Universidade de Aveiro Departamento de Biologia
Ano 2020

ANA RITA PEREIRA GOUVEIA

TÉCNICAS AVANÇADAS DE BIOLOGIA MOLECULAR E TRANSCRITÓMICA APLICADAS A ACTINOMYCETES
ADVANCED TECHNIQUES IN MOLECULAR BIOLOGY AND TRANSCRIPTOMICS APPLIED TO ACTINOMYCETES

Universidade de Aveiro Departamento de Biologia
Ano 2020

ANA RITA PEREIRA GOUVEIA

TÉCNICAS AVANÇADAS DE BIOLOGIA MOLECULAR E TRANSCRITÓMICA APLICADAS A ACTINOMYCETES
ADVANCED TECHNIQUES IN MOLECULAR BIOLOGY AND TRANSCRIPTOMICS APPLIED TO ACTINOMYCETES

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular, realizada sob a orientação do Doutor Antonio Rodríguez García, Investigador do Instituto de Biotecnologia de León e sob a coorientação científica da Doutora Sónia Mendo, Professora Auxiliar com Agregação do Departamento de Biologia da Universidade de Aveiro

Dedico este trabalho aos meus pais e ao seu eterno apoio.

o júri

presidente

Prof. Doutor António José de Brito Fonseca Mendes Calado
professor auxiliar da Universidade de Aveiro

Prof. Doutora Ana Sofia Direito dos Santos Duarte
professora auxiliar da Universidade Católica Portuguesa

Prof. Doutora Sónia Alexandra Leite Velho Mendo Barroso
professora auxiliar com agregação da Universidade de Aveiro

palavras-chave

sRNAs, *Streptomyces*, metabolitos secundários, PhoR-PhoP.

resumo

Hoje em dia, a resistência a antibióticos está a tornar-se numa das maiores preocupações do século. Esta problemática está a arriscar a nossa saúde pública e pode vir a tornar-se, no futuro, num dos problemas de saúde mais letais. Portanto, uma solução é necessária.

Streptomyces é conhecido como o género que é responsável por imensos antibióticos. Porém, nestes últimos anos, o número de antibióticos que chegaram a uso clínico diminuiu, devido a várias razões.

Metabolismo secundário é composto de vários processos que, ainda que não sejam fundamentais para a sobrevivência da célula, dão-lhe imensas vantagens. A maior parte destas vantagens originam-se a partir dos metabolitos secundários, alguns deles conhecidos como antibióticos.

Por outro lado, fósforo é um dos elementos mais importantes para qualquer organismo e, por isso, necessita de um mecanismo que se assegure que ele está sempre regulado. Um destes sistemas depende de PhoR-PhoP.

Com a descoberta de um novo tipo de RNAs, os sRNAs (que têm entre 50 a 400 nucleótidos), a investigação de novos compostos com importância farmacêutica e industrial pode continuar a avançar, dado que algumas destas moléculas podem funcionar como reguladoras do metabolismo secundário e, por isso, estar relacionadas com antibióticos.

O objetivo deste trabalho é identificar sRNAs que estão implicados na regulação de fosfato, a principal forma de assimilação de fósforo. A minha função era experienciar o que era trabalhar num laboratório, ao aprender algumas técnicas de biologia molecular e transcritómica, como introdução de DNA em células e extração de ácidos nucleicos, necessárias para este projeto.

keywords

sRNAs, *Streptomyces*, secondary metabolites, PhoR-PhoP.

abstract

Nowadays, the resistance to antibiotics is turning to be one of the biggest concerns of this century. It is risking our public health and might be, in the future, one of the deadliest health issues. Therefore, a solution is needed. *Streptomyces* is known as a genus that is responsible for many antibiotics. However, in the latest years, the number of antibiotics that reached the clinical use diminished, due to various reasons.

The secondary metabolism is composed by a number of processes that, even though aren't of extreme importance to the cell survival, can give it several advantages. Most of those advantages come as secondary metabolites, and some are known as antibiotics.

On the other hand, phosphorus is one of the most important elements to any organism, and, therefore, a mechanism is needed to make sure that this element is regulated. One of the systems that does that depends on PhoR-PhoP.

With the discovery of a new type of RNAs, the sRNAs (which have between 50 and 400 nucleotides) the investigation of new compounds with pharmaceutical and industrial importance may continue to go forward, since some of those molecules may function as regulators of the secondary metabolism and, therefore, be related to antibiotics.

The aim of this work is to identify sRNAs that are implicated in the phosphate regulation, which is the main assimilation form of phosphorus. My objective in this study was to experience what was like to work in a laboratory, by learning some molecular biology and transcriptomic techniques, such as DNA introduction into cells and nucleic acid extraction, needed for this study.

palabras clave sRNAs, *Streptomyces*, metabolitos secundarios, PhoR-PhoP.

resumen La resistencia a los antibióticos es uno de los grandes problemas de este siglo. Pone en riesgo la salud pública y se prevé que en el futuro sea una de las mayores causas de mortalidad, por lo que se hace necesaria una solución.

El género *Streptomyces* es conocido por ser capaz de producir muchos tipos de antibióticos diferentes. Sin embargo, en los últimos años, el número de antibióticos que llegaran a la práctica clínica ha disminuido, debido a varias razones.

El metabolismo secundario está formado por muchos procesos que, aunque no sean totalmente necesarios a la supervivencia de la célula, le confiere ventajas. La mayoría de esas ventajas son los metabolitos secundarios, algunos conocidos como antibióticos.

Por otro lado, el fósforo es uno de los elementos más importantes para cualquier organismo y por eso necesita tener uno mecanismo que regule su metabolismo. Uno de esos mecanismos depende de PhoR-PhoP.

Con el descubrimiento de un nuevo tipo de ARN, los ARNp (que tienen entre 50 y 400 nucleótidos), la investigación de nuevos compuestos con importancia farmacéutica y industrial puede seguir avanzando, ya que algunas de estas moléculas pueden funcionar como reguladoras de metabolismo secundario y, por eso, estar relacionadas con los antibióticos.

El objetivo de este trabajo es identificar ARNp que están implicados en la regulación por fosfato, la principal forma de asimilación de fósforo. Mis objetivos en este trabajo han sido experimentar lo que es trabajar en laboratorio y aprender algunas de las técnicas transcriptómicas y de biología molecular, como la introducción de ADN en células y extracción de nucleótidos, necesarias para este estudio.
.

Index

List of figures

Introduction

Materials and Methods

Results

List of tables

List of abbreviations

Abbreviation	Meaning
5'PPP	5' triphosphate
ADB	Agarose Dissolving Buffer
BGC	Biosynthetic Gene Cluster
bp	base pair
CCC	Covalently Closed Circular
CIA	Chloroform-Isoamyl Alcohol
CSR	Cluster Situated Regulator
DNA	Deoxyribonucleic Acid
DNApol	DNA polymerase
DNAse	Deoxyribonuclease
dNTP	Deoxyribonucleotide Triphosphate
E-test	Epsilometer test
EDTA	Ethylenediaminetetraacetic Acid
GTE	Glucose-Tris-EDTA
ID	Identification
IT	Transcription Start Site
kb	kilobase
LA	Luria Agar medium
LB	Luria Broth medium
Mb	Megabase
min	minute
ms	millisecond
Miniprep	Minipreparation
mRNA	messenger RNA
nt	nucleotide

OC	Open Circular
OD_{600}	Optical Density at a wavelength of 600 nm
PCR	Polymerase Chain Reaction
pH	potential for Hydrogen
RBS	Ribosome-Binding site
RCF	Relative Centrifugal Force
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
RNase	Ribonuclease
rpm	revolutions per minute
s	second
SARP	*Streptomyces* Antibiotic Regulatory Protein
SDS	Sodium Dodecyl Sulfate
SOB	Super Optimal Broth medium
sRNA	small RNA
TAE	Tris-Acetate-EDTA medium
TB	Terrific Broth medium
Ter	Terminator
TE	Tris-EDTA medium
TPM	Transcripts Per Million
tra	transfer
TSA	Tryptic Soy Agar medium
TSB	Tryptic Soy Broth medium
Tsr	Thiostrepton
U	Unit
UTR	Untranslated Region
WT	Wild-Type

INTRODUCTION

Introduction

1.1. *Streptomyces*

1.1.1. *General characteristics*

Actinomycetes are a heterogeneous group of bacteria that have high guanine and cytosine content (more than 50 %) in their deoxyribonucleic acid (DNA) and are Gram-positive [1], [2]. The high content of guanine and cytosine might come from an adaptative response, since it enables them to fight certain bacteriophages using rare codons [3]. They have at least 350 genera, which makes them one of the biggest bacterial phyla [1]. From this group, *Streptomyces* is the most prevailing genus [4].

These bacteria are classified as both mesophilic and neutrophilic, since their optimal temperature is around 25-35 ºC, and their optimal potential for hydrogen (pH) is around 6.5-8 (but they can live in pHs higher than 9) [5], [6]. They are also considered strict aerobic, but *Streptomyces coelicolor* (*S. coelicolor)* has enzymes that allow it to breathe nitrate (this might also be an adaptative response to living in a place as variable as the soil) [7].

They live normally in the soil (since they need nutrients that come from plants' degradation), but they can also be found in water, air, and other places (including extreme environments as glaciers and deserts) [8], [9]. This happens because they are able to grow in different sources of carbon and nitrogen [10]. In order to be able to adapt to the many changes that the soil suffers, *Streptomyces* has a lot of sigma factors [11].

Terrestrial *Streptomyces* has been over-exploited by various companies and, therefore, its investigation has now turned to its relatives that can be found in those other environments, to those that invade plant tissues and even to non-*Streptomyces* actinomycetes [4].

1.1.2. *Life cycle*

In this section, the complex life cycle of *Streptomyces* cultivated in solid medium will be explained (since, in liquid, it is not completed, since *Streptomyces* is not able to sporulate in this type of medium) [12]. The first stage, spores' sporulation, is characterized by presenting hydrophobic pigmented cells. When the optimal conditions are achieved, germ tubes are originated, that create branched hyphae (which are polynuclear and capable of getting nutrients by penetrating the substrate) [13], [14]. When the hyphae turn into a cell aggregation, a first mycelium is created, which has the proteins responsible for primary metabolism or vegetative mycelium **(Fig. 1)** [15].

The second or reproductive mycelium, or the precursor of aerial mycelium in solid media, is originated after some central cells of the primary mycelium go through apoptosis **(Fig. 1)**. The second mycelium might be characterized as early or late, depending on the hydrophobic covers typical of aerial hyphae; if it has the covers, it is a late secondary mycelium. The secondary mycelium is crucial since it has the proteins needed for secondary metabolism [15].

After this stage, another apoptosis happens in order to create spores, that will be able to reinitiate the cycle [12].

Even though this explanation only refers secondary metabolism in the secondary mycelium, it is also possible for it to happen, and secondary metabolites being originated, in the stage of primary mycelium [12].
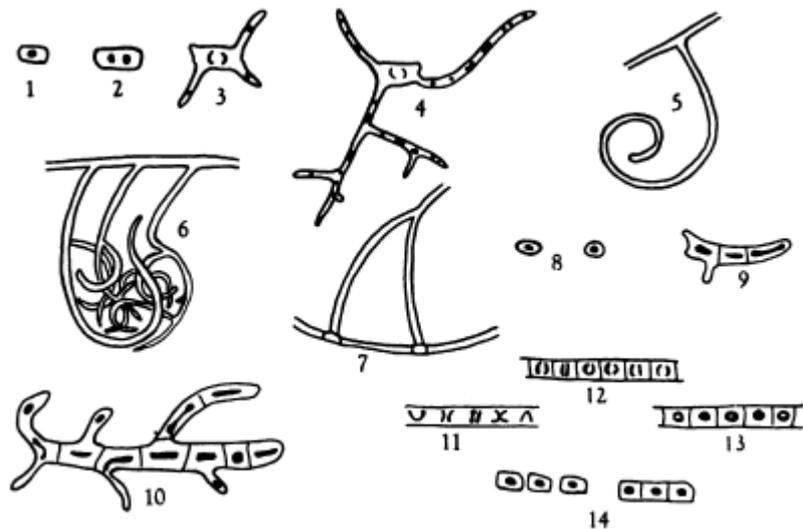
Fig. 1. The life cycle of *Streptomyces*. **(1)** represents a spore, which germinates **(2, 3)** and produces the primary mycelium **(4)**. **(5, 6)** have hyphae, that in **(7)** starts producing initial cells **(8)**. Then, these cells germinate into the secondary mycelium **(9, 10)**. In that mycelium, there are pairs of chromosomes **(11, 12)** able to develop into spores **(13, 14)**. Taken from [16].

1.1.3. *Genetic aspects*

Streptomyces has its genome organized in a large linear chromosome [with more than 7 megabase (Mb)], which doesn't happen in the majority of the actinobacteria [17]. This linearity has been explained by a recombination between a linear plasmid and an ancestor with a circular chromosome, since, when this happens, the result is always a linear molecule (**Fig. 2**) [18]. This chromosome would have its core originated by the ancestor chromosome and its arms by the linear plasmid [19].

In the central core of this chromosome, genes related to primary and central metabolism and, therefore, essential and that are conserved in a lot of species, are present [20]. On the other hand, the genes considered not essential, such as the ones related with secondary metabolism, are located in the arms of the chromosome [21].

There are also elements outside of the chromosome, such as linear and circular plasmids, that are able to replicate in other hosts with variable size and number of copies [22]. They also have a high rate of mutations, most of them negative or neutral, that occur in the terminal regions. These mutations can be explained by the need that these bacteria have to adapt to the different environments [18].
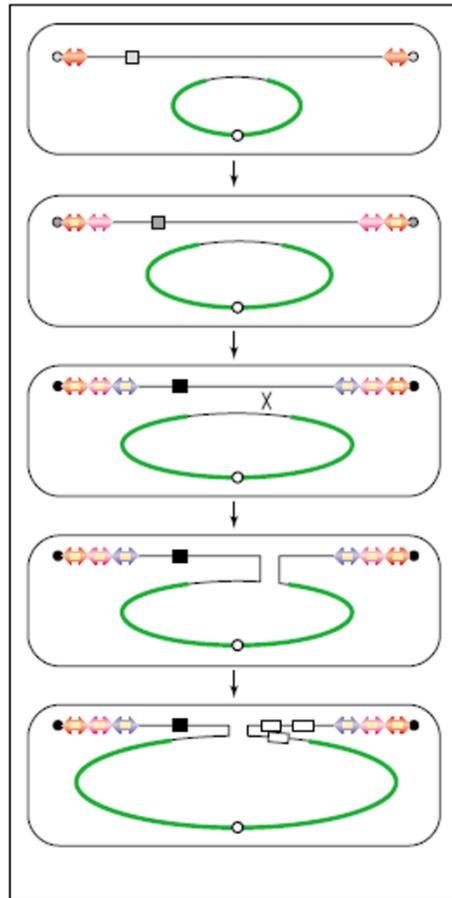
Fig. 2. A model, in *Streptomyces*, for the evolution of the linear chromosome. Here, it can be seen that, during the evolution, circular chromosomes (in green) opened up through recombination with the linear plasmid. Taken from [19].

1.1.4. *Streptomyces coelicolor*

Streptomyces coelicolor is the model organism of the actinomycetes, because it is easy to manipulate, and it also was the first species of *Streptomyces* to have its genome sequenced. Therefore, a lot of known genetic tools that can be used in this species have been already studied [16], [23]. Moreover, it has more codifying sequences than most eukaryotes, which only shows how much it has to adapt. It also has three plasmids: SCP1 (linear and produces methylenomycin); SCP2 (circular and smaller); SLP1 (circular and able to integrate in the chromosome, which enables it to replicate in other species of the genus) [24].

One reason why this species facilitates its studies, compared to other species of *Streptomyces*, is due to the production of pigmented antibiotics, which are easy to see and, therefore, easy to confirm if the bacteria are producing metabolites [25]. If not for this, it would be harder to conclude the same thing. *S. coelicolor* is able to produce various antibiotics, for instance, actinorhodin, produced by genic cluster *act*, which only acts on Gram-positive bacteria and is an acid-base indicator, since in acidic/neutral conditions it has the colour red and in alkaline conditions it is blue [26]. On the other hand, undecylprodigiosin, another antibiotic produced by this species, is red and is produced by the genic cluster *red* [27]. Both of these are detected when there is no phosphate in the medium [28].

There are also antibiotics that depend on calcium levels present in the medium, such as the so-called calcium dependent antibiotic produced by the genic cluster *cda* [29]. There are other antibiotics such as methylenomycin, originated from the genic cluster *mmr* (which acts

specially in genus *Proteus* in acid conditions), and coelimycin P1, originated from *cpk*, able to affect *Bacillus subtilis*, *Escherichia coli* (*E. coli*) and *Kocuria rhizhophila*, and has the colour yellow (**Fig. 3**) [30], [31].
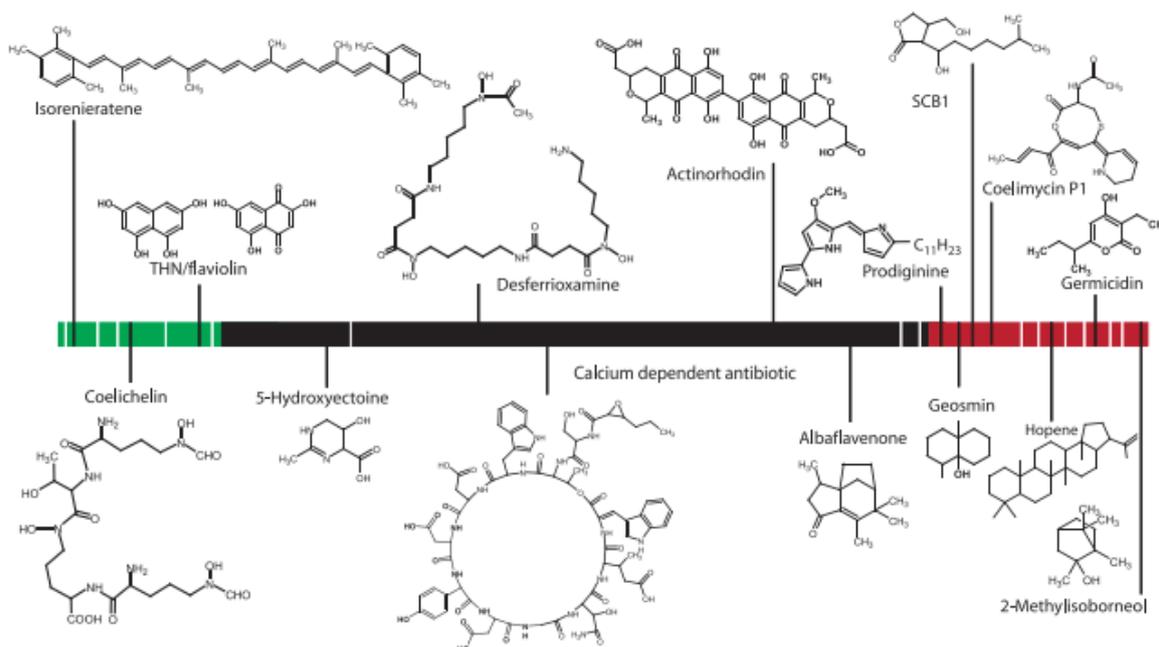
Fig. 3. *S. coelicolor* secondary metabolites in relation to their chromosomal location. Most of the biosynthetic genes are located outside the highly conserved core region (in black). The green and red represent, respectively, the left and right arm. Taken from [25].

1.1.5. *Importance of the* Streptomyces *genus*

Actinobacteria are one of the most notable group of microorganisms, since they alone, and specially the *Streptomyces* genus, represent a tremendous source of valuable chemicals. This genus is responsible for around two-thirds of all naturally derived antibiotics used in various areas, from medicine to agriculture [1], [32]. Until 2005, rare actinomycetes were responsible for the discovery of 2250 new bioactive secondary metabolites, approximately [4].

Although antibiotics are of extreme importance, they are not the only bioactive molecules that the genus *Streptomyces* has provided, but also antiprotozoals, antifungals and antivirals [4].

All of this is only a small percentage of the value of this genus. Until now, only the medical importance has been described. But these bacteria also have a huge industrial value, because their genome is big (which might indicate that more biosynthetic gene clusters, BGCs, can be present), and an ecological value as well, since they are important heterotrophics, which are able to degrade plant biomass (an important carbon source in the terrestrial environment) [33]. None of these, even though the medical importance will be the one discussed the most in this work, should be underestimated.

1.1.6. *Difficulties when dealing with the* Streptomyces *genus*

Natural product research, including the one that aims to discover new drugs that come from *Streptomyces*' secondary metabolites, has been decreasing for various reasons, causing less products to reach the marketplace and have a clinical use. Low production concentration and a challenging isolation of the bacteria, which makes a large-scale cultivation very difficult to achieve, are some of the causes [1], [4].

Everyone thought that this genus had nothing more to offer than these bioactive secondary metabolites that are already known. However, it was discovered that *Streptomyces* might have a plethora of secondary metabolite encoded in the genome that weren't found until now, probably because their products might not be able to be detected by widely used analytical methods or because these genes were not expressed under conventional laboratory conditions (in a laboratory, it is difficult to mimic some aspects of the environment). Therefore, BGCs, contiguous genes that assemble the secondary metabolites, that have remained inhibited under standard cultivation conditions, might be a potential source of new scaffolds to create new antimicrobials [34].

But the difficulties of dealing with *Streptomyces* don't stop there. The mandatory phases to get an antibiotic to clinical use, from preclinical testing to approval for human use, don't facilitate, at all, the process, since it takes around 10-15 years for it to reach the market. This doesn't happen only in the investigation that uses this genus, but with all. There are also economic and scientific factors that keep delaying the appearance of new antibiotics. In general, 1 in each 1000 potential drugs proceed to clinical trials and, of those, around 90 % will fail in the human testing phase [1].

1.2. *Secondary metabolism*

Metabolism is defined as the set of reactions or processes needed by an organism to maintain life.

The primary metabolism consists of the reactions that aim to produce energy. These normally happen during the exponential phase of growth of the cell (**Fig. 4**) [35]. Without the metabolites that come from this process, there would be no life.

On the other hand, the secondary metabolism is responsible for the production of substances, known as secondary metabolites, that are not crucial for the organism survival, but give it several advantages, and it normally happens during the stationary growth phase (**Fig. 4**) [12], [36]. One of the most important secondary metabolites, and the most crucial one for this work, are antibiotics, which is originated from a microorganism and acts in low doses.

The synthesis of secondary metabolites is related with the morphological differentiation, since, for instance, the activation of secondary metabolism happens when the secondary mycelium is developed [15]. Both processes depend on extracellular signals and environmental changes [37].
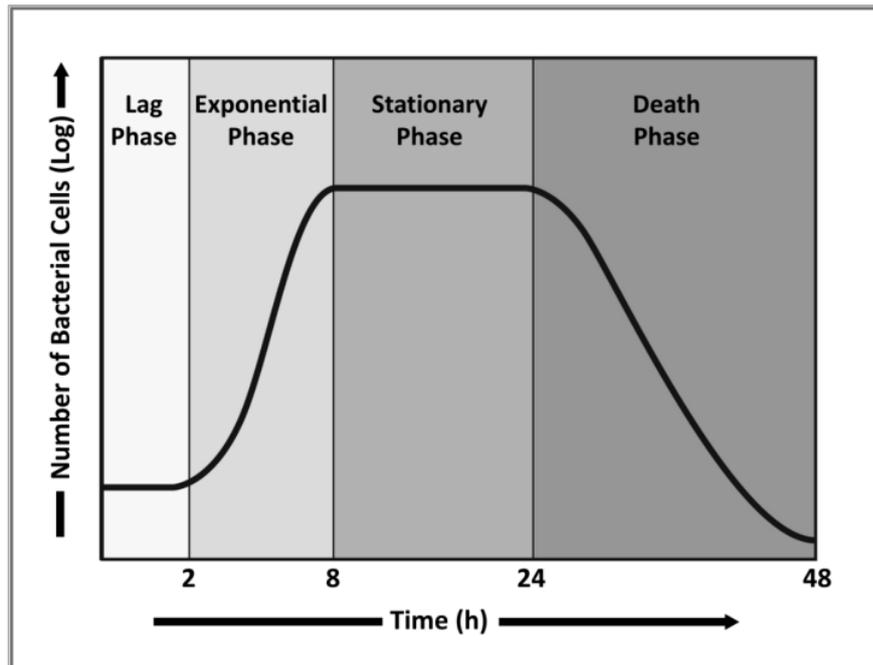
Fig. 4. The various phases of the bacterial growth curve. Taken from [38].

In order to survive, these bacteria, as all living organisms, need to supress their pathogens. In the case of *Streptomyces*, this can be done by producing the antibiotics, antiprotozoals, antifungals and antivirals referred earlier [4]. These natural inhibitors come from secondary metabolites, which are assembled in adjacent chromosomal genes, located in terminal zones of the chromosome (**Fig. 5**) [32], [39]. These contiguous genes are organized in clusters called BGCs, which are able to, individually, codify 35 secondary metabolites [32]. Normally, these clusters are constituted by resistance, transport, structural and regulator genes [40]. These last normally act by enhancing transcription of the structural genes that are in the same cluster as them, which explains why they are called cluster-situated regulators (CSRs) [41].

Most of the BGCs are responsible for the origin of everything (enzymes, regulatory proteins and transporters) needed to generate, process and transport a metabolite, which includes genes with regulatory functions, which are called CSRs (**Fig. 5**) [32], [39]. One BGC has normally one or more CSRs, and the most common one, among *Streptomyces*, is *Streptomyces* antibiotic regulatory protein (SARP) [39].

Outside of these BGCs, there are pleiotropic regulators responsible for controlling, for example, the morphological development [39].

The secondary metabolism needs, as all processes, to be regulated. That regulation is, mainly, done by transcription control. How this control works has already been theorized in two different ways. The first one, the hierarchical theory, says that there are different layers of regulation: an inferior layer, where CSRs control their own metabolites in their own biosynthetic cluster, and a superior one where the CSRs and one or more secondary metabolites are regulated by pleiotropic or global regulators (localized outside of the cluster) [42]. These pleiotropic regulators are capable of also affecting the morphological differentiation, the capture of nutrients, the secondary metabolism, and other [43]. CSRs, pleiotropic and central metabolic genes are regulated by global regulators spread throughout the chromosome. Every single one of them might regulate the biosynthesis of antibiotic [39].

The second theory says that, more than a hierarchy, the regulation of the secondary metabolism is a web and that the different regulators are all connected (**Fig. 5**) [44].
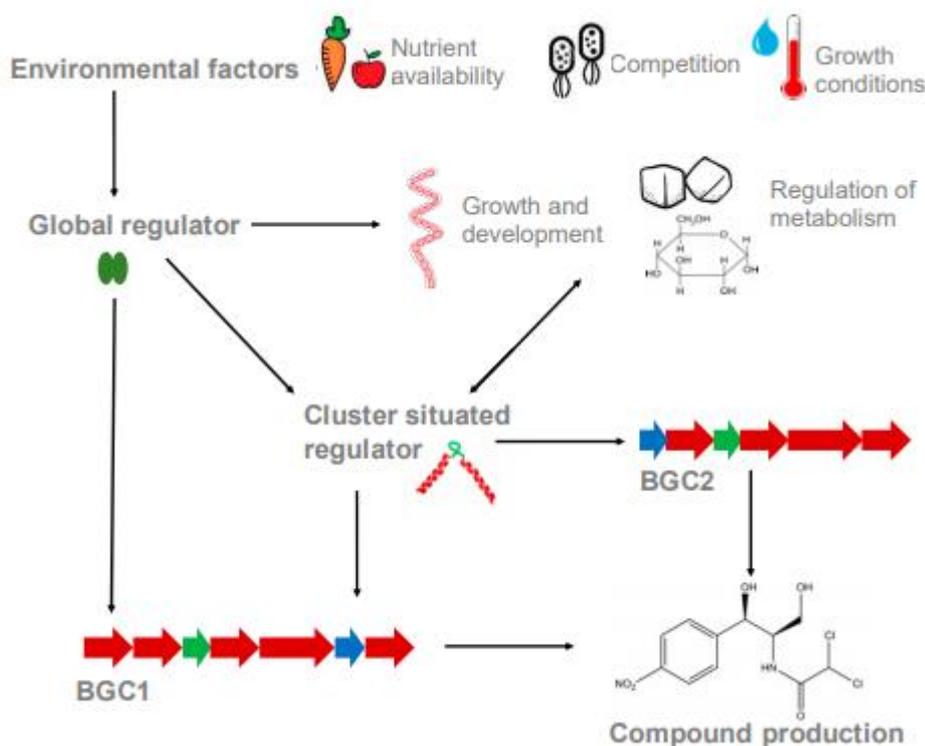
Fig. 5. Cluster situated regulators (CSRs) and global regulatory systems. These last are influenced by external environmental factors and both control the regulation of the expression of the genes related to secondary metabolism. Taken from [45].

1.3. *Two-component system PhoR-PhoP*

Phosphorus is important and it's present in DNA, ribonucleic acid (RNA), in compounds needed for central metabolism (like ATP), signalling and in various other molecules. Moreover, it is involved in post translational regulation and phosphorylation/dephosphorylation of proteins [46], [47].

Therefore, an efficient supply of phosphorus is needed for the cell to survive. However, there are not, for *Streptomyces*, any systems known to deliver organic phosphate in the interior of the cell. So, when there is not enough phosphate, phosphate transporters and enzymes are activated to obtain it. This element is maintained by two strategies: intracellular storing and, under phosphor limiting conditions, scavenging it with the help of enzymes [46].

Phosphorus can be assimilated by bacteria in its inorganic form, phosphate, by, normally, two systems. An exception is, for instance, *S. coelicolor* that has three systems: *pitH1* (expressed when there is a lot of phosphate), *pitH2* and *pstSCAB*, which stands for phosphate specific transport and is the main phosphate transport system in conjugation with *pitH2*, and both are expressed when there is not much phosphate. These last two optimize the transport of phosphate and are regulated by the system PhoR-PhoP [48].

The system PhoR-PhoP (known as PhoR-PhoB in *E. coli*) is a two-component signal transduction system and the main regulator in phosphate metabolism control (mainly in cases

of limitation of phosphate). It is also responsible for the signal transduction from the outside to the inside of the cell, which enables adaptation of bacteria to environmental changes [49].

In this system, PhoR acts as a sensor kinase that receives sensory input, since it has a large extramembrane domain, and has the characteristics of a transmembrane sensor protein needed to adapt the cell to the phosphate limited state [50].

The other member of this system, PhoP, is the response regulator, and as so, has a regulator domain and an effector domain (this last one is responsible for the bound with DNA, making PhoP also a transcription factor) [51]. It is also responsible for the primary and secondary metabolism and morphological development, since it activates/inactivates genes implicated in phosphate metabolism and nitrogen assimilation (and therefore, in the balance between phosphate/nitrogen) [52].

This member, when phosphorylated, binds to DNA sequences known as Pho boxes, which are located in the promotor region of PhoBR regulon genes and is the consensus sequence that is shared by the regulatory areas of the genes in the phosphate (Pho) regulon, which is controlled by this two component system [46], [53]. The Pho regulon is important since it is involved in the response to conditions of limitation of phosphate [54]. These boxes, in these conditions, are activated by either phosphate transporter *pstSCAB* or *phoU* deletion and answer by increasing expression of phosphate transporters (such as *pstSCAB)*, of enzymes needed to scavenge phosphate or by amplifying responses through positive autoregulation of the *phoPR* operons [46].

The transporter *pstSCAB* is also important since its structure determines the autokinase and phosphatase modes of PhoR activity: when it is outward facing closed structure, it promotes autokinase activity (needed in conditions of limited phosphate); and when *pstSCAB* is in an inward facing open structure, the phosphatase activity is promoted (essential for condition of sufficient phosphate) (**Fig. 6**) [46].

In *E. coli*, this system was already studied in extreme conditions, which created the Wanner's Model [55]. In this model, it is explained that, in conditions of excess of phosphate, PstS (protein that unites with phosphate and its promotor is regulated by phosphate dependent of PhoB, PhoP alike) creates an inhibitor complex with transport system p*stSCAB*, protein PhoU and PhoR. Here, PhoR is inhibited and shows phosphatase activity towards PhoB, making sure that the last one is not activated. When phosphate is limited, p*stSCAB* changes its conformation and the inhibition complex is freed [46]. Then, PhoR self-phosphorylates, by transferring its phosphate to PhoB. This response regulator binds to Pho boxes, activating about 30 phosphate-regulated genes (**Fig. 6**) [28].
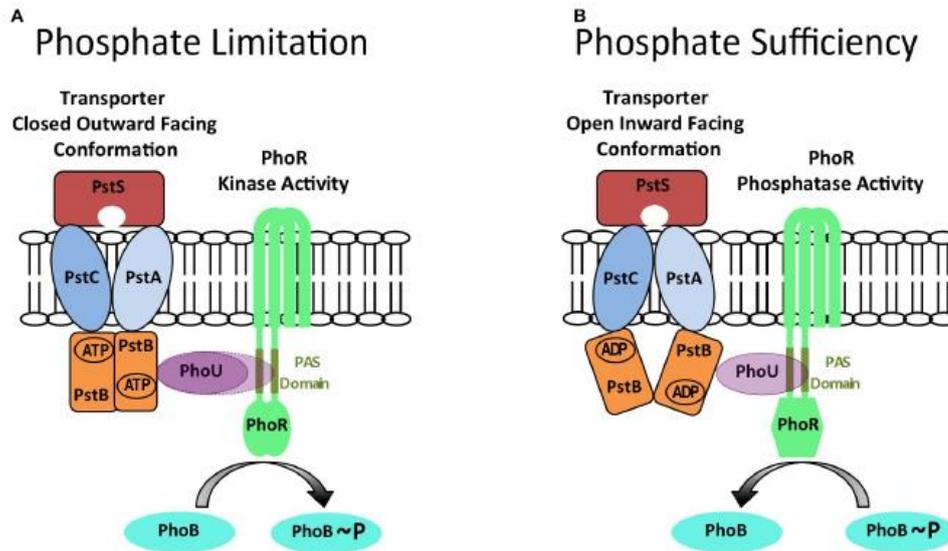
Fig. 6. A model for controlling, in *E. coli*, the balance between PhoR phosphatase and autokinase activities. The autokinase mode exists during conditions of phosphate limitation **(A)**, while the phosphatase mode happens during phosphate sufficiency **(B)**. Those alternate states are proposed to be determined by the different conformations of the *pstSCAB* transporter, which are relayed by PhoU to PhoR, in order to determine its activity. For instance, when *pstSCAB* is in a closed outward facing conformation **(A)**, PhoU adaptor protein is unable to interact with PhoR and the autokinase activity is promoted. On the other hand, in **(B)** the *pstSCAB* is in an open inward facing conformation, which enables the contact between the PhoU adaptor and the PAS domain of PhoR, promoting the phosphatase activity. This way, the PhoB response regulator is dephosphorylated. Taken from [46]**.**

In *Streptomyces*, the proteins from Wanner's model have been described and it is known that, in these bacteria, for the PhoR-PhoP system to work, PhoP needs to be recognized and bounded to DNA. To do that, as explained before, there are DNA sequences, called the Pho boxes, that are described, for *Streptomyces*, as two direct repeat units (DRus) of 11 nucleotides (in which the first seven, "GTTCACC" for *S. coelicolor*, are the most conserved ones and have consensus sequence) (**Fig. 7**). Moreover, the strength of this bond depends on the conservation of the sequence (the more conserved it is, the stronger the bond is) [50].

