THESIS FOR THE DEGREE OF LICENTIATE OF ENGINEERING

Identification of novel antibiotic resistance genes through large-scale data analysis

Fanny Berglund

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
Göteborg, Sweden 2017

Identification of novel antibiotic resistance genes through large-scale data analysis
Fanny Berglund
Göteborg 2017

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg
SE-412 96 Göteborg
Sweden
Telephone +46 (0)31 772 1000

Typeset with LaTeX
Department of Mathematical Sciences
Printed in Göteborg, Sweden 2017

Identification of novel antibiotic resistance genes through large-scale data analysis

Fanny Berglund

Division of Applied Mathematics and Statistics
Department of Mathematical Sciences
Chalmers University of Technology and University of Gothenburg

Abstract

Antibiotic resistance is increasing worldwide, and is considered a serious threat to public health by e.g. the World Health Organization. Antibiotic resistance genes are hypothesized to originate from harmless bacteria in and around us, from where they are horizontally transfered into human pathogens. It is therefore of great importance to explore human-associated and environmental bacterial communities to identify novel antibiotic resistance genes before they reach clinical settings. The three papers presented in this thesis aim to identify novel antibiotic resistance genes in large genomic and metagenomic datasets. In *paper I*, the aim was to identify novel genes of the clinically important subclass B1 metallo-β-lactamases. By analyzing whole bacterial genomes as well as metagenomes from environmental and human-associated bacterial communities, 76 novel putative B1 genes were predicted. Twenty-one of these were selected for experimental validation, whereof 18 expressed the predicted phenotype in *E. coli*. Phylogentic analysis revealed that the novel genes formed 59 previously undescribed gene families. In *paper II*, a large volume of genomic and metagenomic data was searched for novel plasmid-mediated quinolone resistance (*qnr*) genes. In total, 611 *qnr* genes were predicted, of which 20 were putative novel. Nine of these were experimentally tested in *E. coli*, whereof eight expressed the predicted phenotype. In *paper III*, a new method for identification and reconstruction of novel antibiotic resistance genes from fragmented metagenomic data was presented. The method is based on gene specific models, which are optimized for a high sensitivity and specificity. The method is furthermore computationally efficient and can be applied to any class of resistance genes. The results of this thesis provides a deeper insight to the diversity and evolutionary history of two types of clinically relevant antibiotic resistance genes. It also provides new methods for efficient and reliable identification of novel resistance genes in fragmented metagenomic data.

Keywords: antibiotic resistance, metagenomics, big data, carbapenemases, hidden Markov model.

List of papers

The licentiate thesis includes the following papers.

I. **Berglund, F.**, Marathe, N., Österlund, T., Razavi, M., Bengtsson-Palme, J., Kotsakis, S., Flach, C.F., Larsson, D.G.J., Kristiansson, E. (2017). Identification of 76 novel metallo-β-lactamases through large-scale screening of genomic and metagenomic data. *Manuscript.*

II. Boulund, F., **Berglund, F.**, Bengtsson-Palme, J., Larsson, D.G.J., Kristiansson, E. (2017). Computational discovery and validation of novel fluoroquinolone resistance genes in public metagenomic data sets. *Manuscript*

III. **Berglund, F.**, Österlund, T., Boulund, F., Larsson, D.G.J., Kristiansson, E. (2017). A computational method for identification of novel antibiotic resistance genes in metagenomes. *Manuscript*

Publications not included in this licentiate thesis:

- Gustavsson, K., **Berglund, F.**, Jonsson, P.R., Mehlig, B. (2016). Preferential Sampling and Small-Scale Clustering of Gyrotactic Microswimmers in Turbulence. *Physical review letters, 116*(10)

Author contributions

I. Participated in study design, collected the data for- and created the model, collected the genomic and metagenomic data, developed and implemented the analysis pipeline, performed the data analysis, performed the clustering, created the phylogenetic tree, analyzed the results, drafted and edited the manuscript.

II Participated in study design, collected the data, implemented the analysis pipeline, performed the analysis of metagenomic data, performed the clustering, and edited the manuscript.

