

LAÍS SILVA BATALHA

**DIVERSIDADE GENÔMICA, CLASSIFICAÇÃO TAXONÔMICA E DETECÇÃO DE
FAGOS QUE INFECTAM BACTÉRIAS LÁCTEAS *Streptococcus thermophilus***

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de *Doctor Scientiae*.

Orientadora: Monique Renon Eller

Coorientadores: Leandro Licursi de Oliveira
Pedro Marcus Pereira Vidigal

**VIÇOSA - MINAS GERAIS
2022**

Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa

T

B328d
2022 Batalha, Laís Silva, 1991-
Diversidade genômica, classificação taxonômica e detecção
de fagos que infectam bactérias lácteas *Streptococcus
thermophilus* / Laís Silva Batalha. – Viçosa, MG, 2022.
1 tese eletrônica (122 f.); il. (algumas color.).

Orientador: Monique Renon Eller.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Tecnologia de Alimentos, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.645>

Modo de acesso: World Wide Web.

1. Bacteriófagos. 2. Biodiversidade. 3. Reação em cadeia da
polimerase. 4. Imunologia. 5. Polímeros. I. Eller, Monique
Renon, 1986-. II. Universidade Federal de Viçosa. Departamento
de Tecnologia de Alimentos. Programa de Pós-Graduação em
Ciência e Tecnologia de Alimentos. III. Título.

CDD 22. ed. 579.26

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

LAÍS SILVA BATALHA

**DIVERSIDADE GENÔMICA, CLASSIFICAÇÃO TAXONÔMICA E DETECÇÃO DE
FAGOS QUE INFECTAM BACTÉRIAS LÁCTEAS *Streptococcus thermophilus***

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência e Tecnologia de Alimentos, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de agosto de 2022.

Assentimento:



Laís Silva Batalha
Autora



Monique Renon Eller
Orientadora

*A Deus,
meus pais, aos meus irmãos e ao Haroldo
pelo amor, apoio e momentos felizes.
À Pitucha, Puck, Phiphi (in memoriam) e à Pops,*
DEDICO

AGRADECIMENTOS

Agradeço a Deus pela vida, fé e por nunca me deixar sozinha nas horas mais difíceis.

Agradeço aos meus pais, Antônio e Eloisa, pelo amor incondicional, educação, respeito, confiança, apoio e incentivo em todos os momentos da minha vida.

Agradeço aos meus queridos irmãos, Laura e Luide, pelo carinho, amizade e companheirismo em todos os momentos.

Agradeço ao Haroldo por todo apoio, amor, amizade e companheirismo.

Agradeço à Universidade Federal de Viçosa pela oportunidade de cursar minha Graduação, Mestrado e Doutorado.

Agradeço ao Pedro pela imensa contribuição neste trabalho e, principalmente, por democratizar o ensino da bioinformática.

Agradeço ao Marco Túlio pela amizade e colaboração neste trabalho.

Agradeço aos colegas de laboratório Beth, Talita, Danilo e Matheus, pelos momentos de descontração. Em especial à Jeniffer, pela amizade e ajuda na condução dos experimentos.

Agradeço ao prof. Alessio Giacomini pela orientação durante meu Doutorado Sanduíche na Itália.

Agradeço à prof. Monique pela orientação.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos. Este projeto de pesquisa foi financiado pelo CNPq (Processo 409972-2018-5).

Agradeço ao Núcleo de Análise de Biomoléculas (NuBioMol) da Universidade Federal de Viçosa (UFV) pelo suporte na análise dos dados. O NuBioMol é financiado pelas seguintes agências de fomento: Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Fapemig), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Financiadora de Estudos e Projetos (Finep) e Sistema Nacional de Laboratórios em Nanotecnologias (SisNANO)/Ministério da Ciência, Tecnologia e Informação (MCTI).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

A todos aqueles que de alguma forma contribuíram para a minha formação, MUITO OBRIGADA!

“Eu não vim até aqui, pra desistir agora”.
(Engenheiros do Hawaii)

RESUMO

BATALHA, Laís Silva, D.Sc., Universidade Federal de Viçosa, agosto de 2022. **Diversidade genômica, classificação taxonômica e detecção de fagos que infectam bactérias lácteas *Streptococcus thermophilus*.** Orientadora: Monique Renon Eller. Coorientadores: Leandro Licursi de Oliveira e Pedro Marcus Pereira Vidigal.

Bacteriófagos ameaçam constantemente as culturas *starter* de *Streptococcus thermophilus* usadas para a produção de queijos e iogurtes nas indústrias de laticínios. Dessa forma, o desenvolvimento de estratégias eficazes de prevenção e combate à lise por fagos estreptocócicos é fundamental neste campo. Para esse fim, é necessária a caracterização de fagos que infectam *S. thermophilus*, compreensão da dinâmica de interação com os hospedeiros, expansão do conhecimento da evolução desses fagos e melhoramento dos métodos de detecção. Neste estudo, usamos uma ampla gama de métodos complementares, incluindo genômica comparativa, análise do genoma central e filogenia de genes de assinatura, para mostrar que os fagos que infectam *S. thermophilus* estão organizados em 142 espécies e cinco gêneros (três deles novos), e que devido à sua genética diversidade, a classificação em nível de família varia de acordo com os critérios de classificação utilizados. Também desenvolvemos um método de PCR multiplex para detectar, em uma única reação, fagos estreptocócicos. Com base nos resultados da reação de PCR multiplex e na conservação dos genomas, os pares de *primers* desenhados neste estudo foram capazes de detectar 53 (85,48%) dos 62 fagos disponíveis até 2018, destacando-se como uma ferramenta útil para a detecção rápida da maioria dos fagos que infectam *S. thermophilus*. Além disso, iniciamos o desenvolvimento de imunoenaios para detecção de fagos que infectam *S. thermophilus*, sendo um deles de fluxo lateral e o outro baseado em vesículas de polidiacetileno. Este estudo permitiu a ampliação do conhecimento sobre a diversidade genética e evolução dos fagos estreptocócicos, além do desenvolvimento de ferramentas para a detecção desses vírus baseadas em métodos moleculares e imunológicos, fundamentais para promover estratégias de controle e minimizar falhas nos processos de fermentação do leite.

Palavras-chave: Biodiversidade. Cultura *starter*. Imunologia. PCR multiplex. Polidiacetileno.

ABSTRACT

BATALHA, Laís Silva, D.Sc., Universidade Federal de Viçosa, August, 2022. **Genomic diversity, taxonomic classification, and detection of phages infecting the dairy bacterium *Streptococcus thermophilus***. Adviser: Monique Renon Eller. Co-advisers: Leandro Licursi de Oliveira and Pedro Marcus Pereira Vidigal.

Bacteriophages constantly threaten starter cultures of *Streptococcus thermophilus* used for cheese and yogurt production in the dairy industry. Thus, the development of effective strategies to prevent and combat lysis by streptococcal phages is fundamental in this field. To this end, it is necessary to characterize the phages that infect *S. thermophilus*, understand the dynamics of interaction with the hosts, expand the knowledge of the evolution of these phages and improve detection methods. In this study, we used a wide range of complementary methods including comparative genomics, core genome analysis, and signature genes phylogenetics, to show that the *S. thermophilus* phages are organized into 142 species and five genera (three of them new), and that due to their genetic diversity, the classification at family level varies according to the classification criteria used. We also developed a multiplex PCR method to detect, in a single reaction, streptococcal phages. Based on the results of the multiplex PCR reaction and the conservation of genomes, the primer pairs designed in this study were able to detect 53 (85.48%) of the 62 phages available up to 2018, standing out as a useful tool for rapid detection of most phages infecting *S. thermophilus*. In addition, we started the development of immunoassays to detect phages that infect *S. thermophilus*, one of which is lateral flow and the other is based on polydiacetylene vesicles. This study allowed the expansion of knowledge about the genetic diversity and evolution of streptococcal phages, in addition to the development of tools for the detection of these phages based on molecular and immunological methods, fundamental to promote control strategies and minimize failures in milk fermentation processes.

Keywords: Biodiversity. Starter culture. Immunology. Multiplex PCR. Polydiacetylene.

SUMÁRIO

1. INTRODUÇÃO GERAL.....	9
2. CAPÍTULO I. Expanding the genomes insight of <i>Streptococcus thermophilus</i> phages through multifaceted approach	12
3. CAPÍTULO II. Desenvolvimento de PCR multiplex para detecção de fagos que infectam <i>Streptococcus thermophilus</i>	83
4. CAPÍTULO III. Desenvolvimento de imunoenaios para detecção de fagos que infectam <i>Streptococcus thermophilus</i>	101
5. CONCLUSÕES GERAIS	122

INTRODUÇÃO GERAL

Os bacteriófagos (fagos) são um problema persistente nas indústrias de laticínios pois ameaçam o desempenho das culturas *starter* de *Streptococcus thermophilus*, uma bactéria termofílica produtora de ácido láctico (BAL) e fundamental na fabricação de produtos lácteos, como queijos e iogurtes (Calasso et al., 2016; Marco et al., 2017). As consequências da infecção dessa bactéria por fagos geram problemas tecnológicos e de segurança, que podem variar entre fermentação lenta, redução da qualidade do produto final ou falha total do processo (Briggiler Marcó & Mercanti, 2021; Mahony & van Sinderen, 2015).

Os fagos estreptocócicos compartilham características comuns, como o hospedeiro *S. thermophilus*, genoma com estrutura modular, sintenia de funções gênicas e mosaicismos genéticos. No entanto, a caracterização constante de isolados, sequenciamento e análise comparativa de genomas mostrou uma diversidade genômica impressionante entre esses fagos (Lavelle et al., 2018; McDonnell et al., 2016, 2017; Szymczak et al., 2017, 2019). Do ponto de vista industrial, essa diversificação representa uma ameaça constante aos processos de fermentação do leite. Eventos de recombinação entre fagos que infectam diferentes espécies bacterianas podem resultar na extensão da gama de hospedeiros e dificultar o controle eficiente de infecções de fagos em culturas *starter*, que normalmente consistem em misturas de bactérias lácticas (de Melo et al., 2018; Leroy & De Vuyst, 2004; McDonnell et al., 2016; Quiberoni et al., 2010; Szymczak et al., 2017).

Dessa forma, o desenvolvimento de estratégias eficazes de prevenção e combate à lise por fagos que infectam *S. thermophilus* é fundamental nas indústrias de laticínios. A detecção e o monitoramento de fagos no leite, nos coprodutos lácteos e no ambiente industrial representam uma oportunidade para que sejam tomadas medidas de controle antes que ocorram prejuízos à qualidade dos produtos fermentados, financeiros e, eventualmente, à saúde do consumidor (Lavelle et al., 2018; McDonnell et al., 2017; Philippe & Moineau, 2021; Szymczak et al., 2017). Para esse fim, é necessária a caracterização de fagos que infectam *S. thermophilus*, compreensão da dinâmica de interação com os hospedeiros, a expansão do conhecimento da evolução desses fagos e o melhoramento dos métodos de detecção.

Assim, no primeiro capítulo, é descrito o estudo que levou à ampliação do conhecimento sobre a diversidade genética e a evolução dos fagos estreptocócicos por meio do uso de uma ampla gama de métodos complementares, incluindo genômica comparativa, análise do genoma central e análise filogenética de genes assinatura, mostrando como os fagos de *S. thermophilus* estão organizados.

No segundo capítulo, é descrito o desenvolvimento de um método de PCR multiplex para detectar, em uma única reação, fagos estreptocócicos. Essa reação mostrou ser uma ferramenta útil para a detecção rápida da maioria dos fagos que infectam *S. thermophilus*.

