Lysogeny in *Streptococcus dysgalactiae* subsp. *dysgalactiae:* lethargy or failure?

From classical infection approaches to whole-genome sequencing

Mariana Nascimento

Microbiology and Biotechnology Lab (LMB-BioISI|Bugworkers|TecLabs), Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal

A B S T R A C T

Streptococci are mostly commensal bacteria found in warm-blooded animals (including humans), but may also cause localized and systemic infections with severe sequelae. Their vast virulence gene repertoire, in part encoded within mobile genetic elements, greatly contributes to their pathogenic success. Concerningly, cases of streptococci regarded as animal pathogens crossing the barrier to become zoonotic agents have been reported. *Streptococcus dysgalactiae* subsp. *dysgalactiae* (SDSD), an animal pathogen involved in bovine mastitis, seems to be undergoing this process, given its recent involvement in human infections. At the root of this phenomenon may be the high rate of bacteriophage-mediated horizontal gene transfer observed between streptococci, particularly involving the emerging zoonotic agents and known human pathogens.

To test this hypothesis, protocols for bacteriophage induction were performed, producing putative phage lysates which were subsequently used in infection assays, where no productive infection was obtained. Phage presence was then assessed through phage DNA extraction and virion visualization through Atomic Force Microscopy with positive results, albeit phage tails could not be detected. To assess prophage genome integrity, whole-genome third-generation sequencing was employed and putative prophages were detected in all tested SDSD strains, as well as bacteriophage resistance systems and phage-associated virulence factors. The number, the varying degrees of integrity, as well as the array of phage-associated sequences and their homology with sequences found in human pathogens and zoonotic agents, support the initial hypothesis that phage elements not only mediate the cross-talk between streptococci but also ultimately shape their pathogenic potential.

Keywords: *Streptococcus*; prophages; horizontal gene transfer; third-generation sequencing; pathogenicity

1. INTRODUCTION

Streptococci are gram-positive, low G+C content bacteria, first described by Rosenbach, often associated with warm-blooded animals, including humans. Most species establish relationships of a commensal nature with the respective hosts, inhabiting their mucosal surfaces in the oral cavity, upper respiratory tract and gastrointestinal tract among others; however, given the adequate conditions, these microbes can cause both localized and systemic infections (Whiley and Hardie, 2009). This genus comprises human pathogens such as *Streptococcus pyogenes* (the genus type-species) as well as animal pathogens, such as *Streptococcus equi* subsp. *equi*. Recently, streptococci that were regarded as exclusively animal pathogens have crossed the barrier to become zoonotic agents, such as *S. equi* subsp. *zooepidemicus* and *S. dysgalactiae* subsp. *equisimilis* (SDSE) (Murray *et al*., 2009). The remaining *S. dysgalactiae* subspecies, an animal pathogen involved in bovine mastitis, seems to be undergoing the same process: instances of *S. dysgalactiae* subsp. *dysgalactiae* (SDSD) causing disease in humans have been reported. Infections after contact with infected fish and after knee surgery, as well as cases of infected endocarditis caused by this bacterium have been reported (Koh *et al*., 2009; Park *et al*., 2012; Jordal *et al*., 2015). Recent studies showed SDSD cells to have high adherence and internalization to human cells, strengthening the hypothesis that *S. dysgalactiae* subsp. *dysgalactiae* isolates are becoming able to infect a human host (Roma-Rodrigues *et al*., 2016).It is known that the *Streptococcus* genus is quite diverse, in part due to horizontal gene transfer and recombination of genes between its members (particularly in pyogenic β-hemolytic streptococci, a group that encompasses both *S. pyogenes* and *S. dysgalactiae*) (Whiley and Hardie, 2009). Although genetic transfer in streptococci can be mediated by several mechanisms, transduction might be particularly relevant since bacteriophages have been detected in considerable

proportion, especially among GAS, taking up to 12% of the bacterial genome (Canchaya *et al*., 2003). Because phages can encode virulence factors, they contribute to the organism's pathogenicity and thus play a role in adaptation of the microbe to different hosts and different environmental pressures; moreover, the contribution of phage presence has been recognized in the generation of streptococcal strains with increased pathogenic potential (Whiley and Hardie, 2009). Ongoing acquisition of phages between a recognized pathogen and a largely commensal bacterium may not only have drastic effects on the overall population structure of the genus but also result in rapid changes to the pathogenic potential of SDSE and SDSD (Davies *et al.*, 2007). The evolutionary dynamics of pathogenic bacteria are one of the many examples that highlight the dual outcomes of phage presence within a bacterial host, challenging the view of bacteriophages as simply parasitic elements (Cumby *et al*., 2012).

Bacteriophages can carry out their life cycle in different manners, each influencing the bacterial counterpart in diverse ways. Lysogenic phages, however, may prove interesting in matters of HGT, given their extended impact upon the host: these phages integrate into the bacterial chromosome (or acquire plasmid form) to gain control of the host's metabolism, propagating themselves passively as an element of the bacterial genetic patrimony. Nevertheless, they retain the ability to revert to a lytic mode of infection under stressful conditions (Fortier and Sekulovic, 2013). The most common are the tailed phages, a category of viruses with double-stranded DNA and an exclusively proteinaceous capsid. Tailed phages belong to the *Caudovirales* order, which is composed by the *Myoviridae*, *Siphoviridae* and *Podoviridae* families (Abedon, 2009). Generally, phages with siphoviral morphotypes (the most common among the Caudovirales, and thus the most common type of bacteriophages) have synteny among the genes that encode for the virion structure and genes with assembly functions. Despite this conserved arrangement, phages still contain variable regions with sequences of unknown function; in fact, it is estimated that phages might represent the largest reservoir of unexplored genes (Hatfull and Hendrix, 2011).