III Participated in study design, developed and implemented the method, collected reference sequences and created the three models, did the optimization of the models, analyzed the metagenomic data, analyzed the results, drafted and edited the manuscript.

Acknowledgements

First of all I would like to thank my supervisor Erik Kristiansson. Your endless positivity, encouragement and many ideas has guided me through my first years. Tobias Österlund, my excellent co-supervisor, who is always there when I need to explore ideas or discuss something, be it work related or not. I would also like to thank my second co-supervisor Joakim Larsson, who always gives good feedback and support. My thanks also goes to my examiner Olle Nerman, for the good advises you have given me.

I would like to thank everyone I have worked with at Sahlgrenska for discussing the many biological components of my work and for giving me valuable input, Nachiket Marathe, Stathis Kothsakis, Mohammad Razavi, Carl-Fredik Flach and Johan Palme. Thank you Fredrik Boulund for our nice collaboration with the research and the server, I learned a lot from you. Thank you to everyone in our bioinformatics group, Anna Johnning, Mariana Pereira, Viktor Jonsson, Anna Rehammar and Marina Axelson-Fisk. It feels really good to be surrounded by such smart and inspiring people. My room-mate Sandra Eriksson-Barman for being such a good listener, and Olle Elias, Ivar Simonsson and Jonatan Kallus for being a great lunch and fika company. I also want to thank everyone else at the mathematics department for creating a friendly atmosphere. Finally I would like to thank my family and friends for always supporting me.

Contents

1 Background

All today known living organisms depend on their genetic material to function and grow. The genetic information is encoded in two forms of nucleic acids, where deoxyribonucleic acid (DNA) acts as repositories and ribonucleic acid (RNA) act as transmitters. DNA molecules generally contain four types of nucleotides; adenine (A), thymine (T), guanine (G) and cytosine (C), while RNA has the same nucleotides with the exception of the replacement of T for uracil (U). Depending on how these nucleotides are organized, they decide an organism's functions and traits. In each cell in every organism, a copy of the complete set of genetic information is stored and is called the genome. The size of genomes is highly variable with the smallest viruses having a few thousand bases to the more than three billion bases of the human genome, and even more in some species. Most genomes consists of a large number of genes, which are parts of the genome that contains information about the synthesis of RNA and molecular proteins Mathews et al. (2000).

In the early 50's, when it was discovered that proteins were constructed of amino acids in what seemed to be defined orders (Sanger, 1960), the search for the blueprint of the protein creation started. Although the existence of the DNA was proposed shortly thereafter (Watson et al., 1953), it wasn't until 1965 the first nucleic acid molecule in the form of tRNA was sequenced (Holley et al., 1965). The first DNA sequenced was from the bacteriophage λ, which was completed in 1971 and consisted of 12 bases (Wu and Taylor, 1971), and the first whole genome sequenced was another bacteriophage, ϕX with 5375 bases, in 1978. The length of the sequences continued to increase with the first human chromosome sequenced in 1999 (Dunham et al., 1999) and in 2001 the first draft versions of the human genomes 3.2 billion bases were published (Venter et al., 2001; Lander et al., 2001). The advancements in the sequencing techniques had until then been achieved mainly with the Sanger method, also referred to as the "first-generation" sequencing, but during the last decade several new sequencing techniques have emerged. A big breakthrough was the invention of

the high-throughput second-generation or "next-generation" sequencing (NGS) techniques. With the second-generation sequencing came the ability to produce millions of sequence reads in parallel, speeding up the process and reducing the cost of each read substantially. Although the produced sequence reads with second-generation are typically not as long and not as accurate as those produced with Sanger technology, the great throughput of reads makes the coverage of each DNA fragment high and the final sequences more accurate since each DNA nucleotide is sequences several times on average. Recently, the next-generation sequencing technologies have been further developed with the aim to produce even longer reads while maintaining the low price per read achieved with the second-generation technique (van Dijk et al., 2014). The fast pace of the development of sequencing techniques have created an enormous amount of sequence data, and today there are around 86 000 sequenced bacterial genomes publically available (NCBI, 2017), compared to only 300 sequenced bacterial genomes 10 years ago (Binnewies et al., 2006).

