LUCIANO LOPES QUEIROZ

BACTERIÓFAGOS EM ALIMENTOS FERMENTADOS: EXEMPLOS DE DIVERSIDADE E INTERAÇÕES EM QUEIJO CANASTRA E BOZA

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biomédicas da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

São Paulo
2021

LUCIANO LOPES QUEIROZ

BACTERIÓFAGOS EM ALIMENTOS FERMENTADOS: EXEMPLOS DE DIVERSIDADE E INTERAÇÕES EM QUEIJO CANASTRA E BOZA

Tese apresentada ao Programa de Pós-Graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

Área de concentração: Microbiologia

Orientador: Dr. Christian Hoffmann

Versão original

São Paulo
2021

CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
Serviço de Biblioteca e informação Biomédica
do Instituto de Ciências Biomédicas da Universidade de São Paulo

Ficha Catalográfica elaborada pelo(a) autor(a)

Queiroz, Luciano Lopes
Bacteriófagos em alimentos fermentados: exemplos de diversidade e interações em queijo canastra e boza / Luciano Lopes Queiroz; orientador Christian Hoffmann. -- São Paulo, 2021.
137 p.

Tese (Doutorado)) -- Universidade de São Paulo, Instituto de Ciências Biomédicas.

1. Mibrobiologia. 2. Bacteriofágos. 3. Alimentos Fermentados. 4. Ecologia. 5. Evolução. I. Hoffmann, Christian, orientador. II. Título.

LUCIANO LOPES QUEIROZ

BACTERIOPHAGE IN FERMENTED FOODS: EXAMPLES OF DIVERSITY AND INTERACTION IN CANASTRA CHEESE AND BOZA

Thesis presented to the Postgraduate Program in Microbiology of the Instituto de Ciências Biomédicas from Universidade de São Paulo to obtain the degree of Doctor in Science.

Concentration area: Microbiology

Supervisor: Dr. Christian Hoffmann

Original version

São Paulo
2021

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS

Candidato(a): Luciano Lopes Queiroz

Titulo da Tese: Bacteriófagos em alimentos fermentados: exemplos de diversidade e interações em queijo canastra e boza

Orientador: Christian Hoffmann

A Comissão Julgadora dos trabalhos de Defesa da Tese de Doutorado, em sessão publica realizada a **...../...../.....**, considerou o(a) candidato(a):

( ) Aprovado(a) ( ) Reprovado(a)

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Examinador(a): Assinatura: .....
Nome: .....
Instituição: .....

Presidente: Assinatura: .....
Nome: .....
Instituição: .....

M

ICB USP

Cidade Universitária "Armando de Salles Oliveira", Butantã, São Paulo, SP · Av. Professor Lineu Prestes, 2415 - ICB III - 05508 000
Comissão de Ética em Pesquisa - Telefone (11) 3091-7733 - e-mail: cep@icb.usp.br

CERTIFICADO DE ISENÇÃO

Certificamos que o Protocolo CEP-ICB nº **935/2018** referente ao projeto intitulado: "***Viroma do Queijo Canastra: dinâmicas temporais e relação fago-hospedeiro***" sob a responsabilidade de ***Luciano Lopes Queiroz*** e orientação do(a) Prof.(a) Dr.(a) ***Christian Hoffmann,*** do Departamento de Microbiologia, foi analisado pela **CEUA** - Comissão de Ética no Uso de Animais e pelo **CEPSH** – Comitê de Ética em Pesquisa com Seres Humanos, tendo sido deliberado que o referido projeto não utilizará animais que estejam sob a égide da Lei nº 11.794, de 8 de outubro de 2008, nem envolverá procedimentos regulados pela Resolução CONEP nº 466/2012.

São Paulo, 26 de fevereiro de 2018.

Luciane Valéria Sita

Profa. Dra. **Luciane Valéria Sita**
Coordenadora CEUA ICB/USP

Profa. Dra. **Camila Squarzoni Dale**
Coordenadora CEPSH ICB/USP

EXPEDIÇÃO
DATA 9/3/18
REL. N.º 229

ACKNOWLEDGEMENTS

I thank all Brazilian citizens for giving me the opportunity and privilege to dedicate my life to science.

I am very thankful to my advisor, Christian Hoffmann, for helping me when I needed it most, for the ideas and support that contributed so much to my formation as scientist.

A special thanks to Gustavo Lacorte for have supported me in all field trips, receiving me at his home, and for the good discussions about science, life, politics, and other subjects.

I thank all Hoffmann lab members, William, Débora, Fabiana Lima, Fabiana Sardá, Marina, Isabella, and Rebeca that contributed so much to my work. I also thank the members from the Food Microbiology lab, Ana Paulina, Diogo, Marcela, Gabriela, Josi, Katia, and Lúcia, where I developed a large part of my experiments with their help.

I thank my family for their unconditional support in pursuing my dreams.

ACKNOWLEDGEMENTS

This work was supported by grants from the Brazilian National Research Council (CNPq grant: 425157/2018-0), and from the Sao Paulo Research Foundation (FAPESP grant: 2013/07914-8), and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

“Cada qual cuide de seu enterro, impossível não há.”

Quincas Berro D’agua, A morte e a morte de Quincas Berro d’agua,
Jorge Amado

RESUMO

Bacteriófagos (ou fagos), vírus que infectam bactéria, tem sido descritos em diversos tipos de alimentos fermentados, como vinho, kimchi, kefir, iorgute e queijo. Neste contexto, a diversidade de fagos é melhor caracterizada em ambientes de laticínio, indústria na qual eles podem causar impactos econômicos significativos. Este estudo investigou a comunidade de bacteriófagos em dois alimentos fermentados, queijo Canastra, um queijo artesanal brasileiro, e no Boza, uma bebida fermentada tradicional dos Balcãs. As comunidades microbianas e virais foram acessadas por sequenciamento genômico e metagenômico, técnicas de microbiologia clássica também foram utilizadas para o isolamento e caracterização de bactérias e fagos. Nas amostras de queijo Canastra foi observada uma alta diversidade de fagos, a maioria composta por novas sequências. Genomas bacterianos foram montados a partir do metagenoma dessas amostras e se observou a presença de vários mecanismos de defesa anti-fago, evidenciando as interações entre fagos e bactérias neste ecossistema. Um novo fago lítico capaz de infectar bactérias da espécie *Staphylococcus aureus* foi isolado e caracterizado, e também foi observado o surgimento de uma linhagem bacteriana resistente a infecção durante o ensaio de crescimento entre fago-bactéria. A partir de análises de genômica comparativa entre as bactérias sensíveis e resistentes observou-se mutações em genes associados ao reconhecimento do hospedeiro pelo fagos, evitando que a infecção pelo fago ocorra na linhagem resistente. O microbioma do Boza foi caracterizado, demonstrando uma dominancia de espécies de *Leuconostoc*, da qual um genoma foi recuperado, assim como a presença de fagos líticos e lisogênicos no metagenoma. Um fago lisogênico foi detectado e caracterizado no genoma recuperado de *Leuconostoc*. Uma análise filogenômica comparativa com outros fagos integrados em genomas de *Leuconostoc* foi realizada, demonstrando um grande número de fagos lisogênicos neste gênero, assim como a estrutura de seus genomas. Estes resultados abrem novas perspectivas sobre a composição das comunidade de bacteriófagos e as interações fago-bactéria em alimentos fermentados.

ABSTRACT

Bacteriophages (or phages), viruses that infect bacteria, have been described in several types of fermented foods, such as wine, kimchi, kefir, yogurt, and cheese. In this context, phage diversity is most well characterized in dairy environments, where phages can have significant economic impacts in this industry. This study investigates the bacteriophage community in two fermented foods, Canastra cheese, a Brazilian artisanal cheese, and in Boza, a traditional fermented beverage from the Balkans. Microbial and viral communities were assessed by genome and metagenome sequencing, and classical microbial techniques were also used to isolate and characterize bacteria and phages. A high phage-diversity was observed in Canastra cheese, mostly composed of novel sequences. Bacterial metagenome-assembled genomes (MAGs) were recovered from these samples, encoding several anti-phage defence mechanisms, thus evidencing the interactions between phage and bacteria in this ecosystem. A novel lytic phage able to infect *Staphyloccocus aureus* was isolated and characterized, and the emergence of phage resistant bacteria lineage was also observed during the growth dynamics assays between phage and bacteria. Comparative genomic analysis between phage sensitive and phage resistant bacteria showed mutations in genes associated with phage host recognition, preventing phage infection in the mutant bacteria. The Boza microbiome was characterized, showing a dominance of *Leuconostoc* species, from which a Metagenome Assembled Genome (MAG) was obtained, and the presence of lytic and lysogenic phages. A lysogenic phage was detected in this *Leuconostoc* MAG and characterized. A phylogenomic comparative analysis with other phages integrated in *Leuconostoc* genomes was made, demonstrating a high number of lysogenic phages in this genus, as well as their genome structures. These results shed lights into the bacteriophage community composition and the phage-bacteria interactions in fermented foods.

SUMMARY

RESUMO....10

ABSTRACT....11

INTRODUCTION....17

Bacteriophages....17

Life Cycle....17

Infection....18

Phage-Bacteria interactions....18

Phage host range....20

Advances in microbial ecology....20

Microbiome of fermented foods....21

Phages in fermented foods....22

Phage as a problem in dairy industries....23

Cheese production systems....23

Fermented beverage environment....24

Future uses for phages....25

CHAPTER 1 - HIGH LEVEL OF INTERACTION BETWEEN PHAGES AND BACTERIA IN AN ARTISANAL RAW MILK CHEESE MICROBIAL COMMUNITY....27

1.1 INTRODUCTION....27

1.2 MATERIAL AND METHODS....29

1.2.1 Sampling....29

1.2.2 Virus-like particles (VLPs) concentration....29

1.2.3 DNA Extraction....29

1.2.4 Sequencing....30

1.2.5 Viral metagenome bioinformatics analyses....30

1.2.6 Microbial metagenome bioinformatics analyses....32

1.2.7 Statistical analysis....35

1.3 RESULTS....35

1.3.1 VLP-based description of the bacteriophage community present in the Canastra Cheese endogenous starter culture....35

1.3.2 Classification and genome characterization of bacteriophages present in Canastra cheese endogenous starter culture....37

1.3.3 Identification of lactic acid bacteria (LAB) in the endogenous starter cultures and cheese samples accessed by microbial metagenome sequencing 39

1.3.4 Antiphage defense mechanisms found in Canastra cheese MAGs 41

1.3.5 Presence of CRISPR spacers and prophages in MAGs 42

1.3.6 Phages contigs recovery from microbial metagenomes 43

1.3.7 Correlations between phage and MAG populations in cheese metagenomes 43

1.3.8 Evidence of interactions between phages and bacteria during cheese production 44

1.4 DISCUSSION....45

1.5 CONCLUSION 50

1.6 REFERENCES....50

CHAPTER 2 - THE ARMS RACE BETWEEN ROSENBLUMVIRUS PHAGE AND *STAPHYLOCOCCUS AUREUS*: THE SIMULTANEOUS EMERGENCE OF PHAGE INFECTION RESISTANCE AND BETA-LACTAM ANTIBIOTIC SUSCEPTIBILITY..59

2.1 INTRODUCTION....59

2.2 MATERIAL AND METHODS....60

2.2.1 Sampling....60

2.2.2 Bacterial isolation....61

2.2.3 Virus-like particles (VLPs) concentration 61

2.2.4 Culture media, plaque assay, phage isolation and purification 62

2.2.5 Phage DNA extraction and Sequencing 62

2.2.6 Bioinformatics 63

2.2.7 pH and thermal stability of Staphylococcus phage SP01....64

2.2.8 Host range determination 64

2.2.9 Control of *Staphylococcus aureus* population....65
2.2.10 Adsorption, antibiotic and adherence....65
2.2.11 Electronic microscopy....66
2.2.12 Visualizations....66
2.3 RESULTS....67
2.3.1 Isolation of phages and their host from cheese samples....67
2.3.2 *Staphylococcus aureus* SAU39 genome characterization....67
2.3.3 *Staphylococcus* phage vB-SauP_CS44 genome characterization....68
2.3.4 Morphological characterization....69
2.3.5 Determination of host range....70
2.3.6 *Staphylococcus* phage vB-SauP_CS44 stability....71
2.3.7 Control of *Staphylococcus aureus* SAU39 population by vB_SauP-CS44 72
2.3.8 Comparative genomic analysis....73
2.4 DISCUSSION....76
2.5 CONCLUSION....79
2.6 REFERENCES....79
CHAPTER 3 – A PHYLOGENOMIC VIEW OF THE RELATIONSHIP BETWEEN *LEUCONOSTOC* SPP AND THEIR LYSOGENIC PHAGE FROM BOZA, A TRADITIONAL FERMENTED BEVERAGE....87
3.1 INTRODUCTION....87
3.2 MATERIAL AND METHODS....89
3.2.1 Sampling, DNA Extraction and Sequencing....89
3.2.2 Quality filter, taxonomic and functional profiles and contig assembly....89
3.2.3 Metagenome-assembled genomes (MAGs) and pagenome....90
3.2.4 Phages and prophages detection....91
3.2.5 Phylogenomic tree of *Leuconostoc* lysogenic phages....92
3.2.6 Visualizations....92

3.3 RESULTS....92
3.3.1 Taxonomic and functional analysis of Boza metagenome....92
3.3.2 Metagenome assembled genomes (MAGs)....93
3.3.3 Bacteriophages in Boza....93
3.3.4 *Leuconostoc* pangenome....94
3.3.5 Phylogenomic and comparative analysis of *Leuconostoc* prophages....98
3.4 DISCUSSION....100
3.5 CONCLUSION....102
3.6 REFERENCES....102
CONCLUSIONS....108
REFERENCES....111
APPENDIX 1 - Supplementary Information of Chapter 1....119
APPENDIX 2 - Supplementary Information of Chapter 2....129
APPENDIX 3 - Supplementary Information of Chapter 3....136

LIST OF FIGURES

Figure 1.1. Viral diversity recovered by VLP metagenome sequencing 37

Figure 1.2. Viral cluster taxonomy and genome annotation 38

Figure 1.3. Metagenome-assembled genomes (MAGs) composition and antiphage defense mechanisms in endogenous starter and cheese 41

Figure 1.4. Procrustes analysis using PCoA coordenations of viral and bacterial communities 44

Figure 1.5. Phage-bacteria interaction network during cheese production 45

Figure 2.1. Circular genome representation of S. aureus SAU39.....68

Figure 2.2. Lytic bacteriophages of genera Rosenblumvirus (Podoviridae, Podovirinae) 69

Figure 2.3. Morphology of Staphylococcus phage vB-SauP_CS44 obtained by transmission electron microscopy of negative staining of phage.....70

Figure 2.4. Determination of host range to *Staphylococcus* phage SP01 by spots test against Staphylococcus strains 71

Figure 2.5. Thermostability and pH stability of CS44.....72

Figure 2.6. Growth curve of *Staphylococcus aureus* SAU39 interacting with CS44 ..73

Figure 2.7. Mutation in Poly(ribitol-phosphate) beta-N-acetylglucosaminyltransferase TarS protein.....75

Figure 2.8. Mutation in fibronectin-binding protein A (FnBPA; at bottom), fibronectin-binding protein B (FnBPB; at top) proteins 75

Figure 3.1. Genome annotation of phage genomes recovered from boza metagenome. 94

Figure 3.2. Pangenomic analysis of *Leuconostoc* genus. 96

Figure 3.3. Heatmap of Average Nucleotide Identity (ANI) values among genomes used in pangenome.....97

Figure 3.4. Pangenomic analysis of *Leuconostoc lactis* group.....98

Figure 3.5. Phylogenomic tree and genome annotation of *Leuconostoc* lysogenic phages. 99

INTRODUCTION

Bacteriophages

Viruses are present in every ecosystems on Earth with high abundance, genetic and taxonomic diversity, shaping biogeochemical cycles and ecosystem dynamics (PAEZ-ESPINO et al., 2016; SUTTLE, 2016). It is estimated that there are 10^{31} viral particles on Earth, with particle density ranging from 10^{5}-10^{7} particles per millimeter in seawater to 10^{8}-10^{10} per gram of fecal material, with soil containing the largest variation, with 10^{3}-10^{9} particles per gram of soil (CHEVALLEREAU et al., 2021; LIANG; BUSHMAN, 2021). When they infect bacteria, these obligate intracellular parasites are called bacteriophages (or simply phages). Most known phages have a linear, double-stranded DNA (dsDNA) genome, encapsulated by a tailed protein capsid, which is used to inject the genetic material inside the phage's host cell. However, they can also show genomes encoded by single-stranded DNA (ssDNA), double and single stranded RNA (dsRNA and ssRNA, respectively), and be encapsulated within non-tailed virions (DION; OECHSLIN; MOINEAU, 2020).

Life Cycle

Phages life cycles may be classified as chronic, lytic or lysogenic. Phages identified until the moment that have a chronic life cycle belong to the Inoviridae family. These phages infect their host cells to produce new particles but dot not induce cell lyses upon particle release into the environment (CHEVALLEREAU et al., 2021). On the other hand, lytic phages will infect their host cells; produce new viral particles, and release them on the environment inducing host-cell lysis in this process. Phages that exclusively replicate their genome through a lytic cycle are named as virulent phages. Finally, lysogenic phages will infect their host cells and incorporate their genetic material into their host's genome. These integrated phages are called prophages and their replication occur in conjunction with the host's genome, persisting following the host cell division (FEINER et al., 2015; LIANG; BUSHMAN, 2021). The switch between lytic and lysogenic life cycles is influenced by genetic characteristic of each phage genome, but is also shaped by host-associated (e.g. density) and external factors (e.g. temperate and pH) (CORREA et al., 2021).

Phages can also use an alternative replication mode, infecting their host cells but not entering in either a lytic or lysogenic cycle, remaining in the host in a non-replicative state, as an extrachromosomal element. This state has been called as pseudolysogeny and occurs in environments with limitation of nutrients or other resources (CHEVALLEREAU et al., 2021; LIANG; BUSHMAN, 2021). Moreover, this categorical or non-continuous phage life cycle classification has been recently questioned, with the proposal that interactions between virus and bacteria occur across a continuum of infection modalities (CORREA et al., 2021).

Infection

Phage-bacteria interactions tend to be highly specific, due to phage infection strategies that depend on the host binding proteins and the anti-phage defense systems in the host (MAHONY et al., 2016; SAMSON et al., 2013). The recognition of host by phages occurs through receptor binding proteins (RBP) present in the phage structure, such as tail spikes, fibres, capsid filaments that interact with host receptor molecules, which may be a carbohydrate, protein or teichoic acid (MAHONY et al., 2016; NOBREGA et al., 2018). The defense systems present in the bacterial hosts may act in preventing phage adsorption, DNA entry, cutting phage nucleic acid, and aborting the infection by cell lysis (HAMPTON; WATSON; FINERAN, 2020; LABRIE; SAMSON; MOINEAU, 2010). These mechanisms are classified in adaptive immune systems, including several types of CRISPR-Cas systems, and innate immune systems, such as restriction-modification (RM), abortive infection (Abi) (HAMPTON; WATSON; FINERAN, 2020) and, most recently described systems BREX, for Bacteriophage Exclusion (GOLDFARB et al., 2015), and DISARM, for Defense Island System Associated with Restriction-Modification (OFIR et al., 2018), and others (DORON et al., 2018).

Phage-Bacteria interactions

The interactions between phage and bacteria are important for temporal and spatial population dynamics in all microbial ecosystems. One of the most thoroughly studied phage-bacteria growth model is called Kill-the-Winner (KtW), where phage growth follows a host cell density-dependent dynamics (ARKHIPOVA et al., 2017;

CHEVALLEREAU et al., 2021; RODRIGUEZ-BRITO et al., 2010). In the KtW model, virulent phage populations grow as their bacterial host populations increase in size, decreasing the abundance of dominant bacterial populations and regulating their abundance (THINGSTAD, 2000; WINTER et al., 2010). An alternative phage-bacteria growth dynamics model to the KtW model has been proposed for lysogenic phages, called Piggyback-the-Winner (PtW). The increase in host cell density does not necessary produced an increase in phage abundance, instead, more phages enter in a lysogenic cycles, ultimately decreasing their abundances (KNOWLES et al., 2016).

Lysogenic phages, integrated as prophages in their host's chromosome, have been considered more mutualistic than damaging when compared with virulent phages (CHEVALLEREAU et al., 2021). Although being less damaging to their host, the presence of prophages may affect microbial population and community-level dynamics. Some fitness advantages brought by phage-bacteria mutualism are 1) protection of the host from new viral infections, by superinfection immunity and/or exclusion; 2) the increase of their host's competitiveness capacity by conferring new traits, such as genes related with bacterial virulence, bacteriocins (DRAGOŠ et al., 2021), and antibiotic resistance (CORREA et al., 2021).

Although KtW and PtW represent the prototypical phage-bacteria growth dynamics, it is expected that in nature such relationships go well beyond simple predator-prey interactions. For instance, virulent phages may be less efficient at infection; phages may enter in a pseudolysogeny state; closely related phages may simultaneous infected a single host cell (co-infection) (DÍAZ-MUÑOZ, 2017); and phages could also become integrated within the host genome, expressing auxiliary metabolic genes (AMGs) (BREITBART et al., 2018; HOWARD-VARONA et al., 2020).

Finally, the diversity and composition of phage-bacterial communities are influenced by environmental (*e.g.* pH, temperature, salinity) and biological factors (RODRIGUEZ-BRITO et al., 2010; WINTER et al., 2010). Biological factors can be divided in host traits (abundance, size, distribution, physiological status, and range) and virus traits (infection type, virion size, burst size, and latent period) (CHOW; SUTTLE, 2015).

Phage host range

The commonly perceived relationships existing between phages and bacteria establish that phages can infect a narrow range of closely related bacteria (ROSS; WARD; HYMAN, 2016), due their specific strategies of infection, such as host receptor binding proteins (RBPs), and phage defense mechanisms existing in each host (MAHONY et al., 2016; SAMSON et al., 2013). Nevertheless, recent studies have shown the existence of phages capable of infecting hosts from much wider taxonomic ranges (i.e different genera) (PAEZ-ESPINO et al., 2016; ROSS; WARD; HYMAN, 2016). There are several evidences that phage host range is a dynamic and highly evolvable trait, which it can expand or contract, depending on density, diversity and quality of host available in the habitat (HOLTZMAN et al., 2020; MEYER et al., 2016; SANT et al., 2021). The discovery of generalist phages and highly evolvable traits opened a great number of possible interactions between phages and bacteria, such as the control of different bacterial population by the same phage and events of horizontal gene transfer (HGT) (ROSS; WARD; HYMAN, 2016), impacting their taxonomic and functional diversities (CHEVALLEREAU et al., 2021).

Advances in microbial ecology

Microbial ecology has gone through great changes due advances in next-generation sequencing (NGS) technics, and accessibility and development of new bioinformatics tools (MOKILI; ROHWER; DUTILH, 2012). One approach that was developed using NGS is metagenomics, the direct genetic analysis of all genomes present in an environment, providing the taxonomic and functional gene composition of microbial communities (THOMAS et al., 2012). Metagenomics has been applied to several research fields, such as marine, plant, human, and food environments (MOKILI; ROHWER; DUTILH, 2012). It allows the recovery of not-yet-cultivated microbial species, based on the reconstruction of metagenome-assembled genomes (MAGs) for Bacteria, Archaea, yeasts; as well as for viruses (BOWERS et al., 2017; LEMOS et al., 2021; ROUX et al., 2019). Metagenomic analysis can be carried out using the total DNA present in an environment, or pre-processing a sample to enrich on a given set of organisms, such as for viruses (ROUX et al., 2019). Microbial metagenomes are determined using the total DNA from an environment, and we are

able to recovery MAGs and viruses sequences, as well as their compositional and functional annotations. Viral communities, on the other hand, are usually studied using enrichment procedures that aim to exclude bacterial and other cells component, by filtering, treatments with DNase and RNase. This viral enrichment approach is denominated viral metagenomics (ROUX et al., 2019).

Microbiome of fermented foods

Fermented foods are a traditional method to preserve and have been consumed by humans for centuries. They possessed desirable attributes to human consumption such as proteins, vitamins, minerals, probiotic activity and others (TAMANG et al., 2020). The principal microorganisms present in fermented foods are lactic acid bacteria (LAB), such as *Lactococcus lactis*, *Streptococcus thermophilus*, *Leuconostoc* spp. and *Lactobacillus* spp (BRÜSSOW, 2001; MAHONY et al., 2016; REZAC et al., 2018). These foods have been seen as valuable models for the study of microbial community dynamics, due to their simplicity, reproducibility, accessibility, culturability, and easy-of-manipulation (WOLFE; DUTTON, 2015).

Lactobacillus, widely present in fermented foods, are Gram-positive rods or coccobacilli, strictly fermentative, aero-tolerant or anaerobic, that will enter in homo- or heterofermentative when grown in glucose rich environments as source of carbon, producing more than 85% of lactic acid or equal proportions of lactic acid, CO2, and ethanol/acetic acid (HAMMES; VOGEL, 1995). *Lactobacillus lactis* is the most used species as starter culture in industrial dairy environments, such as cheese, yoghurt and fermented milk (DEVEAU et al., 2006). Other species of *Lactobacillus* found in fermented foods are *L. casei*, *L. paracasei*, *L. plantarum*, *L. rhamnosus*, and *L. curvatus* (BERNARDEAU et al., 2008). The genus *Streptococcus* is composed by Gram-positive, spherical or ovoid cells arranged in chains or pairs, homofermentative, facultatively anaerobic, and some species may be found as commensal or pathogenic in humans and other animals (HARDIE; WHILEY, 1995). *Streptococcus thermophilus* is also used extensively as starter culture in the production of fermented milk products (AUCLAIR; ACCOLAS, 1983). Other genera are *Leuconostoc* and *Lactococcus.* These genera are normally present in less abundance, where they act as adjutant starters, for example in cheese production

(KOT et al., 2014; MAHONY et al., 2016). In other fermentative foods they can be used as single or mixed starter cultures (ALTAY et al., 2013; ZORBA et al., 2003). *Leuconostoc* species are Gram-positive, facultative anaerobic, heterofermentative and produce dextran from sucrose (KOT et al., 2014). *Lactococcus* species are coccoid Gram-positive, anaerobic and facultative anaerobic, homofermentative, and produce lactic acid from glucose (TEUBER, 1995).

Important bacterial pathogens that may be found in fermented environments are *Escherichia coli* and *Staphylococcus aureus* (PINEDA et al., 2021). *E. coli* are Gram-negative, facultative anaerobic, coliform bacteria, commonly associated to fecal contamination with commensal and pathogenic species (BRAZ; MELCHIOR; MOREIRA, 2020; JANG et al., 2017). *S. aureus* are Gram-positive bacteria normally found as a common human and animal commensal; although it can be an opportunistic pathogen (VAN DALEN; PESCHEL; VAN SORGE, 2020), causing subclinical and clinical mastitis in cows, for example.

Phages in fermented foods

Bacteriophages have been described in several types of fermented foods, such as wine, kimchi, sauerkraut, kombucha, kefir, yogurt, and cheese (DEVEAU et al., 2006; JUNG et al., 2018). Phage diversity is most well characterized in dairy environments, where phage can have significant economic impacts in this industry (GONÇALVES DE MELO; LEVESQUE; MOINEAU, 2018).

Most bacteriophages found in fermentation environments infect LAB, and belong to the order Caudovirales and family Siphoviridae, with less frequent phages from families Myoviridae and Podoviridae. In dairy environments, the most common genera found are phage belonging to 936, P335 and c2 groups, all within the Siphoviridae family (DEVEAU et al., 2006; MAHONY et al., 2016). These phages commonly infect LAB from *Lactococcus lactis* and *Streptococcus thermophilus* species (MAHONY; VAN SINDEREN, 2014). Other less abundant phage groups are also able to infect bacterial species of adjutant LAB, such as *Leuconostoc*, *Lactobacillus*, and potential pathogenic bacteria, such as *Staphylococcus* and *Listeria* (DEVEAU et al., 2006; MAHONY et al., 2016).

There are only a few studies characterizing bacteriophage community composition in cheese samples using viral metagenomes. Such studies usually focus on whey and cheese rind samples (DUGAT-BONY et al., 2020; MUHAMMED et al., 2017). Recently, a meta-analysis using 184 cheese microbial metagenomes identified a high abundance of phage-associated sequences (WALSH et al., 2020).

Phage as a problem in dairy industries

Dairy industries depend on processes intrinsically linked to the correct use of starter cultures metabolism in fermentative processes. Therefore, phages can be responsible for major problems in dairy industries, if they were to infect strains used in their production (MARCÓ; MOINEAU; QUIBERONI, 2012; VERREAULT et al., 2011). For instance, phage contaminants can affect the cultures used in dairy industrial production plants, causing pH increase, loss of sensorial characteristics, and even inhibition of the fermentation process altogether (MAHONY et al., 2016). The food industry has developed several anti-phage strategies to avoid and diminish phage contaminations, among them, starter cultures composed by phage insensitive mutants, the rotation of the bacterial strains used, and physical and chemical treatments (GONÇALVES DE MELO; LEVESQUE; MOINEAU, 2018; SAMSON; MOINEAU, 2013). However, phages can be important elements of microbial communities in fermented foods produced outside these industrial settings. Examples of such settings include the production of natural fermented beverages and artisanal cheeses, particularly where natural starter cultures (or endogenous starter cultures) are applied.

Cheese production systems

Many cheeses around the world are produced using endogenous starter cultures (WOLFE et al., 2014), a complex microbial community composed of yeasts, bacteria and phage, all of which interact to create the final food product. This procedure often uses the backsloping method, where residual fermented whey is collected during production to be re-inoculated as a starter culture on the next day’s production batch, effectively creating a continuous microbial growth system (HOLZAPFEL, 2002). Brazil produces a wide variety of artisanal cheeses, several of which use the

backsloping method (CAMPOS et al., 2021; PINEDA et al., 2021). These cheeses are produced for the most part using cow's milk, and less frequently, in specific regions of Brazil, with goat, sheep, and buffalo milk. Some examples of artisanal cheeses are Marajó, Caipira, Colonial, Manteiga, Cerro, Araxá, and Canastra cheeses (PINEDA et al., 2021). Canastra cheese is one the most famous Brazilian cheeses and its endogenous starter cultures have recently been characterized using molecular techniques, containing a diverse, but stable microbial community (KAMIMURA et al., 2019). The community stability in these endogenous starters is assumed to be maintained by a continuous diversification process: when one species is excluded, the genetic function is kept present in the environment by another closely related species (WOLFE; DUTTON, 2015). Studies of Canastra cheese and its production process are usually focused on the bacterial and fungus components of this microbial ecosystems (ANDRADE et al., 2017; PERIN et al., 2017); however, no data has been published to date about their bacteriophages.

Fermented beverage environment

Other environments where bacteriophage can be found are fermented beverages, there are few descriptions of phage composition using metagenomic techniques in fermented beverages (LEDORMAND; DESMASURES; DALMASSO, 2020), although there are reports of some isolated phages (PHILIPPE et al., 2021). The bacterial communities present in these fermented beverages are composed by several LAB (ALTAY et al., 2013). Boza is a traditional fermented cereal beverage popular in different countries from the Balkan Peninsula, and it has its microbial community well described, composed by species from *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Enterococcus*, *Pediococcus*, *Oenococcus*, and *Weissella* genera (LEBLANC; TODOROV, 2011; TODOROV; HOLZAPFEL, 2015). Many different species of LAB have been isolated from boza, characterizing their bacteriocinogenic and antimicrobial potential (LEBLANC; TODOROV, 2011). Nevertheless, few studies have focused on understanding the phages present within boza's microbial communities, their life cycles, their impact on beverage production, and interactions with other microbial component of this traditional beverage (LEDORMAND; DESMASURES; DALMASSO, 2020).

Future uses for phages

Antimicrobial resistance (AMR) has been classified by the World Health Organization (WHO) among the top 10 global public health problems faced by humanity (WHO, 2014). The use of bacteriophages as an alternative to traditional antibiotic therapy has recently come to the forefront of antibacterial treatment for human and veterinary infections (KORTRIGHT et al., 2019). Bacteriophage therapy consists in administering phage directly to the patient with the objective to lysing bacterial pathogens that cause an infection (GORDILLO ALTAMIRANO; BARR, 2019). There are several desirable characteristic in phages to be used in bacteriophage therapy, such as the ability to completely lyse a bacterial culture, a wide host range, but limited to a single species, obligatory lytic life cycle, and free of toxin genes (KLUMPP; LOESSNER, 2013). Phage therapy may be done using a single phage (monophage therapy), a cocktail of phages (polyphage therapy) and/or in association with traditional or alternative antibiotics (DUC et al., 2020; GORDILLO ALTAMIRANO; BARR, 2019). Products based on phage cocktails have been recommended due to the emergence of phage resistance in bacteria populations, where more than one phage may infect the bacteria using different receptor binding proteins (RBPs) (DĄBROWSKA; ABEDON, 2019; HYMAN, 2019). The high diversity of phages in natural environments and the advances in molecular technologies makes it easier to discovery, isolate and describe new phages for use in phage therapy, when compared with traditional antibiotic discovery and development (DĄBROWSKA; ABEDON, 2019). Normally, phages were isolated from their host's habitats; some environments that have been explored include seawater, fish farms, fecal material, sewage, meat, and fermented foods (GONÇALVES DE MELO; LEVESQUE; MOINEAU, 2018; HYMAN, 2019).

Antibiotic treatment can be effective to treat *S. aureus* infections but the emergence of antibiotic resistant *S. aureus* strains, such as methicillin resistant *Staphylococcus aureus* (MRSA), has been considered an important public health problem and listed as a high priority for the development of new antibiotics. In a veterinary context, *S. aureus* colonizes and infects livestock, causing mastitis in ruminants (MATUSZEWSKA et al., 2020), accounting for large economic losses in dairy

production, from the isolation of infected cows to be treated with antibiotics and, to high bacterial load in milk-derived products, an especially important effect for the production of raw milk cheese, such as artisanal Canastra cheese (CAMPOS et al., 2021). Thus, the developments of alternative treatments to *S. aureus* infections, such as bacteriophage therapy, are important in human and veterinary contexts.

1 CHAPTER 1 - HIGH LEVEL OF INTERACTION BETWEEN PHAGES AND BACTERIA IN AN ARTISANAL RAW MILK CHEESE MICROBIAL COMMUNITY

> "Ed, don't stay on the trails when you collect insects. Most people take it too easy when they go in the field. They follow the trails and work a short distance into the woods. You'll get only some of the species that way. You should walk in a straight line through the forest. Try to go over any barrier you meet. It's hard, but that's the best way to collect". Darlington said to Wilson in the spring of 1953. "Naturalist", Edward O. Wilson.

1.1 INTRODUCTION

Viruses are abundant in all ecosystems on Earth, presenting high genetic and taxonomic diversities, shaping biogeochemical cycles and ecosystem dynamics (PAEZ-ESPINO et al., 2016a; SUTTLE, 2016). These obligate intracellular parasites are called bacteriophages (or simply phages) when they infect bacteria. The diversity and composition of phage-bacterial communities are influenced by environmental (*e.g.* pH, temperature, salinity) and biological factors (RODRIGUEZ-BRITO et al., 2010; WINTER et al., 2010). Biological factors can be divided in host traits (species abundance, organismal size, distribution, physiological status, and host range) and virus traits (infection type, virion size, burst size, and latent period) (CHOW; SUTTLE, 2015b).

Many raw milk cheeses around the world are produced using endogenous starter cultures (WOLFE et al., 2014), a complex microbial community composed of yeasts, bacteria and phage, all of which interact to create the final food product. This procedure often uses the backsloping method, where residual fermented whey is collected during production to be reinoculated as a starter culture on the next day's production batch, effectively creating a continuous microbial growth system (HOLZAPFEL, 2002). Brazil produces a wide variety of artisanal cheeses, several of which use the backsloping method (CAMPOS et al., 2021; PINEDA et al., 2021). Canastra cheese is one such cheese, and its endogenous starter cultures have recently been characterized using molecular techniques. They contain a diverse, but stable microbial community (KAMIMURA et al., 2019). The community stability in

these endogenous starters is assumed to be maintained by a continuous diversification process: when one species is excluded, the genetic function is kept present in the environment by another closely related species (WOLFE; DUTTON, 2015). Studies of Canastra cheese microbiome are usually focused on the bacterial and fungal components (ANDRADE et al., 2017; PERIN et al., 2017); however, little is known about its virome composition and interactions with the bacterial hosts.

Most bacteriophages found in dairy fermentation environments belong to the Siphoviridae family, such as 936, P335 and c2 groups. These non-enveloped viruses possess icosahedral morphology and non-contractile tails, with their genome encoded in double-stranded DNA (dsDNA), and commonly infect bacteria from the *Lactococcus* genus. Other less abundant phage groups are also able to infect bacteria from the genera *Leuconostoc*, *Lactobacillus*, *Streptococcus*, *Bacillus*, *Staphylococcus* and *Listeria* (DEVEAU et al., 2006; MAHONY et al., 2016). Phage-bacteria interactions tend to be highly specific, due to phage infection strategies that depend on the host binding proteins and the antiphage defense systems in the host (MAHONY et al., 2016; SAMSON et al., 2013). These defense systems are classified in adaptive immune systems, including several types of CRISPR-Cas systems, and innate immune systems, such as restriction-modification (RM), abortive infection (Abi) (HAMPTON; WATSON; FINERAN, 2020) and, most recently described systems BREX for Bacteriophage Exclusion (GOLDFARB et al., 2015) and DISARM for Defense Island System Associated with Restriction-Modification (OFIR et al., 2018).

Here, we present the first description of the virome composition in Brazilian artisanal Canastra cheese and the phage-bacterial interactions in this food system. We identified 1,234 viral operational taxonomic units (vOTU) and explored the interactions with bacteria across seven cheese producing properties using a combination of viral and microbial metagenomic sequencing. We characterized a putative novel species of *Streptococcus* phage 987 group, as well as its potential host, a metagenome assembled genome (MAG) classified as *Streptococcus salivarius*. Finally, the relationships between 15 complete and high-quality phage genomes and 16 MAGs obtained from starter cultures and cheeses were evaluated.

1.2 MATERIAL AND METHODS

1.2.1 Sampling

Samples were collected from seven cheese producers located in São Roque de Minas and Medeiros cities, in the Serra da Canastra region, state of Minas Gerais, Brazil. For the viral metagenome analysis, 50 mL of the endogenous starter culture to be used in the daily production, obtained from the previous day of production, were sampled and aliquoted in sterile polypropylene tubes. Cheeses produced with this starters were also sampled, at 22-day of ripening, when the cheeses are initially released for sale and human consumption. All samples were placed at -20 °C for the duration of the field trips (as long as 48h), and shipped frozen overnight to the laboratory where they were stored at -80 °C until further processing.

