ESCOLA
SUPERIOR
DE SAÚDE
POLITÉCNICO
DO PORTO

P.PORTO

M

MESTRADO
Bioestatística e Bioinformática Aplicadas à Saúde

In silico screening of antibiotic resistance genes in a genomic library of bacteriophages

Sara Catarina Leal Aguiar

09/**2019**

P.PORTO

ESCOLA SUPERIOR DE SAÚDE

In silico screening of antibiotic resistance genes in a genomic library of bacteriophages

Autor

Sara Catarina Leal Aguiar

Orientador(es)

Professor Doutor Rúben Fernandes / Escola Superior de Saúde (ESS), Instituto Politécnico do Porto

Professora Doutora Pilar Baylina / Escola Superior de Saúde (ESS), Instituto Politécnico do Porto

Professor Doutor Vítor Sá / Escola Superior de Saúde (ESS), Instituto Politécnico do Porto

Dissertação apresentada para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Bioestatística e Bioinformática Aplicadas à Saúde pela Escola Superior de Saúde do Instituto Politécnico do Porto.

To the ones I love and always will miss.
The ones I admire and inspire me.

My family and my friends.

Though especially to my dad, my mom and my sister.

My rock, my joy and my dauntless.

I will always throw a few kisses to the air to give you thanks and mean I adore you, until next time.

Para aqueles que amo e sempre sentirei saudade.

Aqueles que admiro e inspiram-me.

Minha família e amigos.

Mas em especial, para o meu Pai, minha Mãe e minha Irmã.

Meu apoio, minha alegria e minha destemida.

Eu vou sempre atirar beijos para o ar para vos agradecer e dizer que vos adoro, até à próxima.

Acknowledgements

First and foremost, I thank Professor Rúben Fernandes, Professor Pilar Baylina and Professor Vítor Sá for the orientation in this dissertation. I thanks to the professors of this Institution for the knowledge given along this master`s that contributed to the delivery of this work.

Lastly, I would like to acknowledge previous mentors, PhD Maria João Carvalho and Professor Timothy Walsh, that though many years have had passed already the knowledge and motivating received still impact me and have positively strengthened this work.

Furthermore, I would like to acknowledge all the authors and publishers' houses that kindly permitted the use of their figures in the following work.

Abstract

Antibiotic resistance (AR) is a worldwide concern that threatens the effective treatment of infections using antibiotics. Nowadays, bacteria can become resistant to almost all antibiotics (1–5).

It is known that AR is often mobile between bacteria of different taxonomic and ecological groups by either transformation, conjugation or transduction (6,7). Here is address the controversial dissemination of AR through the mechanism of transduction by determining the presence of antibiotic resistance genes in bacteriophages genomes through an *in silico* approach.

As such this dissertation presents the screening of 4789 antibiotic resistance genes (ARGs) in all 2051 bacteriophages genomes of the order *Caudovirales* available online using an *in silico* method of choice, and further intends to discuss the challenges in the choice as also discuss the pros and limitations of the method chosen.

Results determined the presence of 32 different antibiotic resistance gene families in 16 different phage hosts at a 99% confidence interval. Although the results showed strong diversity among the phage genomes where ARGs were found, the impact of the spread of AR through the transduction mechanism remains unclear. In order to address these questions, further future work can be carried out, although the relevance might be also questioned.

Keywords: Antibiotic resistance (AR). Antibiotic resistance genes (ARGs). Bacteriophages. Transduction. Bioinformatics.

Resumo

A resistência a antibióticos é uma preocupação mundial que ameaça o tratamento eficaz de infeções usando antibióticos. Atualmente, as bactérias podem se tornar resistentes a quase todos os antibióticos (1–5).

Sabe-se que a resistência a antibióticos é frequentemente móvel entre bactérias de diferentes grupos taxonómicos e ecológicos por transformação, conjugação ou transdução (6,7). Aqui é abordada a controversa disseminação de resistência a antibióticos através do mecanismo de transdução, determinando a presença de genes de resistência a antibióticos em genomas de bacteriófagos através de uma abordagem *in silico*.

Como tal, esta dissertação apresenta um *screening* de 4789 genes de resistência a antibióticos em todos os 2051 genomas de bacteriófagos da ordem *Caudovirales* disponíveis on-line usando um método *in silico* previamente escolhido, e pretende discutir os desafios na escolha como também discutir os prós e limitações do método escolhido.

Os resultados determinaram a presença de 32 diferentes famílias de genes de resistência a antibióticos em 16 bacteriófagos diferentes a um intervalo de confiança de 99%. Embora os resultados mostrem uma forte diversidade entre os genomas de fagos onde foram encontrados genes de resistência a antibióticos, o impacto da disseminação de resistência a antibióticos através do mecanismo de transdução permanece incerto. Para abordar essas questões, trabalhos futuros podem ser realizados, embora a relevância possa ser questionada.

Palavras-chave: Resistência a antibióticos. Genes de resistência a antibióticos. Bacteriófagos. Transdução. Bioinformática.

Table of contents

Abbreviations and Acronyms

AMR – antimicrobial resistance

AR – antibiotic resistance

ARGs – antibiotic resistance genes

BAVS – Bacterial and Archaeal Viruses Subcommittee

BLAST – Basic Local Alignment Search Tool

CARD – The Comprehensive Antibiotic Resistance Database

CRISPR – stands for Clustered Regularly Interspaced Short Palindromic Repeats, is a family of DNA sequences found within the genomes of prokaryotic organisms such as bacteria and archaea.

CRISPR-Cas – The CRISPR-Cas system is a prokaryotic immune system that provides resistance to foreign genetic elements such as those present within plasmids and phages.

DNA – deoxyribonucleic acid

ESKAPE – refers to the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species)

FASTA – a text-based file format for representing either nucleotide sequences or amino acid sequences

HGT – horizontal gene transfer

ICTV – International Committee on Taxonomy of Viruses

LHT – Lwoff, Horne and Tournier virus classification system

MEGA X – Molecular Evolutionary Genetics Analysis software

MGE – mobile genetic elements

MLSB – macrolides lincosamide–streptogramin B

NCBI – National Center for Biotechnology Information

NGS – next-generation sequencing

RNA – ribonucleic acid

SWISS-PROT – Manually annotated database from UniProt (Universal Protein) Knowledgebase; Records with information extracted from literature and curator-evaluated computational analysis

UPGMA – Unweighted Pair Group Method with Arithmetic

UTIs – urinary tract infections

WGS – whole-genome sequencing

List of Tables

List of Figures

1. Introduction

Antibiotic resistance is a worldwide concern, known since the discovery of the first antimicrobial agents (sulphonamides and penicillin) in the early 20th century (1,10–12). The availability of the first antimicrobial agents in the 1940s enabled health care providers to treat previously lethal infections (7). However, despite the discovery and improvement of new antibiotics which significantly changed the health care treatment of infections, the pressure for bacteria to adapt was higher and resistance mechanisms to antibiotics emerged, usually within a few years after the antibiotic was produced (13,14). The loss of effective antibiotics is raising, and this could bring us back to an era when simple infections can lead to untreatable infections with higher mortality (15).

Nowadays, antibiotic resistance ranks among the most urgent threats to global public health, as it easily crosses international boundaries and spreads between continents (1,2,4). The increase of antibiotic use in human medicine, agriculture, and other settings since the mid-20th century have accelerated the evolution of antibiotic resistance mechanisms by serving as a strong selective pressure for the emergence and spread of antibiotic resistance (1).

This has and will continue to undermine the ability to fight infectious diseases and manage common infections, particularly, in vulnerable patients. Bacteria can become resistant to almost all antibiotics, including all beta-lactams, the most prescribed and less toxic antibiotics to humans (5). When treatment options are limited by antibiotic resistance, health care providers are forced to use antibiotics that may be more toxic to the patient, more expensive and use those drugs reserved for the most serious/untreatable infections e.g. colistin (16,17). Globally the ESKAPE pathogens (*Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa,* and *Enterobacter sp.)* and *Escherichia coli* are of particular concern, due to the fact of these bacteria have rapidly acquired resistance to key antibiotics (1,18).

The Review on Antimicrobial Resistance reported in 2014 a low estimated of 700 000 deaths by year linked to antimicrobial resistance and predicted that this global mortality would increase to 10 million deaths surpassing the mortality due to cancer (8.2 million) in 2050 (19,20).

Without a doubt, antibiotic resistance constitutes a colossal problem that has been mismanaged and must be tackled in more rigorously (15). Understanding the means of dissemination of antibiotic resistance between bacteria is crucial in other to approach this problem.

It is known that antibiotic resistance is often mobile and the vehicles for their mobility are mobile genetic elements such as bacteriophages, plasmids and transposons that can be transferred among bacteria of different taxonomic and ecological groups (6). This mobility is possible through either transformation, conjugation or transduction (7). However, regarding transduction, it is controversial the impact of the dissemination of antibiotic resistance through this mechanism. It has been documented

proof of dissemination through this mechanism usually from *in vitro* methods in other to understand the impact *in vivo* (8,9). Though, there is very few effective *in silico* methods in other to screen bacteriophages for the presence of antibiotic resistance genes.

Computational tools have contributed to enlighten several mechanisms in many areas of expertise. In biology is no different. Bioinformatics and computational biology had their begging since the 1960s when computers emerged as important tools in molecular biology (21). Since then, various computational techniques, algorithms, computational programs and online databases have changed and improved significantly the management of biologic data and now several of these computational tools are indispensable (21–24).

Nevertheless, bioinformatic tools have their challenges ranging from the massive management of biological data with the data storage capacity, structure and accurate representation of them, easy access and manipulation of these biological data by user, processing speed, different file formats and integration of multiple tools (21,25). Due to these challenges, numerous databases and computational tools have been developed and are available to the entire scientific community. However, this knowledge is many times fragmented and dispersed through many different computational resources, often in a redundant way, therefore challenging the choice for the most adequate and effective bioinformatic tool to use (21,24).

Hence, this work intends to screen the presence of several antibiotic resistance genes in a library of bacteriophages genomes found online by the means of an *in silico* method of choice, and further discuss the challenges and limitations when choosing and using the *in silico* method chosen.

1.1. Aims and objectives of the dissertation

The purpose of the present dissertation is to give a current comprehensive knowledge of the problematic of antibiotic resistance with a focus in the dissemination of antibiotic resistance through the mechanism of transduction.

It aims to screen the presence of several antibiotic resistance genes in several bacteriophages genomes found online by the means of an *in silico* method of choice, in an effective, time and cost-friendly manner. As also to hopefully contribute data to the controversy of the impact of dissemination of antibiotic resistance through the mechanism of transduction.

As such englobes specific objectives as:

1. To research the theme of antibiotic resistance and collect a state-of-art to give an insight into the different mechanisms of antibiotic resistance and the different mechanisms of dissemination of antibiotic resistance between bacteria.
2. To gather a state-of-art regarding bacteriophages as vehicles of antibiotic resistance genes.

3. To give an overall state-of-art concerning the subject of bioinformatics with a focus in antimicrobial resistance (AMR) bioinformatics tools.
4. To present the *in silico* method chosen and further discuss the challenges that lead to that choice.
5. To screen several antibiotic resistance genes in several bacteriophages genomes in a single analyse.
6. To reflect and discuss the challenges and limitations found when using the *in silico* method chosen.
7. To discuss the results according to their hosts, antibiotic classes and the mechanisms of antibiotic resistance.

1.2. Dissertation synopsis

Taking into account the aims and the scope of this dissertation, in the following written work will be structure in chapters. Namely:

- 2. Methodology – where it will be given a descriptive view in the general literature survey for the present dissertation as well as the description on the methodology used in the data collection and data analyse for this work.
- 3. State-of-Art – Antibiotic resistance – here a comprehensive insight in the greater subject of antibiotic resistance, giving informatic knowledge of the different mechanisms of antibiotic resistance, the diversity in antibacterial agents and the different vehicles of antibiotic resistance and the main mechanisms of horizontal gene transfer between bacteria.
- 4. State-of-Art – Bacteriophages – gives an overall knowledge of what are bacteriophages their diversity and role in the mechanism of transduction.
- 5. State-of-Art – Bioinformatics – touches in what is bioinformatics, the challenges in this field. Focus in bioinformatics algorithms to compare sequences and gives a generic insight into the diversity of AMR bioinformatic tools available.
- 6. Results – states the data results from the data analyses performed.
- 7. Discussion – discusses the challenges in the choice of one bioinformatic tool to perform the screening of big sizes of data in a user and time-friendly manner and reflect in the limitations in the *in silico method* chosen to perform the screening. Discuss the results obtain from this work according to their hosts, antibiotic class and mechanism of antibiotic resistance. Also gives an overall discussion focusing on the scope of this dissertation.
- 8. Conclusions and Future work – summarizes the findings in this work, mentions the contribute and reflects in limitations and future work that can be designed to address remaining and following questions.

2. Methodology

In the following chapter, it will be only presented the methodology used in the *in silico* method chosen and recommend to screen several antibiotic resistance genes in several bacteriophages genomes in a user-friendly, free and rapid manner.

2.1. Literature survey

Mostly the literature that contributed to this work were journal articles, review articles and reports retrieved from secondary online sources, such as PubMed (https://www.ncbi.nlm.nih.gov/pubmed/). Nevertheless, also was included resources such as books both found in physical form and online.

The keywords used range from antibiotic resistance, bacteriophages, bioinformatics, antibiotic resistance genes, transduction and BLAST.

2.2. *In silico* method

It was chosen to retrieve the antibiotic resistance genes from a Bacterial Antimicrobial Resistance Reference Gene Database (Accession: PRJNA313047 ID: 313047) a BioProject from NCBI and perform the screening using the BLAST (Basic Local Alignment Search Tool) computational tool available online in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Further phylogenetic analyse were performed using the free software MEGA-X (Molecular Evolutionary Genetics Analysis available in https://www.megasoftware.net/).

2.2.1. Data collection

A total of 4789 antibiotic resistance genes and their reference DNA sequences in bacteria species available in NCBI from the BioProject with the accession: PRJNA313047 were retrieved into a single FASTA file (5,19 Mb).

It is important to note that a vast number of antibiotic resistance sequences can be found with a significant similarity between each other since there are many different variants found in different antibiotic resistant genes families (To best assess the numbers, please view Table 4 in chapter "6. Results").

2.2.2. Data analyses

2.2.2.1 Screening method: BLAST

In other to perform the screening of several antibiotic resistance genes was determined that the best method to be applied would be using the basic local alignment algorithm performed by the Basic Local Alignment Search Tool (BLAST), a computational tool available online in NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The basic local alignment algorithm focusses in finding regions of local similarity between sequences, finding the portion in the genome where it finds the best alignment to the shorter sequence, instead of dispersing the alignment along the genome (the longer sequence). The program BLAST can compare nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

The FASTA file with all 4789 DNA reference sequences was submitted in BLASTn to be only run against the organisms found in *Caudovirales* (taxid:28883).

As of February 2019, *Caudovirales* order has a total of the 2051 complete genomes in NCBI, corresponding to 21 of *Akermannviridae*, 543 of *Myoviridae*, 355 of *Podoviridae*, 1125 of *Siphoviridae* and 7 unclassified *Caudovirales*. (26) (for more inside in the order please view chapter "4. State-of-Art – Bacteriophages").

The results were found to 99% confidence interval by default and analysis run approximately 55 minutes (for more in results please view chapter "6. Results").

2.2.2.2 Multiple sequence alignment: phylogeny

Opposing the local alignment (alignment of one or more parts of two nucleotide or protein sequences) there is the global alignment (alignment of pairs of nucleotide or protein sequences over the entire length of them). The multiple sequence alignment performs global alignments between sequences and then between several more allowing analyses of phylogeny.

Multiple sequence alignment can be performed online by Clustal Omega computational tool though it was not possible due to an error indicating that the genomes weren't aligned, therefore is preferred to perform the phylogeny analyse in MEGA X software that is freely uploaded in different operating systems.

MEGA-X (Molecular Evolutionary Genetics Analysis available in https://www.megasoftware.net/) is free software that implements many analytical methods and tools for phylogenomic and phylomedicine available in several different operating systems and in two interfaces (graphical or command line).

A total of 15 complete genomes (exception of *Enterobacteria phage* T4 because the complete genome was not available) where it was found antibiotic resistance genes were submitted to multiple sequence alignment by Clustal W in MEGA X (Molecular Evolutionary Genetics Analysis) software. From a set of sequences, Clustal W calculates a series of pairwise alignments scores (comparing each sequence one another) and convert them to two distances (27). Both the pairwise alignment and the followed multiple alignment token about 24 hours each to run. Then from the distances and by a clustering algorithm, it can be built a phylogenetic tree.

The phylogeny method chosen to construct the tree was the Unweighted Pair Group Method with Arithmetic mean (UPGMA). UPGMA is a quantitative method that clusters sequentially the taxonomic operational units with the least evolutionary distance between them (24).

3. State-of-Art - Antibiotic resistance

3.1. Antibiotic resistance

Antibiotic resistance corresponds to the ability of bacteria to withstand the effects of an antibiotic (28). The evolution of resistant strains is a natural phenomenon; however, the misuse of antimicrobial agents can accelerate the process (13).

Antibiotic resistance is not new (6,14), but now threatens to turn back the clock to the pre-antibiotic era (29,30). Initially, resistant bacteria occurred mostly in hospitals. For example, sulfonamide-resistant *Streptococcus pyogenes* emerged in military hospitals in the 1930s and penicillin-resistant *Staphylococcus aureus* confronted London civilian hospitals very soon after the introduction of penicillin in the 1940s. In fact, in late the 1950s to early 1960s, resistance to multiple drugs was detected among enteric bacteria – namely, *Escherichia coli, Shigella* sp. and *Salmonella* sp. (6). Fortunately, the discovery and development of new antibiotics increased significantly in the 1960s. However, ultimately, resistance emerged within a few years after the drugs being in the market (13,14). Antibiotic resistance is inevitable and regardless of where it emerges, it can easily spread into other environments (31). It is a fact that resistant strains appear easily in health care facilities mainly due to the extensive antibiotic use and also for the fact that in those environments many different bacterial strains may come into contact with each other (29). However, resistant strains are becoming more common in the non-clinical community. Not only hospital policies, as well as environmental policies, should be implemented with regards to antibiotic use, hygiene and infection control practices in humans, animals and in agriculture (13,28,31).

Antibiotic resistance may be either presented naturally, so-called inherent or intrinsic or be acquired by the processes of genetic mutation or gene transfer (31–33). Inherent resistance is where the bacteria naturally manifest resistance to antibacterial drugs through a variety of mechanisms (7,32,33). Thereby, resistance to a certain antibiotic is observed in every single individual strain of the same species independently of their growth phase or environment (34). Acquired resistance is where the susceptible bacteria acquire resistance by mutation or by integration of mobile genetic elements (such as bacteriophages, plasmids or transposons) through either transduction, transformation or conjugation (6,32,33). This type of resistance may occur in variable proportions of isolates in their genera or species, being also variable in time (34).

Antibiotic resistance is a certainty. Currently, there are more than 15 classes of antibiotics whose targets are involved in essential physiological or metabolic functions of the bacterial cell and none escaped a resistance mechanism (6). The numbers of antibiotic resistant strains are increasing and globally (4). Millions of kilograms of antimicrobials are used each year both in health care facilities, communities and in farming/fish industries, driving the resistance problem by killing susceptible strains and selecting those strains that are resistant (6).