The conserved arrangement as well as the considerable untapped potential that lies within phage genomes highlights the differences between bacterial and viral evolution. However, although quite different, the evolutionary strategies of one intervenient affect the other in a multitude of ways. Ultimately, a bacterial population is shaped both by the predatory action of phages and by the presence of phage-encoded genes which may enhance bacterial survival, help conquer new ecological niches and maintain previously acquired ones (Cumby *et al*., 2012). Phages and bacteria are then involved in co-evolution cycles that must feature the emergence of phage-insensitive hosts, responsible for preserving bacterial lineages, and the emergence of counter-resistant phages able to threaten new bacterial strains. The need for bacteria to develop defensive action against phage predation has given rise to a set of mechanisms collectively called the "bacteriophage resistome". The resistome includes adsorption resistance mechanisms (which work by diminishing the contact between the viral particle and its host, through loss of receptor molecules, for example), restriction mechanisms (which cause the death of phage particles but preserve the host) and abortive infection mechanisms (which result in the death of both the bacteriophage and the host) (Hyman and Abedon, 2010). Among these, there are some of particular relevance such as abortive infection mechanisms and two restriction mechanisms: restriction-modification systems and CRISPR/Cas systems, well known tools used in genetic engineering (Labrie *et al*., 2010).

Detection of bacteriophages and well as the gauging of their functionality can be carried out using several different techniques: classical infection assays allow the determination of phage infectivity, inferring phage presence; phage DNA extraction and virion microscopy visualization allow a look at the extracellular phase of the phage's life cycle; bacterial whole-genome sequencing (WGS) focuses on the prophage state of the viral life cycle, when integration into the bacterial chromosome occurs. Each technique is not, of course, without its limitations, and as such, they must be integrated to obtain the most information about bacteriophage content inside a given host. The usage of techniques that provide information about phage integrity close to physiological conditions, such as Atomic Force Microscopy, or to more accurately replicate prophage synteny, as does long-read third-generation sequencing, may help clarify results obtained from more classical phage study approaches (Fortier and Moineau, 2009).

2. MATERIALS AND METHODS

2.1 INDUCTION/INFECTION ASSAYS

Bacterial strains

For the first tasks of the present work, four strains from *S. pyogenes* and five strains from

S. dysgalactiae subsp. *dysgalactiae* were used to produce phage lysates. The four SPYO strains (encoded as GAP8, GAP58, GAP88 and GAP826) originate from clinical samples collected from human hosts; three out of the four SDSD strains (VSD5, VSD9 and VSD13) are also of clinical/subclinical origin, and were collected from bovine hosts; one SDSD strain (encoded as GCS-Si) is of clinical origin and was collected from a human host in Singapore, who developed cellulitis upon contact with infected fish (Koh *et al*., 2009). Moreover, two strains of clinical origin in bovine hosts (VSD17 and VSD19) were used as host cells for infection assays. Strains were selected based on their virulence gene repertoire.

Growth conditions and culture media

Bacteria were recovered from cryopreserved cultures maintained in THYE - Todd-Hewitt (BD) supplemented with 1% yeast extract (Oxoid) - with 20% (v/v) glycerol at -80°C. To potentiate growth and verify their hemolysis features, 10 µL of the preserved cultures were streaked onto COS (Columbia Agar with Sheep Blood Plus) from Oxoid; inoculated plates were incubated overnight at 37°C. In latter experiments, bacteria for liquid pre-inocula were taken either from the COS plates or from the cryopreserved cultures and added to one of the following culture media: THYE and M17YE (M17 (BD) supplemented with 1% yeast extract). For standard solid plate growth, each medium was supplemented with 1.5% bacteriological agar (BIOKAR Diagnostics). Although THYE is a standard medium for the growth of streptococci, it contains some of the harmful components for phage infection, namely sodium carbonate and disodium phosphate; in turn, M17YE does not contain either substance (containing disodium-β-glycerophosphate instead, which does not harm the process) and so the two media were used in induction assays.

Bacteriophage induction assays

For phage induction assays, liquid bacterial cultures from all 8 strains were grown overnight at 37°C in THYE and M17YE. Overnight cultures were diluted 1:100 in the fresh corresponding culture medium (for a total volume of 20 mL per culture) and allowed to grow until OD_{600} ≈ 0.2-0.25, to ensure induction occurred in the early exponential growth phase. Mitomycin C (Sigma-Aldrich) was then added to each culture to reach a final concentration of either 0.2 µg/mL or 0.5 µg/mL. Cultures were then incubated at 37°C for 4 hours, with samples being collected at the 2h, 3h and 4h time points. Samples were then centrifuged at 1500 × *g* and 4°C for 15 minutes (using an Eppendorf 5810 R centrifuge). The supernatant was collected and filtered using 0.45 µm pore membrane filters (Sarstedt) and the resulting filtrate was diluted 1:1 in SM buffer 2x (0.06% gelatin, 20mM NaCl, 16 mM $MgSO_4$, 100mM Tris-HCl) and stored at 4°C.

Infection assays

A. Spot assays

Lysates were diluted up to 10^{-4} in SM buffer. Cultures of the VSD17 and VSD19 isolates (host strains) were incubated overnight at 37°C in M17YE. Overnight cultures were then diluted 1:100 in fresh M17YE (for a total volume of 50 mL per culture) and incubated at 37°C until OD_{600} ≈ 0.8. Plates of M17YE (with 1.5% agar and supplemented with $CaCl_2$) were previously prepared, as well as 5 mL aliquots of molten M17YE (with 0,5% agar and supplemented with $CaCl_2$) which were kept stabilized in a 45°C water bath. 200 µL of the host culture were then mixed with the 5 mL aliquot of molten media, which was then plated upon the correspondent bottom layer 1.5% agar medium and left to dry. In each plate, half of the original lysates for a given strain, along with their respective dilutions, were spotted (each spot corresponding to 10 µL of lysate) in a chess pattern, to avoid contact between spots and left to dry. For each host strain, two control plates were made: a plate containing only the host culture and a plate in which the phage lysate was substituted for a solution of mitomycin C in SM buffer at the highest concentration used in the induction assays (0.5 µg/mL).