1.2.2 Virus-like particles (VLPs) concentration

VLPs were obtained by filtering and centrifugation. Briefly, 50 mL of starter culture was vortexed for 60s, and centrifuged at 2500G speed for 10 min. The supernatant was transferred to a new tube and the pH adjusted to ~4.6 with 1M HCl or 1M NaOH as needed (MUHAMMED et al., 2017). Adjusted samples were sequentially filtered using 0.45 and 0.22 µm Millex®–HV PVDF syringe filters (Merck Millipore, Tullagreen, Cork, IRL). Filtered samples were centrifuged at 5000G using an Amicon ® Ultra 15 (100 kDa) (Merck Millipore, Tullagreen, Cork, IRL) following the manufacturer's protocol (30 min of centrifugation for every 12 ml). The final concentrated solution was diluted on the Amicon Filter with SM buffer (NaCl 100 mM, MgSO4•7H2O 8 mM M, Tris-Cl 50 mM pH 7.5), and centrifuged at 5000G up until ~ 3 ml of the concentrated phage solution was recovered (complete protocol are available in Supplementary Information).

1.2.3 DNA Extraction

The concentrated VLP stocks were pH adjusted to 7.5 using HCl of NaOH, if needed, prior to DNA extraction. VLP DNA extraction was made following the protocol produced by JAKOČIŪNĖ and MOODLEY (2018) with the DNeasy Blood & Tissue

Kit (Qiagen Inc.). Briefly, 450 µl of phages concentrated were incubated with 50 µL of DNase I 10x buffer, 1 µL DNase I (1 U/µL), and 1 µL RNase A (10 mg/mL) (Sigma-Aldrich, St. Louis, MO, USA) for 1.5 h at 37 °C. DNase and RNase were inactivated with 20 µl of EDTA 0,5M (Sigma-Aldrich, St. Louis, MO, USA) (final concentration of 10 mM) for 20 min at room temperature. To digest phage protein capsid, 1.25 µL of Proteinase K (20 mg/mL) (Invitrogen, Waltham, MA, EUA) was added and incubated for 1.5 h at 56 °C without agitation. DNA purification was carried out using DNeasy Blood & Tissue Kit with 500 µL of lysed phage to increase the yield of extracted DNA.

Total DNA was extracted using the E.Z.N.A.® Soil DNA Kit (Omega Bio-tek, Inc., Norcross, GA, USA): 250 mg of rind cheese was used for the DNA extraction; 3 mL of starter culture was centrifuged, and the cell pellet was used for extraction according to manufacturer's instructions. The integrity of the DNA extracted was evaluated by electrophorese. DNA was quantified using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Scientific, Waltham, MA, USA) and Synergy H1 Hybrid Reader (BioTek, Winooski, VT, USA).

1.2.4 Sequencing

The Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) was used to generate dual-indexed paired-end Illumina sequencing libraries following the manufacturer’s instructions. Libraries were sequenced using 2 x 150 nt paired-end sequencing runs (4 lanes on separate runs) on NextSeq Genome Sequencer (Illumina) with a NextSeq 500/550 High Output Kit v2.5 at Core Facility for Scientific Research – University of Sao Paulo (CEFAP-USP).

1.2.5 Viral metagenome bioinformatics analyses

The quality of raw sequences was verified using FastQC v0.11.9 (ANDREWS, 2010). NextSeq adapters were removed using BBDuk (BBTools, https://jgi.doe.gov/data-and-tools/bbtools/) with the following parameters: ktrim=r k=23 mink=11 hdist=1. The quality trimming was also done using BBDuk with parameters: qtrim=r trimq=10 minlen=60 ftr=139. To assemble the quality filter reads of each viral metagenome we used SPAdes 3.15.0 (NURK et al., 2017) with metagenomic function

(metaspades.py) and automatic parameters of kmers sizes. Generated contigs were filtered to remove short and redundant sequences using BBMap function dedupe.sh with parameters: minscaf=1000 sort=length minidentity=90 minlengthpercent=90. Open reading frames (ORF) were predicted using Prodigal v2.6.3 (HYATT et al., 2010) in metagenomic mode. The final catalog of viral contigs was generated using similar analysis and criterion of SHKOPOROV et al. (2019), with some adaptations. Briefly, the search for amino acid sequences of predicted proteins we used a Hidden Markov Model (HMM) algorithm (hmmscan from HMMER v3.3) against HMM database of prokaryotic viral orthologous groups (pVOG) (GRAZZIOTIN; KOONIN; KRISTENSEN, 2017) considering the significant hit e-value threshold of 10^{-5}. Ribosomal proteins were searched using Barnnap 0.9 (https://github.com/tseemann/barrnap) with an e-value threshold of 10^{-6}. Contigs were aligned against the viral section of NCBI RefSeq database using BLASTn (ALTSCHUL et al., 1990) of BLAST+ package (CAMACHO et al., 2009) with following parameters: e-value < 10^{-10}, covering > 90% of contig length and > 50% identity. We also used VirSorter v1.0.6 (ROUX et al., 2015) as criteria to predict viral sequences with its standard built-in database of viral sequences, with parameter: --db 1. Contigs that meet at least one of the following criteria were included in the final catalog of viral sequences: 1) VirSorter Positive, 2) BLASTn alignments to viral section of NCBI RefSeq, 3) minimum of three ORFs producing HMM-hits to pVOG database, and 4) be circular.

Contigs selected at filter step (n = 908) were taxonomic assignment using Demovir script (https://github.com/feargalr/Demovir) with default parameters and vConTACT v2.0 (BIN JANG et al., 2019) clustering pipeline, a network-based analytical tool that uses whole genome gene-sharing profiles and distance-based hierarchical clustering to group viral contigs into virus clusters (VCs). Besides our viral contigs, we also included in the pool known viral genomes (NCBI RefSeq database release 88). Integrase and site-specific recombinase genes were identified in HMM hit to pVOG annotation of viral contigs. A counting table of viral contigs was generated, mapping unassembled sequences from each library using BBMap with the following parameters: minid=0.99 ambiguous=random. Reads count to contigs with coverage values less than 1X for 75% of a contig length, were set to zero (ROUX et al., 2017).

The number of sequences mapped on viral contigs was normalized using the DESeq2 package (LOVE; HUBER; ANDERS, 2014). Completeness, contamination and quality of contigs were assessed using CheckV (NAYFACH et al., 2020). We selected 14 contigs classified as Complete and High-quality genomes, annotated them using the multiPhATE2 pipeline (ECALE ZHOU et al., 2021), with ORF prediction by Prodigal, gene annotation with hmmscan using pVOG database.

A phylogenomic tree of *Streptococcus* phages was constructed, based on previously study of PHILIPPE et al. (2020), using VICTOR (MEIER-KOLTHOFF; GO, 2017): Virus Classification and Tree Building Online Resource. Briefly, pairwise comparisons of the nucleotide sequences were realized using the Genome-BLAST Distance Phylogeny (GBDP) method (MEIER-KOLTHOFF et al., 2013) under settings recommended for prokaryotic viruses. The intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.0 (LEFORT; DESPER; GASCUEL, 2015) and branch support was inferred from 100 pseudo-bootstrap replicates each and visualized with FigTree 1.4.4 (RAMBAUT, 2012).

1.2.6 Microbial metagenome bioinformatics analyses

The quality of raw sequences was verified using FastQC v0.11.9. NextSeq adapters were removed using BBDuk (BBTools) with the following parameters: ktrim=r k=23 mink=11 hdist=1. The quality trimming was also done using BBDuk with parameters: qtrim=r trimq=10 minlen=100 ftr=140. Compositional profiles of microbial metagenomes samples were assessed using MetaPhlAn2 (v2.7.5) (TRUONG et al., 2015). To assemble the quality filter reads of each microbial metagenome we used SPAdes 3.15.0 (NURK et al., 2017) with metagenomic function (metaspades.py) and automatic parameters of kmers sizes. Generated contigs were filtered to remove short and redundant sequences using BBMap function dedupe.sh with parameters: minscaf=1000 sort=length minidentity=90 minlengthpercent=90.

We examined strain level metagenome-assembled genomes (MAGs) by co-assembling quality filtered sequences using MEGAHIT assembler (LI et al., 2015), with parameter --presets meta-large, and the contigs generated were filtered using BBMap function dedupe.sh with parameters: minscaf=1000 sort=length

minidentity=90 minlengthpercent=90. The resulting filtered contigs were submitted to Metagenomic Workflow using Anvi' 6.1 (EREN et al., 2015). Briefly, we created contig databases, mapping samples reads against contigs using Bowtie2 (LANGMEAD; SALZBERG, 2012) and converted SAM files to BAM with SAMtools (LI et al., 2009); sequence homologs were searched and added to contigs database with hidden Markov Model (HMM) using HMMER (FINN; CLEMENTS; EDDY, 2011); genes were annotated functionally using NCBI's Clusters of Orthologus Groups (TATUSOV et al., 2001)and taxonomically using Centrifuge (KIM et al., 2016); we created an anvi'o profile database with contig length cutoff of 2,500 bp; contigs binning were made using CONCOCT software (ALNEBERG et al., 2014) and generated bins refined manually using anvio-refine function. We selected 16 refined MAGs following the criterias of >50% of completeness and <10% of redundancy (MAGs contigs are available in Supplementary Data). Taxonomic inference of MAGs was done using CheckM (PARKS et al., 2015) and PhyloPhlAn 3.0 (ASNICAR et al., 2020) with database SGB.Nov19, after that, all complete sequences of each MAG species from NCBI RefSeq were downloaded and compared them with MAG sequences using FastANI (JAIN et al., 2018). A counting table of viral contigs was generated, mapping unassembled sequences from each library using BBMap with following parameters: minid=0.99 ambiguous=random. The number of sequences mapped on MAGs contigs was normalized using the DESeq2 package. Pangenomic analysis of *Streptococcus* genus was performed with 94 complete genomes selected from previous study (GAO et al., 2014) and MAG7 using Anvi'o v6.1 with the pangenomic workflow. Briefly, an anvi'o genome database was created, computing (with flag: --use-ncbi-blast; and parameters: --minbit 0.5 --mcl-inflation 8) and displaying pangenome.

CRISPR spacers of MAGs were identified using PILAR-CR (EDGAR, 2007), the spacers consensus generated were matched with our viral contig catalog and the IMG/VR database using BLASTn of BLAST+ package with the following parameters: -qcov_hsp_perc 80 -task blastn -dust no -soft_masking_false (BEZUIDT et al., 2020). Matches of > 90% sequence identity for viral contig catalog and > 95% identity for IMG/VR database were considered. The antiphage defense mechanisms of MAGs were detected following the methods shown by BEZUIDT et al. (2020). Briefly, we

screened predicted genes for domain similarity of known defense systems against the conserved domains database (CDD) of clusters of orthologous groups (COGs) and protein families (Pfams) using RPS-BLAST (e-value < 10^{-2}) (ALTSCHUL et al., 1990). The results were manually filtered for the identification of phage-specific defense systems (complete list is available in APPENDIX 1). Prophages present in MAGs were detected using the PHASTER web server (ARNDT et al., 2016).

Phage contigs present in microbial metagenomes were assessed using VirSorter v1.0.6 (ROUX et al., 2015) as criteria to predict viral sequences with its standard built-in database of viral sequences, with parameter: --db 1. We selected contigs classified only in categories 1, 2 (Phages), 4 and 5 (Prophages) of VirSorter output (n = 514). Open reading frames (ORF) were predicted using Prodigal in metagenomic mode. Those contigs were submitted to the same process that viral metagenome contigs, ORFs annotation with pVOG, ribosomal proteins searched using Barnnap and contigs aligned against the viral section of NCBI RefSeq database using BLASTn. In this case, we used VirSorter Positive as the only criterion to include in the final catalog of viral contigs from microbial metagenome sequences. The taxonomic assignment of contigs, completeness, contamination, and quality of contigs were also made using Demovir, vConTACT v2.0 and CheckV with the same parameters used to viral metagenome sequences.

Our vOTU table of contigs was created combining the two viral contigs catalogs (viral and microbial metagenome). We compare the Average Nucleotide Identity (ANI) of contigs from viral and bacterial metagenomes within and between them using FastANI, with criterion of ANI≥95% and minFraction>85% (ROUX et al., 2019). A counting table of vOTU contigs was generated mapping unassembled sequences from each library using BBMap with following parameters: minid=0.99 ambiguous=random. Read counts to contigs with coverage values less than 1 x for 75% of a contig length, were set to zero (ROUX et al., 2017). The number of sequences mapped on viral contigs was normalized using the DESeq2 package.

1.2.7 Statistical analysis

All analyses were carried out using the statistical software R (R DEVELOPMENT CORE TEAM, 2020) and specific packages as follows: we estimated alpha-diversity using Shannon (log base 2) and Simpson diversity indexes, and richness using the number of observed viral OTU's for each sample using packages *vegan* (OKSANEN et al., 2019) and *microbiome* (LAHTI et al., 2020) packages. We compared similarities between samples of viral and microbial metagenomes through Principal Coordinates Analysis (PCoA) using Jensen-Shannon divergence, followed by Procrustes analysis using *phyloseq* (MCMURDIE; HOLMES, 2013) package. Correlations between normalized abundance values of vOTUs present in starter culture versus normalized MAG abundances present in the cheese were calculated to explore potential phage-bacteria predation relationships. Correlations were deemed significant if they had a value equal to or lower than -0.8 and p value <= 0.05. Additional phage bacterial relationships were explored using WiSH (GALIEZ et al., 2017) with a null model constructed with 148 crAssphage genomes (they were downloaded using *ncbi-genome-download* from viral database with parameters: -s genbank, -l "all" --taxid 1978007). Network plots were generated using R packages *igraph* (CSARDI; NEPUSZ, 2006) and *ggnetwork* (BRIATTE et al., 2021). Genome diagram figures were prepared using the *GenoPlotR* (GUY et al., 2011) package. Other plots were constructed using *ggplot2* (WICKHAM, 2016), *ggpubr* (KASSAMBARA, 2020) and *pheatmap* (KOLDE, 2019).

1.3 RESULTS

1.3.1 VLP-based description of the bacteriophage community present in the Canastra Cheese endogenous starter culture

We assessed the bacteriophage community present in the endogenous starter cultures used by seven artisanal Canastra Cheese producers in Brazil, located in São Roque de Minas and Medeiros, Minas Gerais, Brazil. Viral-like particles (VLPs) were enriched from these starter cultures using a 100 kDa filter membrane and used for metagenome sequencing. The sequencing reads were assembled and rigorously curated to remove bacterial DNA contaminants, producing a final viral sequence

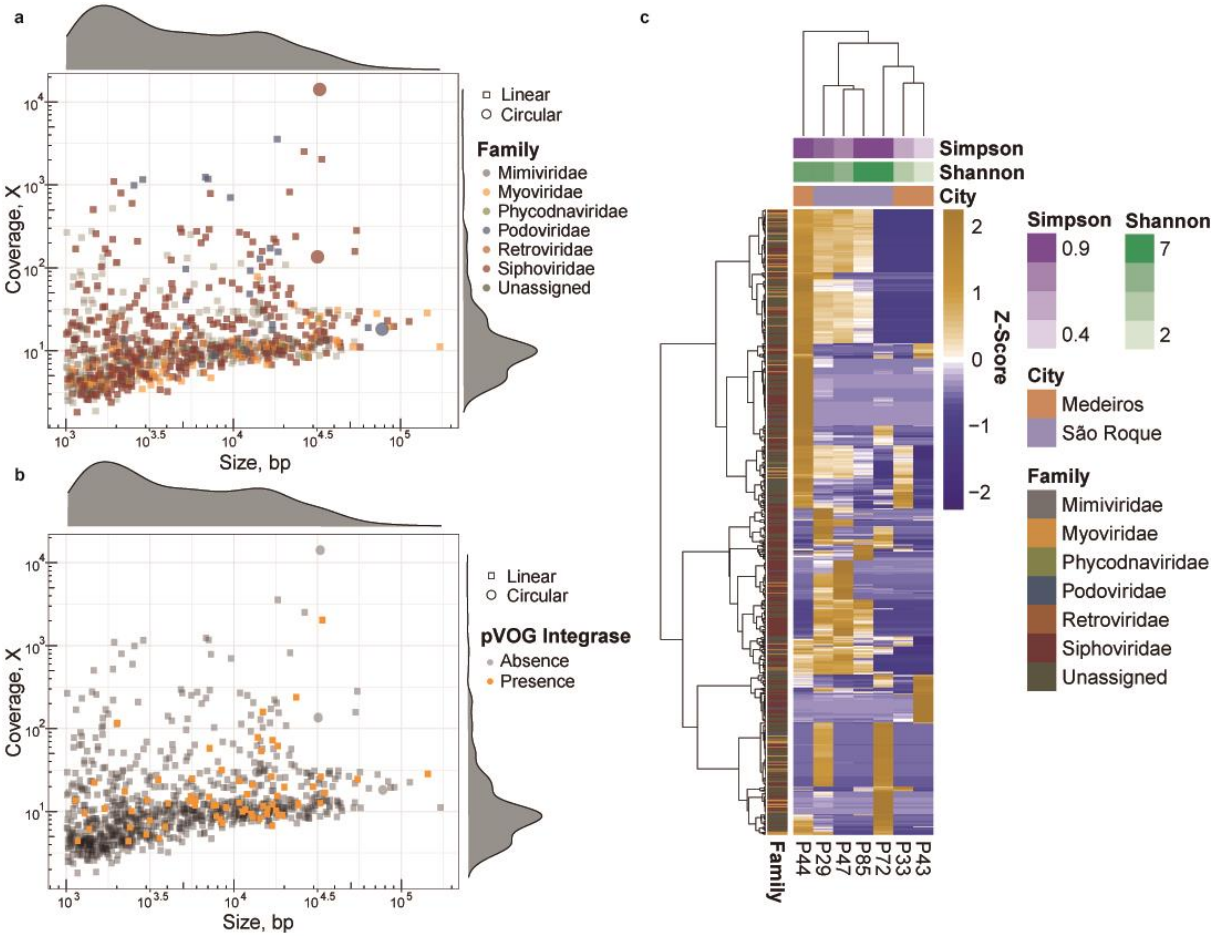
catalog (Appendix 1, Fig. S1). Our final catalog yielded 908 complete and partial viral genomes, with contig sizes ranging from 1×10^3 bp to 1.7×10^5 bp and the coverage between 1.41 - 21108x. Using CheckV software (NAYFACH et al., 2020), as well as the Minimum Information about an Uncultivated Virus Genome (MIUViG) criterium (ROUX et al., 2019), the viral genomes in our catalog were classified as: complete genomes (5), high-quality genomes (12), medium-quality genomes (23), low-quality genomes (584), and not-determined (284).

Family level taxonomic classification for the viral catalog was made using the Demovir pipeline. The order Caudovirales (99%) prevailed, with only a minor number of sequences classified as Algavirales (0.55%) and Imitervirales (0.33%). Contigs were classified at family level as *Siphoviridae* (43.1%), followed by *Myoviridae* (12.1%), *Podoviridae* (3.4%), *Phycodnaviridae* (0.55%), *Mimiviridae* (0.33%), and *Retroviridae* (0.11%). The unassigned contigs corresponded to 40.3% (Fig. 1.1a). Integrase or site-specific recombinase genes were detected in 67 viral contigs (Fig. 1.1b) and VirSorter identified 7.38% of this viral sequence catalog as temperate phages (Appendix 1, Fig. S2).

Alpha diversity analysis of viral metagenomes was calculated using a normalized count table of reads mapped against the viral contigs. Four out of seven samples showed high values of diversity index (> 6 Shannon and > 0.94 Simpson), one sample showed medium values (4.66 Shannon and 0.78 Simpson), and two samples, low values (< 3 Shannon and < 0.6 Simpson) (Appendix 1, Table S1). The sample with lowest diversity values (P43 sample, 1.29 Shannon and 0.39 Simpson) was dominated by one complete, high coverage viral genome (>21000 x) belonging to the *Siphoviridae* family (Fig. 1.1c).

We further refined the classification of our viral catalog by comparing it to the RefSeq complete viral genome database using BLAST and stringent criteria (e-value < 10^{-10}, coverage > 90% of contig length and > 50% identity), and obtained 94 viruses classified at species level. The detected viruses were *Lactococcus* phages 949, asccphi28, bIL285, bIL286, bIL309, bIL312, P078, P162 and ul36. We also found *Lactobacillus* phages, such as phiAQ113 and phiJL1; *Staphylococcus* phages GRCS and phiSA12; and *Streptococcus* phages 9872 and 9874.

Figure 1.1. Viral diversity recovered by VLP metagenome sequencing



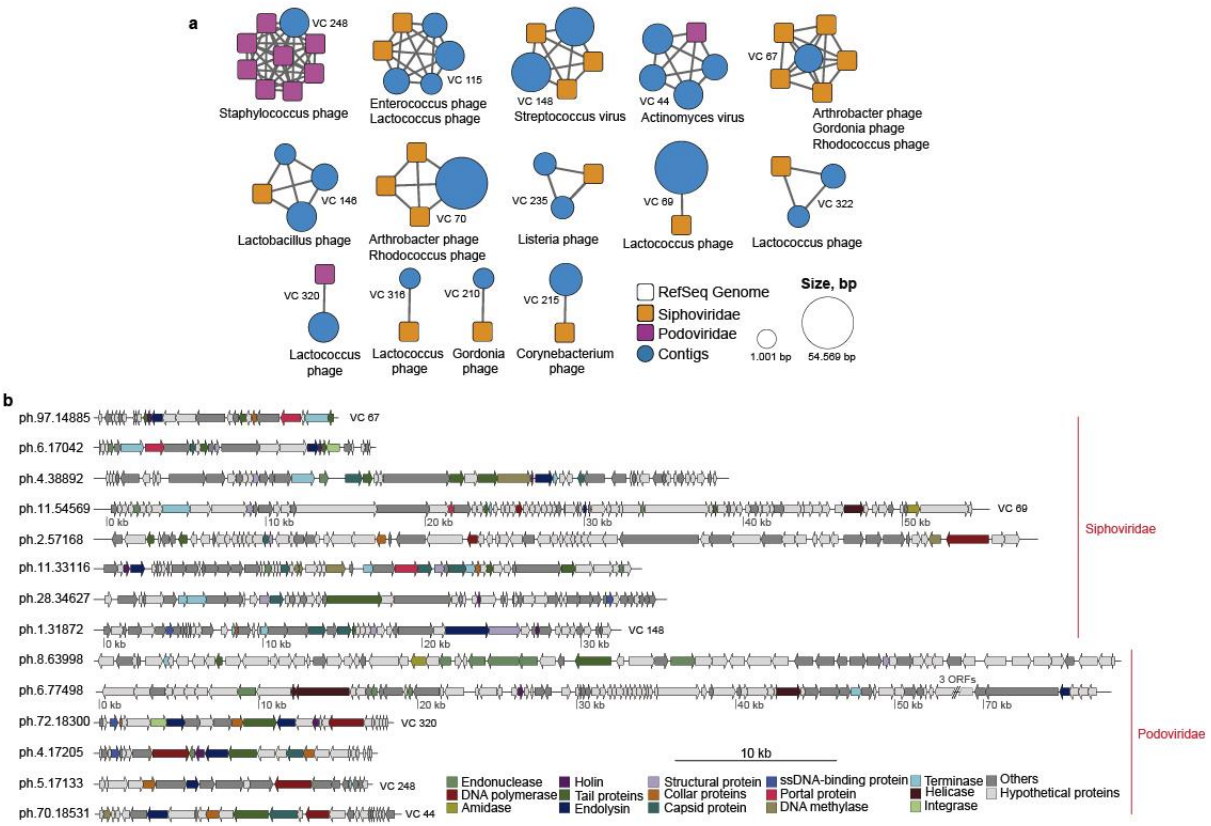
a, Size and coverage distribution of 908 putative viral genomes detected (coverage depth by genome length in bp), color coded by family classification. The families' proportion were Siphoviridae (43.1%), Myoviridae (12.1%), Podoviridae (3.4%), Phycodnaviridae (0.55%), Mimiviridae (0.33%), Retroviridae (0.11%), and Unassigned (40.31%). **b,** Temperate phage distribution as detected by the presence of Integrase/Site-Specific Recombinase genes. **c,** Abundance distribution of viral genomes across analysed starter samples (viral contig abundances plotted as row z-score using normalized abundance values); only contigs present in at least 2 samples are shown in the heatmap (625 contigs).

1.3.2 Classification and genome characterization of bacteriophages present in Canastra cheese endogenous starter culture

We used the vConTACT v2.0 clustering pipeline to refine the taxonomic assignment of our viral genome catalog (BIN JANG et al., 2019). We included in the analysis known viral genomes (NCBI RefSeq database release 88) in addition to our 908 viral contigs. Putative viral genomes from our catalog that clustered at the same Viral Cluster (VC) were considered to be the same virus genera. Matches were found for only 2.75% of our contigs against all viral genomes present in RefSeq, indicating a

large amount of novel viral diversity (Appendix 1, Fig. S3). Viral contigs clustered with RefSeq genomes belong to *Siphoviridae*, *Myoviridae* and *Podoviridae* families and their sizes ranged from 1,001 to 54,569 bp. We observed that some VCs were composed of complete (VC 148: *Streptococcus virus*) and high-quality (VC 248: *Staphylococcus phage*; VC 44: *Actinomyces virus*; VC 67: *Arthrobacter*, *Gordonia*, and *Rhodococcus phages*; and VC 69: *Lactococcus phage*) viral contigs (Fig. 1.2a).

Figure 1.2. Viral cluster taxonomy and genome annotation



a, Network of clusters formed between RefSeq genomes and viral contigs. RefSeq genomes and their family level classification are represented by squares and viral contigs by circles. Square colors indicate viral families and the size of circles represent the viral contig length in base pairs. **b**, Genome annotation of complete and high-quality bacteriophages genomes using pVOG database; some of these genomes were clustered with RefSeq genomes as shown by their VC number. Colors indicate the gene identification.

Some complete and high-quality genomes were clustered with viral genomes present in RefSeq, for instance: three RefSeq genomes of *Streptococcus virus* 9871, 9872, and 9874 that belong to phage 987 group clustered with phage ph.1.31871 and ph.1.31871 (VC148). A phylogenomic analysis of this VC containing 98 *Streptococcus phages* genomes placed our novel genomes firmly within the 987

group (Appendix 1, Fig. S4). Also, a high-quality genome clustered with 7 RefSeq genomes of *Staphylococcus phages* (i.e. GRCS) that belong to the genus *Rosenblumvirus* (VC 248).

We used the multiPhATE2 pipeline (ECALE ZHOU et al., 2021) to annotate 14 of our 17 complete and high-quality putative viral genomes. We used Prodigal software to predict ORFs and the pVOG database for gene annotation (Fig. 1.2b). Eight genomes were classified as *Siphoviridae* with genomes sizes ranging from 14,885 to 57,168 bp, and six as *Podoviridae* with genomes sizes ranging from 17,133 to 77,498 bp. The most frequently annotated genes in these 14 viral genomes were Tail protein (22), Terminase (17), Endonucleases (15), Capsid protein (14), and Endolysin (13). We observed integrase genes in two genomes, including a fully recovered genome which clustered with *Lactococcus phage* asccphi28 belonging to P034 phage species (VC 320).

1.3.3 Identification of lactic acid bacteria (LAB) in the endogenous starter cultures and cheese samples accessed by microbial metagenome sequencing

To better understand the microbial community and the interactions between phage and bacteria in Canastra cheese, we also carried out microbial metagenome sequencing using samples of endogenous starters and cheese produced with these same starters at 22-days of ripening, which is the minimal ripening time required by law in Brazil for the commercialization of these cheeses (DORES; NOBREGA; FERREIRA, 2013). Microbial metagenome sequences were obtained for samples collected in six of the seven studied producers, and were initially analysed using MetaPhlAn2 (TRUONG et al., 2015). The most abundant bacterial species detected using MetaPhlAn2 were *Lactococcus lactis* (average of 30.6%), *Streptococcus thermophilus* (17.4%), *Streptococcus infantarius* (13.7%), *Streptococcus salivarius* (7.3%), and *Corynebacterium variabile* (5.5%) (Appendix 1, Fig. S5). MetaPhlAn2 also detected *Lactococcus* phage ul36 (6.1%) in starter and cheese samples from producer P29 and in the starter from producer P72.

Having established this first characterization of the microbial community, we refined the bacterial species analysis by detecting metagenome-assembled genomes

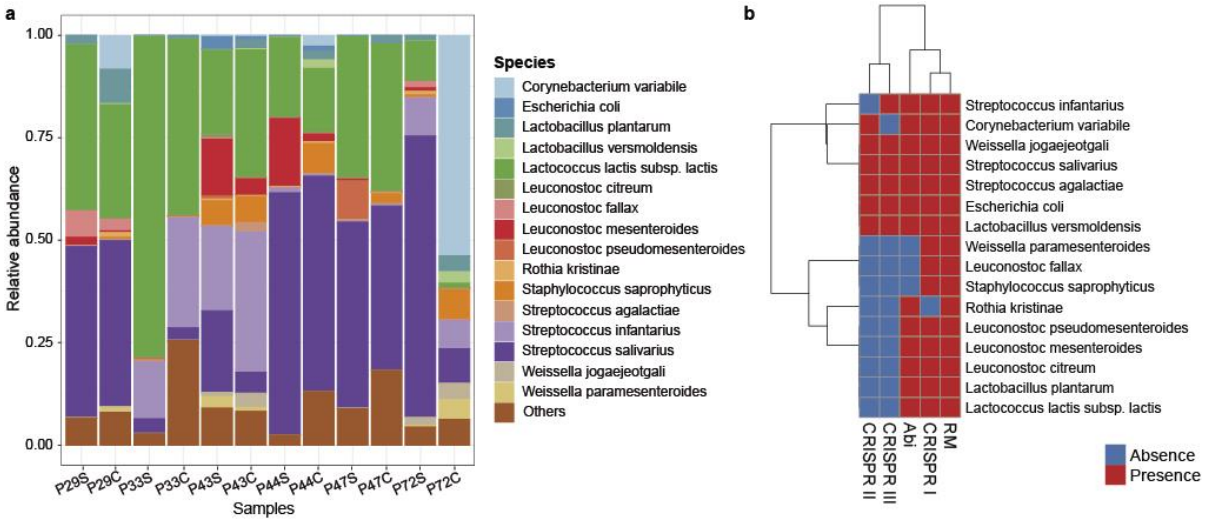
(MAGs). Metagenomic contigs were submitted to the Metagenomic Workflow of Anvi'o 6.1 (EREN et al., 2015). Contig binning was made using CONCOCT (ALNEBERG et al., 2014) and the generated bins were manually refined using *anvio-refine* function. From a total of 50 MAGs obtained, we selected 16 refined MAGs with at least >50% completeness and <10% redundancy, nine of which showed more than 90% completeness. Taxonomic inference of MAGs was done using CheckM (PARKS et al., 2015) and PhyloPhlAn 3.0 (ASNICAR et al., 2020). Finally, we downloaded all complete sequences for each MAG species from NCBI RefSeq and compared them with MAG sequences using FastANI (JAIN et al., 2018).

We identified 13 MAGs belonging to the Firmicutes phylum, two belonging to Actinobacteria, and one to Proteobacteria. The most representative genera were *Leuconostoc* with 4 MAGs, followed by *Streptococcus* with 3, and *Lactobacillus* and *Weissella*, both with 2 MAGs, all of which had at least 95% of average nucleotide identity (ANI) to reference genomes. Additionally, MAG7 was assigned as *Streptococcus salivarius* and showed 94.2% of ANI to *Streptococcus salivarius* BIOML-A24 (genbank accession: GCF_009717045.1). This MAG grouped with S. *salivarius* and *S. infantarius* in a pangenome analysis carried out with 95 reference genomes (Appendix 1, Fig. S6). Other MAGs were classified as *Lactococcus*, *Rothia*, *Staphylococcus*, *Corynebacterium*, and *Escherichia*. Sequence reads from each sample were mapped to the 16 obtained MAGs and the count table generated were normalized using DESeq2 (LOVE; HUBER; ANDERS, 2014) (Fig. 1.3a).

The most abundant MAGs across all samples were classified as *Streptococcus salivarius*, *Lactococcus lactis* subsp. *lactis, and Streptococcus infantarius* (averages of 34.6%, 33%, and 11.4% respectively), detected in all samples. An inverse relationship was detected between *S. infantarius* e *S. salivarius* abundances across all studied samples. We observed a high level of similarity between the community composition observed in the starter and cheese samples within the same producer (Wilcoxon test p value = 0.001, comparison of within versus between producers Bray-Curtis distances), with some species decreasing or increasing in relative abundance, such as the decrease of *L. lactis* subsp. *lactis* and increase of *S. infantarius* in samples from producer P33. A departure from this tight within-producer relationship

between starter and cheese samples was observed only for producer P72, where we observe a marked increase in *Corynebacterium variabile* relative abundance from starter (0.003%) to cheese (56.9%).

Figure 1.3. Metagenome-assembled genomes (MAGs) composition and antiphage defense mechanisms in endogenous starter and cheese.



a, Relative abundance of each MAG classified at species level in 12 paired starter and cheese samples. Others represent reads mapped against lower quality MAGs. **b**, Presence and absence of antiphage defense mechanisms found in each MAG.

1.3.4 Antiphage defense mechanisms found in Canastra cheese MAGs

The 16 MAGs were screened for known bacterial defense system genes and manually filtered for specific antiphage defense mechanisms. We identified a total of 395 defense genes belonging to restriction modification (RM), abortive infection (Abi), CRISPR-type I, II and III mechanisms (Fig. 1.3b). The *Rothia kristinae* genome had no CRISPR-cas system genes and all *Leuconostoc* genomes presented only the CRISPR-I system. *Weissella jogaejeotgali*, *Streptococcus salivarius*, *Streptococcus agalactiae*, *Escherichia coli*, and *Lactobacillus versmoldensis* harbored all five types of defense genes. Another five MAGs harbored three types of defense genes, such as *Lactococcus lactis* subsp. *lactis* and *Leuconostoc mesenteroides*, and four MAGs had only two types of defense genes, *Weissella paramesenteroides*, *Leuconostoc fallax*, and *Staphylococcus saprophyticus* harbored RM and CRISPR-type I, and *Rothia kristinae* with RM and Abi systems. We did not find any defense genes classified as DISARM, BREX, Thoeris, Shedu, Gabija, and others.

1.3.5 Presence of CRISPR spacers and prophages in MAGs

Phage-bacteria interactions can be studied by assessing the CRISPR arrays present in contigs or MAGs obtained from microbial and viral metagenomes, representing the infection history of bacteria present in a given system. Here, we identified CRISPR spacers in our MAGs using PILAR-CR (EDGAR, 2007), and each spacer consensus generated was matched against our viral contig catalog and the IMG/VR database using BLASTn. A total of 10 CRISPR arrays were found in 4 of the 16 MAGs (Supplementary Data). The MAG classified as *Streptococcus salivarius* (MAG7) harbored five arrays with 153 spacers and average length of 212 bp. The second MAG with most CRISPR arrays was MAG3, classified as *Escherichia coli*, which harbored three arrays with 106 spacers and 267 bp of average length. The other two MAGs (*Lactobacillus versmoldensis* and *Weissella jogaejeotgali*) harbored only one array each.

All CRISPR arrays of *Streptococcus salivarius* matched (more than 95% identity) with phages from IMG/VR database, represented by phages from families *Siphoviridae* and *Myoviridae* that have the genera *Streptococcus* and *Streptococcus thermophilus* as predicted host lineages. The array present in the *Weissella jogaejeotgali* MAG matched with phages from family *Siphoviridae*, and no match was found for the array present in the *Lactobacillus versmoldensis* MAG. Two of three arrays detected in the *E. coli* MAG matched with IMG/VR database, containing sequence signatures for *Siphoviridae*, *Myoviridae*, *Podoviridae*, and *Inoviridae* (Tubulavirales) phage families, all of which have *Escherichia, Klebsiella*, and *Xanthomonas* species as predicted host lineages. Only one array, detected in the *E. coli* MAG, matched with phage contigs from our viral catalog, one of these was the ph.11.54569 classified as *Lactococcus* phage and grouped within VC69 (Fig. 1.2).

We identified prophages in our MAGs using PHASTER (ARNDT et al., 2016). Four intact prophage sequences were identified in MAGs classified as *Leuconostoc fallax* (MAG2), *Escherichia coli* (MAG3), *Streptococcus infantarius* (MAG6), and *Streptococcus salivarius* (MAG7). All these sequences were classified at family level as *Siphoviridae*. The most common gene annotations produced by PHASTER were *Lactobacillus* phage Sha1 (7 genes) for the prophage found in MAG2, *Salmonella*

phage SEN34 (6) for the prophage in MAG3, *Streptococcus* phage phiNJ2 (20) for the prophage in MAG6, and *Streptococcus* phage 5093 (9) for the prophage in MAG7.

1.3.6 Phages contigs recovery from microbial metagenomes

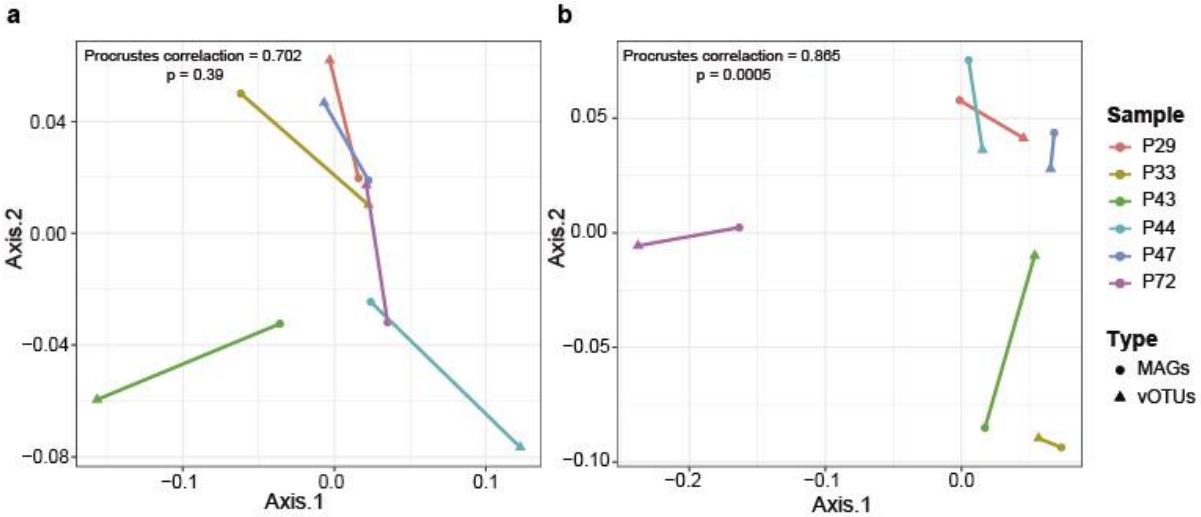
It was possible to recover phage sequences not detected in the VLP-isolated dataset, particularly because cheese samples were not used for VLP isolation, by analysing the microbial metagenome dataset using VirSorter (ROUX et al., 2015). A total of 514 viral contigs were identified from the 12 metagenome samples, and their taxonomic classification at family level revealed a prevalence of *Siphoviridae* (85%), followed by *Myoviridae* (4,2%), and *Podoviridae* (2,5%). We recovered 2 complete and 10 high-quality genomes, while the remaining genome fragments were of medium-quality (27), low-quality (467), and not-determined (8). The two complete phage genomes recovered from the bacterial metagenomes were the same as found when sequencing VLPs obtained from producers P33 and P43 endogenous starter samples (ph.1.31872 and ph.1.32817). Additionally, we also analysed the phage dynamics between starter and cheese samples, as well as their intra- and interspecific interactions.

