations indeed enable their carrier to survive antibiotics treatment, but are not transferrable between bacteria and are – importantly – very

similar to the susceptible variants of the target genes. The latter means that even reads stemming from susceptible ("wild-type") bacteria in a metagenome would map to these "resistance genes", particularly if, e.g., a 90% identity threshold is used. Diluting the database with such genes means that the total resistance gene content will undoubtedly be overestimated, as many of these target genes are ubiquitously occurring essential genes, highly conserved between bacterial species. For example, the *rpoB* gene (the target gene of rifampicin; mutated variants are present in the full CARD database) is present in a single copy in most bacterial species (Dahllöf *et al.* 2000) and has thus been proposed as a possible per-genome normalization gene for metagenomics (Bengtsson-Palme *et al.* 2014a). The presence of around one such "resistance gene" per 16S rRNA in the Swedish lake sediment, as reported when using the full version of CARD (Figure 2) therefore seems reasonable. However, the vast majority of the reads associated with these "resistance genes" will actually derive from antibiotic-sensitive variants of essential target genes.

It is important to realize that this is not a problem related to the CARD database *per se*. The database website clearly states that target genes are present among its sequences, and since 2015 provides a separate dataset with the target genes removed for use in metagenomic studies[1]. Recently, CARD was also updated to fully separate target sequences and functional resistance genes in different files. Still, if care is not taken in examining the content of the database used, this may lead to partially misleading conclusions, with may explain surprising results of some studies (see e.g. Ma *et al.* 2014).

A similar problem is the use of general annotation pipelines, such as the commonly used MG-RAST (Meyer *et al.* 2008), that are not curated with regards to antibiotic resistance. The use of MG-RAST to annotate resistance genes has led to some peculiar reports suggesting that almost one in 25 genes found in human feces would confer antibiotic resistance (Durso *et al.* 2012). The non-stringent identity cutoffs used by default in MG-RAST are likely to be one major cause of these results. Similar use of low identity thresholds in other studies has also led to unexpectedly high estimates of resistance gene abundances in human feces (Nesme *et al.* 2014). This emphasizes the importance of accounting for other factors that could explain unexpected results in metagenomic studies. Overall, there is a clear need for improved stringency with regards to database usage and parameter choices in metagenomics studies aiming to quantify resistance gene abundances.

The influence of fecal contamination

Another complication in the inference of resistance selection in the environment is that the abundance of resistance genes often is tied to the relative proportion of fecal bacteria (Figure 3). This makes it difficult to infer whether an enrichment of resistance genes in a particular sample is due to selection for the resistance factor, or

[1] This dataset was released as a response to a plenary discussion initiated by the author of this thesis at the EDAR3 conference in May 2015.

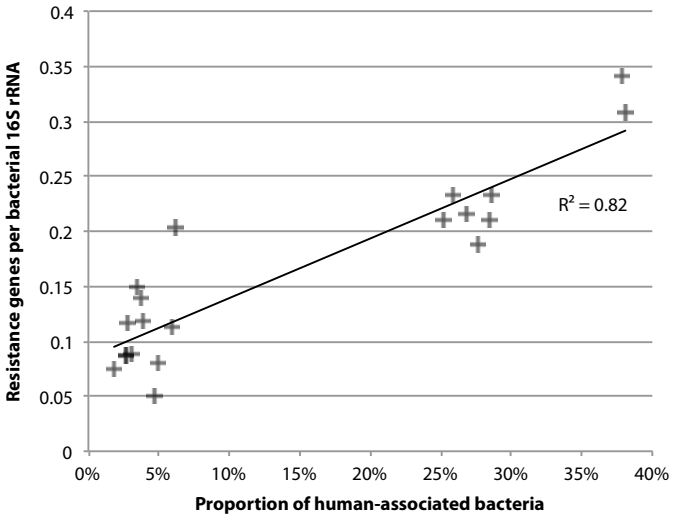


Figure 3. Relationship between the abundances of human-associated bacteria (classified as being present in the Human Microbiome Project genome catalog) and antibiotic resistance genes in the sewage treatment plant samples of Paper IV.

merely the by-product of contamination with human feces. This is also suggested from the sediments investigated in this thesis, where those sampled downstream a Swedish sewage treatment plant (STP) had both higher abundance and diversity of resistance genes than those sampled upstream (Figure 4). Apart from environments contaminated with antibiotics, human feces contains the highest abundances of

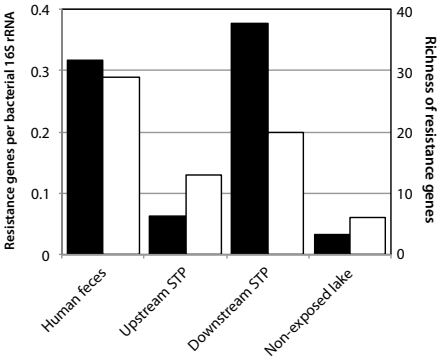


Figure 4. Resistance gene abundances (black bars) and richness (white bars) in sediment samples taken upstream and downstream of a Swedish STP, and human feces from Swedes.

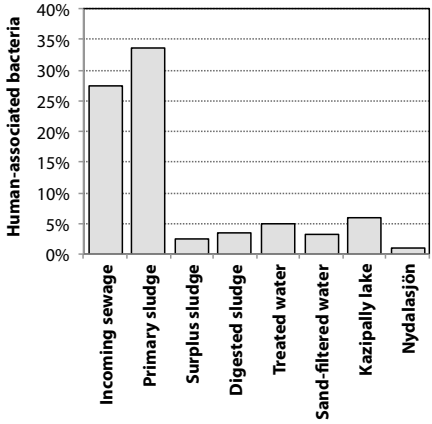


Figure 5. Proportion of human-associated bacteria (classified so by being present in the Human Microbiome Project genome catalog) in STP sample types and the two lakes of Paper V.

known resistance genes investigated in this thesis, and thus the detection of resistance gene enrichments in certain sample types will not tell much about selection unless placed into a taxonomic context, or if the levels detected are substantially above those in human feces, which would also indicate selection for resistance. The latter is the case with, for example, the Indian lake investigated in Paper V, which harbors four times as much fecal bacteria than the Swedish lake (Figure 5), clearly indicating contamination with human feces, but at the same time contains over a thousand times more resistance genes than the Swedish lake, and 80 times more resistance genes than feces from Swedish students (Paper VI).

Because of the relationship between resistance genes and fecal pollution, it becomes important to estimate the proportion of bacteria derived from feces in different environments. There is not any straightforward approach to do this, although several methods have been suggested. Several bacteria have been proposed as marker species for environmental fecal contamination (Roslev & Bukh 2011). The Bacteroidales order could be a suitable target for PCR-based quantification of feces, both specifically from humans (Ashbolt *et al.* 2010; Harwood *et al.* 2014), but also from other animals (Kildare *et al.* 2007). However, it is not certain whether such an assay would be specific enough on short metagenomic read fragments. *Enterococcus* and *Escherichia* have also been suggested as fecal markers (Roslev & Bukh 2011), along with certain enteroviruses (Wong *et al.* 2012). Finally, human mitochondrial DNA (He *et al.* 2015) and even antibiotic resistance gene composition (Whitlock *et al.* 2002) have been used to identify pollution with human feces. Since metagenomics enables detection of a wide diversity of taxa, it has also been proposed to take a larger part of the community composition into account for tracking human feces contamination in the environment (Lee *et al.* 2011). One possibility would thus be to use the bacteria present in the human gut microbiome genome catalog (Human Microbiome Jumpstart Reference Strains Consortium *et al.* 2010) as reference. This approach (used for Figures 3 and 5) will, however, only provide an upper bound for the human-associated bacterial content, as many of the species present in that genome catalog can exist also in the gut microbiome of other species, or in the external environment. Finding appropriate fecal markers remains a hurdle for using metagenomics in environmental resistance gene research, and a perfect solution to the problem may not even be possible.

Unsolved statistical problems for metagenomics

Once gene counts have been established, the next aim is to identify differences in resistance gene abundances between samples. Although this sounds straightforward, a number of technical obstacles remain. The most fundamental problem affecting the statistics of metagenomic data is that the data is high dimensional in the sense that there are generally many more observed genes than biological replicates. Furthermore, the variation between samples in the same group can be fairly large, meaning that even higher numbers of replicates are required to detect statistically significant differences. However, because sequencing is relatively expensive, a tradeoff

exists between sufficient sequencing depth for quantification of genes in each sample and the number of replicate samples sequenced. Finally, since metagenomics generates counts, the resulting data is discrete, and many existing statistical tests assume continuous, normally distributed data. Thus, strategies to turn up the signal and wipe out the noise (Gabriel 2002) are required to gain information from the data. The last few years have seen tremendous development of statistical methods for metagenomic analysis (Jonsson *et al.* 2016), somewhat reminiscent of the early method advances in microarray analysis (Jeffery *et al.* 2006). However, many of those methods provide a descriptive picture of the studied community rather than highlighting statistically significant differences (Dinsdale *et al.* 2013). Interestingly, it took about ten years of microarray usage for statistical methods to "catch up" and become standardized (Bumgarner 2013), and it is reasonable to assume that shotgun metagenomics faces a similar development towards robust standardization within the next few years.

Data transformation approaches

Currently, the statistics for handling metagenomic count data are centered on three fundamentally different approaches: standard tests on transformed counts, tests assuming distributions that account for the features of count data, and non-parametric tests. Data transformations are often used to change the distribution of the data so that it better fits the normal assumptions of standard tests, such as t-tests and ANOVA. For count data, the variance is dependent on the mean, and data transformations can also be used to remove this relationship. Such variance-stabilizing transforms include the square-root transform and various logarithm transforms. Note that logarithm transforms "penalize" very large (and very small) values harder than the square root transform, and thus analysis of logarithm-transformed data is less influenced by the most abundant genes. Transformation methods allow to use standard microarray analysis tools on count data, as implemented in e.g. the Voom package, which estimates and weights the mean-variance relationships of each observation and subsequently analyze the transformed counts using the Limma analysis pipeline (Smyth 2004; Law *et al.* 2014). One problem that becomes apparent when applying a logarithm transform to metagenomic count data is the large number of zeros present. Zeros lead to two problems. The first is practical – zeros cannot be logarithm transformed, and the second is that a zero can either represent that a gene is not present at all, or that it is so rare that the sequencing depth was not sufficient to detect it. The transformation problem can be solved by adding a pseudocount to *all* observations in the dataset. However, the size of the pseudocount will influence effect sizes (and thus statistical significances), particularly when overall gene counts are low, which have led some authors to advise against the use of transformation methods for count data in those cases (O'Hara & Kotze 2010). The latter problem is harder to deal with, and is particularly troublesome when estimating the richness and diversity of taxa or genes, a problem we will return to later. Efforts to handle zero-inflation have been made in,

for example, the metagenomeSeq package, which uses a zero-inflated Gaussian model to correct for undersampling-related bias (Paulson *et al.* 2013).

Non-parametric and count-adapted tests

As an alternative to data transformation, statistical tests that do not make as specific assumptions on the distribution of the data can be used. These are referred to as non-parametric tests (Schlenker 2016), and include e.g. tests based on the ranks of the observation rather than their actual values. These methods are – for better and worse – less sensitive to large variability within the datasets, and are also much more robust to outliers in the data. Other non-parametric tests include permutation tests that resample the data instead of assuming that it follows any particular distribution (Rodriguez-Brito *et al.* 2006; White *et al.* 2009; Bengtsson-Palme *et al.* 2016). Finally, there are also statistical tests designed to better handle count data, usually based on assumptions of Poisson or negative binomial distributed data, such as ShotgunFunctionalizeR (Kristiansson *et al.* 2009), which allows fitting of generalized linear models to metagenomic count data. Such models are also implemented in the RNAseq analysis packages edgeR (Robinson *et al.* 2010) and DESeq (Anders & Huber 2010), which couple the variance and mean either naïvely (edgeR) or by determining the optimal coupling for each individual gene (DESeq). Both these tools are developed for RNAseq data, and although this technique generates similar count data, their assumptions may not be entirely valid for metagenomic analysis. A recent evaluation of different statistical approaches to identify significantly differing genes between metagenomes concludes that the number of replicates, the effect sizes and the gene abundances greatly affect the outcomes of each method, and that no single method is suitable for all metagenomic datasets and questions (Jonsson *et al.* 2016). That said, the methods based on Poisson or negative binomial distributions used for RNAseq overall performed better, particularly with small group sizes, with DESeq and overdispersed Poisson linear models coming out on top. Surprisingly, ordinary square-root transformed t-tests performed relatively robustly also at small group sizes. However, the evaluation also shows that non-transformed methods (standard t-tests, Fisher's exact test and the binomial test) perform poorly and should be avoided. Furthermore, non-parametric methods also perform subpar and should in most cases be replaced by methods based on transformation or appropriate modeling of counts.

Normalization of data to make samples comparable

Another problem related to metagenomics (and many other types of large-scale techniques) is that the sequencing libraries may be of vastly different size, which influences the number of counts from different samples. Furthermore, the composition of samples may be different, and technical factors can bias the sample processing. To make libraries from different samples comparable, normalization is applied. However, depending on the research question, different means of normalization can be appropriate. If one is merely interested in compensating for the different size of the samples, simply dividing each count by the total number of reads

of each library, generating, for example, a count-per-million value, may be sufficient. However, when investigating antibiotic resistance it is often more relevant to determine the counts relative to the bacterial fraction of the sample (trying to exclude contributions from e.g. eukaryotes and viruses). For this purpose, a bacterial marker gene is often used for normalization, most commonly the SSU 16S rRNA, resulting in gene counts per 16S rRNA. However, although the rRNA genes are well studied and often applied for normalization purposes, they can occur in multiple copies within the same genome (Klappenbach *et al.* 2000; Větrovský & Baldrian 2013; Angly *et al.* 2014), and thus other, single-copy, bacterial marker genes have been suggested (Sunagawa *et al.* 2013; Manor & Borenstein 2015), such as the ribosomal protein *rpoB* gene (Dahllöf *et al.* 2000; Bengtsson-Palme *et al.* 2014a). However, since these normalization methods have not yet gained traction, and because of the legacy of qPCR studies, the 16S rRNA remains the most common normalization gene for studies of bacterial communities. One can imagine other relevant normalization strategies, comparing each gene count to, for example, the total plasmid-borne material in the sample, or the total content of resistance genes. Importantly, the choice of normalization method should be based upon the questions asked, and how these questions are best answered. It is also important to consider whether there are variations between samples that will not be compensated for under the normalization method chosen. Such variation may for example be the result of differing 16S rRNA copy numbers, or that not all variants of the marker gene of choice are detected by the methods used (a common problem in metagenomic studies employing short-read sequences, see Paper I). There are also completely different approaches to normalization used in RNAseq, based on minimizing the overall fold-change between experiments, thereby attempting to reduce technical noise (Robinson & Oshlack 2010). Similar thoughts have been carried over into recent metagenomic analysis packages (Sohn *et al.* 2015), although the task of identifying a subset of genes that can be assumed to be stable between samples is not as straightforward in data from communities comprised of a mixture of species.

An additional factor that may also influence gene abundance estimates based on sequences mapped to a reference database, is the length of the reference genes. If this is not compensated for, longer genes may recruit more reads simply by chance. This effect is not relevant to compensate for if one only compares data between samples, but if the abundance levels within each sample are compared, taking gene length into account becomes appropriate. This type of normalization makes sense, but whether or not it is meaningful to compensate for it in real situations is debated (Rapaport *et al.* 2013; Dillies *et al.* 2013). Some authors have suggested that compensating for gene lengths may even be detrimental to differential analyses in RNAseq data (Oshlack & Wakefield 2009), although if the same argument is valid also for metagenomic data is unclear.

Correction for multiple testing

Regardless of which method that is used to determine which genes that are significantly enriched in a group of samples, one p-value will be obtained for each gene tested. This means that with a large reference database, hundreds or thousands of tests will be performed. Since the p-value represents the likelihood of obtaining a certain result by chance, given certain model assumptions (Pearson 1900), performing multiple tests will increase the probability of obtaining false positive observations tremendously (Noble 2009). Therefore, large experiments with many measurements, such as using metagenomics to detect resistance genes, require some form of correction for multiple testing. One way of doing this is to simply multiply each p-value with the number of tests performed (i.e. the number of genes investigated), referred to as the Bonferroni correction (Dunn 1959; 1961). However, in many explorative studies the Bonferroni correction is regarded to be too harsh, and therefore less stringent approaches, such as the Benjamini-Hochberg false discovery rate, are commonly used in large-scale experiments to control the number of false positive observations (Benjamini & Hochberg 1995).

With this in mind, a simplistic approach (log-transforming the data, followed by standard paired or unpaired t-tests) has been chosen for the data analysis in this thesis, which should be reasonably robust against false positive detections. Furthermore, to account for the large number of genes investigated, and thereby the large number of tests performed, the p-values for every gene have been corrected using the Benjamini-Hochberg false discovery rate (Benjamini & Hochberg 1995; Noble 2009).

Abundance and diversity of resistance genes – the risk perspective

Many studies of antibiotic resistance has focused on a few target genes analyzed by e.g. qPCR to determine their relative abundance in different environments (e.g. Knapp *et al.* 2011; Laht *et al.* 2014; Muziasari *et al.* 2014; Rutgersson *et al.* 2014). This provides good estimates for how common these particular genes are in a sample, but contributes no information on whether there are other genes present in the same sample or not. Thus, estimating the total resistance gene abundance and diversity in a sample using e.g. qPCR would require an overwhelming experimental effort. Metagenomics is better suited to achieve a more complete picture both, but is of course still limited to detection of the genes present in the reference database used. From a risk perspective, the relation between abundance and diversity is complicated. High abundances of a few resistance genes towards a single compound would likely be indicative of a direct selection pressure for those genes. If there are only a few highly abundant genes present, but they encode resistance towards different antibiotics, this suggest co-selection, perhaps of a single MGE containing all these genes together. Inferring the selection pressure is in this situation much more complicated, as there could be yet other genes on a plasmid that are selected for, which may have little to do with antibiotic resistance. Finally, high abundances of many different resistance genes against several classes of antibiotics are suggestive of

selection by multiple antibiotics, either at the same time or consecutively. However, this scenario could also arise by extensive co-selection, particularly if the fitness cost of maintaining resistance genes is low.

In the context of risks to human health, this has a number of implications. If we first consider the dissemination of resistant pathogens through the environment, the risks associated with those would be connected primarily to high abundance rather than diversity, simply due to that larger numbers of resistant pathogens increase the likelihood for human exposure to them. The diversity of resistance genes is from this perspective of less concern than the total abundance, or the abundance of resistance genes against the clinically most relevant antibiotics. However, if we instead consider the risk for recruitment of resistance genes into pathogens, high abundances of resistance genes simply implicates that the number of potential donors of transferrable resistance genes is high (Figure 6). Furthermore, high diversity of resistance genes means that there are many potential resistance genes to donate. Thus a combination of high diversity and abundance would be worst from a resistance recruitment perspective, as this means that there are many resistance genes and many potential donors that could transfer their resistance to pathogens or other human-associated bacteria. Finally, we may also consider the implications for resistance emergence. Novel – or more efficient – resistance mechanisms may emerge anywhere, at any time. However, critical for the fixation of new resistance genes in a population is the presence of a selection pressure for maintenance. High diversity of resistance genes suggests that resistance genes are maintained in the community (perhaps even in the absence of direct selection), and thus high diversity rather than abundance would point to risk environments for maintenance of novel resistance factors. It deserves to be noted in this context that there does not seem to be many environments with higher resistance gene abundances than in the human gut microbiome, but lower diversity (Figure 6). There are, however, several environments investigated in this thesis with higher diversity than the human gut, but the same or lower abundance. If a selection pressure is applied to such an environment, there is a large potential that the community can respond by increased resistance gene abundances, and that it thus serves as a reservoir of known, and potentially novel, resistance genes. A wide diversity of resistance genes also suggests a certain degree of resilience of the community to perturbations, particularly in the form of antibiotic exposure, since there is already an arsenal of protective measures present in among its populations. It also suggests that there may be additional, uncharacterized resistance factors present in the community, and in this way known resistance genes may serve as proxies for those not yet described.

The reasoning that a large diversity of resistance genes would render a community more resilient to antibiotics exposure (and perhaps also other perturbations) is an extension of the insurance hypothesis from metacommunity theory (Loreau *et al.* 2003), also referred to as response diversity (Elmqvist *et al.* 2003). This hypothesis postulates that higher species diversity leads to better adaptive capability of the community to disturbances. Whether similar response diversity can

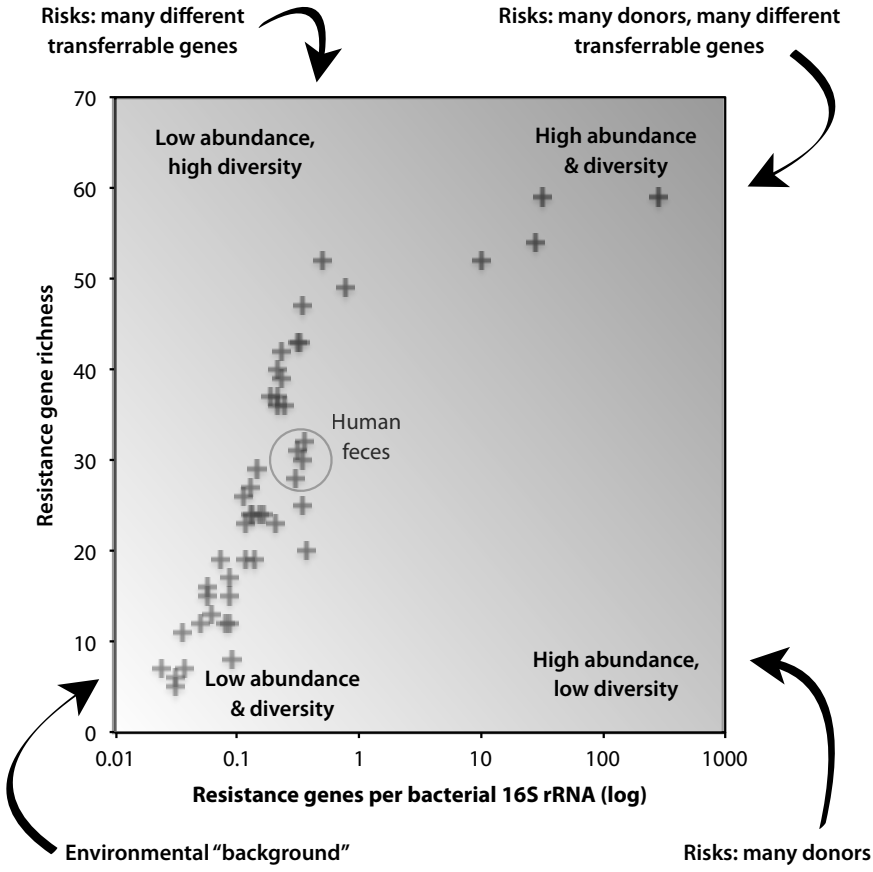


Figure 6. Abundance and richness of resistance genes across environments investigated in this thesis, and the risks to human health associated with resistance recruitment from the environment. Resistance gene richness has been normalized to account for different sequencing depths of samples.

be assumed for resistance genes is not clear, but as they constitute a functional trait, it could be argued that considering resistance gene diversity is a more direct measure than determining species diversity. However, species diversity *per se* is also related to risk for resistance recruitment to pathogens. A taxonomically highly diverse environment increases the numbers of both potential donors and recipients of resistance factors. Thus, large abundance and diversity of resistance genes in a community that also hosts a variety of taxa would, at least theoretically, be of higher risk for human health than if the number of species is low. Finally, it is important to note that while the abundances of resistance genes seem to go down dramatically when selection pressure is removed, there is compelling evidence that complete removal of resistance genes takes vastly longer time to happen, if ever (Levin *et al.* 1997; Enne *et al.* 2001; Björkholm *et al.* 2001; Enne *et al.* 2004; Perron *et al.* 2007;

Andersson & Hughes 2010). Thus, once a resistance gene has been enriched in a community, it may never be eradicated (Andersson 2003). This means that an environment that has once been subjected to antibiotic selection increasing the abundances of a diversity of resistance genes is more likely to remain diverse also after the selection pressure has been removed. It may in this way serve as a reservoir for resistance recruitment long after resistance gene abundances have returned to the levels of comparable environments not exposed in the first place.

Measuring the diversity of resistance genes

It was established above that not only the abundance of resistance genes may be of importance for determining risks, but also the extent of different genes found. However, it is unclear exactly how to establish the diversity of resistance genes, for example whether or not the relative abundances of different genes should be taken into account. Similar difficulties with estimating species richness in different communities have haunted ecology for more than half a century (Magurran 2004). A plethora of diversity indices designed for community ecology exists and are currently in use, each with its own advantages and shortcomings. The most basic such measurement would be to simply count the number of different species (or resistance gene types) encountered, establishing what is called the richness of the sample. This, however, is not without problems. First of all, the richness is intimately connected with sampling effort (in the metagenomics case the size of the sequencing library). One could account for this by normalizing the abundances of each gene in all samples to the size of each sample, thereby making them comparable, and then only count entries with a normalized abundance corresponding to finding at least one copy of the gene or species *in the smallest sample*. However, while this reduces the dependency on library size, it instead introduces a bias towards the most abundant entities. For this reason, rarefaction methods, in which the number of different species (or resistance gene types) encountered are plotted against the sampling effort required to detect them, have instead been suggested to deal with this problem (Hurlbert 1971; Hughes & Hellmann 2005). Other authors have argued that this practice is inadvisable for detecting differences in diversity between communities, and that mixture models are better suited for the task (McMurdie & Holmes 2014).

If we also want to account for the how evenly distributed the resistance genes in a community are, we need to consider the broad variety of ecological diversity indices. Commonly used indices to assess species diversity have been translated into ecology from information theory, although the rationale for their relevance on biological communities is somewhat questionable (Hurlbert 1971). In many studies of species diversity, the Shannon (Shannon & Weaver 1949) or Simpson (Simpson 1949) diversity indices have been employed. Both take the number of different species (or resistance genes) and their relative abundance into account, and thus entwine these two properties in a single number (Magurran 2004). Essentially, the Shannon index measures the entropy of the community (how far from evenly distributed the species/resistance genes are), while the Simpson index reflects the probability that two

sampled individuals in a community would belong to the same species (or resistance gene type).

Both the Simpson and Shannon diversity indices convey information about the structure of the studied community. Yet, are these properties relevant for ranking risks associated with resistance genes? Let us consider what a diversity index means in practical terms. Assume that you enter a forest looking for orchids. Knowing that the forest contains ten different orchid species (the richness) and 100 orchid individuals (total abundance) does not tell you much about whether you will be able to spot all ten species quickly or not, as some of them might be represented by a single individual. A diversity index would provide more information about this, as they also incorporate the aspect of community evenness. If the orchid species are evenly distributed, the diversity index will be high, and spotting all ten should take less time. Now assume that eight out of these ten orchid species are pollinated by a particular type of insect, and that this insect species is eradicated by an ecosystem perturbation. Suddenly, these eight orchid species will now disappear, leaving only the two with alternative pollination means.

What can we learn from this example? An important lesson is that the eight orchid species were selected against by the perturbation event, regardless of their abundance. For human health risks associated with antibiotic resistance genes in the environment, the reason the diversity of resistance genes is important is that a larger diversity presents the community with a larger arsenal of defense systems against antibiotic selection. As with the orchids, if a selection pressure from an antibiotic is applied, only resistant bacteria (either intrinsically, or those carrying resistance genes) will survive. Such events will therefore serve as bottlenecks for fixation of resistance genes in a population. However, as bacteria can also share genes between species and strains, a selection pressure promoting resistance genes may also cause these genes to end up in new species and genetic contexts. Importantly, these selection pressures will apply regardless of if the genes were abundant or not, and crucially without any respect of the evenness of resistance genes in the community. Thus, the usefulness of a diversity index such as Shannon's or Simpson's is negligible from a human health risk perspective. Instead, it makes sense to compare the richness or the shape of rarefaction curves between samples to evaluate diversity-related risks.

Furthermore, we must realize that the studied sample of resistance genes only comprises a subset of the total resistance gene types likely present in a community. Thus, the true richness of the sample is unknown, and information on the abundances associated with lowly abundant genes is either poorly estimated or lacking. This means that it might be informative to account for the unseen resistance genes in some way. Measures for extrapolating richness could be borrowed from ecology, for example the Chao1 (Chao 1984) and ACE (Chao & Lee 1992) estimators. In addition, resampling methods have been suggested to estimate true species richness of samples (Colwell & Coddington 1994). However, these estimators have been shown to fluctuate substantially with changing sample size (Hughes *et al.* 2001). As ecologists and statisticians still struggle with the problem of estimating the number of

rare species in a community, we can conclude that accounting for those is hard, and that for the time being we are probably best off comparing the richness of detected resistance genes in different samples, and hope that those numbers correspond reasonably well to the true richness. In addition, the methods for finding resistance genes using shotgun metagenomics only allow detection of known genes present in a reference database. The yet undiscovered resistance genes, of which there seem to be a multitude in the environment (Riesenfeld *et al.* 2004; Allen *et al.* 2009; Sommer *et al.* 2009; Lang *et al.* 2010; Torres-Cortés *et al.* 2011; Wichmann *et al.* 2014; Munck *et al.* 2015), and which avoid detection regardless of being abundant or rare, are incredibly hard to account for using richness estimators. Once again, one could assume that a large diversity of known resistance genes implies a broad range of unknown resistance factors as well, but to which degree this is true remains unknown.