3.1.1. Mechanisms of antibiotic resistance

Resistance mechanisms may vary. Some are directed at the antibiotic itself, others may target how the drug is transported and others may alter the intracellular target of the drug making the drug unable to inhibit a vital function in the bacterial cell (6). Antibiotic resistant bacteria may present one or more of the following five mechanisms (14,28,32,35,36):

- Enzymatic modification or destruction of the antibiotic: bacteria can acquire genes encoding enzymes, such as beta-lactamases, that can disable the action of the antibacterial agent (7).
- Reduced antibiotic uptake into the bacterium: the alteration of the shape or number of porins can further influence the permeability to the antibacterial agent.
- Increased efflux of antibiotic from the bacterium: bacteria may extrude the antibacterial agent by efflux pumps before it can reach its target site (7).
- Alteration or production of a new target site: target site modification can reduce or unable the affinity of the antibacterial agent with the target site. This mechanism can mediate resistance to a variety of antibiotics such as beta-lactams, glycopeptides and quinolones.
- Over-expression of the drug target: this mechanism leads to the ineffectiveness of the antibacterial agent due to overproduction of the drug target.

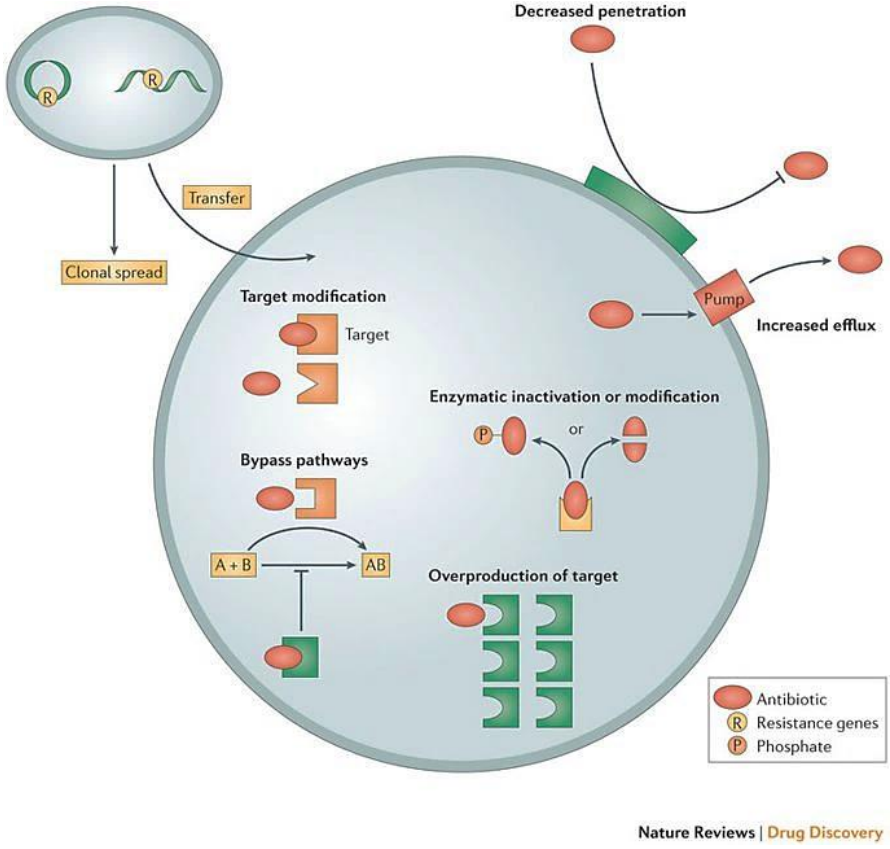


Figure 1 – Representation of the resistance mechanisms previously described in bacterial cells (14,36)

3.2. Antibacterial agents

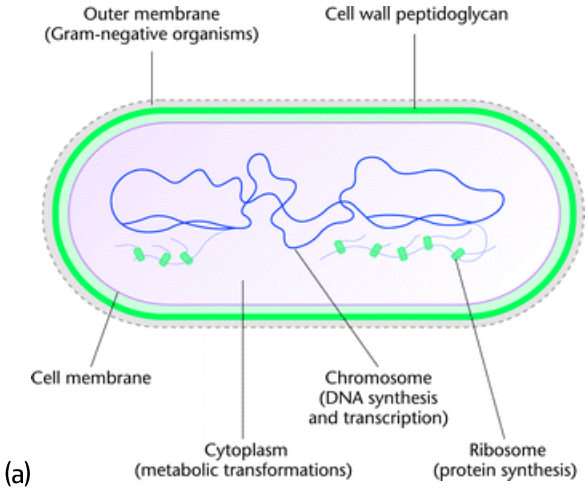
Antibiotics can either be natural, semi-synthetic or synthetic agents that kill (bactericidal) or inhibit the growth of bacteria (bacteriostatic) (31). Nevertheless, the efficacy of the antibiotic depends on the dose, duration of exposure and the type of the invading bacteria (34).

Antimicrobial agents must selectively target the infecting organism without harming the host. The differences between prokaryotic and eukaryotic cells are exploited by antibiotics to achieve their selective effect (7,10).

There are several classes of antibiotics in clinical use and they act upon bacterial cells differently. According to their mechanism of action, antibiotics may be grouped as the following (6,7,31,37,38):

- Inhibitors of cell wall synthesis: beta-lactams (penicillins, cephalosporins, carbapenems, monobactams) and glycopeptides (as vancomycin and teicoplanin);
- Inhibitors of cell membrane function: polymixin B and colistin;
- Inhibitors of ribosome function:
    - Inhibitors of the 30s subunit as aminoglycosides (gentamicin), tetracyclines;
    - Inhibitors of the 50s subunit as macrolides, chloramphenicol, clindamycin, linezolid and streptogramins;
- Inhibitors of nucleic acids synthesis:
    - Inhibitors of DNA synthesis as fluoroquinolones and metronidazole;
    - Inhibitors of RNA synthesis as rifampicin;
- Inhibitors of other metabolic processes:
    - Mycolic acid synthesis inhibitors as isoniazid;
    - Folic acid synthesis inhibitors as sulphonamides and trimethoprim;

In this way, different main structures of the bacterial cell can be targeted by antibiotics, as seen in Figure 2.



(a)

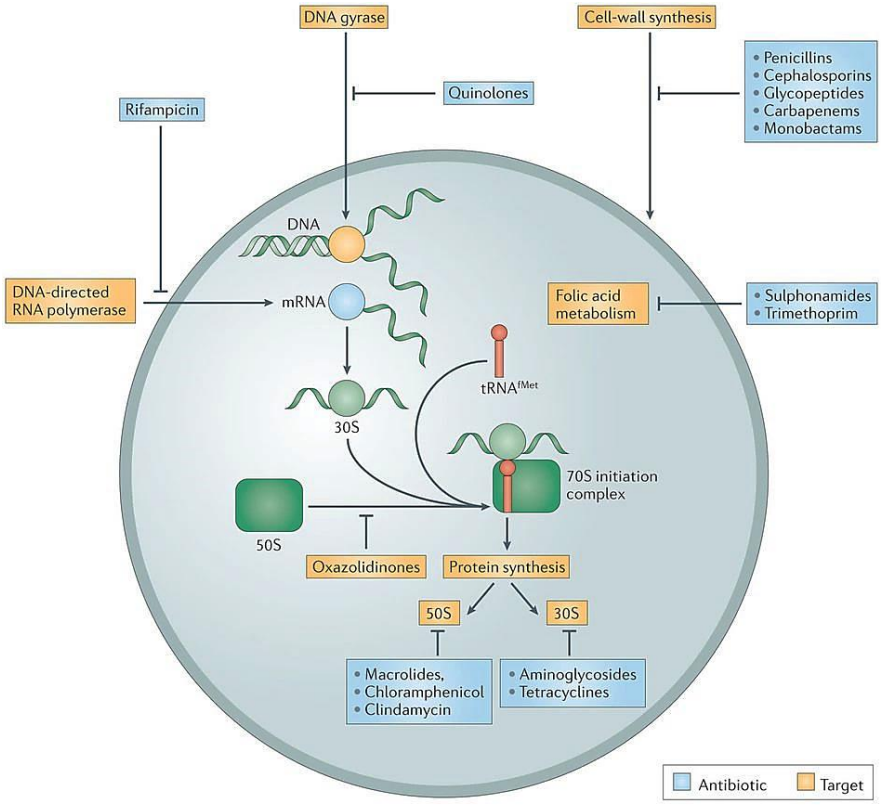


Figure 2 – (a) Bacterial cell and the possible targets for selective attack (10) Copyright © 2001 John Wiley & Sons, Ltd. Obtained permission to use by the author.; (b) In more detail, the targets in yellow boxes and examples of antibiotics in blue boxes (14,36) Copyright © 2013 Macmillan Publishers Ltd. All rights reserved.

Nonetheless, different antibiotics may have the same mechanism of action and achieve inhibition by different pathways (10). For example, the beta-lactams inhibit the synthesis of the bacterial wall by interfering with the enzymes required for the synthesis of the peptidoglycan wall, whereas the glycopeptides bind to the terminal D-alanine residues of the nascent peptidoglycan chain, preventing the cross-linking steps required for stable cell wall synthesis (7).

3.2.1. Beta-lactam antibiotics

Beta-lactams are the most successful and commercialized class of antibiotics developed so far (14,39). They comprise a large group of structurally distinct molecules, being the most diverse class of antibiotics (39). In fact, they are the most commonly used antibiotics in developing countries, mainly due to their availability, relatively low cost, stability in storage, broad antibacterial spectrum and especially low toxicity to the host (11,14,40,41).

The first beta-lactam to be discovered in 1928 when Alexander Fleming noticed the production of a substance with antimicrobial properties that inhibited *Staphylococcus aureus* by the mould *Penicillium notatum*, the penicillin. Since then, many new beta-lactam antibiotics have been developed and improved

(14,37,41). The fact that all beta-lactams are selective inhibitors of bacterial cell wall synthesis, exhibiting selective toxicity against bacterial cells only, made this class of antibiotics "wonder drugs" for many years (11,14,37).

All beta-lactams possess a beta-lactam ring. During the peptidoglycan synthesis, this structure is responsible and indispensable for the inactivation of a set of transpeptidases that catalyse the final cross-link reaction which gives the bacterial cell wall its shape and strength (10,41).

Although antibiotics of this group started as naturally occurring substances, they had their structures modified in order to improve their action and response to new appearing resistance mechanisms. Beta-lactams are a diverse family of antibiotics that can be divided into different subgroups (14,37):

- Penicillins and derivatives: Penicillin was the first discovered beta-lactam. The core compound, 6-aminopenicillanic acid (6-APA), was and still is used as the main starting point for the preparation of numerous other semi-synthetic derivatives (37). This subgroup presents susceptibility to hydrolysis by the bacterial enzymes beta-lactamases, therefore they are often combined with beta-lactamase inhibitors (32).
- Cephalosporins: This subgroup is characterised by semi-synthetic antibiotics based on the original cephalosporin C compound discovered in 1947 by Giuseppe Brotzu in 1947. Cephalosporins are grouped in 1st, 2nd, 3rd or 4th generation according to their time of introduction in therapeutics or spectrum of activity (4th generation cephalosporins display a broader spectrum of activity) (37).
- Cephamycins: This group is very similar to cephalosporins and is often grouped with them (14,37).
- Carbapenems: These compounds are considered broad-spectrum beta-lactam antibiotics (37). In fact, carbapenems possess the broadest spectrum of activity against Gram-positive and Gram-negative bacteria. As a result, they are often used as "antibiotics of last resort" when patients with infections become gravely ill or are suspected of harbouring resistant bacteria (42).
- Monobactams: These are synthetic compounds that contain a beta-lactam ring alone, which is not fused to another ring (37).
- Beta-lactamases inhibitors: These compounds are often combined with other beta-lactams to overcome the action of beta-lactamases (37,43). Examples are clavulanic acid and sulbactam.

Since the beta-lactams' effectiveness relies on their ability to reach and bind to the penicillin-binding proteins (PBPs), resistance to this class of antibiotics is usually linked to the hydrolysis of the beta-lactam ring, modification of PBPs, cellular permeability or and increased levels of antibiotic efflux by means of efflux pumps (44).

Unfortunately, resistance to these antibiotics is very common in both Gram-positive and Gram-negative bacterial pathogens (39,45). Nonetheless, Gram-positive and Gram-negative bacteria exhibit different susceptibility patterns to beta-lactams due to structural differences in their cell walls and their

ability to hydrolyse the beta-lactams (10,11). In fact, Gram-positive are more susceptible than Gram-negative bacteria and the major resistance mechanism against Gram-negative bacteria is drug inactivation by beta-lactamases (39).

3.2.2. A diversity of antibiotic resistance genes: beta-lactamases

In a single antibiotic resistance gene family, as for instance the beta-lactamase gene family, we can have numerous classes of antibiotic resistance genes.

Take the example of beta-lactamases that are enzymes produced by bacteria that confer resistance against beta-lactam antibiotics. There are innumerable beta-lactamases that hydrolyse different subgroups of beta-lactam antibiotics. They have been classified according to their hydrolytic spectrum, susceptibility to inhibitors, whether they are encoded on the chromosome or on plasmids or according to their amino acid sequence (46).

Beta-lactamases can be divided into two major groups according to the presence of a serine in the active site or being zinc-dependent. The former includes classes A, C and D and the latter, metallo-beta-lactamases (class B). These subgroups can be divided further according to other functional or structural characteristics. A more detailed review of the classification system was originally by Bush and Jacoby (46).

Beta-lactamases are a large family and the first reported beta-lactamase was the *Amp*C beta-lactamase in 1945 (14,37). Famous for destroying penicillin, since then many more were discovered that effectively hydrolyse a broader range of beta-lactams. Beta-lactamases are known as narrow-, moderate-, broad-, or extended-spectrum beta-lactamases (ESBLs) - the last two being the most fearful. The broad-spectrum beta-lactamases provide resistance to penicillins and cephalosporins and are not often inhibited by inhibitors such as clavulanic acid or tazobactam. On the other hand, ESBLs confer resistance to penicillins, as also to the first-, second- and third-generation cephalosporins and aztreonam, but not to carbapenems and are inhibited by many beta-lactamase inhibitors (37). One example very representative of this group are the CTX-M (for cefotaxime first isolated in Munich) enzymes first described in 1989, in a clinical isolate of *E. coli* in Germany, and that in the last decade spread all over the world leading to the current endemic worldwide (47,48). CTX-M enzymes are divided in several clusters based on their genetic structure (49) being associated with different genetic elements including the insertion sequences ISEcp1 and ISCR1 (47).

The appearance and spread of carbapenem-hydrolyzing beta-lactamases (carbapenemases) have also intensified this problem (3,37). Carbapenemases confer resistance not only to carbapenems but also to several other beta-lactam groups, thus, the effectiveness of these antibiotics is increasingly compromised (50,51). Members of this group include the *Klebsiella pneumoniae* carbapenemase (KPC), prevalent in *Enterobacteriaceae*, and with growing outbreaks in China, Brazil, and several European

countries as Greece and Italy (3,52). The class D carbapenemases oxacillinases (mainly OXA-48 and OXA-181) is predominant in Europe, South Asia, Central and South America (3). Metallo-beta-lactamases such as IMP were found in Japan at the end of the 1980s and VIM was first reported in Italy, in *Pseudomonas aeruginosa*, in 1997 are now established in the five populated continents (3,52). Historically widespread in *Pseudomonas aeruginosa* and *Acinetobacter* spp. This group are now also predominant in *Enterobacteriaceae* (52). The New Delhi metallo-beta-lactamase was first reported in 2009 and incredibly has already reached every continent except for Antarctica (4).

Adding to the rising concern the antibiotic resistance genes can be subjected of mutations increasing the diversity, in a single antibiotic resistance gene family we can have several different variants as for example of the bla_{OXA} currently with 719 different reference sequence variants in NCBI Reference Gene Browser (https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/oxa).

3.3. Mobile genetic elements: vehicles of antibiotic resistance

Antibiotic resistance is often mobile and the vehicles for their mobility are mobile genetic elements such as bacteriophages, plasmids and transposons that can be transferred among bacteria of different taxonomic and ecological groups. These elements are very important vehicles of antibiotic resistance dissemination since they enable bacteria to adapt quickly to the introduction of antibacterial agents into their environment (6). This ability of bacterial strains to rapidly become resistant to antibiotics has been underestimated (33).

Mobile genetic elements (MGE) can be defined as any segment of DNA that is capable of translocation from one part of a genome to another or between genomes. Therefore, this definition includes a wide range of distinct mobile genetic elements (37):

- Plasmids: plasmids or conjugative elements are extrachromosomal elements that contain their own origin of replication. Also, commonly contain an origin of transfer and genes encoding functions that allow them to transfer to new hosts via conjugation. In addition to functions involved in replication and transfer, plasmids commonly encode several antibiotic resistance genes. Many plasmids have a broad host range and can transfer between different phylogenetic groups whereas others have a much narrower host range and are confined to one genus or species. Additionally, some resistance plasmid types cannot coexist in a microbial cell and this fact gave rise to the division into incompatibility groups. Four major groups have been defined on the basis of genetic relatedness and pilus structure: IncF group (containing IncC, IncD, IncF, IncJ, and IncS), IncI group (including IncB, IncI, and IncK), IncP group (consisting of IncM, IncP, IncU, and IncW), and Ti (37).
- Conjugative transposons: conjugative transposons or integrative conjugative elements (ICE) are a highly heterogeneous group of genetic elements with different properties and host ranges. Like

the conjugative plasmids, these elements contain an origin of transfer and the genes required for conjugation. However, conjugative elements do not contain an origin of replication. Thus, in order to be maintained, they must integrate into a replicon (either plasmid or chromosome). Since they do not have to have the replication machinery that is compatible with the host they tend to have a larger host range than plasmids (37).

- Bacteriophages: the phage particles are packaged with bacterial DNA instead of phage DNA. The type of transfer is transduction and it can be specialised (the DNA adjacent to the phage insertion site is packaged) or generalised (any segment of bacterial DNA is packaged into the phage head) (37).
- Insertion sequences (IS): are the smallest and most abundant transposable elements (47). They consist only of the gene required for element mobility, which encodes the transposase, and the inverted repeat sequences at the ends of the element for integration (37,47).
- Integrons: genetic elements that include components of a site-specific recombination system enabling them to capture and mobilise genes as gene cassettes. The cassettes are mobile elements including a gene (often an antibiotic resistance gene) and an integrase-specific recombination site known as 59-be element (47). Therefore, they are not mobile genetic elements but can become so in association with plasmids or transposons (37).

Factors influencing the acquisition of MGE can include selective pressures in the environment, host factors, and properties of the genetic elements themselves (37).

3.3.1. Horizontal gene transfer

Horizontal gene transfer (HGT) mechanisms are considered responsible for the increased spread of antibiotic resistance between bacteria (30). Conjugation, transformation, and transduction are the main mechanisms by which dissemination of antibiotic resistance genes occurs (9,53–55).