1.3.7 Correlations between phage and MAG populations in cheese metagenomes

We expanded our viral catalog to include the new phage detected using the metagenome dataset, by comparing all contigs obtained from the viral and bacterial metagenomes using FastANI, with criterion of ANI>95% and minFraction>85% (ROUX et al., 2019), creating a final virus list with 1234 unique phage contigs. We then mapped all reads obtained from each sample to this expanded viral list to create a viral OTU (vOTU) table for further comparative analysis. We analysed the global relationships existing between the bacterial and viral communities in the cheese and starter metagenomes using Jensen-Shannon divergence matrices created for each sample type, followed by Procrustes analysis. We did not observe a correlation between viral and bacterial communities in starter samples (0.702 with $p = 0.39$, Fig. 1.4a). However, the correlation between the viral and bacterial communities in the cheese samples was significant (0.865 with $p = 0.0005$, Fig. 1.4b). We identified a

significant correlation between bacterial communities present in starter versus cheese samples, but not for phage communities (Appendix 1, Fig. S7).

Figure 1.4. Procrustes analysis using PCoA coordenations of viral and bacterial communities.



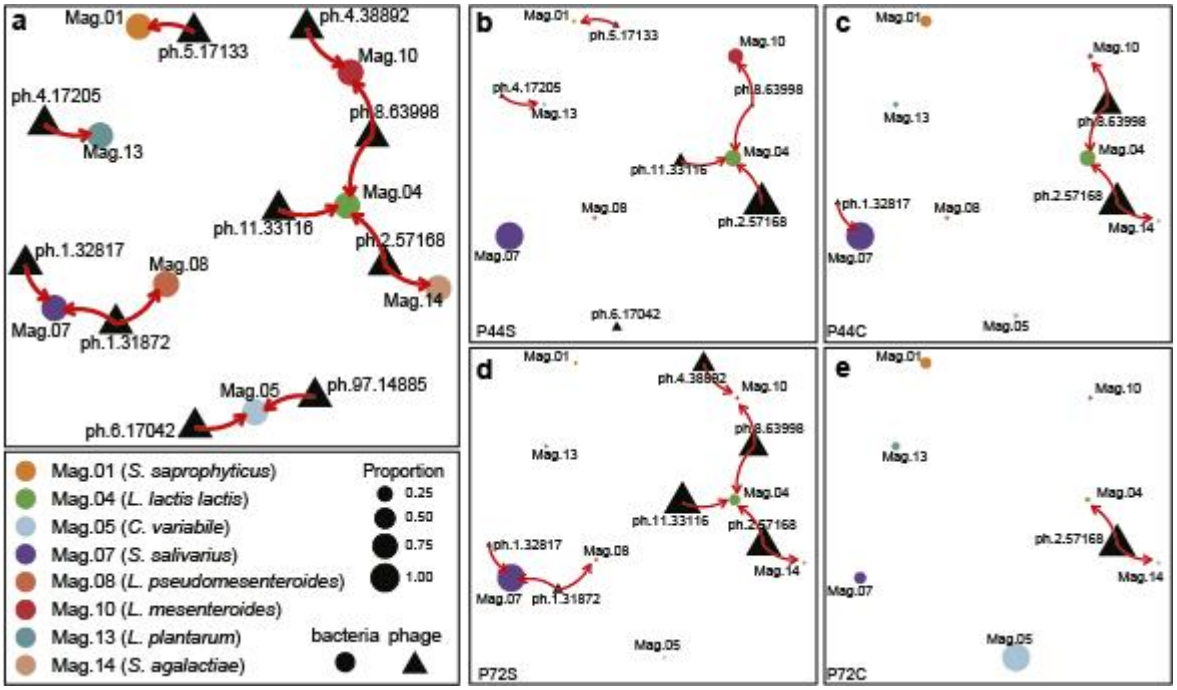
a, Correlations between viral (vOTUs = 1234) and bacterial communities (MAGs = 16) in endogenous starter cultures samples. **b**, Correlations between viral and bacterial communities in cheese samples. Distance matrices were calculated using Jensen-Shannon divergence in both cases.

1.3.8 Evidence of interactions between phages and bacteria during cheese production

We further evaluated the existing relationship between the 15 high quality phage genomes and 16 high quality bacteria MAG in the Canastra cheese production system, using Spearman correlations based on normalized phage/bacteria abundances and on predicted host infection using the software WIsH (GALIEZ et al., 2017). We considered that phages present in the initial starter culture would infect their hosts, and therefore a negative correlation was expected. Indeed, negative correlations were found both in endogenous starters and in cheese samples (Fig. 1.5, and Appendix 1, Table S2). For instance, phages ph.1.32817 (ρ = -0.84, *p* = 0.03) and ph.1.31872 (ρ = -0.88, *p* = 0.02) both showed negative correlations with *S. salivarius* (MAG7), and phage ph.1.31872 was also negatively correlated with *L. pseudomesenteroides* (MAG8) (ρ = -0.83, *p* = 0.04). We also observed negative correlations between phages ph.11.33116 and *L. lactis* subsp. *lactis* (MAG4) (ρ = -0.84, *p* = 0.03); phages ph.8.63998 and *L. mesenteroides* (MAG10) (ρ = -0.88, *p* =

0.01); and ph.2.57168 and *S. agalactiae* (MAG14) (ρ = -0.88, *p* = 0.01). Additionally, the predicted host for phages ph.8.63998 and ph.2.57168 was *L. lactis* subsp. *lactis* (MAG4) (null model *p-value* < 0.05), and the two phages and MAG4 were present in all samples. Other predicted hosts were *S. saprophyticus* (MAG1), for phage ph.5.17133, and *C. variabile* (MAG5) for phages ph.6.17042 and ph.97.14885. Although an overall correlation pattern could be observed across all samples, we also detected a high level of individual variation across all analysed producers, indicating a high level of producer specialization (Fig. 1.5, and Appendix 1, Fig. S8).

Figure 1.5. Phage-bacteria interaction network during cheese production.



a, General network describing the putative phage-bacteria infection interactions present across all samples. **b-e**, Examples of putative phage-bacteria infection interactions for two distinct producers (P44: b-c; and P72: d-e), and for separate starter and cheese samples (starter: b and d; cheese: c and e). Edges represent putative infection relationships between phage and bacteria.

1.4 DISCUSSION

Endogenous starter cultures are used in the production of several cheeses around the world, such as Parmesan, Époisses, and Canastra cheese, in Brazil. Although the bacterial composition of these starters is often well characterized, little is known about phage–bacteria growth dynamics in these cheese production systems, where phages are normally treated as problems, as the viral infections can negatively affect

or even eliminate the starter culture during production. Here, we report the first study of the phage-bacteria structure present in an artisanal cheese produced by the backslopping method, by sampling the endogenous starter culture and the cheese derived from its production batch, from seven distinct artisanal Canastra Cheese producers in Minas Gerais Gerais state, Brazil.

There are only a few studies characterizing bacteriophage community composition in cheese samples using viral metagenomes. Most of these studies have focused on the whey and cheese rind samples (DUGAT-BONY et al., 2020; MUHAMMED et al., 2017). Recently, a meta-analysis using 184 cheese microbial metagenomes identified a high abundance of phage-associated sequences (WALSH et al., 2020). In our study we used both approaches, exploring the microbial communities in starter and cheese samples during cheese production and the interactions among them. We characterized cheese and starter culture samples from several producers in this region by sequencing viral and microbial metagenomes, recovering a total of 1,234 vOTUs, including 18 high-quality or complete viral genomes, and 16 metagenome assembled bacterial genomes (MAGs). The majority of the viral genomes were assigned to *Siphoviridae*, *Myoviridae*, and *Podoviridae* families, which are commonly encountered in dairy systems (DEVEAU et al., 2006; MAHONY et al., 2016; MUHAMMED et al., 2017).

Phage taxonomy remains an important challenge (BIN JANG et al., 2019; ROUX et al., 2019), and most reports have focused on the study of LAB phages, mainly *Lactococcus lactis* and *Streptococcus thermophiles* (MAHONY; VAN SINDEREN, 2014a). Nevertheless, there is still a large prevalence of unknown phages in several environments (DION; OECHSLIN; MOINEAU, 2020; PAEZ-ESPINO et al., 2016a), including dairy and dairy-related environments. We have also observed a high proportion of unclassified contigs or classified only at family level, even when they were considered complete and high-quality contigs.

There is a high level of viral diversity variation across all analysed starter samples, with differences as high as 5-fold being observed for Shannon diversity index. Samples with low viral diversity tended to be dominated by two phages that clustered with *Streptococcus* virus group 987 reference genomes. This phage group has been

recently discovered and described as a novel emerging group of *S. thermophilus* phages. Group 987 phage is thought to have originated from genetic exchange events between *L. lactis* P335 phage group and *S. thermophilus* phages, acquiring morphogenesis related genes and replication modules from each group respectively (MCDONNELL et al., 2016). We are presenting the first description of phage 987 group in a Brazilian cheese, and the observation of a putative novel phage species in this group, considering that both contigs detected in this group showed < 95% ANI values compared to the four available genomes, and all these sequences created a monophyletic clade on our phylogenomic analysis.

One of our most unique findings is the detection of a complete genome, phage ph.72.18300, which clustered with *Lactococcus* phage asccphi28 genome, belonging to the group P034 phage species (family *Podoviridae*), a group rarely found in the dairy industry (DEVEAU et al., 2006; KOTSONIS et al., 2008). *Lactococcus* phage asccphi28 shows more genetic and functional similarities with phages normally infecting *Bacillus subtilis* and *Streptococcus pneumoniae* than other *Lactococcus lactis* phages (KOTSONIS et al., 2008). ANI measured between ph.72.18300 and asccphi28 genomes was 93%, well below the usually accepted threshold for same species classification, indicating that this phage is potentially a novel viral species within the *Lactococcus* phage asccphi28 group. This phage was detected in high abundance in only one of the studied producers. Another phage genome, ph.5.17133, was recovered exclusively from the producer P44 endogenous starter sample, being classified as a *Staphylococcus* phage belonging to the genus *Rosenblumvirus* (*Podoviridae*). Attention has recently been drawn to these phages, due to their potential use in bacteriophage therapy in veterinary medicine, as a means to treat *Staphylococcus*-positive mastitis (BREYNE et al., 2017; TITZE et al., 2020).

Recent studies have highlighted the importance of characterizing microbial strains within dairy-related systems to understand microbiome assembly and function in several habitats, such as cheese rinds (NICCUM et al., 2020; VAN ROSSUM et al., 2020; WALSH et al., 2020). For instance, distinct bacterial strains can respond to environmental and biological stress in different ways. We have characterized 16

MAGs at strain level resolution, including LAB such as *Lactococcus lactis* subsp. *lactis*, *Streptococcus salivarius*, and *Streptococcus infantarius*, followed by less abundant strains of *Leuconostoc*, *Lactobacillus*, and *Weissella*. Among these, a MAG classified as *Streptococcus salivarius* showed less than 95% ANI to any known species. A pangenomic analysis of the *Streptococcus* genus and the phylogenomic tree constructed with 302 single-copy core genes placed our *S. salivarius* MAG between *S. salivarius* and *S. thermophilus* groups, further indicating its potential as a new strain. Therefore, we postulate that the MAG 07 *Streptococcus salivarius* strain could be endogenous to the Canastra region, in Brazil. Furthermore, its growth seems to be modulated by native phages present in this artisanal production system, and this relationship is likely to influence the fermentation dynamics and ultimately the sensorial profile of this cheeses.

We observed a high level of similarity between proportions of predominant microbial species, from starter to cheese samples, indicating a resilient microbial ecosystem. This also highlights that although microorganisms could be acquired during the cheese production and ripening stages, the overall microbiome composition and structure present in Canastra cheese is primarily determined by the starter culture. Nevertheless, one producer seemed to depart from this pattern in our sampling, where *Corynebacterium variabile* dominated the cheese samples while being a minor component of the starter culture. *Corynebacterium variable* is found in smear-ripened cheeses and is responsible for flavor and textural properties during ripening process and strains of this species are known to compose the microbiome of surface cheeses (SCHRÖDER et al., 2011). It is possible that *C. variable* is acquired during the maturation process, when the cheese is in direct contact with several surfaces for a prolonged period of time.

A large amount of evidence for the interaction between phage and bacterial strains in Canastra cheese production system was found. All MAGs showed at least two types of antiphage defense systems, such as CRISPR, restriction-modification (RM) and abortive infection (Abi). These mechanisms are commonly found in bacterial genomes (DORON et al., 2018; PUJATO; QUIBERONI; MERCANTI, 2019), with RM present in about 90% of all bacteria and CRISPR-Cas in 50% of them (HAMPTON;

WATSON; FINERAN, 2020; HILLE et al., 2018). The presence of several methyltransferase genes was observed within the phage contigs, which is a well-known phage evasion mechanism against bacterial RM systems (LABRIE; SAMSON; MOINEAU, 2010; MURPHY et al., 2013). Furthermore, we also identified CRISPR spacer arrays in 4 of the 16 analysed MAGs, indicating previous infections and active evolution of the adaptive immune system in these strains. The MAG identified as *S. salivarius* and *E. coli* showed multiple spacers from different phage species, suggesting multiple infection events (HILLE et al., 2018). Finally, we detected temperate phage sequences inserted in bacterial genomes of 4 MAGs, including the novel putative strain of *S. salivarius*.

We observed a high similarity (as measured by ANI) between phages ph.1.32817 and ph.1.31872, forming the same viral cluster with reference genomes of phage 987 group. However, they did not form a single vOTU, as they presented an alignment fraction less than 85%, indicating a potential strain differentiation. The occurrence of these two phages is negatively correlated with *S. salivarius* MAG7, indicating their ability to infect this species. *S. salivarus* MAG7 was the most abundant *Streptococcus* species in absence of 987 phage strains, and when these phages were present, *S infantarius* MAG6 became the dominant *Streptococcus* species. Thus, the control of *S. salivarius* by the 987 phage provides an adaptive advantage to *S. infantarius*, allowing it to become the dominant *Streptococcus* species in this lactic fermentation ecosystem. Phages belonging to the 987 group are also described as being able to infect some *Lactococcus* species, however we did not observe any evidence for this interaction in our analysis.

For cheese samples, it was possible to detect a global relationship between the composition of phages and bacteria. Biochemical and environmental changes occurring during cheese production and ripening, such as pH and salinity, are known to affect microbiome competition in many types of cheese (CAMPOS et al., 2021; WOLFE et al., 2014). Thus, it is likely that these same factors could influence phage-bacterial interactions. Other factors intrinsic to lactic fermentation systems, and that can influence phage-bacteria interactions, are the metabolism of residual lactose, lactate, citrate, and lipolysis and proteolysis (KHATTAB et al., 2019; MARCÓ;

MOINEAU; QUIBERONI, 2012). Previous studies have demonstrated that there is a balance between active starter cells and lysed cells to control the lactose degradation and proteolysis, respectively, and some milk proteins can interfere in phage activity (CROW et al., 1995; GARCÍA-ANAYA et al., 2020). The matrix structure of cheese can also impact those interactions by firmness and viscosity of cheese (GARCÍA-ANAYA et al., 2020). Finally, phage dispersion rates in a matrix of hard cheese will be different from that in soft cheese or in starter culture and whey solutions.

1.5 CONCLUSION

In conclusion, this study revealed a rich and diverse phage population permeating all cheese and endogenous starter culture samples, with high levels of phage-bacteria interactions in the Canastra cheese production system. We extensively described the main phage members of these microbial ecosystems, and identified a likely novel phage species, belonging to *Streptococcus* phage 987 group, as well as its putative host, a novel strain of *S. salivarius*. We observed a dynamic yet stable microbial ecosystem during cheese production, marked by genomic evidence of continued phage-bacteria interactions, where general patterns emerged, yet maintaining a high level of inter-producer variability. This is a first effort to describe and understand the viral composition and ecological dynamics within the Canastra Cheese production system. We provide a solid background for further mechanistic studies focused on identifying phage-bacteria interactions in artisanal cheeses, with likely impact in similar production systems around the world.

1.6 REFERENCES

ALNEBERG, J. et al. Binning metagenomic contigs by coverage and composition. **Nature Methods**, v. 11, n. 11, p. 1144–1146, 2014.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–10, 1990.

ANDRADE, R. P. et al. Yeasts from Canastra cheese production process: Isolation and evaluation of their potential for cheese whey fermentation. **Food Research International**, v. 91, p. 72–79, 2017.

ANDREWS, S. **FastQC: A Quality Control Tool for High Throughput Sequence Data**, 2010. Disponível em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

ARNDT, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. **Nucleic acids research**, v. 44, n. W1, p. W16–W21, 2016.

ASNICAR, F. et al. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. **Nature Communications**, v. 11, n. 1, p. 1–10, 2020.

BEZUIDT, O. K. I. et al. Phages Actively Challenge Niche Communities in Antarctic Soils. **mSystems**, v. 5, n. 3, p. 1–12, 2020.

BIN JANG, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. **Nature Biotechnology**, v. 37, n. 6, p. 632–639, 2019.

BREYNE, K. et al. Efficacy and safety of a Bovine-Associated Staphylococcus aureus Phage Cocktail in a murine model of mastitis. **Frontiers in Microbiology**, v. 8, n. NOV, p. 1–11, 2017.

BRIATTE, F. et al. ggnetwork. 2021.

CAMACHO, C. et al. BLAST+: Architecture and applications. **BMC Bioinformatics**, v. 10, p. 1–9, 2009.

CAMPOS, G. Z. et al. Microbiological characteristics of canastra cheese during manufacturing and ripening. **Food Control**, v. 121, n. August 2020, p. 107598, 2021.

CHOW, C.-E. T.; SUTTLE, C. A. Biogeography of Viruses in the Sea. **Annual Review of Virology**, v. 2, n. 1, p. 41–66, 2015.

CROW, V. L. et al. The influence of phage-assisted lysis of Lactococcus lactis subsp. lactis ML8 on cheddar cheese ripening. **International Dairy Journal**, v. 5, n. 5, p. 451–472, 1995.

CSARDI, G.; NEPUSZ, T. The igraph software package for complex network research. **InterJournal Complex Systems**, v. Complex Sy, n. 1695, p. 1695, 2006.

DEVEAU, H. et al. Biodiversity and classification of lactococcal phages. **Applied and Environmental Microbiology**, v. 72, n. 6, p. 4338–4346, 2006.

DION, M. B.; OECHSLIN, F.; MOINEAU, S. Phage diversity, genomics and phylogeny. **Nature Reviews Microbiology**, v. 18, n. March, 2020.

DORES, M. T. DAS; NOBREGA, J. E. DA; FERREIRA, C. L. DE L. F. Room temperature aging to guarantee microbiological safety of Brazilian artisan Canastra cheese. **Food Science and Technology (Campinas)**, v. 33, n. 1, p. 180–185, 2013.

DORON, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. **Science**, v. 1008, n. March, p. 1–17, 2018.

DUGAT-BONY, E. et al. Viral metagenomic analysis of the cheese surface: A comparative study of rapid procedures for extracting viral particles. **Food Microbiology**, v. 85, n. July 2019, p. 103278, 2020.

ECALE ZHOU, C. L. et al. MultiPhATE2: code for functional annotation and comparison of phage genomes. **G3 (Bethesda, Md.)**, v. 11, n. 5, p. 1–5, 2021.

EDGAR, R. C. PILER-CR: Fast and accurate identification of CRISPR repeats. **BMC Bioinformatics**, v. 8, p. 1–6, 2007.

EREN, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. **PeerJ**, v. 3, p. e1319, 2015.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, n. suppl, p. W29–W37, 1 jul. 2011.

GALIEZ, C. et al. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. **Bioinformatics**, v. 33, n. 19, p. 3113–3114, 2017.

GAO, X. Y. et al. Comparative genomics of the bacterial genus streptococcus illuminates evolutionary implications of species groups. **PLoS ONE**, v. 9, n. 6, 2014.

GARCÍA-ANAYA, M. C. et al. Phages as biocontrol agents in dairy products. **Trends in Food Science and Technology**, v. 95, n. June 2019, p. 10–20, 2020.

GOLDFARB, T. et al. BREX is a novel phage resistance system widespread in microbial genomes . **The EMBO Journal**, v. 34, n. 2, p. 169–183, 2015.

GRAZZIOTIN, A. L.; KOONIN, E. V.; KRISTENSEN, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): A resource for comparative genomics and protein family annotation. **Nucleic Acids Research**, v. 45, n. D1, p. D491–D498, 2017.

GUY, L. et al. GenoPlotR: comparative gene and genome visualization in R. **Bioinformatics**, v. 27, n. 13, p. 2334–2335, 2011.

HAMPTON, H. G.; WATSON, B. N. J.; FINERAN, P. C. The arms race between bacteria and their phage foes. **Nature**, v. 577, n. 7790, p. 327–336, 15 jan. 2020.

HILLE, F. et al. The Biology of CRISPR-Cas: Backward and Forward. **Cell**, v. 172, n. 6, p. 1239–1259, 2018.

HOLZAPFEL, W. H. Appropriate starter culture technologies for small-scale fermentation in developing countries. **International Journal of Food Microbiology**, v. 75, n. 3, p. 197–212, 2002.

HYATT, D. et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, v. 11, 2010.

JAIN, C. et al. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. **Nature Communications**, v. 9, n. 1, p. 1–8, 2018.

JAKOČIŪNĖ, D.; MOODLEY, A. A Rapid Bacteriophage DNA Extraction Method. **Methods and Protocols**, v. 1, n. 3, p. 27, 2018.

KAMIMURA, B. A. et al. Large-scale mapping of microbial diversity in artisanal Brazilian cheeses. **Food Microbiology**, v. 80, n. October 2018, p. 40–49, 2019.

KASSAMBARA, A. **ggpubr: "ggplot2" Based Publication Ready Plots**, 2020. Disponível em: <https://rpkgs.datanovia.com/ggpubr/>

KHATTAB, A. R. et al. Cheese ripening: A review on modern technologies towards flavor enhancement, process acceleration and improved quality assessment. **Trends in Food Science and Technology**, v. 88, n. January, p. 343–360, 2019.

KIM, D. et al. Centrifuge: Rapid and sensitive classification of metagenomic sequences. **Genome Research**, v. 26, n. 12, p. 1721–1729, 2016.

KOLDE, R. pheatmap. p. 1–8, 2019.

KOTSONIS, S. E. et al. Characterization and genomic analysis of phage asccφ28, a phage of the family Podoviridae infecting Lactococcus lactis. **Applied and Environmental Microbiology**, v. 74, n. 11, p. 3453–3460, 2008.

LABRIE, S. J.; SAMSON, J. E.; MOINEAU, S. Bacteriophage resistance mechanisms. **Nature Reviews Microbiology**, v. 8, n. 5, p. 317–327, 2010.

LAHTI, L. et al. **Tools for microbiome analysis in R. Microbiome package version**, 2020. Disponível em: <http://microbiome.github.io/microbiome>

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 2012.

LEFORT, V.; DESPER, R.; GASCUEL, O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. **Molecular Biology and Evolution**, v. 32, n. 10, p. 2798–2800, 1 out. 2015.

LI, D. et al. MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. **Bioinformatics**, v. 31, n. 10, p. 1674–1676, 2015.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.

LOVE, M. I.; HUBER, W.; ANDERS, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 1–21, 2014.

MAHONY, J. et al. Phage-Host Interactions of Cheese-Making Lactic Acid Bacteria. **Annual Review of Food Science and Technology**, v. 7, n. 1, p. 267–285, 2016.

MAHONY, J.; VAN SINDEREN, D. Current taxonomy of phages infecting lactic acid bacteria. **Frontiers in Microbiology**, v. 5, n. JAN, p. 1–7, 2014.

MARCÓ, M. B.; MOINEAU, S.; QUIBERONI, A. Bacteriophages and dairy fermentations. **Bacteriophage**, v. 2, n. 3, p. 149–158, 2012.

MCDONNELL, B. et al. Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium Streptococcus thermophilus. **Applied and Environmental Microbiology**, v. 82, n. 17, p. 5153–5165, 2016.

MCMURDIE, P. J.; HOLMES, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. **PloS one**, v. 8, n. 4, p. e61217, 2013.

MEIER-KOLTHOFF, J. P. et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. **BMC Bioinformatics 2013 14:1**, v. 14, n. 1, p. 1–14, 21 fev. 2013.

MEIER-KOLTHOFF, J. P.; GO, M. Phylogenetics VICTOR : genome-based phylogeny and classification of prokaryotic viruses. v. 33, n. July, p. 3396–3404, 2017.

MUHAMMED, M. K. et al. Metagenomic Analysis of Dairy Bacteriophages: Extraction Method and Pilot Study on Whey Samples Derived from Using Undefined and Defined Mesophilic Starter Cultures. **Applied and Environmental Microbiology**, v. 83, n. 19, p. e00888-17, 1 out. 2017.

MURPHY, J. et al. Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese-producing plants. **Journal of dairy science**, v. 96, n. 8, p. 4945–57, 2013.

NAYFACH, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. **Nature Biotechnology**, 2020.

NICCUM, B. A. et al. Strain-Level Diversity Impacts Cheese Rind Microbiome Assembly and Function. **mSystems**, v. 5, n. 3, p. 1–18, 16 jun. 2020.

NURK, S. et al. MetaSPAdes: A new versatile metagenomic assembler. **Genome Research**, v. 27, n. 5, p. 824–834, 2017.

OFIR, G. et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. **Nature Microbiology**, v. 3, n. 1, p. 90–98, 2018.

OKSANEN, J. et al. **vegan: Community Ecology Package**, 2019. Disponível em: <http://cran.r-project.org/web/packages/vegan/index.html>

PAEZ-ESPINO, D. et al. Uncovering Earth's virome. **Nature**, v. 536, n. 7617, p. 425–430, 2016.

PARKS, D. H. et al. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. **Genome Research**, v. 25, n. 7, p. 1043–1055, 2015.

PERIN, L. M. et al. Bacterial ecology of artisanal Minas cheeses assessed by culture-dependent and -independent methods. **Food Microbiology**, v. 65, p. 160–169, 2017.

PHILIPPE, C. et al. Novel genus of phages infecting Streptococcus thermophilus: Genomic and morphological characterization. **Applied and Environmental Microbiology**, v. 86, n. 13, p. 1–40, 2020.

PINEDA, A. P. A. et al. Brazilian Artisanal Cheeses: Diversity, Microbiological Safety, and Challenges for the Sector. **Frontiers in Microbiology**, v. 12, n. April, 2021.

PUJATO, S. A.; QUIBERONI, A.; MERCANTI, D. J. Bacteriophages on dairy foods. **Journal of Applied Microbiology**, v. 126, n. 1, p. 14–30, 2019.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.**Vienna, Austria.R Foundation for Statistical Computing, , 2020. Disponível em: <http://www.r-project.org/.>

RAMBAUT, A. **FigTree**, 2012. Disponível em: <http://tree.bio.ed.ac.uk/software/figtree/>

RODRIGUEZ-BRITO, B. et al. Viral and microbial community dynamics in four aquatic environments. **The ISME Journal**, v. 4, n. 6, p. 739–751, 2010.

ROUX, S. et al. VirSorter: Mining viral signal from microbial genomic data. **PeerJ**, v. 2015, n. 5, p. 1–20, 2015.

ROUX, S. et al. Benchmarking viromics: an *in silico* evaluation of metagenome-enabled estimates of viral community composition and diversity. **PeerJ**, v. 5, p. e3817, 2017.

ROUX, S. et al. Minimum information about an uncultivated virus genome (MIUVIG). **Nature Biotechnology**, v. 37, n. 1, p. 29–37, 2019.

SAMSON, J. E. et al. Revenge of the phages: Defeating bacterial defences. **Nature Reviews Microbiology**, v. 11, n. 10, p. 675–687, 2013.

SCHRÖDER, J. et al. Complete genome sequence of Corynebacterium variabile DSM 44702 isolated from the surface of smear-ripened cheeses and insights into cheese ripening and flavor generation. **BMC Genomics**, v. 12, n. 1, p. 545, 2011.

SHKOPOROV, A. N. et al. The Human Gut Virome Is Highly Diverse, Stable, and Individual Specific. **Cell Host & Microbe**, v. 26, n. 4, p. 527- 541.e5, 2019.

SUTTLE, C. A. Environmental microbiology: Viral diversity on the global stage. **Nature Microbiology**, v. 1, n. 11, p. 1–2, 2016.

TATUSOV, R. L. et al. **The COG database: new developments in phylogenetic classification of proteins from complete genomesNucleic Acids Research**. [s.l: s.n.]. Disponível em: <http://www.ncbi.nlm.nih.gov/COG>. Acesso em: 4 maio. 2020.

TITZE, I. et al. Efficacy of bacteriophages against Staphylococcus aureus isolates from bovine mastitis. **Pharmaceuticals**, v. 13, n. 3, 2020.

TRUONG, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. **Nature Methods**, v. 12, n. 10, p. 902–903, 2015.

VAN ROSSUM, T. et al. Diversity within species: interpreting strains in microbiomes. **Nature Reviews Microbiology**, 2020.

WALSH, A. M. et al. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. **Nature Food**, v. 1, n. 8, p. 500–510, 13 ago. 2020.

WICKHAM, H. **ggplot2**. Cham: Springer International Publishing, 2016.

WINTER, C. et al. Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited. **Microbiology and Molecular Biology Reviews**, v. 74, n. 1, p. 42–57, 2010.

WOLFE, B. E. et al. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. **Cell**, v. 158, n. 2, p. 422–433, 2014.

WOLFE, B. E.; DUTTON, R. J. Fermented Foods as Experimentally Tractable Microbial Ecosystems. **Cell**, v. 161, n. 1, p. 49–55, 26 mar. 2015.

2 CHAPTER 2 - THE ARMS RACE BETWEEN ROSENBLUMVIRUS PHAGE AND *STAPHYLOCOCCUS AUREUS*: THE SIMULTANEOUS EMERGENCE OF PHAGE INFECTION RESISTANCE AND BETA-LACTAM ANTIBIOTIC SUSCEPTIBILITY

> "Evolution, thus, is merely contingent on certain processes articulated by Darwin: variation and selection". Ernst Mayr

2.1 INTRODUCTION

Interactions between bacteriophages (or phages) and bacteria can be observed in a wide variety of microbial systems (PAEZ-ESPINO et al., 2016), such as those found in endogenous starter cultures used for cheese production. These microbial communities are normally composed of starter and adjunct starter cultures of Lactic Acid Bacteria (LAB), yeasts and phage, as well as potential food pathogens, such as *Staphylococcus aureus*. Recent studies have highlighted a diverse but stable microbial community in Canastra Cheese, one of the most famous artisanal cheeses produced in Brazil using raw milk (KAMIMURA et al., 2019; PERIN et al., 2017; PINEDA et al., 2021). An important bacterial pathogen that may be found in Canastra cheese is *Staphylococcus aureus*, and little is known about the interaction of *S. aureus* and their phages in Brazilian artisanal cheese environments.

Although *S. aureus* is normally a common human and animal commensal, it can be an opportunistic pathogen (VAN DALEN; PESCHEL; VAN SORGE, 2020). Antibiotic treatment can be effective, but the emergence of antibiotic resistant *S. aureus* strains, such as methicillin resistant *Staphylococcus aureus* (MRSA), has been considered an important public health problem. In a veterinary context, *S. aureus* colonizes and infects livestock, causing mastitis in ruminants (MATUSZEWSKA et al., 2020), accounting for large economic losses in dairy production, from the isolation of infected cows to be treated with antibiotics and, to high bacterial load in milk-derived products, an especially important effect for the production of raw milk cheese, such as artisanal Canastra cheese (CAMPOS et al., 2021).

Bacteriophages have recently come to the forefront of antibacterial treatment as an alternative to treat human and veterinary infections (KORTRIGHT et al., 2019).

Several phage infecting *S. aureus* have been described, normally belonging to order Caudovirales, a group of non-enveloped viruses, with double-strain DNA (dsDNA) genome, possessing icosahedral morphology and thin filamentous tails (MOLLER; LINDSAY; READ, 2019). Most known *S. aureus* phages belong to the family Siphoviridae, and less frequently to families Myoviridae and Podoviridae (DEGHORAIN; VAN MELDEREN, 2012; XIA; WOLZ, 2014). The lysogenic life cycle is common among *S. aureus* phages, integrating their genomes to host chromosomes and contributing to virulence and pathogenesis (NOVICK; CHRISTIE; PENADÉS, 2010). However, to be used in phage therapy, it is desirable that the phage possess some characteristics, such as the ability to completely lyse a bacterial culture and an obligatory lytic life cycle (HYMAN, 2019; KLUMPP; LOESSNER, 2013). Podoviridae phage infecting *S. aureus* have been recently demonstrated to have a high potential to be used as antibacterial agents, such as phages BP39 (BREYNE et al., 2017), GRCS (SUNAGAR; PATIL; CHANDRAKANTH, 2010), and SA46-CTH2 (DUC et al., 2020). Nevertheless, mechanistic studies addressing *S. aureus*-phage interactions are still lacking, particularly those aiming at elucidating different forms of host-cell recognition by receptor-binding proteins (RBPs), how phage and bacteria interact in nature and eventually in a medical setting (MOLLER; LINDSAY; READ, 2019), and the genomic consequences of such interactions.

Our objective in this study was to isolate a new lytic phage able to infect *Staphylococcus aureus* strains isolated from artisanal Canastra Cheese produced with raw milk. We isolated a novel *Rosenblumvirus* phage, called vB_SauP-CS44, genomically characterized it and observed the emergence of infection resistance in its host. We performed comparative genome analysis of susceptible and resistant host bacteria, thus highlighting important elements of phage-bacteria interactions, with potential implications for future use of phage in bacterial infection control.

2.2 MATERIAL AND METHODS

2.2.1 Sampling

We obtained 50 mL of starter culture from the previous day of production, from 12 distinct cheese producers from the Canastra Region (Appendix 2, Table S1), located

in São Roque de Minas and Medeiros, Southwest of Minas Gerais state, Brazil. The samples were aliquoted in sterile polypropylene tubes and kept refrigerated for up to 2 hs on ice while being transported from the field sites. Samples were placed at -20 °C for the duration of the field trips, and shipped overnight at 0 °C to the laboratory where they were stored at -20 °C until further processing.

2.2.2 Bacterial isolation

Staphylococcus aureus SAU39 and SAU66 were isolated previously (CAMPOS et al., 2021), using standard techniques. Briefly, 25 g of each cheese were homogenized with 225 mL of 0.1% sterile peptone water (Oxoid, UK) and decimal dilutions were prepared and plated in duplicates in Petrifilm® STX plates (3M, Minnesota, USA, 2016) for counts of coagulase positive *Staphylococcus*, following the manufacturer's instructions. Selected strains were isolated for further experiments. The remaining bacterial strains used were obtained from the collection of the Laboratory of Microbiology of the School of Pharmaceutical, and at the School of Veterinary Medicine, at the University of São Paulo.

2.2.3 Virus-like particles (VLPs) concentration

VLPs enrichment was realized using Amicon Centrifugal Filter (AGGARWALA; LIANG; BUSHMAN, 2017) without chloroform (KAUFFMAN et al., 2018). Briefly, 50 mL of starter culture was vortexed for 60s and centrifuged at 2500G speed for 10 min. The supernatant was transferred to a new tube and pH adjusted to ~4.6 with 1M HCl or 1M NaOH as needed (MUHAMMED et al., 2017). Adjusted samples were sequentially filtered using 0.45 and 0.22 µm syringe filters (Millex Syringe Filter Unit, polyethersulfone; Millipore, Massachusetts, USA). Filtered samples were centrifuged at 5000G using a Amicon ® Ultra 15 (100 kDa) following the manufacturer's protocol (30 min of centrifugation for every 12 ml). The final concentrated solution was diluted on the Amicon filter with SM buffer (NaCl 100 mM, MgSO4•7H2O 8 mM M, Tris-Cl 50 mM pH 7.5) (Cold Spring Harb Protoc, 2006), and centrifuged at 5000G up until ~ 3 ml of the concentrated phage solution was recovered.

2.2.4 Culture media, plaque assay, phage isolation and purification

Brain Heart Infusion (Oxoid, Basingstoke, UK) broth was used to culture bacteria. Plaque assay experiments were conducted with concentrated VLPs as previously described using an optimized classic agar overlay method (KAUFFMAN; POLZ, 2018). Infectivity tests were carried out using *Staphylococcus aureus* strains SAU39 and SAU66 isolated from Canastra cheeses and *Staphylococcus aureus* ATCC 29213 (Appendix 2, Table S2). Bottom agar plates were prepared with BHI 1% of agar and top agar with 0.3% for culture media. The phage was purified by picking an isolated plaque with a sterile pipette tip and streaking it onto a still-molten agar overlay prepared with the host colony. Inoculated plates were incubated at 37 °C overnight. An agar plug was removed from a new isolated plaque and phage particles were eluted on the same host growth medium at 4 °C overnight. The phage suspension was filtered at 0.22 µm pore size, mixed with glycerol to a final concentration of 50% and stored at -80 °C.