Why do we want to assemble metagenomes?

Depending on where an antibiotic resistance gene is located, its ability to confer resistance, as well as its potency to spread to other bacteria, varies considerably (Dantas & Sommer 2012; Martinez *et al.* 2015). For example, different promoter regions may enhance or reduce the expression of a gene, and interactions with other gene products may influence the resistance function of the gene. Furthermore, a gene that is present on a plasmid or other mobile genetic element is vastly more likely to spread between bacteria than one firmly located on the bacterial chromosome (Martinez 2011; Martinez *et al.* 2015). In addition, the compatibility of a mobile resistance gene with its host also influences whether the gene encodes an efficient resistance mechanism in that specific context. Finally, genes mobilized by integrases may have modified 3' and/or 5' ends, which may also alter their expression in the new context. The latter is thought to have contributed to the efficiency of the NDM-1 carbapenemase gene in a variety of hosts (Dortet *et al.* 2012; Toleman *et al.* 2012). Because of the complex interplay between resistance genes and their genetic environment, it is important to consider the genetic context around resistance genes, as well as the taxonomy of their carriers. For the purpose of investigating the genetic and taxonomic context of resistance genes, both qPCR and functional metagenomics approaches falls short. The former does not provide any information of the genetic regions outside of the PCR primers used, and the latter generally employs too short inserted DNA fragments to enable precise contextual information, although often at least some usable information can be gained (Forsberg *et al.* 2012). To fully understand the genetic context of resistance genes, functional selection of resistant strains or resistance plasmids followed by analysis of their complete sequences is in principle required (Johnning *et al.* 2013; Casali *et al.* 2014; Salipante *et al.* 2015; Holt *et al.* 2015; Flach *et al.* 2015). This is, however, a rather labor-intensive approach, and it is also restricted to isolates that can be cultured and/or plasmids that can be captured by cultivable bacteria. Another approach to gain insights into the context of resistance genes is through the use of metagenomic shotgun sequencing, followed by computational assembly of the reads (Papers IV, V & VI; Ma *et al.* 2016). While this

method is limited to resistance regions abundant in the sample due to the requirement of large sequencing depth, it circumvents the need for cultivation and phenotypic resistance selection.

The current state of assemblers for metagenomic sequence data

The goal of assembling metagenomic data to get better insights into genetic context and synteny is not unique to resistance gene research. The uses for assembled sequence data vary greatly, and includes identification of new proteins (e.g. Kannan *et al.* 2007), studies of uncultivated microbial genomes (e.g. Venter *et al.* 2004; Hugerth *et al.* 2015), discovery of novel plasmids (e.g. Kristiansson *et al.* 2011; Zhang *et al.* 2011), and quantification of species and gene family abundances (see e.g. Qin *et al.* 2010; Karlsson *et al.* 2014; Nielsen *et al.* 2014). Due to the large variety of final aims for metagenomic assembly, several different assembly strategies for metagenomic data exist, and the general applicability of each approach is somewhat limited. Early metagenomics projects, which generated longer and fewer reads, generally utilized the same assemblers as genome projects, such as the Celera assembler for Sanger sequencing data (Myers *et al.* 2000), and the 454 assembler Newbler (Margulies *et al.* 2005) or MIRA (Chevreux *et al.* 1999) for pyrosequencing data. The assemblers used on long-read data are most often based on the overlap-layout-consensus algorithm (Staden 1979), which works well on smaller data sets, but quickly becomes vastly time and memory consuming, as complexity scales roughly quadratic with the number of reads due to the all-to-all comparisons of reads required (Pop 2009; Miller *et al.* 2010). For the massive amount of short-reads generated by e.g. the Illumina platform, such algorithms are thus unsuitable because of the dramatically increased complexity. The first widely used assemblers for short-read data – e.g. SSAKE (Warren *et al.* 2007) – solved this by greedy approaches, which are less computationally expensive, but produce sub-optimal solutions to the assembly problem (Miller *et al.* 2010). However, methods that instead reduce the complexity of the assembly graph by converting it into a de Brujin graph (Idury & Waterman 1995; Pevzner *et al.* 2001) quickly gained traction and remain the most used assembly methods for Illumina data. The de Brujin graph is less complex to build and traverse than the overlap-layout-consensus graph, making the assembly problem easier to solve (Li *et al.* 2012). This has resulted in a plethora of assembly algorithms based on de Brujin graphs, of which some popular examples are Euler (Pevzner *et al.* 2001), Velvet (Zerbino & Birney 2008), ABySS (Simpson *et al.* 2009) and SOAPdenovo (Li *et al.* 2010). Both Velvet and SOAPdenovo have been employed in metagenomic studies in their original forms. However, with increasing popularity of metagenomics, specialized software for metagenomic *de novo* assembly has been developed. These programs are often modified versions of genomic assemblers, such as Meta-Velvet (Namiki *et al.* 2012), Meta-IDBA (Peng *et al.* 2011) and Ray Meta (Boisvert *et al.* 2012). Although these adaptions in theory can improve assembly quality, the discernible difference between assemblies produced by e.g. Velvet and Meta-Velvet is minute (Vázquez-Castellanos *et al.* 2014), which is consistent with our own observations (Bengtsson-Palme,

unpublished data). Benchmarking of different assemblers on data where the true result is known has shown that the N50 metric often used to assess assembly quality is generally useless since an assembler that merges too many reads together will get high N50 values (generally interpreted as good), but does so at the cost of generating chimeric contigs (Salzberg *et al.* 2012; Magoc *et al.* 2013). This problem may be relatively minor for single genome assembly, since the possibilities for manual inspection and correction are fairly large. However, for metagenomic samples where many species are mixed, assessing which contigs that may be chimeric is almost impossible, which makes the numbers of errors a central metric in selecting an assembler software. In this context, it is worrying to note that particularly SOAPdenovo, but also Velvet, produce relatively high number of errors compared to other assemblers (Salzberg *et al.* 2012), such as ABySS and ALLPATHS-LG (Butler *et al.* 2008). However, ALLPATHS-LG requires a very specific set of sequence libraries to operate properly, making it unsuitable as a general-purpose assembly tool Furthermore, other comparisons indicate that ABySS and Velvet perform similarly (and produce comparatively few errors) on short-read data from bacterial genomes (Narzisi & Mishra 2011). A final interesting note is that the SGA assembler, which is based on the somewhat different strategy of using FM-indices derived from Burrows-Wheeler transforms of the data (Simpson & Durbin 2012), performs very well in these evaluations (Salzberg *et al.* 2012; Magoc *et al.* 2013). This shows promise for future method development in the area of metagenomic assembly.

Aside of avoiding assembly errors, another important consideration as metagenomic datasets continue to grow is the issue of scalability. An efficient assembler must not only be able to deliver mostly correct contigs, but must also be able to do so within a reasonable timeframe and within attainable memory limits. Even though assembly is generally carried out on large computer clusters with hundreds of gigabytes of RAM, assembly of some metagenomic datasets is still not feasible with current methods (Scholz *et al.* 2012; Howe *et al.* 2014). This leads to a number of important compromises between the most accurate and most efficient assembly algorithms. One key parameter of large-scale assembly is that the software should be scalable across multiple processor cores and nodes (individual machines) in a computer cluster. Two assemblers have struck a reasonable balance between accuracy and scalability for metagenomic assembly: ABySS and Ray. Both are highly scalable, while still producing results comparable to those of Velvet (Narzisi & Mishra 2011; Paper VI), which is arguably not the most accurate assembler, but certainly not the worst. However, for really large metagenomes neither of these assemblers are sufficiently memory efficient, which has spurred the development of alternative assembly strategies. For example, reads can be binned based on k-mer content prior to assembly reducing the need to assemble all the reads at once (Pell *et al.* 2012). Furthermore, reads from low-coverage regions can be filtered out prior to assembly (Hess *et al.* 2011; Mackelprang *et al.* 2011), or reads from high coverage regions can be set aside, a strategy referred to as digital normalization (Howe *et al.* 2014). Finally, merging of sub-samples of reads assembled individually has been proposed as a

possible, albeit sub-optimal, assembly strategy (Scholz *et al.* 2014). A completely different approach to metagenomic assembly is to target only regions of interest in the metagenome, which also reduces the complexity of assembly. Such approaches has been implemented in assemblers such as EMIRGE for bacterial rRNA (Miller *et al.* 2011), and the general purpose SAT-Assembler (Zhang *et al.* 2014). In addition, Xander provides an approach for parsing the de Brujin graph using hidden Markov models (HMMs) representing genes of interest (Wang *et al.* 2015). While this provides for more accurate assembly of the target genes, it does not solve the complexity problem, as the complete de Brujin graph must still be constructed prior to simplifying the graph using HMMs.

In this thesis, the Velvet assembler was used for the smaller metagenomes of Paper V, while Ray Meta was employed in Papers IV and VI where the sequencing libraries were larger. In addition, the Ray assemblies in Paper VI were compared to Velvet assemblies of the same samples, showing very similar results. In Paper IV, the merging-of-subsamples strategy (Scholz *et al.* 2014) was utilized to assemble the technical replicates from each sample into a single assembly. For this, we used Ray to generate the initial assemblies, and Cap3 (Huang *et al.* 1999) to merge the contigs from each separate assembly together.

Assembly of genes existing in multiple genomic contexts

The above-mentioned problems related to increasing dataset sizes and complexity of metagenomes are common to all metagenomic studies employing assembly approaches. However, there are certain challenges that are relatively unique to the assembly of resistance regions from metagenomic sequencing data (although similar problems also appear in other contexts where the genes of interests are highly conserved but can exist in multiple species). The greatest obstacle to enable assessment of the context of resistance genes identified in metagenomic data is the resistance genes themselves. We are often interested in investigating whether a resistance gene is present on a mobile genetic element or not, as this property is strongly related to the relative risk associated with the gene (Martinez *et al.* 2015; Paper VII). However, resistance genes present on mobile genetic elements are often better conserved between species (since they can be transferred directly) than chromosomal resistance genes. In addition, if they are mobilized in integrative elements they can exist in multiple similar, but not identical, genetic contexts (Frost *et al.* 2005; Norman *et al.* 2009). This presents a problem for assembly software working with short reads. Many times, there can be multiple possible branches out from a highly conserved part of a resistance gene or resistance gene cassette (Figure 7). Almost all assembler software handles this by splitting the contigs at the branching points, although some use coverage information or other external data (such as read-pair information) to avoid unnecessary splits and handle splits more intelligently. Regardless, the result is a fragmented assembly that does not contain much information about the genetic context of the resistance gene of interest. In the example presented in Figure 7, no contextual information is retrieved for resistance

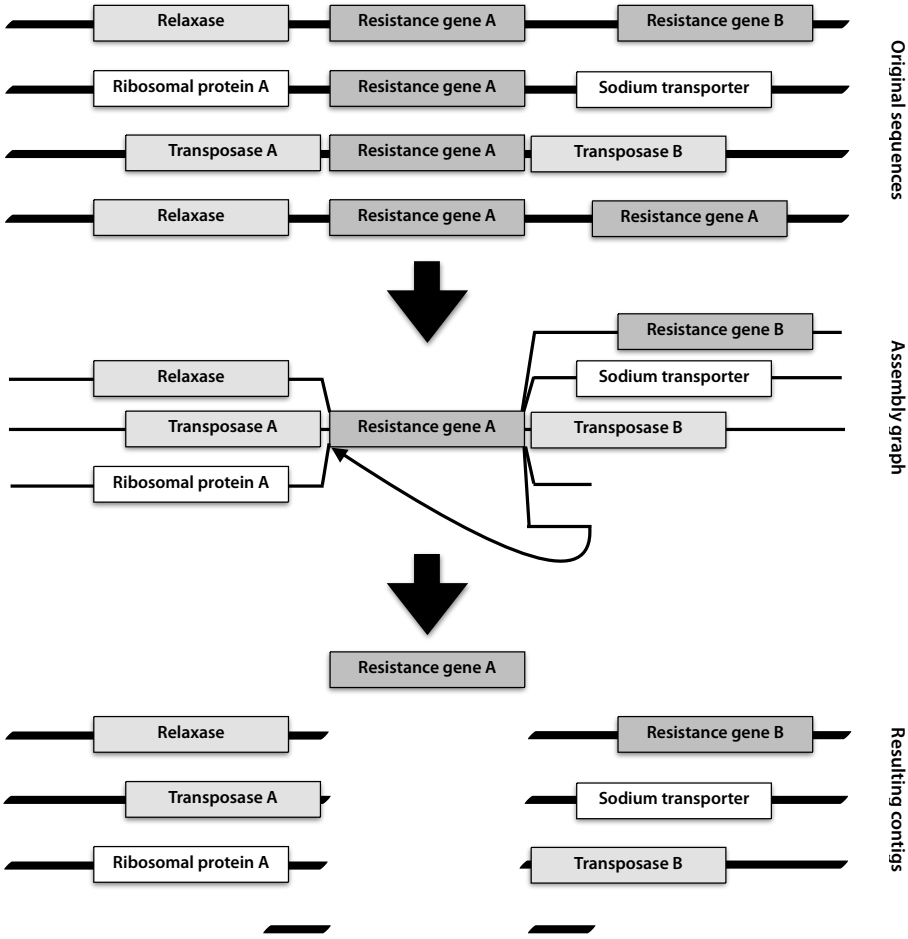


Figure 7. Identical resistance genes may exist in multiple genetic contexts (top), which presents assembly software with serious problems, as they cannot identify which reads that originated from which context (center). Almost all assemblers solve this by splitting the contigs at the ambiguous positions, resulting in a fragmented assembly (bottom). Notice how the repetition of resistance gene A cause a loop in the assembly graph, resulting in two short contigs containing no genes.

gene A, since it ends up on a single contig without any flanking regions. This does not only obscure the information about whether a resistance gene is transferrable between bacteria, but also severely limit our ability to detect resistance genes that are co-localized. In addition, resistance genes are often not identical across their entire length, but rather contain identical regions. In those cases, the individual resistance genes may also be split up on multiple shorter contigs, further complicating the assessment of the assembly (Figure 8).

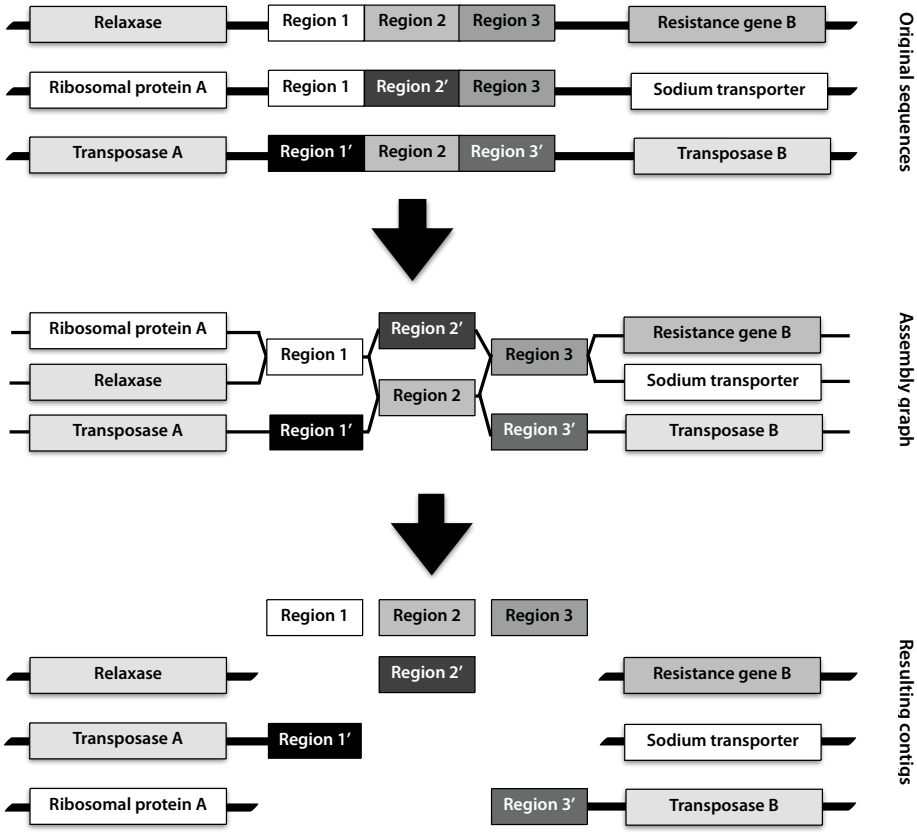


Figure 8. Resistance genes can have certain regions that are identical between variants even if they encode slightly different proteins. This can further split the assembly up in even smaller fragments than in the example of Figure 7. Note that most resistance regions in this example are not assigned to any context, and that no full-length variant of the resistance gene could be assembled.

The problems related to multiple contexts get worse the more common a resistance gene is, since common resistance genes are more likely to be detected in multiple contexts. In addition to these examples where true biological variation causes assembly problems, sequencing errors may also break the assembly up in a similar fashion as in Figure 8, although assemblers are generally better at handling such problems than true biological variation. In addition to resistance genes existing in multiple contexts, integrases and transposases are prone to the same types of problems, and break assemblies up in a similar way, resulting in contigs containing, e.g., one or two resistance genes and a (sometimes partial) ISCR or integrase sequence.

The TriMetAss assembler and further method development

A partial remedy to the problem of genes occurring in multiple contexts is, as mentioned above, to use coverage or read-pair information to traverse the assembly graph (Figures 7 and 8; center). Using coverage data, the original contexts of the example sequences in Figure 7 may be reconstructed, given that the four original sequences are sufficiently differently abundant in the sample (and thus generates differing coverage information. However, coverage information will not be able to solve the loop structure resulting from a repetition of the same gene. Furthermore, in the second example (Figure 8), coverage information alone cannot solve the problem, as parts of the gene of interest are shared between the original sequences and thus the coverage of the different regions will itself be uneven. The more complex the structure of shared regions is, the harder it will be to predict the original contexts using coverage data. In more complicated cases, information about which read pairs that were connected can improve the situation, although the usefulness of this in practice is limited unless several sequencing libraries with different insert sizes have been constructed.

Although the specific problem of highly conserved genes appearing in multiple contexts is fairly unique to metagenomic resistome investigations, a closely related problem exists in *de novo* RNA sequencing of eukaryote organisms. If there is no reference genome for a particular organism, and its gene expression is measured using RNAseq, a *de novo* assembly of the transcripts is often performed to allow estimation of the relative abundances of different mRNAs (Wang *et al.* 2009). In many eukaryotes, the same gene may encode several different mRNAs, so called alternative splicing (Black 2003). This causes the assembly to branch in ways very similar to those in Figure 8. Thus, assembler software adapted for RNAseq data has had to deal with similar types of problems, although only on a local single-gene level. Different software tools have been constructed to deal with mRNA assembly, including Trans-ABySS (Robertson *et al.* 2010), Oases (Schulz *et al.* 2012), and the commonly used Trinity package (Grabherr *et al.* 2011). While it in theory makes sense to apply e.g. Trinity directly on metagenomic data, it turns out that this method is not feasible in practice, since it requires vast amounts of computer memory. To reduce the complexity of the assembly problems to the most interesting regions (in our case corresponding to resistance genes), a targeted assembly approach was developed to improve the assessment of genetic context around resistance genes (Paper V). In this approach, the contigs containing resistance genes from a regular metagenomic assembly are used as seed sequences, and turned into a reference database for a Vmatch (Kurtz 2010) search against the complete set of read pairs. The reads matching to the seed contigs are then assembled using Trinity. The resulting set of contigs from Trinity are then used as seeds for another round of search using Vmatch against the complete set of reads, as above. All matching read pairs, including those matching in the previous round, are then used for another assembly with Trinity. This iterative process is repeated until a stop criterion is fulfilled, for example that no more reads can be assembled using Trinity. The iterative assembly method has been

released as a software package called TriMetAss, and was successful in extending the resistance contigs generated in Paper V. However, this iterative assembly scheme is both rather time-consuming and also does not solve the problem of multiple context genes – it only partially alleviates the issue. Thus, future method development could expand upon the ideas outlined by TriMetAss, by e.g. constructing a complete de Brujin graph for all reads matching to a seed sequence, and then present all possible contigs that this set of reads could produce along with the probability for them being put together by random chance. This probability could be calculated based on a read coverage model, and could also take read pair information into account. The user would then be able to set a probability cutoff for when a contig is to be trusted, and get an estimate for when the set of contigs generated would explain a certain fraction of their reads mapped to, e.g., resistance genes. Unfortunately, the development of such assembly software is not straightforward and is out of scope of the present thesis.

Deducing microbial taxonomy from metagenomic data

Assessing taxonomic composition using metagenomic data

Since changes of the resistance gene composition of a microbial community may not only be caused by direct selection for resistance factors, but also by changing conditions favoring species that happen to carry certain resistance genes, it is important to interpret resistance gene frequencies in the context of taxonomic composition. Furthermore, insights into the types of bacteria inhabiting an environment may provide clues that can guide risk management, e.g. in terms of the presence of pathogenic species or the proportion of human-associated bacteria. There are several approaches to determine the taxonomic structure of a studied microbial community based on metagenomic sequencing data. Three fundamentally different types of methods are commonly used: mapping of reads to reference genomes, binning based on nucleotide k-mer composition, and classification of certain barcoding genes, commonly the 16S or 18S rRNA genes (Zepeda Mendoza *et al.* 2015). The three approaches each have their respective advantages and drawbacks, which will be briefly outlined below.

Genome mapping approaches utilize databases of completely sequenced microbial genomes (or partially sequenced, but the latter may introduce biases that are difficult to compensate for). Common reference genome catalogues include the human microbiome project reference database (Human Microbiome Jumpstart Reference Strains Consortium *et al.* 2010) and the NCBI GenBank and RefSeq databases (O'Leary *et al.* 2016; Clark *et al.* 2016). The raw sequence reads are mapped to the reference database using read mapping tools such as BWA (Li & Durbin 2009), Bowtie (Langmead & Salzberg 2012), STAR (Dobin *et al.* 2013), or less commonly more sensitive alignment programs like BLAST (Altschul *et al.* 1997) or BLAT (Kent 2002). In addition, several tools exist that have streamlined this process, of which some include post-processing steps to account for biases in terms of different genome sizes and GC content, for example MEGAN (Huson *et al.* 2011), CARMA3 (Gerlach & Stoye 2011), Genometa (Davenport *et al.* 2012), MEDUSA (Karlsson *et al.* 2014) and GOTTCHA (Freitas *et al.* 2015). In addition, several more or less automated pipelines for metagenomic analysis exist, which also include whole-metagenome similarity strategies for taxonomic classification, such as the frequently used MG-RAST (Meyer *et al.* 2008) and CAMERA (Seshadri *et al.* 2007) web-portals.

Since the genome mapping approach can be very computationally intensive for large metagenomes, an alternative approach to ease the computational load is often employed, namely taxonomic binning based on sequence composition. This is generally done by counting the occurrences of short nucleotide sequences (k-mers) of e.g. four bases (tetramers). Compositional methods require a database of k-mer frequencies among the reference genomes, which the composition of each read (or assembled contig) can be compared to. Many tools exist for this task, including Kraken (Wood & Salzberg 2014), PhyloPhytia (McHardy *et al.* 2007) and MetaCV

(Liu *et al.* 2012). Furthermore, the Ray assembler (Boisvert *et al.* 2012) includes the option to perform such k-mer based taxonomic assignment of contigs integrated within the assembly process. An alternative way of reducing computational load of classification is to only target specific barcoding genes, most commonly the 16S and 18S rRNA genes, and infer taxonomy based on only these sequences. These methods are generally based on either fishing out the barcoding regions prior to classification, or performing a similarity search of the entire metagenome to a reference database only containing barcoding genes. While the latter approach is conceptually identical to the whole metagenome similarity based methods described above, the reference database of barcoding genes can be substantially smaller, drastically reducing computational time, particularly if heuristic algorithms are applied in the mapper. An upside of using pre-filtration to extract only the barcoding genes from the metagenome is that a more sensitive software tool can be used for classification, as the dataset size usually can be reduced between 100 and 1000 times. Quick extraction of barcoding regions from a metagenome can be performed using, e.g., the MetaRNA (Huang *et al.* 2009), riboPicker (Schmieder *et al.* 2012), SortMeRNA (Kopylova *et al.* 2012) or Metaxa (Bengtsson *et al.* 2011) software packages. For the subsequent classification step, analysis tools for community composition used in PCR-based studies of barcoding genes can be employed. For example, the QIIME (Caporaso *et al.* 2010), Mothur (Schloss *et al.* 2009) and Rtax (Soergel *et al.* 2012) packages support classification of barcoding genes derived from metagenomic DNA fragments. In addition, tools used for classification of complete metagenomic data can also be used on extracted fragments. Commonly used tools for this task are MetaPhlAn (Segata *et al.* 2012), MetaPhyler (Liu *et al.* 2011b) and Phylosift (Darling *et al.* 2014). Furthermore, the RDP naïve Bayesian classifier (Wang *et al.* 2007) is often employed for this classification besides not being designed for classifying fragmentary short read data. It has, however, been shown to perform sub-optimally in these circumstances (Paper I).

Depending on the research questions asked and the sample material under study, the best choice of analysis method may vary. Common to most research endeavors is the need for reliable classifications at a relevant taxonomic level, often corresponding to the species or genus, but sometimes as coarse as the phylum level. In other words, although sensitivity in terms of classifying as many reads as possible is desirable, it is most often more important that the reads that are classified are inferred to the correct taxa. In addition, it is also crucial that the fraction of reads classified represents the distribution of organisms present in the original sample. From a recent evaluation of different classification methods (Peabody *et al.* 2015), the following overall conclusions can be drawn. First, the accuracy of methods varies dramatically. The precision (correct classifications per total classifications made) of the methods ranged from about 5% (Kraken) to over 90% (DiScRIBinATE; Ghosh *et al.* 2010). Second, although variation was large between methods, similarity-based methods generally outperformed k-mer and composition-based approaches in terms of precision. Third, most methods radically overestimated the number of species present in a mock

community with a known number of taxa. Finally, the only barcoding gene based method investigated (MetaPhyler) performed very well in terms of precision, while being among the methods least prone to over-predicting the number of present species. Interestingly, we have found similar large variations of precision among software tools for rRNA analysis (Paper I), suggesting that thorough investigation of which tools that are not up to the task of analyzing short read data is warranted. Taken together, it seems that given sufficient computational resources, the use of similarity or barcoding gene based analyses of taxonomy should be preferred over k-mer and compositional analyses. When choosing between the whole metagenome similarity and barcoding gene strategies, it is important to keep the reference database in mind. Both methodologies require a sequence from a corresponding taxa to be present in the reference database to make a taxonomic classification of a read. However, for many microbial communities, such reference data does not exist. As only a minor fraction of bacteria can be cultured (Amann *et al.* 1995), most microbes have not had their complete genome sequenced. This means that the available reference databases are severely biased in terms of species present, generally with a vast overrepresentation of human-associated bacterial species. Although the same is true for the sequences present in reference databases for barcoding genes, such as SILVA (Yilmaz *et al.* 2014), GreenGenes (McDonald *et al.* 2012) and UNITE (Kõljalg *et al.* 2013), much greater numbers of barcode sequences exist for microbes than full genomes. This means that for communities where a large portion of the taxa present are expected to be represented in a genome database, such as in the case of human gut samples and the HMP reference database (Human Microbiome Jumpstart Reference Strains Consortium *et al.* 2010), it makes sense to use whole metagenome similarity based methods for taxonomic assignment of reads. However, in most environmental communities, we do not expect to have a complete picture of the taxa present – most often far from so – and using a whole metagenome based strategy may thus strongly bias results towards already sequenced species, and accordingly leave a larger portion of reads unclassified (or misclassified). In this case, a barcoding gene based approach would be advisable in order to as far as possible avoid biasing results. An additional advantage of the latter type of methods is that a large set of tools for downstream analysis of diversity and community composition is available in, e.g., QIIME (Caporaso *et al.* 2010), Mothur (Schloss *et al.* 2009), and Vegan (Oksanen *et al.* 2011). Directly utilizing such tools on fragmentary data involves some inherent obstacles however (Bengtsson *et al.* 2012), and is therefore, depending on the desired analysis, not completely straightforward. In this thesis, a barcoding gene based approach – Metaxa2 – has been used (Papers III-VI), complemented with a whole metagenome similarity search when human fecal samples were studied (Paper VI).