- Transformation: happens when naked DNA is released and then is taken up by another bacterium. The antibiotic resistance gene can be integrated either into the chromosome or into a plasmid of the recipient cell (54).
- Transduction: antibiotic resistance genes are transported from one bacterium to another by bacteriophages and then integrated into the chromosome of the recipient cell (lysogeny) (54).
- Conjugation: happens by direct contact between two bacteria: plasmids form a mating bridge across the bacteria and DNA is exchanged. It can result in the acquisition of antibiotic resistance genes by the recipient cell. Transposons are DNA sequences, that carry their own recombination enzymes, and allow for transposition from one location to another; these sequences can also carry antibiotic resistance genes (54).

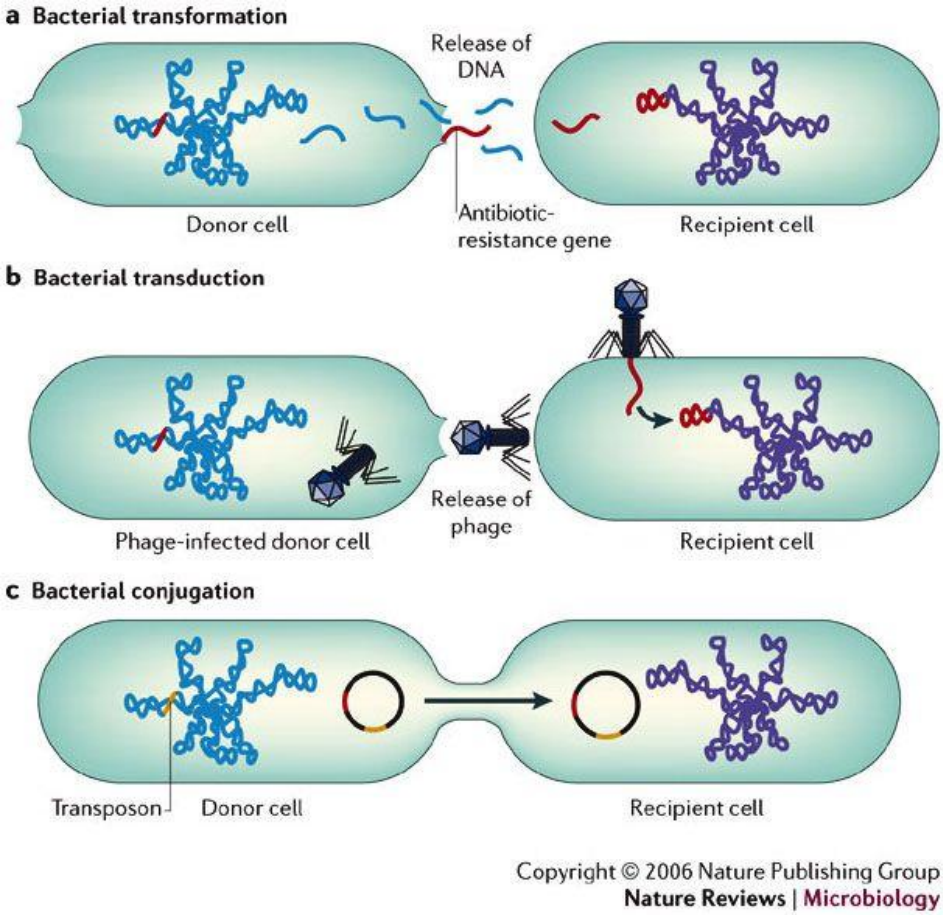


Figure 3 – Representation of the three main mechanisms of horizontal gene transfer between bacteria. (a) Transformation; (b) Transduction; (c) Conjugation; (54) Copyright © 2006 Nature Publishing Group. All rights reserved.

4. State-of-Art - Bacteriophages

The discovery of bacteriophages was linked independently to Frederick Twort in 1915, a British pathologist and to Félix d'Hérelle in 1917, a French–Canadian microbiologist (56). Soon after, phage research conducted in the 1920s and 1930s focused on the development of phage therapy for the treatment of bacterial infections. However, in the late 1930s, was concluded that the efficacy of phage therapy was ambiguous, and that further research was required. These concerns and the emerging of antibiotics led to a decline in interest in phage therapy (56).

Nevertheless, phages had an unforeseen positive impact on our understanding of several biological processes. Since their discovery 100 years ago, phage research contributed to the development of many techniques and reagents that fortify modern biology, from sequencing and genome engineering to the fresh discovery and exploitation of CRISPR–Cas phage resistance systems (56).

4.1. Bacteriophages, or simply phages

Bacteriophages, or simply known as phages, are viruses that infect and replicate only in bacterial cells. They are considered the most abundant biological entities on earth, with an estimated total population of 10^{30} to 10^{32}, and are extremely diverse in size, morphology, and genomic organization (8,57,58).

Though it is important to note that bacteriophages are very species-specific regarding their hosts and usually only infect a single, bacterial species, or even specific strains within a species (59).

All bacteriophages have a DNA or RNA genome encased in a shell of phage-encoded capsid proteins, which protect the genetic material and facilitate its delivery into the next host cell (57,59–62), and detailed visualization of several phages types by electron microscopy soon showed that some appear to have "heads," "legs," and "tails", however, phages are considered non-motile and depend upon a Brownian motion to reach their targets (8,38).

It was electron microscopy, in the 1940s, which led to the recognition of different phage morphologies and the presence of different nucleic acid contents that enabled the evolution of phage taxonomy based on these properties that provided the basis for the first classification systems in 1962, the Lwoff, Horne and Tournier (LHT) virus classification system. In 1966, what is now the International Committee on Taxonomy of Viruses (ICTV) was established to provide a universal viral taxonomy (56,63).

In contrast to other organisms, there is an extremely viral diversity from differences in their genetic material (RNA or DNA), to different configurations (double or single stranded) coupled with a possible orientation of their encoded genes (57,59–62,64,65). The Baltimore classification assigns seven different groups that are based upon their genome configurations as follows: I: dsDNA, II: ssDNA, III: dsRNA, IV: ssRNA (+) sense orientation of genes, V: ssRNA (-), antisense orientation, VI: ssRNA with reverse

transcription of a dsDNA replication intermediate and VII: dsDNA with a ssRNA replication intermediate (64,65).

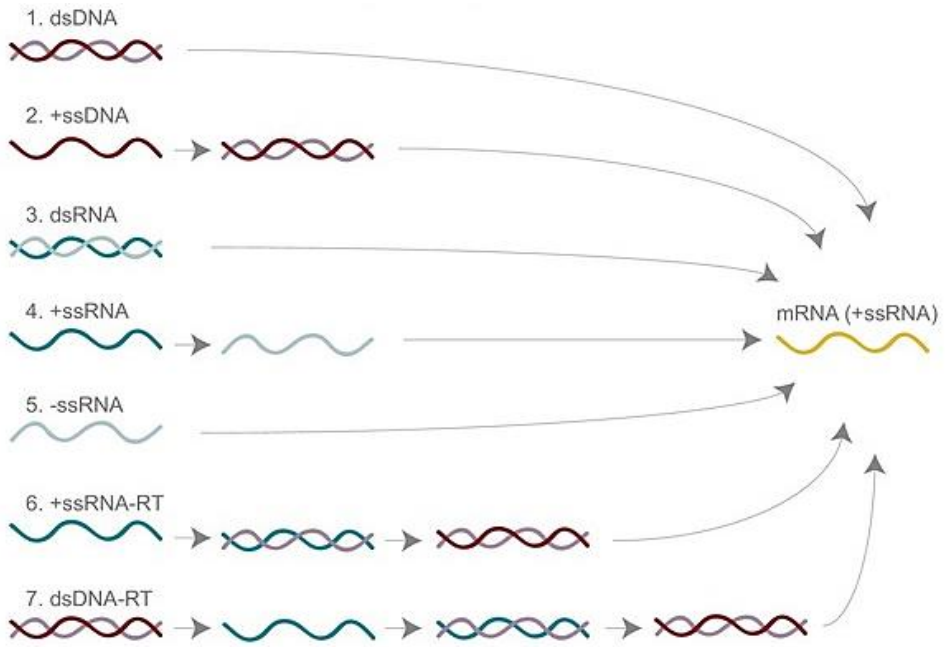


Figure 4 – Classification of Baltimore. Adapted from: Mahmoudabadi, G. *et al.* (65) Copyright © 2018 Mahmoudabadi and Phillips. The figure under the terms of (CC) Creative Commons Attribution License.

In addition, virion morphology and size vary from particles with icosahedral or more complex symmetries or may form filamentous, rectangular, bullet, even bottle-shaped nucleocapsids. Even some viruses are enveloped in a host-derived lipid bilayer (62,64).

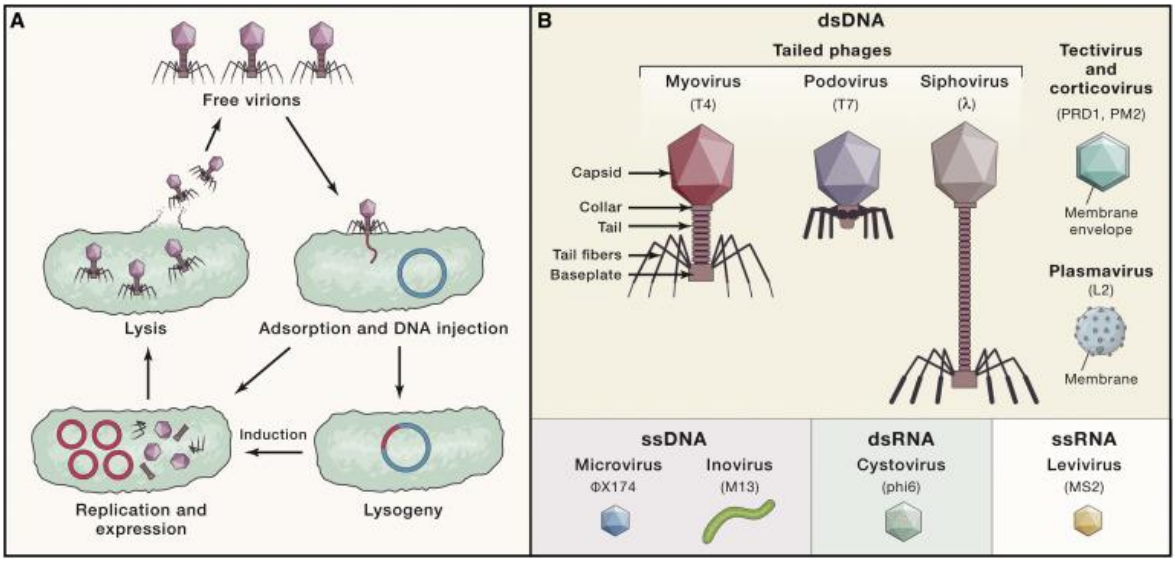


Figure 5 – (A) Phage life cycle; (B) Phage taxonomy and a visual representative based on morphology and genome composition; (62) Copyright © 2017 Elsevier Inc. Obtained permission to use by the author.

Furthermore, viral genomes are hugely variable in size and their complements of genes, ranging from less than 2000 bases encoding 2 genes to 2.5 million base pairs encoding over 2500 genes (64).

Hence, to follow a universal viral taxonomy is extremely necessary.

4.2. Taxonomy

Virus taxonomy is the responsibility of the International Committee on the Taxonomy of Viruses (ICTV, https://talk.ictvonline.org/) that classifies viruses into families, genera and species, and is universally used in virus descriptions (64). Although within ICTV holds, the phage classification is the responsibility of the Bacterial and Archaeal Viruses Subcommittee (BAVS) (63).

Traditionally, virus phenotypic properties such as host range, virion morphology, and replication mechanisms contributed to the assignment of virus in a taxa family level. However, nowadays, gene sequence comparisons provide a clearer guide to their evolutionary relationships and provide the only information that may guide the incorporation of viruses detected in environmental (metagenomic) studies that lack any phenotypic data (64).

Large numbers of new phages are being discovered, named, sequenced, and deposited into public databases. It is important to understand that taxonomy is also able to changes due to the constant flow of new information. In other to name a new bacteriophage several guidelines must be followed although the most important might be "do not use an existing name" (63).

Following are present some of the other instructions that we should follow (63):

- Always name in italic the complete host genus Latin name, rather than higher-order taxa names, followed by a space, followed by the word "phage" in italic, followed by a space, followed by a unique identifier (e.g., *Escherichia phage* T4);
- Must not use an existing unique identifier in the name. Neither do not use Greek letters or start with a numeral or only use a single letter. Identifiers should include strong complexity to easily distinguish your bacteriophage from others;
- Do not use hyphens, slashes, or any type of special character (e.g., %, $, @, etc.). You may only use underscores to separate parts of the designation (e.g., vs_p123_233), but these underscores cannot be carried over into official taxon names.
- Cannot use controversial names/phrases, profanity, names of prominent people, and trademarked names/phrases as unique identifiers.

In other to submit to public databases must be followed also some guidelines (63):

- Must include accurate genomic composition information when no other lineage information is available or otherwise can be inferred. In most cases, it should be possible to place a new isolate within the higher-order dsDNA, ssDNA, dsRNA, or ssRNA lineage groupings.
- Must include lineage information for all submitted sequences. Even if your bacteriophage is novel and does not belong to a described species, provide the most accurate lineage

information possible that places the sequence including genus and/or family. For example, "Viruses; dsDNA viruses, no RNA stage; *Caudovirales; Myoviridae; Tevenvirinae*; T4*virus*".

- Do identify prophages using the "proviral" location descriptor.

In case of difficulty in placing correctly my phage/virus is suggested the following software's (63):

- Nucleotide sequence comparison: NCBI BLASTn, Gegenees (also uses BLASTn), Gepard dotplot.
- Comparison of protein groups and predicted proteomes or identification of signature genes: CoreGenes 3.5, Roary (core and accessory genome analysis), prokaryotic Virus Orthologous Groups resource (pVOGs).
- Multiple alignment and/or phylogenetic analyses: Clustal Omega, MUSCLE, MEGA X, FastTree.
- Visualization: progressiveMAUVE, Easyfig.

4.2.1. Bacteriophage taxonomy

Bacteriophages are considered the most abundant beings in the biosphere, since the total number of phage particles is estimated to be in the order of 10^{30} to 10^{32}, it is 10 times larger than the estimated number of bacterial cells on Earth (8,57,58,66).

There are only a few prokaryotic viruses in comparison to eukaryotic viruses. Considering the scope of this dissertation only the prokaryotic viruses will be considered. In Table 1, is shown the current prokaryotic family viruses as some of the main characteristics.

It is believed that approximately 96% of all the known phages are double-stranded DNA bacteriophages belonging to the order *Caudovirales* (66).

Caudovirales, *Caudo* from Latin cauda, meaning "tail", is known as the order of tailed phages and are thought to be extremely ancient. These phages can infect Eubacteria and Archaea. The virions contain dsDNA and have icosahedral or elongated heads, the tails are helical and generally have fixation structures such as baseplates, spikes or fibers. There is no envelope and the particles adsorb to their hosts and infect them from the outside. The progeny phages are assembled via complex pathways, with phage DNA entering preformed capsids. This order is the most numerous and ubiquitous of all viruses and extremely diverse in size and structure, DNA content and composition, genome structure, proteins, antigenic and biological properties. Virions can be virulent or temperate (67).

This order suffers a recent change in July 2018 (68). Some species before inserted in *Myoviridae* family are now in the new *Ackermannviridae* family. Thus, now *Caudovirales* is divided among four families of tailed bacterial viruses infecting Bacteria and Archaea: *Ackermannviridae*, *Myoviridae*, *Siphoviridae*, and *Podoviridae* (68,69), as can be represented by Figure 6.

As of February of 2019, *Caudovirales* order has a total of the 2051 complete genomes in NCBI, corresponding to 21 of *Akermannviridae*, 543 of *Myoviridae*, 355 of *Podoviridae*, 1125 of *Siphoviridae* and 7 unclassified *Caudovirales* (26).

Table 1 – ICTV classification of prokaryotic (bacterial and archaeal) viruses. Adapted from: Mc Grath, S. *et al.* (70).

Order	Family	Morphology	Nucleic acid	Examples
Caudovirales	*Ackermannviridae*	Nonenveloped, contractile tail	Linear dsDNA	*Klebsiella phage* 0507KN21
	Myoviridae	Nonenveloped, contractile tail	Linear dsDNA	*Enterobacteria phage* T4
	Siphoviridae	Nonenveloped, noncontractile tail (long)	Linear dsDNA	*Enterobacteria phage lambda*
	Podoviridae	Nonenveloped, noncontractile tail (short)	Linear dsDNA	*Enterobacteria phage* T7
Ligamenvirales	*Lipothrixviridae*	Enveloped, rod-shaped	Linear dsDNA	*Acidianus filamentous virus* 1
	Rudiviridae	Nonenveloped, rod-shaped	Linear dsDNA	*Sulfolobus islandicus rod-shaped virus* 1
Unassigned	*Ampullaviridae*	Enveloped, bottle-shaped	Linear dsDNA	*Acidianus bottle-shaped virus*
	Bicaudaviridae	Nonenveloped, lemon-shaped	Circular dsDNA	*Acidianus two-tailed virus*
	Clavaviridae	Nonenveloped, rod-shaped	Circular dsDNA	*Aeropyrum pernix bacilliform virus* 1
	Corticoviridae	Nonenveloped, isometric	Circular dsDNA	*Pseudoalteromonas phage* PM2
	Cystoviridae	Enveloped, spherical	Segmented dsRNA	*Pseudomonas phage* phi6
	Fuselloviridae	Nonenveloped, lemon-shaped	Circular dsDNA	*Sulfolobus spindle-shaped virus* 1
	Globuloviridae	Enveloped, isometric	Linear dsDNA	*Pyrobaculum spherical virus*
	Guttaviridae	Nonenveloped, ovoid	Circular dsDNA	*Aeropyrum pernix ovoid virus* 1
	Inoviridae	Nonenveloped, filamentous	Circular ssDNA	*Enterobacteria phage* M13
	Leviviridae	Nonenveloped, isometric	Linear ssRNA	*Escherichia virus* MS2
	Microviridae	Nonenveloped, isometric	Circular ssDNA	*Enterobacteria phage* phiX174
	Plasmaviridae	Enveloped, pleomorphic	Circular dsDNA	*Acholeplasma phage* L2
	Tectiviridae	Nonenveloped, isometric	Linear dsDNA	*Pseudomonas virus* PRD1

Legend: ds- double stranded; ss – single stranded;