2.2.5 Phage DNA extraction and Sequencing

The volume of 450 µL of phage stocks was incubated with 50 µL DNase I 10x buffer, 1 µL DNase I (1 U/µL) (Sigma-Aldrich, St. Louis, MO, USA) and 1 µL RNase A (10 mg/mL) for 1.5 h at 37 °C (Sigma-Aldrich, St. Louis, MO, USA), without shaking. DNase and RNase were inactivated with 20 µl of EDTA 0,5M (final concentration of 20 mM) for 20 min at room temperature. Phage protein capsid was digested using 1.25 µL Proteinase K (20 mg/mL) (Thermo Scientific, Waltham, MA, USA) and incubated for 1.5 h at 56 °C without shaking (JAKOČIŪNĖ; MOODLEY, 2018). Phage DNA was purified using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. The integrity of the DNA extracted was evaluated electrophoretically (1% agarose gel). DNA was quantified using Quant-iT™ PicoGreen™ dsDNA Assay (Thermo Scientific, Waltham, MA, USA). DNA was sequenced using Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA). Libraries were sequenced using 2 x 150 nt paired-end sequencing runs (4 lanes on separate runs) on NextSeq Genome Sequencer (Illumina) with a NextSeq 500/550 High Output Kit v2.5 at Core Facility for Scientific Research – University of Sao Paulo (CEFAP-USP).

2.2.6 Bioinformatics

The quality of raw sequences was verified using FastQC v0.11.9. NextSeq adapters were removed using BBDuk (BBTools, https://jgi.doe.gov/data-and-tools/bbtools/) with the following parameters: ktrim=r k=23 mink=11 hdist=1. The quality trimming was also done using BBDuk with parameters: qtrim=r trimq=10 minlen=140 ftr=139. To assemble the quality filter reads of *S. aureus* SAU39 wild and resistant we used SPAdes 3.13.0 (BANKEVICH et al., 2012) with automatic parameters of kmers sizes. Generated contigs were filtered to remove short and redundant sequences using BBMap function dedupe.sh with parameters: minscaf=1000 sort=length minidentity=90 minlengthpercent=90. Open Read Frames (ORFs) were predicted with Prokka pipeline (SEEMANN, 2014). tRNA were detected with tRNAscan-SE2 (CHAN et al., 2021) and rRNAs with Barrnap (https://github.com/tseemann/barrnap). CRISPR regions were detected with MinCED (Bland et al. 2007). Plasmid searches were made using metagenomic plasmid function (metaplasmidspades.py)(ANTIPOV et al., 2019). Antibiotic resistance genes were detected using ABRicate (SEEMANN T, https://github.com/tseemann/abricate) with CARD database (ALCOCK et al., 2020). Prophages were detected using the PHASTER web server (ARNDT et al., 2016). Comparative genome analysis was performed using Snippy (SEEMANN T, https://github.com/tseemann/snippy), aligning SAU39 resistant quality reads against assembled contigs of SAU39 wild used as reference genome. Protein alignment was made using MUSCLE (EDGAR, 2004) in software MEGA X (KUMAR et al., 2018).

The single viral genome was *de novo* assembled with Tadpole using automatic kmer length optimization (Tadwrapper, BBTools, jgi.doe.gov/data-and-tools/bbtools). Open Read Frames (ORFs) were predicted for assembled contigs with Prodigal v2.6.3 (HYATT et al., 2010). These predicted ORFs were functionally annotated with BLASTp (ALTSCHUL et al., 1990) against Protein RefSeq Database. tRNA genes were detected with tRNAscan-SE (CHAN et al., 2021). We compared the gene hits table of isolated phage with closely related phage genomes using BLASTn. Maximum Likelihood phylogenetic analysis of whole genomic DNA was made with CS44 and 22 complete genomes of *Staphylococcus* phages available in GenBank (accession numbers of genomes are available in Appendix 2, Table S3). Those

sequences were aligned using Clustal Omega (SIEVERS; HIGGINS, 2018) and a phylogenomic tree generated with neighbor joining with P distance values and 1000 replicate bootstrap using MEGA X (KUMAR et al., 2018).

2.2.7 pH and thermal stability of Staphylococcus phage SP01

The thermostability and pH stability of CS44 were determined by plaque assay with different ranges of temperature and pH. We exposed 10^8 PFU/ml viral particles of CS44 suspensions in 0.9% NaCl to temperatures of 20, 25, 30, 35, 40, 45, 50, 55 e 60 °C at pH 7.0 for 6 hours. To pH stability we exposed viral particles with the same conditions to pHs of 2, 4, 6, 7, 8, 10, 12 at 25 °C for 6 hours (TANAKA et al., 2018). All experiments were made in triplicate. Before incubation they were serially diluted in 0.9% NaCl and titter by double agar layer method with modification (KAUFFMAN; POLZ, 2018). pH values of solutions were adjusted with 1M HCl or 1M NaOH and measured with pH-Indicator Strips (Millipore, Massachusetts, USA). Solutions with phages were incubated without shaking at Eppendorf® Thermomixer® R, dry block heating and cooling shaker (Eppendorf, Hamburgo, Germany).

2.2.8 Host range determination

To evaluate the host range of *Staphylococcus* phage vB-SauP_CS44 we realized plaque assays against 26 strains of *Staphylococcus* spp. (Appendix 2, Table S2). Strains of *S. aureus* Sta1-Sta12, *S. chromogenes* and *S. haemolyticus* were conceded by Dr. Marcus Brian from Faculdade de Medicina Veterinária e Zootecnia of Universidade de São Paulo. All bacteria were inoculated in BHI broth and incubated at 30 °C overnight without shaking. Those cultures were used to plaque assays with an optimized classic agar overlay method (KAUFFMAN; POLZ, 2018). Instead inoculate bacteria and phage at the same time and then add top agar, we inoculated only bacteria, added top agar homenizing bacteria and media, then spotted 10 µl of 1×10^9 PFU/ml of phage in the center of plates, before solidified top agar. Plates were incubated at 30 °C overnight.

2.2.9 Control of *Staphylococcus aureus* population

We evaluated the growth of *S. aureus* SAU39 and ATCC 6538 with and without the presence of CS44 into BHI broth by 24h. From an overnight culture of bacteria, we transferred 100 µl of culture into sterile erlenmeyer tubes with 10 ml of BHI broth. They were cultured at 37 °C with 200 rpm shaking for approximately 2h until reaching 10^8 CFU/ml (0.5 OD_{595}). Exponential growth culture was aliquoted in 96-well polystyrene plate, in each well we put 200 µl of bacterial culture plus 20 µl of phage solution with three different MOI 0.1, 1 and 10 to observe how the bacterial populations are controlled by phage. The plates were incubated at 37 °C without shaking by 24h of growth. OD_{595} was measured at each 30 min using Multiskan™ FC Microplate Photometer (Thermo Scientific). The control and treatments were made in triplicate and the experiment was made two times.

2.2.10 Adsorption, antibiotic and adherence

We carried out the adsorption test as described by DUPLESSIS; MOINEAU (2001) using phage CS44 and two strains of *S. aureus* SAU39, wild-type and resistant. The host strains were incubated in BHI broth until reaching 10^8 CFU/ml (OD_{595} of 0.5). We mixed 100 µl of diluted phages (10^4 PFU/ml) with 900 µl of wild-type, resistant and medium. These mixtures were incubated for 10 min at 25 °C to allow adsorption of the phages to the cells. After that, we centrifuged the mix for 1 min at 14000 rpm, supernatant was recovered and enumerated by plaque assay in triplicate. Antibiotic susceptibility test for the *S. aureus* strains was carried out using a Biomerieux Vitek 2 Compact, at the University of São Paulo Hospital.

The assessment of the biofilm-forming capacity of S. aureus isolates was performed as recommended by STEENACKERS et al. (2011), with some modifications. In a polystyrene 96 well microplate (Nunc, Thermo Scientific, Denmark), wells were filled with 20 µl of bacterium inoculum that was cultivated during 24 hours in Trypticase Soy Broth (TSB) adjusted to the 0.5 McFarland reference. The wells were then completed with TSB, obtaining a final volume of 200 µL. Afterwards, the cover lid containing pegs was placed over the plate which was incubated for 24 h / 37°C without shaking. Every 24 h the medium was replaced with new broth for a total time

of 72 h. Subsequently, the cover was washed with sterile Phosphate-Saline Buffer (PBS) and bacteria adhered to the surface of the cover pegs were fixed with 200 µl of 99% methanol for 15 minutes and then the pegs were set to dry at room temperature. The pegs were stained with 200 µl of a 0.3% (w/v) crystal violet solution for 5 minutes. Excess crystal violet was removed with distilled water and the pegs were dried at room temperature and placed in 33% glacial acetic acid (v/v) for 15 minutes. This last step was performed in order to dilute the violet crystal adhered to the biofilm formed on the surface of the pegs. Optical density was measured at 570 nm using the Multiskan FC microplate reader (Thermo Scientific). The test was performed three times.

Biofilm formation was classified as described by STEPANOVIC et al. (2000) where the OD “cut-off” (ODc) is defined as the OD of the negative control plus three times the standard deviation of the negative control. Biofilm formation of the bacterial strain was then determined following the criteria: OD ≤ ODc, Non-biofilm former; ODc < OD ≤ 2 × ODc, Weak biofilm former; 2 × ODc < OD ≤ 4 × ODc Moderate biofilm former; and 4 × ODc < OD, Strong biofilm former.

2.2.11 Electronic microscopy

We selected aliquots of purified CS44 and culture of SAU39 infected with CS44. Aliquots were adsorbed to glow-discharged carbon film-coated copper grids (400 Mesh, CF400-Cu, Electron Microscopy Sciences), washed with ultrapure water, and negatively stained with uranyl acetate 2% (w/v), being blotted on filter paper after each step. A FEI Tecnai G20 200 kV transmission electron microscope (Department of Cell and Developmental Biology, Institute of Biomedical Sciences, University of São Paulo) was used for image acquisition.

2.2.12 Visualizations

Circular representations of the SAU39 genomes were made using CGview Server (STOTHARD; GRANT; VAN DOMSELAAR, 2018). All analyses were carried out using the statistical software R (R DEVELOPMENT CORE TEAM, 2020), with packages ggplot2 (WICKHAM, 2016) and genoPlotR (GUY et al., 2011).

2.3 RESULTS

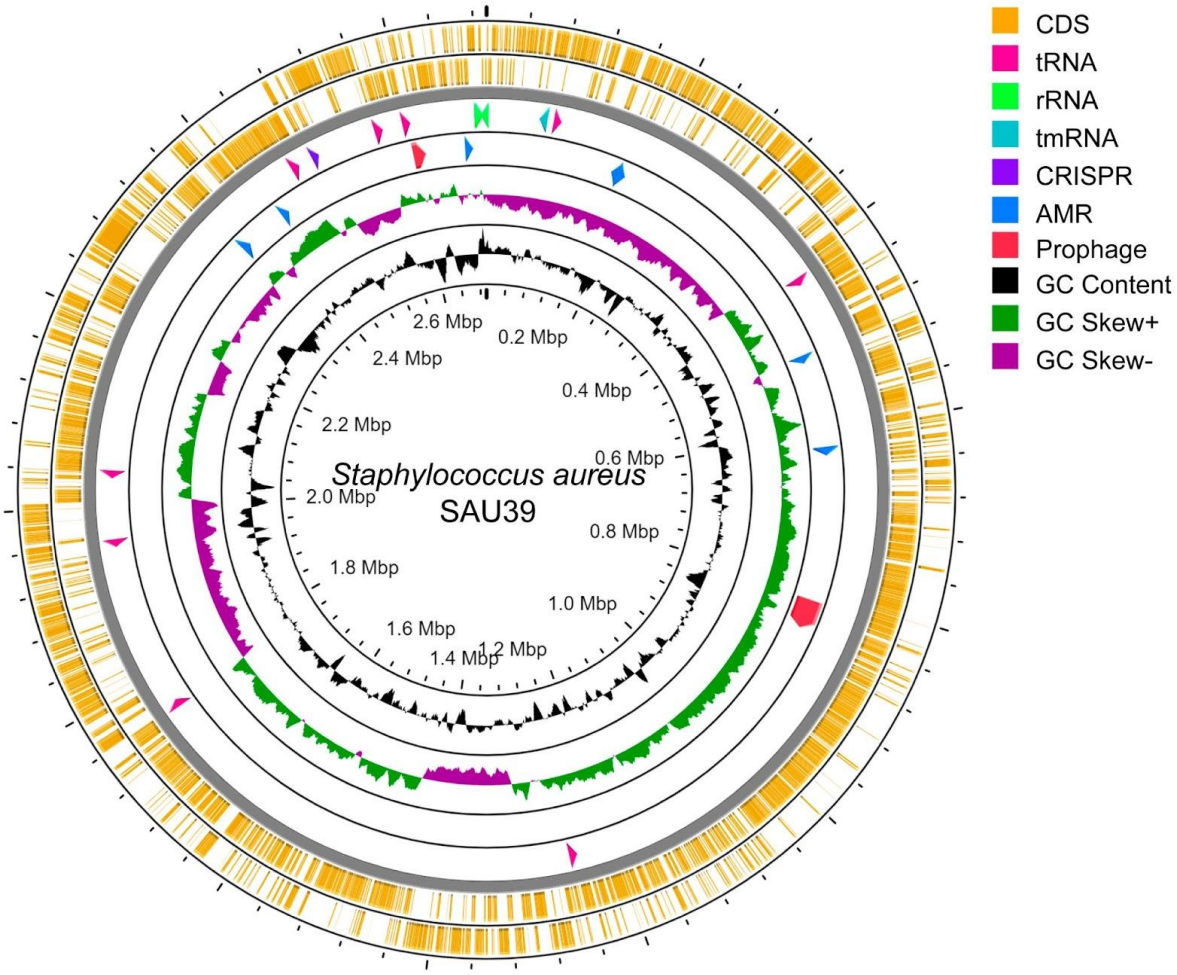
2.3.1 Isolation of phages and their host from cheese samples

Staphylococcus aureus strains from Canastra Cheese were previously isolated by CAMPOS et al. (2020). A selected two representative *Staphylococcus aureus* strain SAU39 for phage infection tests and SAU66 for genome comparative analysis, which were previously confirmed as *Staphylococcus aureus* using standard microbiological procedures. We performed infection tests against *Staphylococcus aureus* SAU39 and *Staphylococcus aureus* ATCC 29213 using virus-like particles (VLP's) obtained from starter cultures of 12 producers of artisanal Canastra cheese (Appendix 2, Table S1). We observed bacteriophage plaque formation in all *S. aureus* strains tested using VLPs obtained from producer P44 (Appendix 2, Fig. S1). A selected bacteriophage plaque was collected, purified and stored for further analysis.

2.3.2 *Staphylococcus aureus* SAU39 genome characterization

The genome of *Staphylococcus aureus* SAU39 isolate was 2,692,760 bp of total length, with 32.64% of GC-content and 2,492 predicted protein-encoding genes. A circular representation of the SAU39 genome is shown in Fig. 2.1. We assembled 45 contigs with N50 of 127,671 bp and 123x average genome coverage, containing 21 tRNA, three pseudogenes and three rRNAs coding regions with complete 16S and 23S genes, and a partial 5S coding region. No plasmids were detected in the assembly and we detected: one CRISPR array with three repeats with 29nt and 151 bp length; 9 antimicrobial resistance (AMR) genes, and two prophage regions.

Figure 2.1. Circular genome representation of *S. aureus* SAU39



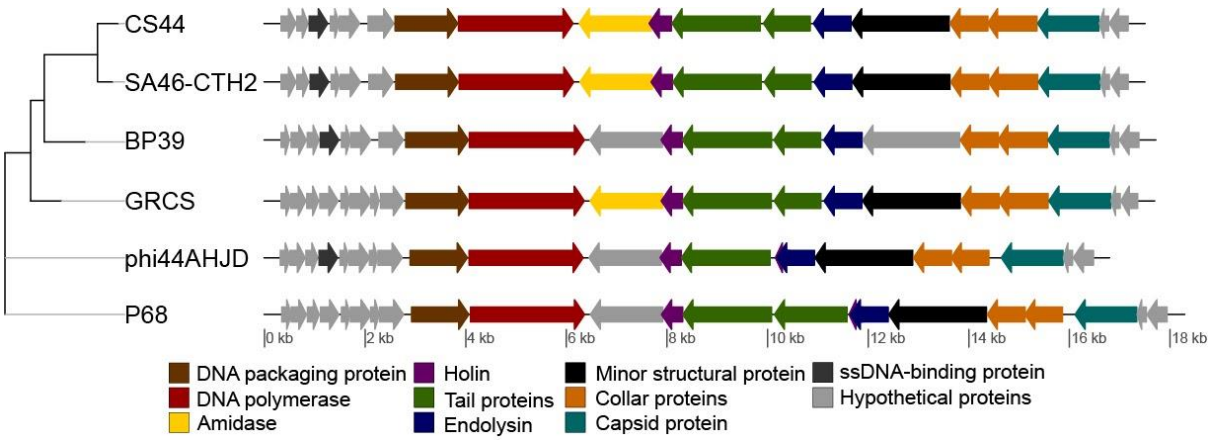
The illustrations were plotted using CGView and are organized in seven rings. The order of rings show the following information (from outermost to inmost): CDSs position in forward and reverse strand; tRNA, rRNA, tmRNA, and CRISPR array; antimicrobial resistance (AMR) and prophages; GC skew+ and GC skew-; and GC content.

2.3.3 *Staphylococcus* phage vB-SauP_CS44 genome characterization

We obtained one contig of 17,311-bp length with 36500x coverage, which was classified as *Staphylococcus* phage from the *Rosenblumvirus* genus (Picoviridae: Picovirinae). This new Podoviridae phage was named *Staphylococcus* phage vB-SauP_CS44, and will be called only as CS44 in the rest of the text. The GC content of CS44 was 29.1%, containing 19 open reading frames (ORFs). The ORFs code for proteins ranging in size from 61 to 760 amino acids, seven of which are classified as hypothetical proteins. The genes with predicted protein functions include those for single-stranded DNA binding protein, DNA packaging protein, DNA polymerase,

amidase, holin, major and minor tail protein, endolysin, lower and upper collar protein, minor structure protein, and major capsid protein (Fig. 2.2). No tRNA coding genes were detected.

Figure 2.2. Lytic bacteriophages of genera Rosenblumvirus (Podoviridae, Podovirinae)



Staphylococcus phage SA46-CTH2 was the most similar with **vB_SauP-CS44**, followed by GRCS and BP39 bacteriophages; 44AHJD and P68 were the least similar. Protein functions were predicted using BLASTp against RefSeq protein database.

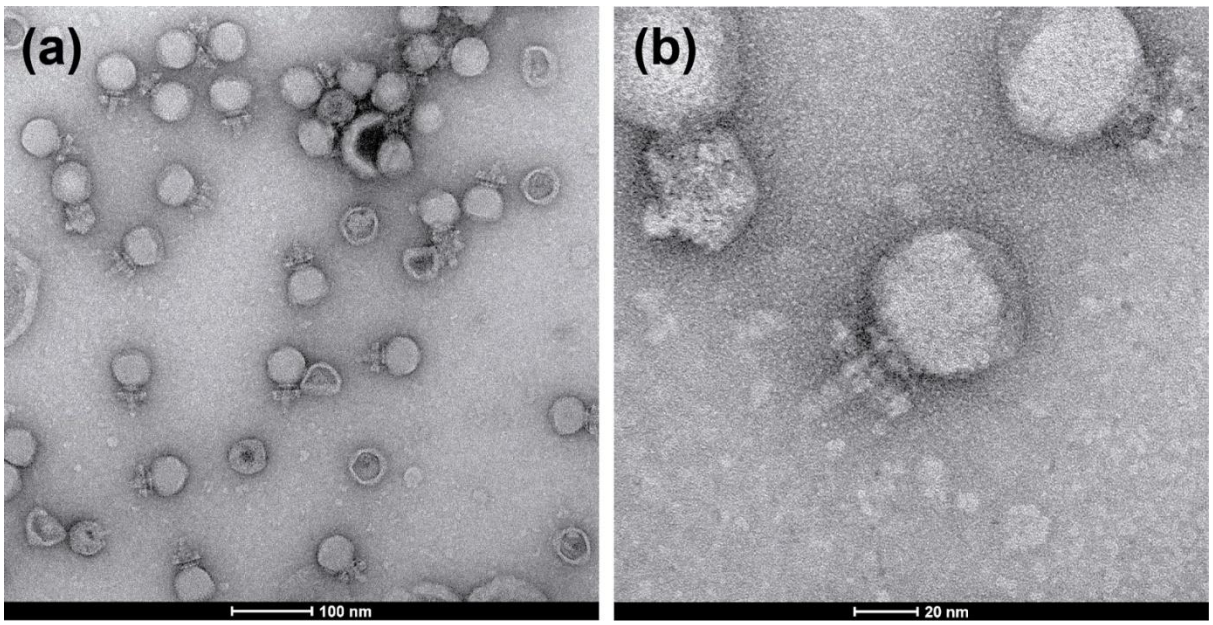
The most similar phages to vB_SauP-CS44 identified using BLASTn were *Staphylococcus* phages SA46-CTH2 (100% coverage and 93.43% identity), GRCS (96% coverage and 85.78% identity), BP39 (91% coverage and 85.30% identity), phi44AHJD (91% coverage and 84.37% identity), and P68 (93% coverage and 84.33% identity) (Appendix 2, Table S4). All these phages belong to *Rosenblumvirus* genus within the Podoviridae family. The genomes of CS44 phage and its closest relative SA46-CTH2 phage, contain 19 predicted proteins, while the next best matches, BP39 and GRCS, contain 21 and 22 genes respectively. All *Rosenblumvirus* genus main proteins are present in CS44 genome. The whole genome phylogenetic analysis demonstrated a cohesive group for all *Rosenblumvirus* genus phages, with a similarity between vB_SauP-CS44 and SA46-CTH2 phages (Appendix 2, Fig. S2).

2.3.4 Morphological characterization

Morphological characterization by transmission electron microscopy (TEM) with negative staining demonstrated that CS44 has an icosahedral head measuring 50

nm in length, and a short, 30 nm non-contractile tail (Fig. 2.3), thus confirming that it belongs to the Podoviridae family, based on genomics and morphological analysis.

Figure 2.3. Morphology of Staphylococcus phage vB-SauP_CS44 obtained by transmission electron microscopy of negative staining of phage.

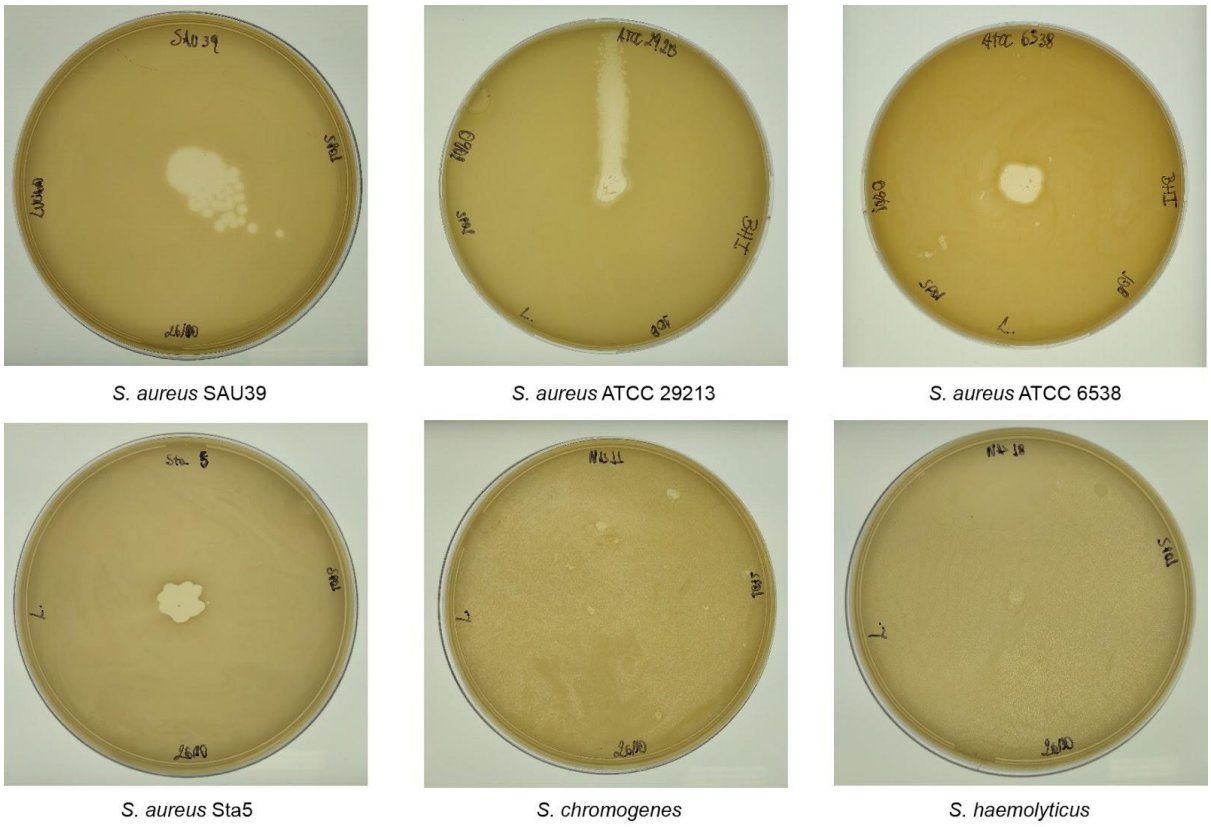


a. several viral particles at a magnification of 62,000 and bar 100 nm; b. detailed viral particles at a magnification of 240,000 and bar 20 nm. Scale bars are shown at the bottom of each image.

2.3.5 Determination of host range

We performed spot tests of CS44 against 26 *Staphylococcus spp.* strains (Appendix 2, Table S2), including a novel *S. aureus* SAU39 strain, obtained from a Canastra cheese sample, three references strains, *S. aureus* ATCC29213, *S. aureus* ATCC25923, and *S. aureus* ATCC6538; and 12 strains isolated from clinical and subclinical bovine mastitis cases. Additionally, we tested two other species, composed of eight strains of *S. chromogenes*, and two strains of *S. haemolyticus*. CS44 was able to infect all *S. aureus* strains tested and none of *S. chromogenes* and *S. haemolyticus* (Fig. 2.4; Appendix 2, Fig. S3), showing a high specificity to *Staphylococcus aureus*.

Figure 2.4. Determination of host range to *Staphylococcus* phage SP01 by spots test against Staphylococcus strains



S. aureus SAU39 | *S. aureus* ATCC 29213 | *S. aureus* ATCC 6538

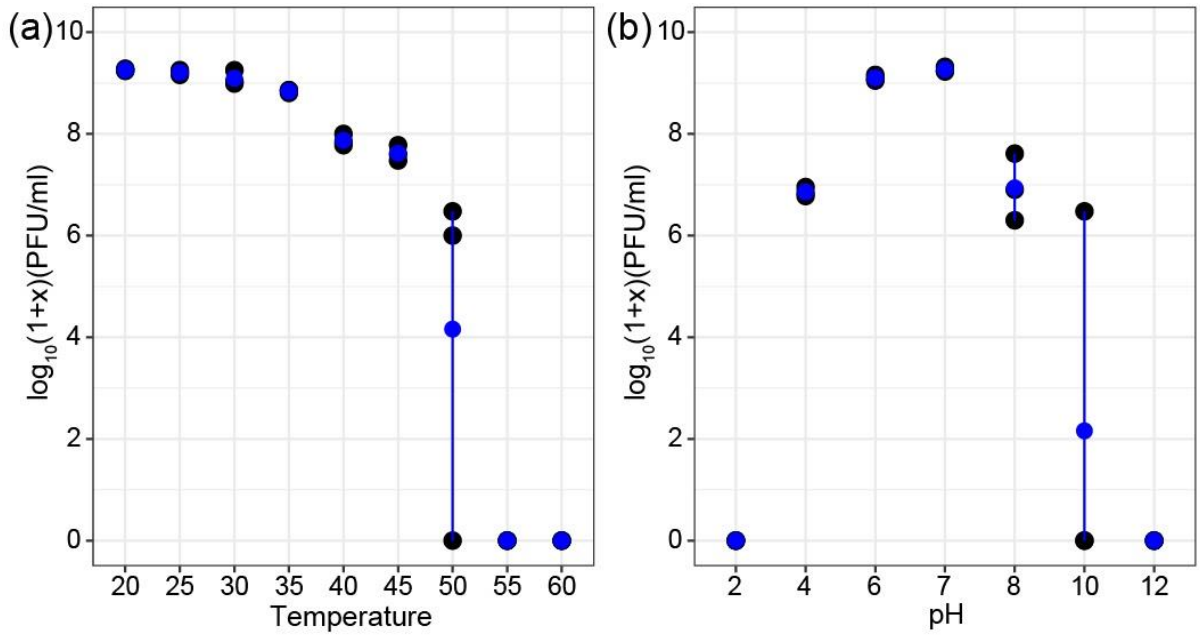
S. aureus Sta5 | *S. chromogenes* | *S. haemolyticus*

Plaque formation was observed on all strains of *Staphylococcus aureus* and none of *Staphylococcus chromogenes* and *Staphylococcus haemolyticus* (See also Appendix 2, Fig. S3).

2.3.6 *Staphylococcus* phage vB-SauP_CS44 stability

Thermostability and pH stability of CS44 were determined using plaque assays at different temperature and pH ranges. Purified viral particles were incubated at temperatures ranging from 20 to 60 °C, at pH 7.0, and at pH ranging from 2 to 12, at 25 °C, for 6 hours, followed by plaque titration. CS44 was stable when incubated at temperature up to 45 °C, albeit with a decrease in PFU formation observed in temperatures above 35 °C, and no plaques observed when incubated at temperatures higher than 50 °C (Fig. 2.5a). Viral particles were stable in pH ranging from 6 to 7, but were still recovered at pH as low as 4 and as high as 10, and were completely inactivated when incubated at pH 2 and 12 (Fig. 2.5b).

Figure 2.5. Thermostability and pH stability of CS44

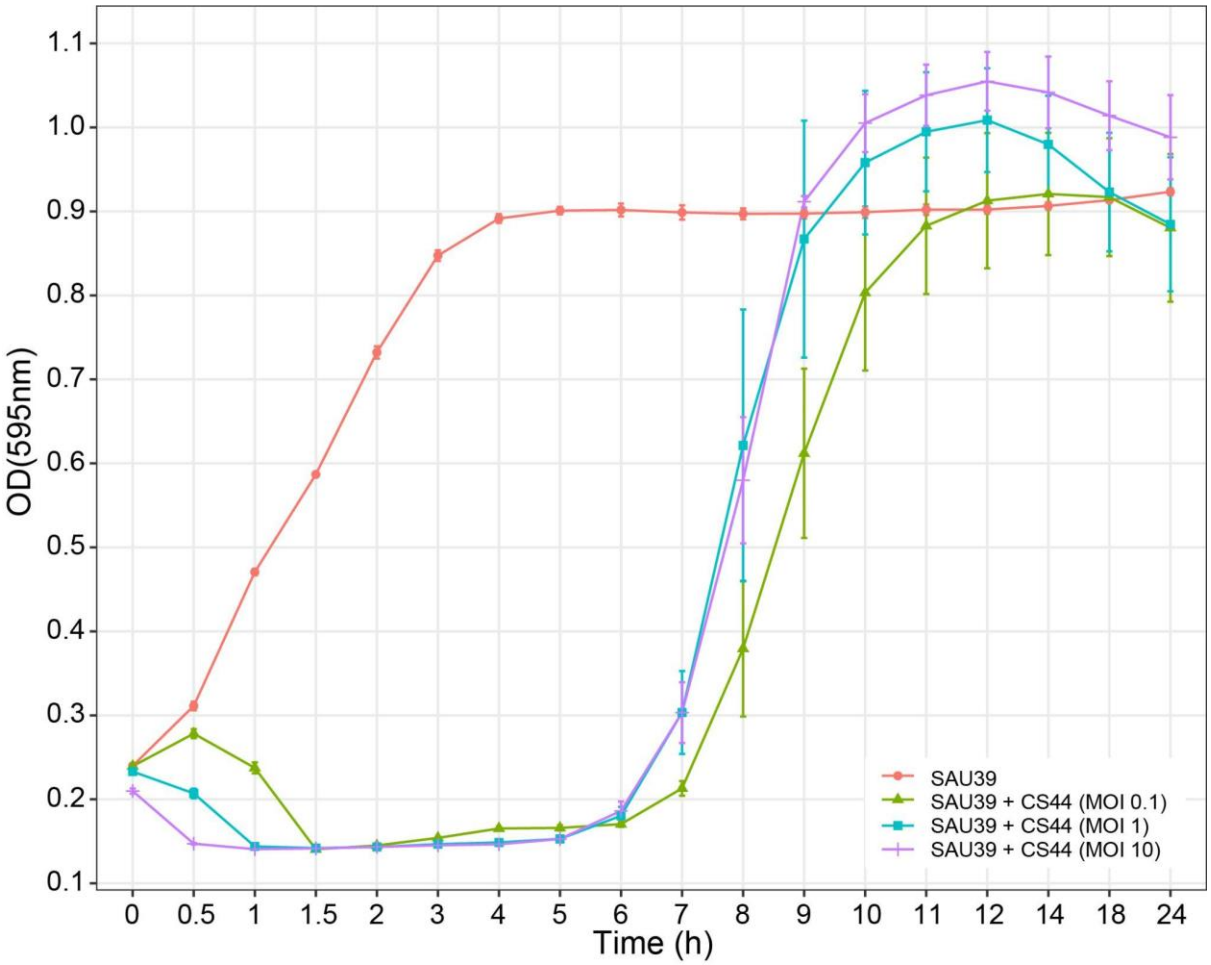


a. Thermal stability; b. pH stability. Values are expressed at $\log_{10}(1+x)$ scale of PFU/ml; blue dots and bars represent the mean and standard deviation (SD), respectively. All plaque assays were realized in triplicate.

2.3.7 Control of *Staphylococcus aureus* SAU39 population by vB_SauP-CS44

We evaluated the growth of *S. aureus* SAU39 for 24h incubating it with and without CS44. Three treatments with different multiplicity of infection (MOI) were used: 0.1, 1 and 10. Stationary phase growth was achieved within 4 to 5 hours for bacteria grown in absence of phage, while bacteria incubated in the presence of phage achieved stationary phase growth after 9 to 10 hours, for bacteria incubated with MOI 0.1 and 1, and after 12 hours for bacteria incubated with MOI 10 PFU (Fig. 2.6). Cell numbers were reduced within 30 minutes, 1h and 1.5h post incubations start, for MOI 10, 1, 0.1, respectively. Culture growth was kept stable (as determined by O.D. change), between 4.5 and 5.5 hours post intubation start. After that, bacterial growth was observed until it reached the stationary phase, indicating that the bacterial population arising in this mixed culture was resistant to the phage infection. The percentage of phage adsorbed onto resistant bacteria was not significantly different from the control (medium only), while sensitive bacteria showed 97% of phage adsorption, confirming the emergence of a phage resistant lineage of *S. aureus* SAU39.

Figure 2.6. Growth curve of *Staphylococcus aureus* SAU39 interacting with CS44



Growth curve of *Staphylococcus aureus* SAU39 strain without and with vB_SauP-CS44 phage at three MOI of 0.1, 1 and 10. The OD of *S. aureus* SAU39 suffer a high decrease at initial hours of experiment and in the next hours a reincrease.

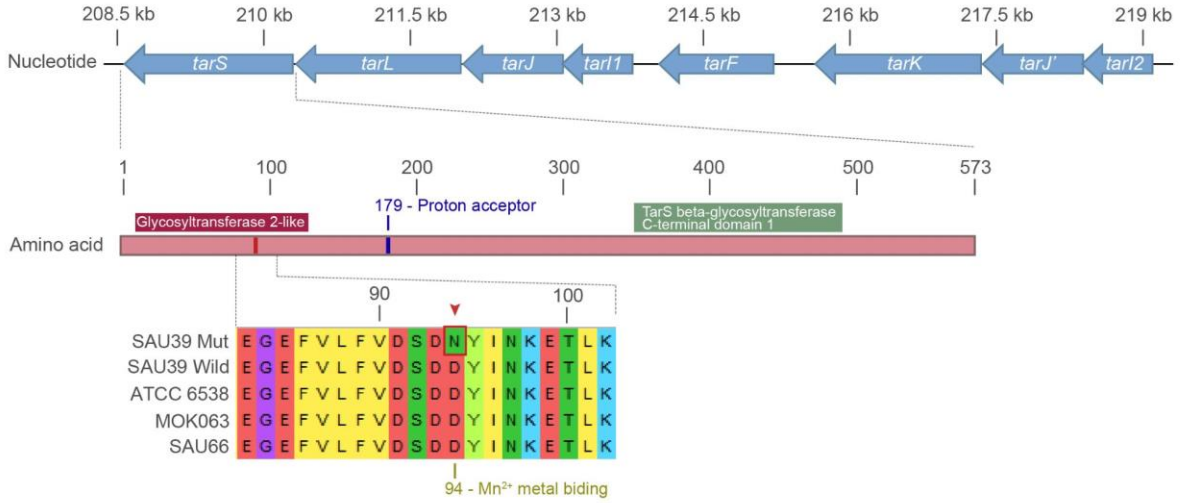
2.3.8 Comparative genomic analysis

To understand the emergence of resistant bacteria after vB_SauP-CS44 infection, we sequenced the genome of the resistant bacteria and compared it against the genome of wild type SAU39. As expected we observed a high level of overall similarity between the resistant mutant (complete description in Appendix 2) and wild type SAU39. We performed an in depth comparative analysis between resistant and susceptible type strains to try to determine the source of the resistant phenotype. Using Snippy software, we found 38 variant calls within 17 genes, and 10 variant calls in non-coding regions. These included 28 single nucleotide polymorphisms (SNP), 7 complex variant calls (combination of SNP and MNP), 2 multiple nucleotide polymorphisms (MNP), and 1 deletion (Appendix 2, Table S5). Ten genes were

annotated as hypothetical protein, two putative lipoprotein, one putative protein, one as putative tRNA-dihydrouridine synthase (*dus_2*), and the other three as fibronectin-binding protein A (*fnbA*), fibronectin-binding protein B (*fnbB*), and Poly(ribitol-phosphate) beta-N-acetylglucosaminyltransferase TarS (*tarS*). Genes *fnbA* and *fnbB* are involved in bacterial adhesion and biofilm formation, and gene *tarS* encodes an enzyme responsible for attachment of beta-O-GlcNAc (beta-O-N-acetyl-D-glucosamine) in poly(RboP)-wall teichoic acids (WTAs). Based on that, we focused our comparative analysis on genes *fnbA*, *fnbB*, and *tarS*.