Improving the accuracy of taxonomic classification of metagenomic data

As mentioned above, central features of methods for taxonomic analysis of microbial communities are that they return results that are representative for the community at large, and that taxonomic classifications are correct. Although a range of tools exist

for classification of barcoding genes, the most commonly used are adapted for amplicon sequencing targeting the same region of the gene, and often expecting each read in the input data to be at least a few hundred basepairs long. This, however, is not the case for shotgun metagenomic data, which is randomly fragmented and often sequenced using technologies yielding reads with a length between 75 and 150 nucleotides each. Furthermore, many existing methods for extracting rRNA reads from shotgun metagenomes perform sub-optimally and either generates excessively many false positive or false negative assignments (Bengtsson-Palme, unpublished data). On top of that, they generally do not sort out archaeal, bacterial, eukaryote, mitochondrial and chloroplast rRNA sequences efficiently, resulting in bias and noise in the downstream analyses (Taberlet *et al.* 2012). An exception to this is the Metaxa software, which was explicitly designed to handle these two types of problems with metagenomic datasets (Bengtsson *et al.* 2011). However, this software tool was designed for Sanger sequencing and pyrosequencing technologies and is thus not very efficient for use on the larger short-read datasets generated by the Illumina platform.

As a solution to these problems, an extensive update to the Metaxa software was made. The updated version – Metaxa2 – is adapted for larger datasets with shorter read lengths, and also includes a taxonomic classification tool that has among the best precision of current barcoding gene classification tools on shotgun metagenomic data (Paper I). Metaxa2 and the evaluation of the software are described in detail in Paper I. Importantly, while all tested classifiers performed well on 1000 basepair reads (except the QIIME implementation of Mothur; Figure 9), precision dropped rapidly at shorter read lengths (Figure 10). Particularly, the performance of Mothur and the RDP classifier, which both rely on naïve Bayesian statistics for classification, deteriorates quickly with shorter read lengths. Notably, at a read length of 100 nucleotides even Metaxa2 only classified 70% of rRNA sequences to the genus level, which was slightly lower than the proportion classified by Rtax. However, almost all assignments of rRNA sequences to a genus by Metaxa2 were correct, while Rtax displayed a large proportion of misclassifications (Figure 10). This evaluation shows the value of utilizing classifiers specifically designed to handle fragmentary short read data from shotgun metagenomes (such as Metaxa2 and Rtax) rather than relying on software originally conceived to deduce taxonomy based on amplicon sequencing data from the same region of the rRNA sequence and with longer lengths (such as Mothur and the RDP Naïve Bayesian Classifier). Finally, we also show that Metaxa2 cuts a reasonable tradeoff between sensitivity and specificity at short read lengths, allowing high-precision taxonomic analysis of shotgun metagenomes.

There are several reasons why Metaxa2 achieves better performance than most other tools on short read data. First of all, Metaxa2 only considers rRNA sequence fragments for which the conserved regions can be detected (Hartmann *et al.* 2010). This largely avoids false positive identifications due to similarity to the more random hypervariable regions of the rRNA genes. Nonetheless, the hypervariable regions are still used in the classification step, as the conserved regions often do not provide sufficient variation between species and genera to allow unambiguous species or

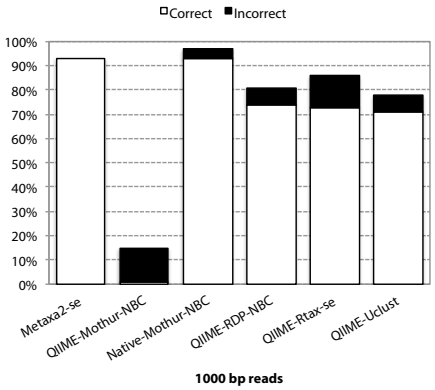


Figure 9. Percent correct and incorrect classifications of Metaxa2 on simulated shotgun metagenomic 1000 bp fragments with known taxonomic origin, compared to other taxonomic classification methods for barcoding genes. Data for the figure taken from Paper I.

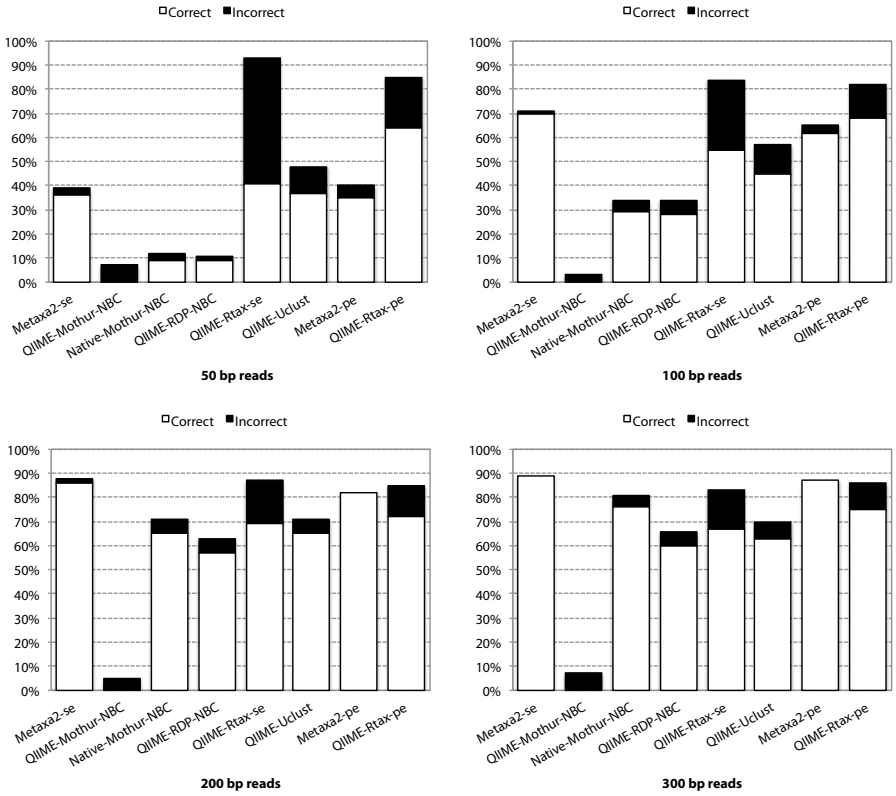


Figure 10. Percent correct and incorrect classifications of Metaxa2 on simulated shotgun metagenomic fragments with known taxonomic origin, compared to other taxonomic classification methods for barcoding genes. The two rightmost bars of each sub-figure represent the performance on paired-end data. Data for the figure taken from Paper I.

genus assignments. Second, the database used by Metaxa2 has been manually curated to only contain full-length reference sequences with high-quality annotation, with a standardized taxonomy. This excludes uninformative entries from e.g. uncultured organisms, as well as organisms where taxonomy is only partially defined. Third, the Metaxa2 classifier is written from the ground up to rather output a less specific taxonomic affiliation than making a precise designation on e.g. the species level that is plausible to be incorrect. Finally, since Metaxa2 separates archaeal, bacterial, eukaryotic, chloroplast and mitochondrial entries, the potential for confusing these disjoint – but homologous – classes of rRNA with each other is lower than for most other software tools. Despite this, the Metaxa2 classifier is not perfect – but clearly performs superior in terms of precision compared to competing software packages (Figures 9 and 10).

Minimal selective concentrations for antibiotics

Methods for determining minimal selective concentrations

To be able to define whether an environment presents its inhabitants with a selection pressure favoring antibiotic resistance, it is necessary to understand which concentrations of antibiotics and other substances that drive resistance selection. Although knowledge in this area is scarce, there have been several attempts to determine the minimal selective concentrations (MSCs) of antibiotics, primarily in laboratory setups (Table 2). Initial work establishing the MSC concept deduced the selective concentrations by letting isogenic bacterial strains differing only in a certain resistance factor (and marker genes) grow in sub-inhibitory concentrations of antibiotics (Gullberg *et al.* 2011). In this work, MSCs at between 1/230 and 1/4 of the minimal inhibitory concentrations (MICs) were measured, signifying that the difference between the two can be substantial, at least in a simplified, but highly controlled, laboratory testing system. Similar experiments have since been repeated for complete resistance plasmids and combinations of antibiotics, biocides and metals (Gullberg *et al.* 2014).

However, the settings for competition experiments are unlikely to match the conditions in environmental compartments subjected to antibiotics. First of all, this is simply due to that in a microbial community, several different species with varying degree of susceptibility to the antibiotic are present. Therefore, community level effects acting on the most sensitive species may occur at concentrations below those measured in laboratory strains. This also makes it difficult to predict to which extent more tolerant species and strains will fill the niches made available by antibiotic selection (O'Brien 2002). Second, most environments are nutrient-poor compared to laboratory systems, which may alter bacterial susceptibility to antibiotics and change the sub-MIC selective window in unpredictable ways. Third, other selective forces, including nutrient availability and predation, may influence the selection process and render selection by low concentrations of antibiotics and other toxicants less of concern for bacteria competing for available resources (Bengtsson-Palme *et al.* 2014a). At the same time, some antibiotics, such as tetracyclines and fluoroquinolones, are not readily degraded in the environment and could thus exert a chronic selection pressure on microbial communities. How long-term chronic low-concentration presence of antibiotics influences resistance development is this far uninvestigated, as is the influence of mixtures of several different antibiotics in combination with other selective agents (Backhaus 2014). For all these reasons, it is important to establish MSCs not only in competition experiments, but also in ecologically relevant testing systems. Some insights can be gained from pioneering research in this direction performed within ecotoxicology (Brosché & Backhaus 2010), using planktonic bacteria as a testing system. However, the main target of study in ecotoxicology is seldom the minimal selective concentration for resistance selection, but rather to measure toxicity in terms of growth inhibition (Blanck 2002). Thus, results from

Table 2. Experimentally determined minimal selective concentrations of antibiotics and other substances[a]

Substance	Strain-specific MSC (µg/L)	Community-based MSC; type of community specified in parenthesis	References
Chloramphenicol		> 49 µg/L (limnic plankton)[b]	Brosché & Backhaus (2010)
Chlortetracyclin		> 4 µg/L (limnic plankton)[b]	Brosché & Backhaus (2010)
Ciprofloxacin	0.1-2.5[c]		Gullberg *et al.* (2011)
Enrofloxacin		< 0.1 mg/kg body weight (rat gut)[d]	Lin *et al.* (2014)
Erythromycin	< 200		Gullberg *et al.* (2011)
Fusidic acid		> 5 µg/L (limnic plankton)[b]	Brosché & Backhaus (2010)
Kanamycin	470		Gullberg *et al.* (2011)
Oxytetracyclin		< 20 µg/kg (soil)	Shentu *et al.* (2015)
Rifampicin		> 23 µg/L (limnic plankton)[b]	Brosché & Backhaus (2010)
Streptomycin	1000	> 5700 µg/L (limnic plankton)[b]	Brosché & Backhaus (2010), Gullberg *et al.* (2011)
Tetracycline	15	< 1 µg/L (aquatic biofilms)	Gullberg *et al.* (2011), Paper III
Trimethoprim	< 2		Gullberg *et al.* (2014)
Arsenite	90 µM		Gullberg *et al.* (2014)
Cu(II) sulfate	90		Gullberg *et al.* (2014)

[a] Note that the MSC here is defined as any endpoint for resistance selection, such as resistance gene enrichment, competitive advantages in two-strain experiments, or increased number of insensitive colonies on selective plates.

[b] The concentrations from Brosché & Backhaus (2010) are based on no effect concentration calculations derived from effect concentration data and have thus only been estimated, not actually measured.

[c] Concentration dependent on which resistance mutation that was introduced.

[d] Since this corresponds to the dose given to rats, interpretation of this concentration is not straightforward. However, 0.1 mg/kg bodyweight corresponds to 1/100 of the therapeutic concentration (Lin *et al.* 2014).

ecotoxicological community studies cannot straightforwardly be used to infer MSCs, and studies directly assessing the MSCs for resistance in microbial communities of different antimicrobial agents are therefore warranted. Indeed attempts aimed at assessing MSCs in complex communities have been made. The MSC of tetracycline in stream periphyton communities has been reported to be below 0.5 µg/L (Quinlan *et al.* 2011), although limited replication and lack of a dose-response relationship for resistance increase make the results somewhat uncertain. Addition of oxytetracycline to arable soil has been shown to increase resistance gene frequencies at 20 µg/kg soil

(Shentu *et al.* 2015) and a dose of 0.1 mg enrofloxacin per kg bodyweight increased the abundance of resistant *E. coli* isolates in the rat gut (Lin *et al.* 2014). Furthermore, in Paper III of this thesis, the MSC for tetracycline resistance in aquatic biofilms is shown to below 1 µg/L. Taken together, there are scattered studies showing effects on resistance selection at sub-lethal antibiotics concentrations. However, as shown in Table 2, the available data is scarce, and more comprehensive studies are needed to enable proper risk assessment and regulation of environmental antibiotic releases (Ågerstrand *et al.* 2015).

Theoretical estimation of selective concentrations of antibiotics

Because of the limited availability of MSC data for antibiotics and the extensive amount of labor required to perform such studies on large scales, efforts to theoretically determine the MSCs for different antibiotics have been attempted (Tello *et al.* 2012). In this study, no-effect concentrations for ciprofloxacin (~0.2 µg/L), erythromycin (~8 µg/L) and tetracycline (~10 µg/L) were estimated based on MIC data taken from the EUCAST database (European Committee on Antimicrobial Susceptibility Testing 2016). However, the incentives to arrive at theoretical selective concentrations for a much larger range of antibiotics are strong, as it would enable implementation of concrete emission limits and environmental standards (Ashbolt *et al.* 2013; Ågerstrand *et al.* 2015).

In Paper II, we extend the approach by Tello *et al.* (2012) and theoretically determine MSCs under the assumption that selective concentrations *a priori* need to be lower than inhibitory concentrations. Thus, the lowest MIC for a particular antibiotic that has been determined for any species, should correspond to the maximum possible MSC for that antibiotic, given that one also compensates for limited species sampling when establishing the lowest MIC. Accordingly, we investigated the MIC data in EUCAST, and for each of 111 antibiotics and 11 combinations of antibiotics present in the database we selected the species with the lowest MIC. For 13 antibiotics, the lowest MIC corresponded to the lowest concentration tested, and for those we estimated a lowest MIC based on the sensitivity distribution for that combination of antibiotic and species compared to that of all other species. This resulted in lowest MICs in the range of 0.69 – 32,000 µg/L, although most lowest MICs were in the 4 – 125 µg/L range. After this had been established, we investigated if there was a link between taxonomic distance (based on rRNA similarity) of two species and the difference in lowest MIC. We found that if such a link exists, it is very weak across all antibiotics investigated together ($R^2 = 0.02$). When we tested the same relationship for each antibiotic separately, eleven had significant associations between rRNA dissimilarity and difference in lowest MIC. However, five of those had negative slopes (indicating that more divergent species would have more similar lowest MICs than closely related ones; a quite counterintuitive finding). Since there was no apparent systematic link between taxonomic distance and lowest MIC, we chose not to compensate for this when estimating MSCs for complex communities. To avoid bias due to the small

Table 3. Predicted no-effect concentrations (PNECs) for different classes of antibacterials. Concentrations given in µg/L

Antibiotics class	Average PNEC	Median PNEC	Minimal PNEC	Maximal PNEC	Antibiotics in class
Beta-lactams	2.888	0.5	0.016	64	48
Quinolones	1.896	0.1875	0.032	16	14
MLS[a]	1.255	1	0.064	4	12
Aminoglycosides	8.056	2	0.5	32	9
Peptidyl transferases	2.413	1	0.064	8	5
Tetracyclines	1.1	1	0.5	2	5
Polypeptides	4.0	2	2	8	3
Glycopeptides	4.25	4.25	0.5	8	2
Lipopeptide	16.5	16.5	1	32	2
Antifolate combinations	0.5	0.5	0.5	0.5	1
DHFR inhibitors	0.5	0.5	0.5	0.5	1

[a] Macrolide-Lincosamide-Streptogramin antibiotics

numbers of tested species for many antibiotics, we subsampled the lowest MIC data for all antibiotics that had been tested against more than 30 species, and thereafter assessed the effect of small sample size on the estimated upper boundaries for MSCs. We used the result of the subsampling to estimate how much lower the actual lowest MIC could be for antibiotics with small number of tested species, to account for that the EUCAST data may correspond to the upper part of the sensitivity distribution, and that complex communities therefore may contain bacteria that could be more sensitive than those reported in EUCAST. Finally, we predicted no-effect concentrations (PNECs) for each antibiotic based on an assessment factor of 10 (see Paper II for a discussion on the choice of this assessment factor). The PNECs closely correspond to MSC estimates, since a PNEC should (given that it is correctly predicted) by definition be slightly lower than the MSC. The final PNECs for antibacterials were in the range of 0.016 µg/L to 64 µg/L (Table 3). It should be noted that the span between minimal and maximal MSC is very large for some antibiotics, such as beta-lactams and quinolones. This is reasonable because these antibiotics have seen substantial development of a variety of subclasses that are effective at very different concentrations.

Validation of the MSC of tetracycline in complex microbial communities

The MSC for tetracycline was estimated in Paper II to be above 1 µg/L in complex communities. To validate this prediction, the MSC for resistance selection of tetracycline was measured experimentally in Paper III. In this paper, the aim was to

Table 4. Experimentally determined minimal selective concentrations of tetracycline for different endpoints in complex microbial communities

Endpoint	MSC range in dose response experiment	MSC established in follow-up experiments
CFU count on R2A plates with 20 µg/mL TC	1-10 µg/L	10 µg/L
MIC range	10-100 µg/L	-
PICT, leucine uptake after short-term TC challenge	100 µg/L	-
Increased resistance gene abundances, metagenomics	0.1-10 µg/L	-
Increased resistance gene abundances, qPCR	1-10 µg/L	≤ 1 µg/L
Changes to taxonomic diversity	n.d.	-
Changes to taxonomic community composition	1-10 µg/L	-

TC = tetracycline; n.d. = no change detected

determine the MSC of tetracycline for a variety of phenotypic and genotypic endpoints representing resistance selection. To this end, bacterial communities were allowed to establish biofilms in aquaria in the laboratory for nine days under different tetracycline concentrations. An initial exposure response experiment was carried out using a 10-fold dilution series of tetracycline levels. This experiment provided concentration ranges in which selection was observed for different endpoints, and constituted an evaluation of endpoints in terms of sensitivity to detect tetracycline-related community changes, including on the taxonomic level. In two follow-up experiments, MSCs were established for a subset of the endpoints, using more replicates in the range where selective concentrations were expected to be based on the initial exposure response experiment. The biofilms in the different aquaria were phenotypically profiled using colony forming unit (CFU) counts, range of measured minimal inhibition concentrations (MICs) for isolates, and pollution-induced community tolerance (PICT) quantified as inhibition of leucine uptake (Blanck 2002). Genotypic profiling was based on the frequencies of antibiotic resistance genes and changes of taxonomic composition between aquaria. In the first experiment genetic changes were quantified using metagenomic shotgun sequencing, and then verified for *tet(A)* using qPCR, but in the two follow-up experiments only qPCR quantification of the two most sensitive resistance genes (*tet(A)* and *tet(G)*) was performed. The most sensitive endpoint for resistance selection by tetracycline in complex bacterial communities was shown to be enrichment of *tet* resistance genes, which responded to 1 µg/L – the lowest concentration tested in the follow-up experiments (Table 4). Thus the MSC for resistance selection of tetracycline was determined to be 1 µg/L or less in complex communities.

The experiments of Paper III validate the prediction of Paper II that the MSC for resistance selection of tetracycline would be around 1 µg/L. They also indicate that

the MSCs for some endpoints may be even lower than predicted in Paper II. However, the work involved in establishing the MSC of one single antibiotic in one single type of complex community highlights the importance of the theoretically established MSCs from Paper II. Verifying the MSCs for all 111 antibiotics in that study would require years of work, or the collective efforts by a large number of research groups. Since scientific data underpinning regulatory documents is needed immediately, the value of the theoretical MSCs in complex communities is high even if some antibiotics would turn out to have MSCs substantially above, or below, those predicted in Paper II.

The many forms of minimal selective concentrations

It is firmly established that antibiotics concentrations below those completely inhibiting bacterial growth can select for resistant bacteria (Gullberg *et al.* 2011; Hughes & Andersson 2012). It is, however, still an open question how low concentrations of different antibiotics need to be to give resistant strains a fitness advantage. Andersson and Hughes (2014) define the MSC as "the lowest concentration of an antibiotic that results in the selection of a resistant mutant in a population over an isogenic susceptible strain." In an assay involving a resistance plasmid, Gullberg *et al.* (2014) instead use the analogous definition that the MSC "is the drug concentration where the fitness cost of the resistance plasmid is balanced by the selective effect of the added drug." The latter definition is slightly more flexible, as it extends the MSC concept to alternative endpoints. In this thesis, the MSC is defined as the lowest drug concentration that promotes enrichment of resistance genes or resistant bacteria in a microbial community. Thus the definition used here is a community MSC rather than a species or strain specific one. Furthermore, as will be discussed in this section, the MSC may represent several different things depending on which endpoint that is studied.

Different endpoints for selective concentrations

A substance that completely inhibits the growth (or kills) certain bacteria will by consequence have a clear selective effect, by providing a dramatic fitness advantage to bacteria that are still able to grow. However, even though complete growth inhibition could be a relevant endpoint on the community level, measuring inhibitory concentrations is likely to be blunt to less dramatic fitness advantages conferred by resistance genes and resistance mutations under a low level of toxicant selection pressure (Baquero *et al.* 1998). Studies comparing the fitness of two bacterial strains – identical in all respects except for the resistance factor – competing with each other under different antibiotic concentrations (Liu *et al.* 2011a; Gullberg *et al.* 2011) allow for very precise determination of the relative fitness cost of the specific resistance factor, and which concentrations that corresponds to a fitness advantage of the resistant strain. However, the ecological relevance of this simplistic testing system is still rather limited, although efforts have been made to take similar competition systems out of the lab and test them *in situ* (Andersson & Hughes 2014).

The question then arises; what would be a relevant endpoint to measure selection by sub-inhibitory levels of antibiotics in complex ecosystems? A number of possible options exist (many of which have been explored in Paper III). A simplistic measure would be to count the number of isolates from a community exposed to low concentrations of an antibiotic that survive treatment with high concentrations of the same compound. This number could then be compared to non-exposed communities (serving as the "background" of resistant bacteria), and a significant increase of the proportion of resistant isolates would indicate that selection has taken place. This way, the MSC_{CFU} of tetracycline was shown to be 10 µg/L in Paper III. However, this method is limited to cultivable bacteria and does therefore not consider the majority of the community exposed. A more direct endpoint, that is not dependent on cultivability, would be the abundances of functional resistance genes. One could then investigate whether the abundances of relevant resistance factors are increased at different concentrations of an antibiotic, compared to non-exposed controls. Increase of mobile resistance genes is a directly selectable trait related to resistance development. Importantly, increases in mobile resistance genes represent a form of increased risks to human health, as higher abundances of these genes means that there are more potential donors of resistance genes to pathogens. Furthermore, measuring gene abundances enables assignment of MSCs to individual resistance genes. For example, the results of Paper III determines both the $MSC_{tet(A)}$ and the $MSC_{tet(G)}$ to be below 1 µg/L tetracycline, while e.g. the $MSC_{tet(E)}$ was around 10 µg/L.

There are also alternative endpoints to measure the selective effects of low antibiotic concentrations on microbial communities, not directly studying phenotypic or genotypic resistance. For example, a concentration of an antibiotic that has selective effects on the community level would be expected to change the taxonomic composition of the community. As shown in Paper III, this endpoint is actually fairly sensitive to changes, with the MSC_{genus} being estimated to be around 1-10 µg/L tetracycline. However, while the composition of bacterial taxa changed, the diversity and richness of genera were largely unaffected, suggesting that resistant bacteria, previously outcompeted, take over the niches of susceptible ones. Nevertheless, it is important to note that while the MSC_{genus} corresponds to a selective effect of an antibiotic, this endpoint may be less related to enrichment of resistance factors, as intrinsically resistant bacteria are as likely to be enriched as those with horizontally acquired resistance genes. Other endpoints that could be explored would be the lowest concentrations that induce changes in gene expression (which could be measured by e.g. metatranscriptomics), induce horizontal gene transfer (studied by e.g. Jutkina *et al.* 2016), cause increased dispersal (for example from a biofilm), or cause a shift in abundance of other genes without resistance function in a microbial community (another suitable target for metagenomics). Notably, in some of these cases, it is unclear if a concentration causing an effect is actually selective, or if it just triggers more random stress responses. Thus, it may be more relevant to refer to minimal concentrations causing effects (MCEs) rather than selective concentrations. Finally, it would also be important to establish the minimal concentrations that can

Table 5. Endpoints for minimal selective concentrations (MSCs) and minimal concentrations causing effects (MCEs) in complex microbial communities

Endpoint	Suggested abbreviation	Measured by
Fitness advantage of competing strains	$MSC_{competition}$	Competition experiments (Gullberg *et al.* 2011)
CFU count	MSC_{CFU}	Plating of isolates (Paper III)
Increased (resistance) gene abundances	MSC_{gene}	Metagenomics or qPCR (Paper III)
Changes of taxonomic composition	MSC_{genus} or $MSC_{species}$	Metagenomics or amplicon sequencing (Paper III)
Increased rate of horizontal gene transfer	MCE_{HGT}	HGT assays (Jutkina *et al.* 2016)
Increased dispersal rate	$MCE_{dispersal}$	Biofilm dispersal assays (Jackson *et al.* 2002; Barraud *et al.* 2006)
Changes of gene expression	$MCE_{expression}$	Metatranscriptomics

co-select for antibiotic resistance. For example, biocides and metals have been suggested to drive antibiotic resistance development (Baker-Austin *et al.* 2006; Pal *et al.* 2014; Wales & Davies 2015), but the concentrations at which they do so are largely unexplored territory. Before determining co-selective concentrations for biocides and metals, it would be valuable to determine which ones that actually have the potential to co-select for antibiotic resistance, as some studies suggest that the potential for co-selection between biocide, metal and antibiotic resistance genes is limited (Pal *et al.* 2015).

The relevance of different endpoints for selective and effect concentrations

Experiments with competing strains have highlighted the importance of the sub-MIC selection landscape in antibiotic resistance evolution (Liu *et al.* 2011a; Gullberg *et al.* 2011; Andersson & Hughes 2012; Gullberg *et al.* 2014). However, it is still uncertain to what extent competition experiments are translatable into selective advantages in complex environmental microbial communities, as the studied system disregards, for example, competition for available niches, nutrient limitations and predation. In addition, the $MSC_{competition}$ is strain specific and can thus not be readily extended to a community with many different taxa. That said, competition experiments between resistant and non-resistant pathogen strains have a strong relevance for investigating the relative fitness of those resistant strains in e.g. the human gut, where we would like to avoid selection for resistant mutants. Such relative fitness advantages are much harder, if at all possible, to capture using a community approach. Recommended treatment regimens with antibiotics aim to reduce the timeframe during which the human microbiome is exposed to sub-inhibitory selection pressures, but avoiding sub-

lethal concentrations in all body compartments entirely is likely impossible. That said, the most pertinent setting to establish MSCs and MCEs in would probably be those where sub-inhibitory concentrations of antibiotics are expected to be present during longer times. This points of course to the use of low amounts of antibiotics as growth promoters in animal farming, but also to the external environment and the complex microbial communities present there. The most straightforward way of establishing MSCs for those communities would be to setup microcosm experiments subjected to a gradient of antibiotics concentrations and investigate the exposed communities for one or several of the endpoints discussed earlier (Table 5). Particularly, settings such as agriculture, sewage and sewage treatment plants, and environments polluted with pharmaceutical waste would be of high relevance to study, since this is where we expect antibiotic concentrations to potentially be high enough to exert a selection pressure. These environments comprise a vast range of ecosystem types, and thus differently composed microbial communities. Thus, the resulting MSCs could be different even if the same endpoint is used, simply because the communities are so disparate. This points to a general concern with MSCs – the concentrations they suggest will always be context dependent, varying with respect to both the chosen endpoint and the studied community or strain.

Environmental antibiotic resistance

Environments that could promote resistance development and dissemination

Given the selective concentrations established in Papers II and III, we next investigated environments that could potentially be exposed to sufficient concentrations of antibiotics to promote resistance development and/or dissemination of resistance genes. Two scenarios for potential selection were investigated: low level exposure in Swedish sewage treatment plants (STPs; Paper IV) and exposure to high concentrations of antibiotics in areas in India subjected to pollution with pharmaceutical waste (Larsson *et al.* 2007; Fick *et al.* 2009; Paper V).

In the three STPs investigated in Paper IV, we found that the concentrations of tetracycline and ciprofloxacin detected in the incoming untreated sewage could be selective for resistance (based on the results from Papers II and III). The concentrations of all other antibiotics examined were below predicted MSCs. However, we did not detect any selection for the corresponding resistance genes, nor for any other particular type of resistance genes throughout the treatment process (except for a non-significant increase of sulfonamide resistance gene abundances; Figure 11). Thus, even if resistance genes towards fluoroquinolones and tetracyclines would be selected for in raw sewage, they seem to be reduced to the same degree as other resistance gene classes in the subsequent treatment steps.