When the genus *Streptomyces* doesn't have enough phosphate, it can suffer effects on a number of processes, for instance, its growth, its morphological differentiation and its production of secondary metabolites [28], [50]. These metabolites are regulated negatively by the phosphate concentration in the medium, which explains why they are, normally, only produced in limiting concentrations of this substance [28]. This might happen due to the fact that, if there is less phosphate, the organism is probably in conditions where it is difficult to grow (it is stressed), and these secondary metabolisms might be an advantage to, for instance, inhibit the competition, or act as biochemical cross talk signals. Therefore, it is a strategy to survive [28]. All of these are reasons why the PhoR-PhoP system is crucial and why it needs to be investigated.
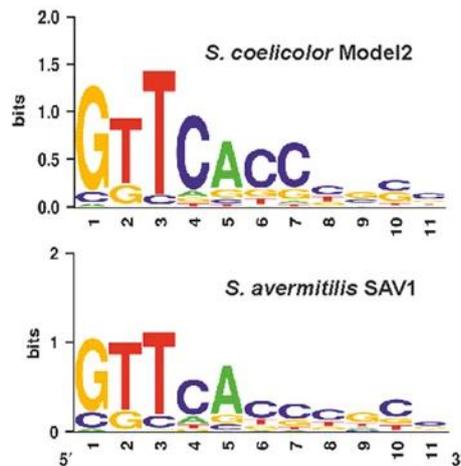
Fig. 7. The DRus in the Pho boxes of *S. coelicolor* and *S. avermitilis*. The height of the letter represents the frequency of that nucleotide in each position. Taken from [50].

1.4. *Small RNAs*

Small RNAs (sRNAs) are small-sized RNA molecules, of length between 50 nucleotides (nt) and 400 nt in prokaryotes (in eukaryotes they are even smaller), that, normally, don't encode proteins and have regulator properties capable of modulating genetic expression by means of interaction with messenger RNA (mRNA) or, less frequently, with proteins [56], [57]. Most of sRNAs are transcriptionally induced under specific conditions, such as cold and heat shock, iron homeostasis, membrane remodelling and sugar and nitrogen metabolism, between others [58].

To the cell, creating regulatory sRNAs instead of proteins is beneficial because sRNAs are synthetized the fastest, since they are smaller and they don't need to be translated [57].

Every single sRNA can have various action mechanisms, which influence every step of a gene expression: since the structure of DNA until the translation [57]. They might inhibit transcription, if they bind with mRNA that's being synthesized [59]. They also act in the translation, by inhibiting it, if: they unite between positions -20 and -15 of the mRNA; or if they target 5'-untranslated regions (UTR) close or in the ribosome-binding site (RBS), thus competing with the 30S ribosomes [59], [60], [61]. On the other hand, they might enhance it, if they bind in the untranslated extremes of the mRNA and leave the mRNA bond site freer for the ribosome. They can also do it by stabilizing the mRNA or even by binding with translational enhancers [62]. Finally, sRNA are also responsible for the posttranscriptional regulation of physiology, stress responses, metabolism and virulence [63].

Even though transcription factors might have similarities with sRNAs, there are crucial differences: transcription factors cannot keep a noise-free silent state and have less recognition sites [63].

sRNAs can be characterized depending on their position in relation to their targets. The antisense sRNAs, or cis-acting sRNAs, are responsible for the regulation of the gene in the opposite chain and are completely complementary to the target gene of the mRNA in one of the extremes, in the middle, or with the complete gene [64].

There are different types of cis-acting sRNAs. One of them is called intragenic sRNAs, which are codified in the reading frame of a coding gene [65]. There are different types of intragenic sRNAs. They can be, for instance, 3' sRNAs, which are codified, as the name itself suggests, in the region non translated 3' of a mRNA. They can be classified as type I, if their promotor is inside the 3' UTR or the mRNA coding sequence, and type II, if they come from

the processing of their parental mRNA. Both types share the 3' end with their corresponding mRNA. On the other hand, their 5' end is different, since there is a 5' triphosphate (5´PPP) in type I and a 5' P in the type II (**Fig. 8**) [66].
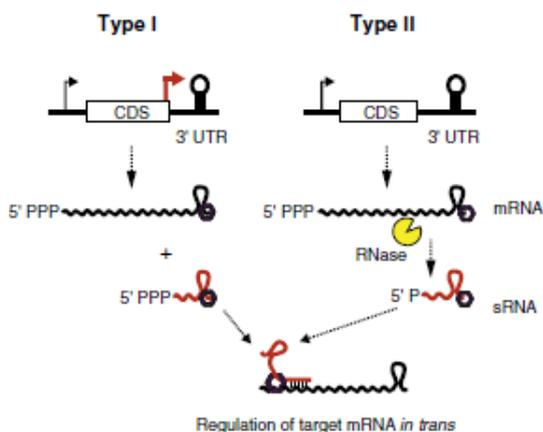
Fig. 8. Two general biogenesis pathways of sRNAs from the 3' region of mRNA loci. Type I, in which the sRNA is transcribed from an mRNA-internal promotor and type II, in which the sRNA is processed from its parental mRNA. Adapted from [66].

On the other hand, riboswitches, another type of cis-acting sRNAs, are located in the non-translated 5' region of the mRNA [63].

Finally, intergenic sRNAs, or trans sRNAs, are codified in a different chromosome than its target. Its complementarity, as opposed to the antisense sRNAs, is not perfect, which might explain why one intergenic sRNA has more than one target. In model Gram-negative bacteria, most of intergenic sRNAs need the chaperone Hfq to have a good regulatory activity, since it can unfold and stabilize RNAs and recruit ribonuclease (RNase) E [64]. Since they have multiple targets, their target gene regulation is determined by the ratio between the number of target RNA and sRNA transcription sites [63].

In Gram-positive bacteria, the sRNAs regulation has some differences, when compared with the regulation in Gram-negative bacteria. For instance, in the Gram-negative, the sRNAs cannot regulate translation and the Hfq chaperone is more important than it is in the Gram-positive bacteria [67].

1.5. *Aim*

The research group I was integrated in was working in a project that aimed to identify, by using transcriptomic techniques and manual annotation, the sRNAs from *S. coelicolor* that were implicated in the regulation by phosphate, dependently or independently of PhoP. These might also be characterized as regulators of the secondary metabolism and, therefore, might be related to the biosynthesis of antibiotics and other secondary metabolites, that show pharmacological activities and industrial interest.

The main objective of my internship with them was learning the routine of a laboratory. In order to do that, a wide variety of techniques, such as techniques of DNA introduction into cells, nucleic acid extraction and gel-analysis were learned. Tasks involving bioinformatics, crucial here for the annotation and evaluation of candidate genes for experimental confirmation, were also done.

MATERIALS AND METHODS

In this section, there are protocols, and their detailed steps, that were done and others that were only seen. Those last will be differentiated from the others with a "*" in front of the name of the technique.

Methods in Microbiology

1. Melting of solid media

Theoretical Foundation

Medium can either be liquid and be used for inoculating microorganisms, or solid, and be used to plate in petri dishes. Both have to be sterilized in an autoclave before being used to prevent unwanted contaminations.

Material

The material here presented is the one needed to make sure that the solid medium is melted and split into petri dishes in a sterilized manner, in order not to contaminate the media.

For situations when little medium is needed

- ➢ Bottles of medium;
- ➢ Microwave.

For situations when much medium is needed

- ➢ Bottles of medium;
- ➢ Pressure cooker;
- ➢ Water;
- ➢ Metal tube rack;

Used when plastic material is placed inside the pressure cooker.

- ➢ Electric heater.

Procedure

Normally, two methods can be used to melt media:

a) If there is not the need for much medium (less than 6 bottles with 100 mL of it), a microwave can be utilized. While using it, to prevent burns, gloves should be used.

1. Before the bottles were placed in the microwave, the lid should be opened slightly, in order to ensure that some of the pressure was taken from it.

2. The moment the medium started to boil, it should be taken from the microwave.

So that none of it leaves the bottle.

3. The lid should be opened a bit more and the bottle agitated.

This is done to liberate a bit more of pressure, that was accumulated thanks to the boiling medium.

4. This must be repeated until the medium was completely melted and there was nothing floating in it.

b) The other method is used to melt more than 6 bottles of medium and it involves a pressure cooker.

1. In this one, the bottles and water, enough to be as tall as the thickness of a finger, should be placed in the cooker.

If what is put in the pressure cooker is made of plastic, something, like a metal tube rack, should be placed under the plastic material.

2. After that, the pressure cooker was put on top of an electric heater, the maximum temperature is chosen, and the pressure cooker was turned on.

3. The pressure regulator at the top of the lid of the pressure cooker would rise with the pressure. The moment the pressure regulator was as high as it could be, the heat in the electric heater was decreased a bit for 5 minutes (min).

4. After that period of time, the cooker was taken from the heater, already turned off, and when the pressure regulator decreased completely, the lid could be open, and the bottles could be taken from the pressure cooker.

This pressure cooker is also used to sterilize medium.

The same as the medium, all material needed for the procedures should be sterile, in order to not contaminate the samples. From here on out, all the material mentioned is sterile.

2. *Inoculation and plating of microorganisms*

Theoretical Foundation

When studying microorganisms, their living conditions, such as nutrients and oxygen, need to be assured, in order to be able to learn more about them. Therefore, they are grown in media, which, theoretically, have all nutrients needed for them to survive and multiply.

Two microorganisms were studied a great deal in this work: *Streptomyces* and *E. coli*. Even though only those two were worked with, many differences were seen, particularly while plating and inoculating them.

Each microorganism has different needs. For instance, there are some that need oxygen and others to which that element is lethal. One difference between the necessities of the two mentioned microorganisms has to do with the fact that *Streptomyces,* since it is filamentous*,* needs baffled flasks to provide more agitation, while *E. coli* doesn't. That is why a brief study on the microorganism should be done before starting working with it.

Media can be differentiated between solid (which is split between different petri dishes) or liquid (that is maintained in either tubes or flasks). Even though the first one is called solid, in order to use it, it needs to be firstly in a liquid state, or else it can't be split into petri dishes. Normally, the way to differentiate them is in the name: if the name ends with an "A", it might

mean that it has agar and, therefore, it's solid; on the other hand, if it ends with an "B" it can be liquid (example, tryptic soy agar, TSA, and tryptic soy broth, TSB).

The methods of placing the microorganism in the medium have different names depending on the type of medium used. When liquid medium is used, it's called inoculation, while with solid it's called plating the microorganism.

There are also differences in the methods when different microorganisms are used, because they can have different growth rates. While plating *Streptomyces,* compared to *E. coli*, more cautions are needed, since it grows slower, which means that is easier to contaminate. Since *E. coli* grows fast, not many other microorganisms that might be in the petri dish or in the flask can keep up with it.

Since *Streptomyces* grows slower, the thickness of the medium in the petri dishes should be higher than the ones used for microorganisms with faster growth rates, because, in the dry oven, the media tend to dry when inside it for some days.

The texture of the colonies is also different. *E. coli*, for instance, has a so-called creamy texture, which means that the colonies easily stick to the toothpick, and touching them is enough to have sufficient amount of cells. On the other hand, *Streptomyces* needs a lot of pressure to be done in the colony, which easily breaks into pieces, in order to have some sample in the toothpick (**Fig. 9**).

Plating for both of them is also different. In *E. coli*, a single line done in the medium with the toothpick with bacteria is enough to, in a few hours, have fully grown colonies. While, in *Streptomyces*, after making sure that the toothpick or the wire loop has some cells, a square or a circle should be done in the medium, and it should be done over and over again, to ensure that some cells are transferred to the new medium (**Fig. 9**).

While inoculating, some differences are also seen when working with both bacteria. If the inoculation is also done with a toothpick, in *E. coli* the toothpick with the sample staying a few seconds in the medium is enough, while with *Streptomyces* it needs to stay for more time and the toothpick should also be moved in order to help the colony separate itself from the toothpick.

Since *Streptomyces* has many morphologies, there are also differences when some of those are used. For instance, mycelium is harder to take a representative volume with the pipette and its absorbance value may not represent the reality. On the other hand, using spores also present some issues, since they develop much slower.
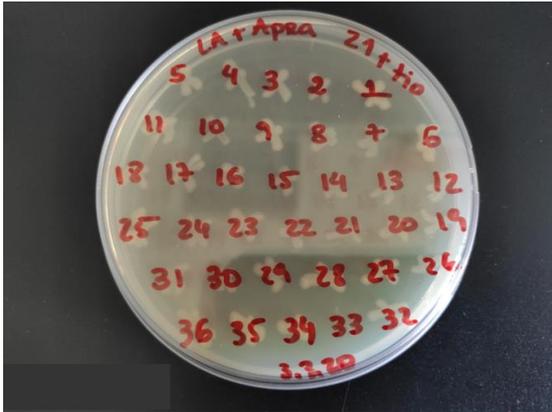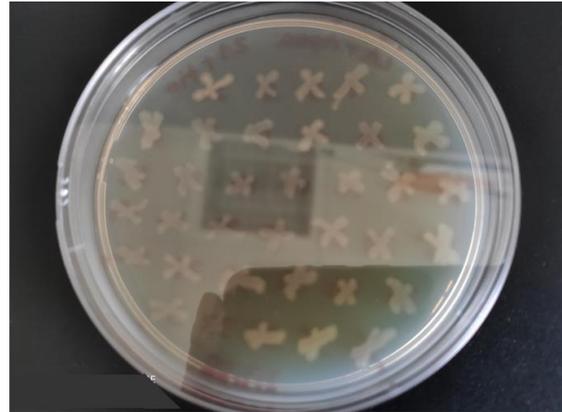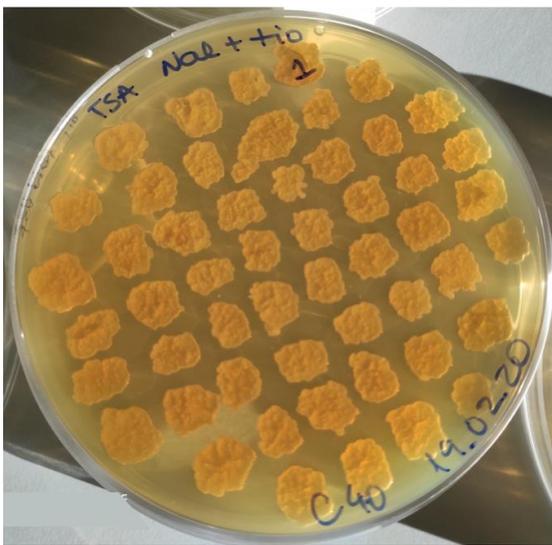
Fig. 9. Petri dishes with *E. coli* (**A** and **B**) and *Streptomyces* (**C** and **D**).

Material

<u>For inoculation</u>

➢ Flow hood;

➢ Samples of the microorganism;

➢ Flask or tube;

➢ Incubator shaker;

➢ Micropipette and tips;

➢ Medium for the microorganism;

➢ Wire loop or toothpicks.

<u>For plating</u>

- ➤ Samples of the microorganism;

- ➤ Petri dishes;

- ➤ Dry oven;

- ➤ Micropipette and tips;

- ➤ Flow hood;

- ➤ Medium for the microorganism;

- ➤ Wire loop, Digralsky spreader or toothpicks.

Procedure

These procedures might be done in a flow hood, depending on the microorganism used and its growth rate (for instance, when dealing with *Streptomyces* this precaution should be considered).

<u>For inoculation</u>

1. Enough medium was placed in a flask or in a tube.

Normally, TSB is used for *Streptomyces* and terrific broth (TB) for *E. coli*.

When inoculating microorganisms, some space should be left without medium, for there to be oxygen for them to breath. For instance, in Falcon tubes (with the capacity of 50 mL) only 5 to 10 mL of their volume should be used.

2. The flask/tube was tilted and the antibiotic, if needed, was added.

When small volumes are added, the plunger button of the micropipette should be pressed up and down several times.

3. A colony was taken with a toothpick from one petri dish where it grown.

Only individual colonies that are grown enough should be picked. This should be done, because if colonies that are joined together are used, there is a chance that there will be different genomes. This shouldn't happen, because all colonies are clones, have the same genome, but this fact is not always true.

4. The flask/tube was tilted and the tip of the toothpick with the colony was placed in the medium.

Tilting is done to ensure that only the tip of the toothpick touches the medium to discard the microorganism.

5. The flask/tube was placed in the incubator shaker at the right conditions.

The optimal temperature for *Streptomyces* is around 30 ºC, while for *E. coli* is at 37 ºC at 250 revolutions per minute (rpm).

To stop the cells from growing more, they should be taken out from the shaker and placed on ice.

The minute the incubations leave the shaker, their cells start to die. Therefore, they should only be taken out when they are needed and when everything else is prepared.

For plating

1. The medium needed was melted and split into petri dishes.

Slipping medium into petri dishes should be done quickly, to avoid its solidification.

If antibiotic is needed, it needs to be added to the medium before it is split into petri dishes and when it's not too hot, or else the antibiotics can be degraded.

2. The medium was let alone to solidify in the petri dishes.

3. This step could be done in various ways, depending on where the culture of bacteria was and what was needed:

a. When the culture was in a liquid medium:

a.1. Some volume of it was added with a micropipette to the petri dish.

If the culture comes from a flask, only what is in the liquid should be used, not what is in the walls.

a.2. Then a spreader was used immediately to spread the bacteria around all the petri dish, in order to have equal conditions.

a.3. Then, they needed to dry.

If more than one volume of culture is plated in different petri dishes, the micropipette should start with the one with less volume. The same thing needs to be done with the spreader.

If the volume of sample added is too little, pure water should also be added to avoid it drying before it is spread.

b. When the culture was in another petri dish:

b.1. An individualized colony was picked either by a toothpick or a wire loop.

The first is normally used when the colony is plated in a small place inside the petri dish, while the second is used to spread the colony around all of it.

b.2. The toothpick or the wire loop spread the colony in the petri dish.

In this method, no bigger pieces of colonies should stay behind, in order for it not interfering in the results.

4. The petri dishes were placed in the dry oven at the optimized temperature for each microorganism.

For *E. coli* is 37 ºC while for *Streptomyces* is from 28 ºC to 30 ºC. Normally, *E. coli* is also able to grow in less than 16 h, while *Streptomyces* might take days, depending on the medium used.

After the microorganism is grown, the medium with it can only be saved one or two weeks in the fridge.

3. *Identification of microorganism**

Theoretical Foundation

Sometimes other microorganisms can be seen along with our samples in petri dishes. If that happens, identifying which microorganism is contaminating the sample might be important.

In this study, samples of microorganisms that were found in wine barrels of a company that produces wine were sent to analyse. This procedure was important in this case because the owner of the company needed to know if the wine that was in those contaminated barrels was drinkable, and, therefore, could be sold, or if the microorganism was nefarious and the wine needed to be discarded in order not to case any health issue.

Material

➢ Samples;

➢ Petri dishes;

➢ Flow hood;

➢ Bottles with sterile TSA medium;

Used since it is a general medium that can be utilized for most bacteria.

➢ Specific medium for lactic acid bacteria;

Only used because it is known that these are lactic acid bacteria.

Two media were used because in TSA different bacteria and contaminants can grow.

➢ 10 mL tubes;

➢ Test tube rack;

➢ Platform to balance the petri dishes;

➢ Micropipettes and tips;

➢ Digralsky spreader;

➢ Water bath;

➢ Computer;

➢ Dry oven.

Procedure

This procedure must be done in a flow hood.

1. TSA medium was placed in petri dishes.

2. From the samples gotten, two dilutions (10^{-1} and 10^{-2}) were done for each sample and for each dilution done, 100 µL were plated in the petri dishes already prepared with TSA.

This step was done in order to ensure that individual colonies would be obtained in some of the petri dishes.

3. The same amount of sample not diluted was plated in the same conditions.

4. These petri dishes were placed in the dry oven at 28 ºC.

This temperature was chosen because most bacteria can grow in these conditions.

5. The bottles of the medium specific for lactic acid bacteria were placed in a water bath at 52 ºC.

This temperature was chosen because it ensured that the medium continued in a liquid state and that the bacteria, when placed in it, survived.

6. From the dilutions and the original tubes of samples, another 100 µL were taken and added to 10 mL of the specific medium. This mix was then split into petri dishes.

This step needs to be done rapidly, in order for the medium not to solidify. Therefore, only one bottle of medium at a time can be taken from the water bath and only when everything else is already ready.

7. After doing the last step to every dilution and every sample, the petri dishes were let alone to dry.

8. While the last petri dishes filled with sample and medium were still wet, the first ones, already dried, were filled with more 10 mL of medium without bacteria.

Again, the necessary quantity of medium is only taken from the water bath when everything else is prepared, so it doesn't solidify.

This step is necessary to ensure that much oxygen doesn't exist, which is preferred by this kind of bacteria. If these were anaerobic, the process would be different*.

9. The petri dishes were placed in a dry oven at 28 ºC.

*If these bacteria were anaerobic, which they are not, more cautions would be needed in order to ensure that much oxygen doesn't enter in contact with them. Working with bacteria that are intolerant to oxygen is difficult and needs a specific incubation chamber for anaerobiosis. On the other hand, a special compound that degrades oxygen is also placed in the petri dish.

The procedure of identifying the microorganism doesn't end here. After the last step, and after colonies have growth in both media, the morphology of each of them is analysed. This is done since a different morphology might mean that different microorganisms are present.

Then, every different morphology is isolated, and its DNA is extracted. After having the DNA, it is sequenced and the obtained sequence is then compared to the other from the different morphologies, to conclude if they indeed come from the same microorganism or not (here the software "ContigExpress" might be a reliable option).

The sequencing procedure produces a file with "electropherogram", or "electrophoregram", which is a record or chart produced by electrophoresis, and it has different peaks and different colours for each nucleotide. The file also computes a quality value, which represents the reliability that the program has that that peak is indeed that nucleotide. When the value is "0", caution is needed. There is also the possibility of having two peaks in the same position, which might be caused by having two templates during polymerization. Smaller peaks at the end can also occur, which might be explained by the loss of intensity that the fluorescent chemical has in that moment. Normally, beginning and endings of the sequence are of low quality. The

vector NTI (Invitrogen) tool recognized by "ContigExpress" can be used to visualize the sequencing files.

Finally, the different sequences, that correspond to different microorganisms, are then run by "Nucleotide BLAST" ("Blastn"), which will finally be able to identify, by comparison with a database, with a certain percentage of identity, the microorganism. Here, the "closest" organism is decided by the "Query Cover", which is the alignment percentage between the two sequences, and the "E value", that is the number of alignments with similar quality gotten randomly.

Some more information on the organism can be retrieved through the program "Mycobank".

4. *Ligation reaction*

Theoretical Foundation

A ligation reaction is the process used to join two specific DNA molecules (insert/fragment and vector), to generally form a plasmid. Only some ligations, those that will originate a functional plasmid, will enter in the cell and multiply [68].

To bind insert (fragment of interest) to vector, there is the need to cut them. In order to do that, restriction enzymes are used. These are proteins that can recognize specific small sequences and can cut them or near them. While cutting them, two different ends can be originated by the enzymes **(Fig. 10)**:

● Blunt ends: Straight cuts that don't produce salient ends.

● Sticky ends: Cuts that produce salient ends that are complementary. These provide better results in a ligation experiment.

A

B
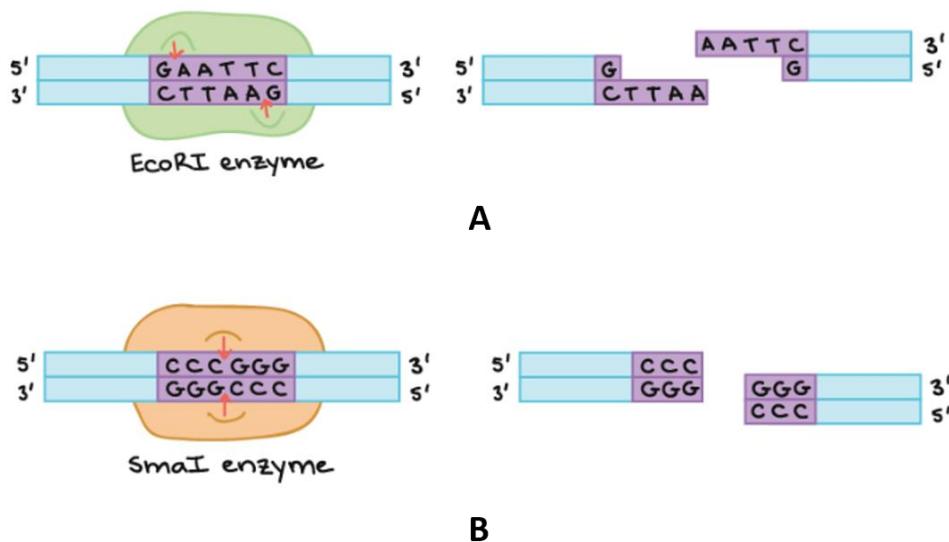
Fig. 10. Difference between sticky **(A)** and blunt ends **(B)**. Taken from [69].

The plasmid originated, if the ligation and transformation are successful, will be similar to the one represented below **(Fig. 11)**.
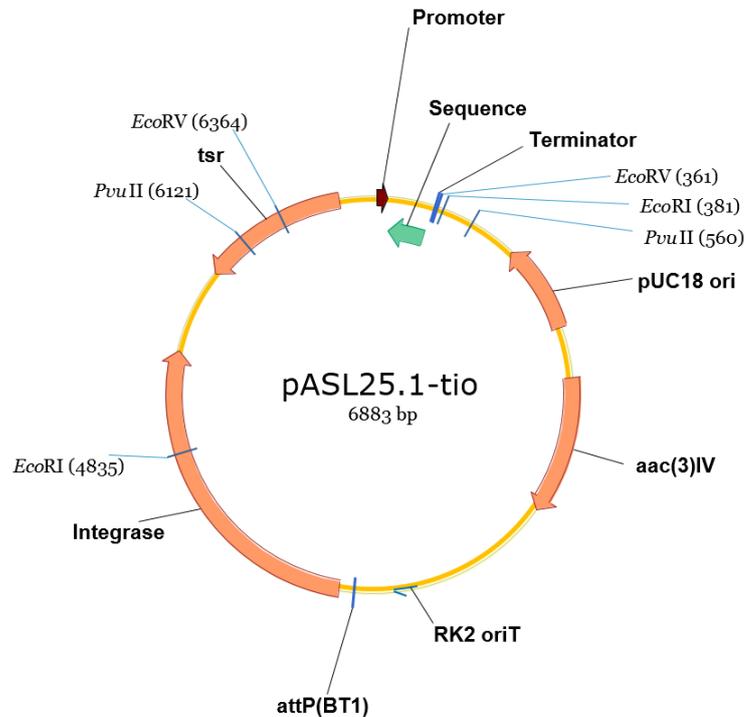
Fig. 11. Schematic representation of the plasmid constructed.

As seen in the image, the vector has resistance genes (in this case to thiostrepton, identified as "tsr" in the image, for *Streptomyces,* and to apramycin for *E. coli,* identified as "aac(3)IV"), that function as selecting markers. It also has one origin of replication for *E. coli* ("pUC18 ori"), a transfer sequence ("RK2 oriT") which enables the mobilization of the plasmid from a donor strain (*E. coli*) to the host strain (*Streptomyces).* There is also a site where the plasmid will bind with the genomic DNA ("attP") of *Streptomyces* and an enzyme to do that bonding, an integrase.

Finally, the plasmid contains an insert ("Promoter"-"Sequence"-"Terminator") that will originate an antisense of the sRNA of interest.

This scheme is also useful to know the size of the fragments gotten in an electrophoresis after digestion with restriction enzymes, since they have the coordinates where the enzyme cuts [for instance, here EcoRI cuts in the coordinate 4835 and in 381 base pairs (bp)]. This way, and by knowing that the whole plasmid has a length of 6883 bp (as seen in the figure), the sizes expected in the electrophoresis can be calculated.

The plasmid was obtained after ligation of a BclI-XbaI vector and a BamHI-SpeI insert with sticky ends, but these restriction enzymes are unable to recognize those sites again. This happens because when the vector and insert ends are united by a ligase, the sequence obtained is not palindromic like the original one was.

After transforming, to confirm if the ligation was successful, extracting the plasmid and digesting it needed. The ultimate verification is done afterwards, through an electrophoresis.

Material

➢ Vector;

➢ Insert DNA;

➢ T4 DNA Ligase;

➢ Buffer for the ligase;

➢ Restriction enzymes;

➢ Pure water;

➢ Eppendorf tubes;

➢ Centrifuge tube rack;

➢ Heat block;

➢ Incubator shaker;

➢ Micropipettes and tips.

Procedure

The procedure can have some small changes depending on the insert and the vector used. To increase the probability of joining the insert to the vector, ligations are normally done with small volumes.

1. Vector and insert were mixed together.

It is important to know the proportion between the quantities of insert and vector used and that is often of 3:1 or 5:1.

2. Ligase buffer and ligase were also added.

3. Two enzymes with different regions of cut (one for each of the vector targets) were used to originate distinguished ends where the insert would bind.

In this type of ligation, the insert can only bind with the vector from one direction and only one ligase is used because it is not specific and doesn't care if the cut regions are the same or not.

4. The samples were incubated.

The temperature and the duration of the incubation depends on which type of cut will happen:

- Since blunt ends are more difficult to bind, the temperature normally used is between 4 and 6 ºC and the ligation incubates overnight.

- On the other hand, sticky ends are normally bound at room temperature (about 22 ºC) for 2 or 3 hours.

5. After this, the ligation was completed and the enzyme was inactivated by placing the samples for 10 min at 65 ºC.

The next step was to transform this ligation inside one bacterium to turn it into a plasmid.

5. _Transformation (chemical procedure)_

Theoretical Foundation

Transformation is a method in which a foreign DNA present in the medium is introduced into a cell **(Fig. 12)** [70]. Even though this occurs naturally in some bacteria, its efficiency changes depending on the species.
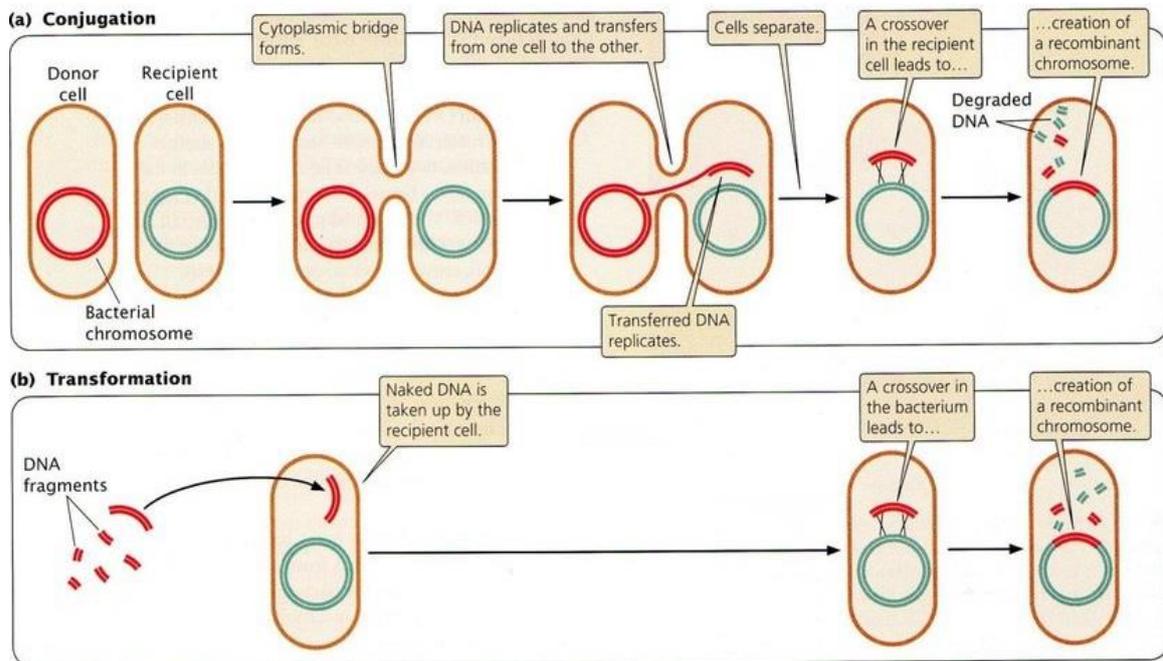
Fig. 12. Schematic representation of transformation and conjugation. Adapted from [70].

In the laboratory, this technique is very helpful, since it allows almost every plasmid, no matter if they are in their circular or in their supercoiled form, to be introduced in nearly all bacteria.

In order for the transformation to happen, the bacteria must be in a so-called competent state, which is only achieved in certain physiological conditions. This state is crucial for the success of the technique because it ensures that both the wall and cell membrane present changes that allow the entrance of nucleic acids in the cell.

Even though there are many species that are not able to present this state, some methods, based on physical and chemical treatments, were created in the laboratory to induce it. These treatments produce micropores in the cell, which allow the introduction of the foreign DNA in an efficient way.

To detect if the transformation was successful or not, the plasmid has a selective marker which will make the bacteria present characteristics, such as resistance to an antibiotic, that will facilitate the differentiation between the cells that were transformed and the ones that weren't. Here, _E. coli_ must present resistance to apramycin, as seen before, in order for the transformation to be concluded as successful. However, having the resistance gene as a selective marker has some problems because it implicates adding antibiotic, which can lead to secondary effects.