No terceiro capítulo, detalhamos sobre o desenvolvimento de imunoenaios para detecção de fagos que infectam *S. thermophilus*, sendo um deles de fluxo lateral e o outro baseado em vesículas de polidiacetileno (PDA). O imunoensaio de fluxo lateral não foi efetivamente testado em função da baixa fluidez do conjugado através da membrana de nitrocelulose. Já o imunoensaio baseado em vesículas de PDA foi efetivo para a detecção de fagos que infectam *S. thermophilus*. Se otimizados, os protótipos de imunoenaios desenvolvidos podem se tornar uma alternativa para a detecção de fagos estreptocócicos na indústria de laticínios.

REFERÊNCIAS BIBLIOGRÁFICAS

- Briggiler Marcó, M., & Mercanti, D. J. (2021). Bacteriophages in dairy plants. Em *Advances in Food and Nutrition Research* (Vol. 97, p. 1–54). Elsevier.
<https://doi.org/10.1016/bs.afnr.2021.02.015>
- Calasso, M., Ercolini, D., Mancini, L., Stellato, G., Minervini, F., Di Cagno, R., De Angelis, M., & Gobbetti, M. (2016). Relationships among house, rind and core microbiotas during manufacture of traditional Italian cheeses at the same dairy plant. *Food Microbiology*, 54, 115–126. <https://doi.org/10.1016/j.fm.2015.10.008>
- de Melo, A. G., Levesque, S., & Moineau, S. (2018). Phages as friends and enemies in food processing. *Current Opinion in Biotechnology*, 49, 185–190.
<https://doi.org/10.1016/j.copbio.2017.09.004>
- Lavelle, K., Martinez, I., Neve, H., Lugli, G., Franz, C., Ventura, M., Bello, F., Sinderen, D., & Mahony, J. (2018). Biodiversity of *Streptococcus thermophilus* Phages in Global Dairy Fermentations. *Viruses*, 10(10), 577. <https://doi.org/10.3390/v10100577>
- Leroy, F., & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology*, 15(2), 67–78.
<https://doi.org/10.1016/j.tifs.2003.09.004>
- Mahony, J., & van Sinderen, D. (2015). Novel strategies to prevent or exploit phages in fermentations, insights from phage–host interactions. *Current Opinion in Biotechnology*, 32, 8–13. <https://doi.org/10.1016/j.copbio.2014.09.006>
- Marco, M. L., Heeney, D., Binda, S., Cifelli, C. J., Cotter, P. D., Foligné, B., Gänzle, M., Kort, R., Pasin, G., Pihlanto, A., Smid, E. J., & Hutkins, R. (2017). Health benefits of fermented foods: Microbiota and beyond. *Current Opinion in Biotechnology*, 44, 94–102. <https://doi.org/10.1016/j.copbio.2016.11.010>
- McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.-P., Lugli, G. A., Ventura, M., Kouwen, T. R., & van Sinderen, D. (2017). Global Survey and Genome Exploration of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus*

- thermophilus*. *Frontiers in Microbiology*, 8, 1754.
<https://doi.org/10.3389/fmicb.2017.01754>
- McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.-P., Kouwen, T., & van Sinderen, D. (2016). Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, 82(17), 5153–5165.
<https://doi.org/10.1128/AEM.00835-16>
- Philippe, C., & Moineau, S. (2021). The endless battle between phages and CRISPR-Cas systems in *Streptococcus thermophilus*. *Biochemistry and Cell Biology*, bcb-2020-0593. <https://doi.org/10.1139/bcb-2020-0593>
- Quiberoni, A., Moineau, S., Rousseau, G. M., Reinheimer, J., & Ackermann, H.-W. (2010). *Streptococcus thermophilus* bacteriophages. *International Dairy Journal*, 20(10), 657–664. <https://doi.org/10.1016/j.idairyj.2010.03.012>
- Szymczak, P., Janzen, T., Neves, A. R., Kot, W., Hansen, L. H., Lametsch, R., Neve, H., Franz, C. M. A. P., & Vogensen, F. K. (2017). Novel Variants of *Streptococcus thermophilus* Bacteriophages Are Indicative of Genetic Recombination among Phages from Different Bacterial Species. *Applied and Environmental Microbiology*, 83(5), e02748-16, e02748-16. <https://doi.org/10.1128/AEM.02748-16>
- Szymczak, P., Rau, M. H., Monteiro, J. M., Pinho, M. G., Filipe, S. R., Vogensen, F. K., Zeidan, A. A., & Janzen, T. (2019). A comparative genomics approach for identifying host-range determinants in *Streptococcus thermophilus* bacteriophages. *Scientific Reports*, 9(1), 7991. <https://doi.org/10.1038/s41598-019-44481-z>

CAPÍTULO I

Expanding the genomes insight of *Streptococcus thermophilus* phages through multifaceted approach

Artigo publicado na Research, Society and Development: Batalha, LS, Vidigal PMP, Gontijo MTP, Eller MR. 2022. Expanding the genomes insight of *Streptococcus thermophilus* phages through a multifaceted approach. Research, Society and Development. v. 11, n. 10, p. e273111032693. <http://dx.doi.org/10.33448/rsd-v11i10.32693>

ABSTRACT

Viruses have complex evolutionary relationships and several strategies have been used in an attempt to classify the phages that infect *S. thermophilus*. In this study, we used a wide range of complementary methods including comparative genomics, core genome analysis, and signature genes phylogenetics, to show that the *S. thermophilus* phages are organized into 142 species and five genera (three of them new), and that due to their genetic diversity, the classification at family level varies according to the classification criteria used. No significantly conserved genes were identified among the 183 genomes evaluated, however, the genes encoding holin protein were conserved in more than 95% of genomes. The holins analysis suggests that at least two α -helix are required for protein function within *S. thermophilus* phages. This study allowed the expansion of knowledge about the genetic diversity and evolution of streptococcal phages, both fundamental to promote control strategies and minimize failures in milk fermentation processes.

Keywords: Biodiversity, core genome, holin, panvirome, taxonomy, signature genes, starter culture.

1. INTRODUCTION

Advances in sequencing technologies and bioinformatics tools applied to metagenomic studies have contributed to the continued expansion of genome databases and the understanding of the genetic diversity of bacteriophages (phages) (Dion et al., 2020). This knowledge is particularly important for the phages that infect *Streptococcus thermophilus*, a thermophilic starter culture fundamental in the manufacture of dairy products such as cheese and yogurt (de Melo et al., 2018). Infection of starter cultures by phages during food fermentation, particularly

in dairy fermentation, can lead to technological, economic, and environmental problems due to delays in fermentation or incomplete fermentation of milk, change in product quality, reduced productivity and disposal of milk, in cases of total process failure (de Melo et al., 2018; Mahony et al., 2020; Pujato et al., 2019).

Streptococcal phages share common features such as the *S. thermophilus* host, genome with modular structure, synteny of gene functions, and genetic mosaicism. However, the constant characterization of isolates, sequencing, and comparative analysis of genomes has shown an impressive genomic diversity among these phages (Lavelle, Martinez, et al., 2018; McDonnell et al., 2016, 2017; Szymczak et al., 2017; Szymczak, Rau, et al., 2019). The expansion of viral diversity at the genomic level is further highlighted by the discovery of P738 group phages (Philippe et al., 2020), in addition to the other four existing groups: *Moineauvirus* (previously called *cos*-containing phages), *Brussowvirus* (previously called *pac*-containing phages), 5093 and 987 (Le Marrec et al., 1997; McDonnell et al., 2016; Mills et al., 2011; Tremblay & Moineau, 1999). Many of these phages have genes originating from non-dairy environments, further highlighting their complex evolutionary relationships (McDonnell et al., 2016; Mills et al., 2011). From an industrial point of view, this diversification represents a constant threat to milk fermentation processes. Recombination events between phages that infect different bacterial species can result in the extension of the host range and hamper the efficient control of phage infections in starter cultures, which normally consist of mixtures of lactic acid bacteria (LAB) (de Melo et al., 2018; Leroy & De Vuyst, 2004; McDonnell et al., 2016; Quiberoni et al., 2010; Szymczak et al., 2017).

The phage genomes that infect *S. thermophilus* are under a constant process of evolution due to the selective pressure of their hosts, both in response to natural mechanisms of resistance to phage infection, as well as by industrialization and the use of commercial starter cultures. The evolution mechanisms of these phages involve horizontal gene transfer, insertions, deletions, point mutations, and exchange of modules, which promote their genetic diversification and complex evolutionary relationships that do not follow the traditional hierarchical phylogeny (Brussow & Desiere, 2001; Desiere et al., 1998; Lavelle, Martinez, et al., 2018; Lavelle, Murphy, et al., 2018; Lucchini et al., 1999a; Philippe et al., 2020; Szymczak et al., 2017). To encompass the full range of genetic diversity, the taxonomy of streptococcal phages must be updated to provide a stable and standardized classification framework, but dynamic, capable of accommodating revisions and reinterpretations of the perceived relationships between these phages as knowledge about new viruses and their genomes advances.

Thus, our classification strategy has included a variety of classification tools that employ very different approaches. Our analyzes ranged from clustering methods based on whole proteome similarity, as for example ViPTree (Nishimura et al., 2017) and VICTOR (Meier-Kolthoff & Göker, 2017), analysis of panvirome concerning gene distribution with Roary (Page et al., 2015), intergenomic similarity using VIRIDIC (Moraru et al., 2020), and detection and phylogeny of signature genes performed with CoreGenes (Zafar et al., 2002) and phylogeny.fr (Dereeper et al., 2008). This multifaceted approach allowed us to gradually descend from clustering in family level to intra-familial relationships. Despite the diversity of applied methods, their results proved to be complementary and predominantly in agreement. All methods converged in robust results from five genera of *S. thermophilus* phages (three of them new), distributed in 142 species. With this study, we allowed the expansion of the number of *S. thermophilus* phages genera and species available in the ICTV database, making sense of the genetic diversity existing among them.

2. MATERIAL AND METHODS

2.1. Genome database

Sequences of phage genomes that infect *S. thermophilus* available in the GenBank/NCBI database (<https://www.ncbi.nlm.nih.gov/genbank>) (Benson et al., 2012) were retrieved in March 2022 along with other associated information. These data were complemented with bibliographic references linked to the genomes and organized in Supplementary Information (SI). Taxonomic classifications of *S. thermophilus* phages were retrieved from the International Committee on Taxonomy of Viruses (ICTV) (<https://talk.ictvonline.org/taxonomy/>). The lifestyle of the phages based on the conserved protein domains was predicted using BACPHLIP (BACterioPHage LIifestyle Predictor) (Hockenberry & Wilke, 2021).

2.2. Panvirome analysis and search for core genome

The panvirome of phages that infect *S. thermophilus* was analyzed using Roary 3.13.0 (Page et al., 2015). For this, the phage genomes were reannotated using PROKKA 1.14.5 (Seemann, 2014), selecting the annotation mode for viral genomes (--kingdom Viruses), and the GFF (General Feature Format) files were used as input by the Roary. Protein sequences annotated on the phage genomes were compared and grouped using a minimum percentage identity of 40% (-i 40). After grouping, the binary matrix of presence/absence of genes that are part of the accessory genome was used to group the phages. This matrix corresponds to the

distribution of genes that encode proteins that were shared between the analyzed genomes. The accessory genome was considered because no genes were included in the core genome. The dendrogram with the grouping of the binary matrix was visualized using FigTree v1.4.4. (Rambaut, 2012). CoreGenes 5.0 (Zafar et al., 2002) was also used to identify the core genome.

2.3. Genomic organization of phages

Gene synteny and conservation of genomic structure of phages that infect *S. thermophilus* were evaluated using Clinker 0.0.21 (Gilchrist & Chooi, 2020). A reference genome was selected, including phages named: Streptococcus phage DT1 (NC_002072), Streptococcus phage O1205 (NC_004303), Streptococcus phage 5093 (NC_012753), Streptococcus virus 9871 (KU678389), and Streptococcus phage P738 (MK911750). These phages were chosen because they were used as a model for the group in several studies (Deveau et al., 2008; Levesque et al., 2005; McDonnell et al., 2017; Mills et al., 2011; Tremblay & Moineau, 1999) or because they were the first representatives of the group (Le Marrec et al., 1997; Lucchini et al., 1999b; McDonnell et al., 2016; Philippe et al., 2020). The products of the predicted ORFs (Open Reading Frames) for each of the phages were manually revised based on the genome annotations available in GenBank/NCBI.