We also assessed whether antibiotic resistance could be co-selected for by other antibiotics, antibacterial biocides, or metals. To investigate this, we assembled the metagenomic data and annotated the contigs for the presence of resistance genes. In total, 776 out of 1,722,659 assembled contigs (0.045%) carried resistance genes (583 carrying antibiotic and 216 carrying biocide and metal resistance genes). Only 122 contigs carried more than one resistance gene and thus showed co-selection potential. The genes that most commonly co-occurred with other resistance genes were the sulfonamide resistance gene *sul1*, the *qacEdelta* resistance gene providing low-level resistance to quaternary ammonium compounds, and the aminoglycoside resistance gene *ant(3")-Ia*. Often, these genes were co-located with the *intI1* class I integrase gene. In addition, the genes constituting the *mer* operon, conferring mercury resistance, were frequently encountered together on contigs from several samples. The comparably limited number of co-occurrences, together with the overall limited evidence for selection of resistance genes (Figure 11), suggests that co-selection of resistance genes of different types is limited in STPs. However, the assembly approach is generally unable to identify genes situated far from each other on the same mobile genetic element.

We also identified a few resistance genes that were significantly enriched through the treatment process. For example, the carbapenemase gene OXA-48 was enriched in surplus and digested sludge, indicating that STPs may select for clinically relevant antibiotic resistance genes, but likely not through a direct selection pressure by antibiotics. Rather, STPs probably select for particular types of bacteria that happen

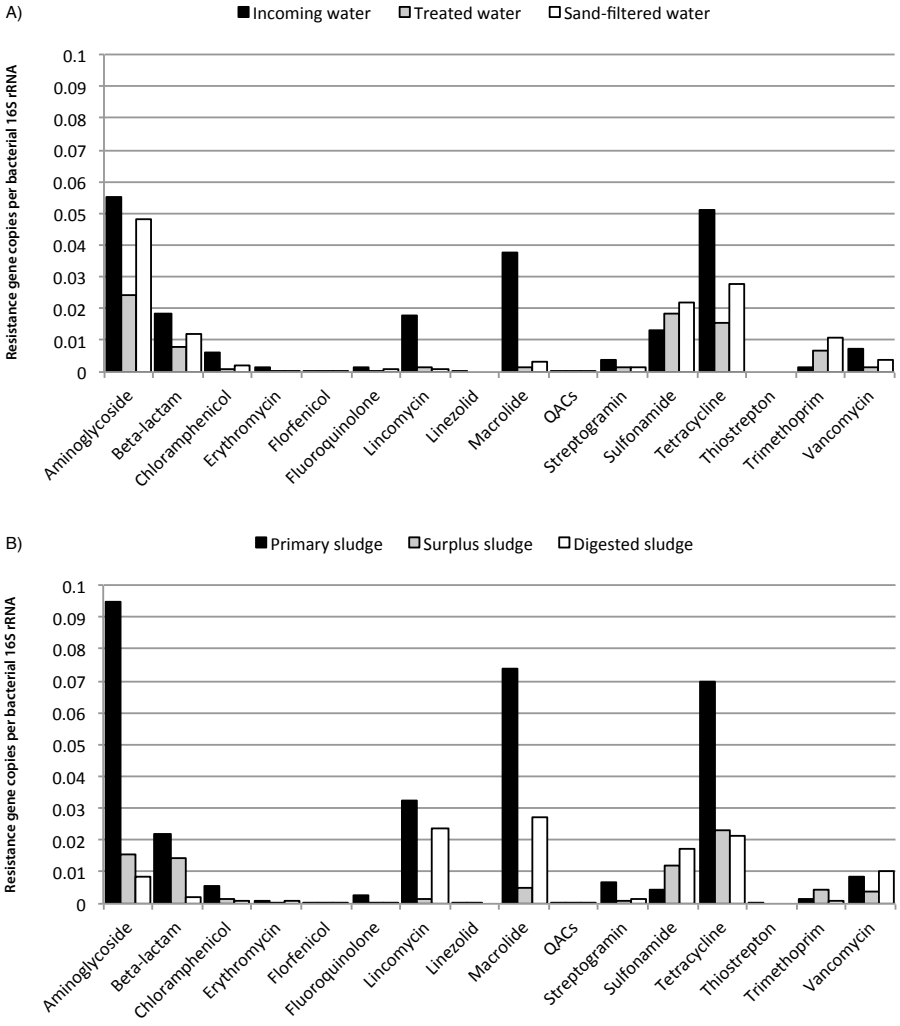


Figure 11. Abundance of antibiotic resistance gene classes in the different treatment steps of three STPs. (A) Resistance gene abundances per bacterial 16S rRNA in incoming, treated and sand-filtered water. (B) Resistance gene abundances per bacterial 16S rRNA in sludge. Figure reproduced from the Paper IV supplement.

to carry the OXA-48 gene, perhaps chromosomally. Nevertheless, this might be important since a rise of OXA-48 carriers increases the number of potential donors of this resistance gene and thus the likelihood for resistance transfer to other bacteria.

In contrast to this low concentration scenario, we have also investigated environments in India subjected to pollution from pharmaceutical manufacturing, In these environments, concentrations of antibiotics can be substantially above the predicted MSCs, and sometimes even reach and exceed therapeutic concentrations (Larsson *et al.* 2007; Fick *et al.* 2009; Kristiansson *et al.* 2011). In this thesis, we have

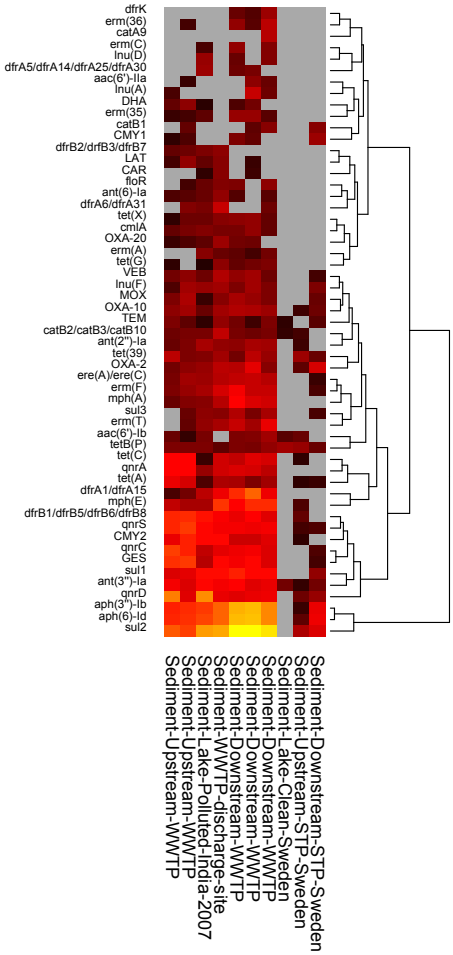


Figure 12. Heatmap of resistance gene abundances in sediment samples from Sweden (three rightmost samples), around a WWTP receiving pharmaceutical production wastewater in India and the Kazipally lake subjected to dumping of pharmaceutical production waste (leftmost samples). Grey corresponds to undetected resistance genes, and abundance is indicated by a scale from black (<10^{-4} copies per bacterial 16S rRNA) through red (around 0.1 copies per 16S rRNA) to yellow (>10 copies per 16S rRNA). Only resistance genes with an abundance of at least 0.005 copies per 16S rRNA in at least one sample are shown. Hierarchical clustering was done on log-transformed abundances using Euclidean distances.

explored the resistome of the Kazipally lake in India and contrasted it to the resistome of a Swedish lake (Nydalasjön near Umeå; Paper V). In addition, we also have analyzed data from a river receiving effluent from a treatment plant handling wastewater from pharmaceutical industries in the Hyderabad area (Kristiansson *et al.* 2011) using Illumina sequencing (unpublished data). Remarkably, these analyses show a vast diversity of resistance genes, not only against the antibiotics detected in the river and lake water (Figure 12). Particularly, the aminoglycoside resistance genes *aph(3")-Ib* and *aph(6)-Id* as well as the sulfonamide resistance gene *sul2* were greatly enriched downstream the WWTP. The dramatically high abundances of these particular genes is most a reflection of that the DNA extraction protocol for the samples included a DNA amplification procedure which introduces bias towards certain regions of DNA (Pinard *et al.* 2006), particularly small plasmids, which would include those carrying the *sul2* and *qnrD* genes (Dr. Nachiket Marathe, personal communication). It is also interesting to note that the diversity of resistance genes was high also upstream from the WWTP, suggesting that pollution occurs also from other sources. Indeed there are reports of illegal dumping of pharmaceutical waste in the river (Greenpeace 2004; Boralkar *et al.* 2005). The situation was similar in the Kazipally lake, which was more thoroughly studied in Paper V. In this lake, the *sul2* gene and the fluoroquinolone resistance gene *qnrD* were present in particularly high abundances. In addition, genes involved in

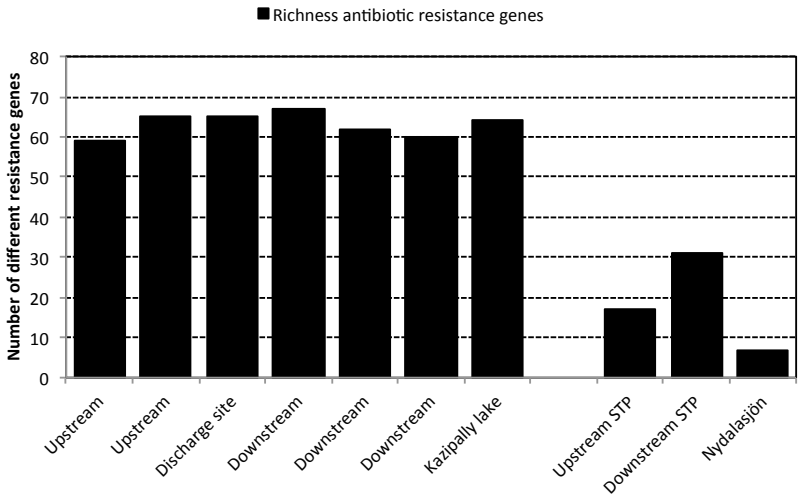


Figure 13. Number of resistance genes with abundance larger than 10^{-3} per bacterial 16S rRNA in Indian and Swedish sediment samples.

horizontal gene transfer of resistance genes, such as integrases, transposons and plasmid conjugation systems, were vastly more abundant and had higher diversity in the Indian lake compared to the Swedish non-polluted lake (Paper V: Figure 4). Assembly of the metagenomic reads from the Indian lake revealed several contigs containing resistance genes, as well as 26 novel putative plasmids. Similar results have been obtained previously from pyrosequencing data of the river sediments near the WWTP (Kristiansson *et al.* 2011).

Community effects of chronic exposure to high levels of antibiotics

As described above, high concentrations of antibiotics exert selection not only for genes conferring resistance to the antibiotics detected, but to a wide array of other resistance genes and mobile genetic elements (Figure 12; Paper V). This warrants further examination of the effects of chronic high-level antibiotics exposure. Paper V shows that in addition to resistance genes against many different classes of antibiotics, genes involved in processes such as genetic transfer, plasmid maintenance, metabolism of macromolecules and viral reproduction were strongly enriched in the polluted Kazipally lake compared to a Swedish lake. However, despite that the overall taxonomic diversity was lower in the polluted lake than in the non-polluted one, this difference was small. This suggests that a multitude of species have acquired antibiotic resistance or are intrinsically resistant, and therefore able to survive and reproduce despite the strong selection pressure from antibiotics (and potentially other chemicals). Possibly, this could be a case of evolutionary rescue, in which the bacterial populations of the lake must either have been inhabited by a fraction of resistant variants already before exposure to antibiotics, or produced such strains rapidly – by mutations, horizontal gene transfer, or both – to recover and avoid extinction

(Gonzalez *et al.* 2013). Importantly, these findings are similar to what was found in contaminated river sediments from the same region, which also show a large diversity of resistance genes (Figure 13), suggesting that this is not an isolated process unique to the lake.

Dissemination of resistance genes through sewage treatment plants

As described earlier, we did not find strong evidence for selection of resistance genes in STPs by antibiotics, nor co-selection by biocides or metals (Paper IV). However, STPs may also play an important role in dissemination of resistance genes and resistant bacteria from the human population to the environment, enabling reinfection with resistant pathogens. To examine this, we quantified the reduction of resistance genes in STPs, and set that in relation to changes of taxonomic composition, from feces to STP effluent and digested sludge. We found that although resistance genes were reduced more than 50 times in effluent compared to influent in terms of volume, their relative abundances per bacterial 16S rRNA were only reduced by 63%. In sludge, the reduction per 16S rRNA was larger, but the reduction per volume was less than 70%. Similar patterns were noted for biocide and metal resistance genes, chromosomal as well as plasmid-borne. The richness of resistance genes in was reduced by around 50% in the effluent, and around 30% in digested sludge. Taken together, this shows that a large quantity of resistance genes are released from STPs into the environment, and that STPs are not efficiently removing resistant bacteria. At the same time, the STPs effectively removed most human-associated bacteria, both from effluent water and sludge. In addition, there was a large discrepancy between the microbial communities in feces and the incoming sewage, due to a shift from obligate anaerobic bacteria to facultative anaerobes (Paper IV: Figure 1). This suggests that the resistance genes in STPs to a large extent are carried by non-fecal bacteria, but their distribution remains to be described. As the abundances of mobile genetic elements were not significantly reduced in the effluent, this presents an opportunity for resistance genes to not only diffuse through the STPs, but also to spread between bacteria, for example if presented with a sufficient selection pressure. Thus, selection *per* se STPs may be limited in STPs, but they can still provide as a dispersal route for resistance genes into the environment.

The role of travel in disseminating resistance genes across the globe

The human microbiome carries a range of both mobile and chromosomal antibiotic resistance genes, including genes not yet encountered in pathogens (Sommer et al. 2010; Forslund et al. 2013; Hu et al. 2013). Since the resistance situation in the world varies, with resistance rates being highest primarily in eastern Asia (Bebell & Muiru 2014), humans traveling around the world can serve as a dissemination route for resistant bacteria between countries and continents. It is already well established that travelers are more prone to carry ESBL-producing enterobacteria when returning from countries with a worse resistance situation (Tängdén et al. 2010; Östholm-Balkhed et al. 2013; Angelin et al. 2015), however the diversity of resistance genes

brought back has remained unknown. To investigate travel as a general dissemination route for resistance genes, we used shotgun metagenomic sequencing of fecal samples taken before and after travel in a cohort of healthcare students traveling from Sweden to the Indian peninsula or central Africa and back again (Paper VI). We found that the overall resistance gene abundance increased significantly after travel, and that resistance genes towards several classes of antibiotics were enriched (Paper VI: Figure 1). However, resistance genes did not increase significantly across all classes of antibiotics, and the most common resistance genes did not increase in abundance after travel. Although this may partially be due to a limited number of studied individuals (35 persons), it indicates that there exists a stable "core resistome" of the human gut that is not much altered by travel, and a variable part of the resistome that can change depending on the environmental conditions individuals are exposed to. Importantly, the changes related to travel occurred in the absence of antibiotics intake (subjects that took antibiotics during or six months prior to their trip were excluded), which has been shown to be an important factor in colonization with ESBL-producers while abroad (Kantele et al. 2015). While we could not identify any significant differences between the resistance gene profiles of travelers to central Africa and the Indian peninsula, only travelers to India acquired ESBL positive strains detectable by culturing. Interestingly, all those isolates were positive for the CTX-M-15 beta-lactamase, but we could not detect corresponding increases of the CTX-M gene (or any other beta-lactamases) in the metagenomes of the same individuals. This suggests that even when shotgun metagenomics utilizes a large sequencing depth, it may be too shallow to detect clinically important resistant bacteria in the human gut. Conversely, culturing for ESBL-producing bacteria was a poor indicator of overall resistance gene abundance and diversity, suggesting that the two techniques complement rather than replace each other. Finally, not only resistance genes, but also integrases and ISCR transposases were enriched after travel. Also, bacteria belonging to the Proteobacteria phylum increased in relative abundance after travel. Many pathogenic species belong to the Proteobacteria, and although not significant after correction for multiple testing there was a tendency for the *Escherichia* genus to increase after travel. However, these taxonomic changes and the increases of resistance genes were not correlated, so changes in taxonomy cannot be the sole driver of resistance gene changes. The findings of Paper VI support that travel facilitates dissemination of a range of resistance genes once they have made it into the human microbiome.

Where is the abundance and diversity of resistance genes largest?

Combining the results of the four studies in this thesis utilizing shotgun metagenomic sequencing makes it is possible to obtain a rough picture of how the resistomes of different environments relate to each other. The most striking feature of such a comparison is how the Indian lake subjected to pharmaceutical pollution stands out in terms of resistance gene abundances (Figure 14). This is almost certainly partly due to that the DNA from the Indian lake was obtained using a DNA "random"

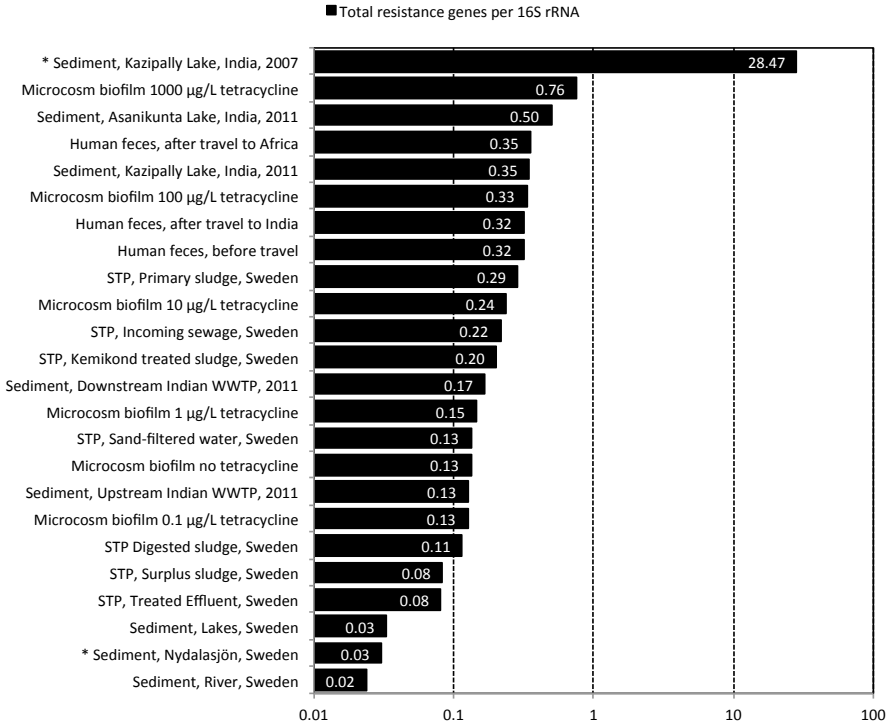


Figure 14. Resistance gene abundances in the metagenomes investigated in this thesis, measured as total resistance gene copies per bacterial 16S rRNA. DNA from samples with an asterisk were extracted using an amplification protocol that biases results and are thus *not* directly comparable to the other samples in terms of abundances.

amplification kit. Thus, the actual abundances of resistance genes in the Indian lake are likely to be substantially lower than those indicated, and should strictly only be compared to other amplified samples, such as the Swedish lake Nydalasjön investigated in Paper V. Aside from the polluted river and lake sediments, the human gut microbiome carries the largest numbers of known resistance genes. The relative abundance of resistance genes in the human gut is even 45% higher than that in incoming sewage to Swedish treatment plants. This comparison between studies also contextualizes the findings of the tetracycline aquarium experiment, in which the controls formed biofilms containing approximately the same resistance gene abundances as digested sludge or sand-filtered effluent from Swedish STPs. Adding 10 µg/L tetracycline almost doubles the abundance of resistance genes compared to controls, bringing it to the same range as incoming sewage. This can be compared to 1000 µg/L aquaria, which host more than twice the resistance genes than found in human feces.

Investigating the richness of resistance genes in different environments (Figure 15) tells a slightly different story, however. The samples taken downstream from Indian WWTP receiving wastewater from pharmaceutical industries stands out with the

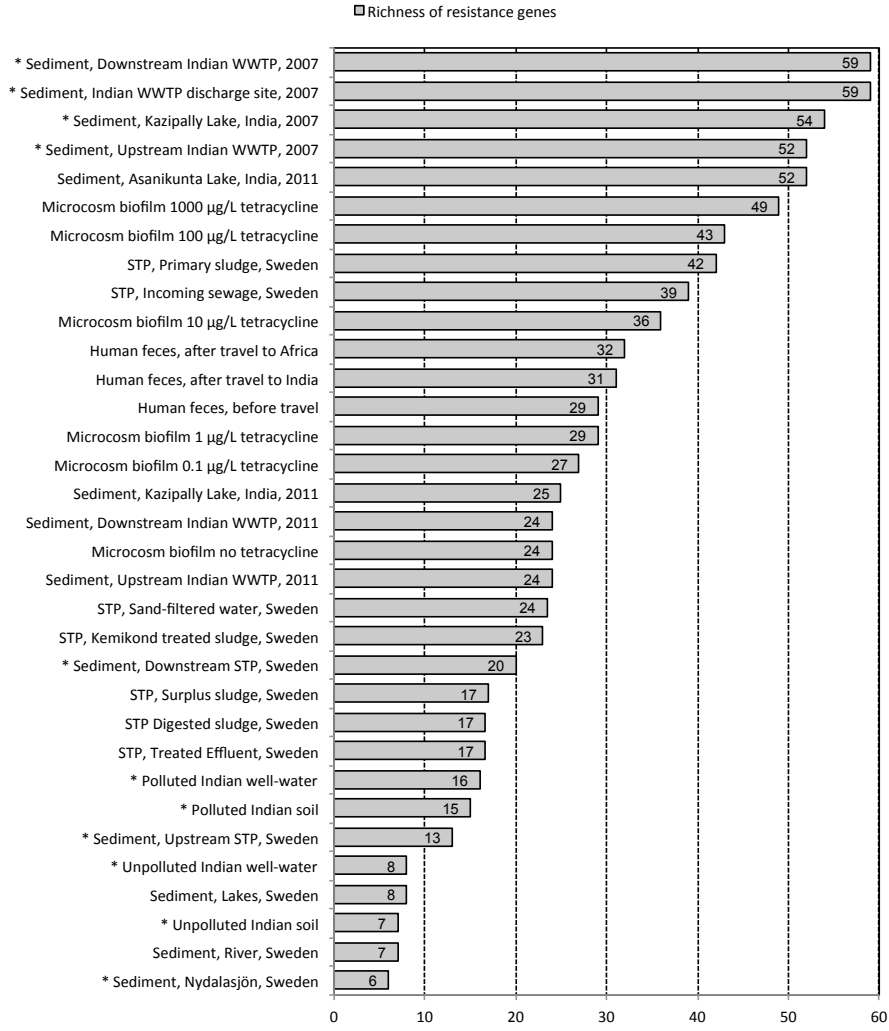
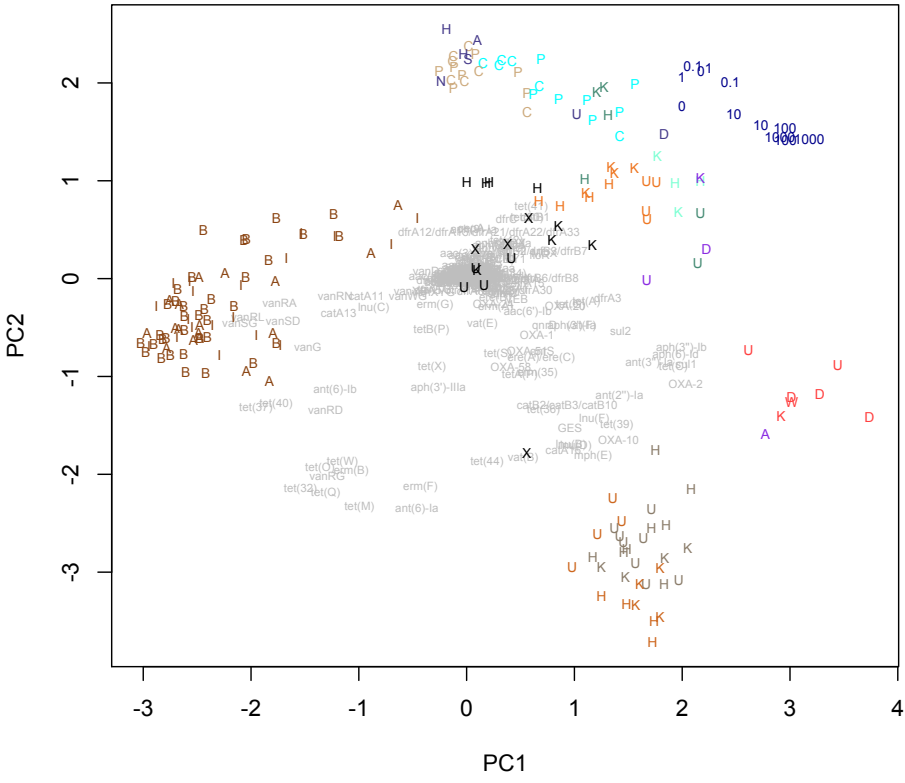


Figure 15. Resistance gene richness in the metagenomes investigated in this thesis. To account for the large differences in sequencing depth, richness was estimated as the number of resistance genes with abundance above 10^{-3} per bacterial 16S rRNA in each metagenome. Thus the actual number of resistance genes in each metagenome is almost guaranteed to be larger, and the numbers shown here should be considered proxies (comparable between samples) for the true numbers. DNA from samples with an asterisk were extracted using an amplification protocol that can bias results and differences in resistance gene richness of those compared to the other samples should be interpreted with some caution.

largest resistance gene richness. Primary sludge and incoming sewage water from the Swedish STPs follows next, together with the biofilms exposed to the highest concentrations of tetracycline. Perhaps surprisingly, incoming sewage contains more different types of resistance genes than human feces, despite that feces contains larger



Human feces:	**B** = before travel; **I** = after travel to India; **A** = after travel to Africa
Incoming sewage:	**H** = Henriksdal; **K** = Käppala; **U** = Uppsala
Primary sludge:	**H** = Henriksdal; **K** = Käppala; **U** = Uppsala
Surplus sludge:	**H** = Henriksdal; **K** = Käppala; **U** = Uppsala
Digested sludge:	**H** = Henriksdal; **K** = Käppala; **U** = Uppsala; **X** = Kemikond treated (Käppala)
Treated effluent:	**H** = Henriksdal; **K** = Käppala; **U** = Uppsala
Sand-filtered effluent:	**H** = Henriksdal; **K** = Käppala
Tetracycline aquariums:	**0** (numbers correspond to tetracycline concentration in µg/L)
Polluted Indian sediments:	**U** = Upstream WWTP; **W** = WWTP discharge site; **D** = Downstream WWTP; **K** = Kazipally lake (2007)
2011 Indian sediments:	**U** = Upstream WWTP; **D** = Downstream WWTP; **K** = Kazipally lake (2011); **A** = Asanikunta lake
Swedish sediments:	**U** = Upstream STP; **D** = Downstream STP; **N** = Nydalasjön; **H** = Härlanda Tjärn; **S** = Stora ån; **A** = Axelsmosse
Indian soil:	**P** = Polluted; **C** = Clean
Indian well water:	**P** = Polluted; **C** = Clean

Figure 16. Principle component analysis (PCA) of resistance gene profiles (presence/absence) in the environments investigated in this thesis. The resistance genes separating the samples are indicated in gray.

abundances of resistance genes. The lowest richness was found in Swedish lake and river sediments and unpolluted soil – the same samples that also hosted the smallest abundances of resistance genes. Here, less than ten resistance genes had higher abundances than 10^{-3} per bacterial 16S rRNA. Notably, tetracycline selection in aquaria gave rise to biofilms carrying a richness of resistance genes close to that of the polluted environments.

Finally, we may also look at how the composition of resistance genes differs between samples from different environments (Figure 16). Here, it is evident that two environmental types stand out in terms of resistance gene composition: polluted sediments and human feces. Notably, the incoming sewage and primary sludge samples are not very similar to human feces, for reasons discussed earlier. Treated STP effluent is reminiscent of the resistomes from Indian soil and well-water, and also that of sediments from Swedish lakes. It is interesting to note that the lowest doses in the tetracycline aquarium experiment correspond well to treated STP effluent and Indian well water in terms of resistance gene content, but with higher tetracycline exposure the resistance profiles become skewed towards the "polluted side" of the figure. Resistance gene composition is still very dissimilar from that of polluted sediments even in the highest tetracycline concentration, but the tendency of a movement in this direction points to that the same set of genes could be involved in resistome changes under antibiotic selection.

An ecological framework for antibiotic resistance

Conventionally, the struggle against antibiotic resistance development has mainly taken place in the clinical and community settings – aiming at preventing selection for resistant bacteria during antibiotics treatment – and in agriculture, restricting use in animals. Recently, the role of the environment as an important piece in the resistance puzzle has been increasingly recognized (Martinez 2008; Wright 2010; Pruden *et al.* 2013; Ashbolt *et al.* 2013; Finley *et al.* 2013; Bondarczuk *et al.* 2015). However, the understanding of the environment as a source and dissemination route for resistance genes and resistant bacteria is still limited. The lack of knowledge of how, and under which circumstances, the environment facilitates resistance development makes mitigating the emergence and dissemination of mobile resistance factors problematic (Berendonk *et al.* 2015). To disentangle the different roles of the environment in these processes, we ultimately need to build models for how resistance emerge and is disseminated. Such models will by necessity be descriptive at first, as most of their parameters remain unknown, but they regardlessly have value as indicators of the most urgent knowledge gaps to fill in order to develop mitigation strategies. Such one-health approaches will be instrumental to succeed in the uphill battle against antibiotic resistance (Collignon 2013; So *et al.* 2015; Collignon 2015). This section will attempt to formalize a framework for environmental antibiotic resistance, set in an ecological context.