The massive amount of genomic data has given us unprecedented opportunities to study all life forms on earth. The applications are numerous, ranging from evolutionary studies to disease preventions. With even better and faster sequencing techniques, together with new algorithms to handle the data, the number of applications will most likely be endless.

1.1 Metagenomics

It is estimated that there are over 10^{30} individual bacterial cells (Whitman et al., 1998) and between 10^7 to 10^9 bacterial species on earth (Curtis et al., 2002; Dykhuizen, 1998). Historically the study of microorganisms has been focused on single species, but the vast majority of all microorganisms cannot be cultured by standard techniques (Hugenholtz et al., 1998). The uncultivable microorganisms represent many diverse organisms living in several unique communities and they are often distantly related to the cultivable ones (Riesenfeld et al., 2004b). It was therefore important to develop culture-independent methods to understand the ecological role of these communities, their genetic diversity and population structure. To tackle this challenge, a technique which today is called *functional metagenomics* was developed (Handelsman et al., 1998). DNA was taken directly from environmental communities, instead of from a single organism, and fragmented into pieces. The DNA fragments were then inserted into cultivable bacteria which were grown under certain conditions such as under antibiotic exposure, to assess the community for specific genes. The collective genome of all organisms present in a community was named *metagenome*.

However, functional metagenomics has some drawbacks; the procedure is time-consuming and not every gene can be expressed in a cultivable host. Furthermore, when searching for a specific type of gene many fragments must be created to ensure that the complete gene is captured by the inserted fragments. But with cheaper and faster sequencing techniques the sequencing of uncultured microorganisms directly from their community became possible (Heather and Chain, 2016). The technique, sometimes called *shotgun metagenomics*, is the direct isolation and sequencing of genomic DNA from a bacterial community through high-throughput DNA sequencing (Wooley et al., 2010). Although the many benefits with metagenomics, it comes with some complications. Compared to single organism sequencing when the whole genome of a specific species is sequenced, the metagenomic data often contain DNA from a large number of species, and due to the high diversity often present in and in-between microbial communities, metagenomes are often undersampled. The data is furthermore often highly fragmented, with fragments sometimes as short as 75 bases, depending on which sequencing technique used. The result is a data set which is hard to reconstruct through sequence assembly, and often it is not possible to reconstruct the complete genomes of the organisms present in the sampled community. The analysis of metagenomic data is therefore challenging, and demands specially designed methods and algorithms.

During the last years, several big projects have been conducted within the metagenomic area, including the study of the gut microbiome of 124 European individuals which resulted in the first human gut gene catalog (Qin et al., 2010). Then came the Human Microbiome Project (Consortium et al., 2012) where scientists discovered, among other things, that the microbial communities between body sites of healthy individuals differed remarkably, and the Tara Ocean project where 7.2 trillion bases of metagenomic data has been sequenced, leading to new picture of the diversity within the ocean with more than 40 million non-redundant genes discovered (Armbrust and Palumbi, 2015). The impact of metagenomics will likely continue to increase, as the amount of metagenomic data continues to accumulate in publically available databases.

1.2 Antibiotic resistance

Antibiotics are substances that can kill or inhibit the growth of bacteria, and have since its discovery saved millions of people with life-threatening bacterial infections and facilitated major improvements in medicine and surgery. The antibiotic can either be a natural product, produced by specific microorganisms,

it can be synthetically constructed, or a combination thereof. During the past 60 years, most antibiotics has been the result of natural production by bacteria or fungi, with a few exceptions of synthetically constructed antibiotics (Walsh, 2003). There are several classed of antibiotics, but the majority of them were discovered between 1940 and 1962, and since then only two new classes have been introduced (Coates et al., 2011).