2.4. Protein analysis

To evaluate the robustness of the Roary and CoreGenes analysis in detecting orthologous proteins in the panvirome, a multiple alignment of holins sequence of the streptococcal phages was performed using MUSCLE 3.8.31 (Edgar, 2004). All holin sequences of the 183 phages were included. In addition, enzymatic catalytic domains (ECDs) of holin families were predicted using HMMER software v3.3.2 (Finn et al., 2011). The structure of representative holins was predicted using RoseTTAFold modeling (Baek et al., 2021). The structures were then compared using FATCAT (Flexible structure Alignment by Chaining Aligned fragment pairs allowing Twists) (<https://fatcat.godziklab.org/>) (Li et al., 2020) and visualized using Mol* Viewer (<https://molstar.org/>) (Sehnal et al., 2021).

2.5. Taxonomic assignment

To determine the family-level classification of the *S. thermophilus* phages, we used (i) the main predicted proteome-based clustering tools: VICTOR (Virus Classification and Tree building Online Resource) (<https://ggdc.dsmz.de/victor.php>) (Meier-Kolthoff & Göker, 2017), VirClust (Virus Clusterer) (<https://rhea.icbm.uni-oldenburg.de/VIRCLUST/>) (Moraru, 2021)

and ViPTree (Viral Proteomic Tree) (<https://www.genome.jp/viptree/>) (Nishimura et al., 2017) and we calculated (ii) shared orthologous genes. On the VICTOR analysis, all pairwise comparisons of the amino acid sequences were conducted using the Genome-BLAST Distance Phylogeny (GBDP) method (Meier-Kolthoff et al., 2013) under settings recommended for prokaryotic viruses (Meier-Kolthoff & Göker, 2017). The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FastME including SPR postprocessing (Lefort et al., 2015) for D6 intergenomic distance formula. Branch support was inferred from 100 pseudo-bootstrap replicates. Trees were rooted at the midpoint (Farris, 1972) and visualized with iTOL (Letunic & Bork, 2021). Taxon boundaries at the species, genus and family level were estimated with the OPTSIL program (Göker et al., 2009) and VICTOR used the following clustering thresholds to suggest taxon boundaries at genus (0.749680), subfamily (0.888940) and family (0.985225) level (Meier-Kolthoff & Göker, 2017). The parameters for VirClust were: (i) protein clustering based on “evalue”, after reciprocal BLASTp hits were removed if e-value >0.0001 and bitscore <50; (ii) hierarchical clustering based on protein clusters (PCs) with a bootstrapping of 1000 replicates. The resulting tree was split into viral genome clusters (VGCs) using an 0.98, 0.90 or 0.85 intergenomic distance threshold to determine the family-level classification of the *S. thermophilus* phages (Moraru, 2021). On the ViPTree analysis, a proteomic tree was constructed from the phage genomes based on genome-wide sequence similarities computed by tBLASTx, using dsDNA nucleic acid type and prokaryote host category. The orthologous genes identification was performed with VirClust, based on PCs calculated with the above parameters, CoreGenes and Roary.

In addition to the amino acid-based VICTOR analysis, the intra-familial relationships were analyzed using (i) nucleic acid-based intergenomic similarities calculated with VIRIDIC (Virus Intergenomic Distance Calculator) (<http://rhea.icbm.uni-oldenburg.de/VIRIDIC/>) (Moraru et al., 2020) and (ii) core protein phylogeny. The thresholds used for species and genus definition in VIRIDIC were 95% and 70% intergenomic similarity, respectively. The core proteins analysis was conducted as follows: (i) core genes were detected and annotated with Roary, VirClust and CoreGenes; (ii) a multiple alignment of core protein sequences was constructed with MUSCLE, and (iii) the phylogenetic tree was constructed with phylogeny.fr in “one click” mode (Dereeper et al., 2008). Phylogenetic trees of signature genes were rooted using a more distant relative (outgroup) and were accompanied by bootstrap values.

3. RESULTS

3.1. General characteristics of phages that infect *S. thermophilus*

3.1.1. Distribution and current taxonomy of *S. thermophilus* phages

As of March 2022, 183 complete phage genomes that infect *S. thermophilus* were available in the GenBank database (SI file 1 Table S1). According to the literature, the phages that infect *S. thermophilus* are distributed into five groups: *Moineauvirus* (59.56%, n = 109), *Brussowvirus* (24.59%, n = 45), 5093 (7.10%, n = 13), 987 (7.65%, n = 14) and more recently, the P738 group (1.09%, n = 2) (Achigar et al., 2017; Ali et al., 2014; Arioli et al., 2018; da Silva Duarte et al., 2018; Desiere et al., 1998; Deveau et al., 2008; Guglielmotti et al., 2009; Hynes et al., 2018; Lavelle, Martinez, et al., 2018; Lavelle, Murphy, et al., 2018; Le Marrec et al., 1997; Levesque et al., 2005; Lucchini et al., 1998; McDonnell et al., 2016, 2017; Mills et al., 2011; Neve et al., 1998; Philippe et al., 2020; Somerville et al., 2019; Stanley et al., 1997; Szymczak et al., 2017; Szymczak, Rau, et al., 2019; Szymczak, Vogensen, et al., 2019; Tremblay & Moineau, 1999).

At the moment, the ICTV recognizes the existence of 10 species of phages that infect *S. thermophilus*, which are classified into two genera: *Moineauvirus* (*Moineauvirus Abc2*, *Moineauvirus DT1*, *Moineauvirus mv7201*, *Moineauvirus Sfi19*, *Moineauvirus Sfi21*) and *Brussowvirus* (*Brussowvirus ALQ132*, *Brussowvirus bv858*, *Brussowvirus bv2972*, *Brussowvirus bvO1205*, *Brussowvirus Sfi11*), all belonging to the *Caudoviricetes* class (<https://talk.ictvonline.org/taxonomy/>, 2021 release). This classification is incomplete and does not represent the genetic diversity among streptococcal phage populations, which requires that a new genome-based classification of the group be proposed along with updating the ICTV database.

3.1.2. Lifestyle of *S. thermophilus* phages

Based on the phage lifestyle assessment (SI file 1 Table S1, Figure 1), a total of 114 (64.05%) and 64 (35.95%) phages were identified as virulent and temperate, respectively, with >50% probability. Five phage genomes from the *Moineauvirus* genus showed undefined classification and 53.84% (n = 56) were classified as temperate. In contrast to the phages of the *Brussowvirus* genus and 5093, 987 and P738 groups, which were classified as virulent in 91.11% (n = 41), 92.30% (n = 12), 78.57% (n = 11) and 100% (n = 2) of the analyzed genomes, respectively, with >50% probability.

3.1.3. Isolation and sampling sources

The main sources of isolation of phages that infect *S. thermophilus* were cheese whey (n = 106) and cheese (n = 51), corresponding together to almost 86% of the total phages isolated from fermentation processes. Both virulent and temperate phages were mostly isolated from these sources, corresponding to 85.96% (n = 98) and 84.37% (n = 54), respectively. Italy (n = 48), France (n = 39) and Ireland (n = 17) are the countries where the most phages were isolated, characterized, and sequenced (SI file 1 Table S1).

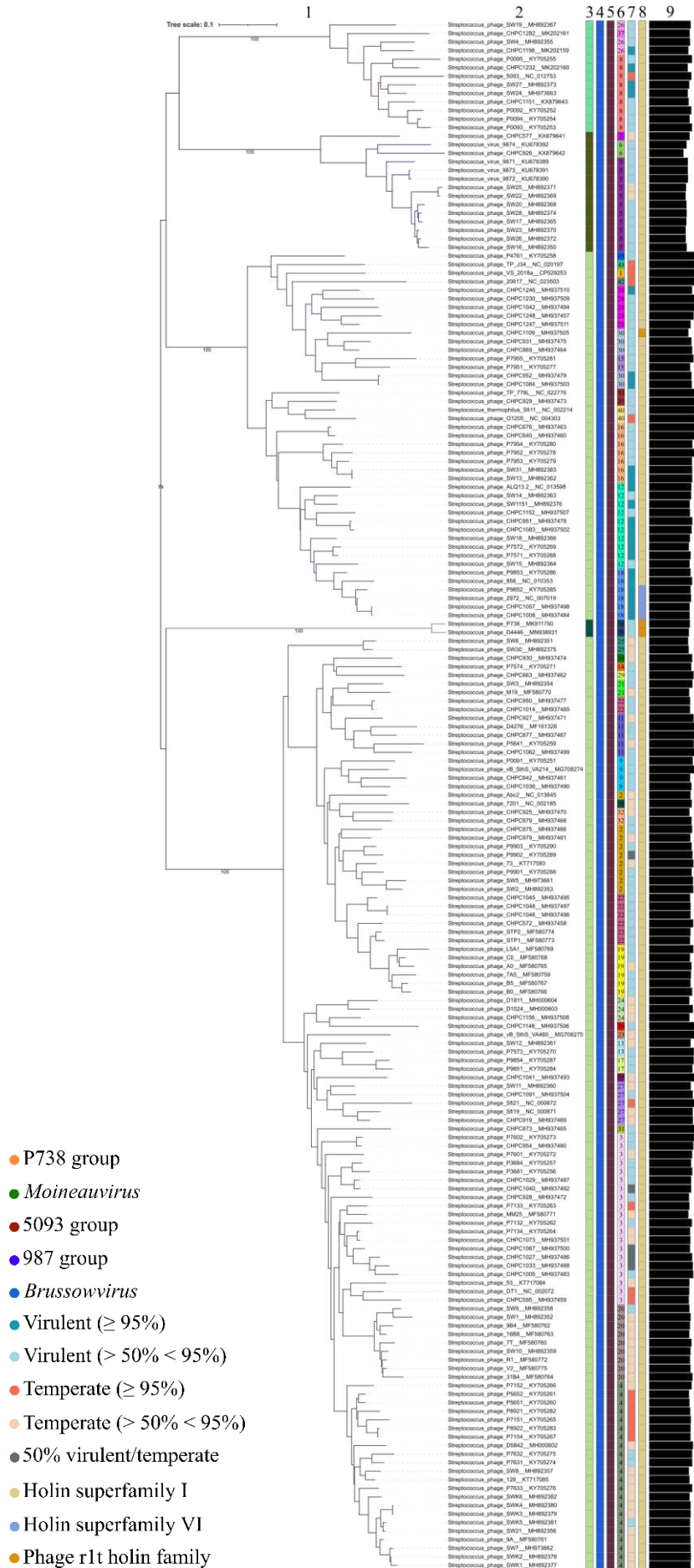


Figure 1 - Phylogenomic tree generated by VICTOR using the amino-acid sequences of the *Streptococcus thermophilus* phages. Proteome-based tree (1), phage name (2), VICTOR genus cluster (3), VICTOR subfamily cluster (4), VICTOR family cluster (5), VIRIDIC genus cluster (6), BACPHLIP lifestyle prediction (7), phage holins classification (8), amino-acids sequence length - Min. (8,861), Max. (14,848) (9). See SI file 3 for high resolution version. During step 'check_and_map_input', the VICTOR analysis skipped duplicate genomes files 'Streptococcus_phage_SW32_MH892384' and 'Streptococcus_phage_SW33_MH892385' (duplicates of 'Streptococcus_phage_SW13_MH892362'). Although these phages are identical, they were isolated in three different countries from different samples (SI file 1 Table S1).