The emergence of mobile resistance factors

As discussed earlier, novel antibiotic resistance factors could emerge anywhere, at any time. The astounding number of bacterial cells on Earth (around 10^{30} – a thousand billion billion billions; Kallmeyer *et al.* 2012), means that essentially anything that can happen in the bacterial world, *will* happen at some point. Thus, emergence of new resistance factors is likely to occur continuously. However, there are two reasons that we are not flooded by novel resistance genes. First, most resistance factors that have just recently become mobilized likely have a fairly high fitness cost associated with them. Thus, they would be selected *against* unless there is a strong selection pressure to maintain them. Second, even if such a resistance factor would have a low or negligible fitness cost, it is unlikely to become fixated in the bacterial population unless there is a selection pressure to maintain it (Martinez 2011). This selection pressure may be weak, but unless it is present the only manner a novel resistance factor would be retained is through genetic drift (Baquero *et al.* 1998).

If we assume that a novel resistance factor could be mobilized anywhere but needs a selection pressure to be retained, the subsequent question becomes: where are selection pressures strong enough to promote maintenance of mobile resistance genes? Considering that most novel resistance factors likely have high costs, environments allowing sustained longevity of a resistance gene regardless of cost would be of particular importance, since it is reasonable to assume that once a resistance gene gains a foothold in a bacterial community, it will rapidly evolve

towards diminished fitness cost (Salyers & Amábile-Cuevas 1997). This highlights the risks associated with situations in which antibiotics concentrations clearly above the MSCs, or even the MICs, as observed in environments polluted with waste from pharmaceutical pollution (Larsson 2014b). In addition, the conditions bacteria face within antibiotic production plants are largely unexplored, and are also likely to be extensively selective, although the number of bacteria present in such settings may be very small. Other settings where exposure to antibiotics is high is in the human and animal gut during treatment.

It is of course possible that resistance genes with considerably lower fitness costs may emerge on mobile genetic elements, and that sub-inhibitory concentrations of an antibiotic will suffice to select for their maintenance. Thus, attention also has to be pointed to raw sewage, agriculture and sewage treatment plants, where concentrations of antibiotics around the predicted MSCs have been determined (Michael *et al.* 2013; Paper II). Finally, one cannot neglect that novel resistance determinants may be selected for naturally, if they e.g. confer a competitive advantage against antibiotics producers, or allow host bacteria to survive higher concentrations of an antibiotic that they themselves produce.

Horizontal gene transfer of resistance factors

Horizontal gene transfer is central for the spread of novel resistance genes as it allows resistance determinants to extend their prevalence beyond a particular clone. This way, gene transfer makes resistance genes available a much larger part of the bacterial population in a particular environment, often beyond species boundaries (Martinez 2011). As for the mobilization of novel resistance factors, transfer of genes between bacteria can in theory occur anywhere. However, for resistance genes to be horizontally transferred to pathogenic bacteria, they need to, at least temporarily, share the same habitat (Matte-Tailliez *et al.* 2002; Wiedenbeck & Cohan 2011). Furthermore, horizontal gene transfer is much more likely to occur between phylogenetically closely related bacteria (Philippot *et al.* 2010; Smillie *et al.* 2011). Finally, transfer processes are induced by stressors such as antibiotics (Beaber *et al.* 2004; Hastings *et al.* 2004; Maiques *et al.* 2006), and antibiotic selection contributes to fixation of transferred resistance genes in their new host. Thus, resistance transfer can be expected to be relatively frequent between human-associated bacteria (Salyers *et al.* 2004), particularly during treatment with antibiotics. This means that once a resistance factor has entered into a human pathogen, it is more likely to further spread between pathogens, than being transferred again into another pathogen from environmental bacteria (as also argued in Paper VII). Moreover, avoiding transfer of resistance between pathogens is likely impossible, since they share habitats, often are phylogenetically related, and mobile resistance factors generally seem to be associated with low fitness costs (Salyers & Amábile-Cuevas 1997; Andersson & Hughes 2010). Somewhat surprisingly, the human microbiome harbors a fairly large number of resistance genes that have not been transferred to human pathogens (Sommer *et al.* 2009; 2010). The reasons for this are unknown, but one can speculate that strong

barriers to transfer are at play. For example, the carriers of those gene may be evolutionary divergent from most human pathogens, or their resistance genes may simply not have been mobilized onto a suitable mobile genetic element, complicating their transfer (Martinez 2011).

However, the vast majority of existing resistance factors are likely not encountered in pathogens and human commensals, but present in environmental bacteria (Allen *et al.* 2010). Bacteria not typically associated with the human microbiome may have opportunity to interact with human-associated species in various settings. One possibility is that environmental bacteria can transiently be present in the human microbiome, through e.g. interaction with wild animals, intake of exotic foods, or drinking of contaminated water (Allen *et al.* 2010; De Boeck *et al.* 2012). The impact of these exposure scenarios is uncertain as the timeframes for interaction is limited and the incentives for transferring resistance genes would in most cases be low, except during antibiotic treatment. That said, there are other settings where human bacteria can interface with animal-associated and environmental ones. A key consideration in these contexts is the length of the dispersal route from those milieus back into the human population (Baquero *et al.* 2009). A pathogen (or commensal) that acquires a novel resistance factor but is eradicated before it can return to a human host never causes any clinical resistance problems, while those that make it back to their hosts may do. An obvious setting that offers interaction opportunities for a range of different bacterial species, and also may present sufficient conditions for resistance selection, is sewage treatment plants (Rizzo *et al.* 2013). Other milieus that may serve as breeding grounds for resistance transfer can be found in agriculture (particularly among livestock; Allen 2014), water bodies (Baquero *et al.* 2008; Lupo *et al.* 2012), and food (Rolain 2013). All these environments have in common that the exposure routes to humans after a potential transfer event are relatively short. STPs generally discharge their effluent (which has been repeatedly been shown to contain resistance genes; see e.g. Paper IV) into water bodies. Humans often use this water for activities such as drinking water supply and recreational swimming. Furthermore, animals drink the water untreated and may subsequently spread resistant pathogens to humans. Global food trade has been shown to also ship pathogenic bacteria around the world, for example in the German 2011 Shiga-toxin-producing *Escherichia coli* (O104:H4) outbreak (Rasko *et al.* 2011; Buchholz *et al.* 2011). Finally, transfer of resistance factors from human pathogens to environmental bacteria is possible, enabling human-associated bacteria to use environmental bacterial populations as reservoirs for resistance genes that can later be re-recruited into the human-associated resistome (Salyers & Amábile-Cuevas 1997; Salyers & Shoemaker 2006). Although this process is nearly impossible to quantify, it is likely of lesser concern than the recruitment of novel resistance factors into pathogens or the dissemination of resistant pathogens through the environment. Furthermore, measures to prevent re-recruitment of resistance genes from the environment would be almost identical to mitigation strategies to avoid spread of novel resistance factors into pathogens.

Dissemination of resistant bacteria

The main route for human spread of resistant pathogens is from other people, either in clinics or through acquisition in the community setting. The typical dispersal routes here are through body contact or indirect contact transmission, aerosols, and food prepared by persons carrying the pathogen (Livermore 2000). These are also the typical transmission routes for infectious bacteria in general, and interventions preventing their circulation are essentially the same as those from classical epidemiology (Rao 1998; Livermore 2000; Lipsitch *et al.* 2000; Levin *et al.* 2014). Importantly, proper hygiene routines constitute the principal dispersal barrier for resistant pathogens, and the significance of sanitation for preventing spread of resistant bacteria between humans cannot be overstated (Mattner *et al.* 2012).

Apart from transmission between humans, environmental dissemination routes for resistant bacteria has also been pointed out as potentially important for the spread of antibiotic resistance (Allen *et al.* 2010; Pruden *et al.* 2013; Finley *et al.* 2013; Levin *et al.* 2014). Again, environments facilitating dissemination of resistant bacteria also enable spread of non-resistant human pathogens. Thus, sewage, wastewater treatment plants, water bodies, food trade and travel, but also air-borne aerosols, are important factors enabling bacterial transmission between hosts through the environment (Fernando *et al.* 2010; Rolain 2013; Molton *et al.* 2013; European Food Safety Authority & European Centre for Disease Prevention and Control 2013; Pruden 2014; McEachran *et al.* 2015; Angelin *et al.* 2015; Barberán *et al.* 2015; Paper VI). Limiting the spread of human-associated bacteria – resistant or not – requires an understanding of the environmental dispersal barriers that exist. Contrary to the case of clinical and community transmitted bacteria, identifying relevant barriers to dispersal is considerably harder in the environment. We may here adopt a metacommunity ecology perspective and consider the human and/or animal hosts of pathogens as habitable patches, while most other external environments would serve as a dispersal matrix (Leibold *et al.* 2004; Table 6). Metacommunity theory suggests that if patches are of equivalent quality, the distance between patches and the dispersal capability of species determine their relative success (Bengtsson 2009). Thus, the quality of the dispersal matrix and the ability to survive between hosts are fundamental properties for pathogens to spread between humans through the environment. Some understanding of how different pathogens survive in the external environment can once again be gained from epidemiology, although this is not a particularly well-studied subject outside of a few select model bacteria. An important factor in these dispersal processes may be the presence of inactive dormant stages, e.g. the highly resilient spores formed by some pathogenic bacteria (Leggett *et al.* 2012). Such dormant life-stages could vastly help the bacteria to survive in the dispersal matrix, almost regardless of matrix quality, and to re-spawn once in a suitable host (Lennon & Jones 2011; Shade *et al.* 2012).

The dispersal routes of bacteria through the environment have not evaded investigated, however. Research on microbial source tracking, usually aiming at identifying the sources and health risks associated with e.g. leaks of untreated sewage,

Table 6. Implications of different metacommunity perspectives for resistance selection

Perspective	Ecological processes	Implications for resistance selection
Species sorting	Gene content determined by local processes. Disturbances decrease local genetic diversity. Dispersal between patches is almost absent.	Antibiotic selection and fitness cost of genes determine gene content. Antibiotic exposure favors specific resistance gene types. Little input of resistance genes from external sources. In the absence of antibiotics, selection will be for resistance genes with low cost, or for loss of resistance genes. Almost independent of matrix quality.
Patch dynamics	Patches of roughly equal quality. Colonization of patches is dependent on distance and dispersal limitations. Cost, detoxification effectiveness, and dispersal ability of carriers determine the success of each resistance gene. Loss of resistance genes is stochastic.	Similar antibiotic exposure, such as in human gut when not under treatment. Human individuals are disconnected and hygiene is a dispersal limitation. Resistance genes that confer relevant resistance and are carried by bacteria that spread easily will be most successful. Selection for resistance genes with low cost and high transmission potential. Large dependence on matrix quality.
Dispersal-driven (mass effects)	Source-sink dynamics. Patches of different quality. Local gene content partially dictated by dispersal. Resistance genes that confer a fitness advantage will be successful.	Resistance genes may "spill over" from environments in which they are better adapted to poorer ones. Some environments may be exposed to antibiotics, and others not. Gene content dependent on in-flow of resistance genes, and subsequent selection for efficient variants. Resistance genes matching antibiotic exposure will be maintained. Large dependence on matrix quality.
Neutral	All resistance genes have similar fitness costs and confer similar resistance patterns. Ecological drift and mutation of resistance genes maintain diversity. Only valid on long time scales.	This could be true for e.g. multidrug efflux pumps. In the absence of selection, genetic drift will govern which resistance factors that are maintained. This may be a relevant perspective in pristine environments. Matrix quality governs differences between environments in terms of resistance gene content.

have generated some knowledge regarding the persistence and re-infection potential of human-associated bacteria in the environment (Harwood *et al.* 2014). Furthermore, it is known that physical forces, such as wind and watershed, move bacteria over large distances. Wild birds and animals in contact with human activities

are known to carry resistance genes, and may also spread them across large areas (Baquero *et al.* 2008; Allen *et al.* 2010). Still, much remains to be understood in terms of dispersal limitations, environmental survival, competitiveness versus environmental species and strains, resistance selection and alternative habitats for human-associated bacteria in the environment. Even less is known about how environmental bacteria harmless to humans carrying resistance genes disperse and interact with human-associated bacteria. The dissemination routes from environments presenting a selection pressure for initial emergence, mobilization and maintenance of resistance genes to humans and/or animals are poorly understood, but constitute important propagation routes for resistance genes into the human population. They thus need to be delineated, along with the factors influencing matrix quality for environmental bacteria. This calls for efforts to monitor the presence of pathogens and resistance genes in a variety of environmental settings to better understand possible dispersal routes. Furthermore, experimental microcosm setups would be necessary to delineate the dispersal parameters in different environmental matrices of human-associated bacteria, environmental bacteria, and individual resistance factors.

Evolutionary processes influencing environmental antibiotic resistance

For the long-term maintenance of antibiotic resistance genes in bacterial communities, two antagonistic evolutionary forces are at play: selection promoting resistance phenotypes, and selection reducing fitness cost. As discussed earlier, gain and fixation of resistance genes in a bacterial population are largely dependent on a direct antibiotic selection pressure (Martinez 2011). The selective forces towards maintenance of resistance genes do not only include direct antibiotic selection pressure, however. Even in the absence of a direct selection pressure from an antibiotic, resistance genes may be favored by co-selection by other substances presents, such as other antimicrobial agents including metals and biocides (Baker-Austin *et al.* 2006). In addition, resistance genes may be maintained because they confer advantages to the cell even in the absence of a selection pressure, in essence allowing the bacteria to perform an intrinsic function more efficiently when they carry the resistance gene (Enne *et al.* 2004). However, carriage of resistance genes usually comes with a cost in terms of reduced fitness, although this cost is sometimes small (Andersson & Hughes 2010). This cost is (together with genetic drift) the sole factor that acts to reduce the frequency of resistance genes in bacterial populations. Random losses of resistance genes happen all the time, but seldom result in complete elimination from the bacterial community, which means that once a selection pressure for resistance re-emerge, resistance development of bacterial populations previously subjected to resistance selection can be quick (Levin *et al.* 1997). Selection pressure acting against resistance is therefore crucial for eradication of resistance factors from a community.

Bacteria typically become resistant to antibiotics via i) up-regulation of efflux pumps exporting the substance from the cell, ii) expression of degradation enzymes that can render the substance harmless, iii) protection of the target of the antibiotic,

iv) alternative means to perform inhibited functions, or v) modifications to the cell wall, reducing permeability for the antibiotic substance (Walsh 2003). Resistance mechanisms associated with efflux pumps and cell wall modifications are often caused by mutations in chromosomal DNA, although some efflux pumps are transferrable between bacteria on plasmids. Degradation enzymes, target protection proteins, and enzymes allowing utilization of alternative enzymatic pathways are more likely to be transmissible on mobile genetic elements as they add functions to the host rather than modify existing ones. Thus, fitness costs associated with the latter three mechanisms are primarily associated with the cost of carrying the resistance plasmid and expressing its genes, while costs of mutations are related to decreased growth rate due to changes in essential genes and/or altered resource usage. In both cases, compensatory mechanisms, such as mutations, can reduce fitness costs over time (Andersson 2003). Under antibiotic selection pressure, evolution of a bacterial population towards mutation-mediated resistance depends on both the population size and the mutation rate (Perron *et al.* 2015a). Certain mutations have little or no fitness cost, but those have been shown to also confer lower degree of resistance than more costly mutations (Melnyk *et al.* 2015). Would the same be true for resistance genes in microbial communities? Recent meta-analysis of fitness costs associated with different types of resistance factors suggests that plasmid-mediated resistance infer a much smaller cost than mutational resistance, and that the fitness reduction by carrying a resistance plasmid is relatively small (Vogwill & MacLean 2015). Indeed there are substantial fitness costs associated with the initial uptake of horizontally transferred genes (Baltrus 2013), but both plasmids and hosts seem to compensate for those costs within a comparably small number of generations through plasmid domestication (Bouma & Lenski 1988; San Millan *et al.* 2015; Vogwill & MacLean 2015). Thus, the majority of horizontally transferred resistance genes may actually present little cost to their host. If that be the case, the advantage of losing a resistance gene would be small for the individual cell, essentially reducing the gene loss mechanism to that of stochasticity. Random losses, however, are likely not sufficient to fully eradicate resistance genes from a population (Levin *et al.* 1997).

Given that most antibiotics in use are derived from natural compounds produced by microorganisms in the environment, the presence of genes conferring resistance to those compounds across a range of habitats is not surprising (Allen *et al.* 2010). Most likely, however, resistance genes did not evolve as a means to fight the high concentrations of antibiotics used in therapy, since such high concentrations are not encountered in environments with no or little anthropogenic impact (Kümmerer 2009a; b). Many antibiotics instead seem to primarily function as pigments, toxins, and effectors in microbial communities (Demain 1998), or be involved in microbial signaling (Linares *et al.* 2006). A curious property of antibiotics is that, at low concentrations, many of them seem to escalate mutation rates and mobilize DNA (Aminov 2009; Blázquez *et al.* 2012). The exact reasons for this remain poorly understood, but it has been hypothesized that higher mutation rates enable quicker niche adaption (Aminov 2009). Thus, a signaling role for antibiotics as secondary

metabolites may be that when resources in the habitat begin to decay, they initiate generation of genetic variability that may be favorable in the search for new suitable niches and habitats. In essence, this would ensure more efficient utilization of resources. In this case, resistance genes may have evolved to balance these needs, or to protect bacteria against such signaling schemes of other species. This implies that there may be advantageous to carry resistance genes regardless of anthropogenic antibiotic selection, and that expecting a reversal of resistance after exposure has ceased may be overly naïve. The almost ubiquitous presence of resistance genes in a vast range of environments (Allen *et al.* 2009; Sommer *et al.* 2009; Lang *et al.* 2010; Martiny *et al.* 2011; D'Costa *et al.* 2011; Forsberg *et al.* 2012; Segawa *et al.* 2012; Munck *et al.* 2015) indeed suggests that this is the case, and that the cost associated with carrying resistance genes is almost negligible unless the niche is extremely resource-poor, with genome streamlining as a result (Yooseph *et al.* 2010; Giovannoni *et al.* 2014).

An ecological framework for antibiotic resistance development

The above reasoning leads to the identification of four important steps on the route to clinically important antibiotic resistance: emergence of novel resistance factors, mobilization, transfer to human pathogens, and dissemination. Notably, all these steps need not to happen in this particular order; transfer to human pathogens may occur after dissemination to the human microbiome, or certain steps in the process may repeat (Figure 17). A crucial factor for a resistance gene to reach human pathogens in that it is maintained throughout all these steps. As argued earlier, resistance genes with high costs are very unlikely to be maintained in the absence of a selection pressure, particularly if located on a mobile genetic element. Furthermore, a scenario with a constant selection pressure by antibiotics, from the environmental emergence of a resistance gene to its transfer to a human pathogen, seems improbable, although one could argue that there are places in the world where this may be possible. Taken together, it seems reasonable that successfully maintained resistance genes have either evolved towards low fitness cost in a mobile context (a sort of evolutionary rescue (Gonzalez *et al.* 2013) on the individual gene level), or were associated with low fitness costs from the beginning. Since loss of resistance genes is likely as long as they bestow their carrier with a significant fitness cost, recently mobilized genes that do not provide an obvious fitness advantage are undoubtedly sorted out early from mobile genetic elements such as plasmids (Baquero *et al.* 2013). This highlights the importance of environments in which resistance genes provide a strong selective advantage, for example milieus subjected to antibiotics pollution (Larsson 2014a; b). Since these environments would also present bacteria with conditions that favor increased mutation frequency, one consequence may be that resistance genes could be present in several slightly different variants, all selected for detoxification efficiency, of which only those with a low fitness cost survive when the selection pressure is removed (for example after dispersal of the host to a non-polluted environment). Given how long the dispersal route from initial mobilization to human pathogens

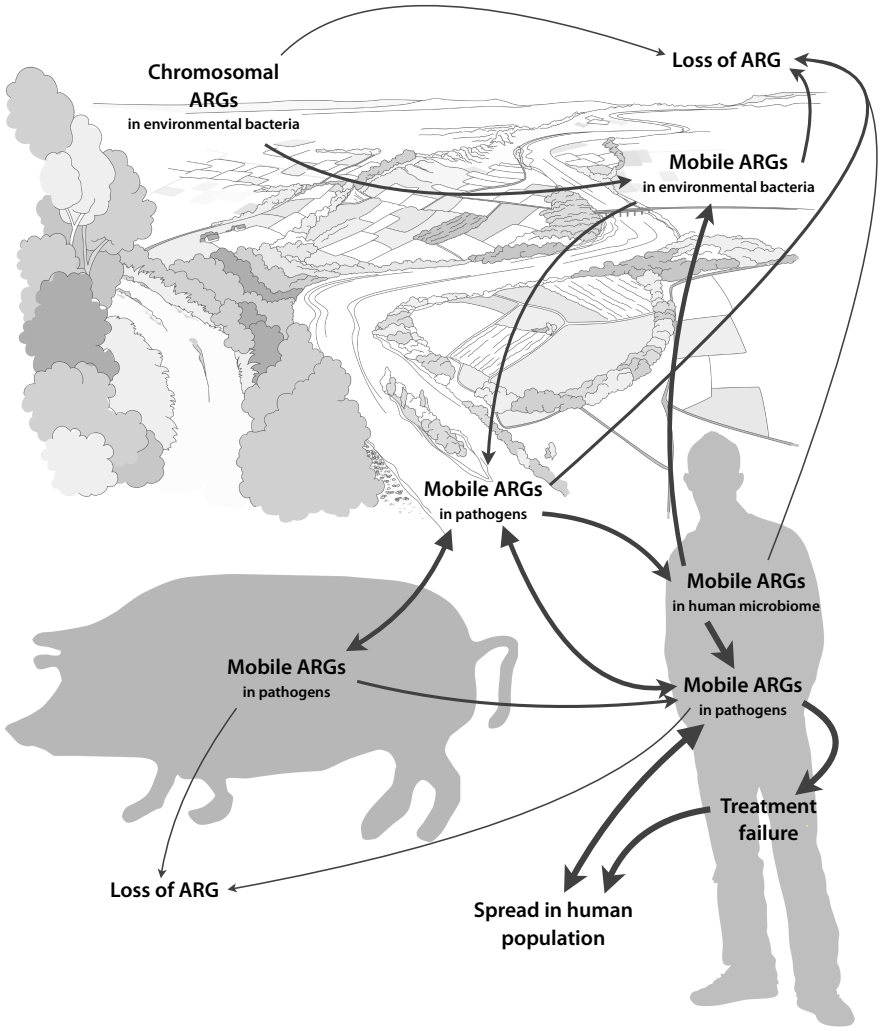


Figure 17. A framework for emergence and transfer of antibiotic resistance genes (ARGs). The width of each arrow roughly corresponds to the assumed frequency of each event. Many events are likely more frequent when antibiotic selection pressure is stronger or recurrent. Still, due to the much larger numbers of environmental bacteria than human- and animal-associated, mobilization of ARGs is probably common on a global scale.

would generally be for a novel resistance gene, it is not surprising that mobile resistance genes found in pathogens today are terribly hard to eliminate from bacterial populations (Levin *et al.* 1997; Andersson 2003; Jernberg *et al.* 2007; Löfmark *et al.* 2008) and seem to bestow little fitness cost on their carriers (Enne *et al.* 2004; Andersson & Hughes 2010; Gullberg *et al.* 2014; Vogwill & MacLean 2015). This suggests that once a resistance gene is widely spread among human pathogens (or even among human commensals), the game is lost and we are restricted to manage the spread from individual cases of infection. Mitigation of the spread of

resistance factors to human pathogens should therefore ideally take place *before* they get a foothold in the human microbiome. Thus, detection of resistance determinants in the environment that are *not yet widespread among clinical bacteria* is a primary concern in risk assessment of antibiotic resistance (Paper VII).

Which environments pose the most pertinent risks to human health?

Ultimately, the main reason to study antibiotic resistance in the environment is to gain further insights into the risks to human health. This knowledge can then be used to design interventions that could prevent or delay the recruitment of resistance factors to pathogens from environmental bacteria. To identify suitable mitigation strategies, we need to first define what environments and scenarios that constitute the most severe risks. This, however, is not completely straightforward. Some researchers have argued that the most severe risk scenarios involve "resistance genes that are already known to contribute to the failure of antibiotic treatment and have previously been reported to reside on mobile genetic elements that are hosted by human bacterial pathogens" (Martinez *et al.* 2015). This is of course the case when such genes are encountered in the human microbiome, but while they are clearly of importance, finding them in environmental bacterial communities is not necessarily indicative of a high-risk situation. Well-known resistance genes present on mobile genetic elements easily spread with human feces, and detection of them in the external environment may simply be an indication of human fecal contamination (Pruden *et al.* 2006; 2012). Risks associated with human fecal pollution should not be neglected, but is almost exclusively related to the dissemination of already resistant bacteria. Furthermore, these genes are already circulating among pathogens, and as argued earlier transfer of them between pathogens within the human microbiome is expected to be vastly more frequent than transfer of the same genes from environmental bacteria. Thus, in terms of future treatment outcomes, the clinical consequences of recruitment of resistance genes from environmental sources that are already present among pathogens are likely to be minor. We therefore think that Martínez *et al.* (2015) overestimate "the risks associated with well-known resistance genes that are already circulating among human pathogens and underappreciates the potential consequences of the transfer of previously unknown resistance determinants from the environmental resistome" (Paper VII).

The risk landscape can essentially be partitioned into three main components: 1) the risks for mobilization and fixation of novel resistance determinants, 2) the risks for recruitment of resistance genes not previously present in human pathogens through horizontal gene transfer, and 3) the risks associated with dissemination of resistant bacteria (pathogens or not) through the environment to the human population (Table 7). We have already learnt that antibiotic selection is likely a central element for all these components (Figure 17). Although mobilization of novel resistance genes could happen anywhere, stronger selection pressures are likely directly related to higher risks for their fixation in bacterial populations, as the costs for carrying recently emerged mobile resistance determinants probably are high. This identifies the human

Table 7. Human health risks associated with environmental antibiotic resistance

Risk scenario	Environments of particular concern	Possible mitigations
Emergence and fixation of novel resistance genes	Human and animal microbiome	Reduce antibiotics usage
	Intensive aquaculture	Ban the use of antibiotics
	Environments subjected to pharmaceutical pollution	Regulate releases from pharmaceutical production
Mobilization and transfer of resistance genes	Environments subjected to pharmaceutical pollution	Regulate releases from pharmaceutical production
	Sewage	Disinfection of treated sewage and sludge
Transfer of resistance genes to human pathogens	Human microbiome	Reduce antibiotics usage, avoid transmission of pathogens
	Animal microbiome	Reduce antibiotics usage
	Sewage treatment plants	Disinfection of treated sewage and sludge
Dissemination of resistant bacteria	Human-to-human contacts	Hygiene
	Hospitals	Hygiene
	Animal agriculture	Reduce antibiotics usage, avoid direct contact between animals and humans, treatment/disinfection of animal feces
	Poorly treated sewage	Implement sewage treatment in developing nations
	Water bodies	Ban the release of untreated sewage into water bodies, regulate industrial releases of chemicals

and animal microbiome during antibiotics treatment, intensive aquaculture assisted by antibiotics (Cabello 2006), as well as environments polluted with high levels of antibiotics, as particular high-risk environments in early resistance emergence. Mobilization and transfer of resistance factors is also driven by antibiotics exposure *per se* (Beaber *et al.* 2004; Hocquet *et al.* 2012), and does not seem to require high concentrations of antibiotics (Dörr *et al.* 2009; Jutkina *et al.* 2016). Thus, polluted environments once again pose a high risk, but e.g. sewage may just as well contain sufficient toxicant concentrations to promote horizontal gene transfer. For the transfer of resistance to human pathogens, the abundance of pathogenic bacteria that can act

as recipients is crucial. This means that the human microbiome is likely to be central in this process, and that human commensals may play important roles as intermediary resistance reservoirs (Sommer *et al.* 2010; Forslund *et al.* 2013). In addition, animals may also serve as intermediate hosts for resistant bacteria, and contribute a breeding ground for resistance transfer to human pathogens (Allen *et al.* 2010). The transfer of novel resistance genes to human pathogens in sewage treatment plants seems somewhat more unlikely (Paper IV), but is certainly not impossible. Finally, risk scenarios for dissemination of human-associated resistant bacteria are the same as those allowing dispersal of pathogens in general: human-to-human contact, hospital settings, animal agriculture, and poorly treated sewage (Livermore 2000; Pruden 2014). Conversely, the dissemination routes for environmental bacteria carrying resistance genes are much less clear. Regardless, the most critical factor is whether there is a quick dispersal route to the human population (Baquero *et al.* 2009). The shorter the "length" of this route, the higher the risks associated with a particular environment.