Antibiotic resistance is the ability of bacteria to withstand the effects of antibiotics, and as a result, medicines previously used to treat infections have become ineffective. Bacteria can withstand antibiotics through different resistance mechanisms which can be classified as intrinsic, acquired or adaptive. Bacteria with intrinsic resistance often have some inherited characteristics which causes them to be unaffected by antibiotics, and this is a feature of all members of the species. This characteristic does not change over time, and can for example be the outer membrane of many Gram-negative bacteria which is impermeable to certain types of antibiotics (Carlos, 2015). In contrast to the intrinsic resistance, bacteria can also acquire resistance via mutations and horizontal gene transfer. Through mobile genetic elements such as plasmids and transposons, these resistance genes can move between bacterial cells and species (Stokes and Gillings, 2011). The acquired resistance can, once obtained, be inherited to daughter cells of the bacteria. The third variant of resistance is the adaptive resistance, or stress response, where sudden changes in the environment of a bacterial community may trigger defense systems which can limit the permeability of the outer membrane, or by overexpressing efflux pumps which leads to a decreased accumulation of antibiotics, and thus a decreased susceptibility (Carlos, 2015).

The first antibiotic in commercial use was the penicillin in the 1940-ies, but almost immediate after its release, a bacterial produced enzyme with the ability to hydrolyze it was discovered. (Abraham and Chain, 1940). New types of antibiotics were developed in response to the increasing resistance, but as soon as a new drug was on the market, a new type of resistance was discovered. Today resistance has been recognized to almost every developed antibiotics (Ventola, 2015) Over time the antibiotic resistant bacteria furthermore have undergone the development from resistance to single classes of antibiotics to being multi-drug resistant, often through horizontal gene transfer. In clinical settings, frequently encountered resistant bacteria are the extended-spectrum β-lactamase (ESBL) positive Enterobacteriaceae, while other common types are the Methicillinresistant *Staphylococcus aereus* (MRSA) and the vancomycin-resistant enterococci (VRE) (Cantas et al., 2013).

As a last resort treatment for a patient infected with a multi-resistant bacteria the β-lactam *carbapenems* are often used. However, recent studies show that

resistance even to this antibiotic is emerging worldwide. The resistance to carbapenems is to a large extent due to the expression of carbapenemases, a class of enzymes where the majority have broad-spectrum substrate profiles (Papp-Wallace et al., 2011). An especially worrisome carbapenemase is the acquired NDM-1, which was first discovered in 2009 (Yong et al., 2009). Only in a few years, the gene has rapidly spread through pathogens and has now been found on several geographical locations and isolated in numerous bacterial species (Walsh et al., 2011).

1.3 Antibiotic resistance and the environment

Antibiotics have been produced naturally by environmental microbial communities since ancient times, with some estimates pointing back to two billion years ago (Hall and Barlow, 2004). Consequently, as a result of evolution, antibiotic resistance have likely been around for a similarly long time (D'Costa et al., 2011). This would have given the microbes plenty of time to develop a large and diverse set of resistance genes, with most of them still being undiscovered.

It is clear that the increase of antibiotic usage in human and veterinary settings, together with the greater movement of people and animals, has led to the increased prevalence and spread of antibiotic resistance bacteria (Martínez, 2008; Cantas et al., 2013). But human pathogens are, in most cases, not originally carriers of antibiotic resistance. Instead, the antibiotic resistance genes encountered in clinical settings are hypothesized to have originated from the environment (Walsh, 2013). Resistance genes similar and identical to those in pathogenic bacteria have been discovered in various environmental communities (Riesenfeld et al., 2004a; Canton, 2009; Boulund et al., 2012), including pristine environments such as glaciers (Segawa et al., 2013) and 30 000 years old permafrost samples (D'Costa et al., 2011). Furthermore, environmental and commensal bacterial communities have been shown to harbor a large diversity of antibiotic resistance genes, among which many have not been identified in clinical settings (Sommer et al., 2009; Forsberg et al., 2012; Wichmann et al., 2014). It has further been shown that selection pressures, such as antibiotic exposure, may enrich the abundance and diversity of resistance genetic elements in these communities (Gillings and Stokes, 2012). It is therefore likely that the environment has and will continue to act as a reservoir for antibiotic resistance genes that can be spread to pathogenic bacteria (Allen et al., 2010).