3.1.4. Phage morphology and host recognition

All phages that infect *S. thermophilus* are to the siphovirus morphotype and have an icosahedral capsid connected to a non-contractile tail (Lavelle, Martinez, et al., 2018; Mahony & van Sinderen, 2014; McDonnell et al., 2017; Philippe et al., 2020). They have a limited host range, as most isolated phages only infect their primary host and a small number infect between two and 14 strains of *S. thermophilus* (Binetti et al., 2005; Lavelle, Martinez, et al., 2018; McDonnell et al., 2017; Philippe et al., 2020; Zinno et al., 2010). Each group of phages has individual characteristics that reflect the genetic content and morphological characteristics of these agents, including host recognition structures at the tip of the tail. Phages from the *Moineauvirus* and *Brussowvirus* genera are morphologically similar, displaying long tails and often a terminal structure at the tip of the tail like a small plaque accompanied by a fiber (Accolas & Spillmann, 1979; Lavelle, Martinez, et al., 2018; Le Marrec et al., 1997). Group 5093 phages have globular appendages attached to the tips of the long tail, and group 987 phages have tails considerably shorter than those of the other phage groups, with a broad appendage at the tip of the tail that resembles that of the phages that infect *Lactococcus lactis* of group P335 (Accolas & Spillmann, 1979; Lucchini et al., 1998; McDonnell et al., 2016, 2017; Mills et al., 2011; Szymczak et al., 2017). On the other hand, the phages of the newly named group P738 have short tails similar to those of the group 987 phages and twisted tail fibers composed of two or three subfibers (Philippe et al., 2020).

3.1.5. Genome type and organization

Streptococcal phage genomes have a modular structure and synteny of gene functions (Le Marrec et al., 1997; McDonnell et al., 2016; Philippe et al., 2020; Szymczak et al., 2017; Tremblay & Moineau, 1999). They have double-stranded DNA (dsDNA) in linear form, and the size and GC content of the genomes range from 30 to 48 kb and 36.6 to 40.2%, respectively.

The number of predicted ORFs (Open Reading Frames) predicted in the phage genomes ranges from 38 to 68 (SI file 1 Table S1).

3.2. Genomic diversity of *S. thermophilus* phages

Panvirome analysis based on gene distribution divided the phages into five main clusters (SI file 2), which correspond to the phage groups already described (*Moineavirus*, *Brussowvirus*, 5093, 987, and P738). In total, 580 genes were identified in the panvirome analysis and, of these, almost 21% (121) are unique genes in the analyzed genomes, which shows the genetic heterogeneity existing among the phages that infect *S. thermophilus*. The panviromes of the *Moineavirus* and *Brussowvirus* genera have greater genetic diversity, with a total of 187 and 162 genes, respectively, and organized into several subclusters, which may be related to the greater number of representatives already characterized. The panviromes of the 5093, 987 and P738 groups have 100, 81 and 50 genes, respectively, but the analysis of the genetic diversity of the phages of these groups, in turn, is limited by the smaller number of sequenced genomes. No significantly conserved genes were identified among the 183 genomes, considering a minimum identity of 40%. Likewise, no core genome was identified with CoreGenes. However, two genes were present in 176 (96.17%) of the 183 analyzed genomes using Roary. Such genes are not present in the genomes of the two phages of the P738 group and in five phages of the *Brussowvirus* genus. One of these genes had its putative function identified and corresponded to the genes encoding holin protein.

Comparative analyzes of the genetic organization and genome content of the reference phages showed that the genomes of these phages share common features such as the modular structure, synteny of gene functions, and high genetic variation between nucleotide sequences (Figure 2). However, no gene product was conserved in the five genomes. The lysis module was the only partially conserved genomic region among the phages from *Moineavirus* and *Brussowvirus* genera, and 5093 and 987 groups phages. Notably, the reference phage of the P738 group was revealed to be genetically distinct from the others evaluated, sharing some gene products in a limited way with the reference phage of the 987 group. This result confirms the previous analysis that the lysis module is the only partially conserved genomic region among the phages from *Moineavirus* and *Brussowvirus* genera, and 5093 and 987 groups phages.

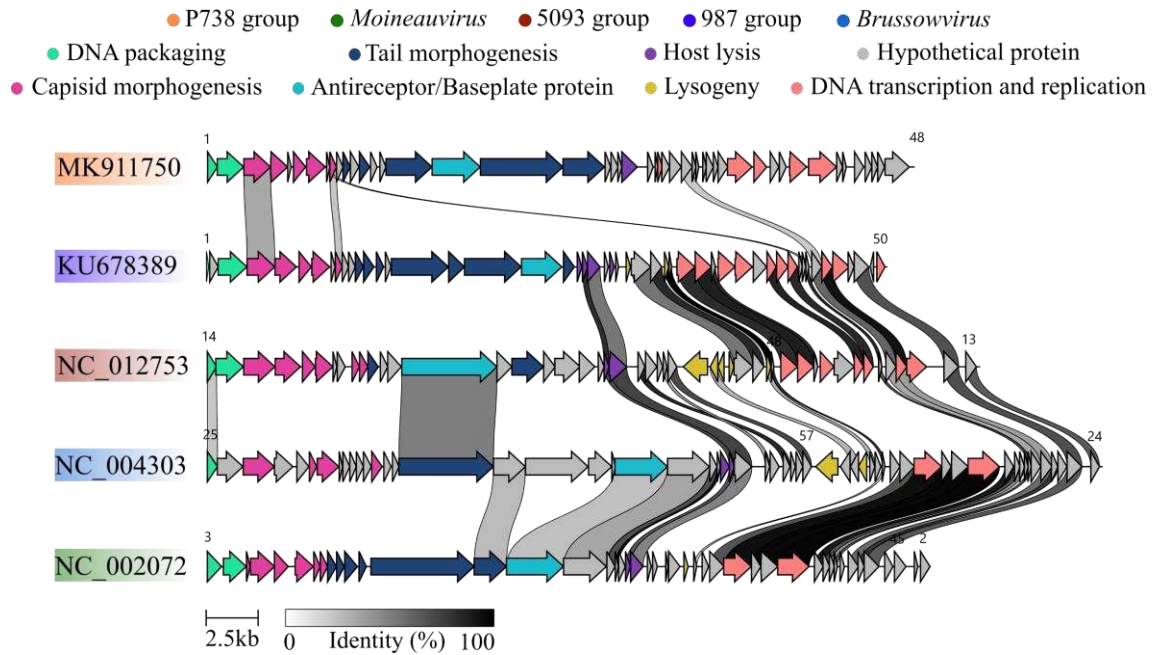


Figure 2 - Comparative analysis of the genetic organization and content of phage genomes that infect *Streptococcus thermophilus*. Predicted ORFs (indicated by arrows) and gene products (putative function indicated by color coding) are aligned with adjacent genomes according to percent amino acid identity (indicated by shaded boxes).

3.3. Characterization of *S. thermophilus* phages-encoded holins

Multiple alignment of holin sequences encoded by the 183 *S. thermophilus* phages revealed that this protein is conserved among the phages *Moineauvirus* genus and 5093 and 987 groups (SI file 4). Holin sequences are also conserved in phages of the *Brussowvirus* genus, with the exception of five of them (MH937505, KY705285, NC_007019, MH937498 and MH937484). The holins encoded by the two phages of the P738 group are conserved with each other, presenting, notably, sequence similarity to the holin encoded by the phage MH937505 (*Streptococcus* phage CHPC1109, *Brussowvirus* genus). The P738 group phages (MK911750.1_00022 and MN938931.1_00022) encode holins belonging to phage r1t holin family (PF16945), as well as one phage (MH937505.1_00021) from the *Brussowvirus* genus (Figure 1). Still in the *Brussowvirus* genus, four other phages (KY705285.1_00024, NC_007019.1_00025, MH937498.1_00026, and MH937484.1_00026) encode holins from holin superfamily VI (PF09682). The other bacteriophages (n = 176) encode holins from holin superfamily I (PF04531).

The two sequences of the holins encoded by the P738 group phages, as well as the five sequences of the *Brussowvirus* genus, were selected to verify the structural prediction to determine possible structural conservation despite low sequence similarity (Figure 3a).

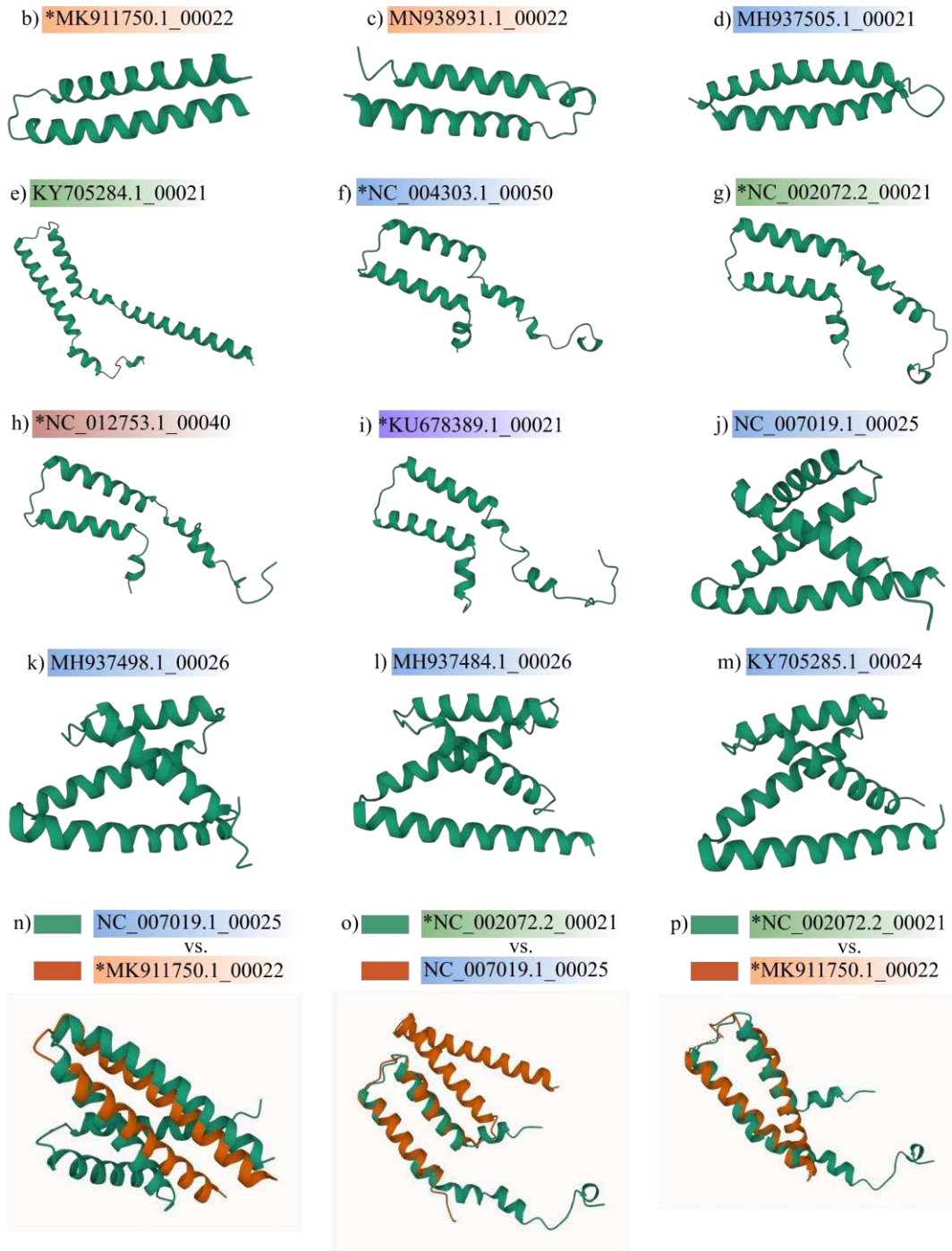
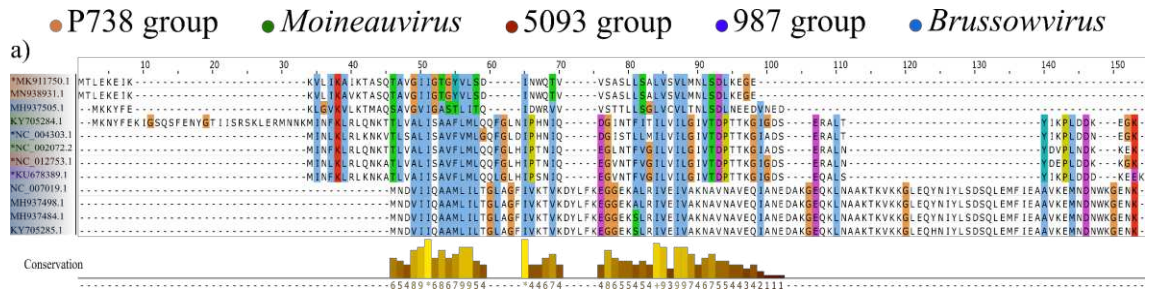