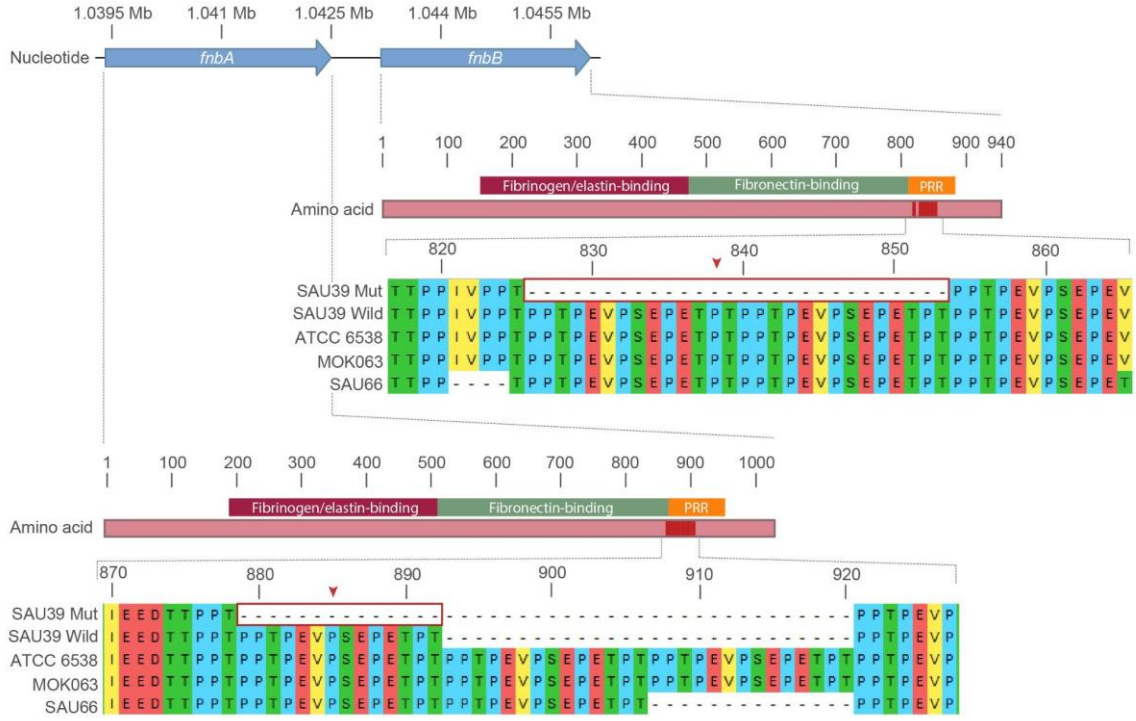
We aligned the protein sequences of *fnbA*, *fnbB* and *tarS* genes from wild type and resistant mutant SAU39, against sequences from strains ATCC 6538, MOK063, and SAU66 (another wild type strain isolated from Canastra cheese). A change in Aspartic acid (D) to Asparagine (N) was observed at position 94 of the TarS protein sequence (Fig. 2.7), responsible for metal coordination. This missense mutation was due to a point mutation at this codon's first position (G to A; evidence: A:307 G:1). We tested the wild type and phage-resistant bacteria against several antibiotics. Although both bacteria present the same sensitivity profile, the phage-resistant mutant presented a 50% decrease in the MIC for oxacillin (SAU39 wt: 0.5 µg/mL; SAU39mut: <= 0.25 µg/mL. For *fnbA* we detected a combination of single and multiple nucleotide polymorphism (evidence: TTTTG:111 GCTTT:0) and for *fnbB,* two SNPs at two positions. For both genes the nucleotide changes resulted in silent mutations, without changes in the codon translation. However, in protein alignment of FnBPA and FnBPB, we are able to observe deletions of several amino acids in a proline-rich repeat (PRR) region of both proteins (Fig. 2.8). *fnbAB* genes are involved in cell adhesion, and we therefore performed a biofilm formation essay to determine if these mutations would have an effect on this cell phenotype. Both isolates were classified as weak biofilm producers and we could not observe a significant difference between them for this phenotype (Appendix 2, Fig. S4).

Figure 2.7. Mutation in Poly(ribitol-phosphate) beta-N-acetylglucosaminyltransferase TarS protein



The TarS protein sequence of SAU39 Mutant was aligned against SAU39 Wild, ATCC 6538, MOK063 and SAU66. The teichoic acid ribitol (tar) biogenesis pathway genes were represented in blue arrows. TarS protein sequence was represented in light red with glycosyltransferase 2-like, TarS beta-glycosyltransferase-terminal domain 1 regions, proton acceptor binding site at position 179, and the point mutation detected in red at position 94 (the Mn^{2+} metal binding). The protein alignment was represented at the bottom of the figure and point mutation is indicating with a red arrow.

Figure 2.8. Mutation in fibronectin-binding protein A (FnBPA; at bottom), fibronectin-binding protein B (FnBPB; at top) proteins



The FnBPs sequence of SAU39 Mutant was aligned against SAU39 Wild, ATCC 6538, MOK063 and SAU66. The *fnbA* and *fnbB* genes were represented in blue arrows. FnBPs sequences were represented in light red with fibrinogen/elastin-binding, fibronectin-binding, and proline-rich region (PRR) region. Mutations detected are represented in red within PRR in both proteins. The protein alignment was represented at the bottom of the amino acid sequence with a red arrow indicating the deletions.

2.4 DISCUSSION

In this study, we characterized a novel lytic bacteriophage called vB-SauP_CS44 capable of infecting *Staphylococcus aureus*, belonging to *Rosenblumvirus* (Podoviridae) genus, isolated from an endogenous starter culture used in the Canastra cheese production region in Brazil. We characterized the genome, morphology, thermo- and pH stability, growth dynamics and host range of CS44. Additionally, we characterized the genome of one of its hosts, a new *S. aureus* strain (SAU39), isolated from mature cheese from the same production region, as well as a phage infection resistant mutant that emerged during the characterization experiments.

Our morphological and genomic analysis allowed us to place CS44 in the Podoviridae family. CS44 has a small genome size, with 17,311-bp and 19 ORFs, an icosahedral morphology and is closely related to phage SA46-CTH2 that has grouped with other phages from *Rosenblumvirus* genus. Several studies have isolated and characterized *Rosenblumvirus,* small lytic phages, such as P68, phi44AHJD (VYBIRAL et al., 2003), BP39 (BREYNE et al., 2017), GRCS (SUNAGAR; PATIL; CHANDRAKANTH, 2010), SCH1 (ALESHKIN et al., 2016), vB-SauP_phiAGO1.3 (GŁOWACKA-RUTKOWSKA et al., 2019), and SA46-CTH2 (DUC et al., 2020). This group is frequently isolated from sewage (BREYNE et al., 2017; SUNAGAR; PATIL; CHANDRAKANTH, 2010), and recently SA46-CTH2 was isolated from poultry meat (DUC et al., 2020). Although several phages have been isolated from dairy environments (MAHONY et al., 2017; MAHONY; VAN SINDEREN, 2014b), CS44 is the first *Rosenblumvirus* isolated from dairy and, more specifically, from endogenous starter culture, that infects *S. aureus* specifically.

The potential to use *Rosenblumvirus* phages in bacteriophage therapy has been demonstrated for some of them *in vitro* (DUC et al., 2020; GŁOWACKA-

RUTKOWSKA et al., 2019) and *in vivo* (BREYNE et al., 2017). These phages showed desired characteristics, such as the ability to completely lyse a bacterial culture, a wide host range, but limited to a single species, obligatory lytic life cycle, free of toxin genes, and high stability (KLUMPP; LOESSNER, 2013; KORTRIGHT et al., 2019). CS44 is lytic, shows a wide host range with high specificity to *S. aureus* strains, has no toxin genes coded in its genome and possesses high thermal and pH stabilities. All these characteristics demonstrate a high potential for *Staphylococcus aureus* pathogenic control, both in food and animal production, particularly in dairy environments, where it was originally isolated from.

Although CS44 is able to lyse bacterial cultures in different conditions, we observed the emergence of phage-resistant bacterial populations during the growth experiments. The emergence of phage-resistant bacteria has been observed in several bacterial species (CHAN et al., 2016; FERNÁNDEZ; RODRÍGUEZ; GARCÍA, 2018; FILIPPOV et al., 2011), including *Staphylococcus aureus* infected by *Rosenblumvirus* phage (DUC et al., 2020; GŁOWACKA-RUTKOWSKA et al., 2019). Some solutions to avoid the dominance and infection continuity of phage-resistant bacteria lineages is the use of phage cocktails instead of a single phage (DĄBROWSKA; ABEDON, 2019; HYMAN, 2019), and a combination of phages with alternative antimicrobial agents, such as bacteriocins (DUC et al., 2020). Although these strategies could be used for the development of stable commercial applications, much could be gained by a more complete understanding of the mechanisms behind the emergence of phage resistant bacteria. There are several examples in nature of the genetic trade-offs, where an organism's trait improves its fitness to a given condition (e.g. selective pressure by phage predation) while reducing performance in others (KORTRIGHT et al., 2019). Some studies with *Yersinia pestis* (FILIPPOV et al., 2011), *Klebsiella pneumoniae* (DE ANGELIS et al., 2021; HESSE et al., 2020), *Vibrio cholerae* (YEN; CAIRNS; CAMILLI, 2017), *Pseudomonas aerugin*osa (HOSSEINIDOUST; VAN DE VEN; TUFENKJI, 2013), and staphylococcal biofilms (GŁOWACKA-RUTKOWSKA et al., 2019) demonstrated that phage-resistant lineages were less virulent than phage sensitive (wild-type). *Rosenblumvirus* phage vB-SauP_phiAGO1.3 resistant bacteria have recently been shown to harbor a nonsense mutation in the *arl* gene, producing a non-functional

ArlS protein (GŁOWACKA-RUTKOWSKA et al., 2019). ArlS composes a two component *S. aureus* regulatory system ArlRS involved in the regulation of over 100 *S. aureus* genes, including virulence and adhesion related genes (FOURNIER; KLIER; RAPOPORT, 2001; JENUL; HORSWILL, 2019).

In our study, we selected a phage-resistant S. aureus SAU39 mutant that contained mutations in *tarS*, *fnbA* and *fnbB genes.* The glycosyltransferase TarS is an important component of teichoic acid ribitol (*tar*) biogenesis pathway, catalyzing β-O-GlcNAcylation of the wall teichoic acid (WTA) polymer (WINSTEL; XIA; PESCHEL, 2014). Here, we observed a missense mutation that affects the *tarS catalytic site,* changing the amino acid D at position 94 to an N, thus modifying its ability to coordinate the ion Mn^{2+} (SOBHANIFAR et al., 2016). This teichoic acid modification is essential for the phage adsorption to *S. aureus'* cell wall, and this mutation likely hinders or prevents a productive infection. The lack of this WTA modification have also been associated with a higher susceptibility to β-lactam antibiotics, including in Methicillin-resistant *Staphylococcus aureus* (MRSA) (BROWN et al., 2012; GERLACH et al., 2018; WINSTEL; XIA; PESCHEL, 2014). Thus, phage such as CS44 could be used to kill antibiotic resistant bacteria, such as MRSA, and any phage-infection resistant mutant arising could be treated with available antibiotics, providing an attractive avenue for treatment of this type of bacterial infection.

Fibronectin binding proteins (FnBPA and FnBPB) are important structures for adhesion and invasion into epithelial cells and biofilm development (FOSTER et al., 2014; JOSSE; LAURENT; DIOT, 2017; SPEZIALE; ARCIOLA; PIETROCOLA, 2019), composing the class of MSCRAMMs (Microbial surface component recognizing adhesive matrix molecules) of surface proteins (FOSTER et al., 2014). Both proteins have similar organizations, containing a proline-rich repeats (PRR) sequence in their C-terminal regions (ARCIOLA; CAMPOCCIA; MONTANARO, 2018). The complete deletion of *fnbA* and *fnbB genes* (GEOGHEGAN et al., 2013) or changes in Fibrinogen/Elastin-binding domain and Fibronectin-binding repeats (KEANE et al., 2007) have been shown to affect *S. aureus* biofilm formation and adhesion. We identified that the *S. aureus* SAU39 phage resistant mutant also has deletions in proline-rich repeats (PRR) of *fnbA* and *fnbB*. A large part of these PRR regions is

encompassed within the *S. aureus* cell wall (STEMBERK et al., 2014), and it is possible it may be associated with peptidoglycan attachment (WILLIAMSON, 1994). Thus, FnBPs with short PRR domain may block the attachment or produce a fragile structure, reducing the adhesion and biofilm development capacity of resistant cells. We could not observe differences between wild type and phage-resistant bacteria biofilm formation. This may be explained by the inherently low biofilm production capacity of these strains, or perhaps that the mutations detected in the PRR are not determinant for biofilm formation. More studies are necessary to fully characterize the importance of the PRR region for biofilm formation, and how other sensitive *S. aureus* strains would respond to *Rosenblumvirus* infection.

2.5 CONCLUSION

We have demonstrated the emergence of phage resistance in a wild type *S. aureus* strain, accompanied by mutations in *tarS*, *fnbA* and *fnbB genes,* corroborating the hypothesis that increasing fitness of some traits may reduce performance in others (KORTRIGHT et al., 2019). *S. aureus* SAU39 resistant to infection will be more susceptible to β-lactam antibiotics. Although we observe the emergence of phage resistance, bacteriophage therapy using CS44 could be carried out in combination with β-lactam antibiotics or with other phages that use different binding receptors. Nevertheless, more studies are required to evaluate the inactivation of *tarS*.

2.6 REFERENCES

AGGARWALA, V.; LIANG, G.; BUSHMAN, F. D. Viral communities of the human gut: metagenomic analysis of composition and dynamics. **Mobile DNA**, v. 8, n. 1, p. 12, 2017.

ALCOCK, B. P. et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. **Nucleic Acids Research**, v. 48, n. D1, p. D517–D525, 2020.

ALESHKIN, A. et al. Phagebiotics in treatment and prophylaxis of healthcareassociated infections. **Bacteriophages: An Overview and Synthesis of a Re-Emerging Field**, v. 6, n. 4, p. 105–122, 2016.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–10, 1990.

ANTIPOV, D. et al. Plasmid detection and assembly in genomic and metagenomic data sets. **Genome Research**, v. 29, n. 6, p. 961–968, 2019.

ARCIOLA, C. R.; CAMPOCCIA, D.; MONTANARO, L. Implant infections: Adhesion, biofilm formation and immune evasion. **Nature Reviews Microbiology**, v. 16, n. 7, p. 397–409, 2018.

ARNDT, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. **Nucleic acids research**, v. 44, n. W1, p. W16–W21, 2016.

BANKEVICH, A. et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. **Journal of Computational Biology**, v. 19, n. 5, p. 455–477, maio 2012.

BREYNE, K. et al. Efficacy and safety of a Bovine-Associated Staphylococcus aureus Phage Cocktail in a murine model of mastitis. **Frontiers in Microbiology**, v. 8, n. NOV, p. 1–11, 2017.

BROWN, S. et al. Methicillin resistance in Staphylococcus aureus requires glycosylated wall teichoic acids. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 46, p. 18909–18914, 2012.

CAMPOS, G. Z. et al. Microbiological characteristics of canastra cheese during manufacturing and ripening. **Food Control**, v. 121, n. August 2020, p. 107598, 2021.

CHAN, B. K. et al. Phage selection restores antibiotic sensitivity in MDR Pseudomonas aeruginosa. **Scientific Reports**, v. 6, n. May, p. 1–8, 2016.

CHAN, P. P. et al. TRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. **bioRxiv**, p. 1–37, 2021.

DĄBROWSKA, K.; ABEDON, S. T. Pharmacologically Aware Phage Therapy: Pharmacodynamic and Pharmacokinetic Obstacles to Phage Antibacterial Action in Animal and Human Bodies. **Microbiology and Molecular Biology Reviews**, v. 83, n. 4, p. 1–25, 30 out. 2019.

DE ANGELIS, L. H. et al. Phage resistance is associated with decreased virulence in kpc-producing klebsiella pneumoniae of the clonal group 258 clade II lineage. **Microorganisms**, v. 9, n. 4, 2021.

DEGHORAIN, M.; VAN MELDEREN, L. The staphylococci phages family: An overview. **Viruses**, v. 4, n. 12, p. 3316–3335, 2012.

DUC, H. M. et al. Isolation and application of bacteriophages alone or in combination with nisin against planktonic and biofilm cells of Staphylococcus aureus. **Applied Microbiology and Biotechnology**, 2020.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792, 2004.

FERNÁNDEZ, L.; RODRÍGUEZ, A.; GARCÍA, P. Phage or foe: an insight into the impact of viral predation on microbial communities. **ISME Journal**, p. 1–9, 2018.

FILIPPOV, A. A. et al. Bacteriophage-resistant mutants in yersinia pestis: Identification of phage receptors and attenuation for mice. **PLoS ONE**, v. 6, n. 9, 2011.

FOSTER, T. J. et al. Adhesion, invasion and evasion: The many functions of the surface proteins of Staphylococcus aureus. **Nature Reviews Microbiology**, v. 12, n. 1, p. 49–62, 2014.

FOURNIER, B.; KLIER, A.; RAPOPORT, G. The two-component system ArlS-ArlR is a regulator of virulence gene expression in Staphylococcus aureus. **Molecular Microbiology**, v. 41, n. 1, p. 247–261, 2001.

GEOGHEGAN, J. A. et al. Subdomains N2N3 of fibronectin binding protein a mediate staphylococcus aureus biofilm formation and adherence to fibrinogen using distinct mechanisms. **Journal of Bacteriology**, v. 195, n. 11, p. 2675–2683, 2013.

GERLACH, D. et al. Methicillin-resistant Staphylococcus aureus alters cell wall glycosylation to evade immunity. **Nature**, v. 563, n. 7733, p. 705–709, 2018.

GŁOWACKA-RUTKOWSKA, A. et al. The Ability of Lytic Staphylococcal Podovirus vB_SauP_phiAGO1.3 to Coexist in Equilibrium With Its Host Facilitates the Selection of Host Mutants of Attenuated Virulence but Does Not Preclude the Phage Antistaphylococcal Activity in a Nematode Infection M. **Frontiers in Microbiology**, v. 9, n. 7, p. e11720, 18 jan. 2019.

GUY, L. et al. GenoPlotR: comparative gene and genome visualization in R. **Bioinformatics**, v. 27, n. 13, p. 2334–2335, 2011.

HESSE, S. et al. Phage Resistance in Multidrug-Resistant Klebsiella pneumoniae ST258 Evolves via Diverse Mutations That Culminate in Impaired Adsorption. **mBio**, v. 11, n. 1, p. 1–14, 25 fev. 2020.

HOSSEINIDOUST, Z.; VAN DE VEN, T. G. M.; TUFENKJI, N. Evolution of pseudomonas aeruginosa virulence as a result of phage predation. **Applied and Environmental Microbiology**, v. 79, n. 19, p. 6110–6116, 2013.

HYATT, D. et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, v. 11, 2010.

HYMAN, P. Phages for phage therapy: Isolation, characterization, and host range breadth. **Pharmaceuticals**, v. 12, n. 1, 2019.

JAKOČIŪNĖ, D.; MOODLEY, A. A Rapid Bacteriophage DNA Extraction Method. **Methods and Protocols**, v. 1, n. 3, p. 27, 2018.

JENUL, C.; HORSWILL, A. R. Regulation of Staphylococcus aureus Virulence . **Microbiology Spectrum**, v. 7, n. 2, p. 1–21, 2019.

JOSSE, J.; LAURENT, F.; DIOT, A. Staphylococcal adhesion and host cell invasion: Fibronectin-binding and other mechanisms. **Frontiers in Microbiology**, v. 8, n. DEC, p. 1–8, 2017.

KAMIMURA, B. A. et al. Large-scale mapping of microbial diversity in artisanal Brazilian cheeses. **Food Microbiology**, v. 80, n. October 2018, p. 40–49, 2019.

KAUFFMAN, K. M. et al. A major lineage of non-tailed dsDNA viruses as unrecognized killers of marine bacteria. **Nature**, v. 554, n. 7690, p. 118–122, 24 jan. 2018.

KAUFFMAN, K. M.; POLZ, M. F. Streamlining standard bacteriophage methods for higher throughput. **MethodsX**, v. 5, p. 159–172, 2018.

KEANE, F. M. et al. Fibrinogen and elastin bind to the same region within the A domain of fibronectin binding protein A, an MSCRAMM of Staphylococcus aureus. **Molecular Microbiology**, v. 63, n. 3, p. 711–723, 2007.

KLUMPP, J.; LOESSNER, M. J. Listeria phages: genomes, evolution, and application. **Bacteriophage**, v. 3.3, n. September, p. 8, 2013.

KORTRIGHT, K. E. et al. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. **Cell Host and Microbe**, v. 25, n. 2, p. 219–232, 2019.

KUMAR, S. et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular Biology and Evolution**, v. 35, n. 6, p. 1547–1549, 1 jun. 2018.

MAHONY, J. et al. Phage biodiversity in artisanal cheese wheys reflects the complexity of the fermentation process. **Viruses**, v. 9, n. 3, p. 1–18, 2017.

MAHONY, J.; VAN SINDEREN, D. Current taxonomy of phages infecting lactic acid bacteria. **Frontiers in Microbiology**, v. 5, n. JAN, p. 1–7, 2014.

MATUSZEWSKA, M. et al. The Evolutionary Genomics of Host Specificity in Staphylococcus aureus. **Trends in Microbiology**, v. 28, n. 6, p. 465–477, 2020.

MOLLER, A. G.; LINDSAY, J. A.; READ, T. D. Determinants of phage host range in *Staphylococcus* species. **Applied and Environmental Microbiology**, v. 85, n. 11, p. 1–16, 2019.

MUHAMMED, M. K. et al. Metagenomic Analysis of Dairy Bacteriophages: Extraction Method and Pilot Study on Whey Samples Derived from Using Undefined and Defined Mesophilic Starter Cultures. **Applied and Environmental Microbiology**, v. 83, n. 19, p. e00888-17, 1 out. 2017.

NOVICK, R. P.; CHRISTIE, G. E.; PENADÉS, J. R. The phage-related chromosomal islands of Gram-positive bacteria. **Nature Reviews Microbiology**, v. 8, n. 8, p. 541–551, 2010.

PAEZ-ESPINO, D. et al. Uncovering Earth's virome. **Nature**, v. 536, n. 7617, p. 425–430, 2016.

PERIN, L. M. et al. Bacterial ecology of artisanal Minas cheeses assessed by culture-dependent and -independent methods. **Food Microbiology**, v. 65, p. 160–169, 2017.

PINEDA, A. P. A. et al. Brazilian Artisanal Cheeses: Diversity, Microbiological Safety, and Challenges for the Sector. **Frontiers in Microbiology**, v. 12, n. April, 2021.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.**Vienna, Austria.R Foundation for Statistical Computing, , 2020. Disponível em: http://www.r-project.org/.

SEEMANN, T. Prokka: Rapid prokaryotic genome annotation. **Bioinformatics**, v. 30, n. 14, p. 2068–2069, 2014.

SIEVERS, F.; HIGGINS, D. G. Clustal Omega for making accurate alignments of many protein sequences. **Protein Science**, v. 27, n. 1, p. 135–145, 2018.

SOBHANIFAR, S. et al. Structure and Mechanism of Staphylococcus aureus TarS, the Wall Teichoic Acid β-glycosyltransferase Involved in Methicillin Resistance. **PLoS Pathogens**, v. 12, n. 12, p. 1–24, 2016.

SPEZIALE, P.; ARCIOLA, C. R.; PIETROCOLA, G. Fibronectin and Its Role in Human Infective Diseases. **Cells**, v. 8, n. 12, 2019.

STEMBERK, V. et al. Evidence for steric regulation of fibrinogen binding to staphylococcus aureus fibronectin-binding protein a(FnBPA). **Journal of Biological Chemistry**, v. 289, n. 18, p. 12842–12851, 2014.

STEENACKERS, et al. Structure–activity relationship of 2-hydroxy-2-aryl-2,3-dihydroimidazo [1,2-a]pyrimidinium salts and 2N-substituted 4(5)-aryl-2-amino-1H-imidazoles as inhibitors of biofilm formationby *Salmonella* Typhimurium and *Pseudomonas aeruginosa*. **Bioorganic Medical Chemistry**, v.19, 3462-3473, 2011.

STEPANOVIC, S.; VUKOVIC, D.; DAKIC, I.; SAVIC, B.; S˘VABIC´-VLAHOVIC, M. A modified microtiter-plate test for quantification of Staphylococcal biofilm formation. **Journal of Microbiological Methods**, v.40, 175-179, 2000.

STOTHARD, P.; GRANT, J. R.; VAN DOMSELAAR, G. Visualizing and comparing circular genomes using the CGView family of tools. **Briefings in Bioinformatics**, v. 20, n. 4, p. 1576–1582, 2018.

SUNAGAR, R.; PATIL, S. A.; CHANDRAKANTH, R. K. Bacteriophage therapy for Staphylococcus aureus bacteremia in streptozotocin-induced diabetic mice. **Research in Microbiology**, v. 161, n. 10, p. 854–860, 2010.

TANAKA, C. et al. A Lytic Bacteriophage for Controlling Pseudomonas lactis in Raw Cow's Milk. **Applied and Environmental Microbiology**, v. 84, n. 18, p. AEM.00111-18, 6 jul. 2018.

VAN DALEN, R.; PESCHEL, A.; VAN SORGE, N. M. Wall Teichoic Acid in Staphylococcus aureus Host Interaction. **Trends in Microbiology**, v. 28, n. 12, p. 985–998, 2020.

VYBIRAL, D. et al. Complete nucleotide sequence and molecular characterization of two lytic Staphylococcus aureus phages: 44AHJD and P68. **FEMS Microbiology Letters**, v. 219, n. 2, p. 275–283, fev. 2003.

WICKHAM, H. **ggplot2**. Cham: Springer International Publishing, 2016.

WILLIAMSON, M. P. The structure and function of proline-rich regions in proteins. **Biochemical Journal**, v. 297, n. 2, p. 249–260, 1994.

WINSTEL, V.; XIA, G.; PESCHEL, A. Pathways and roles of wall teichoic acid glycosylation in Staphylococcus aureus. **International Journal of Medical Microbiology**, v. 304, n. 3–4, p. 215–221, 2014.

XIA, G.; WOLZ, C. Phages of Staphylococcus aureus and their impact on host evolution. **Infection, Genetics and Evolution**, v. 21, p. 593–601, 2014.

YEN, M.; CAIRNS, L. S.; CAMILLI, A. A cocktail of three virulent bacteriophages prevents Vibrio cholerae infection in animal models. **Nature Communications**, v. 8, p. 1–7, 2017.

3 CHAPTER 3 – A PHYLOGENOMIC VIEW OF THE RELATIONSHIP BETWEEN *LEUCONOSTOC* SPP AND THEIR LYSOGENIC PHAGE FROM BOZA, A TRADITIONAL FERMENTED BEVERAGE

> "It's the emotion in the seller's voice that really sells the boza". Mevlut Karataş, boza seller in Istanbul from A Strangeness in My Mind by Orhan Pamuk

3.1 INTRODUCTION

Boza is a traditional fermented cereal beverage, popular in different countries from the region of the Balkan Peninsula. It is made by one or a combination of diverse kinds of grains, such as millet, maize, wheat, or rice semolina or flour. Historical records link this product to ancient civilization from the Middle East region; however, it has been spread over the region by the Ottoman Empire. In the last 20 years several publications have been focused on the microbiological status of boza (LEBLANC; TODOROV, 2011; TODOROV; HOLZAPFEL, 2015).

Several species of Lactic Acid Bacteria (LAB) have been isolated from boza. According to studies that evaluated boza's microbial community from different regions, the boza communities were composed by microorganisms from genera *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Enterococcus*, *Pediococcus*, *Oenococcus*, and *Weissella* (ALTAY et al., 2013). At the species level, ther are several strains of *Lactobacillus acidophilus*, *Lactobacillus fermentum*, *Lactobacillus paracasei*, *Lactobacillus pentosus*, *Lactobacillus plantarum*, *Lactococcus lactis subsp. lactis*, *Leuconostoc mesenteroides*, *Leuconostoc lactis*, *Enterococcus faecium*, *Pediococcus pentosaceus*, *Oenococcus oeni*, *Weissella confusa* and others (TODOROV; HOLZAPFEL, 2015).

Different species of LAB isolated from boza have their bacteriocins and antimicrobial metabolites characterized (LEBLANC; TODOROV, 2011). These bacteriocins showed activity against potential pathogens, such as *Listeria monocytogenes*, *Staphylococcus aureus*, *Escherichia coli*, and *Salmonella typhimurium* (BALCIUNAS et al., 2013). In addition to the well accepted specificity of bacteriocins' mode of action against closely related species (CHIKINDAS et al., 2018), some report have foucused on an unorthodox bacteriocin activity, such as inhibition of *Mycobacterium*

spp., fungi, and some viruses (TODOROV; CHIKINDAS, 2020; TODOROV; DE MELO FRANCO; TAGG, 2019).

Bacteriophages (or phages), viruses that can infect bacteria, have been studied in several types of fermented foods (LEDORMAND; DESMASURES; DALMASSO, 2020; MARCÓ; MOINEAU; QUIBERONI, 2012). Phages can follow a lytic or lysogenic cycle: lytic phages will shut down the defense mechanism of bacteria, proceed with genome replication and phage particle assembly, subsequently bursting the host cell as new phage particles are released on the environment; lysogenic phages can incorporate their genomes in their host's genome and be passed on as the bacteria replicates (FEINER et al., 2015). These genome incorporated phages are named prophages. Lysogenic phages have been observed in LAB, mainly in *Lactococcus* and *Lactobacillus*, and less frequent in *Streptococcus* and *Leuconostoc* strains (MARCÓ; MOINEAU; QUIBERONI, 2012). Depending on the environmental conditions, these prophages can turn to a lytic cycle, killing their bacterial host cell, releasing new viral particles and infecting other susceptible bacteria. However, lysogenic phages can also show positive effects, such as conferring resistance to their host against infection by other phages (CORREA et al., 2021).

Phage-bacteria interaction in fermented foods can affect starter and adjunct starter cultures, causing pH increase, loss of sensorial characteristics and even inhibition of the fermentation process altogether (MAHONY et al., 2016). Dairy environments were the most explored due the problems caused by phage infections in dairy industries (MAHONY et al., 2017; MURPHY et al., 2013). These positive and negative consequences can also be observed in other fermented foods and beverages in industrial and artisanal conditions (HAYES et al., 2017; WÜNSCHE, 1989). Nevertheless, few studies have focused on understanding the phage communities in traditional beverages (LEDORMAND; DESMASURES; DALMASSO, 2020).

Here we investigate the dynamics of bacterial and phage communities in Boza, a traditional fermented beverage from the Balkans. Our study aim was to describe the bacterial and phage community compositions and their potential interactions. We characterized the Boza microbiome using whole-metagenome shotgun sequencing,

determining its taxonomic and functional profiles based on clade-specific markers as well as the viral composition based on virus hallmark genes. We recovered metagenome assembled genomes (MAGs), and performed a comparative genome analysis of the principal MAG obtained, highlighting potential phylogenomic relationships with its lysogenic phage.

3.2 MATERIAL AND METHODS

3.2.1 Sampling, DNA Extraction and Sequencing

The boza sample analyzed was obtained from a local market in the town of Belogradchik, in the North-West of Bulgaria. Total DNA was extracted from 250 µL of one Boza sample using the PowerMax® Soil DNA Isolation Kit (Qiagen Inc., Hilden, Germany) according to the manufacturer's instructions. The integrity of the DNA extracted was evaluated by electrophorese. DNA was quantified using Quant-iT™ PicoGreen™ dsDNA Assay Kit (Thermo Scientific, Waltham, MA, USA) and Synergy H1 Hybrid Reader (BioTek, Winooski, VT, USA). Extracted DNA was used as input for the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) to generate dual-indexed paired-end Illumina sequencing libraries according to the manufacturer's instructions. Paired-end reads were obtained using 2 sequencing runs using the NextSeq Genome Sequencer (Illumina), with a NextSeq 500/550 Mid Output Kit v2.5, at Core Facility for Scientific Research – University of Sao Paulo (CEFAP-USP).

3.2.2 Quality filter, taxonomic and functional profiles and contig assembly

The quality of raw sequences was verified using FastQC v0.11.9. NextSeq adapters were removed using BBDuk (BBTools, https://jgi.doe.gov/data-and-tools/bbtools/) with the following parameters: ktrim=r k=23 mink=11 hdist=1. The quality trimming was also done using BBDuk with parameter trimq=15. Compositional and functional profiles of microbial metagenomes samples were assessed using MetaPhlAn2 (v2.7.5) (TRUONG et al., 2015) and HUMAnN2 (v2.8.1) (FRANZOSA et al., 2018). HUMAnN 2.0 was used to determine the presence/absence and abundance of genetic functions (UniRef90). The obtained gene abundances were regrouped to

KEGG Orthogroups (KOs) and normalized as counts per million (CPM). The microbial metagenome was assembled using the quality filtered reads and the software SPAdes 3.15.0 (NURK et al., 2017), with metagenomic function (metaspades.py) and automatic parameters of kmers sizes. Plasmid searches were made using metagenomic plasmid function (metaplasmidspades.py) (ANTIPOV et al., 2019) and the presence of plasmid genes was verified using the script viralVerify (ANTIPOV et al., 2020) with -p argument. Contigs shorter than 1000 bp were removed using the reformat.sh function from BBTools.

3.2.3 Metagenome-assembled genomes (MAGs) and pagenome

Metagenome-assembled genomes (MAGs) were recovered using filtered contigs from Anvi'o 6.1 Metagenomic Workflow (EREN et al., 2015). Briefly, we created an anvi'o profile database with contig length cutoff of 2,500 bp; filtered reads were mapped to the contigs using Bowtie2 (LANGMEAD; SALZBERG, 2012); SAM files were converted to BAM file using SAMtools (LI et al., 2009); sequence homologs were searched and added to contigs database with hidden Markov Model (HMM) using HMMER (FINN; CLEMENTS; EDDY, 2011); genes were functionally annotated using NCBI's Clusters of Orthologous Groups (TATUSOV et al., 2001) and taxonomically annotated using Centrifuge (KIM et al., 2016). Metagenomic assemble genomes (MAGs) were obtained by contig binning using CONCOCT (ALNEBERG et al., 2014) and the generated bins were refined manually using anvio-refine function. We selected 2 refined MAGs that had at least 50% completeness and no more than 10% redundancy. MAGs scaffolds were uploaded to Microbial Genomes Atlas Online (MiGA) to access genome assembly completeness and to identify the closest bacterial strain to each putative novel genome (RODRIGUEZ-R et al., 2018). The pangenomic analysis of MAG BZ-01 was carried out using Anvi'o v6.1's pangenomic workflow, plus 49 complete *Leuconostoc* genomes downloaded from NCBI using the script 'ncbi-genome-download' with parameters: -s refseq -l complete -R all and -g "Leuconostoc". (https://github.com/kblin/ncbi-genome-download). Briefly, we create an anvi'o genome database, computing (with flag: --use-ncbi-blast; and parameters: --minbit 0.5 --mcl-inflation 8) and displaying pangenome. The pangenomic analysis of

the five selected *Leuconostoc* genomes and BZ1-01 also used the pangenomic workflow with parameter --mcl-inflation 10.

3.2.4 Phages and prophages detection

Phages contigs present in the microbial metagenome were assessed using VirSorter v1.0.6 (ROUX et al., 2015) with its standard built-in database of viral sequences, and parameter --db 1. We selected contigs classified only in categories 1 and 2 (Phages), and 4 and 5 (Prophages) of VirSorter output (n = 21). Open reading frames (ORF) were predicted using Prodigal v2.6.3 (HYATT et al., 2010) in metagenomic mode. ORFs were annotated using a Hidden Markov Model (HMM) algorithm (hmmscan from HMMER v3.3) against HMM database of prokaryotic viral orthologous groups (pVOG) (GRAZZIOTIN; KOONIN; KRISTENSEN, 2017) considering the significant hit e-value threshold of 10^{-5}. Ribosomal proteins were searched using Barnnap 0.9 (https://github.com/tseemann/barrnap) with an e-value threshold of 10^{-6}. Contigs were aligned against the viral section of NCBI RefSeq database using BLASTn (ALTSCHUL et al., 1990) of BLAST+ package (CAMACHO et al., 2009) with following parameters: e-value < 10^{-10}, covering > 90% of contig length and > 50% identity. The taxonomic assignment of contigs, completeness, contamination and quality of contigs were assessed using Demovir script (https://github.com/feargalr/Demovir) with default parameters, CheckV (NAYFACH et al., 2020) and vConTACT v2.0 (BIN JANG et al., 2019) clustering pipeline, a network-based analytical tool that uses whole genome gene-sharing profiles and distance-based hierarchical clustering to group viral contigs into virus clusters (VCs). Besides our viral contigs, we also included in the pool known viral genomes (NCBI RefSeq database release 88). We selected 3 contigs classified as High-quality and Medium-quality genomes, annotated them using the multiPhATE2 pipeline (ECALE ZHOU et al., 2021), with ORF prediction by Prodigal and gene annotation with pVOG database. Lysogenic phages present in *Leuconostoc* reference genomes and MAG BZ1-01 were detected using the PHASTER web server (ARNDT et al., 2016). Only prophages sequences classified as “intact” were selected for further analysis. Comparisons between viral contigs and prophages detected were made using BLASTn hit table.

3.2.5 Phylogenomic tree of *Leuconostoc* lysogenic phages

All intact profages identified in the *Leuconostoc* genomes were used to construct a phylogenomic tree using VICTOR - Virus Classification and Tree Building Online Resource (MEIER-KOLTHOFF; GÖKER, 2017). Briefly, pairwise comparisons of the nucleotide sequences were made using the Genome-BLAST Distance Phylogeny (GBDP) method (MEIER-KOLTHOFF et al., 2013) using settings recommended for prokaryotic viruses. The intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME (LEFORT; DESPER; GASCUEL, 2015) and, branch support was inferred from 100 pseudo-bootstrap replicates each and visualized with FigTree 1.4.4 (RAMBAUT, 2012).

3.2.6 Visualizations

Genome diagrams and comparison figures were prepared using the genoPlotR package (GUY et al., 2011). All analyses were carried out using the statistical software R (R DEVELOPMENT CORE TEAM, 2020), pheatmap (KOLDE, 2019) and Qiimer packages.

3.3 RESULTS

3.3.1 Taxonomic and functional analysis of Boza metagenome

We obtained a total of 1,472.974 reads for the BZ1 sample. The most abundant bacterial genera identified by MetaPhlAn in the sample were *Leuconostoc* (86.2%), *Lactococcus* (10%), *Staphylococcus* (3.2%), and *Acinetobacter* (0.3%). At species level, we identified *Leuconostoc lactis* (82.3%), *Lactococcus lactis* (8.9%), *Leuconostoc citreum* (3.5%), *Staphylococcus aureus* (3.2%), *Lactococcus raffinolactis* (1.1%), and *Leuconostoc pseudomesenteroides* (0.3%).