Historically, the studies of antibiotic resistance have been conducted in clinical settings with pathogenic bacteria in focus. However, to unravel the real diversity

and abundance of resistance genes it is necessary to turn the investigation to the environmental and commensal bacterial communities (Bengtsson-Palme and Larsson, 2015). Because the majority of bacteria inhabiting these communities are uncultivable under standard laboratory conditions, the invention of metagenomics has provided an unprecedented opportunity to explore these environments under unbiased conditions. But it is not straightforward to search for resistance genes in metagenomic data. The data is often large and highly fragmented with reads from several species, so there is a high probability that genes existing in low abundance will pass undetected. Secondly, resistance genes are often located on mobile elements, such as plasmids and transposons, which can be situated in various contexts, making the assembly of these genes notoriously troublesome (Bengtsson-Palme et al., 2014). New methods to analyze metagenomes are therefore needed in order to identify novel resistance genes before they spread to pathogens, and to increase the knowledge of the diversity of specific classes of antibiotic resistance genes.

2 Results

The aim of this thesis is to identify previously uncharacterized antibiotic resistance genes to further elucidate their abundance, their diversity and their origin. The first two papers, *paper I* and *paper II*, were aimed at identifying novel subclass B1 metallo-β-lactamases and plasmid-mediated *qnr* genes respectively. *Paper III* aimed at developing a method to facilitate identification of novel antibiotic resistance genes in fragmented metagenomic data. This chapter provides an introduction to the three papers together with a summary of the overall aims and main findings of each paper.

2.1 Introduction to Paper I

Carbapenems are important broad-spectrum antibiotics often used as a last resort treatment for patients infected with multi-resistant bacteria. Carbapenem resistance is caused by carbapenemases, enzymes that often are able to hydrolyze all known β-lactams in addition to carbapenems. During the last years, the resistance towards carbapenems has increased rapidly in many regions of the world (Papp-Wallace et al., 2011), and carbapenem resistance genes, that up until a couple of years ago had never been seen, is now detected in pathogens basically all over the world. The majority of the mobile carbapenem resistance genes have been identified in clinical settings, but it is hypothesized that these genes originated from environmental bacteria, and from there have made the move into pathogens. Many acquired and clinically relevant carbapenem resistance genes such as VIM, IMP and NDM belong to the class metallo-β-lactamases. This class is further divided based on molecular structure into the three subclasses B1, B2 and B3. In *paper I*, the aim was to identify previously uncharacterized subclass B1 metallo-β-lactamases (B1BL) to obtain a more detailed picture of the origin and diversity of this important subclass. Using a new developed computational pipeline we screened more than 5 terabases of metagenomic data

from human and environmental bacterial communities, as well as all bacterial genomes and plasmids available in the NCBI GenBank database. In order to achieve high accuracy, the classification in the pipeline was done with a Hidden Markov model optimized for finding previously uncharacterized B1BL.

In total, 76 novel B1BL genes were identified and when clustered together with all previously reported B1BL genes using a sequence similarity cut-off of 70%, they formed 59 novel gene families. A phylogenetic tree was created based on one representative sequence from each of the 59 novel gene families together with the previously reported B1BL genes. Analysis of the resulting tree showed that subclass B1 genes can be organized into five groups, mainly defined by the taxonomy of the host. We furthermore noticed that all except one previously identified mobile B1BL clustered together with chromosomally encoded genes of the phylum Proteobacteria. Of the identified novel genes, 21 were selected for experimental verification where the genes where synthetically constructed and inserted into an *Escherichia coli* host. A CarbaNP test (Nordmann et al., 2012) was conducted and showed that 18 of the tested genes had a carbapenemase activity. The results from this paper significally extends the number of identified subclass B1 metallo-β-lactamases, and provide a much more detailed picture of their diversity and evolutionary history.