Figure 3 - Structural prediction of representative holins encoded by *Streptococcus thermophilus* phages. Multiple alignment of representative holin sequences (a). Phage r1t holin family (PF16945) from P738 group (b, c), and *Brussowvirus* genus (d). Holin superfamily I (PF04531) from *Moineauvirus* genus (e, g), *Brussowvirus* genus (f), 5093 group (h), and 987 group (i). Superfamily VI holins (PF09682) from *Brussowvirus* genus (j, k, l, m). Pairwise alignment of two α -helix in holins from superfamily VI and r1t holin family (n). Pairwise alignment of two α -helix in holins from superfamilies I and VI (o). Pairwise alignment of two α -helix in holins from superfamily I and r1t holin family (p).

Furthermore, these sequences were compared with the sequences of the holins encoded by the reference phages and representative holins from each one of the predicted protein families (PF16945, PF09682, and PF04531). Holins belonging to the phage r1t holin family (PF16945) have two α -helix (Figure 3b-d), while superfamily VI holins (PF09682) have four (Figure 3j-m). The structure of holins from holin superfamily I (PF04531) is conserved, with representatives in *Moineauvirus* and *Brussowvirus* genera, and 5093 and 987 groups phages (Figure 3e-i). These holins possess two α -helix, and the superfamily I holin encoded by some *Moineauvirus* genus also have an additional helix at N-terminus (Figure 3e). The angstrom error estimates for the predictions are shown in SI file 5 Figure S1.

The pairwise alignment of the predicted structures revealed a significant pairwise alignment (p -value $1.67e-05$) of two α -helix in holins from superfamily VI and r1t holin family (Figure 3n). Similar results are found for holins from superfamilies I and VI (p -value $8.13e-04$; Figure 3o) and holins from superfamily I and r1t holin family (p -value $1.42e-03$; Figure 3p). The graphic representation of the alignments and the differential distance matrix decomposition are shown in SI file 5 Figure S2. The conserved structural region in all the alignments is within the functional domains of each of the holin protein families, which suggests that at least two α -helix are required for protein function within *S. thermophilus* phages.

3.4. Taxonomic assignment

3.4.1. Support for the description of 142 species of streptococcal phages

The intergenomic similarity between pairs of the 183 phage genomes that infect *S. thermophilus*, as well as the information on the length ratio and aligned fraction of the genomes calculated by VIRIDIC, are presented in SI file 6. The evaluated phages were divided into five clusters, according to the phage groups already described in the literature. Additional divisions into subclusters (internal triangles) in the *Moineauvirus* and *Brussowvirus* genera, again highlight the genetic diversity among their members. VIRIDIC integrates the ICTV virus classification criteria based on nucleotide identity thresholds of 95% for species. The similarity

data of the genome pairs quantified by VIRIDIC suggest that the 183 phages that infect *S. thermophilus* are organized into 142 species. The two phages of the P738 group were considered two different species (SI file 5 Table S1, Figure S3). The 109 phages of the *Moineavirus* genus were organized into 86 species (SI file 5 Tables S2 and S3, Figure S4), while the 45 phages of the *Brussowvirus* genus were organized into 35 species (SI file 5 Tables S4 and S5, Figure S5). The 14 phages of the 987 group were distributed in seven species (SI file 5 Tables S6 and S7, Figure S6) and of the 13 phages of the 5093 group, 12 formed different species (SI file 5 Tables S8 and S9, Figure S7). With this study, we allowed the expansion of the number of *S. thermophilus* phage species available in the ICTV database from 10 to 142, including their taxonomic classification.

3.4.2. Support for the description of five genera of streptococcal phages

The ICTV has established 70% nucleotide identity of the full genome length as the cut-off for genera, in addition to the presence of homologous conserved ‘signature or core genes’ and evaluated using phylogenetics. Based on the proteome-based VICTOR analysis, the division into the five groups is also found in the VICTOR phylogeny (Figure 1). Even though VICTOR has recommended four genera, the light-green-colored one is not monophyletic and was split into two distinct genera as a consequence. That is, the subtree comprising phages "Streptococcus_phage_SW6__MH892351" to "Streptococcus_phage_SWK1__MH892377" formed the fifth genus.

Each genus identified in the VICTOR analysis was investigated for the presence of core genes. The results found by Roary, VirClust and CoreGenes analysis were similar (Table 1, SI file 1 Tables S2-S16). The P738 group phages shared at least 43 homologs proteins (91.49%). These homologs proteins included: small/large terminase subunits, portal protein, minor/major capsid proteins, head scaffolding protein, head-tail adaptor, major tail protein, tape measure protein, host-specificity tail protein, holin, endolysin, DNA helicase, and primase. The phages from this genus have genomes of, on average, 33.85 kb (37.05 %GC), encoding 47 proteins. At the DNA level, these phages share 90.71% DNA similarity (SI file 1 Tables S2-S4, SI file 5 Table S1, SI file 6). The *Moineavirus* phages shared at least 15 homologs proteins (33.33%). These homologs proteins included: small/large terminase subunits, portal protein, major capsid protein, head scaffolding protein, head-tail connector protein, major tail protein, antireceptor, and holin. The phages from this genus have genomes of, on average, 35.90 kb (38.74 %GC) encoding 45 proteins. At the DNA level, these phages share at least 44.35% DNA similarity (SI file 1 Tables S5-S7, SI file 5 Table S2, SI file 6).

Table 1 - Core genes with putative functions identified for each group/genus using the Roary, VirClust and CoreGenes tools

Genus (reference genome)	Roary information			VirClust information			CoreGenes information	
	Gene_ID	Protein length	BLAST_annot	Gene_ID	Protein length	PHROGS_annot	ID	annot
P738 (MK911 750)	gene_00001	159	terminase small subunit	gene_1	160	terminase small subunit	QDP43702.1	putative terminase small subunit
	gene_00002	425	terminase large subunit	gene_2	426	terminase large subunit	QDP43703.1	putative terminase large subunit
	gene_00003	430	portal protein	gene_3	431	portal protein	QDP43704.1	putative portal protein
	gene_00004	271	head protein	gene_4	272	minor head protein	QDP43705.1	putative head morphogenesis protein
	-	-	-	gene_5	55	no_hit	QDP43706.1	hypothetical protein
	gene_00006	195	scaffold protein	gene_6	196	head scaffolding protein	QDP43707.1	putative scaffold protein
	gene_00007	302	major head protein	gene_7	303	major head protein	QDP43708.1	major head protein
	gene_00008	50	hypothetical protein	gene_8	51	Arc-like repressor	QDP43709.1	hypothetical protein
	gene_00009	108	head-tail connector	gene_9	109	head-tail adaptor	QDP43710.1	putative head-tail connector
	gene_00010	103	hypothetical protein	gene_10	104	no_hit	QDP43711.1	hypothetical protein
	gene_00011	116	tail protein	gene_11	117	neck protein Ne1	QDP43712.1	putative tail protein
	gene_00012	136	hypothetical protein	gene_12	137	tail completion Tc1	QDP43713.1	hypothetical protein

gene_00013	165	major tail protein	gene_13	166	major tail protein	QDP43714.1	major tail protein
gene_00014	104	hypothetical protein	gene_14	105	no_hit	QDP43715.1	hypothetical protein
gene_00015	97	hypothetical protein	gene_15	98	no_hit	QDP43716.1	hypothetical protein
gene_00016	738	tape-measure protein	gene_16	739	tail length tape measure protein	QDP43717.1	putative tape-measure protein
gene_00017	773	host-specificity tail protein	gene_17	774	minor tail protein	QDP43718.1	putative host-specificity tail protein
gene_00018	1323	tail fiber protein	gene_18	1324	tail fiber protein and host specificity	QDP43719.1	putative tail fiber protein
gene_00019	666	tail protein	gene_19	667	tail protein	QDP43720.1	putative tail protein
gene_00020	90	hypothetical protein	gene_20	91	no_hit	QDP43721.1	hypothetical protein
gene_00021	116	hypothetical protein	gene_21	117	no_hit	QDP43722.1	hypothetical protein
gene_00022	63	holin	gene_22	64	holin	QDP43723.1	hypothetical protein
gene_00023	248	endolysin	gene_23	249	endolysin	QDP43724.1	putative endolysin
gene_00024	88	hypothetical protein	gene_24	89	no_hit	QDP43725.1	hypothetical protein
gene_00025	41	hypothetical protein	gene_25	42	no_hit	QDP43726.1	hypothetical protein
gene_00026	75	transcriptional regulator	gene_26	76	transcriptional repressor	QDP43727.1	putative transcriptional regulator
gene_00027	120	hypothetical protein	gene_27	123	nucleotide kinase	QDP43728.1	hypothetical protein

gene_00031	48	hypothetical protein	gene_31	49	no_hit	QDP43732.1	hypothetical protein
gene_00033	113	hypothetical protein	gene_33	114	no_hit	QDP43734.1	hypothetical protein
gene_00034	75	hypothetical protein	gene_34	76	no_hit	QDP43735.1	hypothetical protein
gene_00035	154	hypothetical protein	gene_35	155	no_hit	QDP43736.1	hypothetical protein
gene_00036	395	helicase	gene_36	396	DNA helicase	QDP43737.1	putative helicase
gene_00037	215	recombinase	gene_37	207	Erf-like ssDNA annealing protein	QDP43738.1	putative DNA recombination protein
gene_00038	140	hypothetical protein	gene_38	141	no_hit	QDP43739.1	hypothetical protein
gene_00039	112	hypothetical protein	gene_39	113	no_hit	QDP43740.1	hypothetical protein
gene_00040	274	primase	gene_40	275	DNA polymerase/primase	QDP43741.1	putative replication protein
gene_00041	449	virulence-associated protein E	gene_41	450	DNA helicase	QDP43742.1	putative virulence-associated protein E
gene_00042	96	hypothetical protein	gene_42	97	no_hit	QDP43743.1	hypothetical protein
gene_00043	57	hypothetical protein	gene_43	58	no_hit	QDP43744.1	hypothetical protein
gene_00044	138	hypothetical protein	gene_44	139	no_hit	QDP43745.1	hypothetical protein
gene_00045	106	hypothetical protein	gene_45	107	no_hit	QDP43746.1	hypothetical protein
gene_00046	92	hypothetical protein	gene_46	93	no_hit	QDP43747.1	hypothetical protein

	gene_00_047	112	hypothetical protein	gene_47_	113	no_hit	QDP43748.1	hypothetical protein
	gene_00_048	388	hypothetical protein	gene_48_	390	DNA repair exonuclease	QDP43749.1	hypothetical protein
<i>Moineau virus</i> (NC_002072)	gene_00_002	153	terminase small subunit	gene_2_	154	terminase small subunit	NP_049390.1	terminase small subunit P27 family
	gene_00_003	229	terminase large subunit	gene_3_	230	terminase large subunit	NP_049392.1	putative terminase large subunit
	-	-	-	gene_5_	60	no_hit		
	gene_00_005	386	portal protein	gene_6_	387	portal protein	NP_049394.1	portal protein
	gene_00_006	222	scaffolding protein	gene_7_	223	head maturation protease		
	gene_00_007	397	major capsid protein	gene_8_	398	major head protein	NP_049396.1	major head protein
	gene_00_008	104	head-tail connector protein	gene_9_	105	head-tail adaptor Ad1		
	gene_00_009	116	head-tail connector protein	gene_10_	117	no_hit	NP_049398.1	head closure protein
	gene_00_010	140	tail protein	gene_11_	141	no_hit	NP_049399.1	tail protein
	gene_00_011	123	tail protein	gene_12_	124	tail protein		
	gene_00_012	203	major tail protein	gene_13_	204	major tail protein	NP_049401.1	tail protein
	gene_00_013	117	tail chaperone protein	gene_14_	118	tail protein	NP_049402.1	tail protein
	gene_00_014	1656	tape measure protein	gene_15_	1657	minor tail protein	NP_049403.2	putative tail component protein