B. Incorporation assays

Incorporation assays were performed using single-layer agar plates with molten media and only the original phage lysates were tested. Similarly to spot assays, cultures of the VSD17 and VSD19 isolates (host strains) were incubated overnight at 37°C in M17YE. Overnight cultures were then diluted 1:100 in fresh M17YE (for a total volume of 50 mL per culture) and incubated at 37°C until OD_{600} ≈ 0.8. Meanwhile, 5 mL aliquots of M17YE molten media (with 0.5% agar and supplemented with $CaCl_2$) were kept stabilized in a 45°C water bath. 200 µL of the overnight host culture were then mixed with 10 µL of an original lysate and the 5 mL of culture medium and poured onto a small Petri dish. Controls were the same as those used for the spot assay.

C. Crossed assays

Instead of using strains VSD17 and VSD19, the 8 strains used in phage induction assays (GAP8, GAP58, GAP88, GAP826, VSD5, VSD9, VSD13 and GCS-Si) were used as hosts. Cultures of the host strains were incubated overnight at 37°C in M17YE. Overnight cultures were then diluted

1:100 in fresh M17YE (for a total volume of 50 mL per culture) and incubated at 37°C until OD_{600} ≈ 0.8. To reduce the number of plates produced, lysates from the same strain were mixed in equal parts (mixing a total of 12 lysates, with a volume of 70 µL each). The resulting 8 mixed lysates (ML) were tested against all 8 hosts in an incorporation assay, in which 200 µL of host culture were mixed with 60 µL of the ML and 5 mL of growth medium. The same controls from previous infection experiments were applied.

D. Liquid medium assays

Infection experiments were also carried out in liquid medium, using strains VSD17 and VSD19 as hosts. Mixed lysates (total volume of 1 mL per mixed lysate) for each induced strain were prepared. Liquid bacterial cultures (with a volume of 100 mL) from the two host strains were grown overnight at 37°C in M17YE broth. The overnight cultures were then diluted 1:100 in fresh M17YE broth (total volume of 50 mL per culture) and incubated until OD_{600} ≈ 0.2-0.25. At this point, each culture was infected with a ML and then checked hourly to assess bacterial lysis.

Phage elution and purification

In case of putative phage plaque formation, isolated plaques were extracted from the plate and placed in 200 µL of SM buffer 1x (0,03% gelatin, 10mM NaCl, 8 mM $MgSO_4$ and 50 mM Tris-HCl). As for plates with possible confluent lysis, the entire plate was flooded with 2mL of SM buffer 1x and left to elute for 4 hours; the liquid was then collected, filtered through a 0.45 µm pore membrane filter (Sarstedt) and stored at 4°C. Resulting phage elutes were then tested in spot assays and incorporation assays.

2.2 PHAGE DNA EXTRACTION AND VIRAL PARTICLE VISUALIZATION

Bacterial strains

Infection assay results from the Strep project were revisited and strains with diverse viral infection profiles, as well as virulence gene repertoires, were selected. Because the main focus of the project is to investigate HGT from *S. pyogenes* to *S. dysgalactiae* subsp. *dysgalactiae*, the presence of phage particles extracted from SDSD isolates takes priority. Consequently, 4 SDSD strains were carried over from the first phase of this work - VSD13, VSD17, VSD19 and GCS-Si – and a new strain was added - VSD4. This new strain is similar to other "VSD" encoded ones, in that it is also of clinical/subclinical origin, and was collected from a bovine host. A strain of *Escherichia coli* (*E. coli* K12 MG1655) was used, along with the T7 bacteriophage, as a positive control for these experiments.

Growth conditions and culture media

SDSD strains were recovered from cryopreserved cultures maintained in THYE - Todd-Hewitt (BD) supplemented with 1% yeast extract (Oxoid) - with 20% (v/v) Glycerol at -80°C. For subsequent experiments, strains were incubated overnight at 37°C in M17YE broth or M17YE agar (supplemented with 1,5% Bacteriological Agar (BIOKAR Diagnostics)). The *E. coli* strain was recovered from cryopreserved cultures maintained in NB (Nutrient Broth (BIOKAR Diagnostics)) with 20% (v/v) Glycerol at -80°C. For subsequent experiments, the strain was incubated overnight at 37°C in NB or NA (Nutrient Agar – NB supplemented with 1,5% Bacteriological Agar (BIOKAR Diagnostics)).

Modified phage induction assay

For phage induction assays performed in this stage, liquid bacterial cultures from all 5 SDSD strains were incubated overnight at 37°C in M17YE broth. Overnight cultures were diluted 1:100 in fresh M17YE (for a total volume of 400 mL per culture) and allowed to grow until OD_{600} ≈ 0,2-0,25, to ensure induction occurred in the early exponential growth phase. Mitomycin C (Sigma-Aldrich) was then added to each culture to reach a final concentration of 0,5 µg/mL and cultures were then incubated overnight at 37°C to allow lysis. Crude lysates obtained from this procedure were then used for phage DNA extraction and AFM sample preparations.

For the *E. coli* strain, a similar procedure was followed with the adequate culture medium, but instead of mitomycin C, 100 µL of a highly concentrated T7 phage solution were added.