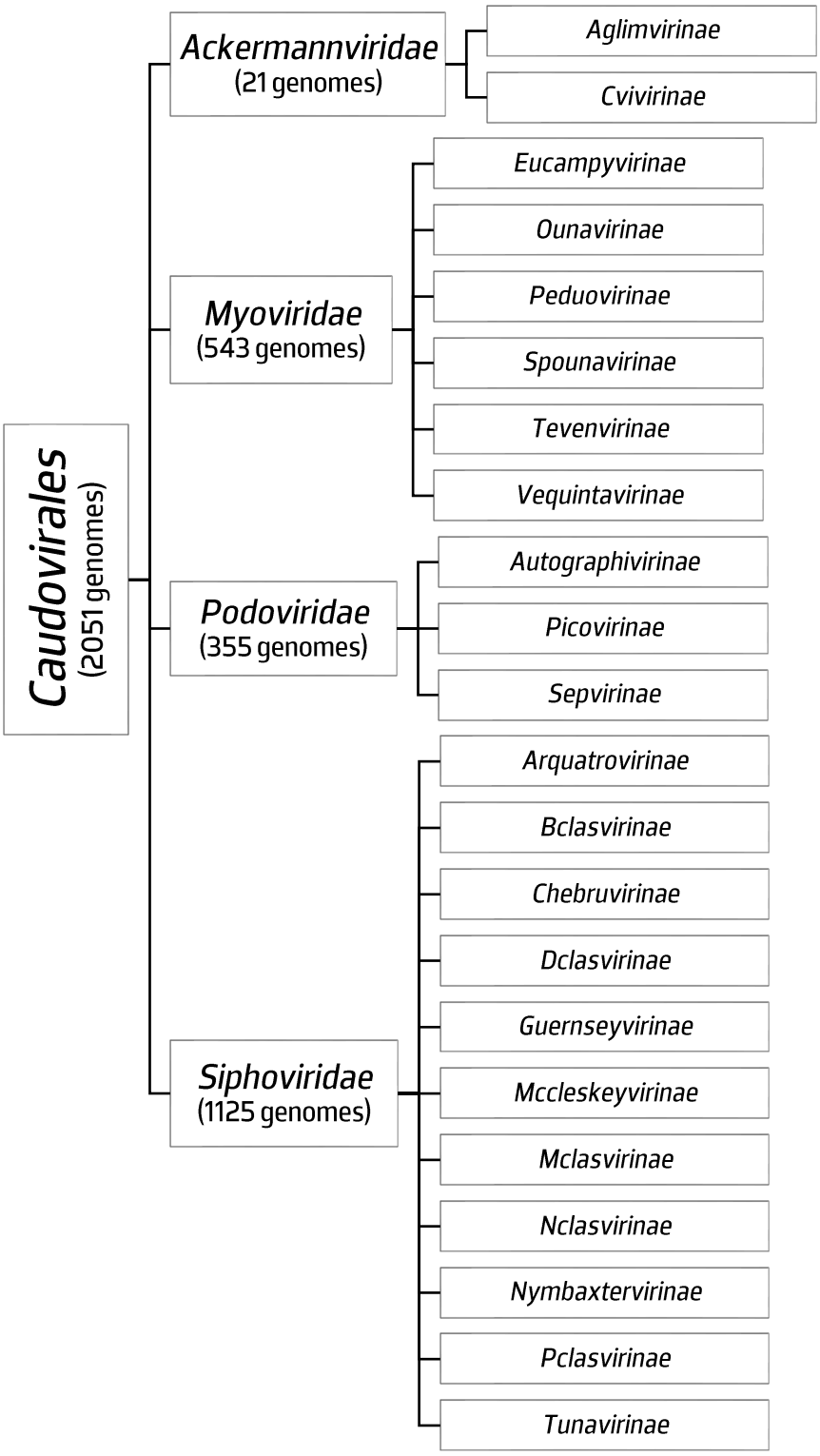


Figure 6 – Schematic representation of taxonomy of the order *Caudovirales* (Adapted from: ICTV (2018): EC 50, Washington, DC.) (68).

4.3. Lytic versus lysogenic cycle

Once a bacteriophage attaches to a susceptible host, it pursues one of three replication strategies: lytic, lysogenic or pseudolysogenic (71).

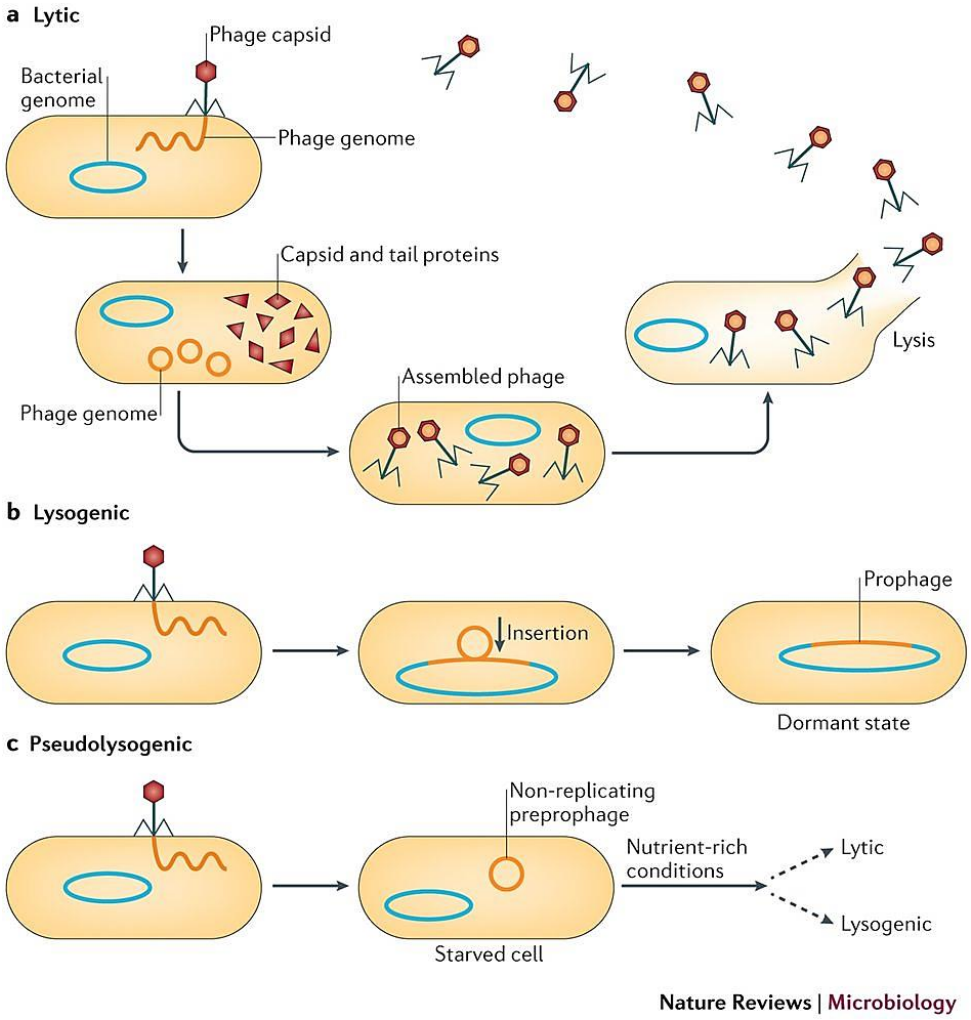


Figure 7 – Representation of phage replication cycles: (a) lytic, (b) lysogenic and (c) pseudolysogenic; (71) Copyright © 2015 Macmillan Publishers Limited. All rights reserved. Obtained permission to use by the author.

In the lytic replication cycle, once the lytic phage introduces its genome into the host cell cytoplasm uses the ribosomes of the host to manufactures its proteins. Quickly, the host cell resources are converted to viral genomes and capsid proteins, which will assemble into new copies of the original phage. Then the host cell will be lysed releasing the new phages into the environment, and this subsequently can infect another bacterium (59,71). Example of a virulent phage is the phage T4.

In the lysogenic replication cycle, once the temperate phage or lysogenic phage introduces its genome into the host cell cytoplasm, the genome will be integrated into the bacterial cell chromosome (becoming a prophage). Remaining in what is considered a phage latent or dormant state, will not promote cell death or production of phage particles. These prophages are replicated together with the bacterial host

chromosome during host cell replication and will only switch into lytic production upon exposure to DNA damage or due to changes in the environment (59,71,72). Example of a temperate phage is the phage λ.

In the pseudolysogenic replication cycle, the phage genome fails to replicate as in previous replication cycles and occur frequently under nutrient-deprived conditions (71). The phage genome remains as a non-integrated preprophage, which resembles an episome (integrative plasmid) until the nutritional status is restored. Only then the phage enters either a lysogenic or a lytic life cycle. Since the pseudolysogenic preprophage does not replicate, the genetic information in preprophage is only inherited by one of the daughter cells following cell division (71).

Note that virulent phages cannot become prophages, thus being always lytic. Although, prophages occasionally exit the bacterial chromosome and can enter the lytic cycle.

4.4. Phage-mediated transduction

Antibiotic resistance genes (ARGs) present in either bacterial chromosomes or plasmids can be transferred between bacteria by the means of phages during the infection cycle (8). Considering the possible phage replication cycles and the mechanism of transduction, in Figure 8, is shown the various possible events that may occur once the bacteriophage attaches to a bacterium (Figure 8 A-C).

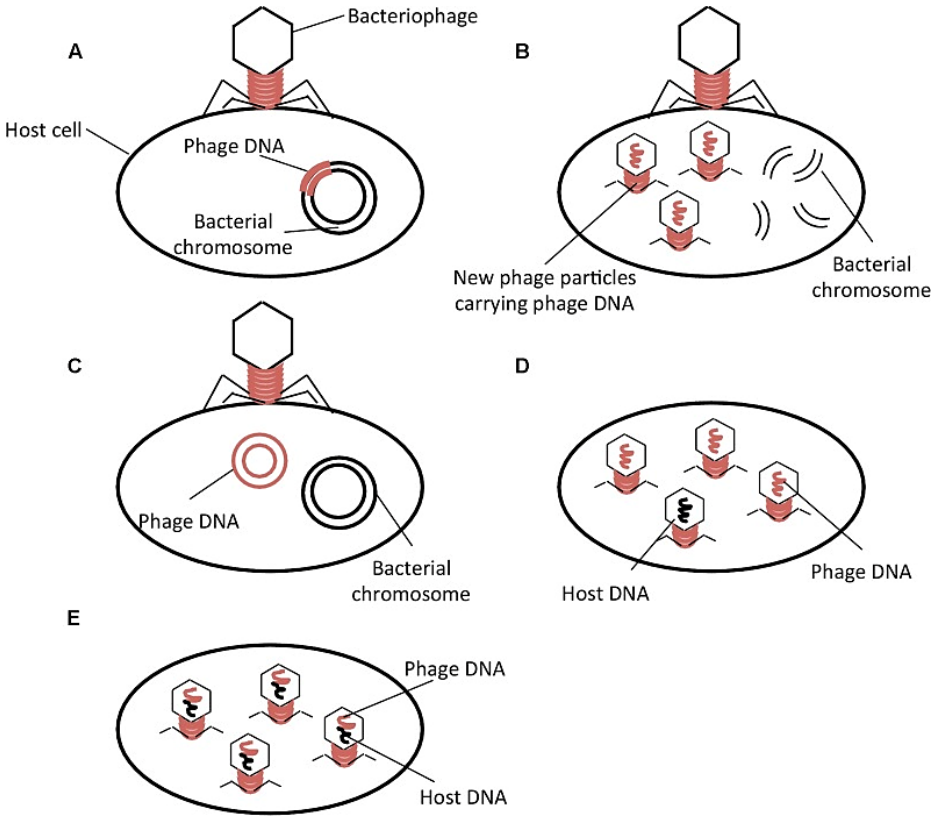


Figure 8 – Phage life cycles and types of phage-mediated transduction. (A) Lysogeny, by temperate phage; (B) Lysis, by virulent phage; (C) Pseudolysogeny; (D) Generalized Transduction; (E) Specialized Transduction; (9) Copyright © 2017 Colavecchio, Cadieux, Lo and Goodridge. The figure under the terms of the (CC BY) Creative Commons Attribution License. Obtained permission to use by the author.

In case of temperate phages, the phage DNA has integrated the host chromosome (Figure 8 A), then the prophage may remain dormant until the excision of the phage from the chromosome resulting in the formation of phage particles and lysis of the host cell (8,9).

On the other hand, virulent phages do not integrate their DNA into the host chromosome. They immediately induce production of phage particles and lysis of the host cell (Figure 8 B) (8,9).

In the few cases of pseudolysogeny, where adverse growth conditions are present, their genome does not degrade but instead remains as a plasmid within the cytoplasm (Figure 8 C) and during bacterial cell division becomes incorporated into only one daughter cell (9,71).

Concerning the mechanism of transduction two types have been described (Figure 8 D and E) (8,9):

- Generalized transduction: concerns the mispackaging of bacterial DNA into the phage capsid (Figure 8 D). The phage may then infect another susceptible host, thereby transferring genetic material to another bacterial cell where it will be integrated by homologous recombination. This type of transduction can be carried out by both virulent and temperate phages during their lytic cycle (8,9).
- Specialized transduction: typical for temperate phages, which insert their genomes into a particular region of the host chromosome. Concerns the improper excision of a prophage from the host chromosome, resulting in the packaging of the bacterial DNA into phages at a higher frequency than generalized transduction (Figure 8 E). The probability that the transferred genes are antibiotic resistance-related is very low (8,9).

4.4.1. Environmental factors associated with bacteriophage transduction of antibiotic resistance genes

As mention previously, phages can partake in the transmission of antibiotic resistance genes (ARGs). This mechanism was described in 1951 and referred as transduction. There are acknowledge two types of transduction mechanisms, generalized and specialized. In generalized transduction, phages can transfer any part of the bacterial chromosome, whereas in specialized transduction only some parts can be transferred, as explained previously (8).

Even though this process has been reported in diverse bacterial species by classic and modern genetic approaches *in vitro*, its contribution to the dissemination of antibiotic resistance *in vivo* remains unclear (8).

Recent findings have shown phage-mediated transduction to be a significant contributor to the dissemination of antibiotic resistance genes (9). However, the impact of this dissemination from this means is still controversial.

It is known the use and misuse of antibiotics in medicine, agriculture, and aquaculture (9).

Researchers agree that phages carrying antibiotic resistance genes present in different environmental biomes, are mobilized to commensal bacteria of animal and human biomes and can be transferred to pathogens of the *Enterobacteriaceae* family, such as *Salmonella* sp. and *E. coli* (9). They also recognise that the extensive use and misuse of antibiotics contribute to selective pressure leading to the incorporation of MGEs by horizontal gene transfer. Also resulting in an increase in mutations and subsequent evolution of antibiotic resistance gene sequences (9).

Studies have been performed in other to reach some conclusions. Some reported the presence of antibiotic resistance genes in phage genomes, such as bla_{TEM} and bla_{CTX-M} beta-lactamases genes in DNA phages isolated from urban sewage and river samples. As for others were able to establish that phages carrying antibiotic resistance genes persist longer in the environment than their bacterial hosts (9,58,73).

Nevertheless, given the ubiquity, abundance, and resilience of phages, this subject should be further explored in future studies. The impact and control of the spread of antibiotic resistance is an ongoing concern that must be tackled (9,58).

5. State-of-Art – Bioinformatics

5.1. What is bioinformatics?

There is not a universal definition for bioinformatics, there are a variety of definitions in the literature and world wide web where some are more inclusive than others.

Bioinformatics can be defined as an interdisciplinary field at the interface between computer science and biological science or between biology and informatics. Meaning that bioinformatics involves the technology that uses computers for storage, retrieval, manipulation, and distribution of information related to biological macromolecules such as DNA, RNA, and proteins. The emphasis here is on the use and development of computational methods to analyse amounts of biological data, such as genetic sequences, cell populations or protein samples, allowing new discoveries and predictions in biology. The computational methods used may include the development of algorithm use, and creation of software's to collect, store and manipulate data for analysis, visualization or prediction (21,23,24,74).

The pivotal goal of bioinformatics is to enlighten knowledge and understanding in living beings as a cell and how it functions at the molecular level by analysing the raw molecular sequence and structural data (24,25). In other to reach this goal, bioinformatics can be divided into two subfields, that are complementary to each other:

- The development of computational tools and databases, that embraces writing software for the sequence, structural, and functional analysis, as well as the building and curating of biological databases (24).
- The application of these tools and databases in generating biological knowledge to understand living systems, that often create new problems and challenges that in turn boost the development of new and improved computational tools (24).

As a side note, bioinformatics differs from a related field known as computational biology. Although there are also different views of how these two fields relate. Some consider that bioinfomatics is limited to sequence, structural, and functional analysis of genes and genomes and their corresponding products. As for, computational biology involves all biological areas that include computation and not necessarily include biological macromolecules. Others for instance, define bioinformatics as the development and application of computational tools in managing all kinds of biological data, whereas computational biology is more confined to the theoretical development of algorithms used for bioinformatics (21,23,24).

5.2. Emerging of bioinformatics

In the 1960s, computers emerged as important tools in molecular biology and genetics giving space to the origin of a new field now known as bioinformatics (21,25). It was mainly due to the researchers frequently need to manage massive volumes of data coming from the sequencing projects, as for the need of new methods of analysis of molecular sequences and the need for rapid and efficient searches in public databases that lead to the formulation of algorithms to solve these and other arising tasks in the biology field (75).

In short period of time, by the end of the 1960s, several computational techniques, algorithms and computational programs for analysis of molecular structure, function, and evolution, as well as rudimentary databases of protein sequences, had already been developed. No differently, in the following decades, new techniques and approaches surged, such as algorithms for sequence alignment, the creation of publicly available databases such as the NCBI, the implementation of rapid banking search systems, the development of more sophisticated systems for the prediction of protein structure, tools for genome annotation and comparison, and functional genome analysis systems (21,75).

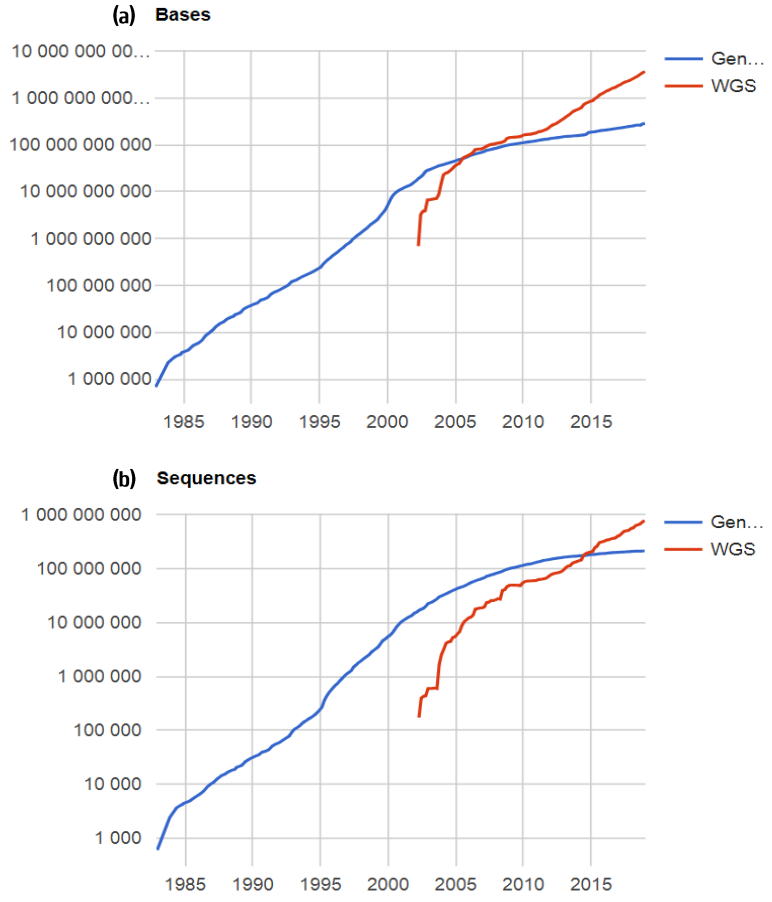


Figure 9 – Graphic representation of the growth of the number of bases (a), and the number of sequences (b), inputs release of GenBank and WGS from 1982 to 2018. Adapted from: https://www.ncbi.nlm.nih.gov/genbank/statistics/ (76).