	-	-	-	gene_16	519	minor tail protein	NP_049405.1	tail family protein
	gene_00_016	914	antireceptor	gene_17	915	tail fiber protein and host specificity	NP_049406.1	tail-host specificity protein
	-	-	-	gene_18	686	minor tail protein		
	-	-	-	gene_19	132	no_hit	NP_049409.1	DUF1366 domain-containing protein
	gene_00_021	80	holin	gene_22	81	holin	NP_049412.1	holin
	-	-	-	gene_43	133	transcriptional regulator	NP_049433.1	DUF1492 domain-containing protein
	gene_00_039	165	DNA binding protein	-	-	-		
	gene_00_041	172	HNH endonuclease	-	-	-	NP_049434.1	HNH endonuclease
	gene_00_043	235	hypothetical protein	-	-	-		
5093 (NC_012753)	gene_00_008	82	hypothetical protein	gene_8	83	no_hit		
	-	-	-	gene_9	52	no_hit		
	gene_00_012	170	DNA binding protein	gene_12	171	DNA binding protein	YP_002925093.1	DNA binding protein
	gene_00_014	97	hypothetical protein	gene_14	98	no_hit		
	gene_00_017	235	domain-containing protein	gene_17	236	no_hit	YP_002925095.1	DUF1340 domain-containing protein
	gene_00_019	148	terminase small subunit	gene_19	149	terminase small subunit	YP_002925097.1	terminase small subunit

gene_00_020	434	terminase large subunit	gene_20_	435	terminase large subunit	YP_002925_098.1	PBSX family terminase large subunit
gene_00_021	502	portal protein	gene_21_	503	portal protein	YP_002925_099.1	portal protein
gene_00_022	407	minor capsid protein	gene_22_	408	minor head protein	YP_002925_100.1	minor capsid protein
gene_00_023	205	scaffolding protein	gene_23_	206	head scaffolding protein	YP_002925_101.1	scaffolding protein
gene_00_024	281	major capsid protein	gene_24_	282	major head protein	YP_002925_102.1	N4-gp56 family major capsid protein
gene_00_025	57	hypothetical protein	gene_25_	62	no_hit	YP_002925_103.1	hypothetical protein
gene_00_026	129	head-tail connector protein	gene_26_	130	head-tail adaptor	YP_002925_104.1	hypothetical protein
gene_00_027	111	minor capsid protein	gene_27_	112	head-tail adaptor		
gene_00_028	119	minor capsid protein	gene_28_	120	minor head protein	YP_002925_105.1	minor capsid protein
gene_00_029	134	minor capsid protein	gene_29_	135	minor head protein	YP_002925_106.1	minor capsid protein
gene_00_030	167	major tail protein	gene_30_	168	major tail protein	YP_002925_107.1	tail protein
gene_00_031	120	tail assembly chaperone protein	gene_31_	121	no_hit	YP_002925_108.1	hypothetical protein
gene_00_032	219	hypothetical protein	gene_32_	220	no_hit	YP_002925_109.1	Gp15 family bacteriophage protein
gene_00_033	1528	tape-measure protein	gene_33_	1529	tail protein	YP_002925_110.1	putative antireceptor protein
gene_00_034	239	distal tail protein	gene_34_	240	no_hit	YP_002925_111.1	hypothetical protein

	gene_00_035	508	tail-associated lysin	gene_35_	509	tail protein	YP_002925_112.1	endolysin
	-	-	-	gene_37_	395	no_hit	YP_002925_114.1	antireceptor
	gene_00_039	91	holin	gene_39_	92	holin	YP_002925_116.1	hypothetical protein
	gene_00_040	80	holin	gene_40_	81	holin	YP_002925_117.1	holin
	-	-	-	gene_41_	282	endolysin	YP_002925_118.1	peptidoglycan hydrolase
987 (KU6783 89)	gene_00_002	462	terminase large subunit	gene_2_	463	terminase large subunit	AMQ65697_1	TerL
	gene_00_003	446	portal protein	gene_3_	447	portal protein	AMQ65698_1	portal protein
	gene_00_004	346	minor capsid protein	gene_4_	347	minor head protein	AMQ65699_1	minor capsid protein
	gene_00_005	199	domain-containing protein	gene_5_	200	head scaffolding protein	AMQ65700_1	scaffolding protein
	gene_00_006	287	major capsid protein	gene_6_	288	major head protein	AMQ65701_1	major capsid protein
	gene_00_007	63	Ig domain containing protein	gene_7_	64	structural protein with Ig domain	AMQ65702_1	hypothetical protein
	gene_00_008	110	head-tail connector protein	gene_8_	111	head-tail adaptor	AMQ65703_1	head-tail connector protein

gene_00_009	103	hypothetical protein	gene_9_	104	minor head protein	AMQ65704 .1	hypothetical protein
gene_00_010	108	capsid and scaffold protein	gene_10_	109	neck protein Ne1	AMQ65705 .1	hypothetical protein
gene_00_011	129	capsid protein	gene_11_	130	tail completion Tc1	AMQ65706 .1	tail terminator protein
gene_00_012	165	major tail protein	gene_12_	166	major tail protein	AMQ65707 .1	major tail protein
gene_00_013	116	tail chaperone protein	gene_13_	117	tail protein	AMQ65708 .1	tail chaperone protein
gene_00_014	89	hypothetical protein	gene_14_	90	tail assembly chaperone	AMQ65709 .1	hypothetical protein
gene_00_015	916	tape measure protein	gene_15_	917	no_hit	AMQ65710 .1	tape measure protein
gene_00_016	253	tail family protein	gene_16_	254	minor tail protein	AMQ65711 .1	distal tail protein
gene_00_017	910	tail-associated lysin	gene_17_	911	minor head protein	AMQ65712 .1	tail-associated lysin
gene_00_018	647	antireceptor	gene_18_	648	virion structural protein	AMQ65713 .1	antireceptor
gene_00_021	81	holin	gene_21_	82	holin	AMQ65716 .1	holin

	gene_00_022	200	lysin	gene_22_22	201	endolysin	AMQ65717.1	lysin
	gene_00_040	51	hypothetical protein	gene_40_40	52	no_hit	AMQ65737.1	hypothetical protein
	gene_00_042	170	DNA-binding protein	gene_42_42	171	DNA binding protein	AMQ65739.1	DNA-binding protein
	gene_00_044	79	hypothetical protein	gene_44_44	97	no_hit	AMQ65741.1	hypothetical protein
	gene_00_045	235	hypothetical protein	gene_45_45	236	no_hit	AMQ65742.1	hypothetical protein
<i>Brussow virus</i> (NC_004303)	-	-	-	gene_20_20	182	DNA binding protein	NP_695098.1	hypothetical protein
	gene_00_022	236	hypothetical protein	gene_22_22	237	no_hit	NP_695101.1	hypothetical protein
	gene_00_025	107	terminase small subunit	gene_25_25	108	no_hit		
	-	-	-	gene_26_26	412	terminase large subunit	NP_695104.1	PBSX family terminase large subunit
	gene_00_027	501	portal protein	gene_27_27	503	portal protein	NP_695105.1	portal protein
	gene_00_028	297	minor capsid protein	gene_28_28	298	minor head protein	NP_695106.1	minor capsid protein
	gene_00_029	193	capsid and scaffold protein	gene_29_29	194	head scaffolding protein	NP_695107.1	DUF4355 domain-containing protein
	gene_00_030	119	major capsid protein	gene_30_30	120	major head protein	NP_695108.1	putative structural protein

gene_00035	114	hypothetical protein	gene_35	115	neck protein Ne1	NP_695113.1	HK97 gp10 family protein
gene_00036	128	hypothetical protein	gene_36	129	tail completion Tc1	NP_695114.1	DUF3168 domain-containing protein
gene_00037	168	major tail protein	gene_37	169	major tail protein	NP_695115.1	major tail protein
gene_00038	117	tail chaperone protein	gene_38	118	tail protein	NP_695116.1	tail assembly chaperone
gene_00039	105	hypothetical protein	gene_39	106	tail assembly chaperone	NP_695117.1	hypothetical protein
gene_00040	1517	tape measure protein	gene_40	1518	tail protein	NP_695118.1	putative tail protein
gene_00041	512	tail protein	gene_41	513	minor tail protein	NP_695119.1	tail family protein
-	-	-	gene_45	844	host range and adsorption protein	NP_695120.1	putative tail protein
gene_00046	654	tail protein	gene_46	670	tail protein	NP_695124.1	hypothetical protein
gene_00047	117	hypothetical protein	gene_47	118	no_hit	NP_695125.1	hypothetical protein

The *Brussowvirus* phages shared at least 15 homologs proteins (31.25%). These homologs proteins included: portal protein, major/minor capsid proteins, head scaffolding protein, major tail protein, and tape measure protein. The phages from this genus have genomes of, on average, 37.32 kb (39.34 %GC) encoding 48 proteins. At the DNA level, these phages share at least 37.0% DNA similarity (SI file 1 Tables S8-S10, SI file 5 Table S4, SI file 6). The 987 group phages shared 23 homologs proteins (50%). These homologs proteins included: large terminase subunit, portal protein, minor/major capsid proteins, head-tail connector protein, head scaffolding protein, antireceptor, major tail protein, tape measure protein, holin, endolysin, and DNA binding protein. The phages from this genus have genomes of, on average, 32.26 kb (36.94 %GC) encoding 46 proteins. At the DNA level, these phages share at least 56.14% DNA similarity (SI file 1 Tables S11-S13, SI file 5 Table S6, SI file 6). The 5093 group phages shared at least 22 homologs proteins (46.80%). These homologs proteins included: small/large terminase subunits, portal protein, minor/major capsid proteins, head scaffolding protein, antireceptor, major tail protein, holin, endolysin, and DNA binding protein. The phages from this genus have genomes of, on average, 34.25 kb (38.26 %GC) encoding 47 proteins. At the DNA level, these phages share at least 64.44% DNA similarity (SI file 1 Tables S14-S16, SI file 5 Table S8, SI file 6).