Bacteriophage DNA extraction

200 mL of the previously obtained crude lysate were treated with DNase I (Sigma-Aldrich) and RNase A (Sigma-Aldrich) with final concentrations of 5 µg/mL and 2 µg/mL, respectively, and incubated for 2h at 37°C. Then, NaCl (Duchefa Biochemie) was added to a final concentration of 1M and lysates were agitated and incubated in ice for 1h. Cell residues were deposited through centrifugation – 15000 × *g* and 4°C for 45 minutes (using a Beckman J2-21 centrifuge equipped with the Beckman JLA-16.250 rotor) – and the supernatants were transferred to new tubes. Phages were then concentrated by precipitation with 10% (w/v) PEG8000 (Sigma-Aldrich) overnight at 4°C. After centrifugation (15000 × *g* and 4°C for 25 minutes), the

resulting pellet was resuspended in 5 mL in SM buffer (0,03% gelatin, 10mM NaCl, 8 mM $MgSO_4$ and 50 mM Tris-HCl). PEG was extracted by adding an equal volume of a 1:1 phenol/chloroform mixture (Sigma-Aldrich) and centrifuging at 4020 × *g* and 4°C for 15 minutes (using an Eppendorf 5810 R centrifuge). The aqueous phase was transferred to a new tube and to it were added: SDS to a final concentration of 0.5%, EDTA pH 8.0 to a final concentration of 0.02 mol/L and proteinase K (Invitrogen) to a final concentration of 0.05 mg/mL. Lysates were then incubated at 37°C for 1h. Phenol extraction was performed by adding 1 vol. of a 1:1 phenol/chloroform mixture, centrifuging at 4020 × *g* and 4°C for 15 minutes, then adding 1 vol. of a 24:1 chloroform/isoamyl alcohol mixture (Carlo Erba Reagents) and centrifuging again at 4020 × *g* and 4°C for 15 minutes. Subsequently, phage DNA was mixed with 1 vol. of isopropanol and left to precipitate overnight at 4°C. In the following day, the samples were centrifuged at 3000 × *g* and 4°C for 10 minutes, washed with 70% (v/v) ethanol and resuspended in 50 µL of TE buffer (10mM Tris, 1 mM EDTA; pH 8.0). Phage DNA (30 µL of each sample) was then submitted to electrophoresis in a 0.8% (w/v) agarose (Invitrogen) gel, with 0.5X TBE buffer (40 mM Tris; 45 mM Boric acid; 1 mM EDTA; pH 8.3) and a constant voltage of 4 V/cm for 1h. DNA was also quantified using the Qubit 2.0 Fluorometer (Invitrogen) with the dsDNA High-Sensitivity Kit according to provider's instructions.

Atomic Force Microscopy sample preparation

Atomic Force Microscopy was carried out in a Multimode 8 HR produced by Bruker, using Peak Force Tapping mode. All measurements were performed by placing a drop (ca. 50 µL) of each sample onto freshly cleaved mica for 20 min, rinsing with ultrapure water and drying with pure N_2. The images were acquired in ambient conditions (ca. 21°C), using etched silicon tips with a spring constant of ca. 0.4 N/m (SCANASYST-AIR, Bruker), at a scan rate of ca. 1.3 Hz.

2.3 THIRD GENERATION WHOLE-GENOME SEQUENCING

Bacterial strains, growth conditions and culture media

The same SDSD strains from section 2.2 were recovered from cryopreserved cultures maintained in THYE – Todd-Hewitt (BD) supplemented with 1% yeast extract (Oxoid) - with 20% (v/v) glycerol at -80°C. For subsequent experiments, strains were incubated overnight at 37°C in M17YE broth, generally in a total volume of 250 mL.

Genomic DNA extraction

DNA extraction was performed using the Wizard® Genomic DNA Purification Kit (Promega), with modifications to the protocol for isolation of genomic DNA from gram-positive bacteria: a lysis cocktail consisting of lysozyme, in a final concentration of 10 mg/mL (Sigma-Aldrich), and mutanolysin, in a final concentration of 0,08 mg/mL (Sigma-Aldrich) was used; final DNA resuspension was performed in 30 µL of nuclease-free water.

Genomic DNA quality control

DNA was quantified using the Qubit 2.0 fluorometer (Invitrogen) with the dsDNA High-Sensitivity Kit according to manufacturer's instructions.

Absorbance scans of genomic DNA were also determined. 5 µL of each sample were added to 495 µL of nuclease-free water. The samples' absorbance scans from 200 nm to 400 nm were then taken using a UNICAM UV2 Spectrometer paired with the Vision V3.32 software. Scans were compared to that of a purified λ phage DNA stock (Invitrogen).

Genomic DNA (30 µL of each sample) was then submitted to electrophoresis in a 0.8% (w/v) agarose (Invitrogen) gel, with 0.5X TBE buffer (40 mM Tris; 45 mM Boric acid; 1 mM EDTA; pH 8.3) and a constant voltage of 5.3 V/cm for 1h.

Library preparation

Third-generation whole genome sequencing was carried out using Oxford Nanopore Technologies' MinION, a portable nanopore sequencer. The two latest chemistries available for this technology were employed. As a result, both 1D (SQK-LSK108 kit with a FLO-MIN106 flow cell for the MinION MK 1B) and $1D^2$ (SQK-LSK308 kit with FLO-MIN107 flow cells for the MinION MK 1B) library preparation protocols were used during sequencing experiments, with minor modifications to the original ONT protocols: 2-2.5 µg of genomic DNA were used, instead of 1-1.5 µg; incubation times during end-repairing of genomic DNA were doubled; the for the $1D^2$ protocol, incubation times with the $1D^2$ adapter and the Barcoded Adapter Mix were tripled.

MinION flow cell set-up

Preparation of the MinION and respective flow cell was performed according to the manufacturer's instructions.