The functional analysis was performed using HUMAnN 2.0, and the metabolic potential reconstructed using KEGG Orthogroups (KOs) (Appendix 3, Fig. S1), for the main genera identified (*Lactococcus*, *Leuconostoc*, *Lactobacillus*, *Staphylococcus*, and *Acinetobacter*). The overall metabolism detected in Boza is shown in Appendix 3, Fig. S2. Noteworth metabolism included the carotenoid biosynthesis pathway in

Leuconostoc, fatty acid elongation in Acinetobacter, and (lipid metabolism) and several vitamin biosynthesis pathways in *Staphylococcus*.

3.3.2 Metagenome assembled genomes (MAGs)

As over 90% of all reads obtained were classified in two single bacterial genera, we proceeded with assembly based analytical methods. Two metagenome-assembled genomes (MAG) were recovered from the BZ1 sample and identified as BZ1-01 and BZ1-02 (Appendix 2, Table S1). Reads that were mapped to contigs belonging to each MAG were re-assembled separately to increase genome quality. MiGA essential genes analysis showed that BZ1-01 genome had 99.1% completeness, 1.9% contamination, and an excellent quality of 89.6%, being classified as *Leuconostoc garlicum* KFRI01 (NZ_CP016329) (ANI: 98.21% and p-value: 0.051). The BZ1-01 genome size is 1,742,767 bp, with average GC content of 43.58%. We found one complete plasmid associated with MAG BZ1-01 with 19,095-bp and similar to pWK40-1 of *Leuconostoc lactis* strain WiKim40. The BZ1-02 reassembled MAG has completeness of 88.7 %, contamination of 4.7%, a high quality of 65.2%, and was classified as *Lactococcus lactis* subsp. *cremoris* IBB477 (NZ_CM007353) (ANI: 96.56% and p-value 0.363), according to the MIGA pipeline analysis. The BZ1-02 draft genome size is 2,037,264 bp, with an average GC content of 36.06%. We used PHASTER to identify prophages in the two MAGs, identifying four phage signatures in the MAG BZ1-01, one of which with the completeness score of 130 and length 23,816 bp. This phage was classified within the Siphoviridae family using Demovir pipeline and *Streptococcus* phages by PHASTER gene annotation. MAG BZ1-02 showed four incomplete phages classified as *Lactococcus* phages.

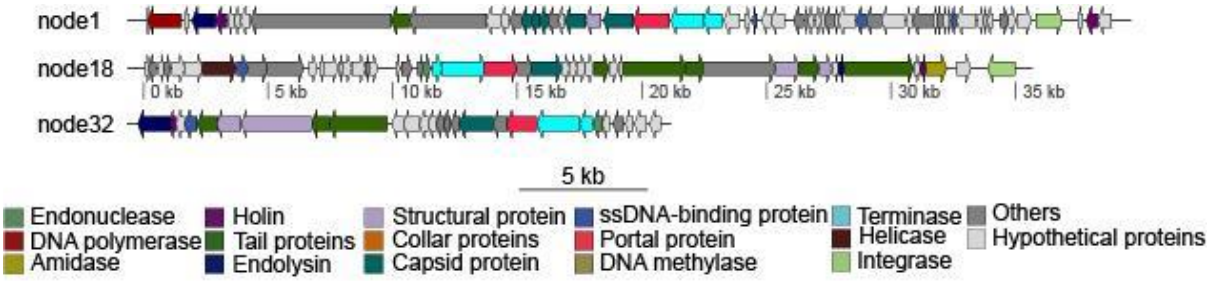
3.3.3 Bacteriophages in Boza

We also investigated the presence of phage-containing contigs in the dataset using VirSorter 1.0.6. A total of 21 viral contigs were identified in the BZ1 sample: 2 high-quality, 1 medium-quality, and 19 low-quality genomes as assessed by CheckV. The one putative phage genome was classified as family *Myoviridae*, and the remaining 20 as family *Siphoviridae*. We also analysed these novel viral contigs using the vConTACT v2.0 clustering pipeline. We included in the analysis the 21 boza phage

contigs and all known viral genomes from NCBI RefSeq. Only five contigs formed viral clusters with RefSeq viral genomes belonging to *Bacillus* and *Lactococcus* phages. The taxonomic classification for these new phages was further refined by using BLAST alignments against RefSeq's complete viral genomes database. We identified *Lactococcus* (16), *Leuconostoc* (2) and *Staphylococcus* (1) phages. Identified species included *Lactococcus* phages bIL285, bIL310, bIL311, bIL312, and ul36; *Leuconostoc* phages phiLN25 and P793; and *Staphylococcus* phage phiNM3.

We selected the 3 contigs classified as High- and Medium-quality genomes and annotated them using the multiPhATE2 pipeline, with ORF prediction by Prodigal and gene annotation with pVOG database (Fig. 3.1). The three genomes were classified by pVOG as *Siphoviridae* and their genomes sizes ranged from 21,433 to 38,712 bp. The genes most frequently annotated in our genomes were tail protein (7), capsid protein (7), terminase (6), endolysin (2), and integrase (2). The presence of integrase genes in Node1 and Node18 is evidence of a lysogenic life cycle. We compared the Average Nucleotide Identity (ANI) of the intact prophage detected in MAG BZ1-01 against the phage contigs recovered from the metagenome and identified an ANI of 100% with Node1. Integrase or site-specific recombinase genes were detected in only six of the remaining 21 contigs, and taken together these results indicate the presence of both lytic and lysogenic phage in Boza.

Figure 3.1. Genome annotation of phage genomes recovered from boza metagenome.



Genome annotation of high- and medium-quality bacteriophages genomes using pVOG database;Colors indicate the gene identification.

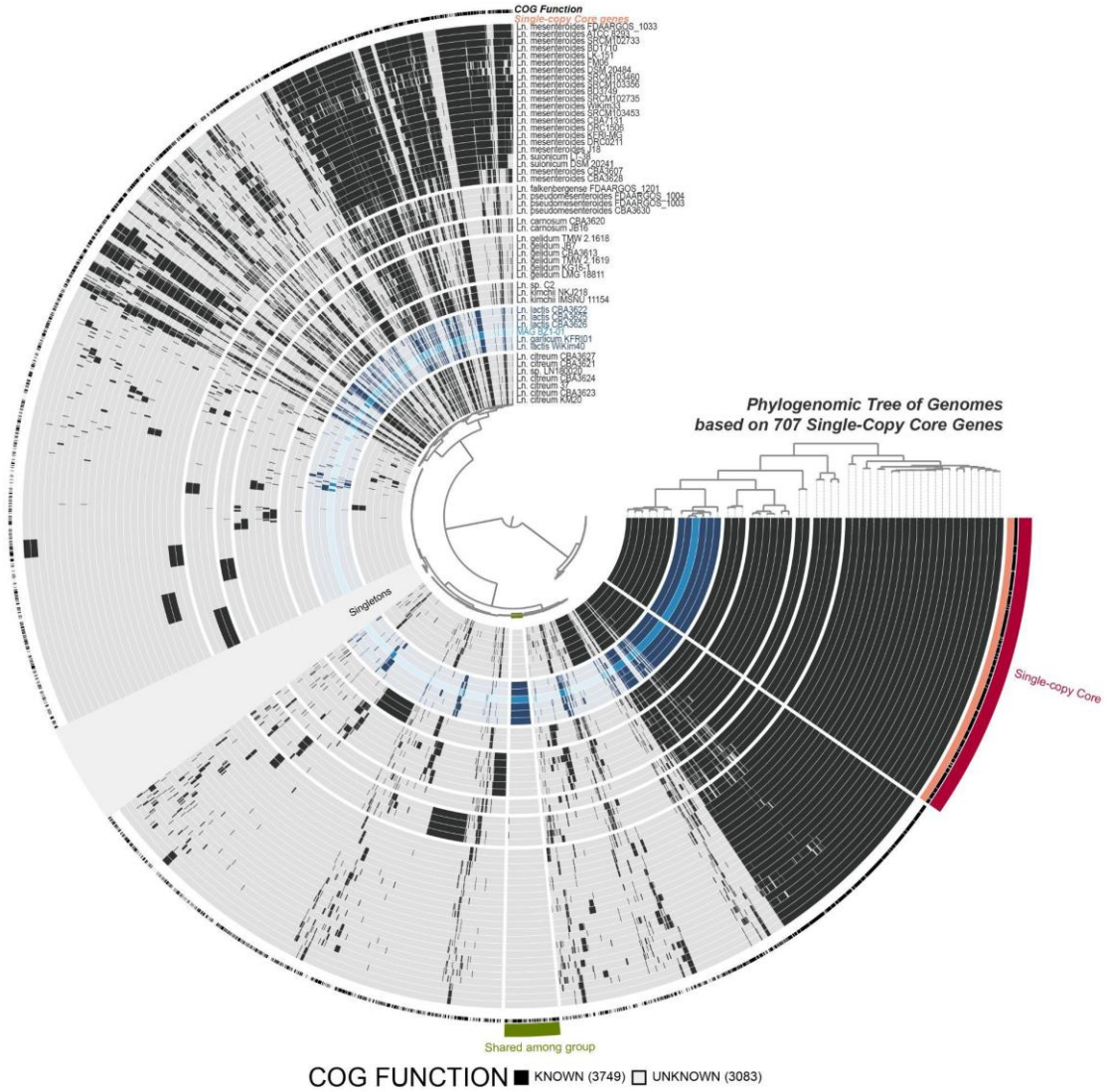
3.3.4 *Leuconostoc* pangenome

A pangenomic analysis was carried out using 49 complete genomes of *Leuconostoc* and the BZ1-01 genome analysis using Anvio 6.1. The *Leuconostoc* pangenome (Fig. 3.2) showed a total of 6,832 gene clusters. We identified 1,584 singleton gene

clusters, unique to only one genome, and 707 gene clusters were identified as single-copy core genes for the *Leuconostoc* genus pangenome. We used the 707 single-copy core genes to construct a phylogenomic tree using a maximum likelihood model and detected 7 distinguishable genome groups, all with ANI >90% (Fig. 3.3). The groups are also homogeneous as their classification, indicating that for the *Leuconostoc* genus, there currently has 7 distinct species, which are separated by minimum within taxon ANI of 90%. For simplicity, we'll refer to these taxonomic groups as species through the manuscript. The novel *Leuconostoc* MAG isolated from Boza is firmly placed within the *L. lactis* species, which also includes the *L. garlicum* genome, indicating that this is likely a strain of *L. lactis*.

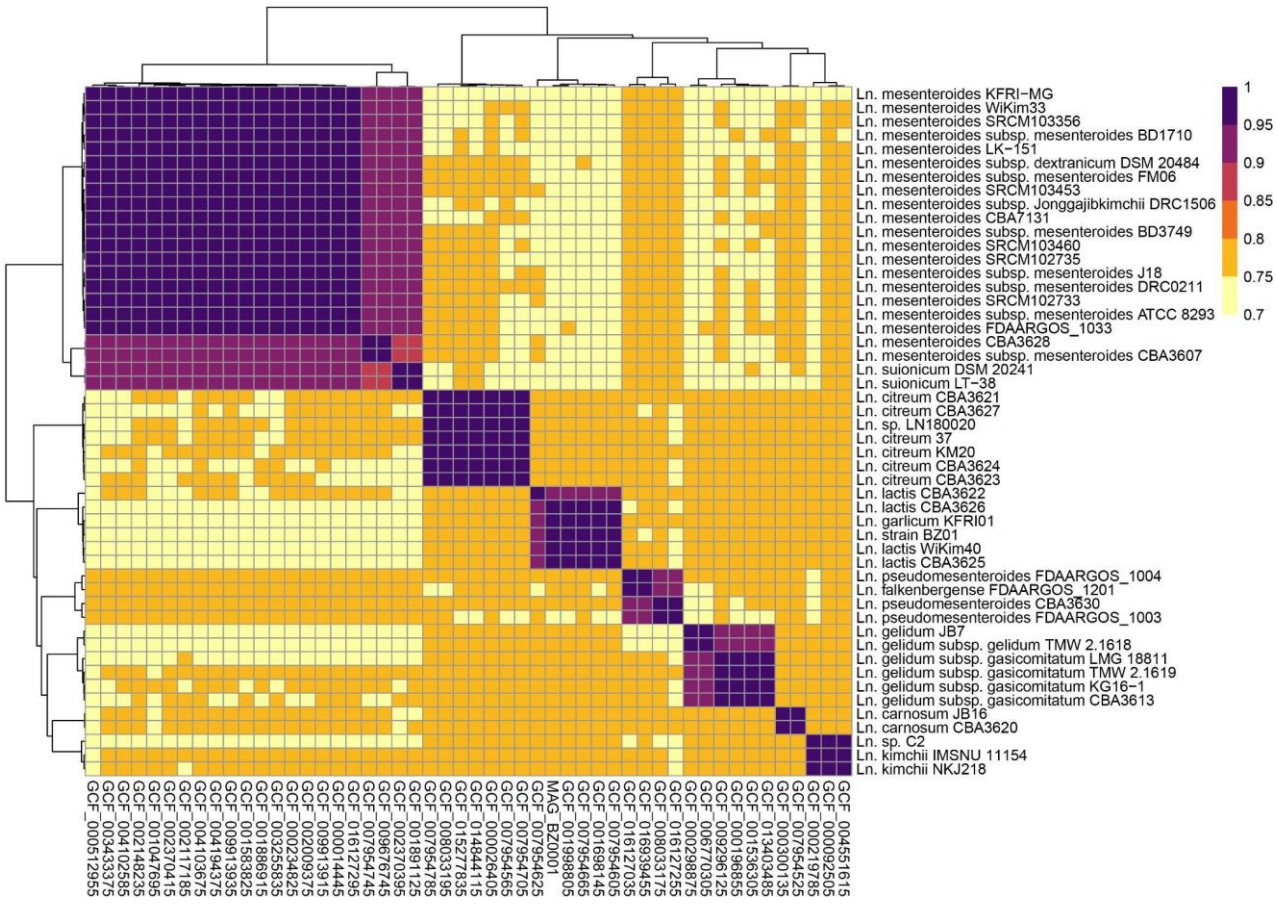
Genes shared between members of this group are associated with Carbohydrate metabolism, such as Galactose mutarotase (COG2017) and Exopolysaccharide biosynthesis protein EpsI (COG5039); with Cell wall, membrane, and envelope biogenesis, such as Sortase (COG3764) and LysM repeat (COG1388); with Defense mechanisms, such as several ABC-type transport systems (COG1132; COG0488) and Beta-lactamase (COG2367); and with Inorganic ion transport and metabolism, such as ABC-type Fe3+-siderophore transport system (COG0609, COG0614).

Figure 3.2. Pangenomic analysis of *Leuconostoc* genus.



The pangenome was performed with 49 complete genomes selected from GeneBank and metagenome-assembled genome BZ1-01 (highlighted in light blue).

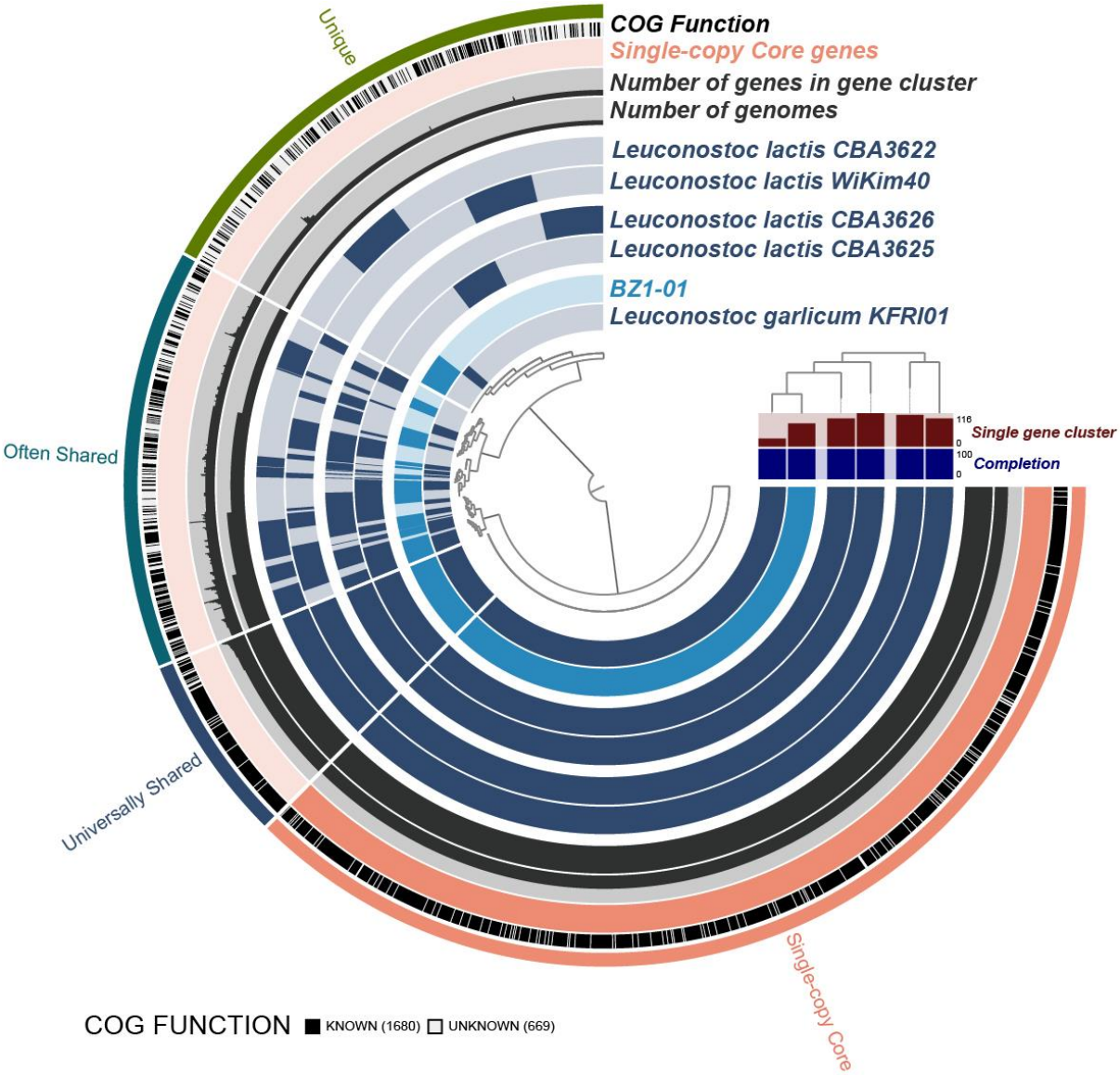
Figure 3.3. Heatmap of Average Nucleotide Identity (ANI) values among genomes used in pangenome.



ANI among 49 genomes selected from GeneBank and metagenome-assembled genome BZ1-01 (identified as Ln. lactis BZ01 at y-axis).

A within *L. lactis* species analysis was completed to better understand the relationship between those five genomes and BZ1-01 (Fig. 3.4). Genes from these 6 genomes were grouped into four bins based on frequency of occurrence: single-copy core (n = 1,172), universally shared (200), often shared (444), and unique (532). MAG BZ1-01 has 81 of the 532 unique gene clusters, some of them were classified as Sortase (COG3764), Transposase (COG3464), post-translational modification (COG2039; COG0071), and phage-related genes (COG4824; COG3728).

Figure 3.4. Pangenomic analysis of *Leuconostoc lactis* group.



The pangenome was performed with genomes clustered into *Leuconostoc lactis* group from pangenome *Leuconostoc* genus, being composed by four *Ln. lactis* genomes, one *Ln. garlicum* and the MAG BZ1-01 (highlighted in light blue).

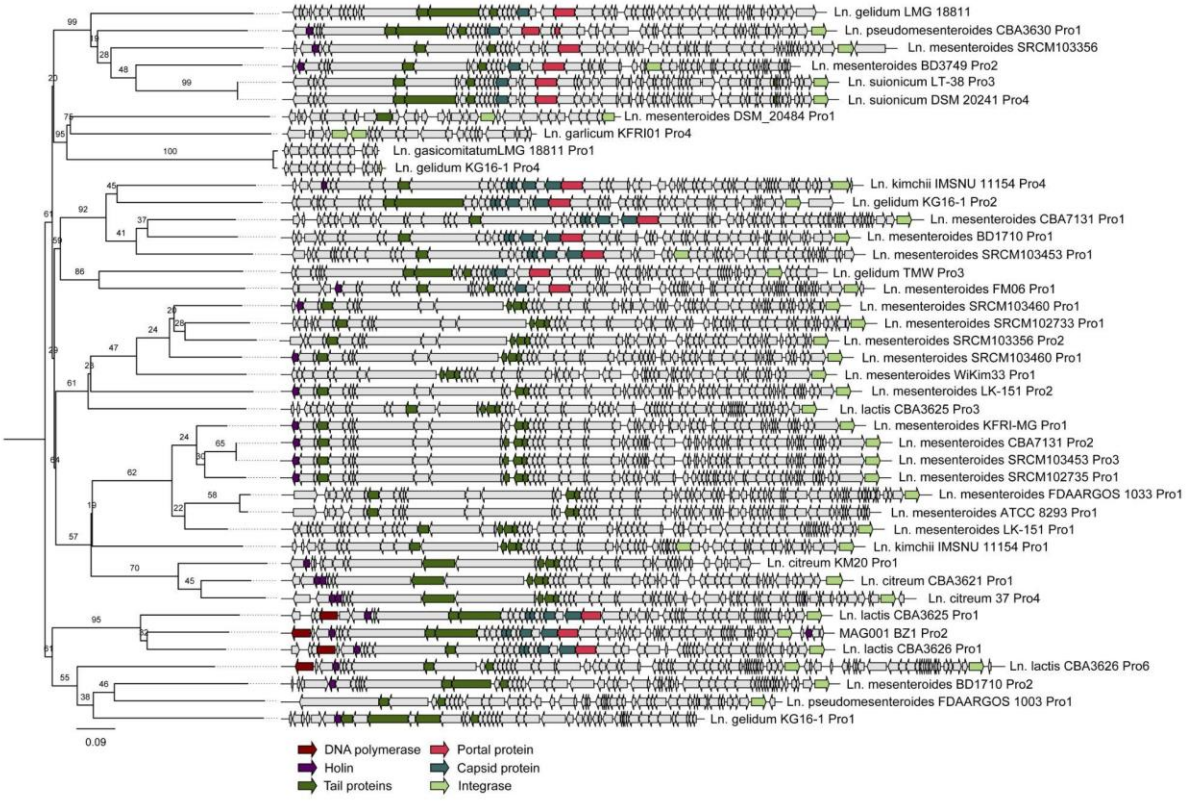
3.3.5 Phylogenomic and comparative analysis of *Leuconostoc* prophages

We next decided to investigate the diversity and phylogenetic relationship of *Leuconostoc* prophages from 49 complete genomes and BZ1-01 MAG. Prophages were identified in *Leuconostoc* genomes using PHASTER software, and we selected for further investigation only prophages classified as intact. We obtained a total of 42 lysogenic phages sequences belonging to 31 bacterial genomes, with a length range from 6847 to 51013 bp, and annotated genomic features using the MultiPHATE2

pipeline. Finally, a phylogenomic tree of *Leuconostoc* lysogenic phages was constructed using VICTOR (Fig. 3.5), a pipeline for genome-based phylogeny and classification of prokaryotic viruses.

Our search reveals a high number of phages integrated into *Leuconostoc* genomes. It is interesting to observe that the tree branches were composed by lysogenic phages infecting different genomes from the same group or closely related and by lysogenic phages infecting different genomes of three or four groups (Fig. 3.5). For example, two branches were composed only by phages integrated in *Ln. mesenteroides* genomes; other was composed of seven phages, six of them integrated at *Ln. mesenteroides* genomes and only one at *Ln. lactis*. On the other hand, there is a branch composed of six phages, integrated to different *Leuconostoc* species, such as *Ln. pseudomesenteroides*, *Ln. mesenteroides*, *Ln. gelidum*, and *Ln. suionicum*. The phage integrated to BZ1-01 formed a branch with two lysogenic phages found in *Ln. lactis*.

Figure 3.5. Phylogenomic tree and genome annotation of *Leuconostoc* lysogenic phages.



Phylogenomic tree of 42 lysogenic phages found in *Leuconostoc* genomes. Arrows represent the genes and the color their annotations.

3.4 DISCUSSION

Our study is the first characterization of the boza microbiome using microbial metagenome approach. We highlight a high dominance of *Leuconostoc* species, however, with a functional partition among bacterial genera contributing to different metabolic pathways. The recovery of metagenomic assembled genomes (MAGs) allows us to observe the presence of prophage sequences in BZ1-01, a MAG classified as *Leuconostoc garlicum*, placed within the *L. lactis* clade upon a pangenomic analysis, with potential to be a putative novel strain.. Phages and prophages have a great influence in lactic acid bacteria (LAB) communities in fermentative foods (SAMSON; MOINEAU, 2013). Thus, to better understand the dynamics of bacteria and phages in boza samples, we recovered 21 phage sequences, identified one of them being the prophage present in BZ1-01 and performed a comparative analysis of 42 prophages present in *Leuconostoc* genomes.

To our knowledge, this is the first characterization of the Boza bacteriophage community. Despite these results being based on one sample, they may direct new efforts to study viromes in Boza and other fermented beverages (LEDORMAND; DESMASURES; DALMASSO, 2020). Here, we observe the presence of phages belonging to *Lactococcus*, *Leuconostoc, Staphylococcus* phage, and *Bacillus* virus groups. The phage composition in Boza showed similar groups when compared with other fermented foods, such as cheese (PUJATO; QUIBERONI; MERCANTI, 2019) and kimchi (JUNG et al., 2011, 2018). The presence of several known lysogenic phages, such as *Lactococcus* phage bIL310, bIL311, bIL312, bIL285, and BK5-T (CHOPIN et al., 2001) and the detection of integrase ORFs is an indicative of interactions between phage and bacteria in Boza, with implications in the microbial dynamics, flavor and aroma in boza and fermentative foods in general. Nevertheless, this could also be a limitation of microbial metagenome methods, allowing us to detect more adsorbed and integrated than free lytic phages (ROUX et al., 2019).

Using a metagenomic approach, we were able to corroborate the previous taxonomic characterization of the LAB community present in boza, based on culture dependent and independent (16S rRNA) analysis, such as *Lactobacillus*, *Lactococcus*,

Leuconostoc, *Enterococcus*, and *Pediococcus* genera, with predominance of *Leuconostoc* (ARICI; DAGLIOGLU, 2002; HANCIOĞLU; KARAPINAR, 1997; OSIMANI et al., 2015; TODOROV, 2010). Here, we reported two high quality MAGs, BZ1-01 and BZ1-02, classified as *Leuconostoc garlicum* KFRI01 (ANI: 98.21%) and *Lactococcus lactis* subsp*. cremoris* IBB477 (ANI: 96.56%), respectively. Several isolates of *Leuconostoc* and *Lactococcus* species were obtained from boza samples (BALCIUNAS et al., 2013) and have their bacteriocins characterized (TODOROV, 2010), but none of complete genomes were sequenced. Using MAGs, we are able to characterize their genomes and the presence of prophages.

The *Leuconostoc* pangenome analysis allows us to understand the difference among genomes, identifying similarities with BZ1-01 and their principal characteristics. We observe the clustering of seven distinct groups of *Leuconostoc* genomes with 90% ANI: a group composed, in their majority, by *Ln. mesenteroides* genomes and two genomes of *Ln. suionicum*; followed by *Ln. citreum* group, *Ln. lactis* group, other formed by three *Ln. pseudomesenteroides* genomes and one *Ln. falkenbergense* genome, and three other groups formed by *Ln. gelidum*, *Ln. carnosium*, and *Ln. kimchi*. Our MAG BZ1-01 were placed within *Ln. lactis* group that also contains a genome of *Ln. garlicum*, mainly associated with kimchi, a Korean vegetable fermented food. These groups were used to compare the lysogenic phages (as prophages or provirus) in *Leuconostoc* genomes.

Lysogeny is commonly observed in LAB strains (MARCÓ; MOINEAU; QUIBERONI, 2012). However, the less studied *Leuconostoc* genera showed few described lysogenic phages (KOT et al., 2014). In our study, we detected lysogenic phages sequences from viral genomes recovered using marker genes, one of these genomes was the same integrated in MAG BZ1-01, being classified using BLASTn as *Leuconostoc* phage P793. Despide the few species isolated and described, the presence of prophages in *Leuconostoc* species was frequently observed, such as in *Ln. mesenteroides* ATCC 8293, *Ln. citreum* KM20, and *Ln. kimchi* IMSNU 11154 (KOT et al., 2014). Recently, CANDELIERE et al., (2021) detected several lysogenic phage sequences in 17 genomes of *Ln. carnosum*. The knowledge about *Leuconostoc* lysogenic phage diversity and their genomes' composition is of great

importance to traditional and industrial fermented food production systems, in addition to their intrinsic importance as starter in boza (ALTAY et al., 2013) and as adjutant LAB in cheese (MAHONY et al., 2016), for example. Based on the phylogenomic and comparative analysis of *Leuconostoc* lysogenic phages, we observed that there is no relationship between phages and their hosts, with similar phages integrated in genomes of the same or different bacterial groups. Genome organization and gene annotation showed similar patterns for each phage tree branch formed; all integrated phages genomes were classified belonging to the *Siphoviridae* family, and based on viral cluster and phylogenomic results the phages grouped at the same branch belong to the same genera.

3.5 CONCLUSION

Our study showed the relationships between phages and bacteria present in boza samples, focused in *Leuconostoc* spp and their lysogenic phages. The recovery of a high-quality MAG of *Leuconostoc lactis/garlicun*, their pangenome analysis and a deep phylogenomic analysis of *Leuconostoc* lysogenic phages may contribute to understanding the importance of those organisms to boza fermentative environment and other similar fermented foods that use natural starter cultures.

3.6 REFERENCES

ALNEBERG, J. et al. Binning metagenomic contigs by coverage and composition. **Nature Methods**, v. 11, n. 11, p. 1144–1146, 2014.

ALTAY, F. et al. A review on traditional Turkish fermented non-alcoholic beverages: Microbiota, fermentation process and quality characteristics. **International Journal of Food Microbiology**, v. 167, n. 1, p. 44–56, 2013.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–10, 1990.

ANTIPOV, D. et al. Plasmid detection and assembly in genomic and metagenomic data sets. **Genome Research**, v. 29, n. 6, p. 961–968, 2019.

ANTIPOV, D. et al. Metaviral SPAdes: Assembly of viruses from metagenomic data. **Bioinformatics**, v. 36, n. 14, p. 4126–4129, 2020.

ARICI, M.; DAGLIOGLU, O. Boza: A lactic acid fermented cereal beverage as a traditional Turkish food. **Food Reviews International**, v. 18, n. 1, p. 39–48, 2002.

ARNDT, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. **Nucleic acids research**, v. 44, n. W1, p. W16–W21, 2016.

BALCIUNAS, E. M. et al. Novel biotechnological applications of bacteriocins: A review. **Food Control**, v. 32, n. 1, p. 134–142, 2013.

BIN JANG, H. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. **Nature Biotechnology**, v. 37, n. 6, p. 632–639, 2019.

CAMACHO, C. et al. BLAST+: Architecture and applications. **BMC Bioinformatics**, v. 10, p. 1–9, 2009.

CANDELIERE, F. et al. Comparative Genomics of Leuconostoc carnosum. **Frontiers in Microbiology**, v. 11, n. January, p. 1–13, 2021.

CHIKINDAS, M. L. et al. Functions and emerging applications of bacteriocins. **Current Opinion in Biotechnology**, v. 49, p. 23–28, 1 fev. 2018.

CHOPIN, A. et al. Analysis of six prophages in Lactococcus lactis IL1403: Different genetic structure of temperate and virulent phage populations. **Nucleic Acids Research**, v. 29, n. 3, p. 644–651, 2001.

CORREA, A. M. S. et al. Revisiting the rules of life for viruses of microorganisms. **Nature Reviews Microbiology**, v. 19, n. August, p. 501–513, 2021.

ECALE ZHOU, C. L. et al. MultiPhATE2: code for functional annotation and comparison of phage genomes. **G3 (Bethesda, Md.)**, v. 11, n. 5, p. 1–5, 2021.

EREN, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. **PeerJ**, v. 3, p. e1319, 2015.

FEINER, R. et al. A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. **Nature Reviews Microbiology**, v. 13, n. 10, p. 641–650, 2015.

FINN, R. D.; CLEMENTS, J.; EDDY, S. R. HMMER web server: interactive sequence similarity searching. **Nucleic Acids Research**, v. 39, n. suppl, p. W29–W37, 1 jul. 2011.

FRANZOSA, E. A. et al. Species-level functional profiling of metagenomes and metatranscriptomes. **Nature Methods**, v. 15, n. 11, p. 962–968, 2018.

GRAZZIOTIN, A. L.; KOONIN, E. V.; KRISTENSEN, D. M. Prokaryotic Virus Orthologous Groups (pVOGs): A resource for comparative genomics and protein family annotation. **Nucleic Acids Research**, v. 45, n. D1, p. D491–D498, 2017.

GUY, L. et al. GenoPlotR: comparative gene and genome visualization in R. **Bioinformatics**, v. 27, n. 13, p. 2334–2335, 2011.

HANCIOĞLU, Ö.; KARAPINAR, M. Microflora of Boza, a traditional fermented Turkish beverage. **International Journal of Food Microbiology**, v. 35, n. 3, p. 271–274, 1997.

HAYES, S. et al. Metagenomic approaches to assess bacteriophages in various environmental niches. **Viruses**, v. 9, n. 6, p. 1–22, 2017.

HYATT, D. et al. Prodigal: Prokaryotic gene recognition and translation initiation site identification. **BMC Bioinformatics**, v. 11, 2010.

JUNG, J. Y. et al. Metagenomic analysis of kimchi, a Traditional Korean fermented food. **Applied and Environmental Microbiology**, v. 77, n. 7, p. 2264–2274, 2011.

JUNG, M. J. et al. Viral community predicts the geographical origin of fermented vegetable foods more precisely than bacterial community. **Food Microbiology**, v. 76, n. June, p. 319–327, 2018.

KIM, D. et al. Centrifuge: Rapid and sensitive classification of metagenomic sequences. **Genome Research**, v. 26, n. 12, p. 1721–1729, 2016.

KOLDE, R. pheatmap. p. 1–8, 2019.

KOT, W. et al. Bacteriophages of Leuconostoc, Oenococcus, and Weissella. **Frontiers in Microbiology**, v. 5, n. APR, p. 1–9, 2014.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 2012.

LEBLANC, J. G.; TODOROV, S. D. Bacteriocins producing lactic acid bacteria isolated from Boza, a traditional fermented beverage from Balkan Peninsula – from isolation to application. In: MÉNDEZ-VILAS, A. (Ed.). . **Science against microbial pathogens: communicating current research and technological advances**. Badajos, Spain: Formatex Research Center, 2011. p. 1311–1320.

LEDORMAND, P.; DESMASURES, N.; DALMASSO, M. Phage community involvement in fermented beverages: an open door to technological advances? **Critical Reviews in Food Science and Nutrition**, v. 0, n. 0, p. 1–10, 2020.

LEFORT, V.; DESPER, R.; GASCUEL, O. FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. **Molecular Biology and Evolution**, v. 32, n. 10, p. 2798–2800, 1 out. 2015.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078–2079, 15 ago. 2009.

MAHONY, J. et al. Phage-Host Interactions of Cheese-Making Lactic Acid Bacteria. **Annual Review of Food Science and Technology**, v. 7, n. 1, p. 267–285, 2016.

MAHONY, J. et al. Phage biodiversity in artisanal cheese wheys reflects the complexity of the fermentation process. **Viruses**, v. 9, n. 3, p. 1–18, 2017.

MARCÓ, M. B.; MOINEAU, S.; QUIBERONI, A. Bacteriophages and dairy fermentations. **Bacteriophage**, v. 2, n. 3, p. 149–158, 2012.

MEIER-KOLTHOFF, J. P. et al. Genome sequence-based species delimitation with confidence intervals and improved distance functions. **BMC Bioinformatics 2013 14:1**, v. 14, n. 1, p. 1–14, 21 fev. 2013.

MEIER-KOLTHOFF, J. P.; GÖKER, M. VICTOR: genome-based phylogeny and classification of prokaryotic viruses. **Bioinformatics**, v. 33, n. 21, p. 3396–3404, 1 nov. 2017.

MURPHY, J. et al. Biodiversity of lactococcal bacteriophages isolated from 3 Gouda-type cheese-producing plants. **Journal of dairy science**, v. 96, n. 8, p. 4945–57, 2013.

NAYFACH, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. **Nature Biotechnology**, 2020.

NURK, S. et al. MetaSPAdes: A new versatile metagenomic assembler. **Genome Research**, v. 27, n. 5, p. 824–834, 2017.

OSIMANI, A. et al. Unpasteurised commercial boza as a source of microbial diversity. **International Journal of Food Microbiology**, v. 194, p. 62–70, 2015.

PUJATO, S. A.; QUIBERONI, A.; MERCANTI, D. J. Bacteriophages on dairy foods. **Journal of Applied Microbiology**, v. 126, n. 1, p. 14–30, 2019.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing.**Vienna, Austria.R Foundation for Statistical Computing, , 2020. Disponível em: <http://www.r-project.org/.>

RODRIGUEZ-R, L. M. et al. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. **Nucleic Acids Research**, v. 46, n. W1, p. W282–W288, 2 jul. 2018.

ROUX, S. et al. VirSorter: Mining viral signal from microbial genomic data. **PeerJ**, v. 2015, n. 5, p. 1–20, 2015.

ROUX, S. et al. Minimum information about an uncultivated virus genome (MIUVIG). **Nature Biotechnology**, v. 37, n. 1, p. 29–37, 2019.

SAMSON, J. E.; MOINEAU, S. Bacteriophages in Food Fermentations: New Frontiers in a Continuous Arms Race. **http://dx.doi.org/10.1146/annurev-food-030212-182541**, v. 4, n. 1, p. 347–368, 6 mar. 2013.

TATUSOV, R. L. et al. **The COG database: new developments in phylogenetic classification of proteins from complete genomesNucleic Acids Research**. [s.l: s.n.]. Disponível em: <http://www.ncbi.nlm.nih.gov/COG>. Acesso em: 4 maio. 2020.

TODOROV, S. D. Diversity of bacteriocinogenic lactic acid bacteria isolated from boza, a cereal-based fermented beverage from Bulgaria. **Food Control**, v. 21, n. 7, p. 1011–1021, 2010.