This protocol was done to two different strains of _E. coli_, since every strain has a specific function. One of them, _E. coli_ DH5-alpha, always used first, methylase positive, is used to

confirm if the ligation was indeed well done, if the plasmid is built the way wanted. This step is needed because the second strain, *E. coli* ET12567 pUZ8002, the one used to conjugate with *Streptomyces*, already has a plasmid (pUZ8002), which makes it harder to analyse the result of a ligation. This plasmid that this strain has is also important for the conjugation, since it has transfer (tra) genes responsible for the transference of vectors that have "RK2 oriT" transference origin **(Fig. 11)**. Another characteristic that makes this strain apt for conjugation is the fact that it is methylase negative, and, if it wasn't, *Streptomyces* would recognize the plasmid as foreign and would reject it. Therefore, when it is planned to introduce a construction in *Streptomyces*, first the result of the ligation is transformed in *E. coli* DH5-alpha, then the transformants are analysed by minipreparations (minipreps). After that, the desired plasmid is isolated and transformed in *E. coli* ET12567 pUZ8002 and the derived strain is used to conjugate with *Streptomyces* (**Fig. 13**).
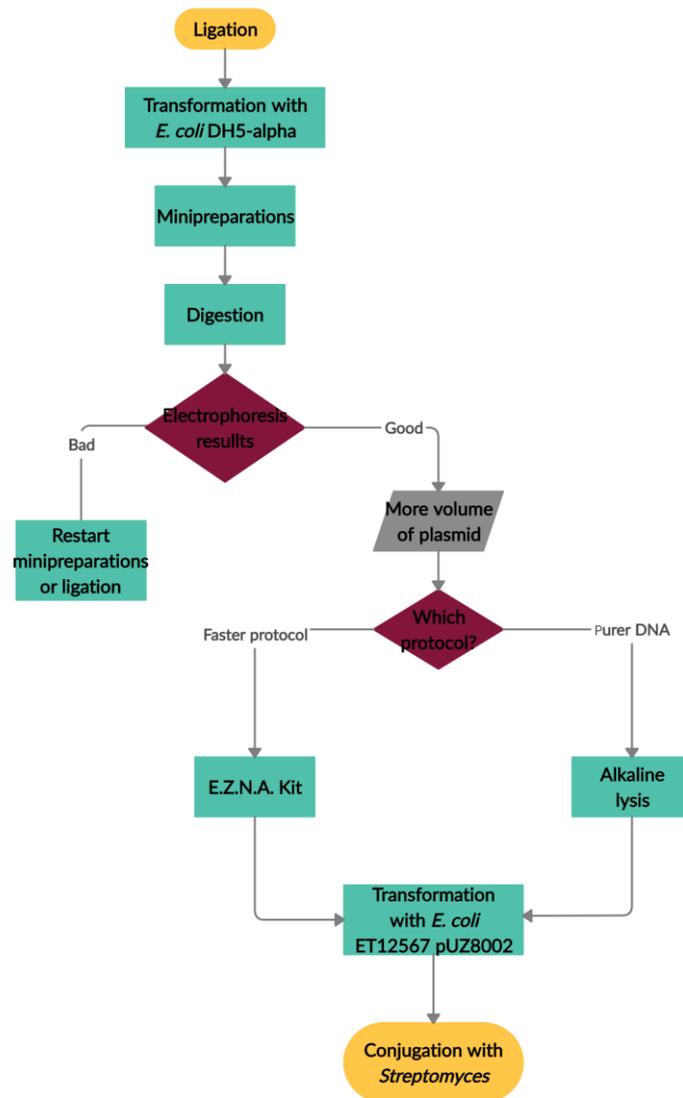
Fig. 13. Flowchart summing up the studies done with the two strains of *E. coli.*

Material

➢ *E. coli* competent cell*s*;

These cells were treated to be able to be transformed. If not for this process, the cells wouldn't accept foreign DNA.

➢ Plasmid DNA;

➢ Eppendorf tubes;

➢ Centrifuge tube rack;

➢ Micropipettes and tips;

➢ Flow hood;

➢ Heat block;

➢ Petri dishes;

➢ Bottle with Super Optimal broth (SOB);

➢ Ice and a container where to put it;

➢ Luria Agar (LA) medium;

➢ Antibiotic;

Here, apramycin was used.

➢ Incubator shaker;

➢ Dry oven.

Procedure

As said before, before beginning this procedure, the cells need to be turned into their competent state. Once this and the ligation of vector and insert are done, the next steps can be carried out.

This procedure has steps that should be done in the flow hood.

1. The tubes with the competent cells were defrost and placed in the ice.

The competent cells need to be in ice because they are sensitive to heat.

2. The plasmid was added to the cells and the mix was kept in ice for 30 min.

The volume of plasmid should never be more than 1/10 of the volume of the competent cells. However, there are some strains that are extremely difficult to transform. When using those, more plasmid can be added to increase the probability of having a successful transformation. Another way to increase that probability is to evaporate some volume of competent cells in order to concentrate them (this can be done by leaving the tubes open in the heat block).

3. A thermal shock of 42 ºC for 45 seconds (s) was submitted to the tubes with the mix. After then, they were returned to the ice for 2 min.

4. SOB was added and the tubes were taken from the ice.

5. The cells were incubated at 37 ⁰C for 1 hour while agitating, to make sure that they would replicate.

6. Apramycin was defrost and added into the LA medium already melted.

7. LA medium was split into petri dishes.

8. Different volumes of the transformation (here were used 20 µL, 150 µL and 400 µL) were plated in petri dishes with LA medium with antibiotic to be able to select those that were successfully transformed.

The antibiotic added to the medium was the one that the vector is resistant to. Therefore, only the bacteria that had incorporated the vector should grow in the petri dishes.

9. The tubes were incubated at 37 ⁰C in a dry oven for 16 h.

If not many transformed colonies are obtained, either the competent cells weren't efficient or not much plasmid was added. Another problem that can happen is related to the so-called satellite colonies, which are small colonies that come from cells that weren't transformed properly, because the plasmid with the resistance gene wasn't able to enter their interior. Therefore, these shouldn't be able to grow in the medium with the antibiotic. However, they grow here because they are surrounded by those who were successfully transformed.

6. *Transformation (electrical procedure)**

Theoretical Foundation

The electroporation (or electrotransformation) consists in an electric transformation, that works by applying a brief and intense electric pulse, that will destabilize the cell membrane, by creating pores where the molecule wanted to insert will enter **(Fig. 14)**. Therefore, this can be considered an electric way to transform cells [71].
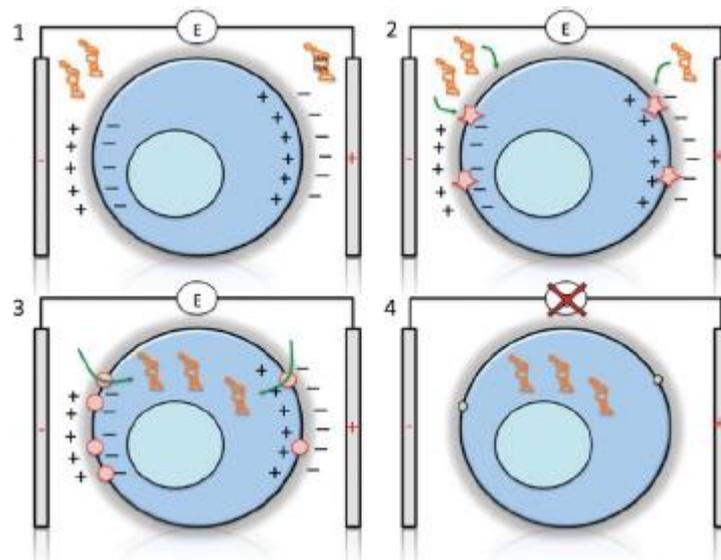
Fig. 14. Main steps of electroporation. **(1)** Short electrical pulses are applied to polarize the cell; **(2)** breakage of the membrane, which creates nanopores; **(3)** entrance of macromolecules; **(4)** the electrical field is switched off and the membrane is resealed. Taken from [71].

Material

➢ Electrocompetent *E. coli* ET12567 pUZ8002;

➢ Plasmid;

➢ Eppendorf tubes;

➢ Centrifuge tube rack;

➢ Test tube rack;

➢ Test tube with cold Luria broth (LB) medium;

➢ Bottle with LB medium;

➢ Antibiotics (chloramphenicol and kanamycin);

Chloramphenicol is used because this strain is resistant to it, while kanamycin is utilized because this strain has a plasmid that is resistant to it.

➢ Antibiotic for plasmid (in this case was apramycin);

Apramycin was added because the plasmid that will be inserted in the strain is resistant to it.

➢ Flask;

➢ Micropipettes and tips;

➢ Electroporation cuvettes (with two electrodes for the current to circulate);

➢ Electroporator;

➢ Flow hood;

➢ Shaking incubator.

Procedure

1. The electrocompetent cells were mixed with the plasmid wanted to insert.

2. The cells were then placed in the electroporation cuvette.

The cuvette has metallic parts, which shouldn't be touched, since it is there that the current will circulate.

When putting the cells here, bubbles shouldn't be formed and the volume in the cuvette shouldn't surpass 50 µL.

3. The cuvette was then inserted in the electroporator.

Here, the cuvette should be placed with its metallic parts directed towards the operator.

4. The electroporator was switched on and its parameters adjusted. For this strain of *E. coli*, the parameters used were:

- Electric resistance: 200 Ω

- Voltage: 2.5 kV

- Electric capacitance: 25 µF

The electroporator gives its results in milliseconds (ms).

5. Cold LB medium was added immediately to the cells electrotransformed.

6. The cells were placed in Eppendorf tubes, which were put in a shaking incubator at 37 ⁰C for 45 min.

7. In a flow hood, the LB medium was transferred from the bottle to the flask.

8. Chloramphenicol, kanamycin and apramycin were added to the same medium.

As said before, these antibiotics were chosen since this strain *E. coli* ET12567 pUZ8002 is resistant to one of them, chloramphenicol, and it has a plasmid resistant to another, kanamycin. When these cells are transformed, they also present resistance to apramycin, due to the ligation. This way, it is possible to select the competent cells that acquired the plasmid.

9. Cells were selected to be added to the medium.

The selection was done by choosing the ones that had the best values in the electroporator (and some that didn't, to understand what went wrong with those). If the value was the same, the parameter used to untie them was choosing the one with the largest volume.

10. Cells were then added to the medium with the antibiotics.

11. The flask was placed in a shaking incubator at 37 ⁰C for 18 h.

7. *Plasmid Extraction*

Plasmids are extrachromosomal DNA molecules, normally present in bacteria, able to replicate independently of the chromosome. Most of them are circular, but there are some that have already been identified as linear. Even though they are not fundamental to the survival of the cell, they can bring numerous selective advantages to it when it is in a competitive environment [72].

Plasmids can be in three conformations: CCC (Covalently Closed Circular) or supercoiled; OC (Open Circular); and linear. The first one corresponds to the native confirmation found *in vivo* and it is almost inaccessible, while the seconds is seen when there is replication. Finally, the linear one is originated when the DNA is cut in both chains at the same place [73].

Plasmids have been used as tools in genetic manipulation and DNA cloning in different areas, as in biotechnology and molecular biology. Their small size, when compared to the chromosome's, makes their extraction and manipulation easier. Moreover, the characteristics of replication and conjugation that the plasmids present are very useful for genetic analysis [72].

There are several methods that can be done to extract plasmid. In this study, three were used: minipreps; E.Z.N.A. kit; and alkaline lysis. Normally, minipreps is done first, to quickly confirm the plasmid construction. However, in the end of that procedure, not much quantity of plasmid is obtained. This way, to perform downstream applications (all that are directed toward the upper 3' end of the DNA or RNA strand), one of the other two methods needs to be done. The E.Z.N.A kit is able to obtain the plasmid quicker than the alkaline lysis, but the latter

gives more of it and purer. Therefore, the choice between which of the last two to use depends on how much plasmid is wanted and what the plasmid will be used for.

Again, these protocols were done to two different strains of *E. coli*. One of them, DH5-alpha, always used first, is used to confirm if the ligation was indeed well done, if the plasmid is built the way wanted, and the second one, ET12567 pUZ8002, is the *E. coli* donor strain, that needs to be transformed, to conjugate with *Streptomyces*.

7.1. *Minipreparations*

Theoretical Foundation

A minipreparation (or miniprep) is the extraction of plasmid DNA from a bacterial culture.

In this minipreps protocol, the *E. coli* transformants are checked to look for the desired plasmid after the ligation. In one petri dish with solid medium and also suspended in liquid medium, each one of the selected colonies is grown, all of them taken from an original petri dish. The suspension in the tubes is later used to analyse if the plasmid of interest is present.

Material

➢ *E. coli* growing in solid LA medium with apramycin;

➢ Petri dish with solid LA medium with apramycin;

➢ Bottle of liquid TB medium;

➢ Phosphate salts;

Needed for the growth of the microorganism.

➢ Apramycin;

➢ STET buffer;

Composed by sucrose, Triton X-100, trisaminomethane-hydrochloric acid (TRIS-HCl) and ethylenediaminetetraacetic acid (EDTA), which binds with magnesium ion (Mg^{2+}), a crucial cofactor needed for the function of nucleases. This way EDTA decreases the DNA degradation.

➢ Lysozyme;

Breaks cell walls.

➢ Sodium acetate;

➢ Isopropanol;

➢ 70-80 % ethanol;

➢ Tris-EDTA (TE) buffer;

➢ Water;

➢ 1.5 mL Eppendorf tubes;

➢ Test tube rack;

➢ Centrifuge tube rack;

➢ Flow hood;

➢ 10 mL test tubes;

➢ Flask;

➢ Spatula;

➢ Chopsticks;

➢ Micropipettes and tips;

➢ Pressure cooker;

➢ Microcentrifuge;

➢ Incubator shaker;

➢ Dry oven.

Procedure

The first part of this protocol (steps 1-8) was done in a flow hood in order to obtain sterile conditions.

1. A petri dish with solid LA medium with apramycin was prepared.

Antibiotic should be added to the medium while it is still in a liquid state, but not hot enough to degrade the antibiotic.

2. The petri dish was identified with all the numbers of minipreps.

3. A sterile bottle of TB medium was prepared, and phosphate salts and the right volume of apramycin were added and mixed.

The antibiotic was added here, because, since the bacteria grew with it, if the medium doesn't have it, the plasmid might disappear.

4. Eppendorf tubes with the same numbers written in the petri dish were prepared.

5. Medium prepared earlier was added to every single tube.

6. A colony (only individual colonies should be used) was picked with a sterile chopstick from the petri dish where *E. coli* was growing and the surface of the petri dish prepared was scratched, on the number assigned to that colony.

7. That same chopstick used was immersed on the Eppendorf tube, filled with medium, with the same number.

8. Once all colonies wanted were picked, the chopsticks were removed and the tubes closed.

9. The petri dish was incubated in a dry oven at 37 ºC for 16 hours (approximately) and the Eppendorf tubes in an incubator shaker at the same temperature during the same amount of time.

The bottle with medium used and the original petri dish were kept in the fridge.

10. After that amount of time, the Eppendorf tubes were put in a rack and centrifugated during 1 min at 12000 rpm, in order to precipitate the cells.

11. The supernatant was removed to eliminate the rest of the medium by overturning the tubes over a flask or a beaker tube. To make sure that there was no more medium, the tubes were hit against a piece of paper. The precipitate stood in the bottom of the tube.

12. Water was boiled in the pressure cooker.

13. STET buffer was added and the cells of every tube were resuspended.

14. Lysozyme (needed to break the cell walls) was added to the lid of every tube.

The tubes cannot be closed in this step, in order to ensure that the substance doesn't enter in contact with the samples at different times.

15. After ensuring that the water was indeed boiling, the tubes were closed and inverted (around 10 times) to mix the lysozyme with the sample.

16. Next, the tubes were put in the boiling water for exactly 40 s.

17. After taking them from the water, a centrifugation during 10 min at the maximum rotations and room temperature was done.

18. The precipitate of cellular remains (as membranes and proteins) was eliminated with a sterile chopstick.

The precipitate should get out easily. If not, something went wrong, and the procedure should be started from scratch.

The suspension that remains in the Eppendorf is filled with nucleic acids, proteins, and others.

19. Sodium acetate and isopropanol were added, in that exact order.

This is done to precipitate the nucleic acids present in the suspension.

20. The tubes were mixed, by inversion, around 20 times.

21. The tubes were left to rest for around 5 to 10 min.

22. The samples were centrifugated during 10 min at the maximum rotations and at room temperature.

23. The supernatant was eliminated by overturning the samples.

The pellet that stays in the bottom of the tubes has RNA, genomic DNA and other molecules (such as proteins) that are of no interest.

24. The precipitate was washed with 70-80 % ethanol, the tubes were mixed, by inversion, 10 times and centrifugated around 3-5 min at the maximum rotations and at room temperature.

25. The supernatant was eliminated by overturning the tubes.

26. The tubes were left open for the precipitate to dry during 15 min so that the remaining of isopropanol or ethanol might evaporate.

27. The samples were resuspended in TE buffer.

7.2. *E.Z.N.A. kit*

Theoretical Foundation

After confirming that the ligation is correctly done (by doing minipreps), the plasmid needs to be extracted, in order to place it in *E. coli* ET12567 pUZ8002 that is used to conjugate with *Streptomyces.*

This protocol is one of those methods, and it consists on getting, in a fast and efficient way, small quantities of plasmid by using an E.Z.N.A. Plasmid DNA Mini Kit [74].

E.Z.N.A. is part of many products that simplify the process of extraction and purification of nucleic acids from various sources [74].

DNA Mini Kit I, the one used here, is one of the many products from E.Z.N.A. able to extract plasmid. It uses a so-called HiBind matrix that can, under optimized conditions, remove contaminants, such as proteins, and bind, specifically and reversibly, DNA or RNA. On the other hand, HiBind DNA Mini Columns enable the processing of multiple samples simultaneously and facilitates steps of binding, washing and elution [74].

After having the purified plasmid DNA, automated fluorescent DNA sequencing and digestion with restriction enzymes can be done, before trying to transform it in *E. coli* ET12567 pUZ8002 needed for conjugation.

The protocol here written is the one that comes with the kit.

Material

➢ Samples of transformed *E. coli* DH5-alpha plated in a medium with antibiotic;

The antibiotic was apramycin, as explained before.

➢ HiBind DNA Mini Columns (comes in the kit);

These should be handled very carefully and only touched at the top.

➢ 2 mL Collection Tubes (comes in the kit);

➢ Solution I (comes in the kit);

➢ Solution II (comes in the kit);

Should be tightly closed when it is not being used to avoid acidification from carbon dioxide present in the air.

➢ Solution III (comes in the kit);

➢ HBC Buffer (comes in the kit);

➢ DNA Wash Buffer (comes in the kit);

➢ RNase A (comes in the kit);

➢ Elution Buffer (comes in the kit);

Needs to be heated to 70 ºC if the plasmid DNA is longer than 10kb.

- ➢ 100 % Ethanol;

- ➢ 100 % Isopropanol;

- ➢ Sodium hydroxide (NaOH);

- ➢ Falcon tubes;

- ➢ 1.5 mL microcentrifuge tubes;

- ➢ Test tube rack;

- ➢ Centrifuge tube rack;

- ➢ Micropipettes and tips;

- ➢ "NanoDrop";

- ➢ LB medium with antibiotic;

Here, as seen before, apramycin was used.

- ➢ Incubator shaker;

- ➢ Centrifuge;

- ➢ Microcentrifuge;

Capable of, at least, 13000 relative centrifugal force (RCF).

- ➢ Vortex;

- ➢ Ice and container where to put it.

Procedure

This procedure should be done while wearing gloves, to avoid contamination.

Before beginning

- o The vail of RNase A was added to the bottle of Solution I, which was then stored at 2-8 ºC.

- o 100 % ethanol was added to the DNA Wash Buffer, which was stored at room temperature.

The volume added depends on the kit used.

- o HBC Buffer was diluted with 100 % isopropanol and stored also at room temperature.

The volumes added also depend on the kit used.

After these preparations are done

1. A single colony from a freshly streaked selective plate was inoculated in LB, that had the appropriate selective antibiotic, and incubated for 12 to 16 h at 37 ºC with vigorous shaking.

The culture volume should not exceed ¼ of the volume of the container.

2. After this amount of time, the samples were placed on ice and were centrifugated at 10000 RCF for 1 min at room temperature.

3. The media was discarded.

4. 250 µL of Solution I with RNase A were added to the samples. To mix it well, vortex was used.

Complete resuspension of cell pellet is crucial to obtain good results.

After using it, the solution used should be immediately stored at 2-8 ºC.

5. The suspension was transferred into a new 1.5 mL microcentrifuge tube.

Only from this step on, can the samples get out of the ice.

6. 250 µL of Solution II were added and the samples were inverted and gently rotated until a clear lysate was obtained.

It is important to do the mixing gently, in order not to shear the chromosomal DNA, which would lower the plasmid purity.

7. An incubation of 2-3 min at room temperature was done.

This incubation shouldn't exceed 5 min.

8. 350 µL of Solution III were added and the tubes were immediately inverted several times until a flocculent white precipitate was formed.

If the solution is not mixed immediately, precipitation can occur.

9. The tubes were then centrifugated at maximum speed (≥13000 RCF) for 10 min.

10. While the centrifugation was in process, the following steps were done to balance the column:

      a) A HiBind DNA Mini Column was placed into a 2 mL Collection Tube.

      b) 100 µL of NaOH were added to the HiBind DNA Mini Column.

      c) A centrifugation was done at maximum speed for 30-60 s.

      d) The filtrate was discarded and the collection tube was reused.

11. The cleared supernatant gotten from Step 9 was carefully transferred into the HiBind DNA Mini Column balanced in the previous step.

The aspiration of the supernatant needs to be done with caution, in order to avoid touching the pellet, which can transfer cellular debris to the HiBind DNA Mini Column.

12. The tubes were centrifugated at maximum speed for 1 min and the filtrate was discarded and the collection tube reused.

13. 500 µL of HBC Buffer diluted with isopropanol were added.

14. The tubes were again centrifugated at maximum speed for 1 min, the filtrate was discarded and the collection tube reused.

15. 700 µL of DNA Wash Buffer with ethanol were added.

16. The tubes were again centrifugated at maximum speed for 1 min, the filtrate was discarded and the collection tube reused.

17. Steps 15 and 16 were repeated.

18. The now empty HiBind DNA Mini Column was centrifugated for 2 min at maximum speed.

This was done to dry rests of ethanol, since they can interfere with downstream applications such as automated fluorescent DNA sequencing and restriction enzyme digestion.

19. The HiBind DNA Mini Column was transferred to a clean 1.5 mL microcentrifuge tube.

20. 30-100 µL of Elution Buffer were added directly to the center of the column membrane.

21. The samples were let sit at room temperature for 1 min and centrifugated at maximum speed for the same amount of time.

22. The samples were placed in the vortex and in the centrifuge for a quick pulse.

23. The samples were analysed through "NanoDrop" (process explained a few sections ahead) and Elution Buffer was used as a blank.

After this, the samples must go through a digestion and electrophoresis (both also explained ahead). Why is this needed when this plasmid was already digested twice in the minipreps protocol? This third digestion, after extracting the plasmid with the kit, is needed to ensure that during this procedure of extraction nothing happened to it.

7.3. *Alkaline lysis**

Theoretical Foundation

In alkaline lysis, both DNAs (chromosomal and plasmidic) are denaturalized and posteriorly renaturalized and, while the plasmid will renaturalize quickly and return to its natural conformation, the chromosomal DNA will not and might be trapped between proteins. With this knowledge, this method is able to differentiate one DNA from the other and extract big amounts of pure plasmidic DNA [75].

Material

➢ Glucose-Tris-EDTA (GTE) buffer;

The glucose maintains the osmotic pressure, while Tris maintains the pH value at 8. EDTA binds with Mg^{2+}, a crucial cofactor needed for the function of nucleases, which decrease the DNA degradation. This way, the degradation of DNA is avoided.

➢ Lysozyme;

Enzyme that breaks the cell wall.

➢ Sodium hydroxide/Sodium Dodecyl Sulfate (NaOH/SDS) buffer;

On its own, NaOH denaturalizes plasmidic and ribosomal DNA. On the other hand, both components mixed induce cellular lysis and denaturalize chromosomal DNA and proteins.

They also induce the release of the plasmid, which is not affected much because of its size and its structure.

➢ Potassium acetate- acetic acid solution;

Induces the precipitation of proteins and of chromosomal DNA.

➢ Ammonium acetate or sodium acetate;

Used to precipitate nucleic acids.

➢ TE buffer;

Increases the DNA stability, by inhibiting deoxyribonucleases (DNases) present in the samples.

➢ Neutral phenol;

Precipitates proteins.

➢ Chloroform-isoamyl alcohol (CIA) 24:1;

Used to denaturalize and to precipitate the proteins and to eliminate rests of phenol.

➢ *E. coli* DH5-alpha;

➢ TB medium;

➢ Phosphate salts;

➢ Lithium chloride;

➢ Isopropanol;

Used to precipitate nucleic acids.

➢ 70 % ethanol;

Used to eliminate proteins.

➢ 80 % ethanol;

➢ Ice and container for it;

➢ Flasks;

➢ Eppendorf tubes;

➢ Test tube rack;

➢ Centrifuge tube rack;

➢ Falcon tubes (50 mL);

➢ Tip of micropipette of 5 mL with hydrophilic cotton inside it;

➢ Micropipettes and tips;

➢ Vortex;

➢ Centrifuge;

➢ Incubator shaker.

Procedure

1. The strain of *E. coli* was grown in TB medium and phosphate salts (with antibiotics, which would be apramycin here), in flasks, at 37 ⁰C and 250 rpm for 16 h.

2. The cells were centrifugated in Falcon tubes for 4 min at 4400 RCF. The supernatant was discarded.

3. GTE, lysozyme and NaOH/SDS were added.

SDS will break the cell walls and denaturalize the chromosomal DNA, and both will precipitate. On the other hand, plasmidic DNA will still be present in the supernatant because its chains don't break when SDS is used.

4. The tubes were inverted 15 times and then maintained at room temperature for 5 to 10 min.

5. Potassium acetate- acetic acid solution (maintained at 0 ⁰C) was added and the mix was strongly shaken.

6. The tubes were incubated in ice for 4 min.

In the step 5, while agitating, foam appears. By maintaining it in the cold, the rest of cellular components will be separated depending on their density. This way, different areas inside the tubes will be distinguished: an upper one, which has rests of membrane; a middle one, the supernatant, with nucleic acids present; and one at the bottom, the precipitate, that also has membranes.

7. The samples were centrifugated for 10 min at 4000 rpm and at a temperature of 0 ⁰C.

While centrifuging, the nucleic acids will stay in the supernatant and the rest of the cell component will precipitate.

Here, the centrifuge is used at 0 ⁰C because there are salts that need to precipitate, while at room temperature is used to obtain DNA.

8. The supernatant was filtered by putting hydrophilic cotton inside the tubes.

9. 0.6 volumes of isopropanol were added to precipitate the samples. The tubes were inverted.

10. The samples were incubated at room temperature for 10-20 min, in order to precipitate the plasmidic DNA and contaminant RNA and separating them from the rest of the cell components.

11. They were then centrifugated for 15 min at 4000 rpm at room temperature.

12. The precipitate was cleaned with 70 % ethanol and then it was resuspended.

13. The tubes were centrifugated 5 min at 4000 rpm.

14. The supernatant was eliminated and the precipitate was let at room temperature to dry a little.

15. TE was added and the tubes were resuspended at room temperature.

16. The samples were placed in Eppendorf tubes and 1 volume of lithium chloride was added, which precipitates RNA.

17. The tubes were let rest for 5 min and then centrifugated for 5 min at 12500 rpm.

18. The supernatant was placed in another tube and 1 volume of isopropanol was added to precipitate it. The tubes were let rest for 5 min.

19. The samples were centrifugated for 5-10 min at 12500 rpm.

20. 80 % ethanol was added. The tubes were inverted and centrifugated for 5 min at 12500 rpm.

21. The samples were resuspended with TE.

22. Phenolization was done.

Phenolization is a process done often to eliminate rests of proteins from nucleic acids preparations. Here, neutral phenol is used to precipitate the proteins, while the DNA stays in the solution. CIA is added to denaturalize and precipitate proteins (in a less effective way than phenol) and eliminate any rests of phenol.

a) The samples were mixed with a volume of neutral phenol and were agitated horizontally for 1 min.

b) The tubes were centrifugated for 5 min at maximum rotations.

c) The supernatant was taken carefully, in order not to drag the interphase that has clusters of proteins.

d) One volume of CIA was added.

e) The samples were agitated horizontally for 1 min and centrifugated for 3 min at maximum rotations.

f) The steps d) and e) were repeated until the interphase appeared translucent.

g) The supernatant was taken and one volume of CIA was added.

h) The samples were agitated horizontally for 1 min and centrifugated for 3 min at maximum rotations.

23. 2.5 volumes of ethanol were added.

24. Samples were maintained at temperature -20 ºC for, at least, 30 min.

25. They were centrifugated for 10 min at 12500 rpm, at 4 ºC.

26. 80 % ethanol was added to the samples to precipitate them.

27. Tubes were inverted and centrifugated for 5 min at 12500 rpm.

28. TE was added to resuspend.

8. *Conjugation*

Theoretical Foundation

This protocol is for an intergeneric conjugation between *E. coli* ET12567 pUZ8002 with the plasmid built and *Streptomyces*.

A conjugation is the process where genetic information is transferred from the donor cell to the receptor one by direct contact between the two **(Fig. 12)** [70].

Producing secondary metabolites has been hampered by lack of cloning vectors and of efficient ways to transfer genes [76]. This technique seems to help in the last parameter, since it has been already successfully used in several *Streptomyces* species [77].

Material

> ➢ Samples of *Streptomyces* (conserved in glycerol at -80 ºC);

> ➢ *E. coli* ET12567 pUZ8002 transformed (conserved in glycerol);

> ➢ Chloramphenicol;

Chloramphenicol is used because this strain of *E. coli* is resistant to it.

> ➢ Kanamycin;

Used because this strain of *E. coli* has a plasmid that is resistant to kanamycin.

> ➢ Antibiotic for plasmid (in this case was apramycin);

Apramycin was added because the plasmid inserted in the strain (during the transformation) is resistant to it.

> ➢ Nalidixic acid;

> ➢ Heat block;

> ➢ Sterile bottle of SM medium;

> ➢ Magnesium chloride ($MgCl_2$);

> ➢ Bottle of 2xTY medium;

> ➢ Petri dishes;

> ➢ TSA medium;

> ➢ Flask;

> ➢ Falcon tubes;

> ➢ Tubes 10 mL;

> ➢ Test tube rack;

> ➢ Digralsky spreader;

- ➢ Toothpicks;

- ➢ Micropipettes and tips;

- ➢ Distilled water (dH_2O);

- ➢ Dry oven;

- ➢ Incubator shaker;

- ➢ Centrifuge;

- ➢ Spectrophotometer and cuvettes;

- ➢ Ice and container to put it;

- ➢ Vortex;

- ➢ Glycerol;

- ➢ Flow hood.

Procedure

This procedure should be done in a flow hood. The method of conjugation itself only starts in "Day 2".

Day 1 (transformation)

1. The plasmid was introduced into *E. coli* ET12567 pUZ8002 and the samples were store in glycerol.

Glycerol is used as a protective agent that prevents the death of the cells when reserved in the freezer. It is always used at 20 %.

Day 2

1. The *E. coli* transformants were transferred to 5 mL 2xTY containing chloramphenicol, kanamycin and antibiotic for the plasmid (apramycin).

This way only transformants will survive in the medium with all these antibiotics and the plasmid will not free itself (which can happen if the antibiotics are not added).

2. The samples were incubated overnight at 37 °C, 220 rpm.

3. SM plates containing $MgCl_2$ were prepared.

Day 3

1. An aliquot of the *E. coli* culture was diluted to measure the absorbance, or optical density at a wavelength of 600 nm (OD_{600}).

A blank cuvette is always needed to calibrate the spectrophotometer, and it must have everything but the microorganism (therefore, it will have the same medium with the same quantity of antibiotics as the *E. coli* culture).

Whenever the samples are not being used, the flasks/tubes need to return to the incubator shaker, to prevent the cells from dying.

A dilution needs to be done because absorbances bigger than 1 are not reliable.

To know the real value of absorbance, the result given by the spectrophotometer has to be multiplied by the inverse of the dilution.

OD_{600} can reach a value of 8 or more. For *E. coli* cell cultures, OD_{600} of 1 corresponds to 8 x 10^8 cells/mL. Such a relation can't be done for *Streptomyces* because of its filamentous morphology.

That range of optical density is important because, more than assuring that the concentration of cells is equivalent, it also ensures that they are in the same stage of growing (at a OD_{600} of 1.2, for instance, they are already in a stationary phase, and don't replicate as much).

2. Overnight culture with OD_{600} of 0.15 were inoculated at 37 ºC 220 rpm in 20 mL 2xTY media containing chloramphenicol, kanamycin and antibiotic on the plasmid.

The samples were incubated until OD_{600} reached 0.4-0.6.

Normally, *E. coli* replicates itself in every 20 min, therefore, if starting OD is 0.15 it will take 2 h and if it is 0.1 it will take 3 h to reach the value wanted of absorbance.

To stop them from overgrowing that range, after been taken from the shaker incubator, the samples should be placed and maintained in ice.

3. Cells were placed in vortex, centrifugated and washed twice with 20 mL 2xTY.

This step is done without antibiotics, since these washes are done to remove them from the samples. Moreover, *Streptomyces*, that will be added, is not resistant to these antibiotics.

Here, it is considered that the plasmid won't disappear, because that only happens after many h without the selective pressure from the antibiotics.

4. Cells were resuspended in 1 mL 2xTY using vortex.

5. Cells were diluted 1/10-1/50 from the 21-24 h culture of *Streptomyces* in 2xTY.

6. The cells were subjected to heat shock of 50 °C for 10 min and were cooled on ice.

7. *Streptomyces* were mixed with *E. coli*.

8. The mixture was centrifugated for 2 min, 6000 rpm, and almost all supernatant was discarded.

9. Cells were resuspended in the remaining supernatant.

10. 0.6 mL of the volume of the samples were discarded after centrifugation.

11. The rest was plated on SM plates containing $MgCl_2$.

Different volumes of cells were plated in the petri dishes: 20, 100 and 300 µL or 20, 60 and 100 µL.

12. The petri dishes were incubated for 16-20 h at 28 °C.

If mycelium is used, incubation should be done for less than that amount of time.

Day 4

1. Each plate was covered with dH_2O containing nalidixic acid and thiostrepton.

One of the petri dishes shouldn't be covered, in order to use as a control.

Thiostrepton is insoluble, therefore it needs to be mixed very well.

Here, the nalidixic acid should be the last one to be added, since it precipitates easily.

In this step, *E. coli* has grown already but individual colonies are still not seen in the petri dish, because there are many of them and the conditions weren't the optimal for this species.