Nowadays, we can determine that the contribution of computational techniques, especially the development of efficient algorithms, become indispensable for a good analysis of the generated data and that the richness of the partnership between informatics and biology, created the need for a deep interaction between computer specialists and biologists, in other to be able to understand and perform rigorously bioinformatic tasks (21,24,75).

The biological data produced increased exponentially since the 90s and due to new methods of genome sequencing, it is possible to verify that the growth is still picking. As shown in Figure 9, the WGS database has been receiving more records of bases and sequences in the last years surpassing the GenBank database (25,76). In addition, as of February 2019, the GenBank repository of nucleic acid sequences had contained 11,546,000 entries and the SWISS-PROT database of protein sequences 559,077 reviewed entries.

As a result of this increase in data flow, bioinformatic tools have become indispensable to the search, management and store of large volumes of biological data, and research of the complex dynamics observed in nature. Currently, there is a wide range of computational tools such as sequence and structural alignment algorithms, database design and data mining programs, phylogenetic tree construction software's, prediction of protein structure and function algorithms, gene finding, and expression data clustering methods to performed rigorously and efficiently biological and mathematical analyses (23,24).

5.3. Challenges in bioinformatics tasks

The massive management of biological data brings with it a range of vital technical and scientific challenges, such as data storage capacity, structure and accurate representation of them, easy access and manipulation of these data by user, processing speed, different file formats and integration of multiple tools (21,25).

Therefore, numerous databases and computational tools have been developed to allow access to the entire scientific community to the different genomic data available, as well as the comparative analysis of them. Currently, there are numerous options for visualization, searching, obtaining and analysing biological data, allowing the achievement of knowledge about the genomes and their respective organisms. However, this knowledge is many times fragmented and dispersed through many different computational resources, often in a redundant way (21,24).

There is a need to unify many of these resources so that we can have an integrated and global view of the biology of all these studied genomes and species. In combining reviewed data obtained *in silico*, it increases the quality of our studies and also improves the possibility to have efficient structure, storage and data processing, allowing dynamic visualization, search, retrieval and analysis of data in a dynamic, flexible and fast way, through a user-friendly graphical interface (21,77).

5.4. Bioinformatic algorithms to compare sequences

Sequence alignment algorithms calculate sequence similarity between query sequence and database and compute the statistical significance. The similarity is a quantitative measure of the proportion of identical matches and conserved substitutions in a pairwise or multiple alignments (24).

There are two types of sequence alignment strategies: global or local alignment. In figure 10 are represented the main differences between global and local pairwise alignment.

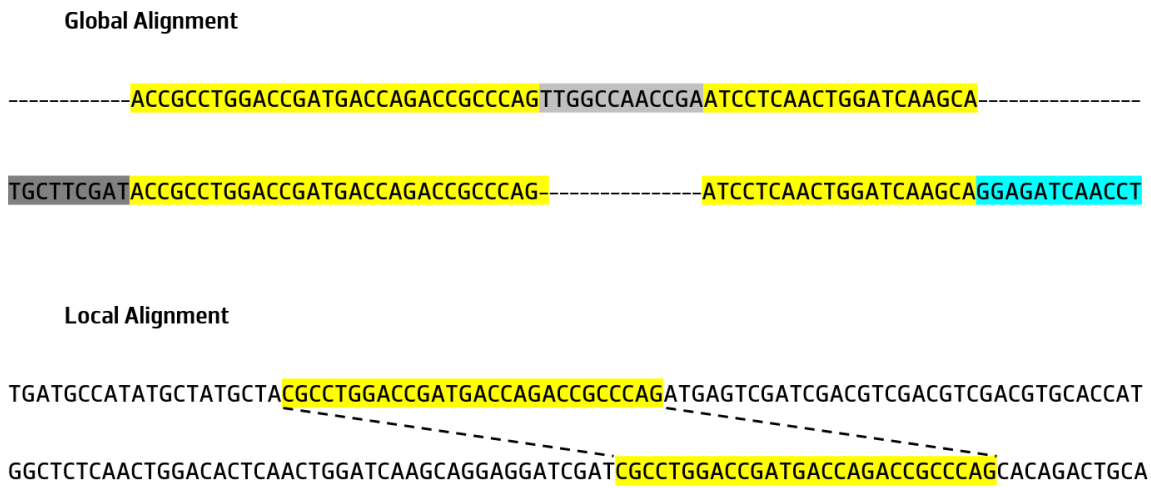


Figure 10 – Representation of global and local alignment.

In global alignment, the two or more sequences to be aligned are presumed to be generally similar over their entire length. Therefore, the alignment is carried out to the entire length of both sequences to find the best possible between the two sequences. This method is more applicable for aligning two closely related sequences of roughly the same length. For divergent sequences and sequences of variable lengths, this method may not be able to generate optimal results because it fails to recognize highly similar local regions between the two sequences and may also include large stretches of low similarity (24,27,78).

The classical global pairwise alignment algorithm that uses dynamic programming is the Needleman–Wunsch algorithm and was the first sequence alignment algorithm to be developed in 1970 (24,27,78).

As a side note, dynamic programming is a comprehensive and quantitative method to find optimal alignments. This method works in three steps. It first produces a sequence versus sequence matrix. Secondly accumulate scores in the matrix. Finally traces back through the matrix in reverse order to identify the highest scoring path. This scoring step involves the use of scoring matrices and gap penalties (24,27).

On the other hand, local alignment does not assume that the two sequences in question have similarity over the entire length. The goal is to get the highest alignment score locally, which may be at the expense of the highest possible overall score for a full-length alignment. It only finds local regions with the highest level of similarity between the two sequences and aligns these regions without regard for the alignment of the rest of the sequence regions. This approach can be used for aligning more divergent sequences with the goal of searching for conserved patterns in DNA or protein sequences. The two sequences to be aligned can be of different lengths (24,27,78).

The first application of dynamic programming in local alignment is the Smith-Waterman algorithm (27). The most widely used local alignment algorithm is BLAST, Basic Local Alignment Search Tool (24).

5.4.1. Basic Local Alignment Search Tool (BLAST)

The Basic Local Alignment Search Tool, or simply known as BLAST, is a program that was developed by Stephen Altschul of NCBI in 1990 and has since become one of the most popular programs for sequence analysis (24,79–81).

BLAST is a sequence similarity search program that can be used via a web interface or as a stand-alone tool to compare a user's query to a database of sequences (24,79–81). Putting it simply BLAST finds regions of local similarity between sequences and calculates the statistical significance of the matches to help decipher the biological significance of the alignment known as the E value ('Expect' value or false-positive rate). As such, BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families (79–81).

Nowadays, the BLAST server at the NCBI has a diverse set of features that can add power to your BLAST searching. In fact, BLAST can be considered a family of programs that includes different variants to compare all combinations nucleotide or protein queries with nucleotide or protein databases, known as BLASTn, BLASTp, BLASTx, tBLASTn, and tBLASTx (24,79–81):

- BLASTn queries nucleotide sequences with a nucleotide sequence database.
- BLASTp uses protein sequences as queries to search against a protein sequence database.
- BLASTx uses nucleotide sequences as queries and translates them in all six reading frames to produce translated protein sequences, which are used to query a protein sequence database.
- tBLASTn queries protein sequences to a nucleotide sequence database with the sequences translated in all six reading frames.
- tBLASTx uses nucleotide sequences, which are translated in all six frames, to search against a nucleotide sequence database that has all the sequences translated in six frames.

Besides these variants, there are also many more different programs as for instance the bl2seq program that performs local alignment of two user-provided input sequences. The graphical output includes horizontal bars and a diagonal in a two-dimensional diagram showing the overall extent of matching between the two sequences (24).

5.5. Bioinformatic tools directed to AMR in phages

Regarding the development of bioinformatics tools directed to the impeding antimicrobial resistance (AMR) in phages, different algorithms and programs were created and thought out from various approaches as can be shown in Figure 11.

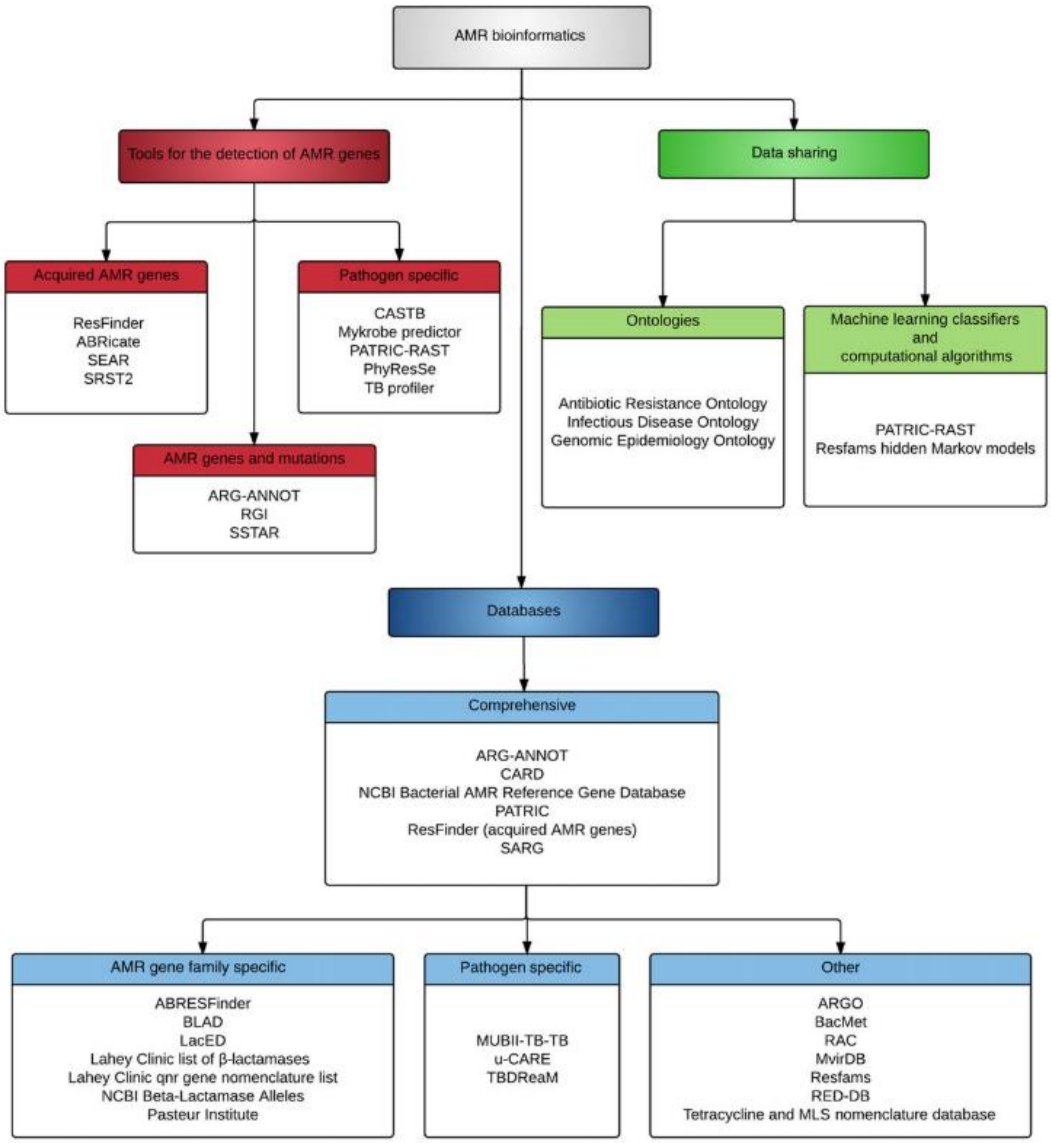


Figure 11 – Schematic representation of antimicrobial resistance bioinformatics tools available, such as software, databases, and data-sharing resources (77) Copyright © 2016 New York Academy of Sciences. Obtained permission to use by the author.

The development of new and improved AMR bioinformatics tools and databases intends to advance our understanding of the molecular mechanisms and improve our ability to accurately predict and screen for antibiotic resistance genes within environmental, agricultural, and clinical settings (77).

The ultimate goal of bioinformatics is to develop new *in silico* methods for the accurate detection of the AMR and subsequent accurate prediction of the antibiogram (the phenotypic range of AMR and susceptibility) from genomic and metagenomic data (77). As such, AMR bioinformatics tools can intervene in different scopes and functions, each with its own strengths and weaknesses (77). Many of them were first developed to help in metagenomic analyses and to search for AMR in other organisms or mobile genetic elements, as for tools like ResFinder that focus mainly on plasmid-borne resistance and detection of known clinical genes and disregard intrinsic resistance encoded by mutations (77).

Metagenomics is transforming our understanding of microbial communities by providing insight into the genetic diversity of uncultured microorganisms. When joined with next-generation sequencing (NGS) and whole-genome sequencing (WGS), the overflowing of sequence data obtained can represent both dominant and rare species in the environment. Meaning, these approaches can afford the genetic basis for understanding the relationship between uncultured bacteria, their phage and the environment (82).

Undoubtedly, *in silico* approaches such as these have been and are fundamental to studying bacteriophages (dsDNA viruses that infect bacteria) since *in vitro* efforts are limited to the reduced number of bacterial hosts that can be cultivated corresponding to approximately to 1% (82,83). However, it is important to understand that despite the advances in bioinformatics databases, software, and tools, there are still limitations and areas of unmet need preventing us from eventually reach rapid, cost-effective, and integrated AMR surveillance, in addition to point-of-care response (77).

6. Results

6.1. Screening results

A FASTA file with the nucleotide sequences of 4789 antibiotic resistance genes available in NCBI Bacterial Antimicrobial Resistance Reference Gene Database (Accession: PRJNA313047; ID: 313047) was retrieved and then submitted to BLASTn against the organisms found in *Caudovirales* (taxid:28883).

It is believed that approximately 96% of all the known phages are double-stranded DNA bacteriophages belonging to the order *Caudovirales*. As of February of 2019, *Caudovirales* order had a total of the 2051 complete genomes in NCBI, corresponding to 1125 of *Siphoviridae*, 543 of *Myoviridae*, 355 of *Podoviridae*, 21 of *Akermannviridae*, and 7 unclassified *Caudovirales* (26).

Table 2 – Phage hosts from *Caudovirales* (taxid:28883) where it was found significant similarity at a 99% confidence interval to different antibiotic resistance genes.

Taxonomy *(Caudovirales)*	**Phage host**	**Bacterial host** (taxonomic family)	**Accession no.**	**Genes**
Myoviridae; P1*virus*.	*Escherichia virus* P1	*Enterobacteriaceae*	MH445380.1	*aadA*; bla_{LEN}; bla_{OKP}; bla_{TEM}; *cmlA*; *cmlB*; *dfr*A14; *qacF*; *qacL*; *mef*(B); *mph*(A); *sul3*;
Myoviridae; P1*virus*.	*Enterobacteria phage* P7	*Enterobacteriaceae*	AF503408.1	$bla_{TEM-237}$;
Myoviridae; P1*virus*; unclassified *Punalikevirus*	*Salmonella phage* SJ46	*Enterobacteriaceae*	KU760857.1	$bla_{CTX-M-215}$; $bla_{CTX-M-223}$
Myoviridae; P1*virus*; unclassified *Punalikevirus*	*Escherichia phage* RCS47	*Enterobacteriaceae*	FO818745.1	bla_{LEN}; bla_{OKP}; $bla_{SHV-111}$;
Myoviridae; *Tevenvirinae*; T4*virus*	*Enterobacteria phage* T4	*Enterobacteriaceae*	KY304109.1	$bla_{CTX-M-186}$
Siphoviridae; *unclassified Siphoviridae*	*Escherichia phage* 1720a-02	*Enterobacteriaceae*	KF030445.1	*aph*(3')-IIIa; *cat*A1;
Siphoviridae; *unclassified Siphoviridae*	*Erysipelothrix phage* phi1605	*Erysipelotrichaceae*	MF172979.1	*mef*(A); *msr*(D); *tet*(M); *tet*(S)
Siphoviridae; *unclassified Siphoviridae*	*Streptococcus phage* IPP61	*Streptococcaceae*	KY065497.1	tet(M); tet(S);
Siphoviridae; *Spbetavirus*; unclassified *Spbetalikevirus*.	*Staphylococcus phage* SPbeta-like	*Staphylococcaceae*	KT429160.1	*aac*(6')-Ie2/*aph*(2")-If2; *aph*(2")-If; *dfr*C;
Siphoviridae; *Wbetavirus*; unclassified *Wbetavirus*	*Bacillus anthracis phage Gamma*	*Bacillaceae*	DQ222853.1	*fos*B2;
Siphoviridae; *Wbetavirus*; unclassified *Wbetavirus*	*Bacillus anthracis phage Cherry*	*Bacillaceae*	DQ222851.1	*fos*B2;
unclassified *Caudovirales*;	*Streptococcus phage* phiJH1301-2	*Streptococcaceae*	KX077896.1	*aac*(6')-Ie2/*aph*(2")-If2; *ant*(6)-Ia; *aph*(2")-If; *mef*(A);
unclassified *Caudovirales*;	*Streptococcus phage* phiSC070807	*Streptococcaceae*	KT336321.1	*ant*(6)-Ia; *aph*(3')-IIIa; *cfr*(C); *erm*(T); *mef*(A); *msr*(D); *sat4*;
unclassified *Caudovirales*;	*Streptococcus phage* phi-SsUD.1	*Streptococcaceae*	FN997652.1	*ant*(6)-Ia; *aph*(3')-IIIa; *cfr*(C); *erm*(T); *erm*(B); *sat4*; *tet*(W);
unclassified *Caudovirales*;	*Streptococcus phage* phi1207.3	*Streptococcaceae*	AY657002.1	*mef*(A); *msr*(D);
unclassified *Caudovirales*;	*Streptococcus phage* phiNJ3	*Streptococcaceae*	KT336320.1	*mef*(A); *msr*(D); *tet*(32); *tet*(O);

The FASTA file submitted had a 5,19 MB in size and it took about 55 minutes to run the analyses.

Results found significant similarity at a 99% confidence interval within 16 different phage host (Table 2) corresponding to a total of 32 different antibiotic resistance genes (Table 3).

As the results can show, in Table 2, in the same phage host many different antibiotic resistance genes can be present.

In the following Table 3 are presented the results for each antibiotic resistance genes where it was found significant similarity with the reference sequence input for each different gene. However, it is important to note that there are various variants present in different antibiotic resistance gene families. To reflect that, in Table 4, are shown some antibiotic resistance gene families and their number of different reference sequences annotated in NCBI Reference Gene Browser (NCBI link: https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/) (84).