TODOROV, S. D.; CHIKINDAS, M. L. Lactic Acid Bacteria Bacteriocins and their Impact on Human Health. In: ALBUQUERQUE, M. A. C. et al. (Eds.). . **Lactic Acid Bacteria**. Boca Raton : CRC Press, Taylor & Francis Group, [2020]: CRC Press, 2020. p. 79–92.

TODOROV, S. D.; DE MELO FRANCO, B. D. G.; TAGG, J. R. Bacteriocins of Gram-positive bacteria having activity spectra extending beyond closely-related species. **Beneficial Microbes**, v. 10, n. 3, p. 315–328, 19 abr. 2019.

TODOROV, S. D.; HOLZAPFEL, W. H. Traditional cereal fermented foods as sources of functional microorganisms. In: **Advances in Fermented Foods and Beverages**. [s.l.] Elsevier, 2015. p. 123–153.

TRUONG, D. T. et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. **Nature Methods**, v. 12, n. 10, p. 902–903, 2015.

WÜNSCHE, L. Importance of bacteriophages in fermentation processes. **Acta Biotechnologica**, v. 9, n. 5, p. 395–419, 1989.

CONCLUSIONS

The main aim of my thesis was to characterize the virome and their interactions at spatial and temporal levels, with the microbiome from fermented foods. For this, I used two types of fermented foods, Canastra cheese, an artisanal Brazilian cheese, and Boza, a traditional Balkans's beverage made with grains.

In Chapter 1, the interactions between bacteriophages and bacteria were obtained from Canastra cheese production samples, seven samples from endogenous starter culture and six from 22-days ripened cheeses. To analyse these samples, I used two approaches, viral metagenome (viral enrichment samples) and microbial metagenome (total DNA) sequencing. The study showed a high diversity of bacteriophages in starter culture and cheese samples, with high levels of phage-bacteria interactions in the Canastra cheese production system. Several virus-like particles (VLPs) were isolated from the endogenous starter cultures, followed by an extensive description of the principal phage members of these microbial ecosystems. The microbiome parcel from these ecosystems was assessed recovering metagenome-assembled genomes (MAGs) from microbial metagenomes of starter culture and cheeses. I identified a likely novel *Streptococcus* phage 987 group species, as well as a putative host, a novel strain of *S. salivarius*. Moreover, were also observed a dynamic yet stable microbial ecosystem during the cheese production, marked by genomic evidence of continued phage-bacteria interactions, such as the presence of bacterial defence mechanisms: Restriction modification (RM), Abortive Infection (Abi), and CRISPR. This was the first effort to describe and understand the viral composition and ecological dynamics within the Canastra Cheese production system, providing a solid background for further mechanistic studies focused on isolating the phages identified here, as well as their bacterial hosts.

In Chapter 2, a novel lytic bacteriophage was isolated from VLP samples from the endogenous starter cultures characterized in Chapter 1. The phage was called vB-SauP_CS44, it is capable of infecting *Staphylococcus aureus*, and belongs to *Rosenblumvirus* genus (*Podoviridae* family). Its genome, morphology, thermo- and pH stability, growth dynamics and host range were characterized. During the growth

dynamics assay I observed the emergence of phage resistance in a wildtype *S. aureus* strain, originally isolated from Canastra cheese. Genomes of phage sensitive and resistant bacteria were sequenced, characterized, and compared. The phage resistant bacteria showed mutations in *tarS*, *fnbA* and *fnbB* genes. The TarS inactivation is related with recognition of host by phages, through receptor binding proteins (RBP) present in the phage structure (head and tail fibers) that interact with host receptor molecules (β-O-GlcNAcylation of the wall teichoic acid). FnbA and FnbB are deemed important for adhence and biofilm formation in bacteria, and despite the observed mutations, I did not foind evidence of a decrease in these capabilities. Moreover, the mutation in *tarS* gene is associated with bacterial susceptibility to β-lactam antibiotics, corroborating the hypothesis that increasing fitness of some traits (phage resistance) may reduce performance in others (susceptibility to β-lactam antibiotics). Although we observe the emergence of phage resistant bacteria, the use of vB-SauP_CS44 in bacteriophage therapy could be carried out in combination with β-lactam antibiotics, which would attack the phage-resistant bacteria, or in combination with other phages that use different binding receptors.

In Chapter 3, the microbiome of Boza was characterized using microbial metagenome sequencing approach, with the recovery of metagenome-assembled genomes (MAGs) and lytic and lysogenic phages. The Boza sample showed a high dominance of *Leuconostoc* species, however, with a functional partition among bacterial genera contributing to different metabolic pathways. The recovery of this MAG, allied with the use of a pangenomic analysis, allowed us to observe the presence of lysogenic phage sequence in BZ1-01, a MAG classified as *Leuconostoc garlicum*, firmly placed within the *Ln. lactis* group, with potential to be a putative novel strain. Twenty one phage sequences were recovered, all of them classified as *Siphorividae* family, and one being the prophage present in MAG BZ1-01. To better understand the lysogeny in *Leucosnotoc* genomes, I performed a comparative analysis of 42 lysogenic phages found in *Leuconostoc* genomes. The genomes of lysogenic phages were annotated and a phylogenomic tree built, evidencing their relationship and distribution in *Leuconostoc* genomes. These analyses may contribute to understanding the importance of phages in boza’s fermentative

environment as well as other similar fermented foods that use natural starter cultures. To our knowledge, this is the first characterization of the Boza bacteriophage community.

The study presented in this Thesis shed light in bacteriophage diversity and dynamics in fermented foods. Particularly, it evidences the importance of phages in endogenous starter cultures used artisanal cheese production, which may be observed in similar producing regions around the world. The isolation of a novel lytic phage and its host interactions, including the description of mechanisms for the emergence of a phage resistant bacterial population, may aid in bacteriophage therapy studies. Finally, the first microbiome and virome characterization of Boza may open new perspectives for microbial studies of fermented beverages.

REFERENCES

ALTAY, F. et al. A review on traditional Turkish fermented non-alcoholic beverages: Microbiota, fermentation process and quality characteristics. **International Journal of Food Microbiology**, v. 167, n. 1, p. 44–56, 2013.

ANDRADE, R. P. et al. Yeasts from Canastra cheese production process: Isolation and evaluation of their potential for cheese whey fermentation. **Food Research International**, v. 91, p. 72–79, 2017.

ARKHIPOVA, K. et al. Temporal dynamics of uncultured viruses: a new dimension in viral diversity. **The ISME Journal**, p. 1–13, 2017.

AUCLAIR, J.; ACCOLAS, J. P. Use of thermophilic lactic starters in the dairy industry. **Antonie van Leeuwenhoek**, v. 49, n. 3, p. 313–326, 1983.

BERNARDEAU, M. et al. Safety assessment of dairy microorganisms: The Lactobacillus genus. **International Journal of Food Microbiology**, v. 126, n. 3, p. 278–285, 1 set. 2008.

BOWERS, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. **Nature Biotechnology**, v. 35, n. 8, p. 725–731, 2017.

BRAZ, V. S.; MELCHIOR, K.; MOREIRA, C. G. Escherichia coli as a Multifaceted Pathogenic and Versatile Bacterium. **Frontiers in Cellular and Infection Microbiology**, v. 10, n. December, p. 1–9, 2020.

BREITBART, M. et al. Phage puppet masters of the marine microbial realm. **Nature Microbiology**, v. 3, n. 7, p. 754–766, 4 jul. 2018.

BRÜSSOW, H. Phages of dairy bacteria. **Annual review of microbiology**, v. 55, p. 283–303, 2001.

CAMPOS, G. Z. et al. Microbiological characteristics of canastra cheese during manufacturing and ripening. **Food Control**, v. 121, n. August 2020, p. 107598, 2021.

CHEVALLEREAU, A. et al. Interactions between bacterial and phage communities in natural environments. **Nature Reviews Microbiology**, v. 0123456789, 9 ago. 2021.

CHOW, C.-E. T.; SUTTLE, C. A. Biogeography of Viruses in the Sea. **Annual Review of Virology**, v. 2, n. 1, p. 41–66, 2015.

CORREA, A. M. S. et al. Revisiting the rules of life for viruses of microorganisms. **Nature Reviews Microbiology**, v. 19, n. August, p. 501–513, 2021.

DĄBROWSKA, K.; ABEDON, S. T. Pharmacologically Aware Phage Therapy: Pharmacodynamic and Pharmacokinetic Obstacles to Phage Antibacterial Action in Animal and Human Bodies. **Microbiology and Molecular Biology Reviews**, v. 83, n. 4, p. 1–25, 30 out. 2019.

DEVEAU, H. et al. Biodiversity and classification of lactococcal phages. **Applied and Environmental Microbiology**, v. 72, n. 6, p. 4338–4346, 2006.

DÍAZ-MUÑOZ, S. L. Viral coinfection is shaped by host ecology and virus–virus interactions across diverse microbial taxa and environments. **Virus Evolution**, v. 3, n. 1, 1 jan. 2017.

DION, M. B.; OECHSLIN, F.; MOINEAU, S. Phage diversity, genomics and phylogeny. **Nature Reviews Microbiology**, v. 18, n. March, 2020.

DORON, S. et al. Systematic discovery of antiphage defense systems in the microbial pangenome. **Science**, v. 359, n. 6379, p. 0–12, 2018.

DRAGOŠ, A. et al. Phages carry interbacterial weapons encoded by biosynthetic gene clusters. **Current Biology**, p. 1–11, 2021.

DUC, H. M. et al. Isolation and application of bacteriophages alone or in combination with nisin against planktonic and biofilm cells of Staphylococcus aureus. **Applied Microbiology and Biotechnology**, 2020.

DUGAT-BONY, E. et al. Viral metagenomic analysis of the cheese surface: A comparative study of rapid procedures for extracting viral particles. **Food Microbiology**, v. 85, n. July 2019, p. 103278, 2020.

FEINER, R. et al. A new perspective on lysogeny: Prophages as active regulatory switches of bacteria. **Nature Reviews Microbiology**, v. 13, n. 10, p. 641–650, 2015.

GOLDFARB, T. et al. BREX is a novel phage resistance system widespread in microbial genomes . **The EMBO Journal**, v. 34, n. 2, p. 169–183, 2015.

GONÇALVES DE MELO, A.; LEVESQUE, S.; MOINEAU, S. Phages as friends and enemies in food processing. **Current Opinion in Biotechnology**, v. 49, p. 185–190, 2018.

GORDILLO ALTAMIRANO, F. L.; BARR, J. J. Phage Therapy in the Postantibiotic Era. **Clinical microbiology reviews**, v. 32, n. 2, p. 1–25, 2019.

HAMMES, W. P.; VOGEL, R. F. The genus Lactobacillus. In: WOOD, B. J. B.; HOLZAPFEL, W. H. (Eds.). . **The Genera of Lactic Acid Bacteria**. Boston, MA: Springer US, 1995. p. 19–54.

HAMPTON, H. G.; WATSON, B. N. J.; FINERAN, P. C. The arms race between bacteria and their phage foes. **Nature**, v. 577, n. 7790, p. 327–336, 15 jan. 2020.

HARDIE, J. M.; WHILEY, R. A. The genus Streptococcus. In: **The Genera of Lactic Acid Bacteria**. Boston, MA: Springer US, 1995. p. 55–124.

HOLTZMAN, T. et al. A continuous evolution system for contracting the host range of bacteriophage T7. **Scientific Reports**, v. 10, n. 1, p. 1–8, 2020.

HOLZAPFEL, W. H. Appropriate starter culture technologies for small-scale fermentation in developing countries. **International Journal of Food Microbiology**, v. 75, n. 3, p. 197–212, 2002.

HOWARD-VARONA, C. et al. Phage-specific metabolic reprogramming of virocells. **ISME Journal**, 2020.

HYMAN, P. Phages for phage therapy: Isolation, characterization, and host range breadth. **Pharmaceuticals**, v. 12, n. 1, 2019.

JANG, J. et al. Environmental Escherichia coli: ecology and public health

implications—a review. **Journal of Applied Microbiology**, v. 123, n. 3, p. 570–581, 2017.

JUNG, M. J. et al. Viral community predicts the geographical origin of fermented vegetable foods more precisely than bacterial community. **Food Microbiology**, v. 76, n. June, p. 319–327, 2018.

KAMIMURA, B. A. et al. Large-scale mapping of microbial diversity in artisanal Brazilian cheeses. **Food Microbiology**, v. 80, n. October 2018, p. 40–49, 2019.

KLUMPP, J.; LOESSNER, M. J. **New research on bacteriophages and food safety**. [s.l.] Woodhead Publishing Limited, 2013. v. 1

KNOWLES, B. et al. Lytic to temperate switching of viral communities. **Nature**, v. 531, n. 7595, p. 466–470, 2016.

KORTRIGHT, K. E. et al. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. **Cell Host & Microbe**, v. 25, n. 2, p. 219–232, fev. 2019.

KOT, W. et al. Bacteriophages of Leuconostoc, Oenococcus, and Weissella. **Frontiers in Microbiology**, v. 5, n. APR, p. 1–9, 2014.

LABRIE, S. J.; SAMSON, J. E.; MOINEAU, S. Bacteriophage resistance mechanisms. **Nature Reviews Microbiology**, v. 8, n. 5, p. 317–327, 2010.

LEBLANC, J. G.; TODOROV, S. D. Bacteriocins producing lactic acid bacteria isolated from Boza, a traditional fermented beverage from Balkan Peninsula – from isolation to application. In: MÉNDEZ-VILAS, A. (Ed.). . **Science against microbial pathogens: communicating current research and technological advances**. Badajos, Spain: Formatex Research Center, 2011. p. 1311–1320.

LEDORMAND, P.; DESMASURES, N.; DALMASSO, M. Phage community involvement in fermented beverages: an open door to technological advances? **Critical Reviews in Food Science and Nutrition**, v. 0, n. 0, p. 1–10, 2020.

LEMOS, L. N. et al. Genome-Resolved Metagenomics Is Essential for Unlocking the

Microbial Black Box of the Soil. **Trends in Microbiology**, v. xx, n. xx, p. 4–7, 2021.

LIANG, G.; BUSHMAN, F. D. The human virome: assembly, composition and host interactions. **Nature Reviews Microbiology**, v. 19, n. August, 2021.

MAHONY, J. et al. Phage-Host Interactions of Cheese-Making Lactic Acid Bacteria. **Annual Review of Food Science and Technology**, v. 7, n. 1, p. 267–285, 2016.

MAHONY, J.; VAN SINDEREN, D. Current taxonomy of phages infecting lactic acid bacteria. **Frontiers in Microbiology**, v. 5, n. JAN, p. 1–7, 2014.

MARCÓ, M. B.; MOINEAU, S.; QUIBERONI, A. Bacteriophages and dairy fermentations. **Bacteriophage**, v. 2, n. 3, p. 149–158, 2012.

MATUSZEWSKA, M. et al. The Evolutionary Genomics of Host Specificity in Staphylococcus aureus. **Trends in Microbiology**, v. xx, n. xx, p. 1–13, 2020.

MEYER, J. R. et al. Ecological speciation of bacteriophage lambda in allopatry and sympatry. **Science**, v. 354, n. 6317, p. 1301–1304, 9 dez. 2016.

MOKILI, J. L.; ROHWER, F.; DUTILH, B. E. Metagenomics and future perspectives in virus discovery. **Current Opinion in Virology**, v. 2, n. 1, p. 63–77, 2012.

MUHAMMED, M. K. et al. Metagenomic Analysis of Dairy Bacteriophages: Extraction Method and Pilot Study on Whey Samples Derived from Using Undefined and Defined Mesophilic Starter Cultures. **Applied and Environmental Microbiology**, v. 83, n. 19, p. e00888-17, 1 out. 2017.

NOBREGA, F. L. et al. Targeting mechanisms of tailed bacteriophages. **Nature Reviews Microbiology**, 2018.

OFIR, G. et al. DISARM is a widespread bacterial defence system with broad anti-phage activities. **Nature Microbiology**, v. 3, n. 1, p. 90–98, 2018.

PAEZ-ESPINO, D. et al. Uncovering Earth's virome. **Nature**, v. 536, n. 7617, p. 425–430, 2016.

PERIN, L. M. et al. Bacterial ecology of artisanal Minas cheeses assessed by culture-

dependent and -independent methods. **Food Microbiology**, v. 65, p. 160–169, 2017.

PHILIPPE, C. et al. Characterization of the First Virulent Phage Infecting Oenococcus oeni, the Queen of the Cellars. **Frontiers in Microbiology**, v. 11, n. January, p. 1–13, 2021.

PINEDA, A. P. A. et al. Brazilian Artisanal Cheeses: Diversity, Microbiological Safety, and Challenges for the Sector. **Frontiers in Microbiology**, v. 12, n. April, 2021.

REZAC, S. et al. Fermented foods as a dietary source of live organisms. **Frontiers in Microbiology**, v. 9, n. AUG, 2018.

RODRIGUEZ-BRITO, B. et al. Viral and microbial community dynamics in four aquatic environments. **The ISME Journal**, v. 4, n. 6, p. 739–751, 2010.

ROSS, A.; WARD, S.; HYMAN, P. More is better: Selecting for broad host range bacteriophages. **Frontiers in Microbiology**, v. 7, n. SEP, p. 1–6, 2016.

ROUX, S. et al. Minimum information about an uncultivated virus genome (MIUVIG). **Nature Biotechnology**, v. 37, n. 1, p. 29–37, 2019.

SAMSON, J. E. et al. Revenge of the phages: Defeating bacterial defences. **Nature Reviews Microbiology**, v. 11, n. 10, p. 675–687, 2013.

SAMSON, J. E.; MOINEAU, S. Bacteriophages in Food Fermentations: New Frontiers in a Continuous Arms Race. **http://dx.doi.org/10.1146/annurev-food-030212-182541**, v. 4, n. 1, p. 347–368, 6 mar. 2013.

SANT, D. G. et al. Host diversity slows bacteriophage adaptation by selecting generalists over specialists. **Nature Ecology & Evolution**, v. 5, n. 3, p. 350–359, 11 mar. 2021.

SUTTLE, C. A. Environmental microbiology: Viral diversity on the global stage. **Nature Microbiology**, v. 1, n. 11, p. 1–2, 2016.

TAMANG, J. P. et al. Fermented foods in a global age: East meets West. **Comprehensive Reviews in Food Science and Food Safety**, v. 19, n. 1, p. 184–

217, 2020.

TEUBER, M. The genus Lactococcus. In: **The Genera of Lactic Acid Bacteria**. Boston, MA: Springer US, 1995. p. 173–234.

THINGSTAD, T. F. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. **Limnology and Oceanography**, v. 45, n. 6, p. 1320–1328, 2000.

THOMAS, T. et al. Metagenomics - a guide from sampling to data analysis. **Microbial Informatics and Experimentation**, v. 2, n. 1, p. 3, 2012.

TODOROV, S. D.; HOLZAPFEL, W. H. Traditional cereal fermented foods as sources of functional microorganisms. In: **Advances in Fermented Foods and Beverages**. [s.l.] Elsevier, 2015. p. 123–153.

VAN DALEN, R.; PESCHEL, A.; VAN SORGE, N. M. Wall Teichoic Acid in Staphylococcus aureus Host Interaction. **Trends in Microbiology**, v. 28, n. 12, p. 985–998, 2020.

VERREAULT, D. et al. Detection of airborne lactococcal bacteriophages in cheese manufacturing plants. **Applied and Environmental Microbiology**, v. 77, n. 2, p. 491–497, 2011.

WALSH, A. M. et al. Meta-analysis of cheese microbiomes highlights contributions to multiple aspects of quality. **Nature Food**, v. 1, n. 8, p. 500–510, 13 ago. 2020.

WHO. **Antimicrobial resistance: global report on surveillanceWorld Health Organization**. [s.l: s.n.]. Disponível em: <http://apps.who.int/iris/handle/10665/112642>.

WINTER, C. et al. Trade-Offs between Competition and Defense Specialists among Unicellular Planktonic Organisms: the “Killing the Winner” Hypothesis Revisited. **Microbiology and Molecular Biology Reviews**, v. 74, n. 1, p. 42–57, 2010.

WOLFE, B. E. et al. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. **Cell**, v. 158, n. 2, p. 422–433, 2014.

WOLFE, B. E.; DUTTON, R. J. Fermented Foods as Experimentally Tractable Microbial Ecosystems. **Cell**, v. 161, n. 1, p. 49–55, 26 mar. 2015.

ZORBA, M. et al. The use of starter cultures in the fermentation of boza, a traditional Turkish beverage. **Process Biochemistry**, v. 38, n. 10, p. 1405–1411, 2003.

APPENDIX 1 - Supplementary Information of Chapter 1

Supplementary Figures

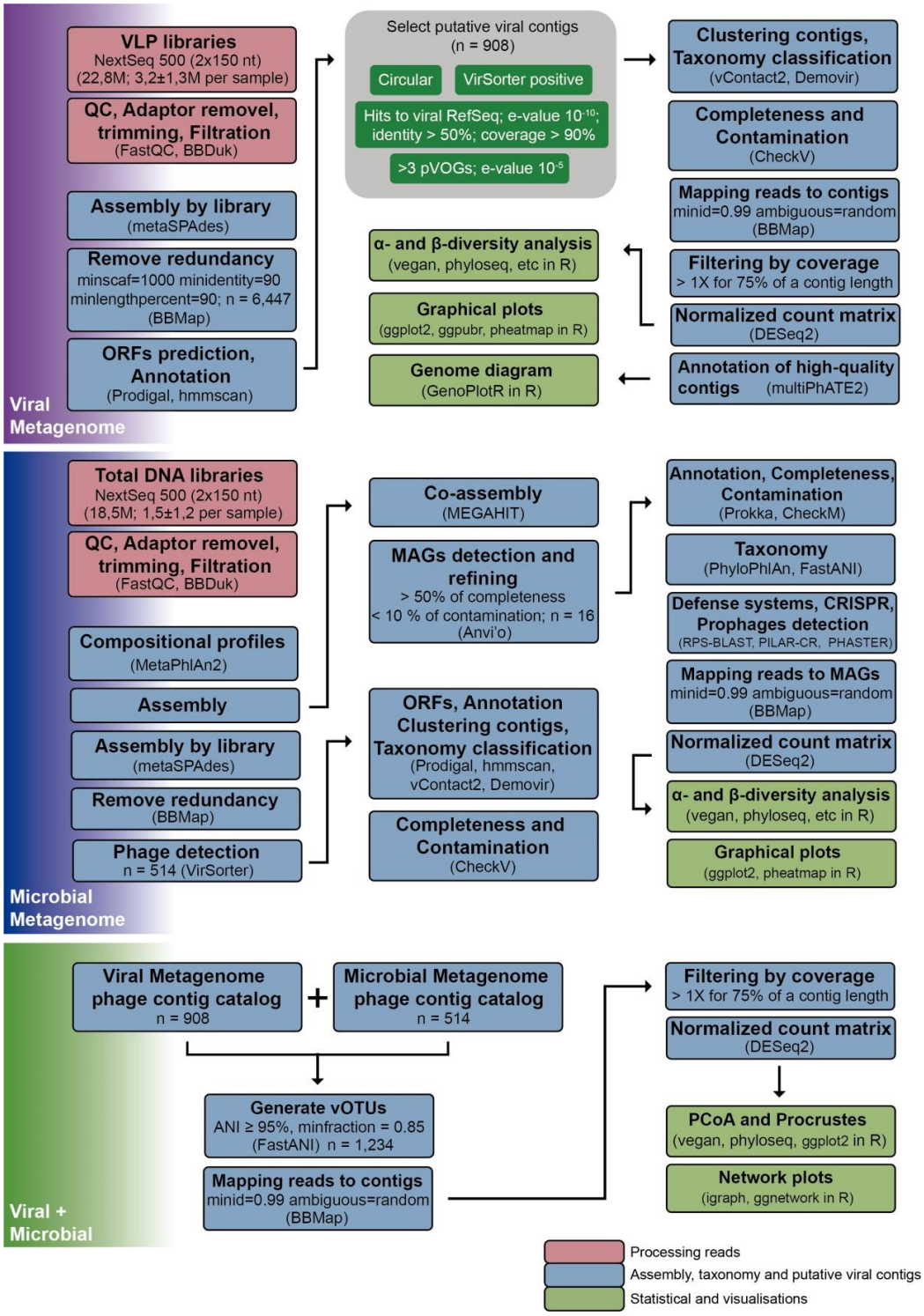


Fig S1. Bioinformatics pipeline and statistics analysis used for processing viral metagenome (VLPs) and microbial metagenome.

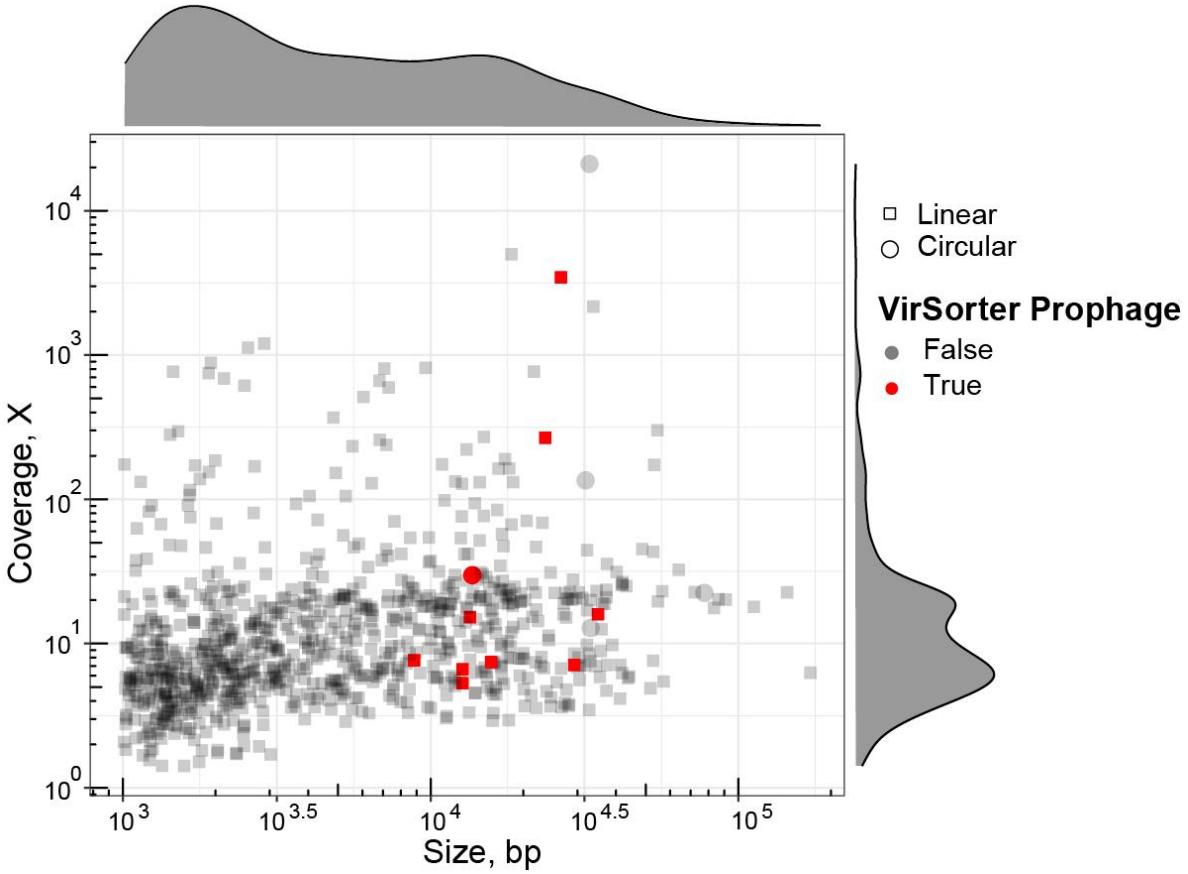


Fig S2. Viral contigs recovered by viral metagenome (VLP) and classified as prophage by VirSorter. Size and coverage distribution of 908 putative viral genomes detected (coverage depth by genome length in bp) and which were classified as prophage by VirSorter.

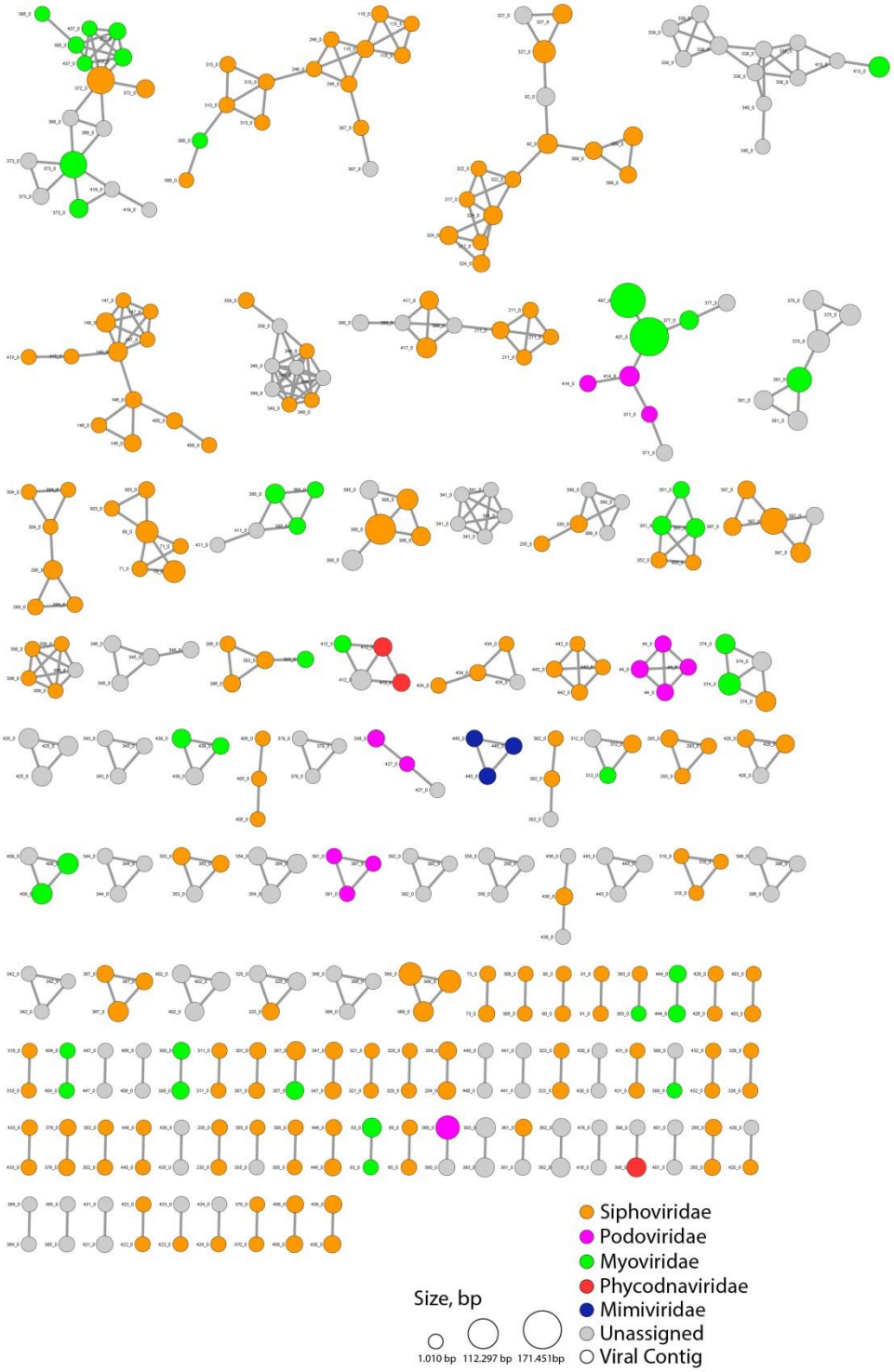


Fig S3. Viral Clusters (VC) formed between viral contigs. Viral contigs and their family level classification are represented by circles. Colors indicate viral families and the size of circles represent the viral contig length in base pairs.

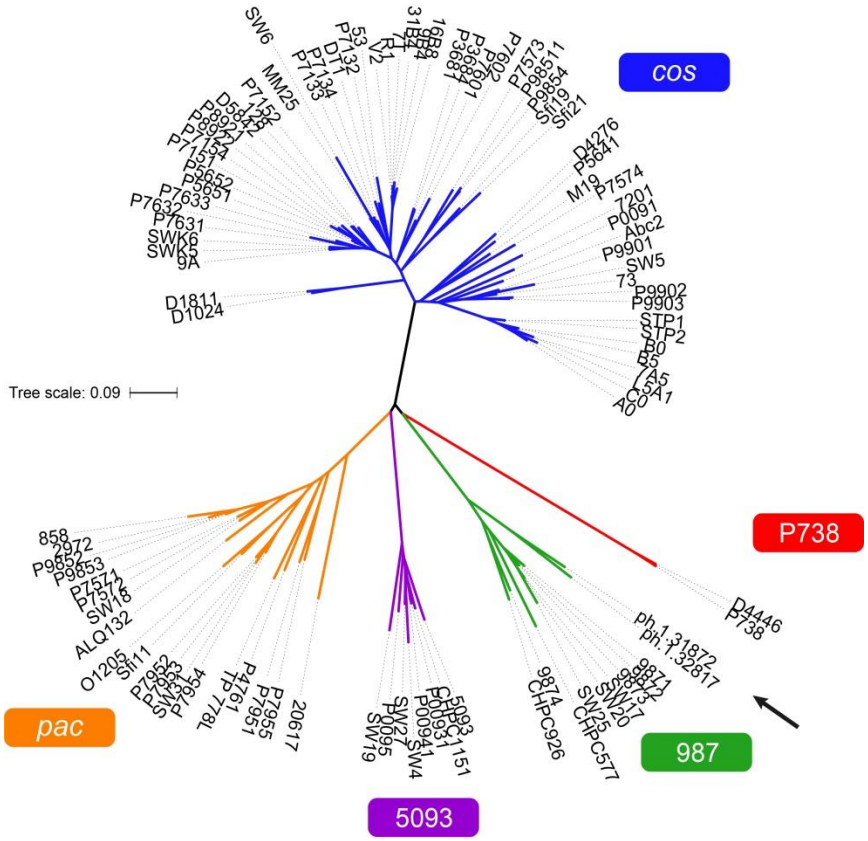


Fig S4. Phylogenomic tree of *Streptococcus* phages. Colors represent the genus of phages and the arrow indicates the genomes of phages ph.1.31872 and ph.1.32817 located within genus 987 group.

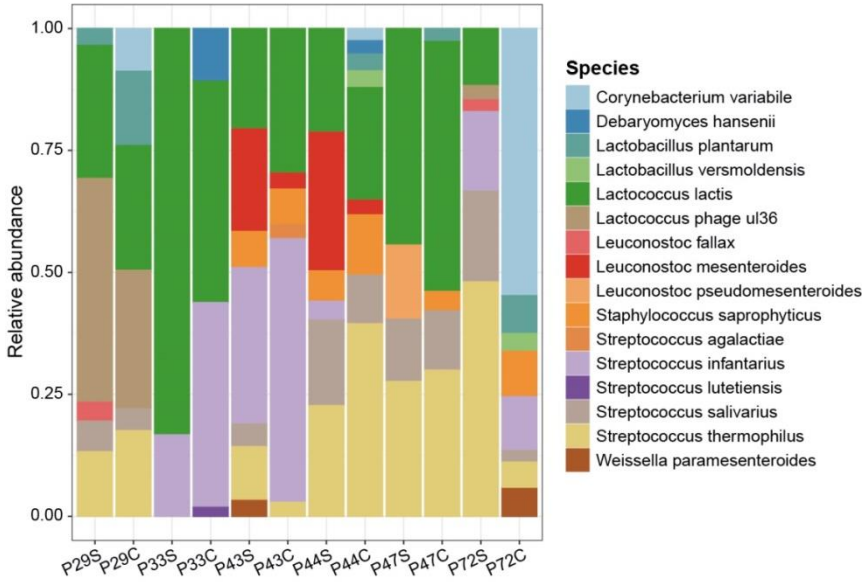


Fig S5. Compositional profiles of microbial metagenomes in endogenous starter culture and cheese. Relative abundance of each species in 12 paired starter culture (e.g. P29S) and cheese samples (e.g. P29C) assessed using MetaPhlAn2**.**

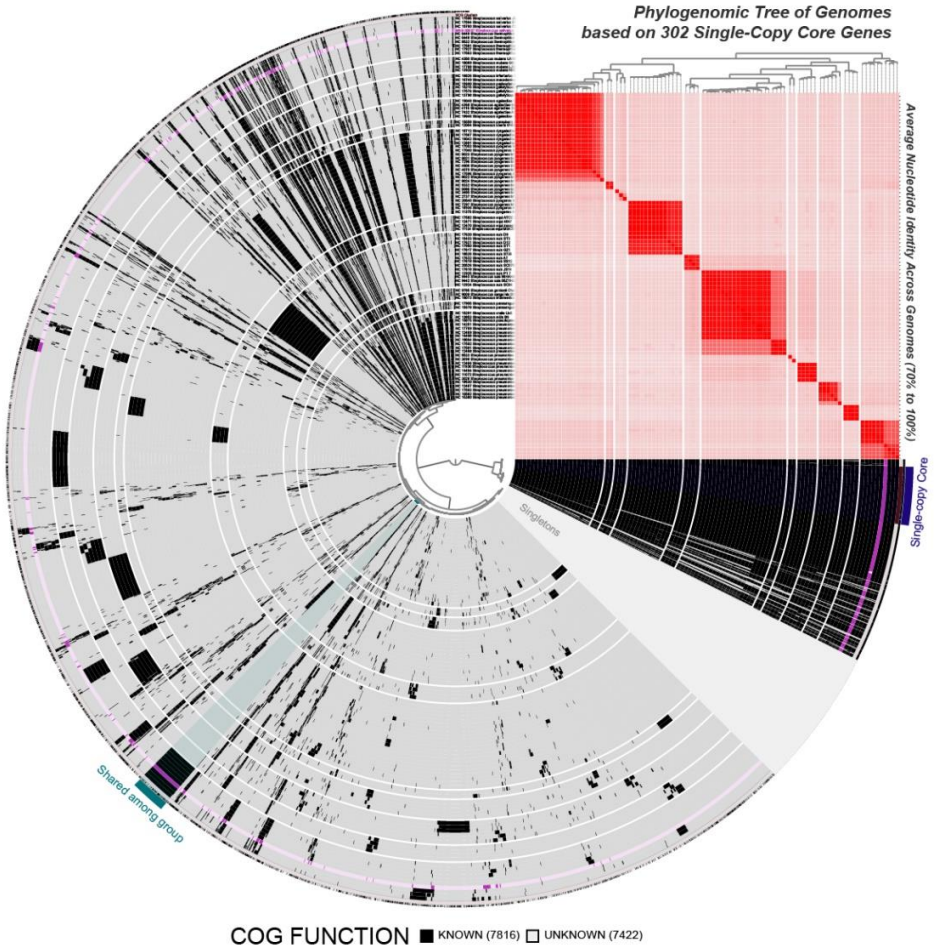


Fig S6. Pangenomic analysis of *Streptococcus* genus. It was performed with 94 complete genomes selected from previous study and metagenome-assembled genomes classified as *Streptococcus salivarius* (MAG7; highlighted in purple).