Taken together, it is not crystal-clear how high-risk settings for human health associated with environmental antibiotic resistance should be defined. However, obvious scenarios where interventions could already be applied are environments with strong selection pressures from antibiotics. Thus, limiting discharges of pharmaceutical waste from antibiotics production and reducing unnecessary use of antibiotics in humans, animals, and aquaculture are extremely important first steps towards mitigation of environmental antibiotic resistance development (Pruden *et al.* 2013; Review on Antimicrobial Resistance 2014; 2015; Bengtsson-Palme & Larsson 2016a; Paper II). Second, identifying and closing down important dispersal routes for resistant bacteria to the human microbiome is also a high priority. For dissemination, targeting critical control points for resistance spread, such as sewage treatment plants, would be of particular importance (Berendonk *et al.* 2015). For example, disinfection of treated effluent could be an efficient means of controlling the dispersal of resistant bacteria. However, building out *any kind* of modern treatment of sewage in developing countries would probably have larger effects on resistance dissemination and would thus be a strategy of even higher priority (Pruden *et al.* 2013; Kookana *et al.* 2014), as the resistance problem is a global issue (see Paper VI).

A future of resistant superbugs?

Apart from the obvious health hazards associated with increased prevalence of resistance genes among human pathogens, there are additional disturbing circumstances suggesting that the future holds an even darker resistance development than what we may currently appreciate. First of all, most resistance genes seem to bestow little fitness cost on their host, and many resistance genes are readily transferred both between bacteria and between plasmids (Normark & Normark 2002). Resistance genes with low cost tend to be maintained, and evolve in response to more efficient variants of the same antibiotic, as observed for cephalosporins and the TEM beta-lactamases (Baquero *et al.* 1998) as well as tigecycline and the *tet*

tetracycline resistance genes (Linkevicius *et al.* 2016). Furthermore, accumulation of resistance genes against several different antibiotics on the same mobile genetic element can happen largely without reducing the fitness of its carrier. Given that several broad-spectrum antibiotics classes are used to treat the same bacteria, such co-localization is more likely than not to happen over time. Disturbingly, this also means that the likelihood that a gene is shared from one plasmid to another increases with time, as there are more plasmids that can act as donors for each gene. Thus, we would expect to see an increase of bacteria with plasmid-borne multi-resistance phenotypes, and that the rate of this increase would also increase with time. Indeed, this is what is currently observed among clinical isolates (Livermore 2009; European Centre for Disease Prevention and Control 2013). Importantly, this increase would appear even if antibiotics usage did not surge. Troublingly, global antibiotics usage is also on the rise (Laxminarayan 2014), likely accelerating the multi-resistance problem even further. The use of biocides and metals as antibacterials may also promote multidrug resistance, although to what extent is still uncertain (Baker-Austin *et al.* 2006; Sütterlin *et al.* 2014; Pal *et al.* 2015).

Multidrug resistance may not be the only problem we will face in the future though. Bacteria can generate genetic diversity through mutations, recombination and horizontal gene transfer. Each of these processes is under balancing selection, where the benefits of generating potentially adaptive genetic variants are weighted against the risk for fitness-reducing mutations (Gillings 2013). Antibiotic exposure has been shown to increase the mutation and recombination frequencies in bacteria, even at sub-inhibitory levels, through the SOS response (Beaber *et al.* 2004; López *et al.* 2007; Blázquez *et al.* 2012). Exposure of environmental bacteria to varying levels of antibiotics is therefore likely to generate variants with higher rate of genetic change, in addition to the selection pressure for resistance. Since bacteria that have higher mutation rates are more likely to get beneficial mutations, and also more likely to quickly generate compensatory genetic changes, antibiotic exposure may select for fixation of bacterial populations with generally higher rates of genetic changes (Gillings & Stokes 2012). In addition, antibiotics are often released into the environment together with bacteria carrying integrons and other mobile genetic elements (Gaze *et al.* 2011; 2013). Since integrase activity is also induced by antibiotics (Maiques *et al.* 2006), this may further increase bacterial evolvability, generating ever more complex mobile genetic elements (Gillings 2014). If those rearrangements come together with a generally increased mutation rate, the net result would be that evolution towards lower fitness cost of resistance genes could happen even more quickly. It is impossible to predict exactly what consequences this may have for the bacterial pangenome. Since integrons and other mobile genetic elements allow bacteria to adapt faster to new niches (Gillings 2014), genes mobilized in the future would likely not be restricted to conferring antibiotic resistance, but may also encompass genes that provide a fitness advantage in terms of adaption to changing environments. Thus genes allowing bacteria to survive highly variable abiotic conditions, handle toxicants, utilize novel carbon sources, compete with other

microbes, adhere to different types of surfaces, re-engineer their ecosystems, and allow formation of highly durable spores would be good candidates for future mobilization. From a human health perspective, it is easy to imagine that selection by antibiotics would favor strains with attributes that are beneficial for colonization and invasion of the human host. This could include mobilization of genes involved in virulence, transmission and pathogenicity (Gillings 2014), but also genes that increase competitive ability with human commensals. This paints a picture of a bleak future in which human pathogens are not only non-treatable by most antibiotics, but also become more aggressive and spread more easily between humans. This signifies the importance of understanding not only the risks for resistance transmission, but also the evolutionary consequences of antibiotics releases into the environment.

Concluding remarks

The risks associated with environmental antibiotic resistance need to be better understood to enable implementation of mitigation strategies to prevent or at least delay resistance gene recruitment to human pathogens. In this thesis, metagenomic DNA sequencing methods have been applied to contribute to this knowledge. To contextualize the results of studies of antibiotic resistance using shotgun metagenomic sequencing, other factors must be considered, such as whether the concentrations of antibiotics present in a studied environment are selective and the taxonomic structure of the investigated communities. To this end, a software tool called Metaxa2 has been developed that can reliably detect and extract rRNA fragments from metagenomes and classify these to higher accuracy than other competing software solutions (Paper I). Furthermore, the minimal selective concentrations in complex bacterial communities of 111 antibiotics were theoretically estimated based on publically available data (Paper II), and the specific selective concentration for tetracycline validated by experimental studies of a variety of endpoints (Paper III).

The results of Papers I-III were used to set the results of the shotgun metagenomic studies into context. In Paper IV, the selection processes for antibiotic resistance in Swedish sewage treatment plants were elucidated, and we found that there is little evidence for direct selection for resistance against particular antibiotics in these settings. Furthermore, we also found that co-selection between resistance to antibiotics, antibacterial biocides and metals appears to be limited. Instead, on the larger scale studied by metagenomics, shifts of taxonomic composition – caused by changes of abiotic factors such as oxygen availability – seem to drive changes of the resistomes of STP communities. These changes limit the interpretability of metagenomic data on resistance gene counts. Thus, comprehensive analyses of the resistance patterns in strains within relevant species are warranted to better understand the selection processes possibly leading to resistance in STPs.

On the contrary, the results of Paper V, as well as other data presented in this thesis, suggest that in environments subjected to high concentrations of antibiotics, such as those exposed to waste from pharmaceutical production, both the abundance and diversity of resistance genes are enriched. In addition, a long range of mobile genetic elements, such as plasmids, integrons and transposons, showed elevated abundances. Many of the resistance genes were also found to be associated with such mobile genetic contexts. This indicates that pharmaceutical pollution can create particularly severe hot-spot environments for resistance development. As evidenced by Paper VI, there seem to be a set of resistance genes that can spread globally by the means of human travel, even in the absence of antibiotics treatment. Those genes are also generally linked to the same mobile genetic elements found to be enriched in Paper V, pointing towards the intricate role of horizontal gene transfer in the development and global dispersal of antibiotic resistance.

When the results of this thesis are combined with other studies of clinical and environmental antibiotic resistance we are able to formalize an ecologically relevant framework for antibiotic resistance. This framework suggests that although emergence of novel resistance factors and mobilization of existing ones probably happen continuously, only few of these determinants are selected for and fixated among bacterial populations. Those that do make it to pathogenic species are likely evolved to convey very little fitness cost to their hosts, and are thus hard to eliminate from pathogen populations. Successful mitigation strategies are therefore in principle limited to avoiding creation of environmental settings that select for, mobilize and fixate resistance genes in bacterial communities, closing down the dispersal routes for resistant bacteria to the human microbiome, and limiting the selection pressure for resistant pathogens (i.e. prudent use of antibiotics for humans and animals). The diversity of resistance genes present in the environment suggests that there are many resistance genes available for pathogens to recruit. Since resistance genes are not likely to be eradicated from the bacterial populations of the human microbiome even in the absence of antibiotics selection, the genes that are already circulating among human pathogens may easily re-emerge during antibiotics treatment. The recruitment of novel resistance genes into pathogens, on the other hand, is likely to have far more devastating consequences for human health, as resistance genes against new antibiotics, or more efficient resistance mechanisms against the ones that already face resistance, would further reduce treatment options (Paper VII). Unfortunately, shotgun metagenomics is ill-posed to identify these yet unknown resistance factors, due to its dependency on reference databases of described resistance genes. Thus, while sequencing metagenomics can provide a snapshot of the environmental resistome, this picture will inevitably be incomplete. Metagenomics will therefore never be a panacea for all environmental antibiotic resistance research, but merely a tool among others. That said, metagenomic sequencing can provide clues to selection pressures, mobility potential of identified genes, taxonomic structure, and other important contextual information that may be impossible to obtain by other means. This enables metagenomics to make particularly valuable contributions to the growing body of knowledge on environmental antibiotic resistance. Such understanding of the environmental resistome is instrumental for future mitigation of antibiotic resistance development to be successful.

Postscript

It may seem hopeless; and indeed the future appears dark. Many experts fear that we are nearing a return to the pre-antibiotic era, which may also represent the end of modern healthcare as we know it (Collignon 2013; Kåhrström 2013; Fowler *et al.* 2014). That said, there are reasons to be optimistic about our ability to treat bacterial infections also in the future. Much research goes into finding antibiotics with novel mechanisms (Butler *et al.* 2013; Hesterkamp 2015), or finding ways of inhibiting the bacterial resistance mechanisms (Cox & Wright 2013), the latter which would re-enable use of antibiotics currently faced with immense resistance problems. Recent approaches have made it possible to discover novel antibiotic substances from non-cultivable bacteria, enabling us to tap into the antibacterial potential of yet uncultivable species (Ling *et al.* 2015). However, the economic incentives for the pharmaceutical industry to pursue the highly costly endeavor to find novel antibiotics are clearly not sufficient, particularly as new substances successfully killing multiresistant bacteria will likely be put on the shelf to use as last-resort antibiotics (Pharmaceutical, Biotechnology and Diagnostics Industries 2016). Thus, efficient new antibiotics is not a lucrative product category to invest in (Cooper & Shlaes 2011; Coates *et al.* 2011; Fernandes 2015). It is clear that governmental funding is required to sustain the development of new antibiotics, particularly as phase III trials have become vastly costly (Cooper & Shlaes 2011; Fernandes 2015). Several suggestions that could alleviate this situation have been made (Pharmaceutical, Biotechnology and Diagnostics Industries 2016), including that the public take all or some of the costs for phase III trials for antibiotics (Cooper & Shlaes 2011), that governments promise in advance to buy a stockpile of a successfully developed antibiotic (Cooper & Shlaes 2011; Brogan & Mossialos 2016), simplified regulation systems for antimicrobial agents (Cooper & Shlaes 2011; Coates *et al.* 2011; Cole 2014; Tomayko *et al.* 2014; Fernandes 2015), or simply increased funding to research into finding novel antibiotics (Cole 2014).

In addition to the efforts to develop new antibiotics, several alternative treatment strategies have emerged. One promising suggestion has been vaccines targeting bacteria. Vaccines have mostly been employed to fight viral infections, but there is (theoretically) no technical reason why similar approaches would not work also for pathogenic bacteria (Mishra *et al.* 2012; García-Quintanilla *et al.* 2016). Furthermore, targeting bacterial features less likely to drive resistance development is investigated, such as specifically attacking virulence (Allen *et al.* 2014). Other, more speculative, strategies involve using bacteriocins – bacterial antimicrobial peptides – as a weapon against bacteria, stimulating bacteria producing particular bacteriocins using prebiotics (Cotter *et al.* 2013), or to use bacteriophages to kill pathogens (Verbeken *et al.* 2014; 2016). Finally, it may also be possible to trigger the human immune system earlier, which could function as an alternative or supplement to traditional antibiotics (Cederlund *et al.* 2011; Agerberth *et al.* 2013). Evidently, we have not run out of ideas on how to tackle bacterial infections in the future. However, all these ideas, as well as

the antibiotics currently in development, need several years in research and clinical trials before they can be used for treatment. Therefore, we need to devise strategies that preserve the currently available antibiotics for as long as possible, which includes prudent antibiotics usage, dramatically reduced use for non-human treatment, and timely action to avoid dissemination of agents selecting for antibiotic resistance into the environment (Table 7). We should also remember that hygiene plays an immensely important role in preventing bacterial infections (Weinstein 2001; Sydnor & Perl 2011) – a fact often forgotten about when the post-antibiotic era is discussed. In fact, mortality rates in infectious diseases had already dropped dramatically when penicillin was introduced in the 1940-ies (Armstrong *et al.* 1999), largely thanks to improved recognition of microbes, implementation of hygiene routines, and access to clean water (Centers for Disease Control and Prevention 1999). It is important to recall that we will not lose these means of fighting microbial infections even if we run out of antibiotic treatment options. The future may be bleak, but there is hope.

Acknowledgements

This doctoral thesis has of course not been the work of a single person. I am greatly indebted to a vast number of people who have helped me or contributed in difference ways to the book your holding in your hand. First of all, I would not be writing this unless my main supervisor **Joakim Larsson** had found a place for me in his group. You have not only taught me about antibiotic resistance, risk assessment, pharmaceutical pollution and fishing, but also on how to write successful grant proposals, how to remove 50% of the discussion of your paper, the importance of bringing note taking equipment to meetings with your supervisor, and the workings of female reproduction. You truly *never* run out of ideas, which has been one of the most stimulating aspects of being in your group. I hope that we can continue this idea spawning together in one way or another, and I am tremendously thankful that you have taken so much time for discussions on my future life choices!

On the other side of the hill – on the Chalmers campus – **Erik Kristiansson** has always kept the door open when data have been poor, mood has been terrible, or statistical methods insufficient. I have never met another person with such a fantastic ability to find a golden grain in the most useless data set and reinvigorate enthusiasm for projects I totally lost faith in. I hope that we will continue our series of "it would be interesting to discuss this more in general" meetings also in the future – it has given me a lot!

Albeit never my formal supervisor, I also need to bring up **Henrik Nilsson** in this context. Henrik has, whether he liked it or not, functioned as my mentor over these years, teaching me about how to write, how to respond to evil review comments, the consequences of not keeping your databases clean, and everything I could ever wish for about fungal taxonomy. Thanks a million for that Henrik!

I have also had the benefit of a large number of great supervisors for my previous work. Particularly I like to thank **Magnus Alm Rosenblad**, who in 2008 got the responsibility to help a second-year student to sort out some metagenomic sequence data. Magnus did the only right thing: a list of stuff that I needed to know, including Linux, Perl and command-line BLAST. Did I ever tell you that I didn't know any of those three? Your mentoring-by-being-there-as-soon-as-a-problem-appeared worked like a charm and is a major reason for that I do what I do today. Thanks for the faith in me! I would also like to thank **Anders Blomberg** who encouraged me to do a bachelor thesis in bioinformatics and metagenomics, and thus was also involved in spawning this career path. Further, **Hans Blanck** and **Martin Eriksson**, who for some reason dared to trust a master student to sort out a periphyton metagenome, also deserves a special place in this section. I also want to thank all of you for not losing contact – it's been great having you around all these years.

Two persons were instrumental in my recruitment to Joakim's group – Erik Kristiansson and **Anna Johnning**. Anna sold Joakim's group to me by promises of fika and crosswords, and I was not disappointed in this regard. (Also, thanks for keeping me in the loop from the inside during the entire recruitment process.) Sadly, we have not been able to keep up these traditions properly since you left the group.

Speaking of Joakim's research group – or more correctly: the Larsson research junta – so many things have made this junta tick over the years. The crosswords, the chocolate Mondays, the now almost forgotten Art of Science battles, the totally epic spex-videos[2] (we even got a prize for being the most creative group at Neuroscience and Physiology), the endless discussions about feces, poo and stool in different forms, and the sport showdowns in curling and boule. Of course, none of this would have been any fun without the fantastic persons in the lab, current as well as former. **Filip**, it was great sharing room with you during your most stressful PhD year! **Fleshy**, although we only got a short time together, your know-it-all in essentially all areas was fascinating to take experience. **Bethany**, so much positive energy in one person – so many stories to tell, I miss you! **Ida**, I trust that you survive fine outside of the academy, but make no mistake – the academy wants you back! **Anna J**, it was great sharing room with you during your most stressful PhD year! **Anna B**, my only rival in taking the longest time to complete one's PhD in the group; getting a second kid is verging on cheating! **Lina**, you won't get rid of me; even in Exeter I keep following you (although nowadays on Skype). Thanks for discussing space travel to get me in a better mood. **Rickard**, thanks for being such an awesome self-sustaining master student while I was at 80% parental leave. A hard task, but you managed phenomenally! I hope you're doing great in Stockholm. **Kim**, being a summer-worker for someone who takes eight weeks of vacation each summer is not the easiest. Still, I have not heard you complaining. Massive. **Carolin**, my solid anchor at work. The oldest (in time in the group) and, thus of course wisest among the group members (particularly when it comes to crosswords). Thanks for taking your time to discuss literally everything from poo to stool, including the latest reports from Klubbkommissionen, while also (in passing) some science. **Flach** – as close to a supervisor a non-supervisor can be, I guess? Whenever I have needed a piece of information on microbiology, you have been there. Whenever I have needed a rapper with a classic Göteborg-accent, you have also been there. Another solid anchor you are. **Chandan**, how awesome that we got this globetrotting Indian guy in the group! I've enjoyed tremendously working with you (even though I realize that at times I have been rather grumpy, particularly now during my last year). Soon your turn to defend, and it will be great sharing room with you during your most stressful PhD year. **Sara**, how I will miss you. Do you realize how much of a perfect fit you were for the group? I mean, fecal discussions were upped by a factor of at least five when you arrived. Let's not lose contact even if you decide to give the academy the finger. **Carl-Johan**, thanks for being the one pointing out that **I** am the strangest person in the group. Both me and Anna appreciated this unexpected reversal of roles! **Martin**, it will be a pleasure to have you here for some additional time (this summer I will not be gone all the time, so we will get to meet more! Win-win situation!) **Nachiket**, it's wonderful how we have gotten India's most amiable researchers to our lab. Thanks for teaching me all about cricket, Indian holidays and the NDM-1 carbapenemase!

[2] No, it says **spex-videos**. You were misreading it. Sorry.

Jekaterina, always there in the morning – either in the lab or by the computer, always with hot coffee ready, always with a smile. You have been keeping my mood up during the writing of this thesis, thanks a lot for that! **Mohammad**, my newest roommate at work, and what a great roommate. Never complaining about my Skype-calls, and always ready to discuss cultural differences between Sweden and Iran. I have learned a lot from you and hope to learn even more! **Maja**, you have not only managed to stir up and sort out the laboratories and practical stuff in the group, you have also brought a lot of positive energy and attitude! On top of this, you and **Samuel** have been very nice dinner company; I hope that we can continue on this tradition when this winter-of-sickness is over! **Gustaf**, thanks for nice lunch discussions about catheters (yes, this is a much more fun subject than it sounds!) **Nadine**, **Stathis**, **Marlies** and **Ida**, thanks for a great month! I extrapolate widely and predict that we will have an upcoming awesome year together!

In addition to the Larsson junta, I have also had a fantastic safety net of great persons at Chalmers in Erik Kristiansson's group. Particularly, **Fredrik** who got the PhD position I thought I wanted (but after all I think we agree that the this outcome was the best for us both, right?) and has been my sounding board for everything bioinformatics all the way through my thesis; **Viktor** who's thoughts on metagenomics and the priorities in life I value tremendously (and you are also the best dance partner!); **Fanny** who similarly to Fredrik keeps me on my toes when it comes to recent bioinformatics and computer developments (it is also so cool that I get to hang around with one of Sweden's best pole-vaulters!); and **Tobbe** who I have had very rewarding discussions on the specificity of metagenomics (and nice dinners) with. Also, lots of love to my favorite Brazilian **Mariana**, the always-positive **Anna R**, the code genius **Anders**, the down-to-earth this-analysis-does-not-mean-anything **Kemal**, who repeatedly has been pointing out that what I'm doing is sometimes rather desperate, as well as **Emma**, **Sanna**, **Stefan**, **Robert** and **Olle** who have always kept the spirit high while I have been visiting Maths.

This acknowledgement section would be severely incomplete without a mention of my partner-in-crime **Kaisa**. I am so happy that we ended up on the same PhD introduction seminar, and look forward to sharing data, grant proposals, paper writing, ice skating, dinners and coffee with you in the future!

I would also like to direct a huge thank you to all my other collaborators in these studies[3]: **Martin Hartmann** on Metaxa2; **Marcus Östman** and **Mats Tysklind** on the tetracycline and sewage papers; **Malin Thoudal** and **Triranta Sircar** on the tetracycline paper; **Jerker Fick** (thanks also for the great train ride from Stockholm to Uppsala after the MistraPharma meeting!) on the sewage and lake studies; **Martin Angelin**, **Mikael Huss**, **Sanela Kjellqvist**, **Helena Palmgren** (thanks for the dinner!) and **Anders Johansson** on the travel study.

[3] Is this the longest Acknowledgements section ever? How do you people keep this reasonably short? I have this problem with word limits…

I have also had great support from the people in the **GoBiG** and **GOTBIN** networks – you know who you are and the interaction with you has been highly appreciated. I have also enjoyed hanging out with the crew at **SciLifeLab** in Stockholm, particularly **Luisa**, and everyone in the **MistraPharma** and **INTERACT** programs. Finally, it's a pity that Obama's lawyers turned my thesis down, but thank you for your effort **Jo**. I look very much forward to the collaboration that hopefully will be the result of it!

I am also very lucky to have lots of good friends that have been cheering, pushing, helping and reminding me about what is important in life (apart from working antibiotics). **Olle** & **Hanna**, thanks for borrowing me space in your freezer! **Hedvig** & **Erik**, thanks for lending me a bed in Stockholm whenever I needed one. **Love**, thanks for asking the hard questions regarding my work (I mean, what's really important with it), the philosophy and the music. Thanks to **Anton** for the larger perspective and the stargazing – you know that somewhere there's a physicist hiding in me too (and a biologist hiding in you). **Emma** & **Julius**; **Emily** & **Fredrik**, thanks for showing how to be awesome parents (and also thanks to Fredrik for discussing alternative uses of Markov models!) Thanks **Fredrik**, **Saga**, **Fia**, **Jennie**, **Andreas**, **Karin**, **Mike**, **Nina**, **Sofi**, **Klara**, **Siri**, **John**, **David**, **Jonas** (listed in some analphabetical order) and everyone else (many who have been mentioned above) for taking long walks, sharing lengthy fikas and/or playing board games with me. These have been great ways of thinking about something else than work. **Clara**, **Per**, and **Robin**, thanks for the coffee – it has been needed.

I also happen to have the world's best family: **Linnéa** – the world's best babysitter and the world's best aunt; **Anders** – my better-than-Wikipedia source of knowledge, the world's best uncle, and without whom my music taste would really suck (and, yes, I will always be envious of your natural curls); **Annette**, who has sort-of become a second mother for me; **Solveig**, who have passed on her curiousness for small and large things in nature to me, and **Bertil**, who have passed on the willing to protect said nature. Thanks to **Klas**, **Yvonne**, **Fredrik** and **Mårten** for acting as a second family when I moved to Gothenburg. Thanks **Bengt** for prospecting the best mushroom walks in the Greggered area. Also, a shoutout to all my cousins and their families on my mother's side and the rest of my clan: it's a pity that we don't meet more often! **Mum & Dad**, thanks for showing me that love is the most important thing in the world. You have been the best – always – and also turned out to be super grandparents. Who knew that?

Maj, how fantastic it is to get a piece of sunshine running around in our lives. You repeatedly show me that there are mort important things in life than handing in theses in time.

Finally, this work is as much yours as mine **Anna**. You have had to live through the ups and downs of the thesis writing process, and will now need to go through it again from the other side. I will do whatever I can to help you; as you have done everything you could to help me. We belong together. Now and forever. I love you.

References

1928 Diagnostics (2012) Resqu: A database of mobile antibiotic resistance genes. *http://www.1928diagnostics.com/resdb/*

Abraham EP, Chain E (1940) An enzyme from bacteria able to destroy penicillin. *Nature*, **146**, 837.

Agerberth B, Bergman P, Gudmundsson GH (2013) Helping the host: induction of antimicrobial peptides as a novel therapeutic strategy against infections. In: *Antimicrobial Peptides and Innate Immunity*, pp. 360–375. Springer, Basel.

Alekshun MN, Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell*, **128**, 1037–1050.

Allen HK (2014) Antibiotic resistance gene discovery in food-producing animals. *Current Opinion in Microbiology*, **19C**, 25–29.

Allen HK, Donato J, Wang HH *et al.* (2010) Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology*, **8**, 251–259.

Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J (2009) Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *The ISME Journal*, **3**, 243–251.

Allen RC, Popat R, Diggle SP, Brown SP (2014) Targeting virulence: can we make evolution-proof drugs? *Nature Reviews Microbiology*, **12**, 300–308.

Altschul SF, Madden TL, Schäffer AA *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, **25**, 3389–3402.

Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiological Reviews*, **59**, 143–169.

Aminov RI (2009) The role of antibiotics and antibiotic resistance in nature. *Environmental Microbiology*, **11**, 2970–2988.

Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology*, **11**, R106.

Andersson DI (2003) Persistence of antibiotic resistant bacteria. *Current Opinion in Microbiology*, **6**, 452–456.

Andersson DI, Hughes D (2010) Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology*, **8**, 260–271.

Andersson DI, Hughes D (2012) Evolution of antibiotic resistance at non-lethal drug concentrations. *Drug Resistance Updates*, **15**, 162–172.

Andersson DI, Hughes D (2014) Microbiological effects of sublethal levels of antibiotics. *Nature Reviews Microbiology*, **12**, 465–478.

Angelin M, Forsell J, Granlund M *et al.* (2015) Risk factors for colonization with extended-spectrum beta-lactamase producing Enterobacteriaceae in healthcare students on clinical assignment abroad: A prospective study. *Travel Medicine and Infectious Disease*, **13**, 223–229.

Angly FE, Dennis PG, Skarshewski A *et al.* (2014) CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome*, **2**, 11.

Armstrong GL, Conn LA, Pinner RW (1999) Trends in infectious disease mortality in the United States during the 20th century. *JAMA*, **281**, 61–66.

Ashbolt NJ, Amézquita A, Backhaus T *et al.* (2013) Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. *Environmental Health Perspectives*, **121**, 993–1001.

Ashbolt NJ, Schoen ME, Soller JA, Roser DJ (2010) Predicting pathogen risks to aid beach management: the real value of quantitative microbial risk assessment (QMRA). *Water research*, **44**, 4692–4703.

Atkinson HJ, Babbitt PC (2009) An atlas of the thioredoxin fold class reveals the complexity of function-enabling adaptations. *PLoS Computational Biology*, **5**, e1000541.

Ågerstrand M, Berg C, Björlenius B *et al.* (2015) Improving environmental risk assessment of human pharmaceuticals. *Environmental Science & Technology*, **49**, 5336–5345.

Babraham Bioinformatics (2012) Trim Galore!
http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Backhaus T (2014) Medicines, shaken and stirred: a critical review on the ecotoxicology of pharmaceutical mixtures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 20130585.

Baker-Austin C, Wright MS, Stepanauskas R, McArthur JV (2006) Co-selection of antibiotic and metal resistance. *Trends in microbiology*, **14**, 176–182.

Baltrus DA (2013) Exploring the costs of horizontal gene transfer. *Trends in Ecology & Evolution*, **28**, 489–495.

Baquero F, Alvarez-Ortega C, Martinez JL (2009) Ecology and evolution of antibiotic resistance. *Environmental Microbiology Reports*, **1**, 469–476.

Baquero F, Martinez JL, Cantón R (2008) Antibiotics and antibiotic resistance in water environments. *Current Opinion in Biotechnology*, **19**, 260–265.

Baquero F, Negri MC, Morosini MI, Blázquez J (1998) Antibiotic-selective environments. *Clinical Infectious Diseases*, **27 Suppl 1**, S5–11.

Baquero F, Tedim AP, Coque TM (2013) Antibiotic resistance shaping multi-level population biology of bacteria. *Frontiers in Microbiology*, **4**, 15.

Barberán A, Ladau J, Leff JW *et al.* (2015) Continental-scale distributions of dust-associated bacteria and fungi. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5756–5761.

Barraud N, Hassett DJ, Hwang S-H *et al.* (2006) Involvement of nitric oxide in biofilm dispersal of *Pseudomonas aeruginosa*. *Journal of Bacteriology*, **188**, 7344–7353.