Since the results showed in many gene families repetitive results for many variants due to the high percentage of similarity between the variants in the same gene family, in Table 3 there are present the results for the gene families disregarding the repetitive results for the many variants.

Table 3 – Antibiotic resistance gene family found to have a significant similarity at 99% confidence interval.

Antibiotic class	Gene family	Accession	Match	Iden.	E value	Query cover	Phage host	Bacterial family
Aminoglycoside	***aadA***	NG_052320.1	MH445380.1	99%	0.0	100%	*Escherichia virus* P1	*Ent.*
	***aac*(6')–Ie2/*aph*(2")–If2**	NG_062220.1	KT429160.1	84%	0.0	98%	*Staphylococcus phage* SPbeta-like	*Staph.*
			KX077896.1	84%	0.0	98%	*Streptococcus phage* phiJH1301-2	*Strep.*
	***ant*(6)–Ia**	NG_047394.1	KT336321.1	99%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	99%	0.0	100%	*Streptococcus phage* phi-SsUD.1	*Strep.*
			KX077896.1	89%	0.0	92%	*Streptococcus phage* phiJH1301-2	*Strep.*
	***aph*(2")–If**	NG_047405.1	KT429160.1	82%	0.0	87%	*Staphylococcus phage SPbeta-like*	*Staph.*
			KX077896.1	82%	0.0	87%	*Streptococcus phage* phiJH1301-2	*Strep.*
	***aph*(3')–IIIa**	NG_047418.1	KF030445.1	99%	0.0	100%	*Escherichia phage* 1720a-02	*Ent.*
			KT336321.1	100%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	100%	0.0	100%	*Streptococcus phage* phi-SsUD.1	*Strep.*
Beta-lactam	bla_{CTX-M}	NG_063838.1	KU760857.1	99%	0.0	100%	*Salmonella phage* SJ46	*Ent.*
			KY304109.1	96%	$5{,}00^{e-148}$	37%	*Enterobacteria phage* T4	*Ent.*
			KY304108.1	95%	$2{,}00^{e-141}$	36%	*Enterobacteria phage* T4	*Ent.*
	bla_{LEN}	NG_060510.1	MH445380.1	91%	0.0	100%	*Escherichia virus* P1	*Ent.*
			F0818745.1	91%	0.0	100%	*Escherichia phage* RCS47	*Ent.*
	bla_{OKP}	NG_049361.1	MH445380.1	88%	0.0	99%	*Escherichia virus* P1	*Ent.*
			F0818745.1	88%	0.0	99%	*Escherichia phage* RCS47	*Ent.*

	***bla*$_{SHV}$**	NG_050002.1	MH445380.1	99%	0.0	100%	*Escherichia virus* P1	*Ent.*
			F0818745.1	99%	0.0	100%	*Escherichia phage* RCS47	*Ent.*
	***bla*$_{TEM}$**	NG_062250.1	MH445380.1	99%	0.0	100%	*Escherichia virus* P1	*Ent.*
			AF503408.1	99%	0.0	100%	*Enterobacteria phage* P7	*Ent.*
Phenicol	***cat*A1**	NG_051704.1	KF030445.1	99%	0.0	100%	*Escherichia phage* 1720a-02	*Ent.*
	***cml*A**	NG_051879.1	MH445380.1	96%	0.0	100%	*Escherichia virus* P1	*Ent.*
	***cml*B**	NG_047657.1	MH445380.1	84%	0.0	99%	*Escherichia virus* P1	*Ent.*
Trimethoprim	***dfr*A14**	NG_047699.2	MH445380.1	99%	0.0	100%	*Escherichia virus* P1	*Ent.*
	***dfr*C**	NG_047752.1	KT429160.1	99%	0.0	100%	*Staphylococcus phage SPbeta-like*	*Staph.*
Fosfomycin	***fos*B2**	NG_050411.1	DQ222853.1	99%	0.0	100%	*Bacillus anthracis phage Gamma*	*Bac.*
			DQ222851.1	99%	0.0	100%	*Bacillus anthracis phage Cherry*	*Bac.*
			DQ289556.1	99%	0.0	100%	*Bacillus anthracis phage Gamma*	*Bac.*
Quaternary ammonium	***qac*F**	NG_052176.2	MH445380.1	91%	8,00^{e-126}	100%	*Escherichia virus* P1	*Ent.*
	***qac*L**	NG_048051.1	MH445380.1	99%	2,00^{e-172}	100%	*Escherichia virus* P1	*Ent.*
Macrolide	***erm*(T)**	NG_047840.1	KT336321.1	99%	8,00^{e-131}	35%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	99%	8,00^{e-131}	35%	*Streptococcus phage* phi-SsUD.1	*Strep.*
	***erm*(B)**	NG_047802.1	KT336321.1	99%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	99%	0.0	100%	*Streptococcus phage* phi-SsUD.1	*Strep.*
	***mef*(A)**	NG_047959.1	MF172979.1	90%	0.0	100%	*Erysipelothrix phage* phi1605	*Erys.*
			KT336321.1	90%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			AY657002.1	90%	0.0	100%	*Streptococcus phage* phi1207.3	*Strep.*
			KX077896.1	89%	0.0	100%	*Streptococcus phage* phiJH1301-2	*Strep.*
			KT336320.1	89%	0.0	100%	*Streptococcus phage* phiNJ3	*Strep.*
	***mef*(B)**	NG_047978.1	MH445380.1	99%	1,00^{e-133}	21%	*Escherichia virus* P1	*Ent.*
	***mph*(A)**	NG_047987.1	MH445380.1	99%	0.0	100%	*Escherichia virus* P1	*Ent.*
	***msr*(D)**	NG_048005.1	MF172979.1	93%	0.0	100%	*Erysipelothrix phage* phi1605	*Erys.*
			AY657002.1	100%	0.0	100%	*Streptococcus phage* phi1207.3	*Strep.*
			KT336321.1	99%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			KT336320.1	99%	0.0	100%	*Streptococcus phage* phiNJ3	*Strep.*
Sulfonamide	***sul*3**	NG_048120.1	MH445380.1	100%	0.0	100%	*Escherichia virus* P1	*Ent.*
Streptothricin	***sat*4**	NG_048073.1	KT336321.1	97%	0.0	100%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	97%	0.0	100%	*Streptococcus phage* phi-SsUD.1	*Strep.*
Tetracycline	***tet*(32)**	NG_048124.1	KT336320.1	94%	0.0	63%	*Streptococcus phage* phiNJ3	*Strep.*
	***tet*(O)**	NG_048267.1	KT336320.1	99%	0.0	100%	*Streptococcus phage* phiNJ3	*Strep.*
	***tet*(M)**	NG_048253.1	MF172979.1	98%	0.0	100%	*Erysipelothrix phage* phi1605	*Erys.*

			KY065497.1	96%	0.0	100%	*Streptococcus phage* IPP61	*Strep.*
	***tet*(S)**	NG_048275.2	MF172979.1	80%	0.0	99%	*Erysipelothrix phage* phi1605	*Erys.*
			KY065497.1	80%	0.0	99%	*Streptococcus phage* IPP61	*Strep.*
	***tet*(W)**	NG_055990.1	FN997652.1	90%	0.0	99%	*Streptococcus phage* phi-SsUD.1	*Strep.*
Macrolides Lincosamide-Streptogramin B (MLSB)	***cfr*(C)**	NG_060579.1	KT336321.1	100%	$3{,}00^{e-36}$	7%	*Streptococcus phage* phiSC070807	*Strep.*
			FN997652.1	100%	$3{,}00^{e-36}$	7%	*Streptococcus phage* phi-SsUD.1	*Strep.*

Legend: Iden. – Identity; E value – "Expect" value or false positive rate; *Bac.* – *Bacillaceae*; *Ent.* – *Enterobacteriaceae*; *Erys.* – *Erysipelotrichaceae*; Staph. – *Staphylococcaceae*; Strep. – *Streptococcaceae*;

From these results, it was, therefore, proven the presence of antibiotic resistance genes in bacteriophages at a 99% confidence interval.

The phage hosts found harbouring antibiotic resistance genes copulate to bacterial hosts from the main bacterial families namely *Enterobacteriaceae, Bacillaceae, Erysipelotrichaceae, Staphylococcaceae* and *Streptococcaceae.*

The antibiotic resistance genes found provide their bacterial host with the resistance to aminoglycosides, β-lactams, phenicols, trimethoprim, fosfomycin, quaternary ammonium compounds, macrolides, sulfonamides, streptothricins, tetracyclines and macrolides lincosamide-streptogramin B (MLSB) antibiotics, corresponding to some of the most important classes of antibiotics that target the bacterium through different ways.

Table 4– Antibiotic resistance gene family names and their number of different reference sequences annotated in NCBI Reference Gene Browser (84).

Antibiotic class	**Gene family (no. var.)**	**Product name**
Aminoglycoside	***aac*(3)-I (19)**	AAC(3)-I family aminoglycoside 3-N-acetyltransferase
	***aac*(6') (111)**	aminoglycoside 6'-N-acetyltransferase
	***aac*(6')-I (33)**	AAC(6')-Ighjkrstuvwx family aminoglycoside N-acetyltransferase
	***aac*(6')-Ib (32)**	AAC(6')-Ib family aminoglycoside 6'-N-acetyltransferase
	***aad*A1 (52)**	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA1
	***aad*A2 (32)**	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA2
	***aad*A5 (20)**	ANT(3")-Ia family aminoglycoside nucleotidyltransferase AadA5
	ant(2")-Ia (19)	aminoglycoside nucleotidyltransferase ANT(2")-Ia
	bla$_{ACT}$ (60)	cephalosporin-hydrolyzing class C beta-lactamase ACT
	bla$_{ADC}$ (114)	class C beta-lactamase ADC
	bla$_{CARB}$ (30)	carbenicillin-hydrolyzing class A beta-lactamase CARB
	bla$_{CMY}$ (146)	class C beta-lactamase CMY
	***bla*$_{CTX-M}$ (202)**	class A extended-spectrum beta-lactamase CTX-M
	bla$_{DHA}$ (26)	class C beta-lactamase DHA

Beta-lactam	*bla*$_{GES}$ (39)	beta-lactamase GES
	bla$_{IMI}$(16)	carbapenem-hydrolyzing class A beta-lactamase IMI
	bla$_{IMP}$ (71)	subclass B1 metallo-beta-lactamase IMP
	bla$_{IND}$(17)	subclass B1 metallo-beta-lactamase IND
	bla$_{KPC}$ (34)	carbapenem-hydrolyzing class A beta-lactamase KPC
	bla$_{L1}$ (17)	L1 family subclass B3 metallo-beta-lactamase
	***bla*$_{LEN}$ (23)**	class A beta-lactamase LEN
	bla$_{MIR}$ (20)	cephalosporin-hydrolyzing class C beta-lactamase MIR
	bla$_{MOX}$ (19)	CMY-1/MOX family class C beta-lactamase CMY
	bla$_{NDM}$(24)	subclass B1 metallo-beta-lactamase NDM
	***bla*$_{OKP}$ (35)**	broad-spectrum beta-lactamase OKP
	bla$_{OXA}$ (673)	oxacillin-hydrolyzing class D beta-lactamase OXA
	bla$_{OXY}$ (28)	class A extended-spectrum beta-lactamase OXY
	bla$_{PDC}$ (295)	cephalosporin-hydrolyzing class C beta-lactamase PDC
	***bla*$_{SHV}$ (178)**	class A beta-lactamase SHV
	***bla*$_{TEM}$ (179)**	class A beta-lactamase TEM
	bla$_{VEB}$ (19)	class A extended-spectrum beta-lactamase VEB
	bla$_{VIM}$ (56)	subclass B1 metallo-beta-lactamase VIM
Colistin	*mcr-1* (18)	phosphoethanolamine--lipid A transferase MCR-1
	mcr-3 (25)	phosphoethanolamine--lipid A transferase MCR-3
	*pmr*B (23)	two-component sensor histidine kinase
Glycopeptide	*van*R (28)	vancomycin resistance DNA-binding response regulator VanR
	*van*S (28)	vancomycin resistance histidine kinase VanS
Macrolide	*mef*(A) (21)	macrolide efflux MFS transporter Mef(A)
Phenicol	***cat*A (19)**	type A chloramphenicol O-acetyltransferase
	*flo*R (18)	chloramphenicol/florfenicol efflux MFS transporter FloR
Polyketide	*tuf*A (19)	translation elongation factor EF-Tu 1
Quinolone	*gyr*A (75)	DNA gyrase subunit A
	*par*C (27)	DNA topoisomerase IV subunit A
	*par*E (20)	DNA topoisomerase IV subunit B
	*qnr*B(81)	quinolone resistance pentapeptide repeat protein QnrB
Sulfonamide	*sul*1(23)	sulfonamide-resistant dihydropteroate synthase Sul1
Tetracycline	*tet*(M) (42)	tetracycline resistance ribosomal protection protein Tet(M)
	tet(W) (24)	tetracycline resistance ribosomal protection protein Tet(W)
Phenicol, Quinolone	*oqx*B (32)	multidrug efflux RND transporter permease subunit OqxB
Aminoglycoside, Tetracycline, Edeine	16S_rrsB (17)	16S ribosomal RNA
Macrolide, Oxazolidinone, Phenicol	23S (19)	23S ribosomal RNA

Legend: At bold the gene family names where results found significant similarity. no. var – number of different variants annoted

6.2. Phylogeny results

A FASTA file with the 15 complete genomes (exception of *Enterobacteria phage* T4 because the complete genome was not available) where it was found antibiotic resistance genes were submitted to multiple sequence alignment by Clustal W in MEGA X software.

The phylogeny method chosen to construct the tree was the UPGMA method. The phylogenetic tree was drawn to scale with the sum of branch length = 8.16137594 as shown in Figure 12. The evolutionary distances were computed using the Maximum Composite Likelihood method and the codon positions included were 1st+2nd+3rd+Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There was a total of 179367 positions in the final dataset.

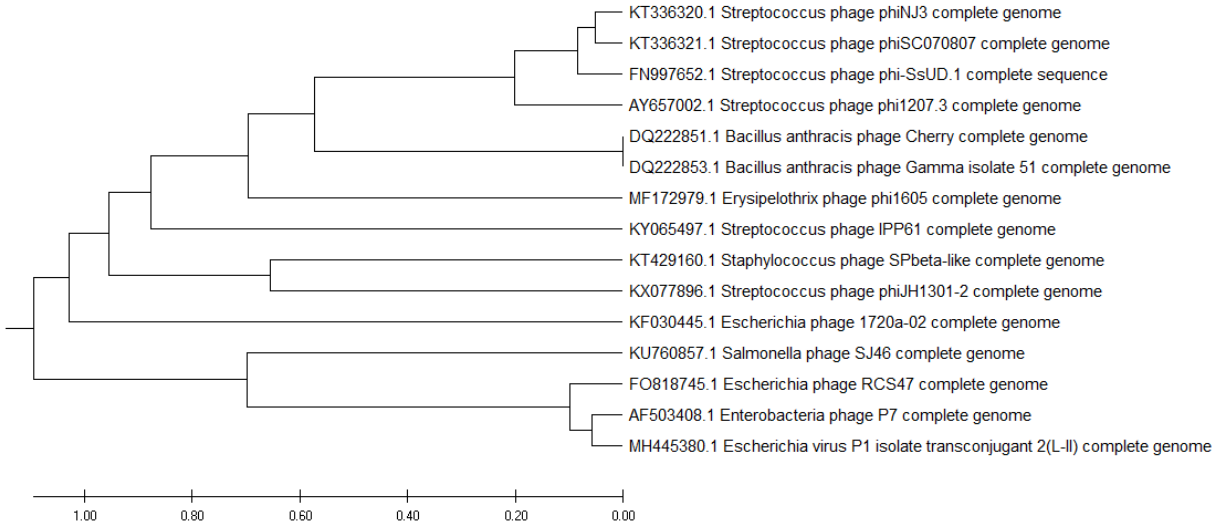


Figure 12 – Representation of the optimal tree of 15 phage genomes where the evolutionary history inferred by UPGMA method. The tree was drawn to scale.

Results indicate diversity between the 15 phage genomes as expected. The clusters where it was found proximity correspond mostly to genomes classified from the same genus bacterial host.

7. Discussion

7.1. Discussion of *in silico* method

In the course of this dissertation arisen some obstacles that lead in the end to the choice of one *in silico* method to perform the screening of the presence of several antibiotic resistance genes in various bacteriophages genomes.

The obstacles circle around the challenges in choosing the best adequate bioinformatic tools to perform the task at hand, as mention previously in chapter "5. State-of-Art – Bioinformatics". The easy access to diverse databases and user-friendly software and bioinformatics tools play heavily in the decision. Ultimately, the aim of this dissertation was reached by screening 4789 antimicrobial resistance genes in 2051 genomes from the *Caudovirales* order by the means of the *in silico* method chosen BLAST. Through this *in silico* method it was possible to prove the presence of ARGs in bacteriophages genomes corroborating with the premise that the horizontal gene transfer between bacteria by the means of transduction, where the antibiotic resistance genes are transported from one bacterium to another by bacteriophages and then integrated into the chromosome of the recipient cell (lysogeny) is possible and occurs *in vivo* (54,85,86).

In the selection process of the *in silico* method to use it was considered and experimented other software, databases and computational tools run in different operating systems. Due to limitations in access, user-friendly to the researchers, up to date and diverse of databases, in the end, was chosen to perform the screening using the computer program BLAST. Since this bioinformatic tool performs the adequate alignment to compare sequences with different lengths, the local alignment where finds regions of local similarity between sequences and calculates the statistical significance of the matches (24). Nonetheless, this *in silico* method is accessible to everyone online, is a user-friendly and time-saving method and most importantly the databases within BLAST is the utmost diverse and up to date databases found (79–81).

For the means of presenting an example, it was performed a screening using the ResFinder program against the same 2050 genomes belonging to the *Caudovirales* order where the results were considerably different. It was only found 2 ARGs among the 2050 genomes, to see in more detail results go to Annexe I. Understandably the results differ because the number of ARGs in their database is inferior to the 4789 ARGs in the NCBI reference sequence databases. Also, ResFinder program was designed to easily detect the antibiotic resistant classes of the ARGs found in genomes.

In addition, the download and treatment of results may differ between the different methods. One limitation in using BLAST, and that can be also found in other computational tools, is that doesn`t allow the download of the data results in formats that permit further the visualization of data or data mining. An example to represent this struggle is the found similarity in many variants to the gene family. When dealing

with many results and different variants it would be more diligent to be possible to download the results in different formats or even a text format to then be possible to analyse them further with a resource to other computer tools.

That being said, it is recommended the use of this *in silico* method in other to perform screening similar to the one performed in this work. However, if the goal is to screen ARGs in genomes not available in NCBI but some obtain *in vitro,* perhaps, it would be recommended to use software best suited for the task at hand, for instance, ResFinder or another options mention in chapter "5. State-of-Art – Bioinformatics". Where further considerations such as up to date and diversity of the database, the operating system, programming language, size of data and the time-consuming versus time-saving task would weigh on the decision.