The nalidixic acid in the coverage is used to inhibit the *E. coli* that might still exist in the petri dish, while the thiostrepton is used to select those *Streptomyces* that already have the plasmid **(Fig. 11)**.

The coverage is not spread with the aid of a Digralsky spreader, since using it will probably move some colonies, that still are not grown enough to see, closer together, which is not wanted, because having individual colonies is what is always ideal. Therefore, to spread the coverage without touching the medium with bacteria, the petri dish should be moved slowly until the coverage reaches every part of the petri dish. If the coverage doesn't get everywhere, some more volume of coverage can be added (but not much, maximum 300 µL more) or some that is "stuck" somewhere can be transferred to the place where there is none (carefully with a micropipette to not touch anything).

Bubbles shouldn't be left behind.

2. The plates were incubated for 5-7 days.

3. After that amount of time, the colonies grown were plated on TSA plates containing the same two antibiotics.

Only colonies that had the coverage on top of them should be used.

4. The petri dishes were incubated for around 2 days more, until the colonies were grown, at 28 °C.

If there is something already grown after only 24 h, it is a false positive and it should be identified in the petri dish, so that it is not used.

5. The colonies that had grown were plated and incubated again in the same conditions with nalidixic acid and thiostrepton.

Only individual colonies with spores should be used, since the others can be contaminants or even *E. coli*.

6. The last two steps were repeated twice more, to ensure that *E. coli* was inhibited (since *E. coli* is lost more and more every time) and only the conjugated *Streptomyces* was able to grow.

7. Some of the grown colonies were selected randomly to plate individually on SM small plates to recover spores and to go through Polymerase Chain Reaction (PCR).

No antibiotics were used here because the plasmid cannot disappear now, since it is already incorporated into the genome. Moreover, it would only slow the growth of the samples, since there would be the need to produce the protein responsible for the resistance to the antibiotics.

This was done individually, because it prevents the spores of one colony from contaminating the ones nearby, which shouldn't matter because the colonies should be genetically the same. But that doesn't happen, because *Streptomyces* has, normally, a very high genetic variability (many insertions, deletions, between other). However, the only mutation that normally is seen is the insertion of the plasmid, which has the sRNA that will be studied.

9. Cup plate method*

Theoretical Foundation

This protocol is done to analyse the inhibitory activity that *Streptomyces* might have in the growth of three microorganisms which are called indicators: *Candida utilis* (yeast), *Bacillus subtilis* and *Kocuria rhizhophila* (both bacteria).

The theoretical foundation of this procedure is that, if *Streptomyces* produces any inhibitory compound, zones of inhibition will be seen [78]. This method only tests the inhibitory activity of extracellular compounds. To analyse intracellular compounds, a more extensive method, that would include extraction, would be required.

Here, mutant strains of *Streptomyces,* in which biosynthesis aggrupations were eliminated, were used to see the difference between their inhibitory activity and the wild type's (with no modifications).

This theoretical foundation is always the same, but the procedure can change, by using, instead of liquid antibiotic, antibiotic disks with determined concentration.

Material

➢ Petri dishes with the grown microorganism that will be analysed (strain of *Streptomyces*);

➢ Bottles with medium (for instance TSA);

➢ Liquid culture of the three indicators microorganisms;

➢ Petri dishes;

➢ Falcon tubes;

➢ Platform;

➢ Pasteur pipette;

➢ Toothpicks;

➢ Micropipettes and tips;

➢ Dry oven;

➢ Heat block.

Procedure

The bottles of medium must be maintained in a heat block, but not in too high of a temperature, since it can kill the microorganisms.

1. A certain volume of medium from the bottles was added in a Falcon tube.

This is done to assure that every petri dish has the same amount of medium (to make the conditions as equal as possible).

2. Each of the three microorganisms was inoculated in one of the tubes already prepared with medium. Then these were carefully agitated so that bubbles wouldn't be formed.

The microorganisms are added here and not plated in the petri dishes, in order to assure that they grow evenly and throughout the thickness of the petri dish, and not only on its surface.

3. One platform was levelled, and the petri dishes were placed in it.

Levelling the platform is important, even more so when dealing with liquid antibiotics, since it will ensure that the compound will spread evenly in the petri dishes and, therefore, the zones of inhibition will be concentric.

4. The medium in each tube was placed in one petri dish. They were let cool down.

5. Cavities in the petri dish where the strains of *Streptomyces* were grown were made with the aid of a Pasteur pipette or a micropipette tip.

6. With the help of a toothpick, the circles made in the previous step were placed in the petri dishes with the indicators microorganisms.

7. The petri dishes were maintained between 30 min and 2 h at 4 ºC.

This step is crucial to enhance the diffusion of the expected component produced by the strain of *Streptomyces* from the circle to the petri dish and to prevent the growth of the indicators microorganisms before that happens. Here, the petri dishes can't be left in the normal way (part with medium and microorganism up), since the circle can fall.

8. The petri dishes were incubated at the temperate best suited for every indicator microorganism.

The petri dishes are placed at the temperate for each indicator, and not at the best temperature for *Streptomyces* because it doesn't matter if the latest grow. The only thing that matters is that the component has already diffused through the medium.

Methods in Molecular and Cellular Biology

1. Plasmid DNA Digestion

Theoretical Foundation

Once the plasmid is extracted, it's time to analyse it. Digestion by restriction enzymes is done frequently to analyse purified plasmid fragments, since they are able to cut in specific sequences [79].

It is known that the plasmid that is normally built has, as already seen, a size of 6 kilobases (kb) and, in order to linearize it in one or more fragments and later run it in an electrophoresis, some enzymes, that have their targets known, are used **(Fig. 11)** [79]. If this treatment wasn't done, the fragment would remain circular, have its 3 different conformations that wouldn't migrate in the same way and, therefore, the electrophoresis would show more bands, which would make it more difficult to analyse. This way, only the linear conformation is obtained.

As seen in the ligation procedure, this plasmid constructed has schematized the coordinates where three different enzymes cut: EcoRI, EcoRV and PvuII. The sequences that each recognizes and cuts are the following (**Table 1**):

Table 1. Example of the cutsites for three enzymes. The arrows represent where the enzymes cut. Information taken from [80].

Enzyme	Cutsite
EcoRI	5´...G▼AATTC...3´ 3´...CTTAA▲G...5´
EcoRV	5´...GAT▼ATC...3´ 3´...CTA▲TAG...5´
PvuII	5´...CAG▼CTG...3´ 3´...GTC▲GAC...5´

Material

> ➢ Plasmid sample;

> ➢ Restriction enzyme;

The enzymes normally come in units (U), which correspond to the quantity of needed enzyme to cut 1 µg of DNA in an hour.

> ➢ Buffer for the enzyme;

> ➢ RNase;

Needed to cut the RNA that might still be present which, if there, will appear in the electrophoresis.

> ➢ Sterilized water;

> ➢ Eppendorf tubes;

> ➢ Centrifuge tube rack;

> ➢ Micropipettes and tips;

> ➢ Dry oven;

> ➢ Freezer.

Procedure

1. The tubes were put in the rack and labelled.

2. A tube, that was used to gather all components in order to later split all the volume into the other tubes, was filled. The first component of the mix was the one with more volume, which normally is water. Then the buffer and the RNase were added.

If any solution is frozen (normally the buffer and RNase are), a wait time is needed, so that they can defrost before using them. To know which buffer to use, the tube of the enzyme should be verified, since the colour of the lid and the brand of the enzyme are the same as the tube that has the buffer needed for that enzyme.

Normally the solution is done for the number of samples that we have. But, to make sure that the risk of the pipettes not being calibrated (especially important since we are dealing with small volumes) is removed, it should be done for that number plus one or two.

3. Lastly, the enzyme was added carefully.

When the volume is taken, the tip of the pipette should not be immersed in the enzyme, since it is slimy and more volume that needed will come glued to the tip. If more volume stays in the tip, more patterns that the ones wanted appear in the electrophoresis, since it begins to cut in places that don't belong to the wanted target. If less than the volume wanted is added, the fragments stay intact or are only cut in one of the ends.

It needs to be added to the bottom of the tube with the solution, since it is immersed in glycerol.

The enzyme is in the freezer in one box filled with other enzymes. Even though only one of them is needed, the whole box should be removed from the freezer. Outside of the freezer, the enzyme should only be touched close to the lid, in order to not warm it. Moreover, the volume should be taken slowly and the enzyme put in the box and the box immediately in the freezer. There is no need to wait for the enzyme to unfreeze. The best thing is to make sure that it is not kept from the cold for too long, or it might lose its activity.

4. The solution prepared was divided between all tubes.

5. The plasmid samples were added to every tube.

6. Then the digestions were put in the oven at 37 ºC for 2 h, at least.

Since the temperature changes depending on the enzyme, a table of the brand with optimal conditions to each enzyme should be analysed first.

After this time, the digestions were prepared to go through electrophoresis.

2. *Electrophoresis*

Theoretical Foundation

A. *What is an electrophoresis gel?*

Electrophoresis is a technique that is used to separate DNA fragments (and other molecules, such as proteins and RNA) depending on their charge and size. It works by having an electric current pass through its gel, making different charged and different length molecules move differently.

Since every DNA molecule has the same quantity of charge for mass, the electrophoresis will separate them basely only on their size, allowing to see how many different fragments are present in the sample and how different they are [72]. The size of each fragment can be deduced by having a marker with known sizes, in order to compare them.

A dye is also used to mix with the marker and with the samples because it gives some weight to them, which avoids them from leaving the wells.

There are other characteristics than can change the movement of the fragments through the gel, other than their size, such as the conformation of the molecule itself. If, for instance, the molecule is circular (as it happens with plasmids that are not digested), it can run differently, and faster, than a linear one, even if their size is the same [72].

B. What's a gel?

A gel, when used for this reason, is made with agarose and has the consistency of jelly. Agarose is a polysaccharide called D-galactose/3,6-anhydro-L-galactose and its nature doesn't allow the separation of molecules that have sizes less than 15 nucleotides [72]. The gel is created when the polysaccharide is mixed with a buffer, such as Tris-Acetate-EDTA (TAE), and both are heated and then cooled down. On a molecular level, this gel is a matrix of agarose molecules united by hydrogen bridges that form small pores.

In one end of the gel some holes called wells need to be made, so that the samples can be placed (**Fig. 15**).

The gel, while still a little hot, is placed in a tray filled with the same buffer that was used to do the gel, in order to conduct the electric current. At each end of the tray there is one electrode: one positive in one extreme; and a negative in the other. The negative needs to be located closer to the wells and the positive at the opposite extreme, since the DNA molecules are negatively charged, due to their phosphate component and, therefore, will move towards the positive electrode (**Fig. 15**).
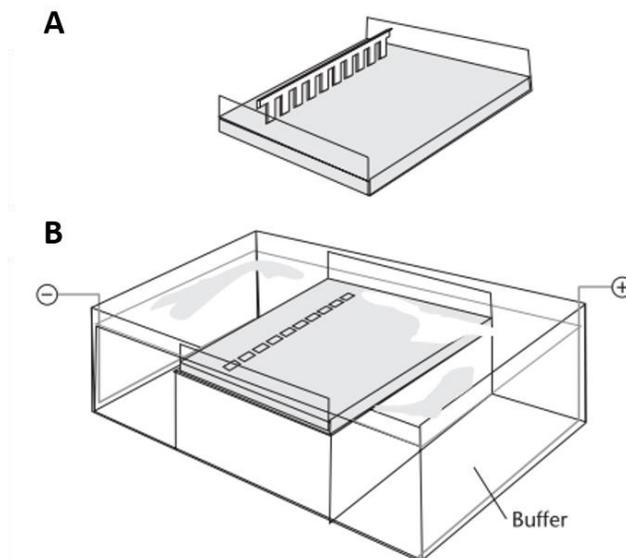
Fig. 15. Schematic representation of a chamber used for agarose gel electrophoresis. (A) Casting tray with comb; and (B) chamber with gel and buffer. The positive and negative electrodes are also represented in (B). Taken from [81].

The gel was used at 0.8 % (which means that, for each 100 mL of TAE buffer, 0.8 g of agarose are needed), but, depending on the case, can either have higher or lower percentage of agarose. A lower percentage is normally used to better differentiate bigger fragments, while a higher percentage is used for smaller fragments. A gel with a percentage higher than 0.8 % is thicker and, therefore, easier to manipulate, but doesn't absorb the ethidium bromide as well.

C. How do the DNA fragments move through the gel?

The agarose matrix, when the electrophoresis is on, functions as a molecular sieve which, with the aid of the current that begins to flow through the gel, facilitates the spread of the samples in the pores and the migration towards the positive electrode. The DNA molecules move towards the positive electrode, due to their phosphate group, which makes them negatively charged. While running, it is expected that the smaller fragments will move farthest in the gel, and the bigger ones will stay closer to the wells where the samples were placed.

A higher voltage will make the gel run faster, but, if on for too long, it might melt the gel. On the other hand, a lower voltage will make the gel run slower, which is an inconvenient if it is needed with urgency. If the electric current stays on for too long, the fragments can escape the gel.

D. Visualization of DNA fragments

After having the fragments separated, the gel can be analysed and the size of the bands there present, which are "lines" well defined, can be known. One band has many fragments of DNA, and all of them with the same size, since one fragment of DNA cannot be seen.

In order to be able to analyse the gel, a pigment able to bind and intercalate between the two strands of DNA must be added (such as ethidium bromide). Then the gel needs to be placed under ultraviolet light, to ensure that the bands can be seen [72].

The rough size of the band can be known by having, as said before, for comparison, a molecular-weight size marker.

Material

➢ TAE buffer 1x;

Constituted by Tris (molecule responsible for the pH regulation), acetate (responsible, as well, for maintaining the wanted pH) and EDTA (chelating agent of divalent cations able to abduct Mg^{2+}, which avoids the appearance of possible nucleases, since they need the element as a cofactor).

➢ Agarose;

➢ Sterilised water;

➢ Eppendorf tubes;

➢ Centrifuge tube rack;

➢ Graduated cylinder;

➢ Flask;

➢ Weight scale;

➢ Microwave;

➢ Mold, tray and comb;

To make the wells.

➢ Power supply;

➢ Samples;

➢ Loading dye;

➢ Molecular weight size marker;

To be able to know the size of the fragments gotten.

- ➢ Micropipette and tips;

- ➢ Ethidium bromide;

This product needs to be handled carefully, due to its mutagenic effects [82].

- ➢ Spatulas;

- ➢ USB flash drive;

- ➢ Electrophoresis machine;

- ➢ Transilluminator.

Procedure

1. The volume of TAE buffer needed was measured in a weight scale.

2. The agarose needed to mix with the volume of TAE measured was weighted. The gel was used at 0.8 %, which means that for each 100 mL of TAE buffer, 0.8 g of agarose are needed.

3. After weighting the agarose, it was placed in the flask and then the measured TAE buffer was added. The solution was weighted in a scale.

4. The flask was placed in the microwave until the solution began to boil. After that, the solution was checked in order to ensure that the agarose was well dissolved.

If not, an extra time in the microwave might be needed.

5. The flask was then again weighted in the scale and sterilized water was added, in order to get the initial volume.

This was needed since, when boiling, some of the volume evaporated.

Water is normally added until the initial volume is surpassed, since the flask is still hot and the evaporating of the mixture continues for a little while.

6. Tray and combs needed were clean with water and dried.

7. The gel was cooled down until it was not hot enough to melt the comb. During that time, the mold was balanced and the tray placed in it.

8. The mold was closed and the comb needed for the number of samples was laid on the notches the mold has on its laterals.

The combs should be put in the tray in order to have the same space between them.

9. The mixture was added, checked to see if there were no bubbles and was let rest for 20 min, at least.

10. Once cold, the tray was taken from the mold and placed in the electrophoresis machine filled with TAE buffer until the gel was completely covered.

11. Molecular-weight size marker was prepared.

It is known that the wells can be filled with 20 µL, without any volume leaking. This solution needs water and loading dye, which is normally prepared at a concentration of 10x and needs

to be at a concentration of 1x to be used in this application. During this step, since this solution of the loading dye is thick, the tip of the micropipette should only touch the surface of the volume. Finally, the DNA from molecular weight size marker is added to the solution.

The rest of the volume was filled with sterilized water until the volume of 20 µL was reached, which should be added first to the mix, since its volume is higher.

12. The same volume of loading dye was added to the DNA samples.

As said before, the loading dye is thick and is used to facilitate filling the wells and to ensure that the samples don't leave them.

13. The samples and the marker were placed in the wells, carefully in order to not form bubbles or break the wells.

14. The lid of the electrophoresis machine was closed and the electrodes were confirmed to be correctly attached to it. The current was switched on (90 Volts) immediately for 1 h.

This was done with some urgency so that the samples don't start leaving from the wells.

The electrophoresis cannot be stopped until it's time to dye the gel, because if the gel stays in the electrophoresis machine after it was stopped, the bands start to spread through the gel, which makes them harder to see. If the hour has already passed but there is no time to dye it, the voltage can be decreased until then.

15. After about an hour, the tray with the gel was taken to the room where the ethidium bromide is used.

In this room, cautions as wearing gloves and making sure that nothing that touches anything placed in that room leaves it are needed. Therefore, the tray cannot touch any surface.

16. Another tray, where the gel was placed inside the room, was filled with ethidium bromide until it covered the gel.

Since this chemical is photo sensible, a weaker light should illuminate the room while using it.

17. The gel was left there for 10-15 min covered with aluminium paper.

18. After that amount of time, the gel was taken from the tray carefully and placed in the transilluminator. The ethidium bromide used returned to the bottle from where it came from.

19. The transilluminator was activated and the white light was switched on in order to place the gel in the centre of the screen.

If it moves, the rest of the ethidium bromide still there should be cleaned.

20. The zoom and the focus were adjusted.

21. The door from the transilluminator was closed, the white light shut down and the ultraviolet one was switched on.

22. A photo was taken and placed in a flash driver. As soon as the photo was taken, the ultraviolet was shut down to not degrade the DNA.

23. The gel was taken from the transilluminator and placed in the designated trash.

3. _Recovery of DNA from agarose gel_

Theoretical Foundation

This technique is done with the sole purpose of quickly retrieving, at high concentrations, a wanted DNA fragment that went through an electrophoresis in an agarose gel. Therefore, it aims to isolate specific DNA fragments from the rest of reaction products. It is suited to use in DNA ligation, PCR, sequencing, and other methods [83].

Most methods of retrieving a wanted DNA fragment that went to electrophoresis either fail to completely remove agarose, causing problems in downstream manipulations, or shear the DNA, which leads to very low yields [83].

This kit can be used in most of the DNA fragments, since it functions with a high range of sizes between 50 bp to 23 kb. However, for sizes between 50 bp to 10 kb, the yield of recovery is between 70-90 %, while for DNA with 11 kb to 23 kb, the recovery decreases to 50-70 % [83].

Material

➢ Agarose gel with the DNA fragment wanted;

➢ Agarose Dissolving Buffer (ADB) (provided by the kit);

➢ DNA Wash Buffer (provided by the kit);

Before starting, 24 mL of 100 % ethanol (or 26 mL of 95 % ethanol) should be added to the bottle of 6 mL of DNA Wash Buffer concentrate. 96 mL of 100 % ethanol (or 104 mL of 95 % ethanol) should also be added to the 24 mL one.

➢ DNA Elution Buffer (provided by the kit);

Done with Tris-HCl and EDTA (EDTA binds with Mg^{2+}, a crucial cofactor needed for the function of nucleases, which decrease the DNA degradation).

➢ Eppendorf tubes;

➢ Centrifuge tube rack;

➢ Zymo-Spin™ Column (provided by the kit);

➢ Collection Tube (provided by the kit);

➢ Micropipettes and tips;

➢ Scalpel:

➢ Heat block;

➢ Microcentrifuge;

➢ Sterile water.

Procedure

1. The DNA fragment was taken from the agarose gel using a scalpel and transferred to a 1.5 mL microcentrifuge tube.

The amount of agarose taken from the gel should be as small as possible.

2. 3 volumes of ADB were added to the agarose excised from the gel.

3. The sample was incubated at 37-55 ºC for 5-10 min, for the gel slice to completely dissolve.

Temperatures above 60 ºC should not be used.

To facilitate the dissolution of the gel, gentle mixing can be done during the incubation.

After this incubation, when the DNA fragments are bigger than 8 kb, one additional volume of water should be added to the mixture, in order to have a better DNA recovery.

4. The melted agarose solution was transferred to a Zymo-Spin™ Column already placed in a collection tube.

5. The sample was centrifuged for 30-60 s between 10000-16000 RCF. The flow-through was removed by aspiration.

6. 200 µL of DNA Wash Buffer were added to the column and it was centrifuged for 30 s in the same conditions.

7. The flow-through was discarded and the wash step repeated.

8. 6 µL of DNA Elution Buffer or of water were added directly to the column matrix.

9. The column was placed in a 1.5 mL tube and centrifuged for 30-60 s between 10000-16000 RCF to elute DNA.

Waiting 1 min prior to elution may improve the yield of DNA larger than 6 kb.

Since elution of DNA from the column is dependent on pH values and temperature, in cases of fragments bigger than 10 kb, eluting the DNA with 60-70 ºC DNA Elution Buffer might improve the results.

4. *Spectrophotometric measurements*

Theoretical Foundation

"NanoDrop" (Thermo Scientific) is the trademark of spectrophotometer that is capable of measuring the range of absorbance with a micro drop (1 µL of sample), which makes it ideal when the volume of the sample is not much. Even more so, since it also makes it possible to retrieve the sample drop after analysing it. For that to happen, it uses superficial tension to maintain the sample in its place, and that place (the pedestal) is hydrophobic, to prevent cross contamination and to help cleaning it.

It allows a fast, reliable and reproducible way of measuring the DNA, RNA and protein purity and has an algorithm that corrects values of sample concentration. It also contains an option to measure the integrity of the samples, by digitally processing images (**Fig. 16**).

Fig. 16. The equipment needed for "NanoDrop".

A. Concentration and purity of samples by spectrophotometry

This equipment can measure, without diluting, very concentrated samples. This characteristic makes it ideal to analyse nucleic acid and determinate its quality. This analysis is calculated by the software with the following proportion:

- 1 $A_{260\,nm}$ = 50 μg/mL of DNA = 40 μg/mL of RNA [84]

- 1 $A_{280\,nm}$ = 1 mg/mL of protein [85]

This way, the concentration of the DNA/RNA sample is calculated taking into account the absorbance value obtained at a wavelength of 260 nm.

On the other hand, the absorbance ratio A260/280 and A260/230 are used to evaluate the purity of the samples (**Table 2**).

The ratio A260/280 is very stable and considers that a pure DNA has a value of between 1.8-2.0. An acceptable DNA has values higher than 1.6, but when it's higher than 2.1 it might mean than there is an RNA contamination. Values below 1.6 also indicate a contamination, but by phenols and proteins. Therefore, the DNA concentration value given by this equipment cannot be fully trusted, since both RNA and proteins can also be absorbed (**Table 2**). To confirm it an electrophoresis gel can be done [86].

If the ratio used is A260/230, the values change, since at 230 nm components like phenols, chaotropic salts and carbohydrates are the ones that are measured. This explains why there is a high peak when a salt that constitutes certain buffers is present. Here, a pure DNA has values of between 2-2.2. Values lower than 1.5 would indicate that the sample is contaminated (**Table 2**). Therefore, to interpret correctly the results obtained, the composition of the resuspension buffer and the extraction protocol used must be considered. For that same reason, when comparing this ratio with the A260/280, it is known that this one is more inconstant [86].

Table 2. Purity values in DNA samples given by "NanoDrop" and its significance. Taken from [86].

Ratio	Value	Purity
A260/280	1.8-2	Pure DNA
	1.6-1.8	Acceptable DNA
	< 1.6	Aromatic compounds present
	> 2.1	Contamination with RNA
A260/230	2-2.2	Pure nucleic acids
	< 1.5	Contamination with phenol, salts and carbohydrates

Looking into the graph given by the equipment is also a good indicator of the purity of the sample. Normally, the DNA samples should look like the image below (**Fig. 17**).
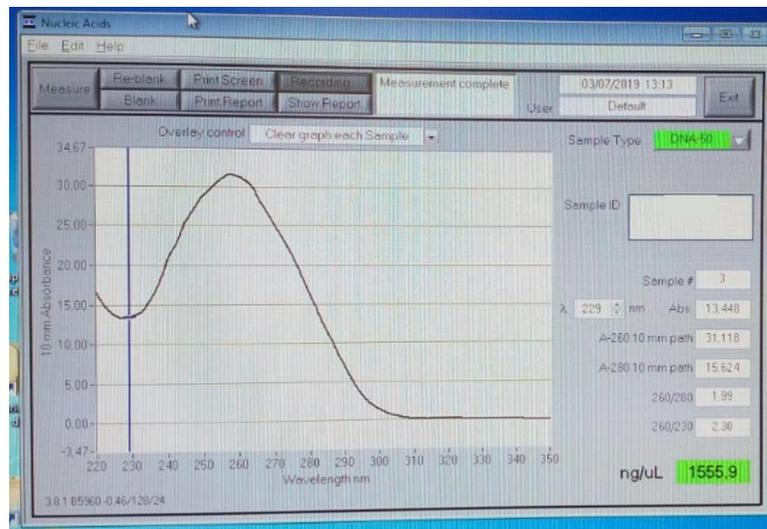
Fig. 17. Graph given by "NanoDrop" for a DNA sample.

If this image doesn't correspond to the one obtained, a contamination might be present.

Material

> ➢ Samples;

> ➢ Centrifuge tube rack;

> ➢ Specific paper to clean the equipment;

> ➢ Sterilized water;

➢ Solution to make the blank;

The buffer/medium where the samples were diluted should be used. If the medium has antibiotics, those must be added in the same concentration as they were in the medium.

➢ 20 µL micropipette and tips;

➢ Computer;

➢ "NanoDrop" equipment.

Procedure

1. The computer was switched on.

2. The plastic protecting the equipment was taken off, the lid was opened and the paper that is folded between the tip and the base of the equipment was also taken aside.

There is a paper folded between the lid and the base so that the place where the sample is placed is not dry and in constant contact with something.

3. The place where the sample will touch was cleaned, by putting sterilised water (2-3 µL are enough) in the base. The lid was closed carefully and immediately opened again.

4. Both the base and the lid were dried with a special paper for the "NanoDrop".

There is no need to do a lot of pressure with the paper.

5. The "ND-1000" desktop icon was selected and then the "NucleicAcid" programme was chosen.

6. The software needed another drop of water to be put in the equipment. This time, not to clean, but to initiate the programme. The procedure was the same as before (opening the lid, putting water, closing it and drying it).

7. The blank solution (2 µL of it) was placed, the lid was closed and the "Blank" option was selected. After doing this, the equipment was dried once again.

8. The samples (1.5 µL of each) were placed in the "NanoDrop".

Before analysing them, the sample type (DNA or RNA) was selected. Moreover, a name was written to be able to associate the result with a sample.

9. "Measure" was selected.

10. Between each sample, the equipment was dried (since there is no need to clean) and the sample name was changed.

11. After there were no more samples to analyse, a drop of water was placed and the lid was closed and opened. The equipment was dried one last time.

12. After this, the folded paper was placed in-between the lid and the base of the "NanoDrop" and it was covered by the plastic. The computer was shut down.

5. *Electrophoresis of RNA**

Theoretical Foundation

Being able to isolate pure and intact RNA is essential for many applications such as reverse transcriptase PCR, northern blotting, RNase protection assays, between others. Recently, another type of RNA, low-molecular-weight RNAs, appears to be promising, but the normally used protocols for the isolation of total RNA or mRNA are not yet optimized to those RNAs [87].

The Agilent 2100 Bioanalyzer determines the integrity and the concentration of total and small RNA, and also DNA, by enabling small-scale gel electrophoretic separation of nucleic acids on microfabricated chips. It presents many advantages that might help considering it as one of the best methods to analyse low-molecular-weight RNAs [88]. Moreover, it is suitable for next-generation sequencing, gene expression, gene editing, and other applications [89].

This kit comes with materials, such as a microchip for RNA or for DNA assays, that can analyse 12 samples at the same time (**Fig. 18**). It is also very useful when the sample doesn't have much volume available, since it only needs a small portion of it (1µL) to function. Moreover, it lasts long, since its stability remains the same for 4 months, which enables many analyses to be done.

Fig. 18. A Bioanalyzer chip for RNA electrophoresis.

This protocol was done by following the instructions that came with the "Agilent RNA 6000 Nano Kit", which can be used to analyse total RNA.

Why is capillary electrophoresis done here instead of the more often used gel electrophoresis? The reason for this rises from the fact that this one has more resolution, which enables for a better discrepancy between bands. This electrophoresis done by this system is called capillary and can also be done with DNA.

Material

All reagents, when not in use, should be kept at 4 ºC. All the materials must be RNA and RNase free.

➢ Sample;

Can only be used if it has more than 10 ng and less than 600 ng of RNA (seen in "NanoDrop"). If the sample has more than that, a dilution needs to be done.

➢ Ladder (provided by the kit);

Needs to be denatured for 2 min at 70 ºC and cooled down in ice. Then it needs to be split into vials with the required amount for a daily use and be stored at -80 ºC. After out of the freezer, it should be kept on ice until it is needed.

➢ Agilent RNA 6000 Nano gel matrix (provided by the kit);

➢ RNA 6000 Nano dye concentrate (provided by the kit);

This material should be handled carefully, because it is sensitive to light, and therefore it can decompose and the signal intensity can be reduced in its presence.

➢ RNA 6000 Nano marker (provided by the kit);

➢ Chip priming station (provided by the kit);

Its syringe needs to be replaced every time a new reagent kit is used.

➢ IKA vortex mixer (provided by the kit);

➢ 16-pin bayonet electrode cartridge (provided by the kit);

These chips have circuits inside of them.

➢ Chip for the electrode cleaner (provided by the kit);

➢ Chip for the water (provided by the kit);

➢ Bioanalyzer;

➢ RNase Zap;

➢ Water;

➢ Micropipettes and tips;

➢ 1.5 mL microcentrifuge tubes;

➢ 0.5 mL microfuge tubes (provided by the kit);

➢ Centrifuge tube rack;

➢ Microcentrifuge;

➢ Heat block;

To denature ladder/sample.

➢ Ice and container where to place it;

➢ "NanoDrop" equipment and all material needed for it (explained in its section).

Procedure

1. All reagents and samples were allowed to equilibrate to room temperature for 30 min before use. Of all the reagents needed, only the dye concentrate and the gel-dye mix (explained ahead) must be protected from light.

2. Samples were analysed in "NanoDrop" to ensure that they have concentrations of more than 10 ng and less than 600 ng of RNA. If the concentration surpasses that value, or it has a turbid aspect, a dilution is needed. In case of having a turbid sample, before diluting, doing a pulse in the centrifuge might be necessary to ensure that most of the particles that were giving the sample its aspect can precipitate. Then, only the supernatant is used for the dilution.

Preparing the gel

1. 550 µL of Agilent RNA 6000 Nano gel matrix were placed into the top receptacle of a spin filter.

2. The spin filter was put in a microcentrifuge and spun for 10 min at 1500 RCF.

3. 65 µL of filtered gel were aliquoted into a 0.5 mL RNase-free microfuge tube included in the kit. The aliquots were stored at 4 °C and used within one month after preparation.

Preparing the gel-dye mix

1. RNA 6000 Nano dye concentrate was mixed with the aid of a vortex for 10 s and agitated.

2. 1 µL of RNA 6000 Nano dye concentrate was added to a 65 µL aliquot of filtered gel already prepared.

3. The tubes were mixed thoroughly in the vortex.

4. The dye concentrate was stored at 4 °C in the dark.

5. Tubes were placed in the centrifuge for 10 min at room temperature at 13000 RCF.

This mix between gel-dye must be used within one day.

Loading the gel-dye mix

1. A new RNA Nano chip was placed in the chip priming station.

2. 9 µL of the gel-dye mix were pipetted to the bottom of the well with a black circle marked with a "G" (**Fig. 18**).

When dispensing, the tip of the pipette must be in the bottom of the chip to prevent a large air bubble forming under the gel-dye mix, which may lead to poor results.

3. The plunger of the syringe of the chip priming station was positioned at 1 mL and the chip priming station was closed.

To ensure that the priming station is properly closed, the lock of the latch must click.

4. The plunger of the syringe was pressed until it was held by the clip.

5. The plunger was freed after exactly 30 s.

6. After the plunger moved back at least to the 0.3 mL mark, 5 s were counted and, after that amount of time, the plunger was placed in the initial 1 mL position.

7. The chip priming station was opened and 9 µL of the gel-dye were mixed in the 1st and 2nd wells of the 4th column, identified by a "G", but not in a black circle (where it was already added) (**Fig. 18**).

Loading the RNA 6000 Nano Marker

1. 5 µL of the RNA 6000 Nano marker were pipetted into the well marked with the ladder symbol and then to each of the 12 wells designed for the samples.

If there are not 12 samples, the wells that are not needed shouldn't be left empty, since the chip may not run properly. Therefore, those should be filled with 5 µL of the RNA 6000 Nano marker and 1 µL of the buffer.

Loading the ladder and samples

1. Before use, ladder aliquots were unfrozen and kept on ice.

Done to avoid extensive warming upon the thawing process.

2. The samples were denatured at 70 °C for 2 min.

This step was crucial to minimize secondary structures.

3. 1 µL of the RNA ladder was placed into the well-marked with the ladder symbol.

4. 1 µL of each sample was pipetted into each of the corresponding 12 sample wells.

5. The chip was placed horizontally in the adapter of the IKA vortex mixer.

Here, there must be some caution to not damage the buldge that fixes the chip during vortexing.

6. The chip was mixed by the vortex for 60 s at 2400 rpm.

The chip must be run within 5 min after everything is placed in it, since the reagents might evaporate, leading to poor results.

Inserting a Chip in the Agilent 2100 Bioanalyzer

1. One of the wells of the electrode cleaner was slowly filled with 350 µL RNase-free Zap.

There is no need to place the Zap in every well, since in these chips, the solution put in one well will diffuse to the others.

2. The electrode cleaner was placed in the Agilent 2100 bioanalyzer and it was left there for about 10 s.

3. After the electrode cleaner was removed, another 10 s were given to allow the Zap on the electrodes to evaporate.

4. The same thing (steps 1-3 in this section) was done with RNase-free water, in its specific chip.

5. The chip with samples was carefully placed into the receptacle.

There is only one way that the chip fits. This step should be done in darkness because the dye concentrate is sensitive to light.

6. The lid was carefully closed.

If too much force is done when the lid is closed, the electrodes may become damaged, and dropping the lid may cause the liquid to spill, which originates bad results. Therefore, closing the lid should be done carefully.