The genes encoding the major tail protein (MTP) and portal protein were detected and annotated in the five genera by all tools (Roary, VirClust and CoreGenes) and, therefore, were used to visualize the clades of genera in the phylogenetic analysis. All genes encoding the MTP and Portal proteins produced a phylogenetic tree in which the genus was represented by a well-supported clade (Figure 4, SI file 7). Based on the results of the phage genomes clustering analyses, detection and phylogeny of the core genes, the *S. thermophilus* phages were classified at genus level according to the criteria described below. The criteria for demarcating a new genus for the phages of the P738 group were: 70% DNA sequence identity and monophyly in the terminase large subunit (TerL), portal, and holin proteins phylogenetic trees (SI file 5 Figures S3 and S8). The criteria for demarcating *Moineauvirus* genus were: at least 44.35% DNA sequence identity and monophyly in the terminase small subunit (TerS), major tail protein (MTP), and portal proteins phylogenetic trees (SI file 5 Figures S4 and S9). The criteria for demarcating *Brussowvirus* genus were: at least 37.0% DNA sequence identity and monophyly in the major capsid protein (MCP), portal protein, and major tail protein (MTP) proteins phylogenetic trees (SI file 5 Figures S5 and S10). The criteria for demarcating a new genus for the phages of the 987 group were: at least 56.3% DNA sequence identity and monophyly in the major capsid protein (MCP), portal protein, and major tail protein (MTP) proteins phylogenetic

Phages from the *Moineauvirus* and *Brussowvirus* were distributed in 23 and 12 genera, respectively, while the phages from groups 5093 and 987 were reorganized into three genera each. However, the VIRIDIC analysis was based only on nucleotide identity thresholds of 70% for genus level demarcation. The presence of a set of conserved genes and monophyly in the signature genes, added to the morphological, genomic, and ecological characteristics shared by all the phages in the same group, allow us to classify the five groups described in the literature as being, in fact, five genera of streptococcal phages. With these results, we allowed the expansion of the number of genera of *S. thermophilus* phages taxonomically classified by ICTV from two to five, adding three new genera within the *Caudoviricetes* class.

3.4.3. Family level clustering of streptococcal phages

Criteria for demarcation of new families and orders are currently being defined by ICTV, however, it has been proposed that the family-level represents a cohesive and monophyletic group in the main predicted proteome-based clustering tools, whose members share a significant number of orthologous genes. The result of the amino acid-based VICTOR analysis showed that *S. thermophilus* phages are organized into one subfamily and one family (Figure 1). However, VICTOR analysis does not identify the orthologous proteins that contributed to the clustering of phages at subfamily and family levels.

It was described that a 0.90 intergenomic distance threshold applied to the PC tree in VirClust delineated most families within *Duplodnaviria* realm (Moraru, 2021). Here, different distance thresholds (0.98, 0.90 and 0.85) were applied to the hierarchical trees produced by VirClust, to determine if one of them is suitable for defining family level clusters. When a 0.98 intergenomic distance threshold was applied to the PC tree in VirClust, the 183 phage genomes clustered into a single VGC (SI file 5 Figure S13), however, the core proteins shared by the phages used to define the VGC were not identified, that is, no PC was formed, generating uncertainty regarding the clustering for demarcation of the family level. Potentially, 0.98 could be used for order level delineation. When protein clustering was performed with the default parameters of VirClust (0.90 intergenomic distance threshold) the resulting tree was split into three VGCs based on PCs (SI file 5 Figure S14). The core proteins, defining each VGC, were identified and annotated (SI file 1 Tables S17-S19). Phage genomes from P738 and *Moineauvirus* genera formed one family each, sharing 44 and 19 PCs, respectively. While phage genomes from *Brussowvirus*, 987 and 5093 genera formed a third family, sharing two PCs. One of these shared PCs had its putative function identified and corresponded to the gene encoding DNA binding protein. However, as only a small proportion of proteins were shared

with other phages in the dataset, it increases clustering uncertainty and can indicate an incorrect clustering (Moraru, 2021). Identical results were obtained with CoreGenes (SI file 1 Table S20). Roary detected a single core gene among the phages of these three genera, encoding a hypothetical protein, also detected by the other tools (SI file 1 Table S21). A total of 107 shell genes (15% \leq strains $<$ 95%) were also found among the phages of these genera. A threshold of 0.85 intergenomic distance applied on the hierarchical tree split the phage genomes into four VGCs (SI file 5 Figure S15). Phage genomes from P738, *Moineavirus* and 5093 genera formed one family each, sharing 44, 19 and 25 PCs, respectively (SI file Tables S22-S24). Phage from *Brussowvirus* and 987 genera formed a fourth family, sharing seven PCs (SI file 1 Table S25). Shared PCs by the phages of the *Brussowvirus* and 987 genera include the DNA binding protein, portal protein, minor head protein, connector, major tail protein, tail protein, and other hypothetical proteins. Similar results were obtained with CoreGenes (SI file 1 Table S26), which identified eight orthologous proteins shared by these phage genera. Roary detected two core genes among the phages of these genera, encoding a hypothetical protein and major tail protein, also detected by the other tools (SI file 1 Table S27). A total of 93 shell genes (15% \leq strains $<$ 95%) were also found among the phages of these genera.

Similar clades were obtained using the main predicted proteome-based clustering tools, confirming that *S. thermophilus* phages are grouped separately into five genera in the hierarchical clustering. However, these clades were assigned to different number of families by the VICTOR (a single family) and VirClust analysis (one, three or four families, according to the intergenomic distance threshold). Phylogenetic analysis of phage proteomes using ViPTree revealed that three of the five genera phages are more closely related to each other than to any other phages (Figure 5). The phages from *Moineavirus*, *Brussowvirus* and 5093 genera formed a cohesive and monophyletic group and appear to form a family separately from the P738 and 987 genera phages. Orthologous proteins shared by the phages of these three genera were identified using Roary. A single core gene (gene encoding hypothetical protein) was included (SI file 1 Table S28). A total of 79 shell genes (15% \leq strains $<$ 95%) and a single soft core gene (95% \leq strains $<$ 99%) were also found among the phages of these genera. On the other hand, the next closest relatives of P738 and 987 genera were phages that infect the pathogenic bacteria *Streptococcus pyogenes* (Streptococcus phage T12, NC_028700 and Streptococcus prophage 315.5, NC_004588) and the lactic acid bacterium *Lactococcus lactis* (Lactococcus phage Tuc2009, NC_002703 and Lactococcus phage P335, DQ838728), respectively. *Moineavirus* and *Brussowvirus* genera shared five orthologous proteins using CoreGenes (SI file 1 Tables S30) and two core genes detected with Roary, encoding hypothetical proteins (SI

file 1 Tables S31). Phages from these genera also shared 70 shell genes (15% \leq strains $<$ 95%) and two soft core genes (95% \leq strains $<$ 99%).

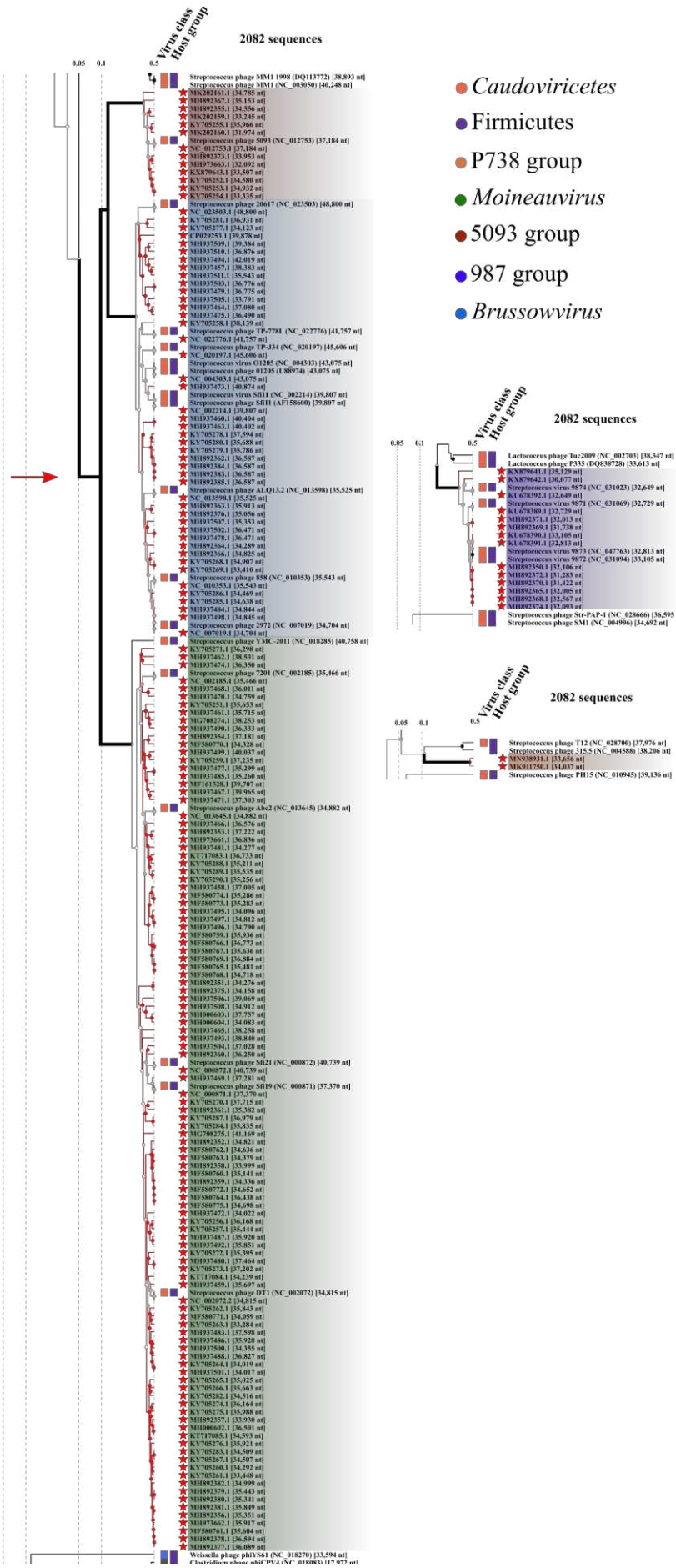


Figure 5 - Parts of the proteomic tree constructed by ViPTree. The tree includes 2082 related sequences with the 183 *Streptococcus thermophilus* phages. The phages of interest in this study are indicated with red stars. The red arrow indicates the possible family-level clustering for phages of *Moineauvirus*, *Brussowvirus* and 5093 genera. The class of the related phages and the phylum of bacterial hosts are also annotated. The colored boxes delineate the genera. The full phylogenetic tree created by ViPTree can be seen in SI file 8.

4. DISCUSSION

Phages have complex evolutionary relationships (Dion et al., 2020; Koonin et al., 2020) and several strategies have been used in an attempt to classify the phages that infect *S. thermophilus*. Historically, the classification of streptococcal phages was based on morphology and the combination of the DNA packaging mode and the number of major structural proteins (Desiere et al., 1999; Le Marrec et al., 1997; Lucchini et al., 1999a, 1999b). This classification limited the division of phages that infect *S. thermophilus* to two groups: *cos*-containing phages, whose phages had cohesive genomic ends when treated with restriction enzymes (*Moineauvirus* genus) and two main structural proteins; and *pac*-containing phages, whose phages showed absence of cohesive ends (*Brussowvirus* genus) and three main structural proteins (Le Marrec et al., 1997). The discovery of phages with new properties resulted in the need to update this classification, which led to the proposition of two new groups, 5093 (Mills et al., 2011) and 987 (McDonnell et al., 2016). Although group 5093 phages have a genomic architecture similar to members of the *Moineauvirus* genus, with two main structural proteins, their genome sequences are more similar to those of phages from the *Brussowvirus* genus and a streptococcal prophage of non-dairy origin (Mills et al., 2011). The genomes of group 987 phages, on the other hand, show genetic exchange events, exhibiting a DNA sequence relationship with the morphogenesis modules of certain phages that infect *L. lactis* of the P335 group and with the replication modules of phages that infect *S. thermophilus* (McDonnell et al., 2016). A fifth group, P738, formed by two genetically distinct phages from the other phages that infect *S. thermophilus* was described. These phages are closely related to each other and share similarities with non-dairy streptococcal phages (Philippe et al., 2020).