7.2. Discussion of results

In the following points, the results found are discussed by their phage host that consequently can copulate with bacteria from the same taxonomic level or even higher taxonomic levels. To then, discuss the results by their antibiotic classes and mechanisms of action targeting the prokaryotic cell, finishing by discussing the results considering the mechanism of antibiotic resistance against the antibiotic classes.

7.2.1. Discussion of results by their hosts

From the search in 2051 phage genomes was only found matches in 15 complete genomes corresponding to less than 1%, agreeing with the findings in Enault *et al.* 2017, that reaffirm phages rarely carry antibiotic resistance genes (87). The phages genomes hosts of these ARGs are found to copulate with bacteria from the *Escherichia, Salmonella, Bacillus, Erysipelothrix, Staphylococcus* and *Streptococcus* genera, corresponding respectively to the bacterial family *Enterobacteriaceae* (both *Escherichia* and *Salmonella* genera), *Bacillaceae, Erysipelotrichaceae, Staphylococcaceae* and *Streptococcaceae*. Therefore, including some of the most concerning pathogenic bacteria from both gram-negative bacteria (*Escherichia* and *Salmonella* genera) and gram-positive bacteria (*Bacillus, Erysipelothrix, Staphylococcus* and *Streptococcus* genera) (45). As for also including genera from the ESKAPE pathogens that are of particular concern, due to the fact of these bacteria have rapidly acquired resistance to key antibiotics (1,18).

Furthermore, the phylogenetic results echoed the presence of ARGs in diverse taxonomic levels of the phage-host despite the mechanism of transduction between phage and bacterium being species-specific, meaning only infect a single, bacterial species, or even specific strains within a species (59,60,88). The dissemination of ARGs easily reached across different taxonomic levels because the horizontal gene transfer between bacteria may occur among diverse taxonomic levels (6,53,89).

7.2.1.1 Gram-negative bacteria: *Enterobacteriaceae*

The *Enterobacteriaceae* family is a big family group of gram-negative with a total of 50 genera and thus including 11 of the most important genera from a clinical point of view (34). This family have been extensively linked to a diverse range of infections specifically urinary tract infections (UTIs), bloodstream infections, nosocomial pneumonia and various intra-abdominal infections. With regards to the *Escherichia* and *Salmonella* genera within this family, the *Escherichia coli* is a frequent cause of intestinal infections and urinary tract infections (UTIs), and *Salmonella* sp. known to cause gastroenteritis and typhoid fever (38,90).

7.2.1.2 Gram-positive bacteria: *Bacillaceae, Erysipelotrichaceae, Staphylococcaceae* and *Streptococcaceae*

In regards to bacterial families that englobe the group of gram-positive bacteria, they can also cause a range of infections in humans. Within the gram-positive bacteria group, they can be divided into two main groups according to their cell shape in cell shape in bacilli, where are included the *Bacillus* and *Erysipelothrix* genera or in cocci, where are included the *Streptococcus* and *Staphylococcus* genera.

Taking into consideration the *Bacillus* and *Erysipelothrix* genera and more specifically the *Bacillus anthracis* is a gram-positive, endospore-forming, rod-shaped bacterium responsible for anthrax, that is a common disease of livestock and, occasionally, of humans. Whereas the *Erysipelothrix rhusiopathiae* infections most commonly present in a mild cutaneous form known as erysipeloid, skin infection in humans frequently found in the hands and fingers (38).

Taking into account the *Streptococcus* and *Staphylococcus* genera, the *Streptococcus* species may cause infections from the respiratory tract such as pneumonia, pharyngitis, tonsillitis and meningitis, or skin infections such as pink eye, erysipelas and necrotizing fasciitis (also known as 'flesh-eating' bacterial infection) (34,38). As for the *Staphylococcus* species are commonly linked to food poisoning and human skin infections, though can also be responsible for meningitis, pneumonia, endocarditis and urinary tract infections (38).

7.2.2. Discussion of results by antibiotic class

Results determined the presence of 32 different antibiotic resistance gene families, excluding the various variants where it was found repetitive results, corresponding to ARGs that are found to provide their bacterial host with the resistance to a diverse group of antibiotic classes, namely aminoglycoside, β-lactam, phenicol, trimethoprim, fosfomycin, quaternary ammonium, macrolide, sulfonamide, streptothricin, tetracycline and macrolide lincosamide streptogramin B (MLSB). Therefore, including the main groups of antibiotics used to combat bacterial infections (12,28,37,91) and corresponding to some of

the most important classes of antibiotics that target the bacterium through different mechanisms of action, please view Figure 2 in the chapter "3. State-of-Art – Antibiotic resistance" (6,7,31,37,38).

Among the five different mechanisms of action that the antibiotics can be divided based on their target in the bacterium, results shown that antibiotics targeting the protein synthesis (aminoglycosides, macrolides, phenicols, tetracyclines and streptothricins), the cell wall synthesis (β-lactams and fosfomycin), the folic acid synthesis (trimethoprim and sulfonamide) and the cell membrane (quaternary ammonium) would be ineffective against bacterial hosts carrying the ARGs found in the phage genomes screened (10,41,44,91,92).

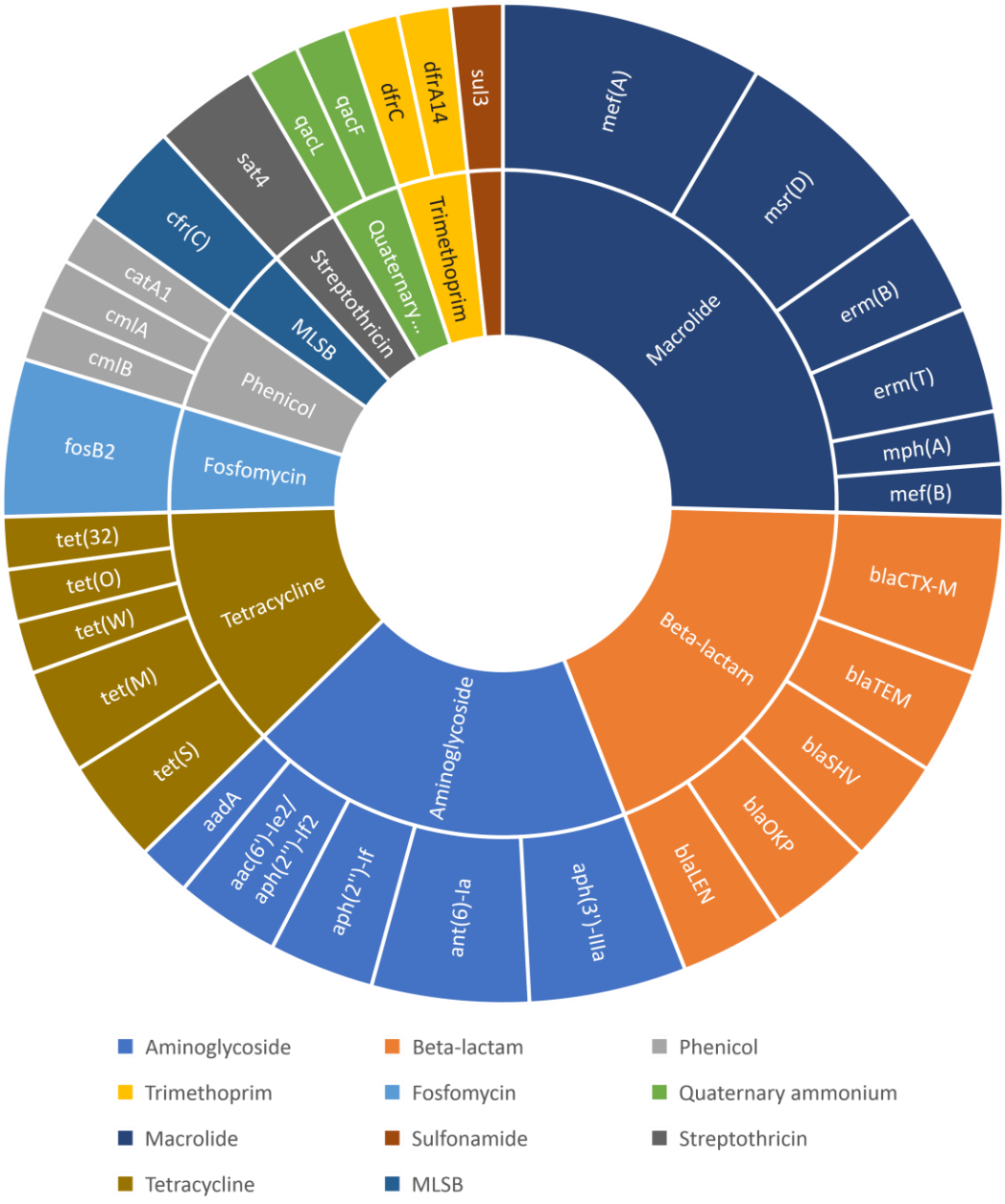


Figure 13 – Graphic representation of the number of matches found in each antibiotic resistance gene families sorted by their antibiotic classes targets.

As view in Figure 13, the macrolides, β-lactams, aminoglycosides and tetracyclines antibiotics correspond to the classes of antibiotics where it was found most diversity of antibiotic resistance gene families and most matches to ARGs annotated in NCBI Bacterial Antimicrobial Resistance Reference Gene Database (Accession: PRJNA313047; ID: 313047). Leaving only two different mechanisms of action left, the inhibitors of cell function (polymixin B and colistin) and inhibitors of nucleic acids synthesis (fluoroquinolones and rifampicin) that are antibiotics use frequently as last resort since are more toxic to the eukaryotic cells (16,17).

7.2.3. Discussion of results by the mechanism of antibiotic resistance

In addition to the diversity of ARGs found and the range of antibiotics classes that can be rendered ineffective to treat infections, there can be at play a diversity of mechanisms of antibiotic resistance.

Results showed that ARGs found in phage genomes provide their bacterial hosts with resistance to various antibiotics with different mechanisms of action, though the mechanisms of antibiotic resistance may also differ, please view Figure 1 in the chapter "3. State-of-Art – Antibiotic resistance" (14,28,32,35,36).

Among the antibiotic resistance gene families found in the phage genomes screened, they can provide the bacterium with resistance against antibiotics classes through different mechanisms of antibiotic resistance. Thus in Table 5, shows the mechanism of antibiotic resistance for each antibiotic resistance gene family found, as well as mentions the bacterial host families where it was found results and the antibiotic classes that are the target of this acquired resistance. More comprehensive information for each ARGs, as well as information on the mechanism of antibiotic resistance and mechanism of action at play can be easily and freely accessed online by The Comprehensive Antibiotic Resistance Database (CARD in link: https://card.mcmaster.ca/home) (93,94).

Overall, results reveal that a variety of mechanism of antibiotic resistance can be at play. According to The Comprehensive Antibiotic Resistance Database (CARD), the mechanisms of antibiotic resistance that can be found may be divided by seven, namely antibiotic target alteration, antibiotic target replacement, antibiotic target protection, antibiotic inactivation, antibiotic efflux, reduced permeability to antibiotic and resistance by absence (93,94). Therefore, from the results found, five of the seven mechanisms of antibiotic resistance can be at play.

The mechanisms of antibiotic resistance present where the antibiotic inactivation (antibiotic resistance achieved by enzymatic inactivation of antibiotic), the antibiotic efflux (antibiotic resistance via the transport of antibiotics out of the cell), the antibiotic target alteration (antibiotic resistance achieved by mutational alteration or enzymatic modification of antibiotic target), the antibiotic target replacement (antibiotic resistance reached by replacement or substitution of antibiotic action target) and the antibiotic target protection (protection of antibiotic action target from antibiotic binding, which process will result in

antibiotic resistance). Hence only leaving the resistance mechanism of reduced permeability to antibiotic (antibiotic resistance conquered generally through reduced production of porins) and resistance by absence (this mechanism of antibiotic resistance is conferred by deletion of the gene, typically a porin) as the mechanisms of antibiotic resistance not included within the results found (93,94).

Table 5 – Mechanism of resistance for each antibiotic resistance gene family found in phage genomes screened grouped by their antibiotic class.

Antibiotic class	Target	Gene family	Mechanism of resistance	Bacterial host family
Aminoglycoside	Protein synthesis	*aad*A *aac*(6')–Ie2/*aph*(2'')–If2 *ant*(6)–Ia *aph*(2'')–If *aph*(3')–IIIa	antibiotic inactivation	*Enterobacteriaceae; Staphylococcaceae; Streptococcaceae*
Beta–lactam	Cell wall synthesis	bla_{CTX-M} bla_{LEN} bla_{OKP} bla_{SHV} bla_{TEM}	antibiotic inactivation	*Enterobacteriaceae*
Phenicol	Protein synthesis	*cat*A1	antibiotic inactivation	*Enterobacteriaceae*
		*cml*A; *cml*B	antibiotic efflux	
Trimethoprim	Folic acid synthesis	*dfr*A14; *dfr*C	antibiotic target replacement	*Enterobacteriaceae; Staphylococcaceae*
Fosfomycin	Cell wall synthesis	*fos*B2	antibiotic inactivation	*Bacillaceae*
Quaternary ammonium	Cell membrane	*qac*F; *qac*L	antibiotic efflux	*Enterobacteriaceae*
Macrolide	Protein synthesis	*erm*(T); *erm*(B)	antibiotic target alteration	*Streptococcaceae; Erysipelotrichaceae*
		mef(A); *mef*(B)	antibiotic efflux	*Streptococcaceae; Enterobacteriaceae*
		mph(A)	antibiotic inactivation	*Enterobacteriaceae*
		msr(D)	antibiotic target protection	*Erysipelotrichaceae; Streptococcaceae*
Sulfonamide	Folic acid synthesis	*sul*3	antibiotic target replacement	*Enterobacteriaceae*
Streptothricin	Protein synthesis	*sat*4	antibiotic inactivation	*Streptococcaceae*
Tetracycline	Protein synthesis	*tet*(32); *tet*(O); *tet*(M); *tet*(S); *tet*(W)	antibiotic target protection	*Streptococcaceae; Erysipelotrichaceae*
Macrolides Lincosamide–Streptogramin B (MLSB)	Protein synthesis	*cfr*(C)	antibiotic target alteration	*Streptococcaceae*

7.3. Discussion of scope of the dissertation

Reflecting at the scope of this dissertation, without doubt, antibiotic resistance has been getting stronger and stronger. Since the discovery of antibiotics, the resistance to them has been rising and the production of antibiotics has been decreasing considerably since 1983, leading also to a declining in a choice of antibiotics (12,91,95–98).

The scope of this dissertation intended to focus in the dissemination of antibiotic resistance between bacteria through the mechanism of transduction since the horizontal gene transfer of antimicrobial resistance genes between clinical isolates via transduction is poorly understood (85,86).

As mentioned before bacteriophages are considered the most abundant biological entities on Earth, with an estimated total population of 10^{31}, 10 times larger than the estimated number of bacterial cells on Earth and are extremely diverse in size, morphology, and genomic organization (8,57,58,66). Although the transduction mechanism due occurs in the environment, it occurs at a low frequency, probably 1000 times less common than conjugation, therefore, the impact of dissemination of antibiotic resistance through this mechanism is very controversial (8,9,58,87).

In other to address this controversy was proposed to screen the presence of several antibiotic resistance genes in a library of bacteriophages by the means of computational tools and to hopefully chose the best *in silico* method or computational tool available to perform the screening.

Ultimately despite the diversity in AMR bioinformatic tools one of the first challenges was that no to very few ARGs in genomes were found when using different bioinformatic tools that have different databases many times not updated. Another limitation was the difficulty in using some of the software in different languages. Software language written in C++ allows faster code execution compared with the Perl and Python languages (24,83). Algorithms in these languages run faster and allow mining and management of big sizes of data. However, bioinformatics software often requires prior knowledge of UNIX-like operating systems, the utilization of command lines (for both installation and usage) and the installation of several software libraries (dependencies) before being usable, which can be unintuitive even for skilled bioinformaticians (24,25). On the positive side, more developers attempt to make their tools available to the scientific community through user-friendly graphical web servers, allowing to analyse data without having to perform demanding installation procedures (25). These pros and cons lead to the choice of the computational tool BLAST, mainly because is free web resource, user-friendly and very importantly with access to up to date and reviewed databases. Without reviewed and up to date databases the results obtained can be misleading. The reasons for that are that despite the experimental biology (that uses *in vitro* methods) and bioinformatics (that uses *in silico* methods) are independent activities, they are also complementary activities. Bioinformatics relies on experimental biology to produce raw data for analysis and in turn, provides useful interpretation and visualization of data that can

lead to further experimental research (24,25). Therefore, the tasks and analysis may often contain errors and the outcome also rests on the computer power available. In addition, it is important to understand the role that bioinformatics plays in the research. Most algorithms lack the ability and sophistication to accurately reflect reality. As such, caution should always be exercised when interpreting results and the use of multiple programs, if they are available, to perform multiple evaluations are advised (24).

Concerning the AMR bioinformatic tools, besides the challenges related to big sizes of data, velocity, variety, veracity and up to date data, comes also along with the challenges with data storage and data sharing. Most of AMR bioinformatics tools focus on the detection of known AMR sequences, however as it is known AMR genomic data are never stationary. As history has shown and has been showing AMR genes can suffer constant mutations generating several variants besides the emergence of new AMR genes (77). Hence, the significance and importance of using *in silico* methods like the one represented in this dissertation to efficiently and rapidly perform pairwise comparisons between emergent AMR genes to assess the changes between the AMR genes.

8. Conclusions and Future Work

Antibiotic resistance is threatening the way that we defend and prevent ourselves against pathogenic bacteria. In other to tackle this problem many fields work together to find understanding and manage the fight against the arising of antibiotic resistance. Bioinformatics pushed innovation in various fields such as experimental biology, molecular and cellular biology, and particularly contributed to an extensive insight in uncultured microorganisms.

This dissertation intended to screen the presence of several antibiotic resistance genes in a library of bacteriophages through the means of bioinformatic tools. The main goal was achieved and allow to confirm and validate the presence of ARGs in phages genomes through the *in silico* method chosen BLAST. Results determined the presence of 32 different antibiotic resistance gene families in 16 different phage hosts at a 99% confidence interval. From the search in 2051 genomes was only found matches in 15 complete genomes corresponding to less than 1%. Although the results showed strong diversity among the phage genomes where ARGs were found and included genera from the ESKAPE pathogens that are of concern, the impact of the spread of antibiotic resistance through the transduction mechanism remains unclear.

Nevertheless, this work contributed to assessing the use of bioinformatics tools to screen various ARGs in various genomes in a cost and time-friendly manner. Also, through studies such as this, it is easily possible to perform quick analyses and even preliminary studies in other to foreshow the pertinence of more comprehensive studies.

In hindsight, though this work didn`t have many expectations, had definitely some challenges and limitations that ultimately lead to the conclusion that it is important to consider different bioinformatic tools, but most of them work in different operating systems through different computer languages and have access to limited databases, therefore tangling the task of choosing the appropriate bioinformatic tool to use. Besides that, and as discussed before, the effectiveness of using *in silico* methods, as for instance BLAST, to screen big sizes of data can be challenging. Using the BLAST tool online is manageable with small sizes of data or with the obtention of a few results. That can`t be said with the opposite.