2.2 Introduction to Paper II

The fully synthetic antibiotic class quinolones were taken into clinical use in 1962, and later in the 1980s an addition of a fluorine on the quinolone yielded the fluoroquinolones. Since the antibiotic was not derived from natural components, it was considered unlikely that fluoroquinolone resistance would emerge. Nevertheless, in 1998 the first plasmid-mediated quinolone resistance (*qnr*) gene was discovered. Since then, the *qnr* genes have rapidly spread and are now distributed globally in many bacterial genera. As with most other antibiotic resistance genes, the *qnr* genes are assumed to have originated from the environment. Although several families of *qnr* genes have been discovered over the last decade, their true abundance and diversity is still unknown.

The aim of *paper II* was to screen a high number of genomic and metagenomic data in order to discover previously unknown *qnr* genes and to estimate their abundance in environmental and commensal bacterial communities. A total of almost 13 terabases of genomic and metagenomic data was analyzed using a computational pipeline based on a Hidden Markov model. A total of 362 843 *qnr* gene fragments were identified, and 611 *qnr* genes were predicted. Among

the 611 identified *qnr* genes, 20 was not present or annotated as fluoroquinolone resistance genes in Genbank, and therefore considered putatively novel *qnr* genes. Nine of these were selected for experimental verification, whereof eight showed an increased minimum inhibitory concentration (MIC) to ciprofloxacin when expressed in *E. coli*. This paper provided the most exhausting search for *qnr* genes in both genomic and metagenomic data ever conducted, and contributed to the extended knowledge about of their relative abundance in various bacterial communities.

2.3 Introduction to Paper III

The environment is hypothesized to act as reservoir for antibiotic resistance genes, but the majority of the environmental bacteria are not possible to cultivate under standard laboratory conditions. Metagenomics, however, facilitates the analysis of these bacteria and their corresponding gene pool. However, metagenomic data is often large and highly fragmented, with many fragments originating from several different species which makes it challenging to work with.

In *paper III* we present a new method for identifying and reconstructing novel antibiotic resistance genes from fragmented metagenomic datasets. The method is an improved version of the methods used in *paper I* and *paper II*, and is constructed to handle large datasets and to create and optimize models for identification of any antibiotic resistance gene. The method does a first classification on fragments, with a threshold score optimized for short sequences, to decide if they originate from the modeled resistance genes. This first classification is significantly reducing the amount of data and enables the method to be computationally efficient. The fragments which are classified as positives are assembled using a paired-end approach where both reads in the pair are used even if only one of the reads was positively classified. In the final step of the method, the assembled contigs are analyzed with the model using a threshold score optimized for full-length genes. To demonstrate the methods performance, three models representing three clinically relevant gene classes were created and used on four metagenomic datasets, which resulted in the identification of 72 novel antibiotic resistance genes. The proposed method is the first efficient way of reliable identification and reconstruction of novel antibiotic resistance genes from fragmented metagenomic data.

3 Future work

Of the 76 novel subclass B1 metallo-β-lactamases discovered in *paper I*, 21 were phenotypically tested, and the next step will be to extend this number. The structures of the discovered enzymes will also be, if possible, established through X-ray crystallization studies. The two identified B1 genes with previously unseen changes of the binding sites will be further investigated to study how the changes affects their substrate spectrum. We furthermore discovered two putative mobile genes, here a deeper investigation of the genetical context of the remaining genes would be of interest to investigate their potential mobility. As the overall concern is the risk that these genes will end up in pathogens and create a clinical problem, the fitness cost they may carry is another aspect we are interested in. Low fitness cost, high mobility and a strong resistance phenotype in Enterobacteriaceae would speak in favor of an increased risk.