Nanopore sequencing data analysis

Sequencing data analysis was performed using both local software and server-based tools. All local software was installed according to the developer's instructions and ran

on a command-line based interface on an Ubuntu System 14.04 LTS. Basecalling was performed after the sequencing run was completed. Albacore v.1.1.2 was used for both R9.4 and R9.5 data; R9.4 requires linear basecalling only while R9.5 implies an additional step where linear basecall results are recalled, detecting potential read pairs. For R9.5 data, only the $1D^2$ paired reads were used downstream. Afterwards, NanoPlot v0.17.4 was used to assess statistics of sequencing data. Japsa v1.7 was then used for read filtering, excluding reads with a quality score (QScore, an indication of how well the raw data fits into the basecalling model that does not fit the usual Phred error rates) below 10 or smaller than 1000 bp in length. The QScore minimum was defined considering the widely used live-basecalling platform Metrichor and a threshold of 10 was established. As for read length, setting a minimum of 1000 bp was defined. Next, assembly was performed using Canu v1.5 (Koren *et al*., 2016) in its full pipeline version (comprising read correction, read trimming and unitig construction steps) with standard parameters for uncorrected nanopore reads; the expected genome size – a parameter required for assembly - was estimated to be around 2.2 Mb based on available SDSD and SDSE genomes on NCBI. For quality checkpoints, QUAST v4.5 (Gurevich *et al*., 2013) and the MUMmer v3.23 function “dnadiff” (Kurtz *et al*., 2004) were used to compare ongoing assemblies to available SDSD and SDSE reference genomes: *Streptococcus dysgalactiae* subsp. *dysgalactiae* strain ATCC 27957 and *Streptococcus dysgalactiae* subsp. *equisimilis* strain ATCC 12394 (NCBI accession numbers: NZ_AEGO00000000.1 and NC_017567.1, respectively). Following the assembly evaluation, Nanopolish v0.7.0 (Loman *et al*., 2015) was used for polishing with default parameters, through the “variants --consensus” subprogram. Usage of the Nanopolish algorithm implies previous indexing and aligning using the Burrows-Wheeler Aligner (BWA) (Li, 2013) as well as Sequence Alignment/Map tools (SAMtools) (Li *et al*., 2009) for necessary file format conversions. A second quality checkpoint was performed. Polished genomes were then annotated using the RAST online server (Aziz *et al*., 2008) with the Classic RAST annotation scheme while enabling the frameshift fix and automatic error fix. Subsequently, phage prediction was performed using PhiSpy v3.2 (Akhter *et al*., 2012) without a specified training set and Prophinder v0.4 (Lima-Mendez *et al*., 2008) with default parameters. Genome visualization, as well as phage-region sequence retrieval was performed using the Integrative Genomics Viewer v2.3 (IGV) Java application.

3. RESULTS & DISCUSSION

3.1 INDUCTION/INFECTION ASSAYS

Results from the previous Strep project suggested productive infection was possible within this selection of *Streptococcus* strains. However, reproducibility of such results was a problem during the first project, and to assess whether these inconsistencies were due to abiotic factors influencing the infection process, the first phase of this work consisted in testing different induction and infection conditions. Induction assays occurred as expected, with cultures responding appropriately to the introduction of mitomycin C. The obtained lysates were then used in different infection assays. Putative phage plaques were detected in all three types of molten media infection assays; strains VSD5, VSD9 and GCS-Si had seemingly positive results in more than one type of infection experiment and strain VSD17 was the only host in spot and incorporation assays to register possibly positive results. Plaques were consequently eluted and purified. After purification, putative phage elutes were re-tested through spot, incorporation and cross assays; yet, productive phage infection was never achieved. The seemingly negative results across all attempted approaches suggested that no bacteriophages with plaquing ability were present. However, absence of plaque-forming ability is not necessarily equivalent to absence of a productive infection and broth-based host range determination might help determine whether productive infection is really occurring (Hyman and Abedon, 2010). Assays in liquid growth media were also performed, but negative results persisted and the putative phages exhibited inability to clear liquid cultures. If present, isolated phages were unable to conduct a productive infection. Even though not all possible experimental conditions were exhausted during the first phase of this work, this approach proved to be extremely time-consuming and led to the conclusion that obtaining productive lysogenic particles from this collection did not seem to be possible. Yet, it provided no information on whether bacteriophages were or not present in the obtained lysates, for lack of infection productivity does not equal absence of phages. Considering these results, the most suitable approach seems to be the confirmation of both the presence and integrity of phage particles themselves, rather than assessing their functionality through classic infection experiments.

3.2 PHAGE DNA EXTRACTION AND VIRAL PARTICLE VISUALIZATION

The presence of fragments similar to the λ phage DNA and the T7 phage DNA controls suggests that induction experiments were successful and the SDSD strains do contain prophage sequences integrated in their genomes. Furthermore, it suggests that they capable of excision from the bacterial genome and successful encapsidation. Phage DNA samples were quantified using the Qubit 2.0 Fluorometer and yielded the following results: T7 phage - 33.2 ng/µL; VSD4 – 150.8 ng/µL; VSD13 – 71.6 ng/µL; VSD17 - 26.2 ng/µL; VSD19 – 31.6 ng/µL; GCS-Si - 21 ng/µL. Although these results confirm phage presence in SDSD strains, with phage genome fragments appearing within the expected sizes (average sizes of *Siphoviridae* members), the phages' physical integrity remains unknown. To assess the physical structure of phage particles, PEG precipitated samples were viewed using Atomic Force Microscopy. To this end, three out of the six lysates were chosen to undergo AFM: the T7 phage lysate (which served as a positive control), and the VSD13 and VSD17 lysates (which had intermediate concentration values expected to be more suited for this technique). As can be seen in **Fig. 1**, globular structures peaking in height were detected across all samples. To assess their size, several sections from each 2D image were inspected and measured using the NanoScope Software. These structures were consistent in size (averaging at about 60 nm in diameter) as well as morphology and their abundance in the samples seemed to reflect that of phage DNA. Because the samples are mounted onto a hydrophilic surface, slight deviations from the canonic icosahedral structure and TEM-obtained dimensions (expectable capsid diameters are around 50 nm, although sizes do vary) are predictable – the adherence of phage capsid proteins to the surface may cause them to appear larger and to lose their shape. The long period of exposure to PEG (a highly hydrophilic compound) the samples were subjected to can also affect capsid shape. Irregularities in the background are due to the complexity of the sample, which still contains leftover culture medium, PEG8000 and SM buffer. Proteins and other compounds present will adhere to the hydrophilic support and create irregularities in the surface. Although washing steps (applied to sample VSD13) did contribute to eliminate this effect, dilution of samples in a cleaner buffer is advised, to both adjust concentration and get rid of background irregularities, resulting in clearer images. Phage tails could not be observed in any of the samples submitted to AFM. While for the T7 phage this could just be due to its morphology - T7 is a member of the *Podoviridae* family, characterized by very small non-contractile tails – the same does not apply to SDSD samples, given that streptococci are most commonly infected by *Siphoviridae* phages, with long flexible non-contractile tails. Absence of phage tails in AFM images could either be an artifact caused by the lack of sample purification procedures and prolonged exposure to