Starting the Chip Run

1. The appropriate assay from the Assay menu was selected (in this case was "RNA nano prokaryotic").

2. Information like samples' name and comments were written.

3. The Start button was clicked to start the chip run.

During the run, people shouldn't be around the equipment nor any tasks should be done near it, since the Bioanalyzer is very sensible to vibrations and the results can be affected.

4. After the chip run was finished, the chip was removed from the receptacle of the bioanalyzer and disposed of.

A chip shouldn't stay in the bioanalyzer more than 1 hour, since the electrodes may become contaminated.

Cleaning up after a run

1. The electrode cleaner was placed in the Agilent 2100 bioanalyzer and it was left there for about 10 s.

2. After the electrode cleaner was removed, another 10 s were given to allow the water on the electrodes to evaporate.

6. *Polymerase chain reaction (PCR)**

Theoretical Foundation

Polymerase chain reaction (PCR) consists in a technique that amplifies a specific DNA sample *in vitro,* originating millions of copies of the fragment being studied. This happens because PCR is composed by, basically, three steps that are repeated. This way, after every cycle, the number of DNA molecules doubles (**Fig. 19**) [72].
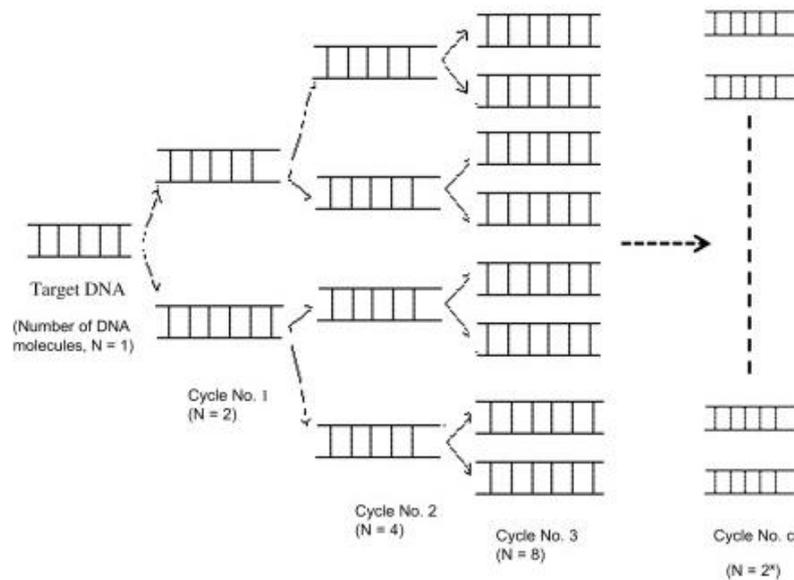
Fig. 19. Exponential amplification in the PCR technique. Taken from [90].

This procedure is done automatically in a device named thermocycler that repeats various times the following steps (**Fig. 20**):

1. **Denaturation:** The denaturation of the DNA template is done by increasing the temperate to 94 ºC during between 30 s and 2 min [72].

2. **Annealing:** In this step, the temperature lowers to between 40-60 ºC during between 30 s and 2 min for the primers to hybridize with their targets sequence. The temperature during this step depends on many factors, such as *primers* sequence, their specificity, and others [72].

To confirm which temperature will lead to a more efficient hybridization, a thermocycler able to applicate different temperatures to each column in the plate is used. Moreover, since every column can have many tubes, other parameters (such as magnesium concentration) can be optimized. The conclusion is done by considering the results from electrophoresis.

3. **Extension:** The temperature here is around 72 ºC for the complementary nucleotides to be added. The duration of this step varies depending on the fragment size and the extension speed of the enzyme. If done for too long, the enzyme might begin to amplify areas that are not wanted [72].

The first denaturation and the last extension should take longer than the ones done in the other cycles, to ensure that the chains denature completely right at the beginning of the process and that, at the end, the enzyme has time to fully complete the chains. The number of cycles depends on the initial amount of targets and complexity of the sequence [72].

Time and temperature used in PCR changes depending on: G-C content; length of the fragment that we want to amplify; and speed of processing by DNA polymerase (DNApol) [72].
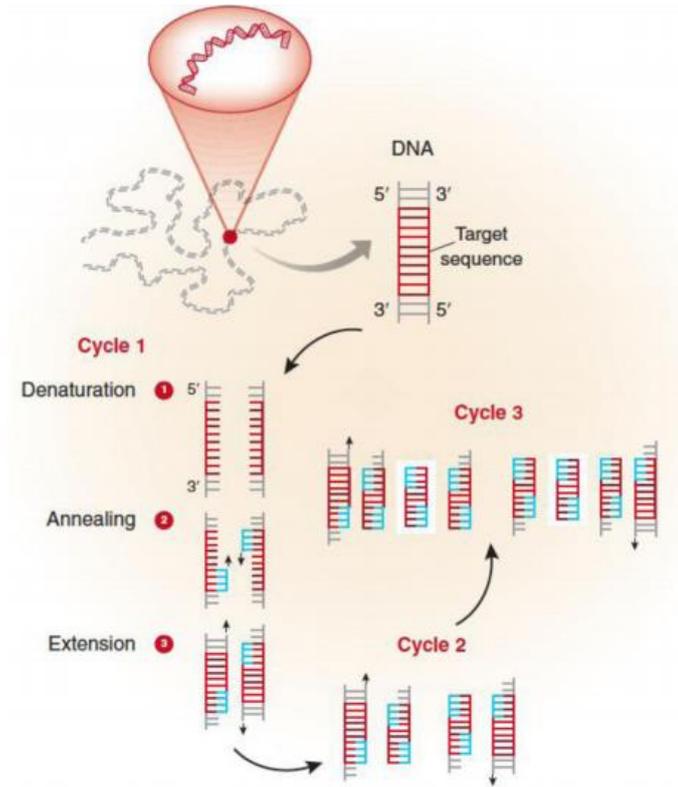
Fig. 20. Representation of the three-step principle followed by PCR. Taken from [91].

In this study, PCR was done to ensure that the recombination in the exconjugants was done correctly. Therefore, two studies where done on PCR: (1) with primers that would amplify the area where the *Streptomyces* would bind with the plasmid (and, therefore, should not exist, if the conjugation was successful); (2) with primers that amplify an area of the plasmid already incorporated in the *Streptomyces'* genome (only will amplify the samples that were conjugated correctly) **(Fig. 21)**.
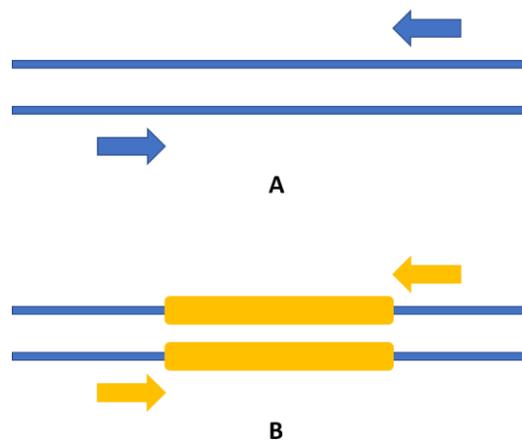
Fig. 21. Representation of the two studies done in PCR. (A) Primers that amplify the area where the *Streptomyces* would bind with the plasmid; (B) Primers that amplify the area of the plasmid already incorporated in the genome. The arrows represent the primers. The blue lines represent the *Streptomyces'* genome, while the yellow ones represent the plasmid.

Material

All materials needed for this procedure should be handled carefully with gloves, in order to minimize the risk of any bacterium or genetic material that might be present in our skin to contaminate our sample.

> ➢ DNA template;

The template can come from purified DNA or from cells that have the DNA that will be studied. If that last one is chosen, the cells need to be boiled before using them, in order to break the cell wall and get the genetic material out in the medium.

Here, the templates used were inoculations of conjugated *Streptomyces*.

> ➢ Primers;

A primer is a short sequence of nucleotides that provides a starting point for DNA synthesis. These are normally designed with previous knowledge of the DNA sequence that needs to be amplified, and by having into account that they shouldn't hybridize with any other genome zone that is not the one wanted. They can't also hybridize between themselves nor can they originate secondary structures [72].

In *Streptomyces,* which has high content in cytosines and guanines, interesting areas with a high proportion of thymine and adenine are searched, since these nucleotides give specificity to the sequence. This characteristic makes doing PCR on these microorganisms more difficult, since it is difficult to design primers in the desired regions due to the scarcity of thymine and adenine nucleotides in the sequence [72].

Two pairs of primers, as said earlier, were used: (1) that will amplify area where the *Streptomyces* would bind with the plasmid (in the "attP" site in **Fig. 11**); (2) that will amplify the plasmid itself inserted in the genome.

> ➢ Deoxyribonucleotide triphosphate (dNTPs);

These are needed since DNApol needs them to synthesize DNA. Normally these are added in equimolar concentrations, to assure that all components have the same number of moles (amount of substance) [72]. But, since *Streptomyces* has a high content of cytosine and guanine, these nucleotides are often added in a higher concentration.

These dNTPs should be added carefully, because if too many of them are provided the polymerase can commit more mistakes [72].

> ➢ DNApol;

This component is needed to synthesize DNA. There are many types of polymerase, each with a different length limit and range of sizes that they can amplify. The Taq polymerase, for instance, is able to amplify, specifically, a single copy of a gene. However, it should only be used when the amplifications that will be done don't need a very high specificity, since it doesn't have a 3'-5' proofreading activity needed to correct the sequence [72]. This way, Taq polymerase can be used, for example, when the sequence doesn't matter much, but a single fragment wants to be seen in electrophoresis. Therefore, the best enzyme for each sample needs to be chosen.

> ➢ Buffer for the enzyme;

The buffer is specific for each DNApol and, for some brands, it might have already the dye for the electrophoresis, or the magnesium chloride, added.

➢ Magnesium chloride;

The concentration of magnesium can greatly affect the PCR final result, since a higher concentration can decrease the specificity of the polymerase and, therefore, regions not wanted can be amplified [72].

➢ Dimethyl sulfoxide (DMSO);

As said before, the DNA that will be amplified has many cytosines and guanines, which can hinder the separation of the DNA chains or enhance the probability of the primers originating secondary structures. All of these hinder the hybridization between template and primers and affect the activity of the polymerase. Therefore, DMSO is useful here because it reduces the quantity of energy needed to separate and hybridize the DNA chains.

If not for DMSO, the temperature used would have to be higher.

➢ Ultrapure water specific for PCR;

Used both for the master mix and as a negative control.

➢ Plasmid used in the transformation;

Used as a positive control only for the primers that would recognize the plasmid already bound to the *Streptomyces'* genome.

➢ Wild-type (WT) strain of *Streptomyces*;

Used as a positive control only for the primers that would bind to the part of the *Streptomyces'* genome that should have bound with the plasmid.

➢ Mix of plasmid and WT strain of *Streptomyces*;

Used as a positive control for both pairs of primers used.

➢ Eppendorfs tubes for PCR;

➢ Micropipettes and tips with filter;

➢ Centrifuge tube rack;

➢ Vortex;

➢ Small petri dishes;

➢ SM medium;

➢ Samples;

➢ Ice and a container to put it;

➢ Electrophoresis gel and equipment (explained in its section);

➢ Thermocycler.

Procedure

Every solution needed, except the DMSO, should always be placed in ice.

1. All solutions were placed in the vortex and mixed well.

Especially the one with magnesium chloride, because it agglomerates easily.

2. To make a master mix, all components needed were added, except for DNA, DNApol and primers.

As usual, the order of components mixed will depend on their volume and, therefore, the first will always be the one with higher volume. This rule was followed, except for the polymerase and primers, which were the last ones added, to ensure that there weren't unwanted amplifications.

When components are added, tips for micropipettes should have filters, since even the tip and the micropipette themselves are sources of contamination and might contain DNA and DNase. By doing this, the risk is lowered.

3. The mix was split into the individual reaction microtubes and the boiled *Streptomyces* was added to each of the tubes.

Controls are also needed. Three positives were used: (1) with the plasmid (should be positive to one of the primers, the one that amplifies the area where the plasmid bound with *Streptomyces*); (2) with the WT, not mutated (should be positive for the other pair of primers); (3) with both (positive for both pairs of primers). The negative control was made with water.

Every tube should be identified carefully to know which sample and which pair of primers each of the tube has.

4. Once all components were mixed inside the tubes, a short centrifugation was done.

5. To avoid water condensation in the tubes during PCR, the thermocycler was turned on before the tubes were ready to be placed in it.

This is needed because the equipment warms the part were the tubes will be placed before the lid, which takes longer to reach the wanted temperature. If the tubes are placed before the lid has the adequate temperature (5 ºC higher than the denaturation temperature), the content in the tubes would evaporate.

6. The program of the equipment with the cycles correctly programmed was switched on.

7. After the cycles ended, the samples were prepared, as seen in the electrophoresis section, and went through that process.

8. The exconjugants with the correct pattern were plated individually in small petri dishes with SM medium, in order for them to get spores or mycelium to store and to work with.

Methods in Bioinformatics

In this part of the work, different samples were analysed: "Selectos242_28", which are the 28 candidates of sRNA that have already been studied experimentally and, therefore, already have the constructs needed done; and "No28", which are the rest of the candidates that have not been experimentally studied yet.

1. _Annotation_

Annotating genes, or sRNAs in this case, is extremely important since, in this process, they are named and their characteristics are written. If not for this, analysing them would be an arduous task, because there wouldn't exist a concise strategy to understand which gene/sRNA was being referred to.

The first thing to do in order to annotate, was to be sure that the candidate is indeed a sRNA. To consider a candidate a sRNA, there were a few conditions that needed to be fulfilled. The first one was that a transcription start site (IT) had to exist. The second condition was to have read coverage (which is the sum of the depth of reads for each genomic position, that represents the number of transcripts, sequencing reads, obtained from a RNA-seq experiment) and the third was to show a transcriptional profile (detected by RNA-seq and microarray experiments). The first two parameters were analysed in Integrative Genomics Viewer (IGV). There, a higher coverage means that that gene/sRNA is being highly expressed (explained a bit more ahead). Finally, more than having a profile, it must be different than adjacent ones.

The manual inspection of the data visualized on IGV was summarized in a numerical number ranging from 0 (definitely not sRNA) to 1 (is without a doubt one), and values between the two (0.5, 0.55, 0.6, 0.7, 0.8 and 0.9), given depending on the characteristics of the candidate (**Table 3**).

Table 3. Values given to the candidates depending on their characteristics.

Value	**Characteristics of the candidate**
_0_	_Definitely not a sRNA_
_0.5_	_It is a sRNA, but there is no clear ending_
_0.55_	_It is a sRNA, but there is no clear ending nor it has a distinguish profile_
_0.6_	_It is a sRNA, but there is no clear ending nor is it expressed much_
_0.7_	_It is a sRNA, but it seems to have no ending_
_0.8_	_It is a sRNA, but there is no clear ending, no distinguish profile, nor is it expressed much_
_0.9_	_Can be a sRNA, but there is no data that proves it_
_1_	_Definitely a sRNA_

In order to annotate, some parameters were needed, such as:

a) *Profile of sRNA*

Microarrays and RNA sequencing (RNA-seq) were done to know how the expression of the sRNA would change depending on the phosphate concentration. Those results were turned into graphs to show the transcription values across the time series, also called transcriptional profiles (**Fig. 22**).

Two strains of *S. coelicolor* were used to study if there were sRNA genes regulated by a phosphate-dependent mechanism: *S. coelicolor* M145, the model strain; and *S. coelicolor* INB201, a deletion mutant of the *phoP* regulator (the first strain is the parental of this mutant). Cultivation of these strains were carried out in flasks using defined medium MG3.2 [92]. In this medium, phosphate was limiting the growth of the cells and was added as potassium salt. Culture samples were taken from the medium to monitor the depletion of phosphate from the medium. In the same cultures where the cells experiment phosphate limitation at 37 h of culture, potassium phosphate was added at 42 h to observe the effect of the excess of phosphate on the regulatory responses triggered by the phosphate limitation. Genes under upregulation by the PhoR-PhoP system would be higher transcribed in the WT strain, as shown previously [93]. Therefore, the WT strain was used as control, to elucidate if the differential expression between both strains was due to PhoP control (**Fig. 22**).

The graphs used to inspect the transcriptional profiles of sRNA candidates depict two types of transcriptional values: the M_g and the Transcripts Per Million (TPM). Both quantify the abundance of transcripts, but while M_g values were obtained by microarrays analysis, TPM values were originated by RNA-seq. For the representation of both M_g and TPM values in the same graph, TPM were logarithmically transformed, to match the logarithmic nature that M_g values have (**Fig. 22**).

In those graphs, there was also the plot of the M_g values from a WT culture without addition of phosphate, to reflect the normal behaviour of the genes in an unsupplemented batch culture (**Fig. 22**).
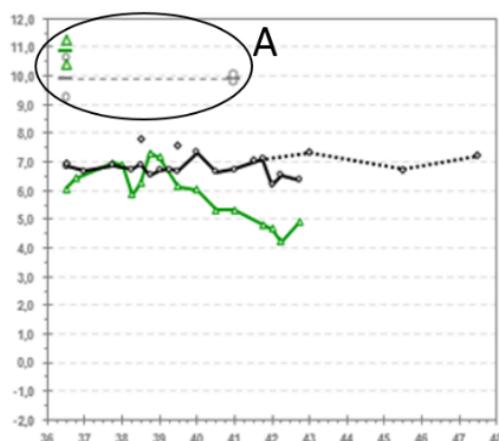
Fig. 22. An example of a profile for a sRNA. Two types of data can be seen: the one obtained from microarrays, the M_g values, and the other obtained by RNA-seq, the TPM values. The last ones are represented by the lines and points that are inside the circle A. The rest of the graph refers to M_g values. Both black lines are related to the WT strain but, while the discontinuous one represents the gene expression in this strain in an undisturbed culture, the continuous one depicts gene expression in a culture that phosphate salts were added at 42 h for a final concentration of 11 mM. On the other hand, the green line represents the same values as the black continuous one but to the mutant strain. Finally, the grey discontinuous line enclosed by the circle, that represents the TPM average expression in the WT strain, calculated from the values of two replicates (open circles); in addition, the values of two replicates at 36.5 h, obtained from the mutant strain cultures, are also plotted (green triangles).

The profiles were written in the following code:

$$W_1M_1\text{-}x|W_2D_2,$$

"W" represented the WT strain, "M" the mutant one, the numbers represented the hours and "x" represented the delay between the two strains (that might not exist). On the other hand, "1" represented the expression changes when phosphate became depleted, while "2" represented the expression changes after it was added. "W" and "M" were substituted by one of the following three characters, depending on the graph: "A" (if the values rose), "0" (if they stayed constant) and "D" (if they descended). "1 or 2" was placed instead of the "x" to indicate subtypes of profiles with varying kinetics of regulation.

Some of the most usual seen profiles are represented in the following image **(Fig. 23)**.
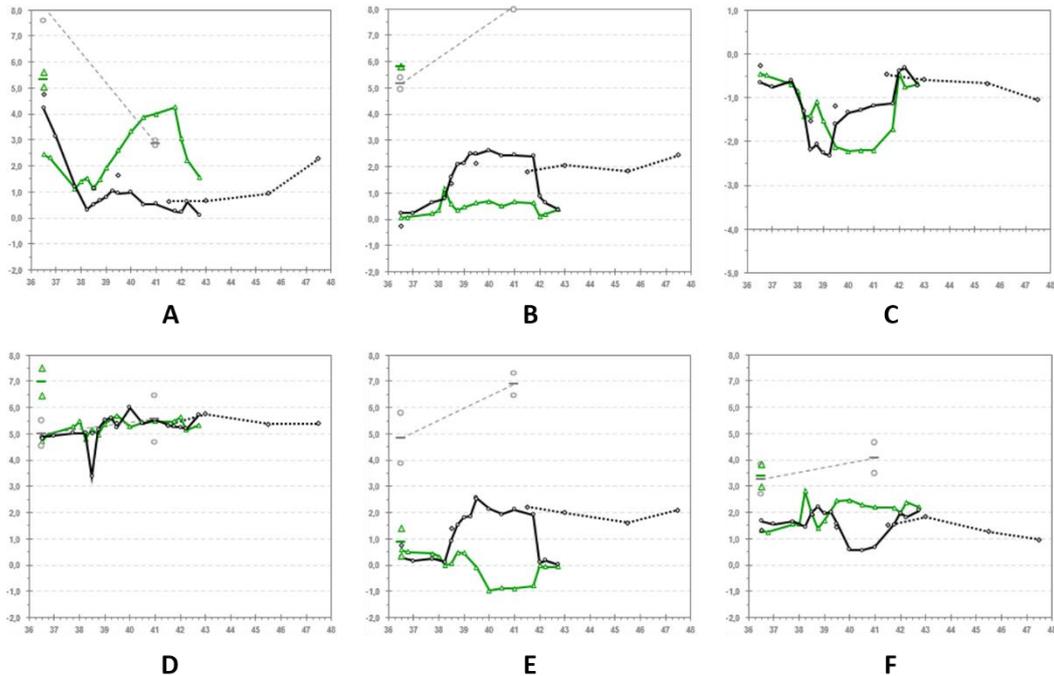
Fig. 23. The most usual seen profiles. **(A)** Profile 0A-0|0D; **(B)** Profile A0-0|D0; **(C)** Profile DD-1|AA; **(D)** Profile 0|0; **(E)** Profile AD-0|DA; **(F)** Profile D0-0|A0.

b) *Length of sRNA*

This one was analysed by using IGV, that, when coordinates for the gene of interest were given, showed the coverage of the genes, or of the sRNAs, if they existed in those coordinates.

There are several coverages in IGV, such as "ScoR5 RIBO" (that represents the relationship between ribosomes and the sRNA and, therefore, the translation, and comes from the data of Jeong et al., 2016), "ScoMG" and "ScpMG" (which represent RNA-seq coverage of the same experiment that produced the TPM values plotted on profile graphs for the WT and mutant strains, respectively), the "ScoMG_Mix" (which comes from another RNA-seq analysis using a mix of RNA extracted from different time points) [94]. This last one was made in order to have a wide representation of transcripts (**Fig. 24**).

There are numbers after "ScoMg" and "ScpMg" that represent how many hours the sample was cultivated (for instance, "ScoMg 36 h"). There are also two more coverages ("ScoR5+" and "ScoVar") which come from another study. While "ScoR5+" is the coverage from the

sample in medium R5, "ScoVar" is the coverage that comes from a mix of RNA preparations (mix of mediums), which makes it unreliable (**Fig. 24**).

There is another parameter to analyse in IGV and that is the probes, that are in a line named ("Scoe_SondasMIMA") (**Fig. 24**). These microarray probes were designed by knowing, through RNA-seq, where the transcribed regions were, in order to be able to bind these probes with those regions. These are particularly important in cases where the end of transcription is not easy to decipher.

Finally, there are transcription end sites or terminators (named as "Ter" in IGV), that were marked by probes as well, which should define where the transcription ends (**Fig. 24**). But that is often not the case. Normally in IGV, the end of the candidate is when the maximum value of coverage lowers more than 90 % (for instance, if the maximum of the coverage is 3, the end of the sRNA is when the coverage has values lower than 0.3).

But in order to have a length, as much as the end, the beginning is also needed. Therefore, the IT, one of the three requirements for considering a candidate a sRNA, was required. Even though, IGV normally says where the sRNA is ("Prediction of the sRNA" in **Fig. 24**), but that representation is not to be trusted. It is best to decide where the beginning of the sRNA is, by seeing where the IT is, which was also marked by probes.

There are different types of IT, all seen in IGV, and IT11 is considered to be more reliable than IT13. Normally, the IT matches the rise of the coverage and represents the beginning of the transcription. If both didn't match, using the coordinate where the coverage increases was more reliable and the sRNA was considered to have many beginnings.
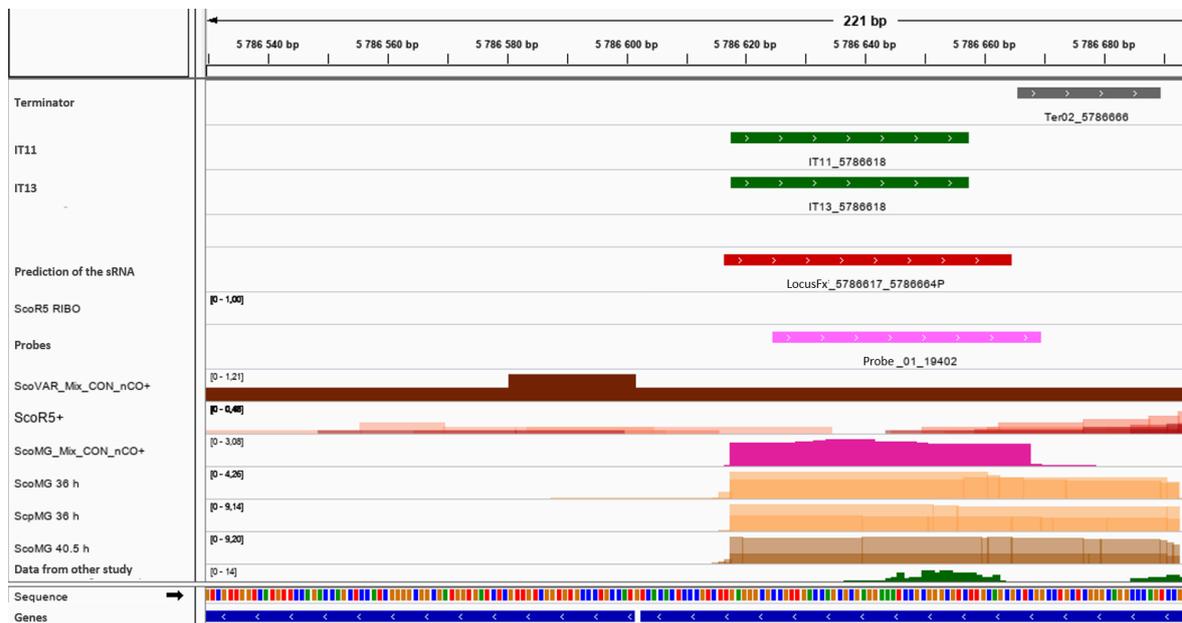
Fig. 24. Template of IGV. Here, it can be seen where the terminators, ITs, types of coverage, genes and the sRNA predicted are. By looking at this image, it is evident how clear the increases of the number of transcripts (coverage) is in ScoMG 36h, ScpMG 36h and ScoMG 40.5h.

c) *Types of sRNA*

Normally sRNAs will bind with the 5' end of the mRNA. But that doesn't always happen.

If the sRNA was inside a gene, it could be classified as intragenic/intern (if it was in the same chain) or antisense (if it was in the opposite chain). It could be antisense to a gene or to another sRNA, and if it was both it's only annotated as an antisense to a gene. An intragenic

sRNA could be classified as an intern (if the sRNA was in the middle of the gene), intern5 (if it was located in the 5' end of the gene) or intern3 (if it was the 3' end) (**Fig. 25**).

It could also be in-between genes and, if that was the case, it was called intergenic. A caution was needed when this type of sRNA appeared, because there was a chance it might just be a beginning of the nearest gene, and not a sRNA at all.
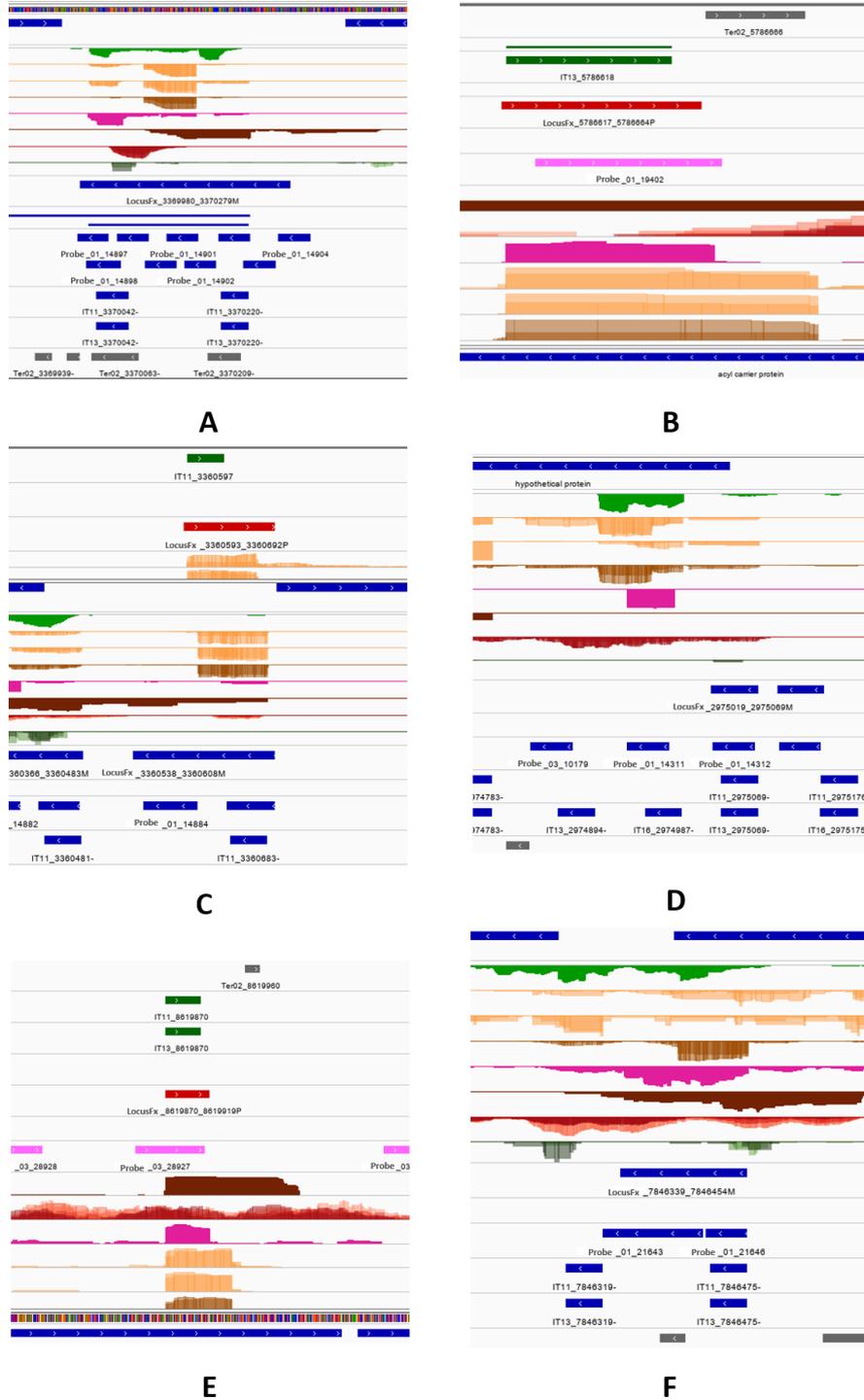
Fig. 25. The most usual seen types of sRNAs. **(A)** Intergenic; **(B)** Antisense to a gene; **(C)** Antisense to a sRNA; **(D)** Intern5; **(E)** Intern; **(F)** Intern3.

d) *Naming the sRNAs*

If there is even a small probability of the candidate being one sRNA (which means every value but 0), it should be named. In order to do that, a format was used to unify, as much as possible, the name given to all the sRNAs candidates. The format that will be used here will be: "sRNA.xxxx". If the sRNA overlapped with a gene (it didn't matter if it was antisense or not), the "xxxx" was substituted by the name of the gene (that normally was called "*scoxxxx*"). If they didn't overlap, the gene used for the sRNA name would be the one that first appeared on its right, independently of the chain it belonged to.

e) *If not a sRNA, then what is it?*

When the candidate was not, with certainty, a sRNA, it could be classified into other classes. One of them was "Fused", which happens when the candidate is fused with another sRNA. Another was named "ITgenGB", which happens when the candidate is not a sRNA but the beginning of a gene.

On the other hand, when a candidate didn't have an IT, the possibility of being a sRNA was discarded and the chance that it might just be "RestGenGB", which is, in other worlds, the rest of the gene, arose.

2. Prioritization

After all the sRNAs candidates have been annotated, only clear sRNAs (value 1) went through a selection, a prioritization, to understand which of them should be further analysed.

This prioritization was done considering, mainly, three aspects of the candidate, which were: expression; profile; and target.

Expression was prioritized using coverage, M_g and TPM values. If these three didn't agree between them (which means that two gave high values and the other low ones, or vice versa), the TPM values were the ones used to make a decision. Therefore, values of 1 to 3 in TPM (that correspond to values of 2 to 8 in M_g) were considered low, and that candidate was classified as not being expressed. A value of 3 in TPM could be considered enough, if the profile of the same candidate was good. A candidate with a value higher than 3 is considered as having a considerable expression.

It was decided that only "AA", "A0" profiles would be considered "interesting", since the first represents a PhoP independent regulation of phosphate and the second a dependent one. "0A" profiles were also prioritized.

Finally, the targets for each candidate were investigated. An interesting target was considered to be the one related either to the secondary metabolism, the primary (only if its function was known) or the central one. The primary metabolism might be important because there are substrates from it that can affect the secondary metabolism. On the other hand, the central one is responsible for the production routes of precursors crucial for glucoses, oxidative phosphorylation, nucleotide production, between other, some of which can also be precursors of secondary metabolites.

In order to know the target, the type of sRNA was crucial. If the sRNA was antisense, the only thing needed to do was to check if the antisense gene is or not related to those processes. That was done with a database that already had the genes classified as "regulator" and "genes belonging to biosynthesis clusters" by articles that concluded that, when these genes were activated/repressed, changes in the secondary metabolism occurred. If the gene had one of these functions, its antisense sRNA could be assumed as having the same function. However, this relationship between gene and sRNA could only be done with

antisense sRNAs. To all the other types, more research on their targets was needed to make such a conclusion.

For instance, to reach a conclusion on sRNAs antisense to other sRNAs (and not to a gene) or on internal sRNAs, only the profile was revised, because there was no other way to get to know the target. If the profile was interesting, as explained before, these would be considered related to secondary metabolism.

The only ones left were the intergenic sRNAs. Those had another different process to search their targets (explained ahead).

With this is mind, values of "1", "2", "10", "20", "30", "50" and "100" were given to prioritize the candidates, with "1" being the most interesting candidates and "100" the ones that were completely uninteresting.

"100" was given to those candidates that seemed to have random profiles, full of artefacts and were neither being expressed nor were they related to secondary metabolism.

To those candidates that had an uninteresting profile, weren't expressed and weren't related to secondary metabolism, a "50" was given. The difference between giving a "100" and a "50" is very little so, normally, only the value of "50" was used.

On the other hand, a "30" was given to those candidates that either had an interesting profile, a good expression or were related to secondary metabolism.