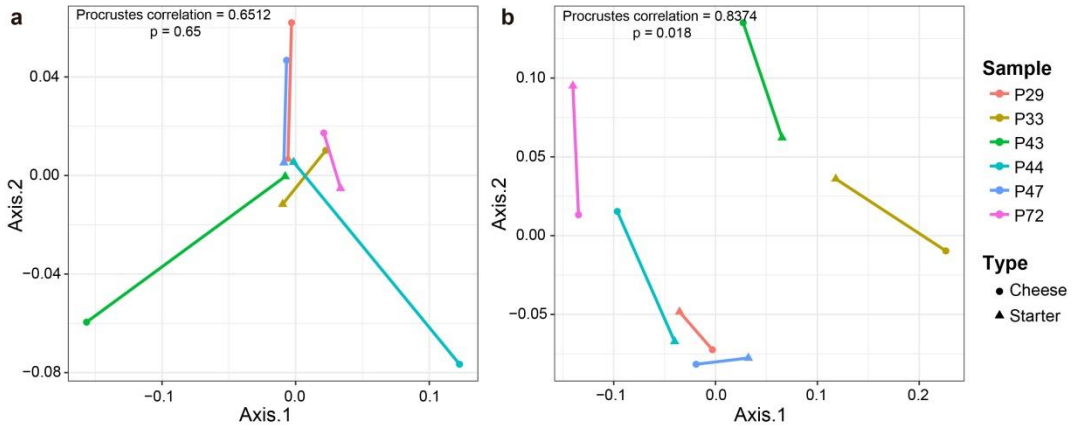


Fig S7. Procrustes analysis using PCoA coordenations of starter culture and cheese communities. a, Correlations between viral community in starter cultures and cheese samples. **b**, Correlations between bacterial communities in starter cultures and cheese samples. Distance matrices were calculated using Jensen-Shannon divergence in both cases.

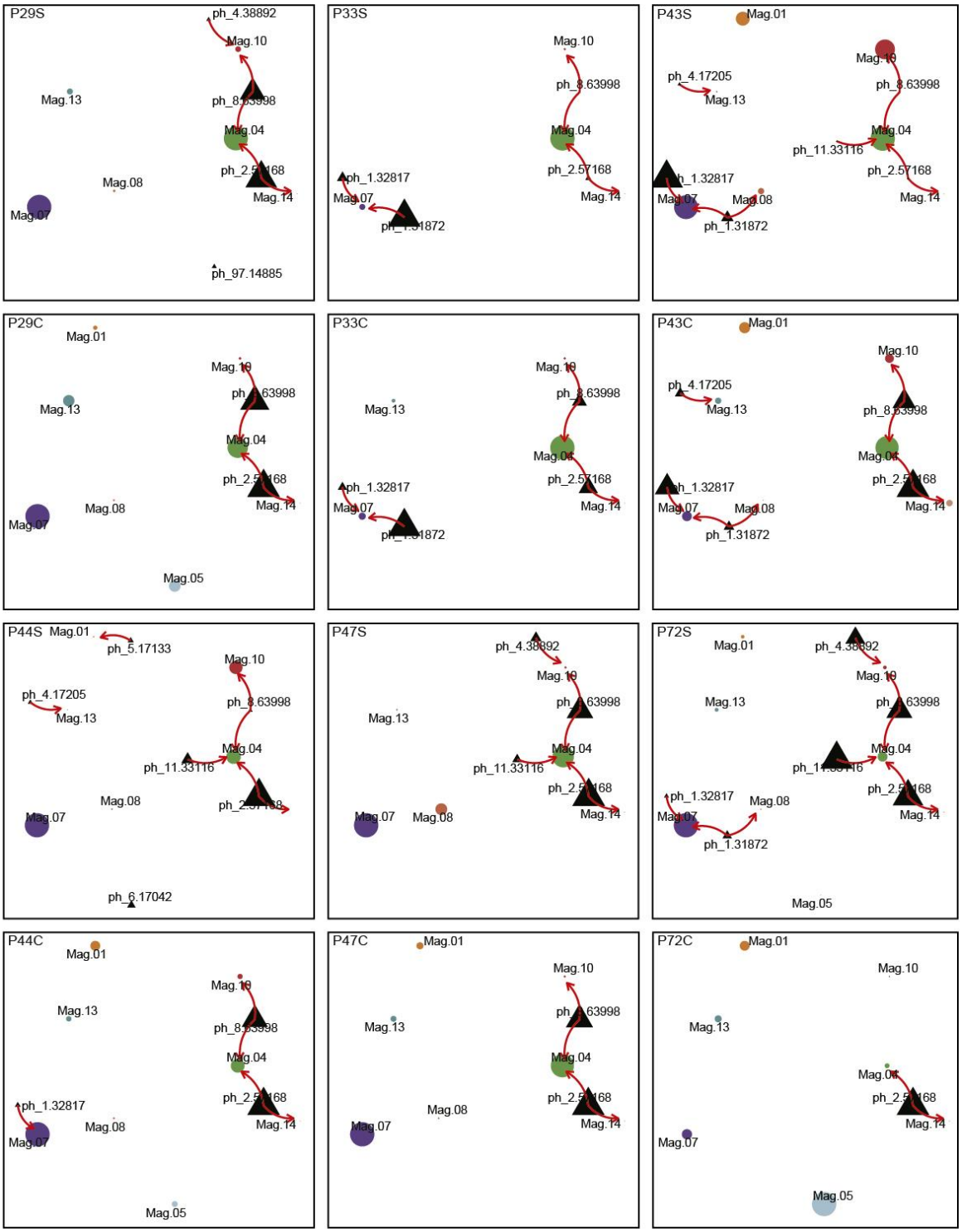


Fig S8. Phage-bacteria interaction network during cheese production. Putative phage-bacteria infection interactions for all producers (e.g. P29 and P33), and for separate starter and cheese samples (e.g. P29S and P29C). Nodes represent each MAG and phage analysed, edges represent putative infection relationships between phage and bacteria. This figure refers to Fig. 1.5.

Supplementary Tables

The supplementary tables with more than 10 columns or rows are available in preprint format at bioRxiv (QUEIROZ et al. 2021).

DOI https://doi.org/10.1101/2021.08.03.454940

Table S1. Alpha-diversity of viral metagenome

Sample	Reads	Observed	Shannon	Simpson
P29	3561923	759	6.83812	0.94896
P33	53071	108	2.64524	0.53693
P43	7500568	108	1.29953	0.39765
P44	438574	465	6.73909	0.97449
P47	1493347	384	4.66591	0.78968
P72	37684	286	7.46798	0.992
P85	234953	313	7.35305	0.98982

Table S2. Significant correlations between phage and MAG abundances from starter to cheese

Phage	MAG	cor	pval
NODE_11_length_33116_cov_15.986812	CANASTRA_MAG_00004	-0.845154255	0.034109423
NODE_1_length_31872_cov_135.581324	CANASTRA_MAG_00007	-0.880406274	0.020598735
NODE_1_length_32817_cov_14173.771931	CANASTRA_MAG_00007	-0.880406274	0.020598735
NODE_1_length_31872_cov_135.581324	CANASTRA_MAG_00008	-0.83165549	0.04012438
NODE_8_length_63998_cov_19.026352	CANASTRA_MAG_00010	-0.885714286	0.018845481
NODE_2_length_57168_cov_11.104757	CANASTRA_MAG_00014	-0.885714286	0.018845481
NODE_1_length_32817_cov_14173.771931	CANASTRA_MAG_00007	-0.845154255	0.034109423
NODE_4_length_38892_cov_19.980174	CANASTRA_MAG_00010	-0.880406274	0.020598735
NODE_4_length_17205_cov_17.695569	CANASTRA_MAG_00013	-0.845154255	0.034109423

Supplementary Methods

List of phage-specific defense systems used to manual filtering identification of marker genes in MAGs genomes. This list was based on BEZUIDT et al. (2020).

R.M	*hsdR*, *hsdM/hsdS*, *yhdJ, ssl2, mcrA, yeeA , mrr, COG2810, mcrBC*
DISARM	*drmA, drmB, drmC , drmD*, *drmMI, drmMII*
BREX	*brxA, brxC*, *brxHI, brxL, brxP, pglX*, *pglW*, *pglZ*
Druantia	*druE, druM*
Abi	*abiEi*, *abiEii*, *abiF*, *abiTii*, *abiH*, *abiU2*, *abiJ*, *abiG*, *abiA*, *abiV*, *abiGi*, abiGii, abiC
Zorya	*zorA*/*zorB, zorC,* zorD, zorE
Septu	*ptuA*, *ptuB*
Gabija	*gajA, gajB*
Theoris	*thsA, thsB*
CRISPR-cas Type I	*cas3 , cas5, csp1, csp2*
Type I-A	*cas8a1, csx13*
Type I-B	*cas1-HMARI*, *cas1-MYXAN, csh2, cst2, cmx8*
Type I-C	*cas5d, csd1, csx17*
Type I-D	*csc1, csc2, csc3, cas10d, cas8c*
Type I-E	*cse1, cse2, cas5e, PRK13921*

Type I-F	*csy1, csy2, csy3, csy4, cas1-YPEST*
Type I-U	*GSU0052, GSU0053, GSU0054, csb1, csb2, csb3, csx15*
CRISPR-cas Type II	*cas1, csn1, cas-NMENI, cas9*
Type II-B	*csx12*
CRISPR-cas Type III	*cas10, cas6, csx1, csx3*
Type III-A	*csm1, csm2, csm3, csm4, csm5, csm6, TM1806*
Type III-B	*cmr1, cmr3, cmr4, cmr5, cmr6, csx1*
Type III-BC	*cmr1, cmr3, cmr4, cmr5, cmr6, csx1*
Type III-D	*csx10*

Supplementary Results

Bacteriophage taxonomy classification. Using CheckV, as well as the Minimum Information about an Uncultivated Virus Genome (MIUViG) criterium (ROUX et al. 2019), we classified our genomes as: complete (5), high-quality (12), medium-quality (23), low-quality (584), and not-determined quality (284). We further refined the classification of our viral catalog by comparing it to the RefSeq complete viral genome database using BLAST and stringent criteria (e-value < 10^{-10}, coverage > 90% of contig length and > 50% identity), and obtained 94 viruses classified at species level. The detected viruses were *Lactococcus* phages 949, asccphi28, bIL285, bIL286, bIL309, bIL312, P078, P162 and ul36. We also found *Lactobacillus* phages, such as phiAQ113 and phiJL1; *Staphylococcus* phages GRCS and phiSA12; and *Streptococcus* phages 9872 and 9874.

Viral contigs clustered with RefSeq genomes belong to *Siphoviridae*, *Myoviridae* and *Podoviridae* families and their sizes ranged from 1,001 to 54,569 bp. We observed that some VCs were composed of complete (VC 148: *Streptococcus virus*) and high-quality (VC 248: *Staphylococcus phage*; VC 44: *Actinomyces virus*; VC 67: *Arthrobacter*, *Gordonia*, and *Rhodococcus phages*; and VC 69: *Lactococcus phage*) viral contigs. Also, a high-quality genome clustered with 7 RefSeq genomes of *Staphylococcus phages* (i.e. GRCS) that belong to the genus *Rosenblumvirus* (VC 248).

The ANI measured between ph.72.18300 (VC320) and asccphi28 genomes was 93%, well below the usually accepted threshold for same species classification, indicating that this phage is potentially a novel viral species within the *Lactococcus* phage asccphi28 group. This phage was detected in high abundance in only one of the studied producers.

Prophage characterization. Four intact prophage sequences were identified in MAGs classified as *Leuconostoc fallax* (MAG2), *Escherichia coli* (MAG3), *Streptococcus infantarius* (MAG6), and *Streptococcus salivarius* (MAG7). All these sequences were classified at family level as *Siphoviridae*. The most common gene annotations produced by PHASTER were *Lactobacillus* phage Sha1 (7 genes) for the prophage found in MAG2, *Salmonella* phage SEN34 (6) for the prophage in MAG3, *Streptococcus* phage phiNJ2 (20) for the prophage in MAG6, and *Streptococcus* phage 5093 (9) for the prophage in MAG7.

Supplementary References

BEZUIDT, O. K. I. et al. Phages Actively Challenge Niche Communities in Antarctic Soils. mSystems, v. 5, n. 3, p. 1–12, 2020.

QUEIROZ, L. L. et al. High Level of Interaction between Phages and Bacteria in an Artisanal Raw Milk Cheese Microbial Community. bioRxiv, p. 1–32, 2021.

ROUX, S. et al. Minimum information about an uncultivated virus genome (MIUVIG). Nature Biotechnology, v. 37, n. 1, p. 29–37, 2019.

APPENDIX 2 - Supplementary Information of Chapter 2

Supplementary Figures

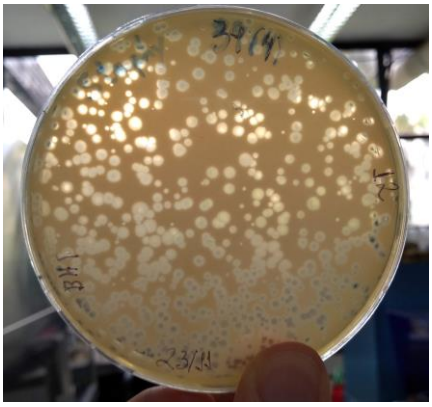


Fig S1. Agar overlay plaque assay and phage isolation from starter sample. The phage CS44 was isolated from producer P44 infecting *Staphylococcus aureus* host.

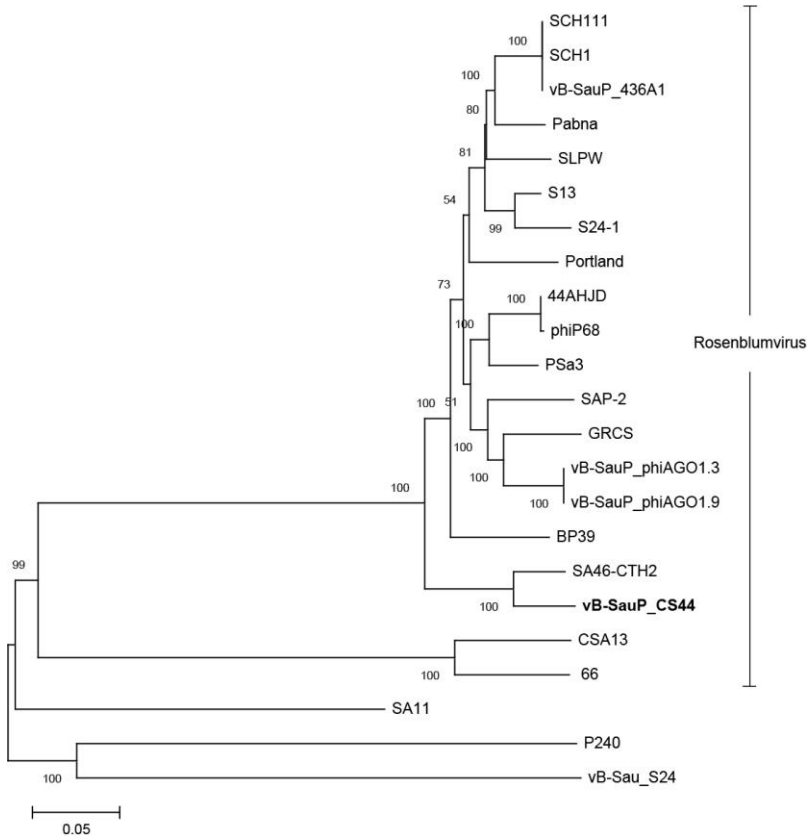


Fig. S2. Phylogenomic tree of *Rosenblumvirus* genus. The tree was generated using neighbor joining with P distance values and 1000 replicate bootstrap.

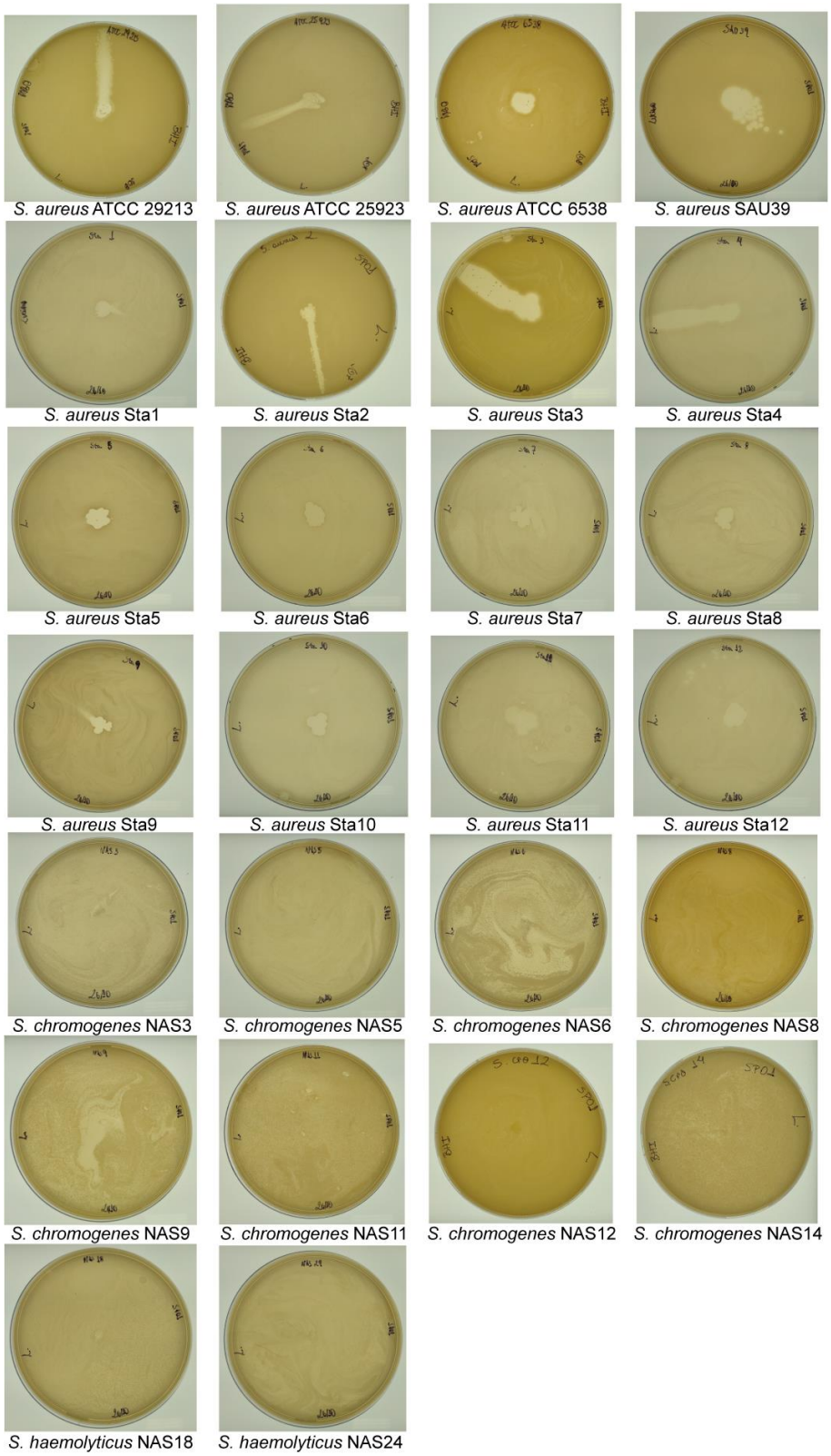


Fig. S3. Host range determination for phage CS44 by spots test against *Staphylococcus* strains. Plaque formation was observed on all strains of *Staphylococcus aureus* and none of *Staphylococcus chromogenes* and *Staphylococcus haemolyticus*.

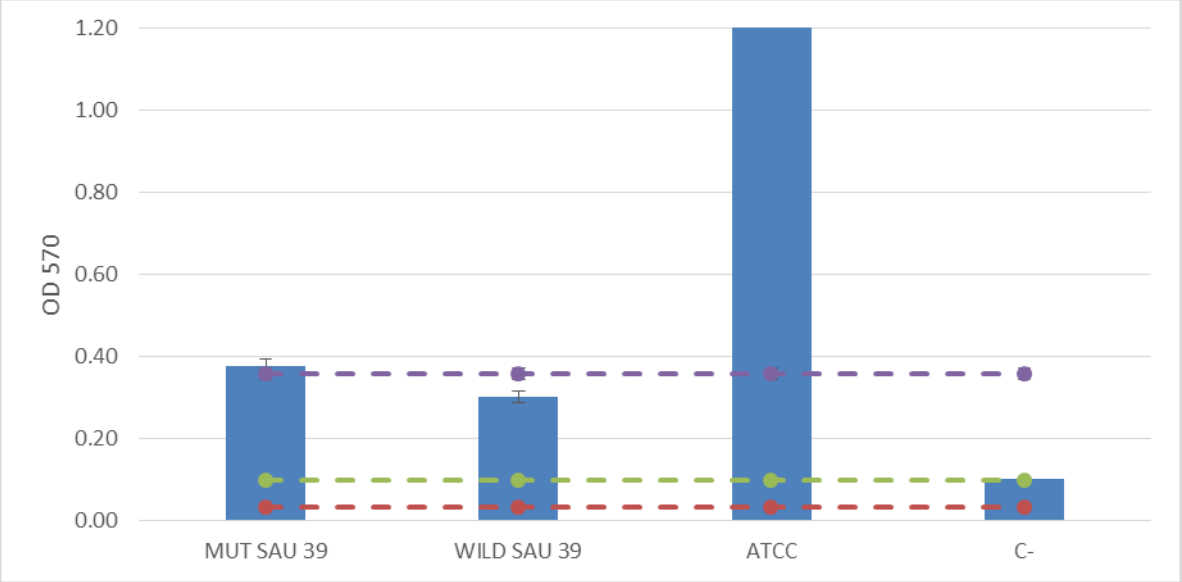


Fig. S4. Biofilm formation assay. The dotted lines delimit the biofilm production ranges, below the first (DOc = 0.03) the culture was classified as non-biofilm forming; between the first and second row (2 × OCD = 0.07), as weakly forming; between the second and third lines (4 × DOc = 0.26) as moderately formative; and above the third line as strongly formative. These experiments were carried out in collaboration with the co-authors of the manuscript Ana Paulina Pineda and Uelinton Pinto.

Supplementary Tables

Table S1. Samples collected location, sample types and study methods applied for each one.

Producer	Location	Sample type	VLPs	Plaque Assay	Phage Isolated	Bacteria Isolated
P29	São Roque de Minas	Starter	+	+	-	-
P30	São Roque de Minas	Starter	+	+	-	-
P32	Medeiros	Starter	+	+	-	-
P33	Medeiros	Starter	+	+	-	-
P37	São Roque de Minas	Starter	+	+	-	-
P39	São Roque de Minas	Starter and Cheese	-	-	-	+
P42	Medeiros	Starter	+	+	-	-
P43	Medeiros	Starter	+	+	-	-
P44	Medeiros	Starter	+	+	+	-
P47	São Roque de Minas	Starter	+	+	-	-

P66	Medeiros	Starter and Cheese	+	+	-	+
P72	São Roque de Minas	Starter	+	+	-	-
P85	São Roque de Minas	Starter	+	+	-	-

Table S2. Bacterial strain used in this study and to host range determination.

Strain	Infection	Source[a]
Staphylococcus aureus SAU39	+	This study
Staphylococcus aureus SAU66	NA	This study
Staphylococcus aureus ATCC 29213	+	Lab. collection
Staphylococcus aureus ATCC 25923	+	Lab. collection
Staphylococcus aureus ATCC 6538	+	Lab. collection
Staphylococcus aureus Sta1	+	FMVZ
Staphylococcus aureus Sta2	+	FMVZ
Staphylococcus aureus Sta3	+	FMVZ
Staphylococcus aureus Sta4	+	FMVZ
Staphylococcus aureus Sta5	+	FMVZ
Staphylococcus aureus Sta6	+	FMVZ
Staphylococcus aureus Sta7	+	FMVZ
Staphylococcus aureus Sta8	+	FMVZ
Staphylococcus aureus Sta9	+	FMVZ
Staphylococcus aureus Sta10	+	FMVZ
Staphylococcus aureus Sta11	+	FMVZ
Staphylococcus aureus Sta12	+	FMVZ
Staphylococcus chromogenes NAS3	-	FMVZ
Staphylococcus chromogenes NAS5	-	FMVZ
Staphylococcus chromogenes NAS6	-	FMVZ
Staphylococcus chromogenes NAS8	-	FMVZ
Staphylococcus chromogenes NAS9	-	FMVZ
Staphylococcus chromogenes NAS11	-	FMVZ
Staphylococcus chromogenes NAS12	-	FMVZ
Staphylococcus chromogenes NAS14	-	FMVZ
Staphylococcus haemolyticus NAS18	-	FMVZ
Staphylococcus haemolyticus NAS24	-	FMVZ

[a] FMVZ, Faculdade de Medicina Veterinária e Zootecnia from Universidade de São Paulo (Brazil).

Table S3. Complete genomes of Staphylococcus phage used to build the maximum likelihood phylogenetic tree and accession code.

Genome	Accession
Staphylococcus phage SA46-CL1	NC_055802.1
Staphylococcus phage GRCS	NC_023550.1

Staphylococcus phage BP39	NC_031046.1
Staphylococcus phage vB_SauP_phiAGO1.3	NC_047919.1
Staphylococcus phage CSA13	NC_048159.1
Staphylococcus phage Portland	NC_055814.1
Staphylococcus phage SAP-2	NC_009875.1
Staphylococcus phage phiP68	NC_004679.1
Staphylococcus phage 66	NC_007046.1
Staphylococcus phage S24-1	NC_016565.1
Staphylococcus phage 44AHJD	NC_004678.1
Staphylococcus phage Pabna	NC_048107.1
Staphylococcus phage SCH1	NC_047788.1
Staphylococcus phage PSa3	NC_047855.1
Staphylococcus phage SLPW	NC_031008.1
Staphylococcus phage S13'	AB626963.1
Staphylococcus phage vB_SauP_phiAGO1.9	MG766219.2
Staphylococcus phage SCH111	KY000085.1
Staphylococcus phage vB_SauP-436A1	MN150710.1
Staphylococcus phage SA11	NC_019511.1
Staphylococcus phage P240	KY056620.1
Staphylococcus phage vB_Sau_S24	KY794643.1

Table S4. Values of coverage and identity of query sequence of *Staphylococcus* phage CS44 against most similar bacteriophages genomes.

Strain	Query Coverage	Percentage of identity
SA46-CTH2	100%	93.28%
GRCS	97%	85.52%
BP39	99%	84.78%
vB-SauP_phiAGO1.3	98%	84.66%
CSA13	94%	83.79%
Portland	97%	85.80%
SAP-2	98%	82.71%
P68	96%	84.00%
P66	95%	83.49%
S24-1	96%	83.46%
phi44AHJD	92%	84.08%
Pabna	96%	83.07%
SCH1	95%	83.05%
PSa3	93%	84.77%
SLPW	92%	84.31%

Table S5. Comparative analysis and search for variant calling between SAU39 wild and resistant.

CONTIG	POS	TYPE	REF	ALT	EVIDENCE	GENE	PRODUCT
NODE_1	59149	snp	G	A	A:13 G:1		hypothetical protein
NODE_1	132923	snp	G	T	T:172 G:2		hypothetical protein
NODE_1	300803	snp	T	G	G:69 T:5		No codon region
NODE_1	300839	complex	GCTA	ACTG	ACTG:11 GCTA:0		No codon region
NODE_2	78	snp	T	C	C:12 T:0		No codon region
NODE_2	83	snp	C	T	T:10 C:1		No codon region
NODE_2	146127	snp	T	C	C:87 T:0	dus_2	putative tRNA-dihydrouridine synthase
NODE_2	146162	snp	G	A	A:156 G:0	dus_2	putative tRNA-dihydrouridine synthase
NODE_2	159212	snp	A	G	G:355 A:0		putative lipoprotein
NODE_3	35619	snp	G	A	A:307 G:1	tarS	Poly(ribitol-phosphate) beta-N-acetylglucosaminyltransferase TarS
NODE_3	80184	snp	C	T	T:339 C:0		hypothetical protein
NODE_4	147679	mnp	AC	GT	GT:63 AC:0		hypothetical protein
NODE_4	147697	snp	A	G	G:29 A:0		hypothetical protein
NODE_6	98408	snp	G	A	A:194 G:0		No codon region
NODE_8	47383	snp	G	A	A:211 G:0		No codon region
NODE_8	87276	complex	NNN	ATT	ATT:13 NNN:0		No codon region
NODE_8	92527	snp	G	A	A:46 G:0	fnbB	Fibronectin-binding protein B
NODE_8	92551	snp	A	G	G:61 A:0	fnbB	Fibronectin-binding protein B
NODE_8	95398	complex	GCTTT	TTTTG	TTTTG:111 GCTTT:0	fnbA	Fibronectin-binding protein A
NODE_9	43763	snp	C	T	T:21 C:0		hypothetical protein
NODE_9	43769	snp	A	T	T:13 A:0		hypothetical protein
NODE_9	44744	snp	T	C	C:11 T:0		hypothetical protein
NODE_9	46196	snp	A	T	T:15 A:0		hypothetical protein
NODE_9	46205	snp	G	A	A:34 G:0		hypothetical protein
NODE_9	47145	complex	TGGA	AGGC	AGGC:33 TGGA:0		hypothetical protein
NODE_9	47154	snp	C	A	A:28 C:1		hypothetical protein
NODE_11	77915	complex	TGGTCAAG	CGGCCAAA	CGGCCAA A:17 TGGTCAA G:2		No codon region
NODE_14	52925	snp	T	C	C:179 T:19		hypothetical protein

NODE_16	42340	snp	A	C	C:86 A:1		putative protein
NODE_16	42357	snp	T	C	C:119 T:0		putative protein
NODE_16	42367	snp	A	T	T:127 A:2		putative protein
NODE_16	48882	del	TG	T	T:163 TG:0 AAGGA:10	lpl2_8	putative lipoprotein
NODE_16	48895	complex	TAGAT	AAGGA	9 TAGAT:0	lpl2_8	putative lipoprotein
NODE_16	48910	snp	G	A	A:143 G:0 AA:10	lpl2_8	putative lipoprotein
NODE_23	11021	complex	NN	AA	NN:0 CT:48		No codon region
NODE_34	3755	mnp	NN	CT	NN:0		No codon region
NODE_36	5684	snp	C	T	T:11 C:0		hypothetical protein
NODE_36	5693	snp	C	T	T:17 C:0		hypothetical protein

APPENDIX 3 - Supplementary Information of Chapter 3

Supplementary Figures

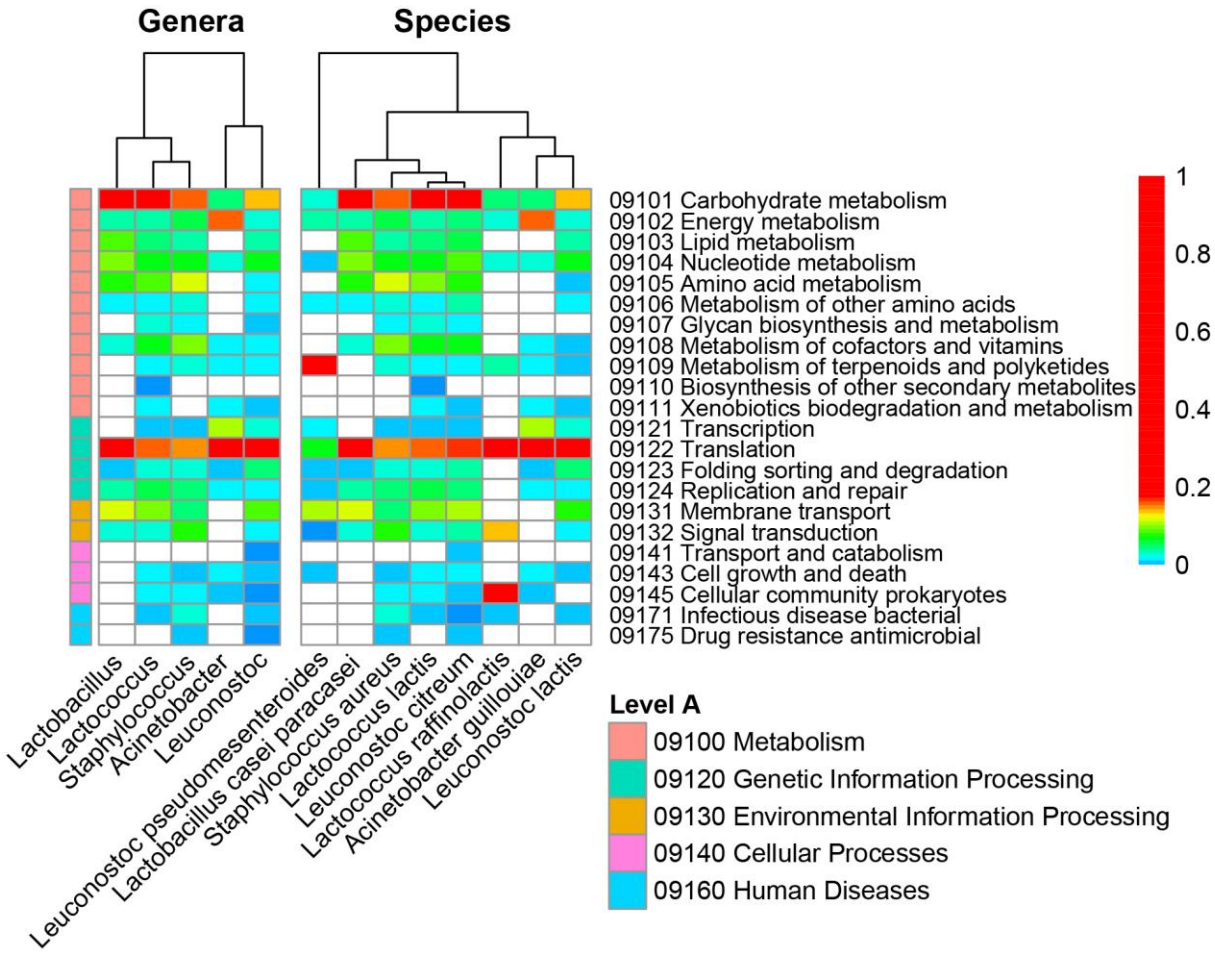


Fig. S1. KEGG functional categories found in Boza metagenome using HUMAnN 2.0.

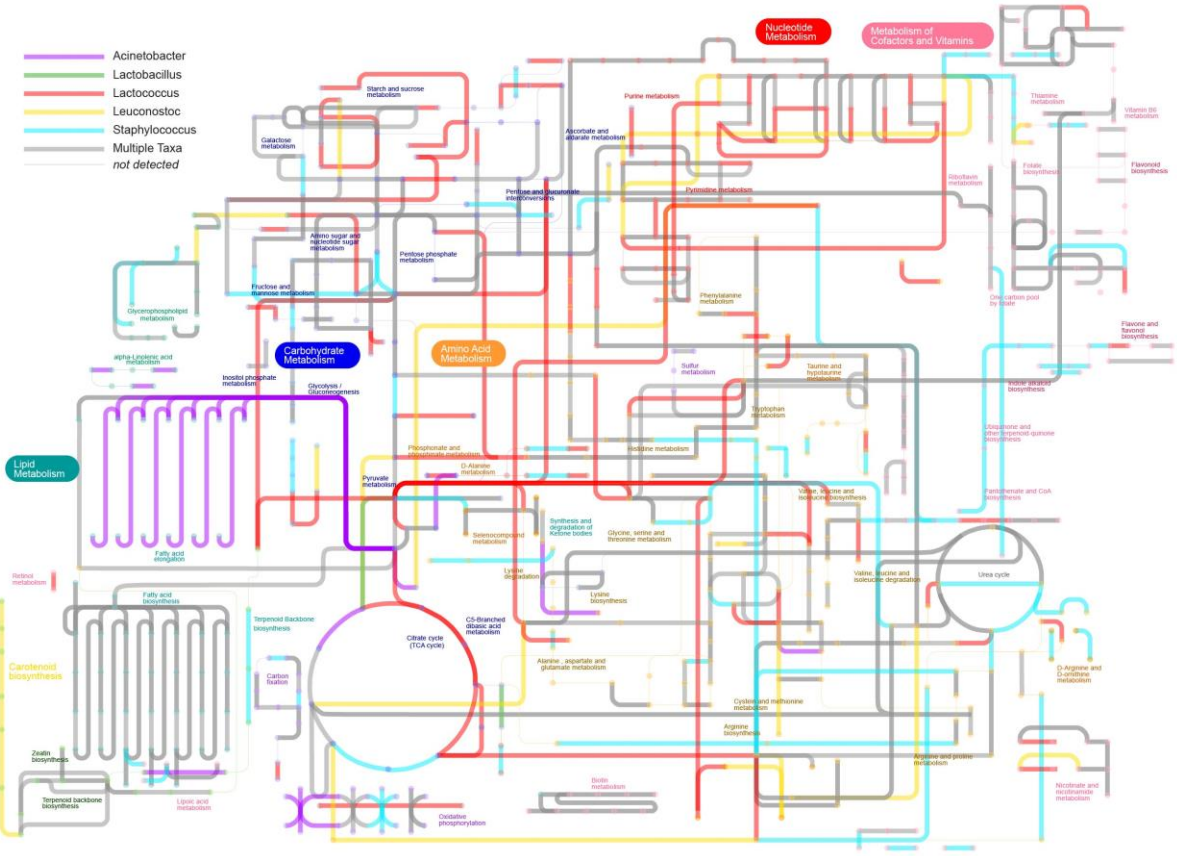


Fig. S2. Metabolic pathway of main bacteria genus found in boza metagenome.

Supplementary Tables

Table S1. Metagenome-assembled genomes (MAGs) from boza sample.

Bins	Taxon	Length (bp)	# contigs	N50	% GC	% completion	% redundancy
BZ1-01	*Leuconostoc*	1874001	102	93718	43.14	100	4.225352
BZ1-02	*Lactococcus*	2050569	418	10142	36.18	94.36	5.633803