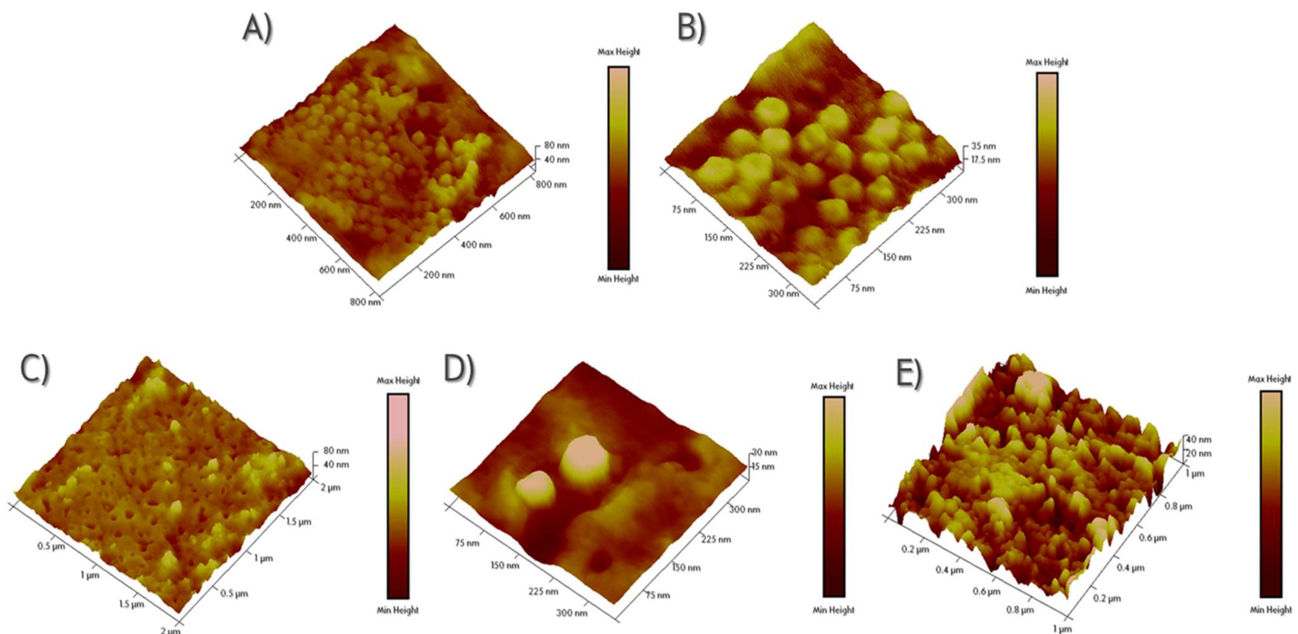

Fig. 1 - AFM 3D images. Images **A)** and **B)** correspond to the VSD13 sample; images **C)** and **D)** correspond to the VSD17 sample; image **E)** corresponds to the T7 phage control sample.

PEG or by a genomic abnormality rendering bacteriophages uncappable of synthesizing or correctly assembling tails. Even assuming that virions are indeed intact, the answer as to why these viral particles are incapable of conducting successful infection may still lie in a genomic approach, by looking not only at the phages' genomes, but also their bacterial counterparts.

3.3 THIRD GENERATION WHOLE-GENOME SEQUENCING

The success of nanopore sequencing runs seems to depend on quite a few factors, with DNA quality being of major influence. Overall, the most successful DNA extraction was that of strain GCS-Si, both in terms of DNA quality and yield; strains VSD13 and VSD19 follow, with reasonably high quality, while strains VSD4 and VSD17 are lower in quality. Runtimes and the resulting coverages seem to differ substantially between 1D and 1D^2 sequencing protocols, with strain VSD17 having the longest sequencing run out of all 5 strains (16 h 17 min) and producing the least amount of coverage (298 x). Interestingly, strain VSD19 registers the smallest runtime and the highest coverage. Besides DNA quality, flow cell state also factors in to the success of sequencing runs and the number of functional pores has been previously found to directly influence data production. Remarkably, despite the low number of available pores, 1D^2 sequencing runs still produced a considerable amount of data, especially when compared to the 665 Mb produced for strain VSD17 through 1D sequencing. Strain VSD13 registers the minimum amount R9.5 sequencing data, at 1.5 Gb, while strain VSD19 had the biggest yield, at about 2.1 Gb.