Those candidates that had not one, but two of these characteristics were given a "20".

Lastly, a "10" was distributed to those candidates that complied with all three characteristics. The best two values ("1" and "2") were reserved for a later phase of the analysis.

Only those with prioritization of 20 (and were intergenic) and 10 went through the next phase of analysis. Moreover, the group of "Selectos242_28", independently of the value of prioritization, were also studied since they had already been revised and studied until some extent.

3. _Searching for targets_

In order to continue prioritizing these candidates (the ones with value 10, the intergenic with value 20, and all of the 28 candidates already somewhat analysed), their targets had to be known, to see to which extent they were related to secondary metabolism or not. As explained before, the antisense sRNAs didn't need to go through this search.

The issue with predicting targets is that a sRNA doesn't have only one target, since the pairing is not continuous and it's also not perfect. To facilitate this challenge, bioinformatic tools such as "GLASSgo" (GLobal Automated sRNA Search go) and "CopraRNA" were used.

"CopraRNA" is a software that computes whole genome predictions to foresee the targets for each sRNA candidate, while "GLASSgo" finds sRNA homologues by combining pairwise identity filtering, iterative Basic Local Alignment Search Tool ("BLAST") searches and structure based clustering. Another option, instead of "GLASSgo", would be using BLAST, however, while this last one only looks for similarities in the sequence, "GLASSgo" also looks for similarities in the secondary structure. That's why "GLASSgo" was chosen [95], [96], [97], [98].

In order to use both "CopraRNA" and "GLASSgo", every sequence must be turned into FASTA format, which was not more that writing the NCBI Reference Sequence (RefSeq) of the organism to which the sequence belonged to (following the symbol ">"). RefSeq provides

information for many organisms, needed for various studies and is found in the website of National Center for Biotechnology Information (NCBI) [99]. After the RefSeq, the GenBack accession number was placed, which is a unique identifier shared by three databases [European Nucleotide Archive (ENA), DNA Data Bank of Japan (DDBJ) and GenBank] that doesn't change even if the annotation or sequence of the record is altered. After that, another line with the sequence itself is was needed and there could be no spaces between the two lines (**Fig. 26**).

```
>NZ_LT629768
GCGCCTGCCGAGTATGTGACGTACGGGTACGTGACATCGTCGTCGGGCCGCACGCATGAGGATGCCCGGTCGGTGGCCCGATCGGC
>NZ_CP016795
GCGCCTGCCGAGTATGTGACGTACGGGTACGTGACATCGCCGTCGGGCCGCACGCATGAGGATGCCCGGTCGGTGGCCCGATCGGC
>NZ_HE971709
GCGCCTGCTGAGTACGTCGACGTAATTCGACGGTGTCTCGACGGGCCGCACGTATGAGGATGCCCGGTCGGTGGCCCGATCGGC
```

Fig. 26. Example of three sequences in the FASTA format.

The sequence of the sRNA candidate, needed for FASTA, was obtained in the NCBI page of the whole genome of the microorganism studied, by changing the region shown with the ending and the beginning of the sRNA seen in IGV. In the same website, the FASTA format of the same sequence could also be asked. If the sRNA belonged to the complementary chain of the one being studied, an additional step was needed in order to have the complement reverse sequence, which could be done in various websites.

When the sequence of the candidate sRNA was obtained, "GLASSgo" could be done. In this software, the only thing that needed to be put was the sequence in FASTA format and the taxon selection must be changed, in order to coincide better with the studied organism. In *Streptomyces*, the taxon selection that was closest to it was the "Bacteria" one. One disadvantage that this software has is the fact that if the sequence is bigger than 800 nucleotides, "GLASSgo" is unable to analyse it [95], [96]. This shouldn't be a worry, because sRNAs are smaller than that.

After "GLASSgo" gave the results, the homologues needed to be analysed in order to decide which ones would go through "CopraRNA". In order to do that, a few conditions were drawn. For example, the homologues needed to have a valid RefSeq, so that "CopraRNA" could recognize it (which meant that the RefSeq either had to start with "NZ" or "NC"). They also had to belong to either *Streptomyces* or *Mycobacterium*, and never to pathogenic strains (such as *Streptomyces scabiei).* They also needed to have a complete genome and belong to different organisms. Finally, homologues with 100 % of valuation (which evaluates how close the homologue sequence is to the candidate's) should not be used (this normally happened if the homologue belonged to *Streptomyces lividans*). All of these conditions, except for the valid RefSeq, came written in the "GLASSgo" results. The RefSeq can be looked for in the NCBI webpage, by searching the gene identification. Gene identification (or gene "ID") is a stable identification of a particular locus in an organism, which doesn't change even when the locus suffers alterations [100]. By searching the gene ID in NCBI, information, from the literature and from other databases, will be given. In the middle of all that information, the RefSeq was also found.

Therefore, only homologues that obeyed these parameters were chosen. Since "CopraRNA" can only analyse between 3 and 5 sequences at a time (and one of them is the candidate's sequence), there might be more than that number of homologues that comply with the rules. This way, a selection was needed. Normally, they were chosen randomly but, if there were many of them, more than one search could be done in "CopraRNA" with different homologues, to see if there was any difference in "CopraRNA" results with different homologues. What was often done was a search with the homologues with the best valuation and another with some of the worst ones.

If "GLASSgo" didn't give the amount of homologues needed for "CopraRNA" (which is 3, and one of them would be the candidate's) or gave a reasonable amount but there weren't 2 of

them that could be used (because they didn't comply with the conditions written earlier), another search in the software had to be done, by changing some of the parameters: parameter setup had to be turned to manually; and the minimum allowed identity decreased to 40 %. If even then, the homologues were not enough to go through "CopraRNA", a final attempt was done by decreasing the minimum allowed identity to 30 %. If, even then, there weren´t homologues, the candidate must go through "IntaRNA" (another tool in the same website as "CopraRNA").

"IntaRNA" is a fast and accurate software that predicts interactions between two RNA molecules and the target sites for non-coding RNAs. This is especially helpful when "GLASSgo" doesn't find the number of homologues needed so that they can be analysed by "CopraRNA". "IntaRNA" doesn't need 2 homologues to compare them to the candidate sequence in order to achieve a result. This software only needs two sequences (one is the query and the other is the target one). However, in this project, the target RNA wasn't known so its sequence couldn't be provided to "IntaRNA". In cases like these, this tool is able to only work with the query sequence and the Target NCBI RefSeq Identification (the one from the organism being studied), in order to be able to get target RNA sequences from prokaryotic NCBI reference genome. All the other parameters present in "IntaRNA" were maintained as they were. However, the same as "GLASSgo", it also has one disadvantage: "IntaRNA" is only able to analyse sequences with less than 750 nucleotides [96], [98], [101], [102].

On the other hand, when the amount of homologues is enough, "CopraRNA" can and should be done. The difference between "CopraRNA" and "IntaRNA", is that "CopraRNA" uses "IntaRNA" to individually analyse the sequences and then compares them, looking for similarities. Because of this, "IntaRNA" is considered less reliable than "CopraRNA". For this last software, it was only needed to place the sequences (of the homologues and the candidate sRNA) in FASTA format, select the organism studied and running the software. Therefore, if the sRNA has the requirements to go through "CopraRNA", this software should always be the choice [97], [98], [96].

4. *Prioritization of targets*

In this section, the results given by every "CopraRNA" and "IntaRNA" were analysed, in order to study the likeness of the targets, given by those software to each sRNA, being indeed targets. Moreover, the most interesting targets also needed to be chosen. By doing this, the most interesting sRNAs would be prioritized depending on their targets. For instance, if a sRNA had targets that were related to the secondary metabolism, or any type of regulation, they would be considered interesting and their prioritization (explained in an earlier section) would rise. The more prioritization an sRNA has, the more chance it has on being experimentally studied, and a conclusion on how correct was the prediction done by "CopraRNA"/"IntaRNA" can be done.

In this part of the study, all different lengths and all different homologues (some with more percentage of identity than other) for each candidate searched on both software were compared, in order to see if a conclusion could be made on how the size of the sequence and the percentage of identity of the homologues could influence the results on "CopraRNA"/"IntaRNA".

The first 31 targets from each length/set of homologues for each sRNA given by the software were analysed. This number was chosen because it included, most of the time, every target that was considered reliable (defined as having a value of fdr smaller than 0.1) and it was a number big enough to understand if there were some false positives (here, false positives are the targets that might be the correct ones, but "CopraRNA" predicted a worse fdr value). Therefore, having a value of fdr smaller than 0.1 is of great importance, but not the only important parameter, because, even if the fdr is bigger than that value, targets such as "RNA polymerase" and "sensory kinase" are interesting, since they are regulators. That is why

there were other characteristics studied for each target, such as the description given by GenBank and RefSeq.

On the other hand, profiles were also analysed. The ones that were considered the most interesting were the ones that were the "reverse profile" of the candidate sRNA. This means that if the candidate had a profile "A0", the profiles that were "D0" had to be prioritized. However, this classification of profiles is very strict and, in order not to discard other targets that might be important, the definition of "reverse profile" was broaden. This way, if the candidate sRNA had more expression of the WT strain ("A0", for instance), the interesting targets only would have to have more expression in the mutant one. Only profiles that were the reverse of the sRNA's were considered interesting because they might mean that some negative regulation might be happening. However, this parameter couldn't be used to either confirm nor to discard targets.

To help this differentiation between which were opposites profiles, differential profiles were used. These were done by subtracting from WT the expression of the mutant strains (**Fig. 27**).
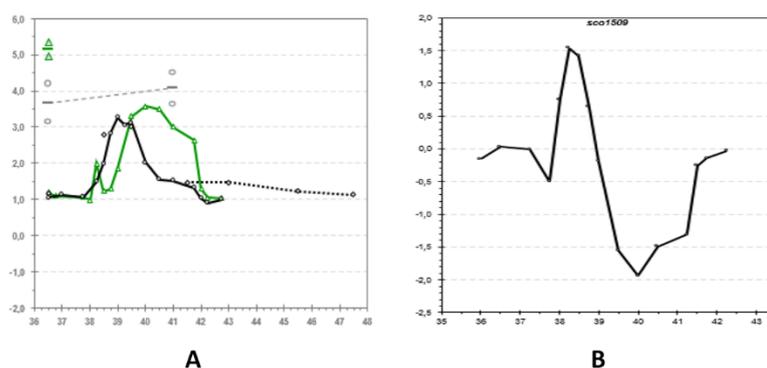

Fig. 27. Example of a profile of a gene **(A)** and its differential profile **(B)**.

After that, characteristics given by "CopraRNA"/"IntaRNA", such as energy and coordinates where the sRNA binds with the mRNA were studied (column "PositionsRNA" and "Start(Q)" in **Fig. 28**). The first one is relevant to understand how easy that sRNA binds with the mRNA (the lower the energy is, the easier it is to hybridize), while the second parameter is used to see if there were targets that had the same position. This last one is important because, the more similar the position of the sRNA between targets and between its antisense (when there is one) is, the more reliable is the prediction that those targets are indeed targets. Moreover, if the coordinates are repeated in many targets, that site might be very conserved (**Fig. 28**).

Then, there was the need to confirm if the sRNA really existed. To do that, the interaction sites where the mRNA hybridizes with the sRNA (the opposite of the one explained earlier, column "PositionmRNA" and "Start(T)" in **Fig. 28**) predicted by "CopraRNA"/"IntaRNA" had to be investigated by looking for an IT. To do that, in Copra, the values given by the results, must be subtracted to 200, because the software sets 200 nt upstream of the start codon and 100 downstream of the stop codon. This was done only to those targets that had coordinates that started before 200 (**Fig. 28**) [97], [98], [96].

The same happens in "IntaRNA". The difference between "IntaRNA" and "CopraRNA" is that the number of nt "IntaRNA" places upstream of the start codon and downstream of the stop codon is 75 in both. Therefore, as it happens in "CopraRNA", the targets that have starts that began before 75 nt had the possibility that their site of interaction was "made up" by the software [98], [96], [102], [101].

Rank	CopraRNA p-value	CopraRNA fdr value	Locus Tag	Energy [kcal/mol]	IntaRNA p-value	Position mRNA	Position sRNA	Annotation
1	0	0	Gene A	-182.47	0	1 -- 121	1 -- 121	SCP8,30 conserved hypothetical protein len: 260 aa; similar to TR:O05856 (EMBL:Z95120) Mycobacterium tuberculosis hypothetical 25,1 kD protein MTCY07D11,16 231 aa; fasta scores: opt: 599 z-score: 647,2 E(): 1,4e-28; 42,5% identity in 233 aa overlap, Contains 4x degenerate repeat:T(G/A)G
2	3.22E-02	0.09117	Gene B	-17.77	0.017052	46 -- 69	101 -- 123	Sun-family protein
3	0.0003	0.4359	Gene C	-26.78	0.000526	162 -- 237	10 -- 70	cytochrome P450
4	0.000333	0.4359	Gene D	-20.52	0.005881	170 -- 237	62 -- 123	SC4G1,30 possible hypothetical protein len: 298 aa, Contains possible N-terminal region signal peptide sequence possibly cleavable
5	0.000385	0.4359	Gene E	-18.92	0.01094	29 -- 47	21 -- 39	thiosulfate sulfurtransferase
6	0.000542	0.4478	Gene F	-15.65	0.038539	162 -- 238	6 -- 72	SC9C5,04c conserved hypothetical protein len: 107 aa; highly similar to TR:P96802 (EMBL:U75344) Mycobacterium smegmatis integration host factor MihF 105 aa; fasta scores: opt: 378 z-score: 470,3 E(): 1e-1; 58,1% identity in 105 aa overlap
7	0.000624	0.4478	Gene G	-17.11	0.022	210 -- 276	20 -- 74	transferase

A

Rank	Target	Start(T)	End(T)	Start(Q)	End(Q)	Energy	p-value	fdr	annotation
1	Gene H	121	147	13	48	-23.49	9E-07	6.71E-03	SCBAC36F5.04 hypothetical protein len: 452 aa C-terminal part similar to TR:Q9FC64 (EMBL:AL391515) Streptomyces coelicolor acetyltransferase SC4B10.23 163 aa fasta scores: opt: 287 Z-score: 300.0 bits: 63.3 E(): 4.2e-09 42.857% identity in 133 aa overlap. Contains Pfam match to entry PF01842 ACT ACT domain and PF00583 Acetyltransf Acetyltransferase (GNAT) family
2	Gene I	111	148	14	52	-22.05	1.15E-05	2.85E-02	acyl-CoA dehydrogenase
3	Gene J	1	28	15	40	-21.91	1.43E-05	2.85E-02	ATP-GTP binding protein
4	Gene K	31	56	25	52	-21.86	1.53E-05	2.85E-02	SCC61A.01 conserved hypothetical protein partial CDS len: >153 aa similar to others of unknown function from Streptomyces coelicolor eg. TR:O86768 (EMBL:AL031035) ATP /GTP-binding protein (886 aa) fasta scores opt: 457 z-score: 460.9 E(): 3.2e-18 48.6% identity in 146 aa overlap. Searches suggest a frameshift mutation between CDSs 01 and 02. SC6D10.23 hypothetical protein (partial) len: >85 aa highly similar to N-terminal region of TR:O86768 (EMBL:AL031035) Streptomyces coelicolor ATP /GTP-binding= 333 z-score: 478.0 E(): 3.2e-19 59.4% identity in 69 aa overlap

B

Fig. 28. Example of part of the results given by "CopraRNA" **(A)** and by "IntaRNA" **(B)**.

After this was done, the targets that were considered interesting, either because of their fdr value or because of their profiles, no matter if they had an IT where the software had predicted it, were searched on a study already done, where some of those sRNAs had already been investigated experimentally. This was done because some of the targets gotten from the software to these candidates had already been concluded, experimentally, as a member of the Pho regulon or as having Pho boxes, which turned them immensely more interesting and important. However, only a small percentage of the targets were described in that study.

Then, the annotation given by RefSeq (always better than the one from the GenBank, because RefSeq can only be given to complete and refined genomes) was searched on articles, to understand if there was a study that was able to relate that function to the secondary metabolism or the production of antibiotics. If, in the previous step, a target was found in that one study and was characterized as having Pho boxes, this gene would be prioritized when looking in other articles. This way, a regulator function might be predicted.

RESULTS AND DISCUSSION

Results obtained for methods in Microbiology

7. Plasmid Extraction

7.1. Minipreps

After doing the digestion of the minipreps gotten, electrophoresis is done. Here, no matter the enzyme used, fragments should be seen. Minipreps is a quick and "dirty" procedure to analyse genetic material, therefore, some genomic DNA (a signal with low mobility in the upper part of the gel) or distortion (a tail below the DNA fragment) might appear in the electrophoresis.

Partial digestions might also happen when the enzyme doesn't completely cut the plasmid. This might occur because the volume of enzyme is too little, the period of time incubating is too short, the DNA is "dirty", between other. When this happens, it is seen that the electrophoresis has bands that run too much and appear darker than the others that didn't run as much.

As seen before, the cut sites of the enzyme are accessible on the internet and, by knowing the sequence of the insert and plasmid, the fragment sizes can be calculated. The number of fragments always corresponds to the number of sites that the restriction enzyme cuts. In this case, the coordinates where the different enzymes cut are schematized in the image present in the methods (**Fig. 11**).

The first digestion done to these minipreps used EcoRI as the restriction enzyme and λPst as the marker (**Fig. 29**). By analysing the representation of the plasmid in the methods section, it can be seen that, with EcoRI, two fragments should be obtained (**Fig. 11**).
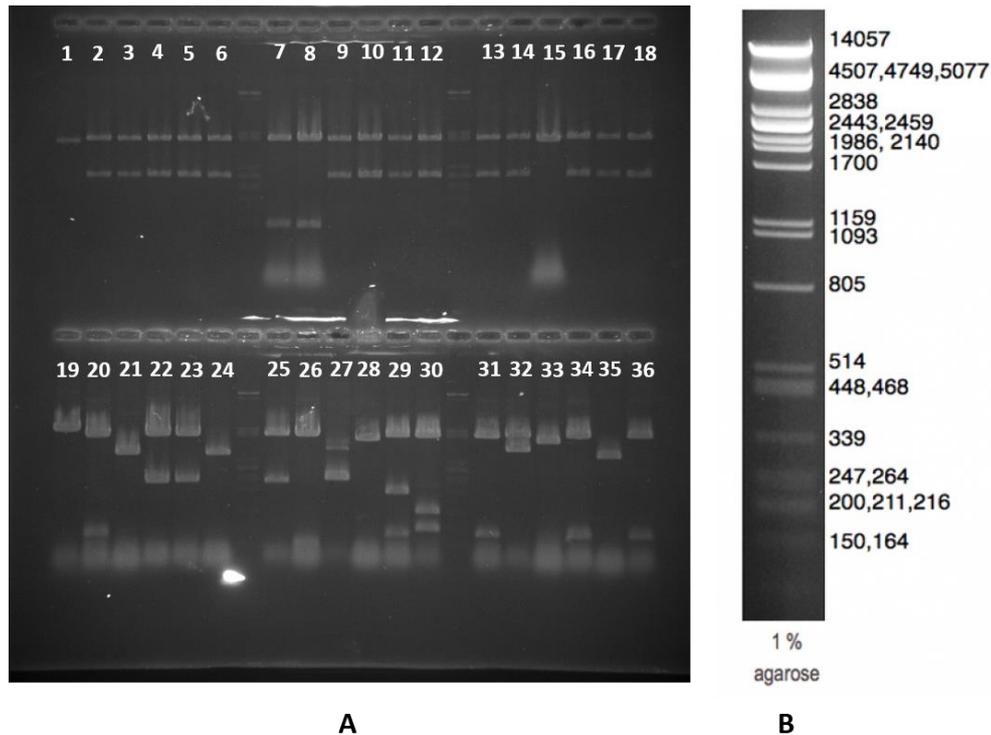
Fig. 29. Electrophoresis of the first digestion done to the minipreps **(A)** and λPst used as marker **(B)**, which was taken from [103].

As it can be seen in the image, there are some samples that got one or three fragments. These correspond to non-desired plasmids, because if the enzyme cuts in two sites, two bands should be obtained.

There are also samples with two fragments, which is good, but there are some of them that have different sizes. To see which were right, the sizes that were calculated were compared to the fragments gotten from the marker.

According to **Fig. 11**, EcoRI cuts in the coordinates 381 bp and 4835 bp of this plasmid, and therefore two fragments of 2048 bp and 4454 bp should be obtained. In the previous image, it can be seen that most of the samples in the first comb (samples in the wells 2-6, 9-12, 13-14 and 16-18) fit with the wanted pattern. There are also some in the second comb (samples in the wells 22-23 and 25) that could also have the right pattern. This is inferred since all these samples have fragments that, based on the marker, have sizes of between 1986-2140 bp and 2838-4507 bp, which it is what it is wanted.

Normally, if the right fragment pattern is not achieved (as it happens with the rest of the samples that weren't mentioned until now), it has something to do with the ligation itself and not the minipreps procedure. This is said because, when the minipreps are not done properly, no fragments are seen in the electrophoresis, which doesn't happen here. Moreover, in this case, since there are samples that have the expected fragments, it can't be said that there is something wrong with the ligation. What can be concluded is that the vectors from some of the samples didn't bind with the insert and, therefore didn't present the wanted pattern, which is normal to happen to some extent.

In some of the samples, a very light diffuse or many diffuse bands can be seen below the 300 bp. Normally these are rests of RNA and they don't interfere with the interpretation, unless there is a fragment of interest that has less than that size. There will always be RNA in the samples because this protocol drags a lot of it.

However, in order to be sure that the construction is correct on those samples that had the right pattern, some more digestions, with different enzymes, must be done. Therefore, some samples that seemed to be right were chosen (here, the samples 2 to 6, were used) to go through a second digestion (**Fig. 30**).


A   **B**
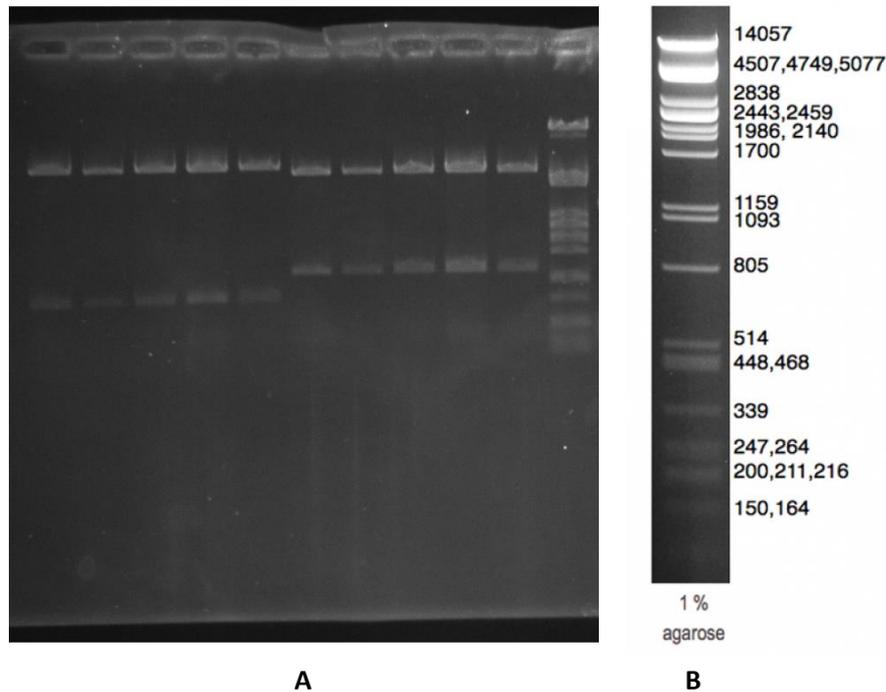
Fig. 30. Electrophoresis of the second digestion done to some of the minipreps **(A)** and the marker λPst used **(B)**. The first five were cut with EcoRV, while the other five were cut with PvuII. **(B)** was taken from [103].

In this image, two digestions with the samples chosen before were done. The first five used EcoRV and the last used PvuII. The marker was the same as the one used in the last digestion.

The sizes that had to be obtained can be seen in the image of the ligation protocol (**Fig. 11**). For instance, the enzyme EcoRV cuts in the coordinates 361 bp and 6121 bp, and since the plasmid size is 6883 bp, the fragments originated by this enzyme should be 5760 bp and 1123 bp. The bands that are present in **Fig. 30**, by comparing them to λPst, have sizes between 5077 bp and 11501 bp (the first band) and between 1093 bp and 1159 bp (the second band), which correspond to what was calculated.

On the other hand, the other five samples in **Fig. 30** were cut by PvuII. According to **Fig. 11**, this enzyme should cut in the coordinates 560 bp and 6121 bp and, therefore, originate fragments of sizes 5561 bp and 1322 bp. When the bands seen in this image are compared to the ones seen in λPst, it can be concluded that the first band has sizes of between 5077 bp and 11501 bp, while the second has between 1159 bp and 1700 bp. Both are around the sizes that were expected.

Therefore, these samples were indeed correct, which means that the plasmid is correctly built, and they can go through the next step of extracting plasmid, with E.Z.N.A. kit or alkaline lysis, depending on how much plasmid and how much purity is required.

7.2. *E.Z.N.A. kit*

Several samples went through this protocol during this project. Here, it will be only presented results from two of them.

As seen in the protocol, after extracting the plasmid, the samples were analysed in the "NanoDrop". The results were the following:

Table 4. Results on "NanoDrop" of the extraction of plasmid by E.Z.N.A. kit.

Sample	**Concentration of DNA (ng/µL)**
14	*250*
19	*400*

The two samples have different concentrations probably because they come from different cultures.

To check the preparation of the plasmid, after going through the protocol, electrophoresis was done with λHindIII as molecular-weight size marker (**Fig. 31**).
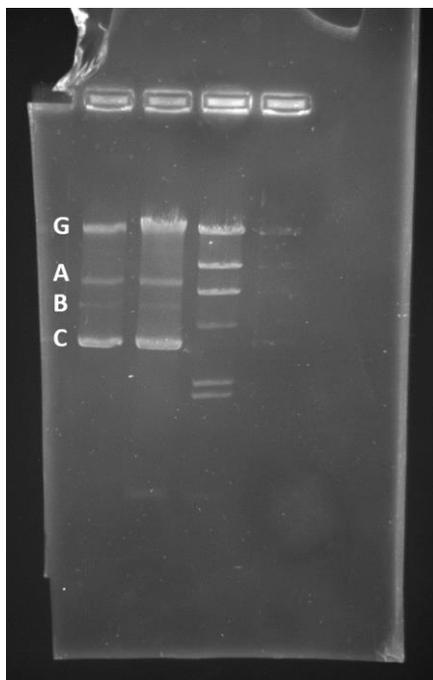
Fig. 31. Result from the electrophoresis done for two plasmids extracted. This was done with λHindIII as the marker. The "G" represents the genomic DNA and the other three letters (A, B, C) symbolize the three conformations.

In this figure, four bands are seen for each plasmid. The first one is genomic DNA. This might have happened because there is a crucial step in the protocol, that, if done incorrectly, might interfere with the results. The step mentioned is the one where the sample has to be agitated slowly and carefully (step 6). If some more force than needed is done here, the genomic DNA can break and stay in the sample together with the plasmidic DNA.

The other three bands were from the three different conformations that a plasmid has. The more intense was the one that ran the most and corresponded to the CCC form (C in the figure). The one behind was the OC form (B) and, finally, the one left represented the linear form (A). This difference seen in the electrophoresis between conformations had nothing to do with size, but with the conformations themselves and the mobility that each of them originates through the gel.

Even though normally a 1.5 mL overnight culture in LB medium will have 3 to 12 µL of plasmid DNA, yield and quality of the plasmid DNA depends on many factors, such as *E. coli* strain, site of insert, plasmid copy number, the kit's binding capacity and growth conditions (as culture medium and volume). For instance, subculturing directly from liquid cultures or glycerol stock might lead to uneven yields or plasmid loss. Therefore, one single isolated colony from a freshly transformed or freshly streaked plate should always be used.

9. *Cup plate method**

The aim of this protocol was seeing if the different mutant strains of *Streptomyces*, which had lack of biosynthetic clusters for unknown compounds, had any impact in the growth of other microorganisms, and that is seen through halos, or inhibition zones, in the petri dish. A halo is an area in the solid culture in a petri dish where the indicator microorganism couldn't grow, because of a compound that *Streptomyces* might be producing that can inhibit it. This halo can be seen easily, since it differentiates from the medium as having a lighter shade of colour.

After the method done and the cultures let grown, it was seen that, from all the indicator microorganisms (*Candida utilis*, *Bacillus subtilis* and *Kocuria rhizhophila*), halos were only seen in *Kocuria rhizhophila*. Moreover, as seen in the image below, the most substantial halos were seen in the WT strain and in one of the mutants (M1), while the other mutant strains didn't have much of an inhibition zone (**Fig. 32**). The difference between the three mutant strains was that M1 had one biosynthesis aggrupation erased, while M2 and M3 had eliminations in other aggrupations as detected by their genome sequence.
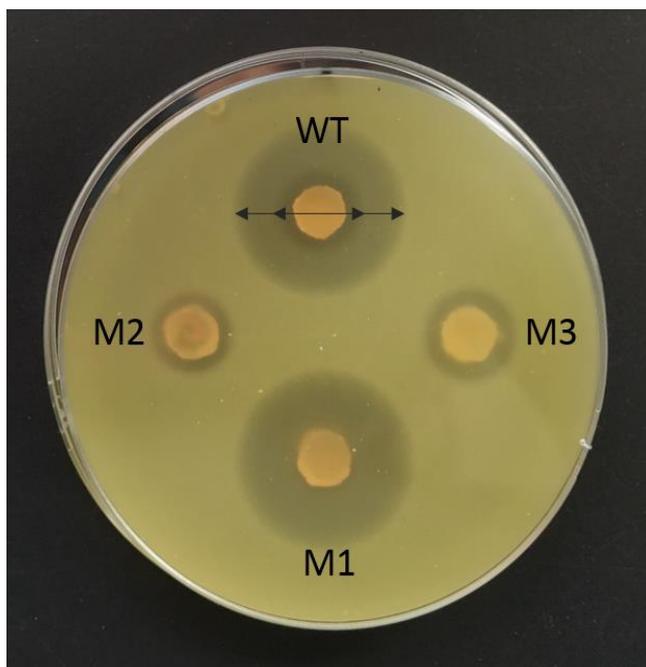

Fig. 32. Result from the cup plate method. The method was done in *Kocuria rhizhophila* against the WT of the *Streptomyces* strain and three mutant ones (M1, M2 and M3) with different biosynthesis aggrupations eliminated. The arrows in WT differentiate the two halos seen.

Moreover, when analysing M1, it can be concluded that there was another biosynthesis aggrupation capable of producing an inhibitory compound, because, even though this strain had one eliminated, there was still a halo.

The difference between the halos seen in the mutant strains might mean that the other aggrupations that were eliminated in M2 and M3 are compromising the synthesis of the inhibitory compound that works against *Kocuria rhizhophila*.

In this picture, it can also be seen that the WT had two inhibitory zones (differentiated by the two arrows). Therefore, the WT is responsible for not just one inhibitory compound, but two.

The results obtained with this method are qualitative. This means that the only information it gives is if the microorganism produces a decrease in the growth or not. Therefore, it can't be said that a bigger halo means that there is more concentration of the inhibitory compound or more inhibitory ability. However, when the difference between halos is very different, normally one of those facts is the responsible for it.

In order to have quantitative results, the method should have a crescent antibiotic concentration, in order to have numbers to create an equation that would relate the concentration of the compound with the diameter of the halo (such as an Epsilometer test, or E-test, for instance).

What could be done next? Next, these inhibitory compounds could be studied to understand what they were and what chemical structure they had.

Results obtained for methods in Molecular and Cellular Biology

5. *Electrophoresis of RNA**

The system used in this study to do electrophoresis of RNA samples, evaluates the RNA through the RNA Integrity Number (RIN), which is an algorithm that takes the entire electrophoretic trace into account and evaluates it. It rates the samples of RNA based on a numbering system from 1 to 10, with 1 being the most degraded and 10 being the most intact RNA [104]. In this project, only samples with RIN higher than 7 could be used.

This algorithm was also and developed to remove individual interpretation in RNA quality control, which means that, when using it, interpreting an electropherogram and comparing samples is facilitated and repeatability of experiments is ensured [104].

The following image gives what is expected of an RNA sample rated with a RIN of 10 (**Fig. 33**).
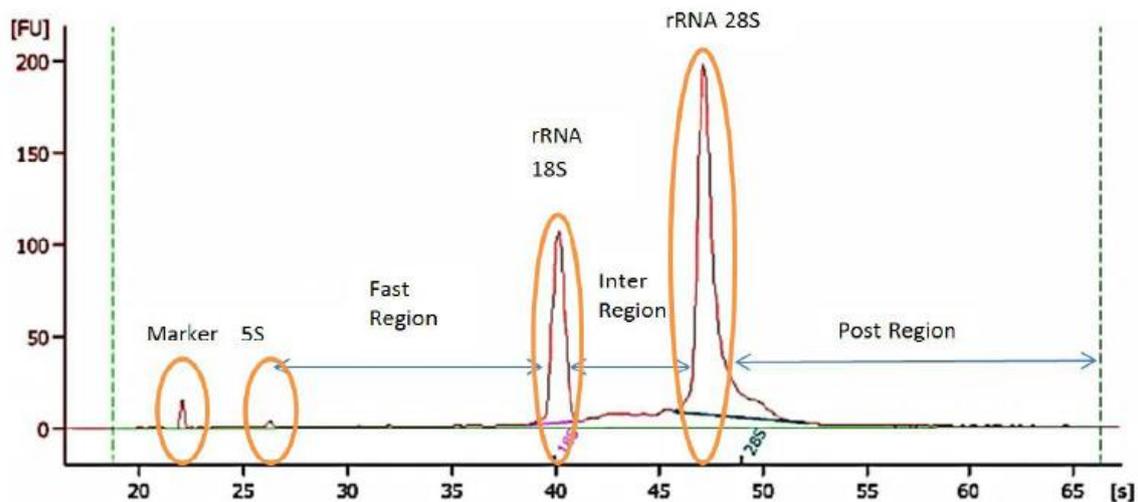
Fig. 33. Electropherogram of a perfect sample (RIN of 10) for eukaryotes. Taken from [105].

In this image, the different peaks are seen. Every single one has to do with fluorescence detected in the samples. The first one is normally the one that corresponds to the internal marker. This peak must be identified correctly, or else the RIN might be affected and the bands of the electrophoresis will appear higher/lower than supposed. If the software doesn't find it or identifies it in the wrong place, the operator can and should use the software to change its identification.