Without doubt studies such as this that of most importance to enlighten and help to search for ARGs in different microorganisms through the means of computational tools, and in a manner that is cost and time-friendly. However, the determination of the impact of dissemination of antibiotic resistance *in vivo* by transduction mechanism remains unclear. Further work can be designed to find answers to questions such as this that are still very much controversial.

For instance, since most algorithms have been developed for analysing bacteria. Future efforts could concentrate in the development of big data algorithms to compute large-scale pairwise comparisons between viromes to be able to create an organized resource for searching and identifying

unknown sequences or to rapidly compare complete viromes to large-scale reference databases that are also of need. In another view, future work could envelop the development of a resource design to help document, identified and understand the viral sequences unknow, a resource able to provide access to tools and databases specific to viruses that could be continually refined by the community (82).

In another hand, as history has shown, the potential of bacteriophages is unpredictable. Surely, we can expect much more in the coming years. Past applications with the utmost impact have been derived from phage biology (such as restriction enzymes and CRISPR–Cas) originated from curiosity-driven research on fundamental phage phenomena. Therefore, expect the unexpected, the results from one study might not answer that question but can answer other unforeseen questions.

To take as a final thought, tackling antimicrobial resistance become crucial nowadays. It needs to be viewed as a health threat but also as a major economic and security threat. As such it must be an issue of priority for heads of state, finance ministers, agriculture ministers, and health ministers (77).

References

1. Schürch AC, van Schaik W. Challenges and opportunities for whole-genome sequencing–based surveillance of antibiotic resistance. Ann N Y Acad Sci. 2017 Jan 1;1388(1):108–20.
2. Kumarasamy KK, Toleman MA, Walsh TR, Bagaria J, Butt F, Balakrishnan R, et al. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. Lancet Infect Dis. 2010 Sep;10(9):602–597.
3. Walsh TR. Emerging carbapenemases: a global perspective. Int J Antimicrob Agents. 2010 Nov 1;36 Suppl 3:S14–8.
4. Johnson AP, Woodford N. Global spread of antibiotic resistance: the example of New Delhi metallo-β-lactamase (NDM)-mediated carbapenem resistance. J Med Microbiol. 2013 Apr 1;62(Pt 4):513–499.
5. Livermore DM, Warner M, Mushtaq S, Doumith M, Zhang J, Woodford N. What remains against carbapenem-resistant Enterobacteriaceae? Evaluation of chloramphenicol, ciprofloxacin, colistin, fosfomycin, minocycline, nitrofurantoin, temocillin and tigecycline. Int J Antimicrob Agents. 2011 May;37(5):415–9.
6. Levy SB, Marshall B. Antibacterial resistance worldwide: causes, challenges and responses. Nat Med. 2004 Nov 30;10(12 Suppl):S122–9.
7. Tenover FC. Mechanisms of antimicrobial resistance in bacteria. Am J Infect Control. 2006 Jun;34(5 Suppl 1):S10–3; discussion S73–64.
8. Torres-Barceló C. The disparate effects of bacteriophages on antibiotic-resistant bacteria. Emerg Microbes Infect. 2018 Dec 10;7(1):168.
9. Colavecchio A, Cadieux B, Lo A, Goodridge LD. Bacteriophages Contribute to the Spread of Antibiotic Resistance Genes among Foodborne Pathogens of the Enterobacteriaceae Family – A Review. Front Microbiol. 2017;8:1108.
10. Chopra I, Greenwood D. Antibacterial Agents: Basis of Action. In: eLS. John Wiley & Sons, Ltd; 2001.
11. Jawetz E, Melnick JL, Adelberg EA, Brooks GF, Butel JS, Ornston LN. Medical Microbiology. 18th ed. Prentice Hall International inc.; 1989. 592 p.
12. Mohr KI. History of Antibiotics Research. In Springer, Cham; 2016. p. 237–72.
13. Larson E. Community factors in the development of antibiotic resistance. Annu Rev Public Health. 2007 Jan 16;28:435–47.
14. Lewis K. Platforms for antibiotic discovery. Nat Rev Drug Discov. 2013 May 30;12(5):387–371.
15. CDC. ANTIBIOTIC RESISTANCE THREATS in the United States, 2013. Centers Dis Control Prev. 2013;1–114.

16. Mazer-Amirshahi M, Pourmand A, May L. Newly approved antibiotics and antibiotics reserved for resistant infections: Implications for emergency medicine. Am J Emerg Med. 2017 Jan 1;35(1):154–8.
17. Van Boeckel TP, Gandra S, Ashok A, Caudron Q, Grenfell BT, Levin SA, et al. Global antibiotic consumption 2000 to 2010: An analysis of national pharmaceutical sales data. Lancet Infect Dis. 2014 Aug;14(8):742–50.
18. Tommasi R, Brown DG, Walkup GK, Manchester JI, Miller AA. ESKAPEing the labyrinth of antibacterial discovery. Nat Rev Drug Discov. 2015 Aug 3;14(8):529–42.
19. O'Neill J. Antimicrobial Resistance : Tackling a crisis for the health and wealth of nations. Review on Antimicrobial Resistance. 2014.
20. O'Neill J. TACKLING DRUG-RESISTANT INFECTIONS GLOBALLY: FINAL REPORT AND RECOMMENDATIONS. THE REVIEW ON ANTIMICROBIAL RESISTANCE. 2016.
21. Catanho MP, Miranda AB de, Degrave WMS. Comparando genomas: bancos de dados e ferramentas computacionais para a análise comparativa de genomas procarióticos. Reciis. 2007 Dec 31;1(2):335–58.
22. Huang J, Ru B, Dai P, Huang J, Ru B, Dai P. Bioinformatics Resources and Tools for Phage Display. Molecules. 2011 Jan 18;16(1):694–709.
23. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001;40(4):346–58.
24. Jin Xiong. Essential Bioinformatics. Vol. 53, Essential Bioinformatics. Cambridge University Press; 2013. 1689–1699 p.
25. Gauthier J, Vincent AT, Charette SJ, Derome N. A brief history of bioinformatics. Brief Bioinform. 2018 Aug 3; p. 1–16
26. NCBI. Complete genomes: Caudovirales [Internet]. [cited 2019 Feb 2]. Available from: https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=28883
27. Lambert C, Van Campenhout J, DeBolle X, Depiereux E. Review of Common Sequence Alignment Methods: Clues to Enhance Reliability. Curr Genomics. 2003;4:131–46.
28. Davies J, Davies D. Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev. 2010 Sep 1;74(3):417–33.
29. Leeb M. Antibiotics: A shot in the arm. Nature. 2004 Oct 21;431(7011):893–892.
30. Brown ED, Wright GD. Antibacterial drug discovery in the resistance era. Nature. 2016 Jan 21;529(7586):336–43.
31. Doyle MP, Busta F, Cords BR, Davidson PM, Hawke J, Hurd HS, et al. Antimicrobial resistance: Implications for the food system: An expert report, funded by the IFT Foundation. Compr Rev Food Sci Food Saf. 2006 Jul 1;5(3):71–137.

32. Barker KF. Antibiotic resistance: a current perspective. Br J Clin Pharmacol. 1999 Aug 1;48(2):124–109.
33. Bennett PM. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. Br J Pharmacol. 2008 Mar 1;153 Suppl(January):S347–57.
34. Ferreira, Wanda F. Canas, Sousa, Joao Carlos F., Lima N. Microbiologia. 1st ed. Lidel; 2010. 640 p.
35. Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ V. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol. 2015 Jan;13(1):42–51.
36. Coates A, Hu Y, Bax R, Page C. The future challenges facing the development of new antimicrobial drugs. Nat Rev Drug Discov. 2002 Nov;1(11):895–910.
37. van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP, Aarts HJM, et al. Acquired antibiotic resistance genes: an overview. Front Microbiol. 2011 Jan; 2:203.
38. Barroso H, Meliço-Silvestre, António. Taveira N. Microbiologia Médica – Volume 1. 1st ed. Lidel; 2014. 544 p.
39. Poole K. Resistance to β -lactam antibiotics. Cell Mol Life Sci C. 2004;61(17):2223–2200.
40. Richmond MH. B -lactam antibiotics – The background to their use as therapeutic agents. Frankfurt (Main): Hoechst Aktiengesellschaft; 1981.
41. Fernandes R, Amador P, Prudêncio C. β -Lactams: Chemical structure, mode of action and mechanisms of resistance. Vol. 24, Reviews in Medical Microbiology. 2013. p. 7–17.
42. Papp-Wallace KM, Endimiani A, Taracila MA, Bonomo RA. Carbapenems: past, present, and future. Antimicrob Agents Chemother. 2011 Nov;55(11):4943–60.
43. Worthington RJ, Melander C. Overcoming Resistance to β -Lactam Antibiotics. J Org Chem. 2013 May 3;78(9):4207–13.
44. Byarugaba D. Mechanisms of Antimicrobial Resistance. In: Sosa A de J, Byarugaba DK, Amábile-Cuevas CF, Hsueh P-R, Kariuki S, Okeke IN, editors. Antimicrobial Resistance in Developing Countries SE – 2. Springer New York; 2010. p. 26–15.
45. Rossolini GM, Arena F, Pecile P, Pollini S. Update on the antibiotic resistance crisis. Curr Opin Pharmacol. 2014 Oct 1;18:56–60.
46. Bush K, Jacoby GA. Updated functional classification of beta-lactamases. Antimicrob Agents Chemother. 2010 Mar;54(3):969–76.
47. Poirel L, Bonnin R a, Nordmann P. Genetic support and diversity of acquired extended-spectrum β -lactamases in Gram-negative rods. Infect Genet Evol. 2012 Jul;12(5):883–93.
48. Naseer U, Sundsfjord A. The CTX-M conundrum: dissemination of plasmids and Escherichia coli clones. Microb Drug Resist. 2011 Mar;17(1):97–83.
49. Zhao W-H, Hu Z-Q. Epidemiology and genetics of CTX-M extended-spectrum β -lactamases in Gram-negative bacteria. Crit Rev Microbiol. 2013 Feb 20;39(1):101–79.

50. Walsh TR, Toleman MA. The new medical challenge: why NDM-1? Why Indian? Expert Rev Anti Infect Ther. 2011 Feb 23;9(2):137–41.
51. Toleman MA, Spencer J, Jones L, Walsh TR. blaNDM-1 is a chimera likely constructed in Acinetobacter baumannii. Antimicrob Agents Chemother. 2012 May 1;56(5):2773–6.
52. Livermore DM. Current epidemiology and growing resistance of gram-negative pathogens. Korean J Intern Med. 2012 Jun;27(2):128–42.
53. Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet. 2015 Aug 1;16(8):472–82.
54. Furuya EY, Lowy FD. Antimicrobial-resistant bacteria in the community setting. Nat Rev Microbiol. 2006 Jan 1;4(1):36–45.
55. Sommer MOA, Munck C, Toft-Kehler RV, Andersson DI. Prediction of antibiotic resistance: time for a new preclinical paradigm? Nat Rev Microbiol. 2017 Jul 31;15(11):689–96.
56. Salmond GPCC, Fineran PC. A century of the phage: Past, present and future. Nat Rev Microbiol. 2015 Dec 9;13(12):777–86.
57. Barroso H, Meliço-Silvestre, António. Taveira N. Microbiologia Medica - Volume 2. 1st ed. Lisboa: Lidel; 2014. 592 p.
58. Balcazar JL. Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. PLoS Pathog. 2014 Jul;10(7):e1004219.
59. Kasman LM, Porter LD. Bacteriophages. [Updated 2019 Aug 12]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2019 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK493185/
60. Hatfull GF, Hendrix RW. Bacteriophages and their genomes. Curr Opin Virol. 2011;1(4):298–303.
61. Boyd EF. Bacteriophage-Encoded Bacterial Virulence Factors and Phage-Pathogenicity Island Interactions. In: Advances in Virus Research. 1st ed. Elsevier Inc.; 2012. p. 91–118.
62. Ofir G, Sorek R. Contemporary Phage Biology: From Classic Models to New Insights. Cell. 2018;172(6):1260–70.
63. Adriaenssens EM, Rodney Brister J. How to name and classify your phage: An informal guide. Viruses. 2017;9(4):1–9.
64. Aiewsakun P, Simmonds P. The genomic underpinnings of eukaryotic virus taxonomy: Creating a sequence-based framework for family-level virus classification. Microbiome. 2018;6(1):1–24.
65. Mahmoudabadi G, Phillips R. A comprehensive and quantitative exploration of thousands of viral genomes. Elife. 2018 Apr 19;7.
66. Fokine A, Rossmann MG. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage. 2014 Jan 1;4(1):e28281.
67. Ackermann H-W. Phage Classification and Characterization. In Humana Press; 2009. p. 127–40.

68. King AMQ, Lefkowitz EJ, Mushegian AR, Adams MJ, Dutilh BE, Gorbalenya AE, et al. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2018). Arch Virol. 2018 Sep 12;163(9):2601–31.
69. Pietilä MK, Demina TA, Atanasova NS, Oksanen HM, Bamford DH. Archaeal viruses and bacteriophages: comparisons and contrasts. Trends Microbiol. 2014 Jun 1;22(6):334–44.
70. Mc Grath S, Sinderen D van. Bacteriophage : genetics and molecular biology. 1st ed. Caister Academic Press; 2007. 344 p.
71. Feiner R, Argov T, Rabinovich L, Sigal N, Borovok I, Herskovits AA. A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. Vol. 13, Nature Reviews Microbiology. Nature Publishing Group; 2015. p. 641–50.
72. Casjens S. Prophages and bacterial genomics: What have we learned so far? Mol Microbiol. 2003;49(2):277–300.
73. Colomer-Lluch M, Jofre J, Muniesa M. Antibiotic Resistance Genes in the Bacteriophage DNA Fraction of Environmental Samples. Aziz R, editor. PLoS One. 2011 Mar 3;6(3): e17549.
74. Nature. Computational biology and bioinformatics - Latest research and news | Nature [Internet]. [cited 2019 Feb 1]. Available from: https://www.nature.com/subjects/computational-biology-and-bioinformatics
75. Ribeiro Junior HL, Oliveira RTG de, Ceccatto VM. Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. Acta Sci Educ. 2012 May 18;34(1):129–40.
76. NCBI. GenBank and WGS Statistics [Internet]. [cited 2019 Feb 2]. Available from: https://www.ncbi.nlm.nih.gov/genbank/statistics/
77. McArthur AG, Tsang KK. Antimicrobial resistance surveillance in the genomic age. Ann N Y Acad Sci. 2017 Jan 1;1388(1):78–91.
78. Issa M, Hassanien AE, Helmi A, Ziedan I, Alzohairy A. Pairwise Global Sequence Alignment Using Sine-Cosine Optimization Algorithm. Adv Intell Syst Comput. 2018; 723:102–11.
79. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004 Jul 1;32(Web Server): W20–5.
80. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, Mcginnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36:5–9.
81. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2019;
82. Hurwitz BL, U'Ren JM, Youens-Clark K. Computational prospecting the great viral unknown. Millard A, editor. FEMS Microbiol Lett. 2016 May 1;363(10):1–12.

83. Almeida OGG, De Martinis ECP. Bioinformatics tools to assess metagenomic data for applied microbiology. Appl Microbiol Biotechnol. 2018;103(1):69–82.
84. NCBI. Isolates Browser - Pathogen Detection - NCBI [Internet]. 2019 [cited 2019 Jun 12]. Available from: https://www.ncbi.nlm.nih.gov/pathogens/isolates#/refgene/
85. Stanczak-Mrozek KI, Manne A, Knight GM, Gould K, Witney AA, Lindsay JA. Within-host diversity of MRSA antimicrobial resistances. J Antimicrob Chemother. 2015 Aug 1;70(8):2191–8.
86. Stanczak-Mrozek KI, Laing KG, Lindsay JA. Resistance gene transfer: induction of transducing phage by sub-inhibitory concentrations of antimicrobials is not correlated to induction of lytic phage. J Antimicrob Chemother. 2017 Jun 1;72(6):1624–31.
87. Enault F, Briet A, Bouteille L, Roux S, Sullivan MB, Petit M-AA. Phages rarely encode antibiotic resistance genes: a cautionary tale for virome analyses. ISME J. 2017;11(1):237–47.
88. Maniloff, J. Bacteriophages. In: eLS. Chichester, UK: John Wiley & Sons, Ltd; 2012. DOI: 10.1002/9780470015902.a0000773.pub3
89. Aminov RI, Mackie RI. Evolution and ecology of antibiotic resistance genes. FEMS Microbiol Lett. 2007 Jun 1;271(2):147–61.
90. Paterson DL. Resistance in gram-negative bacteria: enterobacteriaceae. Am J Med. 2006 Jun 1;119(6 Suppl 1): S20-8; discussion S62-70.
91. Aminov R. History of antimicrobial drug discovery: Major classes and health impact. Biochem Pharmacol. 2017 Jun 1;133:4–19.
92. Jennings MC, Minbiole KPC, Wuest WM. Quaternary Ammonium Compounds: An Antimicrobial Mainstay and Platform for Innovation to Address Bacterial Resistance. ACS Infect Dis. 2015 Jul 10;1(7):288–303.
93. Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. Nucleic Acids Res. 2017 Jan 4;45(D1):D566–73.
94. The Comprehensive Antibiotic Resistance Database [Internet]. [cited 2019 Jun 27]. Available from: https://card.mcmaster.ca/ontology/36007
95. Conly J, Johnston B. Where are all the new antibiotics? The new antibiotic paradox. Can J Infect Dis Med Microbiol = J Can des Mal Infect la Microbiol medicale. 2005 May;16(3):159–60.
96. Spellberg B, Powers JH, Brass EP, Miller LG, Edwards JE. Trends in Antimicrobial Drug Development: Implications for the Future. Clin Infect Dis. 2004 May 1;38(9):1279–86.
97. Andrei S, Valeanu L, Chirvasuta R, Stefan M-G. New FDA approved antibacterial drugs: 2015-2017. Discoveries. 2018 Apr 4;6(1):e81.
98. Norrby SR, Nord CE, Finch R. Lack of development of new antimicrobial drugs: a potential serious threat to public health. Lancet Infect Dis. 2005 Feb 1;5(2):115–9.

ANNEXES

Annexe 1. ResFinder results

ResFinder is a computer program available online in https://cge.cbs.dtu.dk//services/ResFinder/. This program intends in the identification of acquired antimicrobial resistance genes.

It was submitted a FASTA file (size 172.74 MB) with all 2050 complete genomes of the order *Caudovirales* available in NCBI (https://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=10239&sort=taxonomy) to be run against the antimicrobial resistance genes inserted in ResFinder database (2019-02-20).

Results found zero results to 21 genomes of the *Ackermannviridae* family, 355 of the *Podovirirdae* family and 7 of unclassified *Caudovirales* genomes. Found 1 result to the 542 of *Myroviridae* family corresponding to the $bla_{CTX-M-174}$ accession no. KT997886 (100% identity to 876/852 query/HSP). Found 2 result to the 1125 of *Siphoviridae* family corresponding to the aac(6')-aph(2'') gene (aminoglycoside) (99.9% identity to 1440/1440 accession no. M13771) and to fosB2 (99.73% to 420/366 accession no. AE016879). All analyses had a selected %ID threshold of 90 % and a selected minimum length of 60 %.