Beaber JW, Hochhut B, Waldor MK (2004) SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature*, **427**, 72–74.

Bebell LM, Muiru AN (2014) Antibiotic use and emerging resistance: how can resource-limited countries turn the tide? *Global Heart*, **9**, 347–358.

Bengtsson J (2009) Applied (meta)community ecology: diversity and ecosystem services at the intersection of local and regional processes. In: *Community Ecology: Processes, Models, and Applications* (eds Verhoef HA, Morin PJ), pp. 115–130. Oxford University Press, Oxford.

Bengtsson J, Eriksson KM, Hartmann M *et al.* (2011) Metaxa: a software tool for automated detection and discrimination among ribosomal small subunit (12S/16S/18S) sequences of archaea, bacteria, eukaryotes, mitochondria, and chloroplasts in metagenomes and environmental sequencing datasets. *Antonie van Leeuwenhoek*, **100**, 471–475.

Bengtsson J, Hartmann M, Unterseher M *et al.* (2012) Megraft: a software package to graft ribosomal small subunit (16S/18S) fragments onto full-length sequences for accurate species richness and sequencing depth analysis in pyrosequencing-length metagenomes and similar environmental datasets. *Research in Microbiology*, **163**, 407–412.

Bengtsson-Palme J (2012) PETKit. *http://microbiology.se/software/petkit/*

Bengtsson-Palme J, Larsson DGJ (2015) Antibiotic resistance genes in the environment: prioritizing risks. *Nature Reviews Microbiology*, **13**, 396. **Paper VII.**

Bengtsson-Palme J, Larsson DGJ (2016a) Time to regulate antibiotic pollution. *The Medicine Maker*, in press.

Bengtsson-Palme J, Larsson DGJ (2016b) Concentrations of antibiotics predicted to select for resistant bacteria: Proposed limits for environmental regulation. *Environment International*, **86**, 140–149. **Paper II.**

Bengtsson-Palme J, Alm Rosenblad M, Molin M, Blomberg A (2014a) Metagenomics reveals that detoxification systems are underrepresented in marine bacterial communities. *BMC Genomics*, **15**, 749.

Bengtsson-Palme J, Angelin M, Huss M *et al.* (2015a) The Human Gut Microbiome as a Transporter of Antibiotic Resistance Genes between Continents. *Antimicrobial Agents and Chemotherapy*, **59**, 6551–6560. **Paper VI.**

Bengtsson-Palme J, Boulund F, Fick J, Kristiansson E, Larsson DGJ (2014b) Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Frontiers in Microbiology*, **5**, 648. **Paper V.**

Bengtsson-Palme J, Hammarén R, Pal C *et al.* (2016a) Elucidating selection processes for antibiotic resistance in sewage treatment plants using metagenomics. *Manuscript in preparation*. **Paper IV.**

Bengtsson-Palme J, Hartmann M, Eriksson KM *et al.* (2015b) Metaxa2: Improved identification and taxonomic classification of small and large subunit rRNA in metagenomic data. *Molecular Ecology Resources*, **15**, 1403–1414. **Paper I.**

Bengtsson-Palme J, Thorell K, Wurzbacher C, Sjöling Å, Nilsson RH (2016b) Metaxa2 Diversity Tools: Easing microbial community analysis with Metaxa2. *Ecological Informatics. doi: 10.1016/j.ecoinf. 2016.04.004*

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Berendonk TU, Manaia CM, Merlin C *et al.* (2015) Tackling antibiotic resistance: the environmental framework. *Nature Reviews Microbiology*, **13**, 310–317.

Bianchi L, Díez-Sampedro A (2010) A single amino acid change converts the sugar sensor SGLT3 into a sugar transporter. *PLoS ONE*, **5**, e10241.

Björkholm B, Sjölund M, Falk PG *et al.* (2001) Mutation frequency and biological cost of antibiotic resistance in Helicobacter pylori. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 14607–14612.

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annual Review of Biochemistry*, **72**, 291–336.

Blanck H (2002) A Critical Review of Procedures and Approaches Used for Assessing Pollution-Induced Community Tolerance (PICT) in Biotic Communities. *Human and Ecological Risk Assessment*, **8**, 1003–1034.

Blázquez J, Couce A, Rodríguez-Beltrán J, Rodríguez-Rojas A (2012) Antimicrobials as promoters of genetic variation. *Current Opinion in Microbiology*, **15**, 561–569.

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J (2012) Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biology*, **13**, R122.

Bondarczuk K, Markowicz A, Piotrowska-Seget Z (2015) The urgent need for risk assessment on the antibiotic resistance spread via sewage sludge land application. *Environment International*, **87**, 49–55.

Boralkar DB, Alvares C, Devotta S, Sharma PN, Thyagarajan G (2005) *Report of visit to Hyderabad (A.P.) 19-20 October 2004*. Supreme Court of India, New Dehli.

Boulund F, Johnning A, Pereira MB, Larsson DGJ, Kristiansson E (2012) A novel method to discover fluoroquinolone antibiotic resistance (*qnr*) genes in fragmented nucleotide sequences. *BMC Genomics*, **13**, 695.

Bouma JE, Lenski RE (1988) Evolution of a bacteria/plasmid association. *Nature*, **335**, 351–352.

Brogan DM, Mossialos E (2016) Systems, not pills: The options market for antibiotics seeks to rejuvenate the antibiotic pipeline. *Social Science & Medicine*, **151**, 167–172.

Brosché S, Backhaus T (2010) Toxicity of five protein synthesis inhibiting antibiotics and their mixture to limnic bacterial communities. *Aquatic Toxicology*, **99**, 457–465.

Buchholz U, Bernard H, Werber D *et al.* (2011) German outbreak of *Escherichia coli* O104:H4 associated with sprouts. *The New England Journal of Medicine*, **365**, 1763–1770.

Bumgarner R (2013) Overview of DNA microarrays: types, applications, and their future. *Current Protocols in Molecular Biology*, **22**, 22.1.

Bush K (2012) Improving known classes of antibiotics: an optimistic approach for the future. *Current Opinion in Pharmacology*, **12**, 527–534.

Butler J, Maccallum I, Kleber M *et al.* (2008) ALLPATHS: *de novo* assembly of whole-genome shotgun microreads. *Genome Research*, **18**, 810–820.

Butler MS, Blaskovich MA, Cooper MA (2013) Antibiotics in the clinical pipeline in 2013. *The Journal of Antibiotics*, **66**, 571–591.

Cabello FC (2006) Heavy use of prophylactic antibiotics in aquaculture: a growing problem for human and animal health and for the environment. *Environmental Microbiology*, **8**, 1137–1144.

Cai L, Yu K, Yang Y *et al.* (2013) Metagenomic exploration reveals high levels of microbial arsenic metabolism genes in activated sludge and coastal sediments. *Applied Microbiology and Biotechnology*, **97**, 9579–9588.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Casali N, Nikolayevskyy V, Balabanova Y *et al.* (2014) Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nature Genetics*, **46**, 279–286.

Cederlund A, Gudmundsson GH, Agerberth B (2011) Antimicrobial peptides important in innate immunity. *The FEBS Journal*, **278**, 3942–3951.

Centers for Disease Control and Prevention (1999) Control of infectious diseases. *Morbidity and Mortality Weekly Report*, **48**, 621–629.

Chain E, Florey HW, Gardner AD *et al.* (1940) Penicillin as a chemotherapeutic agent. *Lancet*, **239**, 226–228.

Chao A (1984) Nonparametric Estimation of the Number of Classes in a Population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chao A, Lee S-M (1992) Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, **99**, 45–56.

Choo JM, Leong LE, Rogers GB (2015) Sample storage conditions significantly influence faecal microbiome profiles. *Scientific Reports*, **5**, 16350.

Clardy J, Fischbach MA, Currie CR (2009) The natural history of antibiotics. *Current Biology*, **19**, R437–41.

Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2016) GenBank. *Nucleic Acids Research*, **44**, D67–72.

Coates ARM, Halls G, Hu Y (2011) Novel classes of antibiotics or more of the same? *British Journal of Pharmacology*, **163**, 184–194.

Cole ST (2014) Who will develop new antibacterial agents? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 20130430.

Collignon P (2013) The importance of a One Health approach to preventing the development and spread of antibiotic resistance. *Current Topics in Microbiology and Immunology*, **366**, 19–36.

Collignon P (2015) Antibiotic resistance: are we all doomed? *Internal Medicine Journal*, **45**, 1109–1115.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, **345**, 101–118.

Cooper MA, Shlaes D (2011) Fix the antibiotics pipeline. *Nature*, **472**, 32.

Cotter PD, Ross RP, Hill C (2013) Bacteriocins - a viable alternative to antibiotics? *Nature Reviews Microbiology*, **11**, 95–105.

Cox G, Wright GD (2013) Intrinsic antibiotic resistance: mechanisms, origins, challenges and solutions. *International Journal of Medical Microbiology*, **303**, 287–292.

D'Costa VM, King CE, Kalan L *et al.* (2011) Antibiotic resistance is ancient. *Nature*, **477**, 457–461.

Dabrazhynetskaya A, Brendler T, Ji X, Austin S (2009) Switching protein-DNA recognition specificity by single-amino-acid substitutions in the P1 par family of plasmid partition elements. *Journal of Bacteriology*, **191**, 1126–1131.

Dahllöf I, Baillie H, Kjelleberg S (2000) *rpoB*-based microbial community analysis avoids limitations inherent in 16S rRNA gene intraspecies heterogeneity. *Applied and Environmental Microbiology*, **66**, 3376–3380.

Dantas G, Sommer MO (2012) Context matters - the complex interplay between resistome genotypes and resistance phenotypes. *Current Opinion in Microbiology*, **15**, 577–582.

Darling AE, Jospin G, Lowe E *et al.* (2014) PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, **2**, e243.

Davenport CF, Neugebauer J, Beckmann N *et al.* (2012) Genometa - a fast and accurate classifier for short metagenomic shotgun reads. *PLoS ONE*, **7**, e41224.

De Boeck H, Miwanda B, Lunguya-Metila O *et al.* (2012) ESBL-positive Enterobacteria isolates in drinking water. *Emerging Infectious Diseases*, **18**, 1019–1020.

Demain AL (1998) Induction of microbial secondary metabolism. *International Microbiology*, **1**, 259–264.

Dillies M-A, Rau A, Aubert J *et al.* (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, **14**, 671–683.

Dinsdale EA, Edwards RA, Bailey BA *et al.* (2013) Multivariate analysis of functional metagenomes. *Frontiers in Genetics*, **4**, 41.

Dobin A, Davis CA, Schlesinger F *et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

Dohm JC, Lottaz C, Borodina T, Himmelbauer H (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, **36**, e105.

Dortet L, Nordmann P, Poirel L (2012) Association of the emerging carbapenemase NDM-1 with a bleomycin resistance protein in Enterobacteriaceae and *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, **56**, 1693–1697.

Dörr T, Lewis K, Vulić M (2009) SOS response induces persistence to fluoroquinolones in *Escherichia coli*. *PLoS Genetics*, **5**, e1000760.

Dunn OJ (1959) Estimation of the medians for dependent variables. *The Annals of Mathematical Statistics*, **30**, 192–197.

Dunn OJ (1961) Multiple comparisons among means. *Journal of the American Statistical Association*, **56**, 52–64.

Durso LM, Miller DN, Wienhold BJ (2012) Distribution and Quantification of Antibiotic Resistant Genes and Bacteria across Agricultural and Non-Agricultural Metagenomes. *PLoS ONE*, **7**, e48325.

Elmqvist T, Folke C, Nyström M *et al.* (2003) Response Diversity, Ecosystem Change, and Resilience. *Frontiers in Ecology and the Environment*, **1**, 488.

Enne VI, Bennett PM, Livermore DM, Hall LMC (2004) Enhancement of host fitness by the *sul2*-coding plasmid p9123 in the absence of selective pressure. *The Journal of Antimicrobial Chemotherapy*, **53**, 958–963.

Enne VI, Livermore DM, Stephens P, Hall LM (2001) Persistence of sulphonamide resistance in *Escherichia coli* in the UK despite national prescribing restriction. *Lancet*, **357**, 1325–1328.

European Centre for Disease Prevention and Control (2013) *Antimicrobial resistance surveillance in Europe 2012. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net)*. European Centre for Disease Prevention and Control, Stockholm.

European Committee on Antimicrobial Susceptibility Testing (2016) EUCAST. *European Society of Clinical Microbiology and Infectious Diseases*. *http://mic.eucast.org/Eucast2/*

European Food Safety Authority, European Centre for Disease Prevention and Control (2013) The European Union Summary Report on antimicrobial resistance inzoonotic and indicator bacteria from humans, animals and food in 2011. *EFSA Journal*, **11**, 3196.

Fernandes P (2015) The global challenge of new classes of antibacterial agents: an industry perspective. *Current Opinion in Pharmacology*, **24**, 7–11.

Fernando GATP, Collignon PJ, Bell JM (2010) A risk for returned travellers: the "post-antibiotic era". *The Medical Journal of Australia*, **193**, 59.

Fick J, Söderström H, Lindberg RH *et al.* (2009) Contamination of surface, ground, and drinking water from pharmaceutical production. *Environmental Toxicology and Chemistry*, **28**, 2522–2527.

Finley RL, Collignon P, Larsson DGJ *et al.* (2013) The scourge of antibiotic resistance: the important role of the environment. *Clinical Infectious Diseases*, **57**, 704–710.

Flach C-F, Boulund F, Kristiansson E, Larsson DGJ (2013) Functional verification of computationally predicted *qnr* genes. *Annals of Clinical Microbiology and Antimicrobials*, **12**, 34.

Flach C-F, Johnning A, Nilsson I *et al.* (2015) Isolation of novel IncA/C and IncN fluoroquinolone resistance plasmids from an antibiotic-polluted lake. *Journal of Antimicrobial Chemotherapy*, **70**, 2709–2717.

Fleming A (1929) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. *The British Journal of Experimental Pathology*, **10**, 226–236.

Forsberg KJ, Reyes A, Wang B *et al.* (2012) The shared antibiotic resistome of soil bacteria and human pathogens. *Science*, **337**, 1107–1111.

Forslund K, Sunagawa S, Kultima JR *et al.* (2013) Country-specific antibiotic use practices impact the human gut resistome. *Genome Research*, **23**, 1163–1169.

Fowler T, Walker D, Davies SC (2014) The risk/benefit of predicting a post-antibiotic era: is the alarm working? *Annals of the New York Academy of Sciences*, **1323**, 1–10.

Freitas TAK, Li P-E, Scholz MB, Chain PSG (2015) Accurate read-based metagenome characterization using a hierarchical suite of unique signatures. *Nucleic Acids Research*, **43**, e69.

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nature Reviews Microbiology*, **3**, 722–732.

Gabriel P (2002) *Up* (P Gabriel, Ed,). Real World Records.

Gao P, Munir M, Xagoraraki I (2012) Correlation of tetracycline and sulfonamide antibiotics with corresponding resistance genes and resistant bacteria in a conventional municipal wastewater treatment plant. *The Science of the Total Environment*, **421-422**, 173–183.

García-Quintanilla M, Pulido MR, Carretero-Ledesma M, McConnell MJ (2016) Vaccines for Antibiotic-Resistant Bacteria: Possibility or Pipe Dream? *Trends in Pharmacological Sciences*, **37**, 143–152.

Gaze WH, Krone SM, Larsson DGJ *et al.* (2013) Influence of humans on evolution and mobilization of environmental antibiotic resistome. *Emerging Infectious Diseases*, **19**.

Gaze WH, Zhang L, Abdouslam NA *et al.* (2011) Impacts of anthropogenic activity on the ecology of class 1 integrons and integron-associated genes in the environment. *The ISME Journal*, **5**, 1253–1261.

Gerlach W, Stoye J (2011) Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic Acids Research*, **39**, e91.

Ghosh TS, Monzoorul Haque M, Mande SS (2010) DiScRIBinATE: a rapid method for accurate taxonomic classification of metagenomic sequences. *BMC Bioinformatics*, **11 Suppl 7**, S14.

Gillings MR (2013) Evolutionary consequences of antibiotic use for the resistome, mobilome and microbial pangenome. *Frontiers in Microbiology*, **4**, 4.

Gillings MR (2014) Integrons: past, present, and future. *Microbiology and Molecular Biology Reviews*, **78**, 257–277.

Gillings MR, Stokes HW (2012) Are humans increasing bacterial evolvability? *Trends in Ecology & Evolution*, **27**, 346–352.

Giovannoni SJ, Cameron Thrash J, Temperton B (2014) Implications of streamlining theory for microbial ecology. *The ISME Journal*, **8**, 1553–1565.

Glaser L, Stevens J, Zamarin D *et al.* (2005) A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *Journal of Virology*, **79**, 11533–11536.

Glenn TC (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, **11**, 759–769.

Gonzalez A, Ronce O, Ferriere R, Hochberg ME (2013) Evolutionary rescue: an emerging focus at the intersection between ecology and evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **368**, 20120404.

Grabherr MG, Haas BJ, Yassour M *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**, 644–652.

Greenpeace (2004) *State of Community Health at Medak District*. Greenpeace.

Guerin E, Cambray G, Sanchez-Alberola N *et al.* (2009) The SOS response controls integron recombination. *Science*, **324**, 1034.

Gullberg E, Albrecht LM, Karlsson C, Sandegren L, Andersson DI (2014) Selection of a multidrug resistance plasmid by sublethal levels of antibiotics and heavy metals. *mBio*, **5**, e01918–14.

Gullberg E, Cao S, Berg OG *et al.* (2011) Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathogens*, **7**, e1002158.

Gupta SK, Padmanabhan BR, Diene SM *et al.* (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial Agents and Chemotherapy*, **58**, 212–220.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, **5**, R245–9.

Hartmann M, Howes CG, Abarenkov K, Mohn WW, Nilsson RH (2010) V-Xtractor: an open-source, high-throughput software tool to identify and extract hypervariable regions of small subunit (16S/18S) ribosomal RNA gene sequences. *Journal of Microbiological Methods*, **83**, 250–253.

Harwood VJ, Staley C, Badgley BD, Borges K, Korajkic A (2014) Microbial source tracking markers for detection of fecal contamination in environmental waters: relationships between pathogens and human health outcomes. *FEMS Microbiology Reviews*, **38**, 1–40.

Hastings PJ, Rosenberg SM, Slack A (2004) Antibiotic-induced lateral transfer of antibiotic resistance. *Trends in Microbiology*, **12**, 401–404.

Hatosy SM, Martiny AC (2015) The Ocean as a Global Reservoir of Antibiotic Resistance Genes. *Applied and Environmental Microbiology*, **81**, 7593–7599.

Hayden EC (2013) Gene sequencing leaves the laboratory. *Nature*, **494**, 290–291.

He X, Chen H, Shi W, Cui Y, Zhang X-X (2015) Persistence of mitochondrial DNA markers as fecal indicators in water environments. *The Science of the Total Environment*, **533**, 383–390.

Heather JM, Chain B (2016) The sequence of sequencers: The history of sequencing DNA. *Genomics*, **107**, 1–8.

Heid CA, Stevens J, Livak KJ, Williams PM (1996) Real time quantitative PCR. *Genome Research*, **6**, 986–994.

Hellman J, Aspevall O, Bengtsson B, Pringle M (2014) *SWEDRES-SVARM 2014. Consumption of antimicrobials and occurrence of antimicrobial resistance in Sweden*. Public Health Agency of Sweden and National Veterinary Institute, Solna/Uppsala, Sweden.

Hess M, Sczyrba A, Egan R *et al.* (2011) Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science*, **331**, 463–467.

Hesterkamp T (2015) Antibiotics Clinical Development and Pipeline. *Current Topics in Microbiology and Immunology. doi: 10.1007/82_2015_451*

Hocquet D, Llanes C, Thouverez M *et al.* (2012) Evidence for induction of integron-based antibiotic resistance by the SOS response in a clinical setting. *PLoS Pathogens*, **8**, e1002778.

Hollis A, Ahmed Z (2013) Preserving antibiotics, rationally. *The New England Journal of Medicine*, **369**, 2474–2476.

Holt KE, Wertheim H, Zadoks RN *et al.* (2015) Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, E3574–81.

Howe AC, Jansson JK, Malfatti SA *et al.* (2014) Tackling soil diversity with the assembly of large, complex metagenomes. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 4904–4909.

Hu Y, Yang X, Qin J *et al.* (2013) Metagenome-wide analysis of antibiotic resistance genes in a large cohort of human gut microbiota. *Nature Communications*, **4**, 2151.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Research*, 9, 868–877.

Huang Y, Gilna P, Li W (2009) Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics*, **25**, 1338–1340.

Hugerth LW, Larsson J, Alneberg J *et al.* (2015) Metagenome-assembled genomes uncover a global brackish microbiome. *Genome Biology*, **16**, 279.

Hughes D, Andersson DI (2012) Selection of resistance at lethal and non-lethal antibiotic concentrations. *Current Opinion in Microbiology*, **15**, 555–560.

Hughes JB, Hellmann JJ (2005) The application of rarefaction techniques to molecular inventories of microbial diversity. *Methods in Enzymology*, **397**, 292–308.

Hughes JB, Hellmann JJ, Ricketts TH, Bohannan BJ (2001) Counting the uncountable: statistical approaches to estimating microbial diversity. *Applied and Environmental Microbiology*, **67**, 4399–4406.

Human Microbiome Jumpstart Reference Strains Consortium, Nelson KE, Weinstock GM *et al.* (2010) A catalog of reference genomes from the human microbiome. *Science*, **328**, 994–999.

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. *Genome Research*, **21**, 1552–1560.

Idury RM, Waterman MS (1995) A new algorithm for DNA sequence assembly. *Journal of Computational Biology*, **2**, 291–306.

Il'ina TS (2012) Mobile ISCR elements: structure, functions, and role in the emergence, increasing and spreading of blocks of bacterial genes of multiple antibiotic resistance. *Molekuliarnaia Genetika, Mikrobiologiia i Virusologiia*, 3–13.

Jackson DW, Suzuki K, Oakford L *et al.* (2002) Biofilm formation and dispersal under the influence of the global regulator CsrA of *Escherichia coli*. *Journal of Bacteriology*, **184**, 290–301.

Jeffery IB, Higgins DG, Culhane AC (2006) Comparison and evaluation of methods for generating differentially expressed gene lists from microarray data. *BMC Bioinformatics*, **7**, 359.

Jernberg C, Löfmark S, Edlund C, Jansson JK (2007) Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *The ISME Journal*, **1**, 56–66.

Johnning A, Moore ERB, Svensson-Stadler L *et al.* (2013) Acquired genetic mechanisms of a multiresistant bacterium isolated from a treatment plant receiving wastewater from antibiotic production. *Applied and Environmental Microbiology*, **79**, 7256–7263.

Johnson APA, Woodford NN (2013) Global spread of antibiotic resistance: the example of New Delhi metallo-β-lactamase (NDM)-mediated carbapenem resistance. *Journal of Medical Microbiology*, **62**, 499–513.

Johnson ET, Ryu S, Yi H *et al.* (2001) Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *The Plant Journal*, **25**, 325–333.

Johnson TJ, Singer RS, Isaacson RE *et al.* (2015) *In Vivo* Transmission of an IncA/C Plasmid in *Escherichia coli* Depends on Tetracycline Concentration, and Acquisition of the Plasmid Results in a Variable Cost of Fitness. *Applied and Environmental Microbiology*, **81**, 3561–3570.

Jonsson V, Österlund T, Nerman O, Kristiansson E (2016) Statistical evaluation of methods for identification of differentially abundant genes in comparative metagenomics. *BMC Genomics*, **17**, 78.

Jutkina J, Rutgersson C, Flach C-F, Larsson DGJ (2016) An assay for determining minimal concentrations of antibiotics that drive horizontal transfer of resistance. *The Science of the Total Environment*, **548-549**, 131–138.

Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S (2012) Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16213–16216.

Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G (2007) Structural and functional diversity of the microbial kinome. *PLoS Biology*, **5**, e17.

Kantele A, Lääveri T, Mero S *et al.* (2015) Antimicrobials increase travelers' risk of colonization by extended-spectrum betalactamase-producing enterobacteriaceae. *Clinical Infectious Diseases*, **60**, 837–846.

Karlsson FH, Nookaew I, Nielsen J (2014) Metagenomic Data Utilization and Analysis (MEDUSA) and Construction of a Global Gut Microbial Gene Catalogue. *PLoS Computational Biology*, **10**, e1003706.

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.

Kåhrström CT (2013) Entering a post-antibiotic era? *Nature Reviews Microbiology*, **11**, 146.

Kent WJ (2002) BLAT--the BLAST-like alignment tool. *Genome Research*, **12**, 656–664.

Kildare BJ, Leutenegger CM, McSwain BS *et al.* (2007) 16S rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal Bacteroidales: a Bayesian approach. *Water Research*, **41**, 3701–3715.

Klappenbach JA, Dunbar JM, Schmidt TM (2000) rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, **66**, 1328–1333.

Knapp CW, McCluskey SM, Singh BK *et al.* (2011) Antibiotic resistance gene abundances correlate with metal and geochemical conditions in archived Scottish soils. *PLoS ONE*, **6**, e27300.

Knauth S, Schmidt H, Tippkötter R (2013) Comparison of commercial kits for the extraction of DNA from paddy soils. *Letters in Applied Microbiology*, **56**, 222–228.

Kookana RS, Williams M, Boxall ABA *et al.* (2014) Potential ecological footprints of active pharmaceutical ingredients: an examination of risk factors in low-, middle- and high-income countries. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**, 20130586.

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, **28**, 3211–3217.

Kõljalg U, Nilsson RH, Abarenkov K *et al.* (2013) Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, **22**, 5271–5277.

Kristiansson E, Fick J, Janzon A *et al.* (2011) Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. *PLoS ONE*, **6**, e17038.

Kristiansson E, Hugenholtz P, Dalevi D (2009) ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics*, **25**, 2737–2738.

Kurtz S (2010) The Vmatch large scale sequence analysis software. *http://vmatch.de*.

Kümmerer K (2009a) Antibiotics in the aquatic environment--a review--part I. *Chemosphere*, **75**, 417–434.

Kümmerer K (2009b) Antibiotics in the aquatic environment--a review--part II. *Chemosphere*, **75**, 435–441.

Laht M, Karkman A, Voolaid V *et al.* (2014) Abundances of Tetracycline, Sulphonamide and Beta-Lactam Antibiotic Resistance Genes in Conventional Wastewater Treatment Plants (WWTPs) with Different Waste Load. *PLoS ONE*, **9**, e103705.

Lang KS, Anderson JM, Schwarz S *et al.* (2010) Novel florfenicol and chloramphenicol resistance gene discovered in Alaskan soil by using functional metagenomics. *Applied and Environmental Microbiology*, **76**, 5321–5326.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Larsson DGJ (2014a) Antibiotics in the environment. *Upsala Journal of Medical Sciences*, **119**, 108–112.

Larsson DGJ (2014b) Pollution from drug manufacturing: review and perspectives. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **369**, 20130571.

Larsson DGJ, de Pedro C, Paxeus N (2007) Effluent from drug manufactures contains extremely high levels of pharmaceuticals. *Journal of Hazardous Materials*, **148**, 751–755.

Law CW, Chen Y, Shi W, Smyth GK (2014) Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, **15**, R29.

Laxminarayan R (2014) Antibiotic effectiveness: balancing conservation against innovation. *Science*, **345**, 1299–1301.

Lee JE, Lee S, Sung J, Ko G (2011) Analysis of human and animal fecal microbiota for microbial source tracking. *The ISME Journal*, **5**, 362–365.

Leggett MJ, McDonnell G, Denyer SP, Setlow P, Maillard JY (2012) Bacterial spore structures and their protective role in biocide resistance. *Journal of Applied Microbiology*, **113**, 485–498.

Leibold MA, Holyoak M, Mouquet N *et al.* (2004) The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, **7**, 601–613.

Lennon JT, Jones SE (2011) Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nature Reviews Microbiology*, **9**, 119–130.

Levin BR, Baquero F, Johnsen PJ (2014) A model-guided analysis and perspective on the evolution and epidemiology of antibiotic resistance and its future. *Current Opinion in Microbiology*, **19C**, 83–89.

Levin BR, Lipsitch M, Perrot V *et al.* (1997) The population genetics of antibiotic resistance. *Clinical Infectious Diseases*, **24 Suppl 1**, S9–16.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li L-G, Cai L, Zhang X-X, Zhang T (2014) Potentially novel copper resistance genes in copper-enriched activated sludge revealed by metagenomic analysis. *Applied Microbiology and Biotechnology*, **98**, 10255–10266.

Li R, Zhu H, Ruan J *et al.* (2010) *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research*, **20**, 265–272.

Li Z, Chen Y, Mu D *et al.* (2012) Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics*, **11**, 25–37.

Lin D, Chen K, Li R *et al.* (2014) Selection of target mutation in rat gastrointestinal tract *E. coli* by minute dosage of enrofloxacin. *Frontiers in Microbiology*, **5**, 468.

Linares JF, Gustafsson I, Baquero F, Martinez JL (2006) Antibiotics as intermicrobial signaling agents instead of weapons. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19484–19489.