The method presented in *paper III* has several features under implementation. Since the highest variability in the genes often occurs in the start and end regions, the highest abundance of retrieved fragments aligns to the center part of the HMM, often leading to a couple of codons missing in the start and end positions of the retrieved genes. Here, an elongation step involving mapping reads to the assembled gene together with an ORF-prediction step have been proposed and will be implemented. Furthermore, the method will be made to handle several gene models simultaneously, which will further speed up the processing time when searching for several classes of resistance genes. New models representing other gene classes will also be developed using the creation and optimization framework, and the aim is then to apply the methods to several new metagenomic datasets. As a next step, the aim is to develop more refined models, which can make the distinction between phenotypically different, but molecularly very similar genes. This could for example be the distinction between carbapenem hydrolyzing class A-β-lactamases and other class A β-lactamases.

As the amount of public metagenomic data continues to grow, there will be endless opportunities to compare abundances, investigate trends and explore the evolutionary origins of antibiotic resistance genes. Furthermore, as the development of the third generation sequencing techniques moves forward, the amount of longer sequence reads will increase. It would therefore be of interest to integrate those longer reads in the gene prediction methods.

Bibliography

Abraham, E. P. and Chain, E. (1940). An enzyme from bacteria able to destroy penicillin. *Nature*, 146(3713):837.

Allen, H. K., Donato, J., Wang, H. H., Cloud-Hansen, K. A., Davies, J., and Handelsman, J. (2010). Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology*, 8(4):251–259.

Armbrust, E. V. and Palumbi, S. R. (2015). Uncovering hidden worlds of ocean biodiversity. *Science*, 348(6237):865–867.

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

Bengtsson-Palme, J. and Larsson, D. J. (2015). Antibiotic resistance genes in the environment: prioritizing risks. *Nature Reviews Microbiology*, 13(6):396–396.

Binnewies, T. T., Motro, Y., Hallin, P. F., Lund, O., Dunn, D., La, T., Hampson, D. J., Bellgard, M., Wassenaar, T. M., and Ussery, D. W. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Functional & Integrative Genomics*, 6(3):165–185.

Boulund, F., Johnning, A., Pereira, M. B., Larsson, D. J., and Kristiansson, E. (2012). A novel method to discover fluoroquinolone antibiotic resistance (qnr) genes in fragmented nucleotide sequences. *BMC genomics*, 13(1):695.

Cantas, L., Shah, S. Q., Cavaco, L., Manaia, C., Walsh, F., Popowska, M., Garelick, H., Bürgmann, H., and Sørum, H. (2013). A brief multi-disciplinary review on antimicrobial resistance in medicine and its linkage to the global environmental microbiota. *Frontiers in Microbiology*, 4:96.

Canton, R. (2009). Antibiotic resistance genes from the environment: a perspective through newly identified antibiotic resistance mechanisms in the clinical setting. *Clinical Microbiology and Infection*, 15(s1):20–25.

Carlos, F. A.-C. (2015). *Antibiotics and Antibiotic Resistance in the Environment*. CRC Press.

Coates, A. R., Halls, G., and Hu, Y. (2011). Novel classes of antibiotics or more of the same? *British journal of pharmacology*, 163(1):184–194.

Consortium, H. M. P. et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402):207–214.

Curtis, T. P., Sloan, W. T., and Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences*, 99(16):10494–10499.

D'Costa, V. M., King, C. E., Kalan, L., Morar, M., Sung, W. W., Schwarz, C., Froese, D., Zazula, G., Calmels, F., Debruyne, R., et al. (2011). Antibiotic resistance is ancient. *Nature*, 477(7365):457–461.

Dunham, I., Hunt, A., Collins, J., Bruskiewich, R., Beare, D., Clamp, M., Smink, L., Ainscough, R., Almeida, J., Babbage, A., et al. (1999). The dna sequence of human chromosome 22. *Nature*, 402(6761):489–495.

Dykhuizen, D. E. (1998). Santa rosalia revisited: why are there so many species of bacteria? *Antonie van Leeuwenhoek*, 73(1):25–33.

Forsberg, K. J., Reyes, A., Wang, B., Selleck, E. M., Sommer, M. O., and Dantas, G. (2012). The shared antibiotic resistome of soil bacteria and human pathogens. *Science*, 337(6098):1107–1111.