Then, the ribosomal subunit 5S is identified. This peak might not appear, but it is the one that matters the most, since if its aspect is more of a "mountain", it might also correspond to sRNAs.

The third peak belongs to the small ribosomal subunit (18S for eukaryotes and 16S for prokaryotes) while the last one corresponds to the large ribosomal subunit (28S for eukaryotes and 23S for prokaryotes). For the RIN to be as big as possible, the peak from the small subunit should never be higher than the one from the large subunit.

The two spaces between the peaks ("fast region" and "inter region") should be a baseline, as seen in the image, or else it means that there is some degradation.

Finally, the "post region" is not used by RIN to evaluate the RNA sample, but it is important because, if it is not a baseline, there might be DNA present in the sample, which can interfere with the electropherogram of the following sample in the Chip.

Even though this electropherogram was from a eukaryotic RNA, a prokaryotic sRNA rated with a RIN of 10 would have the same aspect.

These "optimal" electropherograms change depending on the microorganism (for instance, *Leishmania* normally has 3 peaks from the ribosomal subunits). However, the electrophoresis protocol is always the same.

As seen in the protocol, a ladder is also placed in the Chip, right beside the samples. This ladder is the first one to be analysed by the system, and functions as the molecular-weight size marker in a normal electrophoresis, which means that it is needed in order to compare the fragments and to know their size.

Normally, the electropherogram from the ladder will present 6 peaks and 1 "mountain", and the spaces between the peaks will be base lines. If the ladder (provided by the company of the Bioanalyzer, Agilent) doesn't present an electropherogram with this aspect, the cause might

reside from the fact that the ladder itself is very easily altered, is labile. This means that, while it was being split and frozen, the ladder could have suffered some alterations. If this happens, it cannot be used.

The ladder used in this study had the following electropherogram (**Fig. 34**):
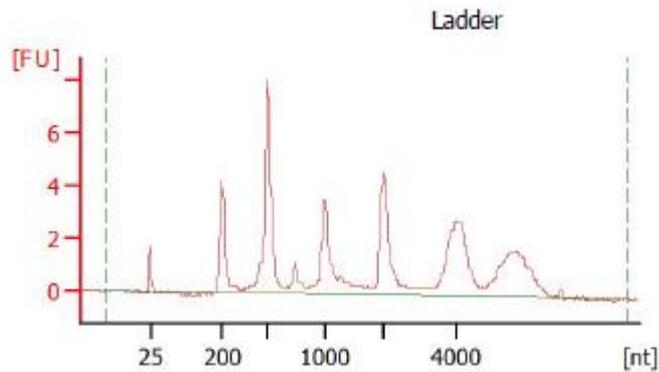
Fig. 34. Electropherogram of the ladder used.

With this image (**Fig. 34**), it can be seen that there is a small peak (the fourth one) that shouldn't be there. It is not known for certain what caused it, because more than one different tube of ladder was used and the result was always the same. However, it probably might be caused by the reason just explained.

In the following image, four examples of four different electrophoretic profiles obtained are seen (**Fig. 35**).

A

B

C

D
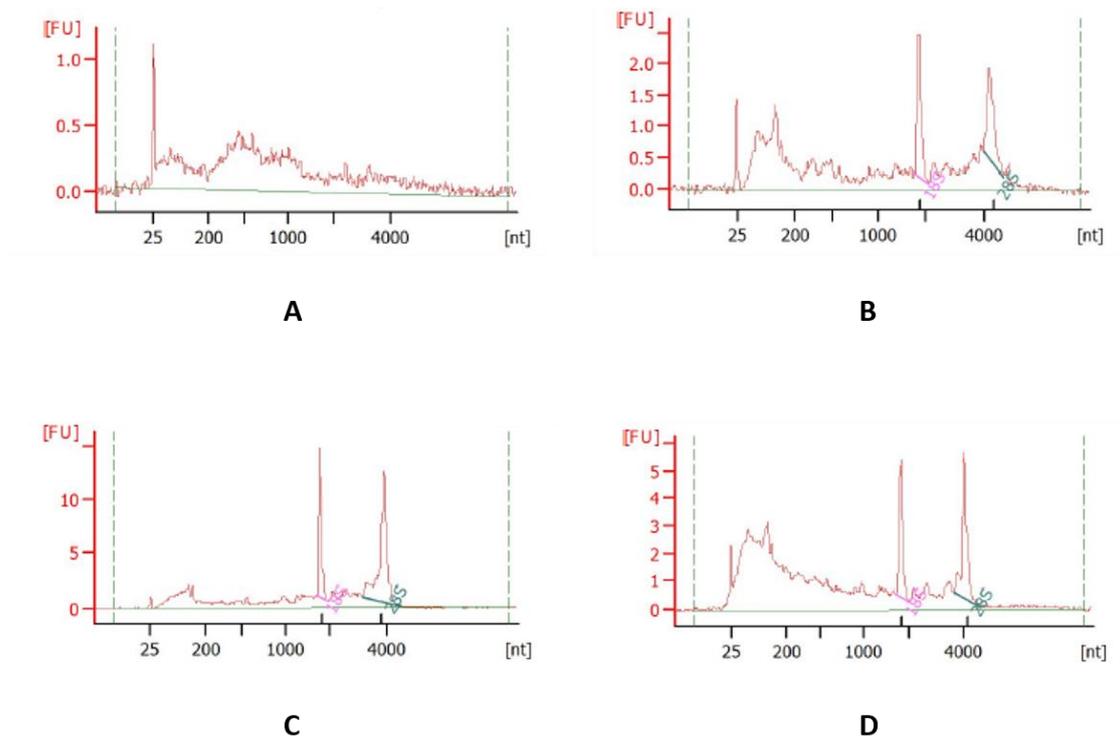
Fig. 35. Some of the electropherograms obtained in this study.

As seen in the image, the first electropherogram, A, as expected when compared to the one from the **Fig. 33**, had a very small RIN (of around 2.3), since it was basically composed by the peak from the marker and degradation.

On the other hand, B had a RIN of 6.5. This value can be explained by the peak originated by 5S, which the algorithm considered as degradation, because of the sRNAs that turned the peak into a "mountain". Moreover, there was also degradation from the large subunit, which can be seen even before the peak from the small subunit, since there was no baseline.

C only had a RIN of 7.7, even though it was a very clean electropherogram, exactly because the larger ribosomal subunit had a smaller amplitude, which means that there was some degradation. There was also an issue with the 5S region, since it was broadened, presumably due to the presence of sRNA. All these problems lowered the RIN value.

Finally, D didn't have a RIN value, probably because the algorithm wasn't able to find the peak from the marker, since it was very close to the peak from the 5S. However, if it did valuate this sample, the RIN would have been, theoretically, bigger than 6.5, the value given to B. This inference can be done because the electropherograms from B and D are very similar, however D has a bigger amplitude for the large ribosomal subunit, while B has degradation in that peak. Therefore, RIN for D would be better than B, but smaller than the one obtained for C.

Even though RIN can be used to ensure the repeatability of gene expression experiments in the step of sample extraction, it cannot predict the usefulness of gene expression data without prior validation work [104]. For instance, in this study, the sRNAs are the elements that are the most important ones and, therefore, the value given by RIN, that negatively evaluates that peak, should not be overvalued and should be analysed with caution.

This method has a quantitative range of 25-500 ng/μL and a qualitative range of 5-500 ng/μL. It is also a very sensitive technique, since the results can vary depending on salt content, cell fixation method and tissue stain. A common problem in laboratories using the Bioanalyzer is the chip shortage with excess reagents, particularly in laboratory service environments that experience both high usage and fluctuations in the demand for different kit types.

After having done electrophoresis to the RNA samples, they go through a transcriptomic experiment, RNA-seq, to study the cell expression.

6. *Polymerase chain reaction (PCR)**

Three different types of *Streptomyces* exconjugants were analysed by this method, and the difference in all rely in the particular insertions introduced in the very same point of the chromosome.

As explained before, two studies were done using this technique (**Fig. 21**). The first one had primers that amplified the area where the *Streptomyces* would integrate the plasmid (first comb of electrophoresis), while the second had primers that amplified a region of the plasmid already incorporated in the *Streptomyces'* genome (second comb). The results are represented in the following image (**Fig. 36**):
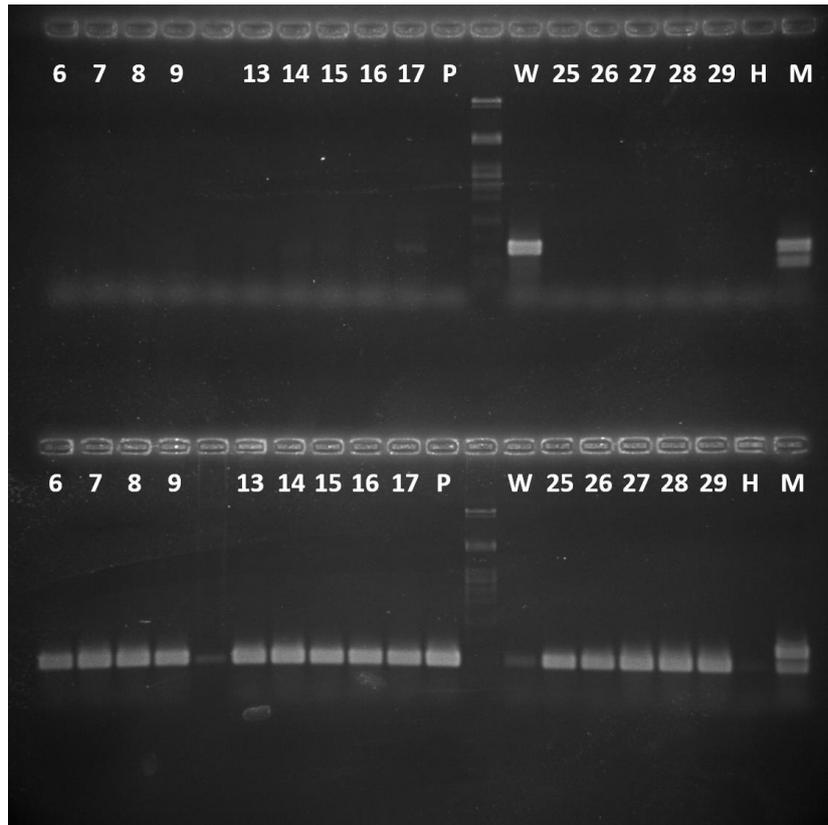
Fig. 36. Electrophoresis of the samples after PCR. (6-9), (13-17) and (25-29) represent the three different exconjugants while the letters represent the controls: P represents the plasmid, W the WT strain, H water and finally M represents the mix with both plasmid and the WT strain. The first comb represents the amplification of the site where the *Streptomyces* would integrate the plasmid, while the second corresponds to the amplification of the plasmid already incorporated in the *Streptomyces*' genome.

As seen in the picture, the controls gave the results wanted: water never had a band; the WT presented a band in the first study (which means that the insertion site is intact) while the plasmid had a band in the second one; and the mix had one band in both studies. In the second comb, however, the water sample had amplification. It is thought that some of the sample in the surrounding wells might have gone to its well while loading the gel.

It is also seen in the picture that all samples gave the patterns wanted, since, in the first study, the samples have no bands, because the place in the *Streptomyces*' genome that binds with the plasmid, after a successful conjugation, doesn't exist anymore. On the other hand, one band appeared in all samples in the second study, which means that the plasmid is integrated in the genome, because if not, since it doesn't have a replicative origin, the band wouldn't exist. These results mean that the conjugation was successful and recombination between plasmid and chromosome did happened.

However, sometimes the PCR shows that something didn't go along with the plans. If the electrophoresis didn't show these results, some hypotheses could be made in order to understand what went wrong. One of those possible hypotheses would be that something went wrong with the method (for instance, the thermocycler wasn't working properly). Another hypothesis, the one that happens most of the time, is that the recombination didn't worked as planned. Finally, there is also the probability of the morphology of *Streptomyces* used being to blame. Here, for instance, spores were used, which require more steps in the protocol (such as the rupture of the membranes) and, therefore, more mistakes can be done. To confirm that

the spores are not responsible for any of the undesirable results, this method should be done with non-sporulated colonies (in medium TSA, that doesn't let sporulation happen).

If using spores might lead to mistakes, why some other *Streptomyces'* morphology is not used instead? Filaments could be used, however it is even more risky than spores, since every section of it (and each filament has many sections) can have a different number of chromosomes (or even not have any). Moreover, these chromosomes, due to the genetic variability characteristic of the *Streptomyces*, might all be different. In that aspect, spores only have a single chromosome. Therefore, if filaments were used instead of spores, it would have to be ensured that they come from a single spore, so that all the chromosomes on the filament could be considered identical.

Results obtained in Bioinformatics analysis

1. Annotation

As explained before, there are two sets of samples: those sRNAs that had already been experimentally studied (the so-called "Selectos242_28") and those that haven't (called "No28"). The annotation during this work was only done for the "Selectos242_28" group. However, the method was the same as the one used for the other group of candidates.

Only 24 of the 28 were annotated. From those, 19 were considered sRNAs with some certainty, while 5 were discarded as not being sRNAs.

Here, a few examples of the analysis done for all of them is explained.

Firstly, a candidate "A" was studied. In the following image there is the fundamental data needed to analyze this sRNA (**Fig. 37**).
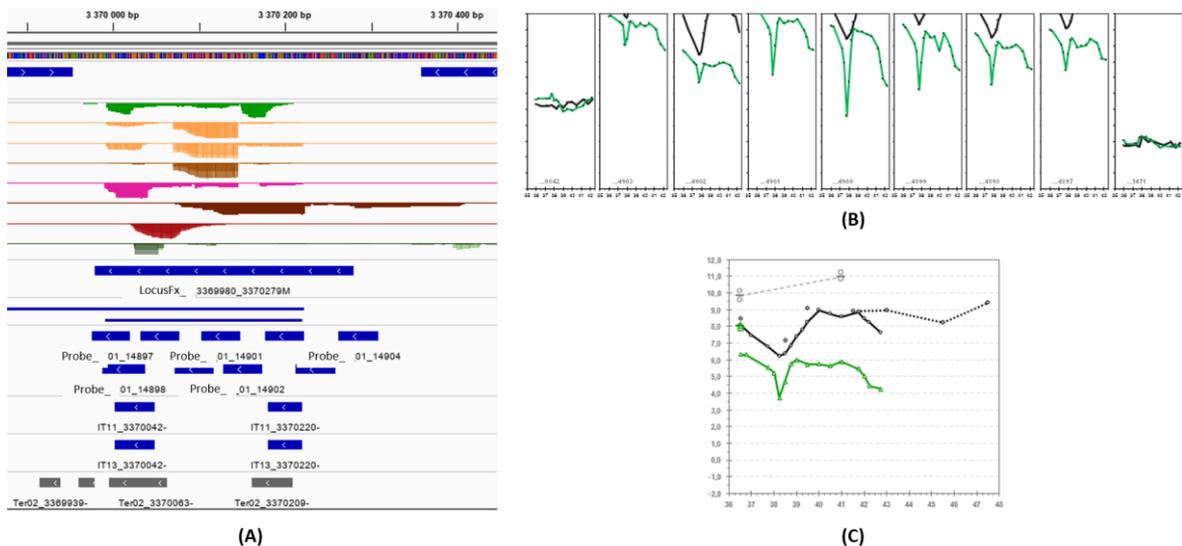
Fig. 37. Information for the analysis of Candidate "A". **(A)** IGV of the candidate "A", the black lines delimitate the sRNA and the circle the IT; **(B)** Some of the profiles of the probes seen in IGV; **(C)** Profile of the candidate "A".

As explained in the methods section, there are three conditions to consider a candidate a sRNA: expression; IT; and profile. In the section (A) of the image it can be seen that there is a rise on the coverage, which means that that site is being more expressed than its surrounding. To confirm this, the profile [section (C) of the same image] is needed. In it, it can be seen that,

in TPM values, the one that is most reliable, the candidate has a value of 10, which corresponds to 2^{10}, which equals to 1024. Therefore, the first condition is fulfilled.

On the other hand, in the circle of the part (A) of the image, there are ITs present. This way, the second conditions is also complied.

Finally, in the (B) part of the image, it can be seen that the profiles of the probes that are in the area near the predicted sRNA are different from the other (from the second until the eighth, every profile is equal, while the first and last are different from those). Therefore, it can be said that this candidate is indeed a sRNA.

Moreover, with those same probes, the length of the sRNA could be obtained, otherwise with so many ITs and terminators the job could be harder. The length of around 200bp was determined (as it can be seen in the part (A) of the image) with the probes aid, by looking for those that had similar profiles.

The following example was also considered an sRNA, but with a difference: the one explained until now was given a value of 1, while this one (the candidate "B") was given 0.9 (**Fig. 38**). Normally, this difference in value is explained by the last one not having a very clearly identified ending, which is not exactly the case. In those other cases, normally, the most reliable since to do is to confide in the coverages and not in the Ter.
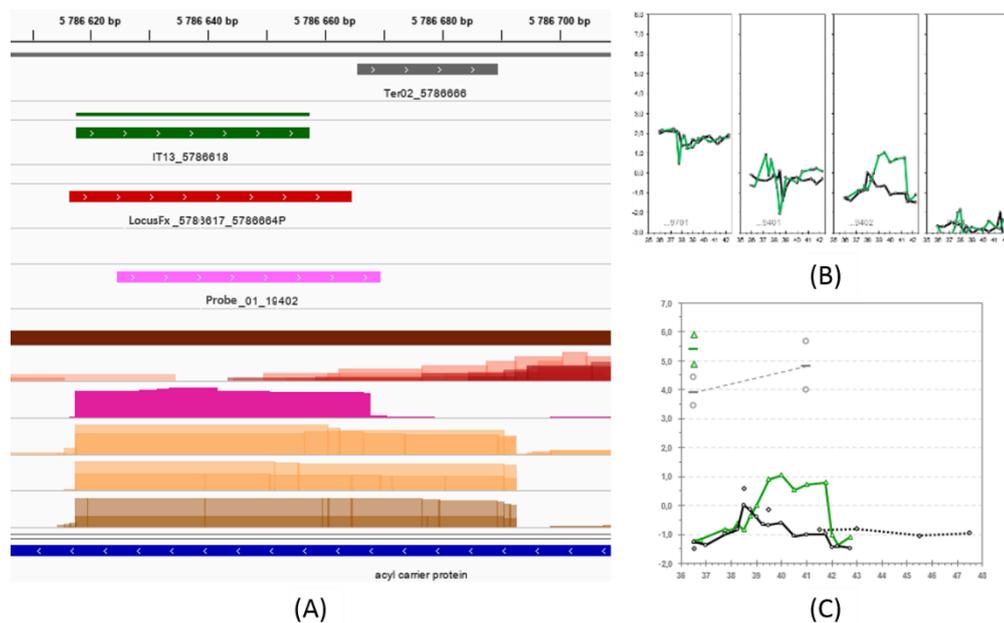
Fig. 38. Information for the analysis of Candidate "B". **(A)** IGV of the candidate "B"; **(B)** Some of the profiles of the probes seen in IGV; **(C)** Profile of the candidate "B".

Here, the three conditions for being considered a sRNA are still fulfilled: it is being expressed, even though it is less than the anterior ($2^5=32$, instead of 1024 as seen before); it has an IT, as seen in the (A) part of the image; and it has a different profile than its surroundings, as seen in the (B) part, where the third profile, that belongs to the probe of the candidate, is different from the others. Therefore, these parameters are not the cause of the decrease of the value from 1 to 0.9. The issue is the length of the sRNA, which can be analyzed in the (A) part of the image.

There, it can be seen that the length is of around 70 nucleotides, which is suspicious, because that is below the minimum read length when Illumina is used. This means that it is

better not to trust the terminator, because, if this is happening, it might just be a secondary structure. This is why the value dropped to 0.9.

Normally, in cases such as this (or when there are lengths higher than 400 nucleotides, which is also suspicious since sRNAs are defined as RNAs smaller than that value), the coverages shouldn't be totally trusted. In those cases, is better to use the probes that mark the sRNA, by selecting some of them that are before, after and in the foreseen sRNA. However, this doesn't work in this example because, while analysing the probes (part (B)), it can be seen that there is not one profile near the candidate that is similar to the one from the sRNA's.

There are candidates, such as the next one (the candidate "C"), which was given the value of 0.5 (**Fig. 39**).
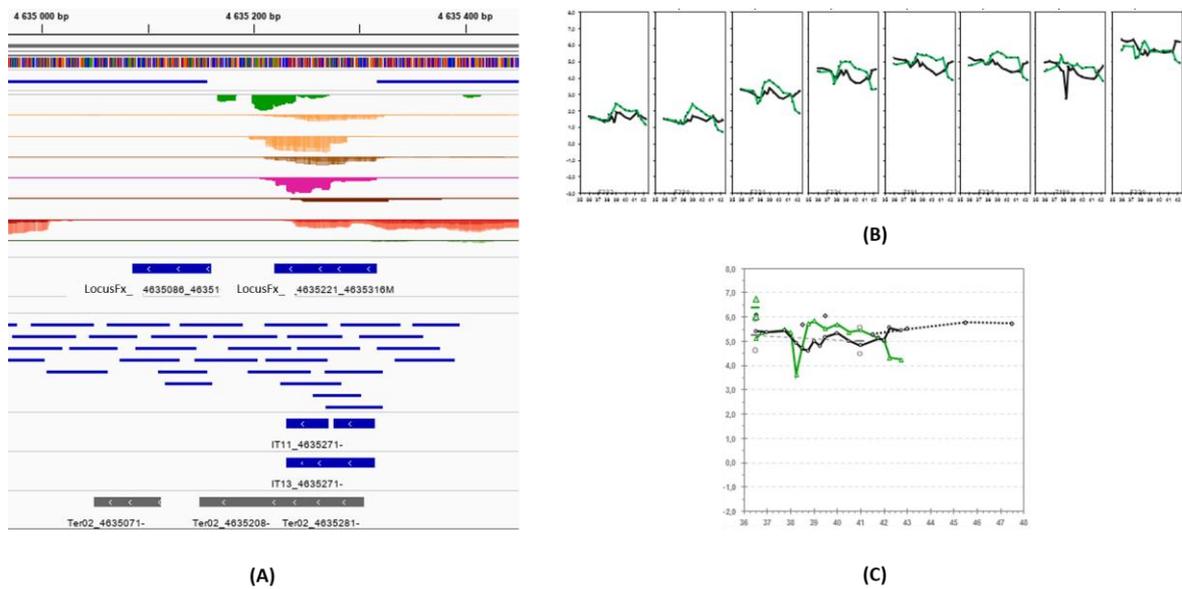
Fig. 39. Information for the analysis of Candidate "C". **(A)** IGV of the candidate "C"; **(B)** Some of the profiles of the probes seen in IGV; **(C)** Profile of the candidate "C".

If those three parameters needed to consider a candidate a sRNA are analyzed, it can be seen that two of them are fulfilled, since it has an IT [part (A) of the image] and a TPM of 5, what corresponds to 32 [part (C) of the image]. However, as it can be seen in the (B) part of the image, there is not one probe that is that different from the other. There is also the fact that, even if there is an IT well identified in the IGV, the profiles of the probes are the same as the ones that are before the IT, which might mean that the sRNA begins earlier than the IT indicates. Therefore, the value of 0.5 was given to this candidate.

Finally, an example of a candidate "D", which is certainty not a sRNA (value 0), will be given (**Fig. 40**).
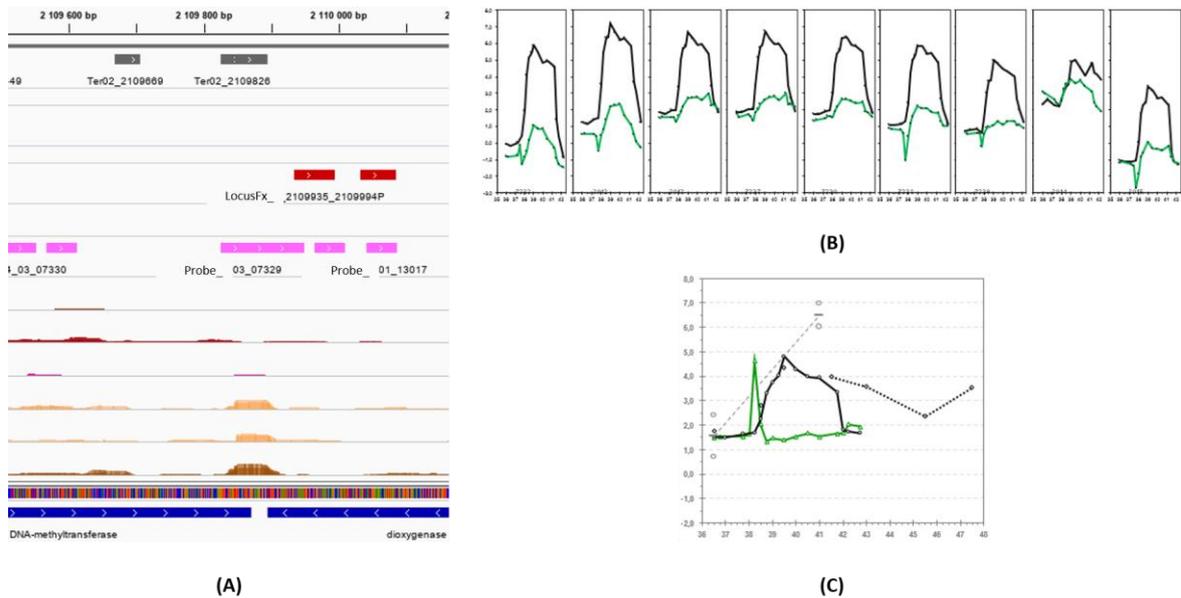
Fig. 40. Information for the analysis of Candidate "D". **(A)** IGV of the candidate "D"; **(B)** Some of the profiles of the probes seen in IGV; **(C)** Profile of the candidate "D".

In IGV [part (A) of the image], it can be seen that where the sRNA was predicted (the red boxes) there is not an IT present. Therefore, one of the three parameters needed to a candidate be considered a sRNA is already unfulfilled. But it is not the only one. In the (B) part of the image, it can be seen that none of the probes has a different profile, which is another of the three parameters. Because of this, this candidate "D" was not considered a sRNA, but the rest of a gene.

The difference between this candidate and the one before (the one with value 0.5), is that this one doesn't have an IT close to where the sRNA was predicted, while the other has one (even though it is unclear if it is correct).

However, this analysis is not the only thing that is required in order to annotate. As explained in the methods, it is also needed to describe the type of sRNA (therefore, "D" won't be analyzed here). As it can be seen in the parts (A) of each example, "A" and "C" are intergenic while "B" is antisense.

To annotate, profiles also must be described. "A" has a A0-0|DD profile, even though it has a very narrow peak in the WT strain, that is considered an artifact and, therefore, ignored. On the other hand, "B" is considered to have a profile AA-1|0D. Finally, "C" has a profile 0|0, which might be confusing since there are some increases and decreases of expression. However, these are so slight (and the biggest increase happens before the limitation of phosphate at 37 h) that they are ignored.

This classification of profiles used until now is exact, but not precise, because a small rise is not differentiated from a big one (a "a", instead of the "A", should be used to characterize these cases, for instance), as it can be seen in "B", where the WT is considered an "A" even though its expression didn't rise as much as the mutant strain in "A.

2. *Prioritization*

In this section, only the "No28" group was used, since all those candidates present in the other group are considered important and, therefore, will go through the next step of analysis, even if not prioritized.

After the analysis was done, the following results were gotten (**Table 5**):

Table 5. Representation of the results gotten for the prioritization of sRNAs.

Prioritization	Frequency
50	13
30	65
20	64
10	9
Total:	151

Therefore, the majority of the candidates were given the value of 30 and 20.

Some of these values might be overestimated because of the sRNAs that were internal and antisense to other sRNAs, since, as explained before, the only characteristic analyzed to prioritize them was the profile. Therefore, if the profiles of these sRNAs were interesting, they were considered as related to secondary metabolism. This method is very unreliable.

From those 151, only 31, the ones that had prioritization of 20 and were intergenic, went through the next step.

3. *Searching for targets*

From the 31 candidates that came from the prioritization of the "No28" group, 11 didn't have two homologues that obeyed the parameters and had to go through "IntaRNA". The same happened to 9 of the 24 candidates from the "Selectos242_28" group. One example of a candidate that had to go through "IntaRNA" is given in **Table 6**.

Table 6. Example of homologues given by "GLASSgo" and their analysis.

Delimitation_Chromosome	Valuation	ID_GenBank	Has admitted NZ?	Has admitted NC?	Is it Streptomyces?	Manual selection for CopraRNA
>A:c6225658-6225509	100.0%	A	1	0	True	0
>B:c167965-167816	100.0%	B	0	0	True	0
>C:2396378-2396527	100.0%	C	1	0	True	0
>D:c5822682-5822533	82.93%	D	1	0	True	1
>D:2583397-2583546	80.12%	D	1	0	True	0

Here, from the 5 homologues gotten, three of them had valuation of 100 %, which discarded them so that they weren't used in "CopraRNA". Moreover, the last two homologues belonged to the same organism and, therefore, only one could be used. Thus, only 1 of the five could go through "CopraRNA" (seen in the column "Manual selection"), which is not enough. Therefore, this sRNA went through "IntaRNA" instead.

Even though "CopraRNA" is more reliable at predicting targets than "IntaRNA", the number of nt both software place upstream and downstream of the start or stop codon is different. "IntaRNA" adds 75 nt both upstream and downstream, while "CopraRNA" adds 200 nt upstream and 100 downstream. This means that "CopraRNA" adds more artificial nucleotides, which increases the probability of predicting a binding site that doesn't exist. However, this problem still isn't enough to consider "IntaRNA" a better option [97], [98], [96], [102], [101].

4. _Prioritization of targets_

Here, an example on how the targets given by "CopraRNA"/"IntaRNA" were prioritized will be explained.

31 targets given by "CopraRNA"/"IntaRNA" from every length/set of homologues were analysed. After seeing how many reliable targets there were (fdr lower than 0.1), it was decided which of the 31 targets analysed had interesting profiles, as seen in the image (**Fig. 41**). If many targets had a good fdr value, it would mean that that sequence was very conserved.

Even though these many targets were studied, there is still the probability that some false positives and negatives might appear. This means that in the table given as result by the software, those that appear first, with a better fdr, might be false positives, and there might exist some further in the table, and therefore with worse fdr, that are false negatives.
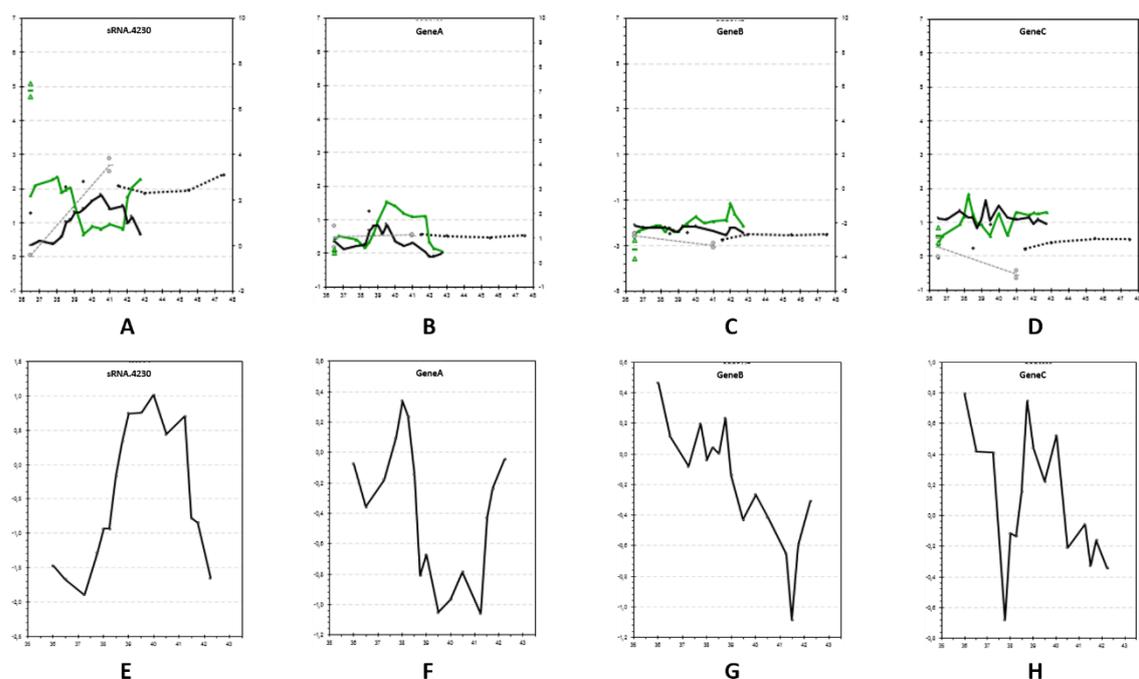

Fig. 41. The profiles of the sRNA and some of its targets and their differential profiles. **(A, E)** Profiles of the candidate; **(B, F)** Profiles of the target GeneA; **(C, G)** Profiles of the target GeneB; **(D, H)** Profiles of the target GeneC.

As explained before, interesting profiles are the ones that are considered the reverse of the candidate's profile (profile A). Two profiles that seem to fulfil that requirement were studied (GeneA and GeneB). However, when the differential profiles were analysed, only the one in GeneA is undoubtedly a reserve profile, because when the candidate's profile rises, this target descends and vice versa. On the other hand, GeneB is not as clear. Finally, the differential profile of GeneC is more similar to the candidate's than to GeneA's and, therefore, shouldn't be considered an interesting profile, because it is not the reverse of the candidate's. However, in this case, GeneC is the antisense of the candidate, which is a characteristic that will always make it interesting.

There are profiles that aren't as interesting as they seem. There is, for instance, cases such as the one in the figure below, which would probably be considered an interesting profile if the candidate had a profile DA (**Fig. 42**). In that profile there is a rise right after a

descension. This rise would normally be related to the Pho regulon, but, here, it might just be caused by another generic regulation (such as the one done by sigma factor).
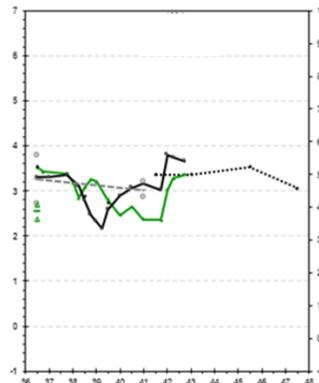

Fig. 42. Profile that might not be as interesting as it seems for cases of candidates with profile DA.

On the other hand, a rise in the profiles, if only transitory, might also not be interesting, because other mechanisms, that try to compensate the need of phosphate, might be the reason why this phenomenon happens and not the Pho regulon.