Filtered data subsets were then used for genome assembly and polishing. Albeit only a small fraction of the obtained data was featured in these subsets, it sufficed to assemble reads into one single contig representing chromosomal DNA in all 5 assembly experiments. During polishing, the total data from each sequencing run is used to polish the previously obtained draft assembly, calculating an improved consensus sequence. However, due to the higher error-rate of nanopore sequencing, both indels and single-nucleotide polymorphisms can only be corrected up to a point. Overall, polishing was considered to improve the assembly, and as such, polished assemblies were used for annotation and phage prediction. As depicted in **Fig. 2**, strains VSD13 and GCS-Si appeared to have the most striking differences from their respective references, with pronounced inverted segments. Interestingly, strain VSD13 is also the subject of *in vitro* and *in vivo* pathogen-host assays in the Strep-hosp project and it has been found to hold remarkable pathogenic potential on *in vivo* assays in zebra fish as well as *in vitro* infection experiments with keratinocytes (Roma-Rodrigues *et al*., 2016). Strain GCS-Si, as previously mentioned, was isolated from a human host who developed cellulitis after contacting with infected fish (Koh *et al*., 2009).

Polished assemblies were subsequently annotated using RAST. Sequencing runs were successful enough to yield over 75% annotated proteins for all strains. Between 47-50% of annotated proteins were covered by RAST subsystems, with slight variations in the number of subsystems ticked for each strain (between 321 and 329). Annotated genome assemblies were then submitted to prophage prediction, using two different tools: Prophinder and PhiSpy, and only the consensus between the two was considered for further analysis, and its results are represented in **Fig 3**. Strain GCS-Si presents the most phage content, followed by strain VSD13, strains VSD19 and VSD17 and finally strain VSD4. There seems to be some correlation with the results from the alignment of assemblies against their closest reference, where strains GCS-Si and VSD13

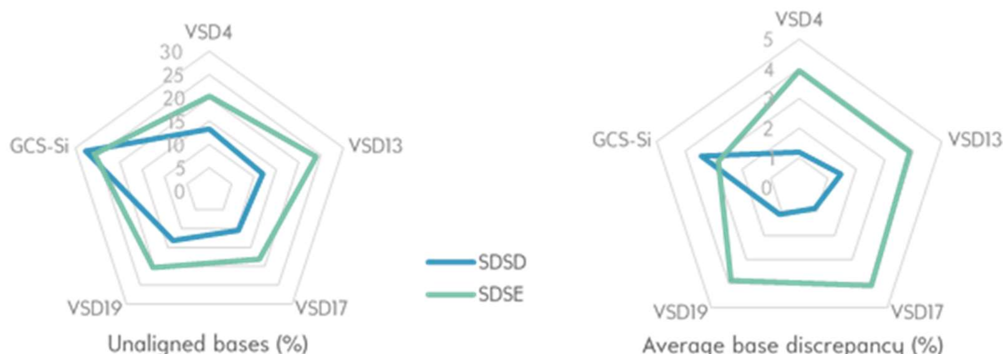

Fig. 2– Assembly discrepancies with S. dysgalactiae subsp. dysgalactiae (SDSD) and S. dysgalactiae subsp. equisimilis (SDSE) reference genomes. Disparities between the five polished assemblies and the SDSD and SDSE reference genomes are represented in terms of the percentage of overall unaligned bases as well as the percentage of discrepant bases between the reference and a given assembly within aligned sequence blocks.

diverged the most. Most putative prophage sequences have sizes within the expected range for *Siphoviridae*, with some of its smallest members scoring about 21000 bp in size (Hatfull and Hendrix, 2011). Some of the predictions on strains GCS-Si and VSD19, however, are much smaller and may represent phage remnants present in their respective genomes.

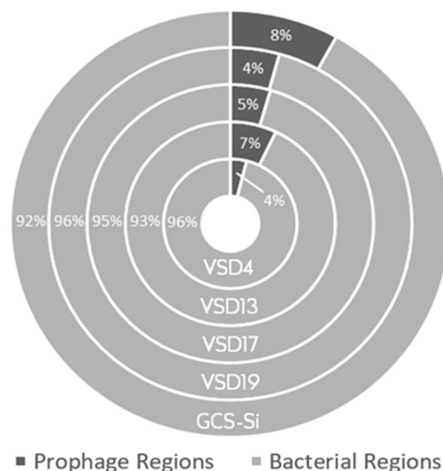

Fig. 3- Consensual prophage content in bacterial genome assemblies.

To better comprehend the phage-host interplay within the five sequenced strains, annotation analysis, as well as sequence homology searches were performed both on detected prophage sequences as well as bacteriophage resistome associated regions. Assessment of phage sequence completeness was performed by scanning annotation for all expected phage modules, as well as checking for homology with functional bacteriophages (Canchaya *et al*., 2003). Strain VSD4 appears to harbor only functional sequences; strain VSD13, although with less certainty, should also serve as host to functional phages only. Strain VSD19 has two functional prophages and one sequence that is most likely a phage remnant, considering these results and the sequence size. Strain GCS-Si's phage patrimony appears to encompass half functional and half defective phages. Lastly, strain VSD17 appears to have no functional phage sequences.

As for these elements' role in bacterial pathogenicity towards humans, strains VSD13 and GCS-Si would be the most affected ones, since they report the most sequences related to human pathogens or zoonotic agents (*S. pyogenes, S. agalactiae, S. suis, S. dysgalactiae* subsp. *equisimilis*), followed by strains VSD17 and then VSD4 and VSD19. These predictions agree with previously mentioned findings about the increased virulence of strains GCS-Si and VSD13. Even if these sequences exist only as genome-integrated phage remnants, they can still impact host fitness if the virulence genes present prove to be functional.