Gillings, M. R. and Stokes, H. (2012). Are humans increasing bacterial evolvability? *Trends in ecology & evolution*, 27(6):346–352.

Hall, B. G. and Barlow, M. (2004). Evolution of the serine β-lactamases: past, present and future. *Drug Resistance Updates*, 7(2):111–123.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., and Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & biology*, 5(10):R245–R249.

Heather, J. M. and Chain, B. (2016). The sequence of sequencers: The history of sequencing dna. *Genomics*, 107(1):1–8.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R., and Zamir, A. (1965). Structure of a ribonucleic acid. *Science*, 147(3664):1462–1465.

Hugenholtz, P., Goebel, B. M., and Pace, N. R. (1998). Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of bacteriology*, 180(18):4765–4774.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921.

Martínez, J. L. (2008). Antibiotics and antibiotic resistance genes in natural environments. *Science*, 321(5887):365–367.

Mathews, C. K., van Holde, K. E., and Ahern, K. G. (2000). *Biochemistry*. Benjamin/Cummings, San Francisco, California, 3 edition.

NCBI (2017). National Center for Biotechnology Information Genome Browser. `https://www.ncbi.nlm.nih.gov/genome/browse/`. [Online; accessed 2017-January-23].

Nordmann, P., Poirel, L., and Dortet, L. (2012). Rapid detection of carbapenemase-producing enterobacteriaceae. *Emerging Infectious Disease*, 18(9).

Papp-Wallace, K. M., Endimiani, A., Taracila, M. A., and Bonomo, R. A. (2011). Carbapenems: past, present, and future. *Antimicrobial agents and chemotherapy*, 55(11):4943–4960.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285):59–65.

Riesenfeld, C. S., Goodman, R. M., and Handelsman, J. (2004a). Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental microbiology*, 6(9):981–989.

Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004b). Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.*, 38:525–552.

Sanger, F. (1960). Chemistry of insulin. *British medical bulletin*, 16(3):183–188.

Segawa, T., Takeuchi, N., Rivera, A., Yamada, A., Yoshimura, Y., Barcaza, G., Shinbori, K., Motoyama, H., Kohshima, S., and Ushida, K. (2013). Distribution of antibiotic resistance genes in glacier environments. *Environmental microbiology reports*, 5(1):127–134.

Sommer, M. O., Dantas, G., and Church, G. M. (2009). Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*, 325(5944):1128–1131.

Stokes, H. W. and Gillings, M. R. (2011). Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into gram-negative pathogens. *FEMS microbiology reviews*, 35(5):790–819.

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9):418–426.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., et al. (2001). The sequence of the human genome. *Science*, 291(5507):1304–1351.

Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *Pharmacy and Therapeutics*, 40(4):277.

Walsh, C. (2003). *Antibiotics*. American Society of Microbiology.

Walsh, F. (2013). Investigating antibiotic resistance in non-clinical environments. *Frontiers in Microbiology*, 4:19.

Walsh, T. R., Weeks, J., Livermore, D. M., and Toleman, M. A. (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(5):355–362.

Watson, J. D., Crick, F. H., et al. (1953). Molecular structure of nucleic acids. *Nature*, 171(4356):737–738.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12):6578–6583.

Wichmann, F., Udikovic-Kolic, N., Andrew, S., and Handelsman, J. (2014). Diverse antibiotic resistance genes in dairy cow manure. *MBio*, 5(2):e01017–13.

Wooley, J. C., Godzik, A., and Friedberg, I. (2010). A primer on metagenomics. *PLoS Comput Biol*, 6(2):e1000667.

Wu, R. and Taylor, E. (1971). Nucleotide sequence analysis of dna: Ii. complete nucleotide sequence of the cohesive ends of bacteriophage λ dna. *Journal of molecular biology*, 57(3):491–511.

Yong, D., Toleman, M. A., Giske, C. G., Cho, H. S., Sundman, K., Lee, K., and Walsh, T. R. (2009). Characterization of a new metallo-β-lactamase gene, blandm-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(12):5046–5054.