These sudden changes in the profiles suggest that some artifact is the cause of it. However, even if it is not impossible, it is very unlikely to happen, since the only artefacts that might appear here are the mistakes done with the operator.

There are also cases of profiles where both strains overlap each other, which might mean that there are two promotors dependent on phosphate, and not a Pho regulon.

Finally, one other fact that makes this selection of interesting profiles harder is that if the profile is seen very often, in many targets, even though it seems interesting, it is considered to not give that much information and, therefore, it is not very useful.

Another parameter that needs to be studied has to do with the fact that the IT might not appear where "CopraRNA"/"IntaRNA" said it should be. This might happen due to several reasons, such as: that part of the gene is not being expressed; it wasn't detected; or it's located in a polycistron (which is a messenger RNA that encodes two or more proteins) and, therefore the IT is probably in the beginning of the polycistron. (**Fig. 43**).


A
B
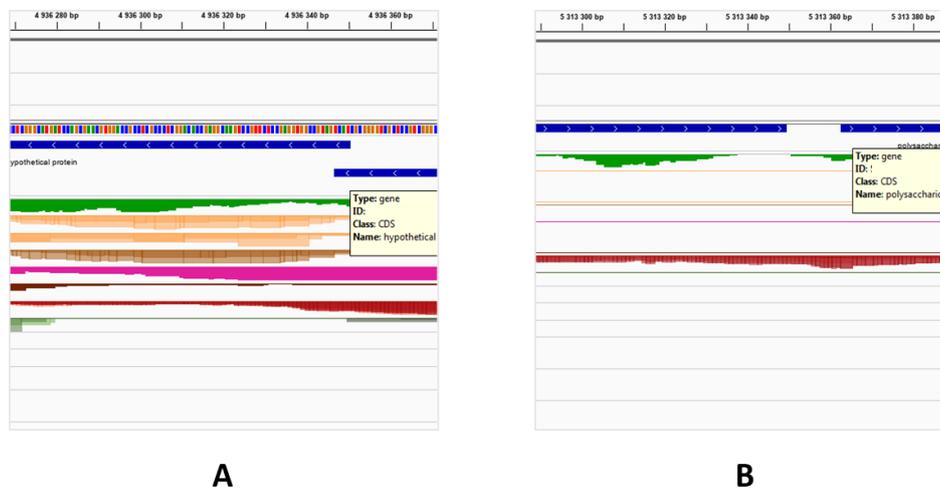
Fig. 43. Targets that didn't have an IT where the software said it should be. The reasons for that normally reside in them being polycistrons **(A)** or not being expressed **(B)**.

Then, finally, the regulatory function of the targets is looked in articles, to understand if they can be related to primary, secondary or central mechanism. The reason why the profiles can be of relevance here is because genes and sRNAs with the same profile should be corelated with the same mechanism or the same function, but not with the same protein. However, that doesn't occur most of the times (**Fig. 44**).
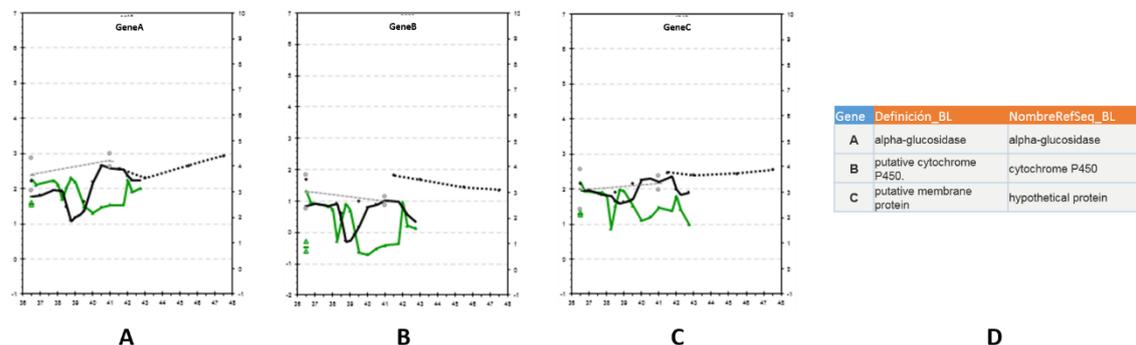

Fig. 44. Profiles of gene A, B and C and their annotation.

In this image, it can be seen that, even though the three targets have very similar profiles, they don't share the same annotation.

As also seen in the previous image, two annotations were used. However, the annotation given by RefSeq was considered more reliable, because, even though both RefSeq and GenBank belong to NCBI, they have differences in terms of details. RefSeq is considered to be the best since it can only be used in complete and refined genomes.

When the annotation of the targets refers to "phosphate transporter" or "phosphatase", the sRNA can be considered to be involved with the system PhoR-PhoP. This is done because it is thought that when phosphate is low, the bacterium either tries to grow by capturing more of it or by going to its reservoir (using, for that, phosphatases, for instance). When phosphate is not available at all, the cells have no choice but to differentiate into spores.

These annotations have different levels of precision and accuracy depending on what it is said. For instance, describing a target as a glutamate/aspartate transmembrane protein is being more precise than just describing it as transmembrane protein. On the other hand, in terms of accuracy, characterizing it as a transmembrane protein is more likely accurate than as a glutamate/aspartate transmembrane protein.

Some of the sRNAs were searched with different sets of homologues, one with more percentage of identity than the other (the ones with more percentage are referred as "most similar" in the next table), and different lengths, one longer than the other, in both software, in order to understand if those characteristics were able to change the results obtained and which length/set was better. A summary of the results gotten for those sRNAs can be seen in the following table (**Table 7**).

Table 7. Summary of the results gotten for those sRNAs that had different homologues/lengths searched. Here, "?" means that it's not certain which set/length is better, while "N.D." means that there is no big difference between the two sets/lengths.

sRNA	"CopraRNA" or "IntaRNA"?		What length was best?		What set of homologues was best?	
	"CopraRNA"	"IntaRNA"	Longest	Shortest	Most similar	Most different
sRNA.3929		X	X			
sRNA.5842	X					X
sRNA.6089	X			X		
sRNA.3017	X				?	?
sRNA.3123	X				X	
sRNA.5478	X				?	?
sRNA.5555	X				X	
sRNA.5649		X		X		
sRNA.6372		X		X		
sRNA.7057		X	N.D.	N.D.		

To reach a decision on which set/length was better, the number of reliable targets and interesting profiles were normally the decisive parameters. In cases where those two parameters weren't enough to decide, there were also other important characteristics seen that could be used, such as if any of the sets/lengths had a Pho box experimentally confirmed and if the regulative function of any of the targets in the set/length could be related to the secondary metabolism. The number of reliable targets and interesting profiles were the decisive parameters for most of the sRNAs, except for sRNA.3929, in which the decisive characteristic was the number of interesting regulative functions.

As seen in the table, there were two sRNAs (sRNA.3017 and sRNA.5478) for which a conclusion on which set wasn't made, because, in both, one set had more reliable targets while the other had more interesting profiles. Therefore, the targets and their regulative function had to be studied. However, there was still some uncertainty, even on that parameter, in making a call on which was better.

In sRNA.3017, there were three targets on one of the sets that could be related to the secondary metabolism. On the other hand, the other set had one target that could be related to the secondary metabolism and other two to the primary and central metabolism. Both had targets that could be members of the Pho regulon, but they weren't experimentally confirmed. Therefore, a clear conclusion on which set was better wasn't made.

On the other hand, sRNA.5478 had both sets with interesting regulative functions in, at least, three targets. In one of them there were more functions related to the secondary metabolism and the other had more related to the primary and central metabolism. However, one of these sets actually had a Pho box experimentally confirmed, while the other only had predictions that there were targets able to bind with PhoP or declared as a possible member of the Pho regulon. This fact that one of the sets has a Pho box experimentally confirmed is not enough to decide if that set is better, because the other might not have Pho boxes confirmed experimentally only because some of its targets weren't studied yet. Therefore, a conclusion on which set is better couldn't be made with certainty.

Finally, there was one of the sRNAs, sRNA.7057, that the different lengths didn't lead to big consequences in the results given by "CopraRNA" (**Table 7**). This might be due to the fact that the lengths were not that different: one has 458 nt, while the other has 497 nt, which is a difference of less than 1/10 of the total length. Another reason for that to happen, and a more

likely one, is that these very similar results in different lengths are due to the "seed" being in the conserved section of both lengths.

Even if the sampling is not that large (this comparison was only done to 10 sRNAs), a conclusion, by analyzing the table, can be done for those samples. It seems that the results provided by the software are normally a bit better (in this case better means that it has more interesting and reliable targets) in the shorter lengths and most similar sets of homologues. However, to confirm this, more lengths and sets of different sRNAs should be searched.

CONCLUSION

Conclusion

During this internship, much was learned, namely several basic techniques of microbiology and molecular and cellular biology and also about the critical analysis of bioinformatic results.

Other objectives, such as knowing how to work in a lab, how to work with a team and how to work in a collaborative study were improved during the time in the internship. A few more abilities were also enhanced, such as learning to communicate science in a new language.

Right now, the understanding, that didn't exist before, on how bacteria are grown, how they are identified, how their genome can be edited and how their characteristics can be studied (between many other) was obtained. Therefore, because of these months, it is impossible to feel anything but a bit more prepared for this career and whatever steps forward in the future.

As my performance in the laboratory and in most of the procedures was quite satisfactory, the precautions needed to have in order to achieve successful procedures were understood. As seen in the results, the minipreps protocol had a number of samples that had the correctly built plasmid, and the PCR also gave the planned results. On the other hand, the plasmid extraction done with the E.Z.N.A. kit, even though it showed the three expected bands, also resulted in some genomic DNA, which, of course is not wanted. In the cup plate method, the halos were only seen against *Kocuria rhizhophila* in the WT strain and in one of the mutants. Finally, after analyzing the different sets/lengths it seems like the results given were normally a bit better in the shorter lengths and in the most similar sets of homologues.

But there are, of course, a few things that could be changed in order to improve these results and their discussion. For instance, while constructing a plasmid, having the resistance gene as a selective marker has some problems because it implicates adding antibiotic, which can lead to secondary effects. Therefore, a new method to select the colonies that have the plasmid should be considered.

On the other hand, the results obtained with the E.Z.N.A. kit can also be improved if the slow agitation step during the protocol is done with caution.

The same can be said about the bioinformatic part of this work. For instance, the classification of profiles can be improved, in order to differentiate a small variation from a big one. On the other hand, the prioritization of candidates also had issues related to those that were internal or antisense to other sRNAs, since the only characteristic analyzed to prioritize them was the profile, which is very unreliable.

Prioritizing targets can also be improved, since it depends on, mainly, the fdr value, profiles and function. This task has issues, because deciding which profiles are interesting is very subjective and, those chosen, if they are seen in a lot of targets, might not be that informative. Then, there is also the risk of analyzing false positives, instead of the false negatives that could be further down in the results given by "CopraRNA" and "IntaRNA".

In the long run, to confirm that the shorter lengths and most similar sets of homologues do give better targets with better fdr and more interesting profiles, more lengths and sets of different sRNAs should be searched. This problem can be solved when the group "No28" has its "CopraRNA"/"IntaRNA" results analyzed.

With these results obtained, what could be done next? What questions are still unanswered? Well, firstly, the inhibitory compounds seen in the cup-plate method could be studied to understand their structure and to identify them. On the other hand, the samples chosen in the electrophoresis of RNA could go through RNA-seq, to know if they are being expressed. Finally, the conclusion, in the prioritization of targets, that the best results were obtained from the most similar homologues or shortest lengths were only the ones that came from "Selectos242_28" group. This means that, if there were results from the other group (since it has more candidates), the conclusion could be different.

BIBLIOGRAPHY

Bibliography

[1]   Y. Takahashi and T. Nakashima, "Actinomycetes, an Inexhaustible Source of Naturally," *Antibiotics*, vol. 7, no. 2, pp. 1–17, 2018.

[2]   M. Ventura *et al.*, "Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum," *Microbiol. Mol. Biol. Rev.*, vol. 71, no. 3, pp. 495–548, 2007.

[3]   K. Chater and G. Chandra, "The Use of the Rare UUA Codon to Define 'Expression Space' for Genes Involved in Secondary Metabolism, Development and Environmental Adaptation in *Streptomyces*," *J. Microbiol.*, vol. 46, no. 1, pp. 1–11, 2008.

[4]   J. J. Hug, C. D. Bader, M. Remskar, K. Cirnski, and R. Müller, "Concepts and Methods to Access Novel Antibiotics from Actinomycetes," *Antibiotics*, vol. 7, no. 2, pp. 1–47, 2018.

[5]   M. A. Akond, M. N. Jahan, N. Sultana, and F. Rahman, "Effect of Temperature, pH and NaCl on the Isolates of Actinomycetes from Straw and Compost Samples from Savar, Dhaka, Bangladesh," *Am. J. Microbiol. Immunol.*, vol. 1, no. 2, pp. 10–15, 2016.

[6]   L. Terra *et al.*, "A Novel Alkaliphilic *Streptomyces* Inhibits ESKAPE Pathogens," *Front. Microbiol.*, vol. 9, pp. 1–13, 2018.

[7]   M. Fischer, J. Alderson, G. Van Keulen, J. White, and R. G. Sawers, "The obligate aerobe *Streptomyces coelicolor* A3(2) synthesizes three active respiratory nitrate reductases," *Microbiology*, vol. 156, pp. 3166–3179, 2010.

[8]   A. B. Lloyd, "Dispersal of Streptomycetes in Air," *J. Gen. Microbiol.*, vol. 57, no. 1, pp. 35–40, 1969.

[9]   P. Sivalingam, K. Hong, J. Pote, and K. Prabakar, "Extreme Environment *Streptomyces*: Potential Sources for New Antibacterial and Anticancer Drug Leads?," *Int. J. Microbiol.*, vol. 2019, pp. 1–20, 2019.

[10]   D. A. Hodgson, "Primary Metabolism and its Control in Streptomycetes: A Most Unusual Group of Bacteria," *Adv. Microb. Physiol.*, vol. 42, pp. 47–238, 2000.

[11]   S. D. Bentley *et al.*, "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)," *Nature*, vol. 417, no. 2, pp. 141–147, 2002.

[12]   A. Manteca, R. Alvarez, N. Salazar, P. Yague, and J. Sanchez, "Mycelium Differentiation and Antibiotic Production in Submerged Cultures of *Streptomyces coelicolor*," *Appl. Environ. Microbiol.*, vol. 74, no. 12, pp. 3877–3886, 2008.

[13]   R. F. Seipke, M. Kaltenpoth, and M. I. Hutchings, "*Streptomyces* as symbionts: an emerging and widespread theme?," *FEMS Microbiol. Rev.*, vol. 36, no. 4, pp. 862–876, 2012.

[14]   M. Wolánski, R. Wali, E. Tilley, D. Jakimowicz, J. Zakrzewska-Czerwinska, and P. Herron, "Replisome Trafficking in Growing Vegetative Hyphae of *Streptomyces coelicolor* A3(2)," *J. Bacteriol.*, vol. 193, no. 5, pp. 1273–1275, 2011.

[15]   A. Manteca, J. Sanchez, H. R. Jung, V. Schwammle, and O. N. Jensen, "Quantitative Proteomics Analysis of *Streptomyces coelicolor* Development Demonstrates That Onset of Secondary Metabolism Coincides with Hypha Differentiation," *Mol. Cell. Proteomics*, vol. 9, no. 7, pp. 1423–1436, 2010.

[16]   D. A. Hopwood, "Forty years of genetics with *Streptomyces*: From in vivo through in vitro to in silico," *Microbiology*, vol. 145, pp. 2183–2202, 1999.

[17]   R. Kirby, "Chromosome diversity and similarity within the *Actinomycetales*," *FEMS Microbiol. Lett.*, vol. 319, no. 1, pp. 1–10, 2011.

[18]   J. N. Volff and J. Altenbuchner, "Genetic instability of the *Streptomyces* chromosome," *Mol. Microbiol.*, vol. 27, no. 2, pp. 239–246, 1998.

[19]   C. W. Chen, C. H. Huang, H. H. Lee, H. H. Tsai, and R. Kirby, "Once the circle has been broken: dynamics and evolution of *Streptomyces* chromosomes," *Trends Genet.*, vol. 18, no. 10, pp. 522–529, 2002.

[20]   J. N. Kim, Y. Kim, Y. Jeong, J. H. Roe, B. G. Kim, and B. K. Cho, "Comparative Genomics Reveals the Core and Accessory Genomes of *Streptomyces* Species," *J. Microbiol. Biotechnol.*, vol. 25, no. 10, pp. 1599–1605, 2015.

[21]   H. Ikeda *et al.*, "Complete genome sequence and comparative analysis of the industrial

microorganism *Streptomyces avermitilis*," *Nat. Biotechnol.*, vol. 21, no. 5, pp. 526–531, 2003.

[22] M. Medema *et al.*, "The Sequence of a 1.8-Mb Bacterial Linear Plasmid Reveals a Rich Evolutionary Reservoir of Secondary Metabolic Pathways," *Genome Biol. Evol.*, vol. 2, pp. 212–224, 2010.

[23] M. Bibb, "The regulation of antibiotic production in *Streptomyces coelicolor* A3(2)," *Microbiology*, vol. 142, no. 6, pp. 1335–1344, 1996.

[24] S. Peng *et al.*, "Three functional replication origins of the linear and artificially circularized plasmid SCP1 of *Streptomyces coelicolor*," *Microbiology*, vol. 159, pp. 2127–2140, 2013.

[25] A. Craney, S. Ahmed, and J. Nodwell, "Towards a new science of secondary metabolism," *J. Antibiot. (Tokyo).*, vol. 66, no. 7, pp. 387–400, 2013.

[26] B. Rudd and D. Hopwood, "Genetics of Actinorhodin Biosynthesis by *Streptomyces coelicolor* A3(2)," *J. Gen. Microbiol.*, vol. 114, no. 1, pp. 35–43, 1979.

[27] T. Schaberle, A. Orland, and G. M. Konig, "Enhanced production of undecylprodigiosin in *Streptomyces coelicolor* by co-cultivation with the corallopyronin A-producing myxobacterium, *Corallococcus coralloides*," *Biotechnol. Lett.*, vol. 36, no. 3, pp. 641–648, 2014.

[28] J. F. Martín, "Phosphate Control of the Biosynthesis of Antibiotics and Other Secondary Metabolites Is Mediated by the PhoR-PhoP System: an Unfinished Story," *J. Bacteriol.*, vol. 186, no. 16, pp. 5197–5201, 2004.

[29] P. P. Chong *et al.*, "Physical Identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2)," *Microbiology*, vol. 144, pp. 193–199, 1998.

[30] M. Gottelt, S. Kol, J. P. Gomez-Escribano, M. Bibb, and E. Takano, "Deletion of a regulatory gene within the cpk gene cluster reveals novel antibacterial activity in *Streptomyces coelicolor* A3(2)," *Microbiology,* vol. 156, pp. 2343–2353, 2010.

[31] T. Haneishi, N. Kitahara, Y. Takiguchi, and M. Arai, "New Antibiotics, Methylenomycins A and B," *J. Antibiot. (Tokyo).*, vol. 27, no. 6, pp. 386–392, 1974.

[32] K. Jakubiec-Krzesniak, A. Rajnisz-Mateusiak, A. Guspiel, J. Ziemska, and J. Solecka, "Secondary Metabolites of Actinomycetes and their Antibacterial, Antifungal and Antiviral Properties," *Polish J. Microbiol.*, vol. 67, no. 3, pp. 259–272, 2018.

[33] A. J. Mccarthy and S. T. Williams, "Actinomycetes as agents of biodegradation in the environment - a review," *Gene*, vol. 115, pp. 189–192, 1992.

[34] A. P. Tyurin, V. A. Alferova, and V. A. Korshun, "Chemical Elicitors of Antibiotic Biosynthesis in Actinomycetes," *Microorganisms*, vol. 6, no. 52, pp. 1–10, 2018.

[35] S. Sanchez and A. L. Demain, "Metabolic regulation and overproduction of primary metabolites," *Microb. Biotechnol.*, vol. 1, no. 4, pp. 283–319, 2008.

[36] J. F. Martin and A. L. Demain, "Control of Antibiotic Biosynthesis," *Microbiol. Rev.*, vol. 44, no. 2, pp. 230–251, 1980.

[37] M. J. Bibb, "Regulation of secondary metabolism in streptomycetes," *Curr. Opin. Microbiol.*, vol. 8, no. 2, pp. 208–215, 2005.

[38] A. T. Garrison and R. W. Huigens III, "Eradicating Bacterial Biofilms with Natural Products and Their Inspired Analogues that Operate Through Unique Mechanisms," *Curr. Top. Med. Chem.*, vol. 17, no. 14, pp. 1–11, 2017.

[39] J. Wei, L. He, and G. Niu, "Regulation of antibiotic biosynthesis in actinomycetes: Perspectives and challenges," *Synth. Syst. Biotechnol.*, vol. 3, no. 4, pp. 229–235, 2018.

[40] P. N. Tran, M. Yen, C. Chiang, H. Lin, and P. Chen, "Detecting and prioritizing biosynthetic gene clusters for bioactive compounds in bacteria and fungi," pp. 3277–3287, 2019.

[41] B. Hou *et al.*, "LmbU, a Cluster-Situated Regulator for Lincomycin, Consists of a DNA-Binding Domain, an Auto-Inhibitory Domain, and Forms Homodimer," *Front. Microbiol.*, vol. 10, pp. 1–13, 2019.

[42] G. Niu, K. F. Chater, Y. Tian, J. Zhang, and H. Tan, "Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp.," *FEMS Microbiol. Rev.*, vol. 40, no. 4, pp. 554–573, 2016.

[43] D. Shu *et al.*, "afsQ1-Q2-sigQ is a pleiotropic but conditionally required signal transduction system for both secondary metabolism and morphological development in *Streptomyces coelicolor*," *Appl. Microbiol. Biotechnol.*, vol. 81, no. 6, pp. 1149–1160, 2009.

[44] J. F. Martín and P. Liras, "Engineering of regulatory cascades and networks controlling

antibiotic biosynthesis in *Streptomyces*," *Curr. Opin. Microbiol.*, vol. 13, no. 3, pp. 263–273, 2010.

[45]   T. C. Mclean, B. Wilkinson, M. I. Hutchings, and R. Devine, "Dissolution of the Disparate: Co-ordinate Regulation in Antibiotic Biosynthesis," *Antibiotics*, vol. 8, no. 2, pp. 1–17, 2019.

[46]   K. M. Devine, "Activation of the PhoPR-Mediated Response to Phosphate Limitation Is Regulated by Wall Teichoic Acid Metabolism in *Bacillus subtilis*," *Front. Microbiol.*, vol. 9, no. 2678, pp. 1–7, 2018.

[47]   T. Hunter, "Why nature chose phosphate to modify proteins," *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, vol. 367, no. 1602, pp. 2513–2516, 2012.

[48]   F. Santos-Beneit, A. Rodríguez-García, E. Franco-Domínguez, and J. F. Martín, "Phosphate-dependent regulation of the low- and high-affinity transport systems in the model actinomycete *Streptomyces coelicolor*," *Microbiology*, vol. 154, pp. 2356–2370, 2008.

[49]   L. T. Fernández-Martínez, F. Santos-Beneit, and J. F. Martín, "Is PhoR–PhoP partner fidelity strict? PhoR is required for the activation of the pho regulon in *Streptomyces coelicolor*," *Mol. Genet. Genomics*, vol. 287, no. 7, pp. 565–573, 2012.

[50]   J. F. Martín, A. Rodríguez-García, and P. Liras, "The master regulator PhoP coordinates phosphate and nitrogen metabolism, respiration, cell differentiation and antibiotic biosynthesis: comparison in *Streptomyces coelicolor* and *Streptomyces avermitilis*," *J. Antibiot. (Tokyo).*, vol. 70, no. 5, pp. 534–541, 2017.

[51]   A. M. Stock, V. L. Robinson, and P. N. Goudreau, "Two-component signal transduction," *Annu. Rev. Biochem.*, vol. 69, pp. 183–215, 2000.

[52]   A. Rodríguez-García, A. Sola-Landa, K. Apel, F. Santos-Beneit, and J. F. Martín, "Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of glnR, glnA, glnII and amtB expression by the response regulator PhoP," *Nucleic Acids Res.*, vol. 37, no. 10, pp. 3230–3242, 2009.

[53]   S. Kimura, K. Makino, H. Shinagawa, M. Amemura, and A. Nakata, "Regulation of the phosphate regulon of *Escherichia coli*: Characterization of the promotor in the pstS gene," *Mol. Genet. Genomics*, vol. 215, no. 3, pp. 374–380, 1989.

[54]   F. Santos-Beneit and L. T. Fernández-Martínez, "Transcriptional response to vancomycin in a highly vancomycin-resistant *Streptomyces coelicolor* mutant," *Future Microbiol.*, vol. 9, no. 5, pp. 603–622, 2014.

[55]   B. Wanner, "Phosphate signalling and control of gene expression in *Escherichia coli*," in *Metal ions in gene regulation*, 1997, pp. 104–128.

[56]   S. Gottesman and G. Storz, "Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations," *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 12, pp. 1–16, 2011.

[57]   L. S. Waters and G. Storz, "Regulatory RNAs in Bacteria," *Cell*, vol. 136, no. 4, pp. 615–628, 2009.

[58]   E. Holmqvist and E. G. H. Wagner, "Impact of bacterial sRNAs in stress responses," *Biochem. Soc. Trans.*, vol. 45, no. 6, pp. 1203–1212, 2017.

[59]   M. Bouvier, C. M. Sharma, F. Mika, K. H. Nierhaus, and J. Vogel, "Small RNA Binding to 5' mRNA Coding Region Inhibits Translational Initiation," *Mol. Cell*, vol. 32, no. 6, pp. 827–837, 2008.

[60]   M. Bobrovskyy and C. K. Vanderpool, "Regulation of Bacterial Metabolism by Small RNAs Using Diverse Mechanisms," *Annu. Rev. Genet.*, vol. 47, pp. 209–234, 2013.

[61]   C. M. Sharma, F. Darfeuille, T. H. Plantinga, and J. Vogel, "A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites," *Genes Dev.*, vol. 21, no. 21, pp. 2804–2817, 2007.

[62]   K. Papenfort and C. K. Vanderpool, "Target activation by regulatory RNAs in bacteria," *FEMS Microbiol. Rev.*, vol. 39, no. 3, pp. 362–378, 2015.

[63]   E. G. H. Wagner and P. Romby, "Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It," *Adv. Genet.*, vol. 90, pp. 1–76, 2015.

[64]   G. Storz, S. Altuvia, and K. M. Wassarman, "An abundance of RNA regulators," *Annu. Rev. Biochem.*, vol. 74, pp. 199–217, 2005.

[65]   M. C. Lybecker and D. S. Samuels, "Small RNAs of *Borrelia burgdorferi*: Characterizing Functional Regulators in a Sea of sRNAs," *Yale J. Biol. Med.*, vol. 90, no. 2, pp. 317–323, 2017.

[66]   M. Miyakoshi, Y. Chao, and J. Vogel, "Regulatory small RNAs from the 3' regions of bacterial mRNAs," *Curr. Opin. Microbiol.*, vol. 24, pp. 132–139, 2015.

[67] P. Romby and E. Charpentier, "An overview of RNAs with regulatory functions in gram-positive bacteria," *Cell. Mol. Life Sci.*, vol. 67, no. 2, pp. 217–237, 2010.

[68] S. J. Kodumal and D. V Santi, "DNA ligation by selection," *Biotechniques*, vol. 37, no. 1, pp. 34–40, 2004.

[69] "Restriction enzymes & DNA ligase." [Online]. Available: https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-cloning-tutorial/a/restriction-enzymes-dna-ligase.

[70] T. Seecharran, "Evolution in action, High frequency DNA exchange between MRSA in conditions mimicking infection of a human host," 2013.

[71] C. Luft and R. Ketteler, "Electroporation Knows No Boundaries: The Use of Electrostimulation for siRNA Delivery in Cells and Tissues," *J. Biomol. Screen.*, vol. 20, no. 8, pp. 932–942, 2015.

[72] E. van Pelt-Verkuil, A. van Belkum, and J. Hays, *Principles and Technical Aspects on PCR Amplification*. 2008.

[73] C. H. Lin, Y. C. Chen, and T. M. Pan, "Quantification Bias Caused by Plasmid DNA Conformation in Quantitative Real-Time PCR Assay," *PLoS One*, vol. 6, no. 12, pp. 1–10, 2011.

[74] Omega, "E.Z.N.A. Plasmid DNA Mini Kit I." pp. 1–88, 2019.

[75] L. Ciccolini, P. A. Shamlou, J. Ward, and P. Dunnill, "Rheological properties of chromosomal and plasmid DNA during alkaline lysis reaction," *Bioprocess Eng.*, vol. 21, pp. 231–237, 1999.

[76] R. H. Baltz, "Genetic manipulation of antibiotic-producing *Steptomyces*," *Trends Microbiol.*, vol. 6, no. 2, pp. 76–83, 1998.

[77] S. Paranthaman and K. Dharmalingam, "Intergeneric Conjugation in *Streptomyces peucetius* and *Streptomyces* sp. Strain C5: Chromosomal Integration and Expression of Recombinant Plasmids Carrying the chiC Gene," *Appl. Environ. Microbiol.*, vol. 69, no. 1, pp. 84–91, 2003.

[78] A. Bacharach and W. Cuthbertson, "The Cup-Plate Method in Microbiological Assay, with Special Reference to Riboflavine and Aneurine," *Analyst*, vol. 73, pp. 334–340, 1948.

[79] G. Stuchbury and G. Mu, "Optimizing the generation of stable neuronal cell lines via pre-transfection restriction enzyme digestion of plasmid DNA," *Cytotechnology*, vol. 62, no. 3, pp. 189–194, 2010.

[80] Available: https://international.neb.com/.

[81] R. Westermeier, "Gel Electrophoresis," in *Encyclopedia of life sciences*, 2005, pp. 1–6.

[82] K. Motohashi, "Development of highly sensitive and low-cost DNA agarose gel electrophoresis detection systems, and evaluation of non-mutagenic and loading dye-type DNA-staining reagents," *PLoS One*, vol. 14, no. 9, pp. 1–13, 2019.

[83] Z. Research, "Zymoclean Gel DNA Recovery Kit."

[84] A. O. Nwokeoji, P. M. Kilby, D. E. Portwood, and M. J. Dickman, "Accurate Quantification of Nucleic Acids Using Hypochromicity Measurements in Conjunction with UV Spectrophotometry," *Anal. Chem.*, vol. 89, no. 24, pp. 13567–13574, 2017.

[85] E. Karlsson and I. Hirsh, "Ion Exchange Chromatography," in *Protein Purification: Principles, High Resolution Methods, and Applications*, 2011, pp. 93–133.

[86] G. Lucena-Aguilar, A. M. Sánchez-López, C. Barberán-Aceituno, J. A. Carrillo-Ávila, J. A. López-Guerrero, and R. Aguilar-Quesada, "DNA Source Selection for Downstream Applications Based on DNA Quality Indicators Analysis," *Biopreserv. Biobank.*, vol. 14, no. 4, pp. 264–270, 2016.

[87] A. Masotti and T. Preckel, "Analysis of small RNAs with the Agilent 2100 Bioanalyzer," *Nat. Methods*, vol. 3, no. 8, pp. 3–4, 2006.

[88] J. Davies, T. Denyer, and J. Hadfield, "Bioanalyzer chips can be used interchangeably for many analyses of DNA or RNA," *Biotechniques*, vol. 60, no. 4, pp. 197–199, 2016.

[89] "Bioanalyzer RNA Kits & Reagents." [Online]. Available: https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-rna-kits-reagents.

[90] A. Singh, M. Singh, V. Sharma, H. Verma, and K. Arora, "Chapter 13- Molecular Techniques," in *Chemical Analysis of Food: Techniques and Applications*, 2012, pp. 407–461.

[91] L. Garibyan and N. Avashia, "Research Techniques Made Simple: Polymerase Chain

Reaction (PCR)," *J. Invest. Dermatol.*, vol. 133, no. 3, pp. 1–8, 2013.

[92]   G. P. Wagner, K. Kin, and V. J. Lynch, "Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples," *Theory Biosci.*, vol. 131, no. 4, pp. 281–285, 2012.

[93]   A. Rodríguez-García, C. Barreiro, F. Santos-Beneit, A. Sola-Landa, and J. F. Martín, "Genome-wide transcriptomic and proteomic analysis of the primary response to phosphate limitation in *Streptomyces coelicolor* M145 and in a DeltaphoP mutant," *Proteomics*, vol. 7, no. 14, pp. 2410–2429, 2007.

[94]   Y. Jeong *et al.*, "The dynamic transcriptional and translational landscape of the model antibiotic producer *Streptomyces coelicolor* A3(2)," *Nat. Commun.*, vol. 7, pp. 1–11, 2016.

[95]   S. C. Lott *et al.*, "GLASSgo – Automated and Reliable Detection of sRNA Homologs From a Single Input Sequence," *Front. Genet.*, vol. 9, pp. 1–12, 2018.

[96]   M. Raden *et al.*, "Freiburg RNA tools: a central online resource for RNA-focused research and teaching," *Nucleic Acids Res.*, vol. 46, pp. 25–29, 2018.

[97]   P. R. Wright *et al.*, "Comparative genomics boosts target prediction for bacterial small RNAs," *PNAS*, vol. 110, no. 37, pp. 3487–3496, 2013.

[98]   P. R. Wright *et al.*, "CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains," *Nucleic Acids Res.*, vol. 42, pp. 119–123, 2014.

[99]   K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res.*, vol. 35, pp. 61–65, 2007.

[100] "Sample Entrez Gene Record." [Online]. Available: https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/Genes/sample_entrez_gene_record.html.

[101] A. Busch, A. S. Richter, and R. Backofen, "IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions," *Bioinformatics*, vol. 24, no. 24, pp. 2849–2856, 2008.

[102] M. Mann, P. R. Wright, and R. Backofen, "IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions," *Nucleic Acids Res.*, vol. 45, pp. 435–439, 2017.

[103] "DNA Marker Lambda/Pst I." [Online]. Available: https://www.aabiot.com/en/dna-marker-lambda-pst-i.

[104] O. Mueller, S. Lightfoot, and A. Schroeder, "RNA Integrity Number (RIN) – Standardization of RNA Quality Control Application," 2004.

[105] J. Norhazlin *et al.*, "Effect of DNase treatment on RNA extraction from preimplantation murine embryos," *Genet. Mol. Res.*, vol. 14, no. 3, pp. 10172–10184, 2015.