Ling LL, Schneider T, Peoples AJ *et al.* (2015) A new antibiotic kills pathogens without detectable resistance. *Nature*, **517**, 455–459.

Linkevicius M, Sandegren L, Andersson DI (2016) Potential of Tetracycline Resistance Proteins To Evolve Tigecycline Resistance. *Antimicrobial Agents and Chemotherapy*, **60**, 789–796.

Lipsitch M, Bergstrom CT, Levin BR (2000) The epidemiology of antibiotic resistance in hospitals: paradoxes and prescriptions. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 1938–1943.

Liu A, Fong A, Becket E *et al.* (2011a) Selective advantage of resistant strains at trace levels of antibiotics: a simple and ultrasensitive color test for detection of antibiotics and genotoxic agents. *Antimicrobial Agents and Chemotherapy*, **55**, 1204–1210.

Liu B, Pop M (2009) ARDB--Antibiotic Resistance Genes Database. *Nucleic Acids Research*, **37**, D443–7.

Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M (2011b) Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics*, **12 Suppl 2**, S4.

Liu J, Wang H, Yang H et al. (2013) Composition-based classification of short metagenomic sequences elucidates the landscapes of taxonomic and functional enrichment of microorganisms. *Nucleic Acids Research*, **41**, e3.

Liu Y-Y, Wang Y, Walsh TR et al. (2016) Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet Infectious Diseases*, **16**, 161–168.

Livermore DM (2000) Epidemiology of antibiotic resistance. *Intensive Care Medicine*, **26 Suppl 1**, S14–21.

Livermore DM (2009) Has the era of untreatable infections arrived? *Journal of Antimicrobial Chemotherapy*, **64 Suppl 1**, i29–36.

Loreau M, Mouquet N, Gonzalez A (2003) Biodiversity as spatial insurance in heterogeneous landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 12765–12770.

López E, Blázquez J (2009) Effect of subinhibitory concentrations of antibiotics on intrachromosomal homologous recombination in *Escherichia coli*. *Antimicrobial Agents and Chemotherapy*, **53**, 3411–3415.

López E, Elez M, Matic I, Blázquez J (2007) Antibiotic-mediated recombination: ciprofloxacin stimulates SOS-independent recombination of divergent sequences in *Escherichia coli*. *Molecular Microbiology*, **64**, 83–93.

Löfmark S, Jernberg C, Billström H, Andersson DI, Edlund C (2008) Restored fitness leads to long-term persistence of resistant *Bacteroides* strains in the human intestine. *Anaerobe*, **14**, 157–160.

Lundström SV, Östman M, Bengtsson-Palme J *et al.* (2016) Minimal selective concentrations of tetracycline in complex aquatic bacterial biofilms. *The Science of the Total Environment*, 553, 587–595. **Paper III.**

Lupo A, Coyne S, Berendonk TU (2012) Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. *Frontiers in Microbiology*, **3**, 18.

Ma L, Li B, Zhang T (2014) Abundant rifampin resistance genes and significant correlations of antibiotic resistance genes and plasmids in various environments revealed by metagenomic analysis. *Applied Microbiology and Biotechnology*, **98**, 5195–5204.

Ma L, Xia Y, Li B *et al.* (2016) Metagenomic Assembly Reveals Hosts of Antibiotic Resistance Genes and the Shared Resistome in Pig, Chicken, and Human Feces. *Environmental Science & Technology*, **50**, 420–427.

Mackelprang R, Waldrop MP, Deangelis KM *et al.* (2011) Metagenomic analysis of a permafrost microbial community reveals a rapid response to thaw. *Nature*, **480**, 368–371.

Magoc T, Pabinger S, Canzar S *et al.* (2013) GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics*, **29**, 1718–1725.

Magurran AE (2004) *Measuring Biological Diversity*. Blackwell Science Ltd, Oxford, UK.

Maiques E, Ubeda C, Campoy S *et al.* (2006) beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *Journal of Bacteriology*, **188**, 2726–2729.

Manor O, Borenstein E (2015) MUSiCC: a marker genes based framework for metagenomic normalization and accurate profiling of gene abundances in the microbiome. *Genome Biology*, **16**, 53.

Marcinek H, Wirth R, Muscholl-Silberhorn A, Gauer M (1998) *Enterococcus faecalis* gene transfer under natural conditions in municipal sewage water treatment plants. *Applied and Environmental Microbiology*, **64**, 626–632.

Margulies M, Egholm M, Altman WE *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.

Martinez JL (2008) Antibiotics and antibiotic resistance genes in natural environments. *Science*, **321**, 365–367.

Martinez JL (2011) Bottlenecks in the transferability of antibiotic resistance from natural ecosystems to human bacterial pathogens. *Frontiers in Microbiology*, **2**, 265.

Martinez JL, Coque TM, Baquero F (2015) What is a resistance gene? Ranking risk in resistomes. *Nature Reviews Microbiology*, **13**, 116–123.

Martiny AC, Martiny JBH, Weihe C, Field A, Ellis JC (2011) Functional metagenomics reveals previously unrecognized diversity of antibiotic resistance genes in gulls. *Frontiers in Microbiology*, **2**, 238.

Matte-Tailliez O, Brochier C, Forterre P, Philippe H (2002) Archaeal phylogeny based on ribosomal proteins. *Molecular Biology and Evolution*, **19**, 631–639.

Mattner F, Bange F-C, Meyer E *et al.* (2012) Preventing the spread of multidrug-resistant gram-negative pathogens: recommendations of an expert panel of the German Society For Hygiene and Microbiology. *Deutsches Ärzteblatt International*, **109**, 39–45.

McArthur AG, Waglechner N, Nizam F *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrobial Agents and Chemotherapy*, **57**, 3348–3357.

McCarthy A, Chiang E, Schmidt ML, Denef VJ (2015) RNA preservation agents and nucleic acid extraction method bias perceived bacterial community composition. *PLoS ONE*, **10**, e0121659.

McDonald D, Price MN, Goodrich J *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME Journal*, **6**, 610–618.

McEachran AD, Blackwell BR, Hanson JD *et al.* (2015) Antibiotics, Bacteria, and Antibiotic Resistance Genes: Aerial Transport from Cattle Feed Yards via Particulate Matter. *Environmental Health Perspectives*, **123**, 337–343.

McHardy AC, Martín HG, Tsirigos A, Hugenholtz P, Rigoutsos I (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods*, **4**, 63–72.

McMurdie PJ, Holmes S (2014) Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Computational Biology*, **10**, e1003531.

Melnyk AH, Wong A, Kassen R (2015) The fitness costs of antibiotic resistance mutations. *Evolutionary Applications*, **8**, 273–283.

Metzker ML (2010) Sequencing technologies - the next generation. *Nature Reviews Genetics*, **11**, 31–46.

Meyer F, Paarmann D, Dsouza M *et al.* (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, **9**, 386.

Michael I, Rizzo L, McArdell CS *et al.* (2013) Urban wastewater treatment plants as hotspots for the release of antibiotics in the environment: a review. *Water Research*, **47**, 957–995.

Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF (2011) EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biology*, **12**, R44.

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. *Genomics*, **95**, 315–327.

Mishra RP, Oviedo-Orta E, Prachi P, Rappuoli R, Bagnoli F (2012) Vaccines and antibiotic resistance. *Current Opinion in Microbiology*, **15**, 596–602.

Molton JS, Tambyah PA, Ang BSP, Ling ML, Fisher DA (2013) The global spread of healthcare-associated multidrug-resistant bacteria: a perspective from Asia. *Clinical Infectious Diseases*, **56**, 1310–1318.

Moore AM, Munck C, Sommer MOA, Dantas G (2011) Functional metagenomic investigations of the human intestinal microbiota. *Frontiers in Microbiology*, **2**, 188.

Munck C, Albertsen M, Telke A *et al.* (2015) Limited dissemination of the wastewater treatment plant core resistome. *Nature Communications*, **6**, 8452.

Muziasari WI, Managaki S, Pärnänen K *et al.* (2014) Sulphonamide and trimethoprim resistance genes persist in sediments at Baltic Sea aquaculture farms but are not detected in the surrounding environment. *PLoS ONE*, **9**, e92702.

Myers EW, Sutton GG, Delcher AL *et al.* (2000) A whole-genome assembly of *Drosophila*. *Science*, **287**, 2196–2204.

Nagel M, Reuter T, Jansen A, Szekat C, Bierbaum G (2011) Influence of ciprofloxacin and vancomycin on mutation rate and transposition of IS256 in *Staphylococcus aureus. International Journal of Medical Microbiology*, **301**, 229–236.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012) MetaVelvet: an extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads. *Nucleic Acids Research*, **40**, e155.

Narzisi G, Mishra B (2011) Comparing *de novo* genome assembly: the long and short of it. *PLoS ONE*, **6**, e19175.

NCBI Resource Coordinators (2015) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, **43**, D6-17.

Nesme J, Cécillon S, Delmont TO *et al.* (2014) Large-scale metagenomic-based study of antibiotic resistance in the environment. *Current Biology*, **24**, 1096–1100.

Nielsen HB, Almeida M, Juncker AS *et al.* (2014) Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nature Biotechnology*, **32**, 822–828.

Nikaido H (2009) Multidrug resistance in bacteria. *Annual Review of Biochemistry*, **78**, 119–146.

Noble WS (2009) How does multiple testing correction work? *Nature Biotechnology*, **27**, 1135–1137.

Norman A, Hansen LH, Sørensen SJ (2009) Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 2275–2289.

Normark BH, Normark S (2002) Evolution and spread of antibiotic resistance. *Journal of Internal Medicine*, **252**, 91–106.

Norrby R, Powell M, Aronsson B *et al.* (2009) *The bacterial challenge: time to react*. European Centre for Disease Prevention and Control, Stockholm.

O'Brien TF (2002) Emergence, spread, and environmental effect of antimicrobial resistance: how use of an antimicrobial anywhere can increase resistance to any antimicrobial anywhere else. *Clinical Infectious Diseases*, **34 Suppl 3**, S78–84.

O'Leary NA, Wright MW, Brister JR *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, **44**, D733–45.

Oksanen J, Blanchet FG, Kindt R *et al.* (2011) vegan: Community Ecology Package.

Oliveira VDC, Rubio FG, Almeida MTG, Nogueira MCL, Pignatari ACC (2015) Trends of 9,416 multidrug-resistant Gram-negative bacteria. *Revista da Associação Médica Brasileira*, **61**, 244–249.

Oshlack A, Wakefield MJ (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, **4**, 14.

Östholm-Balkhed Å, Tärnberg M, Nilsson M *et al.* (2013) Travel-associated faecal colonization with ESBL-producing Enterobacteriaceae: incidence and risk factors. *Journal of Antimicrobial Chemotherapy*, **68**, 2144–2153.

O'Hara RB, Kotze DJ (2010) Do not log-transform count data. *Methods in Ecology and Evolution*, **1**, 118–122.

Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DGJ (2015) Co-occurrence of resistance genes to antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC Genomics*, **16**, 964.

Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DGJ (2014) BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Research*, **42**, D737–43.

Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*, **26**, 715–721.

Paulson JN, Stine OC, Bravo HC, Pop M (2013) Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, **10**, 1200–1202.

Peabody MA, Van Rossum T, Lo R, Brinkman FSL (2015) Evaluation of shotgun metagenomics sequence classification methods using *in silico* and *in vitro* simulated communities. *BMC Bioinformatics*, **16**, 363.

Pearson K (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, **50**, 157–175.

Pell J, Hintze A, Canino-Koning R *et al.* (2012) Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 13272–13277.

Pendleton JN, Gorman SP, Gilmore BF (2013) Clinical relevance of the ESKAPE pathogens. *Expert Review of Anti-infective Therapy*, **11**, 297–308.

Peng Y, Leung HCM, Yiu SM, Chin FYL (2011) Meta-IDBA: a *de Novo* assembler for metagenomic data. *Bioinformatics*, **27**, i94–101.

Perron GG, Gonzalez A, Buckling A (2007) Source-sink dynamics shape the evolution of antibiotic resistance and its pleiotropic fitness cost. *Proceedings Biological Sciences*, **274**, 2351–2356.

Perron GG, Inglis RF, Pennings PS, Cobey S (2015a) Fighting microbial drug resistance: a primer on the role of evolutionary biology in public health. *Evolutionary Applications*, **8**, 211–222.

Perron GG, Whyte L, Turnbaugh PJ *et al.* (2015b) Functional characterization of bacteria isolated from ancient arctic soil exposes diverse resistance mechanisms to modern antibiotics. *PLoS ONE*, **10**, e0069533.

Pevzner PA, Tang H, Waterman MS (2001) An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 9748–9753.

Pharmaceutical, Biotechnology and Diagnostics Industries (2016) *Declaration by the Pharmaceutical, Biotechnology and Diagnostics Industries on Combating Antimicrobial Resistance*. World Economic Forum, Davos, Switzerland. *http://amr-review.org/*

Philippot L, Andersson SGE, Battin TJ *et al.* (2010) The ecological coherence of high bacterial taxonomic ranks. *Nature Reviews Microbiology*, **8**, 523–529.

Pinard R, de Winter A, Sarkis GJ *et al.* (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics*, **7**, 216.

Poirel L, Carrër A, Pitout JD, Nordmann P (2009) Integron mobilization unit as a source of mobility of antibiotic resistance genes. *Antimicrobial Agents and Chemotherapy*, **53**, 2492–2498.

Pop M (2009) Genome assembly reborn: recent computational challenges. *Briefings in Bioinformatics*, **10**, 354–366.

Port JA, Wallace JC, Griffith WC, Faustman EM (2012) Metagenomic profiling of microbial composition and antibiotic resistance determinants in Puget Sound. *PLoS ONE*, **7**, e48000.

Pruden A (2014) Balancing water sustainability and public health goals in the face of growing concerns about antibiotic resistance. *Environmental Science & Technology*, **48**, 5–14.

Pruden A, Arabi M, Storteboom HN (2012) Correlation between upstream human activities and riverine antibiotic resistance genes. *Environmental Science & Technology*, **46**, 11541–11549.

Pruden A, Larsson DGJ, Amézquita A *et al.* (2013) Management options for reducing the release of antibiotics and antibiotic resistance genes to the environment. *Environmental Health Perspectives*, **121**, 878–885.

Pruden A, Pei R, Storteboom H, Carlson KH (2006) Antibiotic resistance genes as emerging contaminants: studies in northern Colorado. *Environmental Science & Technology*, **40**, 7445–7450.

Prudhomme M, Attaiech L, Sanchez G, Martin B, Claverys J-P (2006) Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae*. *Science*, **313**, 89–92.

Qin J, Li R, Raes J *et al.* (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**, 59–65.

Quinlan EL, Nietch CT, Blocksom K *et al.* (2011) Temporal Dynamics of Periphyton Exposed to Tetracycline in Stream Mesocosms. *Environmental Science & Technology*, **45**, 10684–10690.

Rao GG (1998) Risk factors for the spread of antibiotic-resistant bacteria. *Drugs*, **55**, 323–330.

Rapaport F, Khanin R, Liang Y *et al.* (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, **14**, R95.

Rasko DA, Webster DR, Sahl JW *et al.* (2011) Origins of the *E. coli* strain causing an outbreak of hemolytic-uremic syndrome in Germany. *The New England Journal of Medicine*, **365**, 709–717.

Review on Antimicrobial Resistance (2014) *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations* (J O'Neill, Ed,). *http://amr-review.org/*

Review on Antimicrobial Resistance (2015) *Antimicrobials in agriculture and the environment: Reducing unnecessary use and waste* (J O'Neill, Ed,). *http://amr-review.org/*

Riesenfeld CS, Goodman RM, Handelsman J (2004) Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*, **6**, 981–989.

Rizzo L, Manaia C, Merlin C *et al.* (2013) Urban wastewater treatment plants as hotspots for antibiotic resistant bacteria and genes spread into the environment: a review. *The Science of the Total Environment*, **447**, 345–360.

Robertson G, Schein J, Chiu R *et al.* (2010) *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, **7**, 909–912.

Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, **11**, R25.

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rodriguez-Brito B, Rohwer F, Edwards RA (2006) An application of statistics to comparative metagenomics. *BMC Bioinformatics*, **7**, 162.

Rolain J-M (2013) Food and human gut as reservoirs of transferable antibiotic resistance encoding genes. *Frontiers in Microbiology*, **4**, 173.

Roslev P, Bukh AS (2011) State of the art molecular markers for fecal pollution source tracking in water. *Applied Microbiology and Biotechnology*, **89**, 1341–1355.

Rutgersson C, Fick J, Marathe N *et al.* (2014) Fluoroquinolones and *qnr* Genes in Sediment, Water, Soil, and Human Fecal Flora in an Environment Polluted by Manufacturing Discharges. *Environmental Science & Technology*, **48**, 7825–7832.

Salipante SJ, Roach DJ, Kitzman JO *et al.* (2015) Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. *Genome Research*, **25**, 119–128.

Salyers AA, Amábile-Cuevas CF (1997) Why are antibiotic resistance genes so resistant to elimination? *Antimicrobial Agents and Chemotherapy*, **41**, 2321–2325.

Salyers AA, Gupta A, Wang Y (2004) Human intestinal bacteria as reservoirs for antibiotic resistance genes. *Trends in Microbiology*, **12**, 412–416.

Salyers AA, Shoemaker NB (2006) Reservoirs of antibiotic resistance genes. *Animal Biotechnology*, **17**, 137–146.

Salzberg SL, Phillippy AM, Zimin A *et al.* (2012) GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, **22**, 557–567.

San Millan A, Santos-Lopez A, Ortega-Huedo R *et al.* (2015) Small-plasmid-mediated antibiotic resistance is enhanced by increases in plasmid copy number and bacterial fitness. *Antimicrobial Agents and Chemotherapy*, **59**, 3335–3341.

Schlenker E (2016) Tips and Tricks for Successful Application of Statistical Methods to Biological Data. *Methods in Molecular Biology*, **1366**, 271–285.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Schmieder R, Edwards R (2012) Insights into antibiotic resistance through metagenomic approaches. *Future Microbiology*, **7**, 73–89.

Schmieder R, Lim YW, Edwards R (2012) Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics*, **28**, 433–435.

Scholz MB, Lo C-C, Chain PSG (2012) Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current Opinion in Biotechnology*, **23**, 9–15.

Scholz MB, Lo C-C, Chain PSG (2014) Improved Assemblies Using a Source-Agnostic Pipeline for MetaGenomic Assembly by Merging (MeGAMerge) of Contigs. *Scientific Reports*, **4**, 6480.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, **28**, 1086–1092.

Segata N, Waldron L, Ballarini A *et al.* (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods*, **9**, 811–814.

Segawa T, Takeuchi N, Rivera A *et al.* (2012) Distribution of antibiotic resistance genes in glacier environments. *Environmental Microbiology Reports*, **5**, 127–134.

Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: a community resource for metagenomics. *PLoS Biology*, **5**, e75.

Shade A, Peter H, Allison SD *et al.* (2012) Fundamentals of microbial community resistance and resilience. *Frontiers in Microbiology*, **3**, 417.

Shannon C, Weaver W (1949) *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Shentu J-L, Zhang K, Shen D-S, Wang M-Z, Feng H-J (2015) Effect from low-level exposure of oxytetracycline on abundance of tetracycline resistance genes in arable soils. *Environmental Science and Pollution Research International*, **22**, 13102–13110.

Shi P, Jia S, Zhang X-X *et al.* (2013) Metagenomic insights into chlorination effects on microbial antibiotic resistance in drinking water. *Water Research*, **47**, 111–120.

Simpson EH (1949) Measurement of Diversity. *Nature*, **163**, 688–688.

Simpson JT, Durbin R (2012) Efficient *de novo* assembly of large genomes using compressed data structures. *Genome Research*, **22**, 549–556.

Simpson JT, Wong K, Jackman SD *et al.* (2009) ABySS: a parallel assembler for short read sequence data. *Genome Research*, **19**, 1117–1123.

Smillie CS, Smith MB, Friedman J *et al.* (2011) Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, **480**, 241–244.

Smooker PM, Whisstock JC, Irving JA *et al.* (2000) A single amino acid substitution affects substrate specificity in cysteine proteinases from *Fasciola hepatica*. *Protein Science*, **9**, 2567–2572.

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, **3**, 3.

So AD, Shah TA, Roach S, Ling Chee Y, Nachman KE (2015) An Integrated Systems Approach is Needed to Ensure the Sustainability of Antibiotic Effectiveness for Both Humans and Animals. *The Journal of Law, Medicine & Ethics*, **43 Suppl 3**, 38–45.

Soergel DAW, Dey N, Knight R, Brenner SE (2012) Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *The ISME Journal*, **6**, 1440–1444.

Sohn MB, Du R, An L (2015) A robust approach for identifying differentially abundant features in metagenomic samples. *Bioinformatics*, **31**, 2269–2275.

Sommer MOA, Church GM, Dantas G (2010) The human microbiome harbors a diverse reservoir of antibiotic resistance genes. *Virulence*, **1**, 299–303.

Sommer MOA, Dantas G, Church GM (2009) Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*, **325**, 1128–1131.

Staden R (1979) A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, **6**, 2601–2610.

Stokes HW, Gillings MR (2011) Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. *FEMS Microbiology Reviews*, **35**, 790–819.

Sunagawa S, Mende DR, Zeller G *et al.* (2013) Metagenomic species profiling using universal phylogenetic marker genes. *Nature Methods*, **10**, 1196–1199.

Sütterlin S, Edquist P, Sandegren L *et al.* (2014) Silver resistance genes are overrepresented among *Escherichia coli* isolates with CTX-M production. *Applied and Environmental Microbiology*, **80**, 6863–6869.

Sydnor ERM, Perl TM (2011) Hospital epidemiology and infection control in acute-care settings. *Clinical Microbiology Reviews*, **24**, 141–173.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, **21**, 2045–2050.

Tamminen M, Karkman A, Lõhmus A *et al.* (2011) Tetracycline resistance genes persist at aquaculture farms in the absence of selection pressure. *Environmental Science & Technology*, **45**, 386–391.

Tängdén T, Cars O, Melhus A, Löwdin E (2010) Foreign travel is a major risk factor for colonization with *Escherichia coli* producing CTX-M-type extended-spectrum beta-lactamases: a prospective study with Swedish volunteers. *Antimicrobial Agents and Chemotherapy*, **54**, 3564–3568.

Tello A, Austin B, Telfer TC (2012) Selective Pressure of Antibiotic Pollution on Bacteria of Importance to Public Health. *Environmental Health Perspectives*, **120**, 1100–1106.

Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Reviews Microbiology*, **3**, 711–721.

Toleman MA, Spencer J, Jones L, Walsh TR (2012) blaNDM-1 is a chimera likely constructed in *Acinetobacter baumannii*. *Antimicrobial Agents and Chemotherapy*, **56**, 2773–2776.

Tomayko JF, Rex JH, Tenero DM, Goldberger M, Eisenstein BI (2014) The challenge of antimicrobial resistance: new regulatory tools to support product development. *Clinical Pharmacology and Therapeutics*, **96**, 166–168.

Torres-Barceló C, Kojadinovic M, Moxon R, MacLean RC (2015) The SOS response increases bacterial fitness, but not evolvability, under a sublethal dose of antibiotic. *Proceedings Biological Sciences*, **282**, 20150885.

Torres-Cortés G, Millán V, Ramírez-Saad HC *et al.* (2011) Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environmental Microbiology*, **13**, 1101–1114.

Udikovic-Kolic N, Wichmann F, Broderick NA, Handelsman J (2014) Bloom of resident antibiotic-resistant bacteria in soil following manure fertilization. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 15202–15207.

UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Research*, **43**, D204–12.

van Hoek AHAM, Mevius D, Guerra B *et al.* (2011) Acquired antibiotic resistance genes: an overview. *Frontiers in Microbiology*, **2**, 203.

Vázquez-Castellanos JF, García-López R, Pérez-Brocal V, Pignatelli M, Moya A (2014) Comparison of different assembly and annotation tools on analysis of simulated viral metagenomic communities in the gut. *BMC Genomics*, **15**, 37.

Venter JC, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, **304**, 66–74.

Verbeken G, Huys I, Ceulemans C *et al.* (2016) Bacteriophage therapy: Fast-forward to the past lessons identified from the advanced therapy regulation. *Burns : Journal of the International Society for Burn Injuries*, **42**, 11–12.

Verbeken G, Huys I, Pirnay J-P *et al.* (2014) Taking bacteriophage therapy seriously: a moral argument. *BioMed Research International*, **2014**, 621316.

Větrovský T, Baldrian P (2013) The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE*, **8**, e57923.

Vogwill T, MacLean RC (2015) The genetic basis of the fitness costs of antimicrobial resistance: a meta-analysis approach. *Evolutionary Applications*, **8**, 284–295.

Wales A, Davies R (2015) Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. *Antibiotics*, **4**, 567–604.

Walsh C (2003) *Antibiotics: Actions, Origins, Resistance*. ASM Press.

Walsh TR, Weeks J, Livermore DM, Toleman MA (2011) Dissemination of NDM-1 positive bacteria in the New Delhi environment and its implications for human health: an environmental point prevalence study. *The Lancet Infectious Diseases*, **11**, 355–362.

Wang Q, Fish JA, Gilman M *et al.* (2015) Xander: employing a novel method for efficient gene-targeted metagenomic assembly. *Microbiome*, **3**, 32.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57–63.

Warren RL, Sutton GG, Jones SJM, Holt RA (2007) Assembling millions of short DNA sequences using SSAKE. *Bioinformatics*, **23**, 500–501.

Wattal C, Goel N (2014) Tackling antibiotic resistance in India. *Expert Review of Anti-infective Therapy*, **12**, 1427–1440.

Weinstein RA (2001) Controlling antimicrobial resistance in hospitals: infection control and use of antibiotics. *Emerging Infectious Diseases*, **7**, 188–192.

Wellington EM, Boxall AB, Cross P *et al.* (2013) The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *The Lancet Infectious Diseases*, **13**, 155–165.

White JR, Nagarajan N, Pop M (2009) Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Computational Biology*, **5**, e1000352.

Whitlock JE, Jones DT, Harwood VJ (2002) Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. *Water Research*, **36**, 4273–4282.

WHO (2014) *Antimicrobial resistance: global report on surveillance 2014*. World Health Organization, Geneva, Switzerland.

Wichmann F, Udikovic-Kolic N, Andrew S, Handelsman J (2014) Diverse antibiotic resistance genes in dairy cow manure. *mBio*, **5**, e01017.

Wiedenbeck J, Cohan FM (2011) Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, **35**, 957–976.

Wilson ME, Chen LH (2012) NDM-1 and the Role of Travel in Its Dissemination. *Current Infectious Disease Reports*, **14**, 213–226.

Wong K, Fong T-T, Bibby K, Molina M (2012) Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environment International*, **45**, 151–164.

Wood DE, Salzberg SL (2014) Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, **15**, R46.

Wooley JC, Godzik A, Friedberg I (2010) A primer on metagenomics. *PLoS Computational Biology*, **6**, e1000667.

Wright GD (2010) Antibiotic resistance in the environment: a link to the clinic? *Current Opinion in Microbiology*, **13**, 589–594.

Yang Y, Li B, Ju F, Zhang T (2013) Exploring variation of antibiotic resistance genes in activated sludge over a four-year period through a metagenomic approach. *Environmental Science & Technology*, **47**, 10197–10205.

Yang Y, Li B, Zou S, Fang HHP, Zhang T (2014) Fate of antibiotic resistance genes in sewage treatment plant revealed by metagenomic approach. *Water Research*, **62**, 97–106.

Yilmaz P, Parfrey LW, Yarza P *et al.* (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, **42**, D643–8.

Yong D, Toleman MA, Giske CG *et al.* (2009) Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in *Klebsiella pneumoniae* sequence type 14 from India. *Antimicrobial Agents and Chemotherapy*, **53**, 5046–5054.

Yooseph S, Nealson KH, Rusch DB *et al.* (2010) Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature*, **468**, 60–66.

Zaffiri L, Gardner J, Toledo-Pereyra LH (2012) History of antibiotics. From salvarsan to cephalosporins. *Journal of Investigative Surgery*, **25**, 67–77.

Zankari E, Hasman H, Cosentino S *et al.* (2012) Identification of acquired antimicrobial resistance genes. *Journal of Antimicrobial Chemotherapy*, **67**, 2640–2644.

Zepeda Mendoza ML, Sicheritz-Ponten T, Gilbert MTP (2015) Environmental genes and genomes: understanding the differences and challenges in the approaches and software for their analyses. *Briefings in Bioinformatics*, **16**, 745–758.

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Research*, **18**, 821–829.

Zhang T, Zhang X-X, Ye L (2011) Plasmid metagenome reveals high levels of antibiotic resistance genes and mobile genetic elements in activated sludge. *PLoS ONE*, **6**, e26041.

Zhang Y, Sun Y, Cole JR (2014) A Scalable and Accurate Targeted Gene Assembly Tool (SAT-Assembler) for Next-Generation Sequencing Data. *PLoS Computational Biology*, **10**, e1003737.

Zhu Y-G, Johnson TA, Su J-Q *et al.* (2013) Diverse and abundant antibiotic resistance genes in Chinese swine farms. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 3435–3440.