Integrity of resistome-associated sequences, on the other hand, was confirmed by checking annotation files against expected system structures. Strain VSD4 contains a CRISPR I-C, a CRISPR II-A and an AbiG abortive infection system; strain VSD13 contains only the AbiG system; strain VSD17 contains a CRISPR I-C, a CRISPR II-A and a restriction-modification system; strain VSD19 shows the same repertoire as strain VSD4 and finally strain GCS-Si contains two restriction-modification systems and an AbiG system. However, because prophages genomes are flexible, their evolution in response to the selective pressures of resistance mechanisms is fast, meaning that no resistance mechanism is universally efficient. As such, the best defensive approach maybe the rotation between different mechanisms and no fixed combination of resistome sequences outperforms others indefinitely (Durmaz and Klaenhammer, 1995). Attempts to theoretically predict strain resistance to bacteriophages from this data alone are then limited. For example, strains VSD4 and VSD19 share the same bacteriophage resistome and nonetheless, strain VSD4 was found to be resistant to all bacteriophages in infection assays, while strain VSD19 acted as a successful host for infection in some of the experiments performed during the first Strep project.

WGS was performed on SDSD strains to answer the question left AFM results: is the lack of productive phage infection caused by defective phage tails or other factors?

Sequencing results revealed indeed putative prophage sequences lacking tail components (and additional functional modules, in some cases), as well as sequences appearing to be fully functional at a genomic level, attributing lack of productive infection to phage defectiveness as well as lack of phage counter-resistance to bacterial defenses. Performing WGS on the tested strains allowed a glimpse at the host's side of phage infection, stressing the complexity of the phage-host evolutionary arms-race.

4. CONCLUSIONS

Most of all, data points towards the previously posed hypothesis that crosstalk between known streptococcal human pathogens, zoonotic agents and SDSD strains does occur and can indeed enhance their pathogenic potential towards new mammalian hosts. It is not the the potential of each employed approach by itself, but rather the integration of both classical and computational-based methodologies

that shows true promise in the characterization of phage-host dynamics, and in unveiling its complexity. Far more than a question of lethargy of infection or failure to infect, bacteriophage interactions investigated during this dissertation proved to be a mix of both at the very least. However, such an assessment was only possible when looking at different sets of results as a whole. Although sequencing data represents a vast repository of information, it should be a stepping stone in the study of phage-bacterium interactions. Because of its inherent versatility, this data can also be exploited for purposes other than the study of phages, representing a substantial asset in the study of these SDSD strains.

5. ACKNOWLEDGEMENTS

Support from FCT project PTDC/CVT-EPI/6685/2014 and BioISI (FCT/UID/Multi/04046/2013) is acknowledged.

6. REFERENCES

Abedon, S.T. (2009). Phage evolution and ecology (San Diego: Academic Press).

Akhter, S., Aziz, R.K., and Edwards, R.A. (2012). PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity-and composition-based strategies. Nucleic Acids Res. *40*, 1–13.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., *et al.* (2008). The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics *9*, 75.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., and Brüssow, H. (2003). Prophage genomics. Microbiol. Mol. Biol. Rev. *67*, 238–276, table of contents.

Cumby, N., Davidson, A.R., and Maxwell, K.L. (2012). The moron comes of age. Bacteriophage *2*, 225–228.

Durmaz, E., and Klaenhammer, T.R. (1995). A Starter Culture Rotation Strategy Incorporating Paired Restriction/ Modification and Abortive Infection Bacteriophage Defenses in a Single *Lactococcus lactis* Strain. Appl. Environ. Microbiol. *61*, 1266–1273.

Fortier, L.-C., and Moineau, S. (2009). Volume 1: Isolation, Characterization and Interactions. In Bacteriophages: Methods and Protocols, M.R.J. Clokie, and A.M. Kropinski, eds. (New York: Humana Press), pp. 203–219.

Fortier, L.-C., and Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. Virulence *4*, 354–365.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.

Hatfull, G.F., and Hendrix, R.W. (2011). Bacteriophages and their genomes. Curr. Opin. Virol. *1*, 298–303.

Hyman, P., and Abedon, S.T. (2010). Bacteriophage host range and bacterial resistance (San Diego: Academic Press).

Jordal, S., Glambek, M., Oppegaard, O., and Kittang, B.R. (2015). New tricks from an old cow: Infective endocarditis caused by *Streptococcus dysgalactiae* subsp. *dysgalactiae*. J. Clin. Microbiol. *53*, 731–734.

Koh, T.H., Sng, L.H., Yuen, S.M., Thomas, C.K., Tan, P.L., Tan, S.H., and Wong, N.S. (2009). Streptococcal cellulitis following preparation of fresh raw seafood. Zoonoses Public Health *56*, 206–208.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2016). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. 1–35.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *5*.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008). Prophinder: A computational tool for prophage prediction in prokaryotic genomes. Bioinformatics *24*, 863–865.

Loman, N.J., Quick, J., and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. Doi.org 15552.

Murray, P.R., Rosenthal, K.S., and Pfaller, M.A. (2009). Medical Microbiology (Philadelphia: Mosby/Elsevier).

Park, M.J., Eun, I.-S., Jung, C.-Y., Ko, Y.-C., Kim, Y.-J., Kim, C.-K., and Kang, E.-J. (2012). *Streptococcus dysgalactiae* subspecies *dysgalactiae* infection after total knee arthroplasty: a case report. Knee Surg. Relat. Res. *24*, 120–123.

Roma-Rodrigues, C., Alves-Barroco, C., Raposo, L.R., Costa, M.N., Fortunato, E., Baptista, P.V., Fernandes, A.R., and Santos-Sanches, I. (2016). Infection of human keratinocytes by *Streptococcus dysgalactiae* subspecies *dysgalactiae* isolated from milk of the bovine udder. Microbes Infect. *18*, 290–293.

Whiley, R.A., and Hardie, J.M. (2009). Volume 3: The Firmicutes - Genus I *Streptococcus.* In Bergey's Manual of Systematic Bacteriology, P. Vos, G. Garrity, D. Jones, N.R. Krieg, W. Ludwig, F. Rainey, K.-H. Schleifer, and W.B. Whitman, eds. (New York, NY: Springer New York), pp. 655–