Changes in virus classification, taxonomy, and nomenclature occur annually (Walker et al., 2021) as a result of an increasing increase in sequenced genomes and the rise of metagenomic studies (Al-Shayeb et al., 2020; Dion et al., 2020; Simmonds et al., 2017; Turner et al., 2021). To accommodate the full spectrum of virus genetic divergence, ICTV changed the hierarchical taxonomic classification structure to 15 positions, including eight major classifications (realm, kingdom, phylum, class, order, family, genus, and species) and seven

classifications derived from main (ICTV, 2020). At the realm level, viruses were grouped in *Adnaviria*, *Duplodnaviria*, *Monodnaviria*, *Riboviria*, *Ribozyviria*, and *Varidnaviria* (Walker et al., 2021). For tailed phages, it was proposed to eliminate the order *Caudovirales* and the families *Myoviridae*, *Siphoviridae*, and *Podoviridae* (Turner et al., 2021). Without these taxonomic levels, the *Caudoviricetes* class automatically integrated all tailed phages and the criteria for demarcation of new families and orders are currently being defined by ICTV, however, it has been proposed that they should be based on analyzes of entire viral proteomes and consider shared orthologous proteins (Koonin et al., 2020; Turner et al., 2021).

In this study, we used a wide range of complementary methods including comparative genomics, core genome analysis, and signature genes phylogenetics, to show that the *S. thermophilus* phages are organized into 142 species and five genera, and that due to their genetic diversity, the classification at family level varies according to the classification criteria used. In the process of the taxons evaluation, we explored the feasibility of different demarcation criteria and critically evaluated the usefulness of our methods for phage classification. The convergence of results, drawing a consistent and comprehensive picture of five well-supported clades, regardless of method, demonstrates that the tools applied here are particularly useful in *S. thermophilus* phages taxonomy at species and genus levels. At the family level classification, the *S. thermophilus* phages from *Brussowvirus*, 987 and 5093 genera formed a group sharing at least one ortholog gene, or yet, *Brussowvirus* and 987 genera formed a group sharing at least two orthologous genes. These groupings corroborates a recent study that suggested the evolution/emergence of 5093 and 987 phages through recombination with temperate *Brussowviruses*, based on phylogenetic analyzes of streptococcal phage replication modules (Hanemaaijer et al., 2021). On the other hand, the phages from *Moineauvirus*, *Brussowvirus* and 5093 genera formed a monophyletic group in the viral proteomic tree sharing at least one ortholog gene. The relationship among the phages of these three genera has been described previously (Mills et al., 2011). It was reported that the first phage of the genus 5093 (Streptococcus phage 5093, NC_012753) evolved from gene exchanges with *pac*-containing phages (*Brussowvirus* genus), representing a hybrid phage, and that although have a genomic architecture similar to members of the *Moineauvirus* genus, their genome sequences are more similar to those of phages from the *Brussowvirus* genus.

Five well-supported clades were obtained using the main classification tools and the convergence of their results for the analyzed taxa, confirm that *S. thermophilus* phages are grouped separately into five genera in the hierarchical clustering, whose members do not share a sufficiently conserved gene. However, the genes encoding holin protein is conserved in more

than 95% of genomes. Holin superfamily I, in particular, is widespread among LAB (Fujimoto et al., 2020). Mills et al. (2011) also observed that both endolysin and holin genes are highly conserved in *S. thermophilus* phages, mainly consisting on holin superfamily I. Our study highlights two other holin families in *S. thermophilus* phages, rIt family and holin superfamily VI. Labrie et al. (2004) were the authors that first described the rIt holin family in *Lactococcus lactis* phages. The authors observed that the lysis cassette of the phage rIt has the same general lysis module features as other *Lactococcus lactis* phages.

Comparative genomic analyzes have shown that the genomes of these phages exhibit a high degree of conservation within the genera defined here, particularly in the structural modules (Lavelle, Martinez, et al., 2018; McDonnell et al., 2017; Philippe et al., 2020; Szymczak, Rau, et al., 2019), but also exhibit nucleotide divergence between genera and between phage within the same genus. This particularity in genomes of sharing regions of high sequence similarity with abrupt transitions into adjacent regions without detectable similarity is described as genetic mosaicism (Dion et al., 2020; Hendrix et al., 1999). These regions (genes and gene blocks) have distinct evolutionary histories, due to multiple genetic exchange events that suffer in response to the selective pressure of their hosts, which drives their diversity (Dion et al., 2020). Streptococcal phage genomes are undergoing a process of evolution through, as studies indicate, horizontal gene transfer, insertions, deletions, point mutations, and exchange of genetic modules (Brussow & Desiere, 2001; Desiere et al., 1998; Lavelle, Martinez, et al., 2018; Lavelle, Murphy, et al., 2018; Lucchini et al., 1999a; Philippe et al., 2020; Szymczak et al., 2017). From an industrial point of view, this diversification represents a constant threat to milk fermentation processes, as it can result in the extension of the host range and make it difficult to efficiently control phage infections in starter cultures, which normally consist of mixtures of lactic acid bacteria (LAB) (de Melo et al., 2018; Leroy & De Vuyst, 2004; McDonnell et al., 2016; Quiberoni et al., 2010; Szymczak et al., 2017). As the 183 phages evaluated were mainly isolated from industrial fermentation environments, industrialization and the use of commercial starter cultures have likely limited the diversity of bacterial strains in dairy environments, resulting in a high gene flow capable of generating phages recombinants with potentially expanded lytic activity.

5. CONCLUSION

The phages that infect *S. thermophilus* have a remarkable genetic diversity that can result in the extension of their host range, which poses a threat to milk fermentation processes. Considering all the advantages and limitations of the classification tools used here, and the

convergence of their results for the analyzed taxa, our study represents a contribution to the expansion of knowledge about the genetic diversity and evolution of streptococcal phages. Our work also highlights the importance of monitoring, isolating, and sequencing streptococcal phages in industrial environments to promote control strategies and minimize failures in milk fermentation processes.

6. REFERENCES

- Accolas, J.-P., & Spillmann, H. (1979). The Morphology of Six Bacteriophages of *Streptococcus thermophilus*. *Journal of Applied Bacteriology*, 47(1), 135–144. <https://doi.org/10.1111/j.1365-2672.1979.tb01177.x>
- Achigar, R., Magadán, A. H., Tremblay, D. M., Julia Pianzzola, M., & Moineau, S. (2017). Phage-host interactions in *Streptococcus thermophilus*: Genome analysis of phages isolated in Uruguay and ectopic spacer acquisition in CRISPR array. *Scientific Reports*, 7(1), 43438. <https://doi.org/10.1038/srep43438>
- Ali, Y., Koberg, S., Heřner, S., Sun, X., Rabe, B., Back, A., Neve, H., & Heller, K. J. (2014). Temperate *Streptococcus thermophilus* phages expressing superinfection exclusion proteins of the Ltp type. *Frontiers in Microbiology*, 5. <https://doi.org/10.3389/fmicb.2014.00098>
- Al-Shayeb, B., Sachdeva, R., Chen, L.-X., Ward, F., Munk, P., Devoto, A., Castelle, C. J., Olm, M. R., Bouma-Gregson, K., Amano, Y., He, C., Méheust, R., Brooks, B., Thomas, A., Lavy, A., Matheus-Carnevali, P., Sun, C., Goltsman, D. S. A., Borton, M. A., ... Banfield, J. F. (2020). Clades of huge phages from across Earth's ecosystems. *Nature*, 578(7795), 425–431. <https://doi.org/10.1038/s41586-020-2007-4>
- Arioli, S., Eraclio, G., Della Scala, G., Neri, E., Colombo, S., Scaloni, A., Fortina, M. G., & Mora, D. (2018). Role of Temperate Bacteriophage ϕ 20617 on *Streptococcus thermophilus* DSM 20617T Autolysis and Biology. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.02719>
- Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., van Dijk, A. A., Ebrecht, A. C., ... Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557), 871–876. <https://doi.org/10.1126/science.abj8754>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41(D1), D36–D42. <https://doi.org/10.1093/nar/gks1195>
- Binetti, A. G., Del Río, B., Martín, M. C., & Álvarez, M. A. (2005). Detection and Characterization of *Streptococcus thermophilus* Bacteriophages by Use of the Antireceptor Gene Sequence. *Applied and Environmental Microbiology*, 71(10), 6096–6103. <https://doi.org/10.1128/AEM.71.10.6096-6103.2005>

- Brussow, H., & Desiere, F. (2001). Comparative phage genomics and the evolution of *Siphoviridae*: Insights from dairy phages. *Molecular Microbiology*, *39*(2), 213–223. <https://doi.org/10.1046/j.1365-2958.2001.02228.x>
- da Silva Duarte, V., Giaretta, S., Campanaro, S., Treu, L., Armani, A., Tarrah, A., Oliveira de Paula, S., Giacomini, A., & Corich, V. (2018). A Cryptic Non-Inducible Prophage Confers Phage-Immunity on the *Streptococcus thermophilus* M17PTZA496. *Viruses*, *11*(1), 7. <https://doi.org/10.3390/v11010007>
- de Melo, A. G., Levesque, S., & Moineau, S. (2018). Phages as friends and enemies in food processing. *Current Opinion in Biotechnology*, *49*, 185–190. <https://doi.org/10.1016/j.copbio.2017.09.004>
- Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., & Gascuel, O. (2008). Phylogeny.fr: Robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*, *36*(Web Server issue), W465–469. <https://doi.org/10.1093/nar/gkn180>
- Desiere, F., Lucchini, S., & Brüssow, H. (1998). Evolution of *Streptococcus thermophilus* Bacteriophage Genomes by Modular Exchanges Followed by Point Mutations and Small Deletions and Insertions. *Virology*, *241*(2), 345–356. <https://doi.org/10.1006/viro.1997.8959>
- Desiere, F., Lucchini, S., & Brüssow, H. (1999). Comparative Sequence Analysis of the DNA Packaging, Head, and Tail Morphogenesis Modules in the Temperate *cos*-Site *Streptococcus thermophilus* Bacteriophage Sfi21. *Virology*, *260*(2), 244–253. <https://doi.org/10.1006/viro.1999.9830>
- Deveau, H., Barrangou, R., Garneau, J. E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., & Moineau, S. (2008). Phage Response to CRISPR-Encoded Resistance in *Streptococcus thermophilus*. *Journal of Bacteriology*, *190*(4), 1390–1400. <https://doi.org/10.1128/JB.01412-07>
- Dion, M. B., Oechslin, F., & Moineau, S. (2020). Phage diversity, genomics and phylogeny. *Nature Reviews Microbiology*, *18*(3), 125–138. <https://doi.org/10.1038/s41579-019-0311-5>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Farris, J. S. (1972). Estimating Phylogenetic Trees from Distance Matrices. *The American Naturalist*, *106*(951), 645–668. <https://www.jstor.org/stable/2459725>
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Research*, *39*(suppl_2), W29–W37. <https://doi.org/10.1093/nar/gkr367>
- Fujimoto, K., Kimura, Y., Shimohigoshi, M., Satoh, T., Sato, S., Tremmel, G., Uematsu, M., Kawaguchi, Y., Usui, Y., Nakano, Y., Hayashi, T., Kashima, K., Yuki, Y., Yamaguchi, K., Furukawa, Y., Kakuta, M., Akiyama, Y., Yamaguchi, R., Crowe, S. E., ... Uematsu, S. (2020). Metagenome Data on Intestinal Phage-Bacteria

- Associations Aids the Development of Phage Therapy against Pathobionts. *Cell Host & Microbe*, 28(3), 380-389.e9. <https://doi.org/10.1016/j.chom.2020.06.005>
- Gilchrist, C. L. M., & Chooi, Y.-H. (2020). clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *BioRxiv*, 2020.11.08.370650. <https://doi.org/10.1101/2020.11.08.370650>
- Göker, M., García-Blázquez, G., Voglmayr, H., Tellería, M. T., & Martín, M. P. (2009). Molecular Taxonomy of Phytopathogenic Fungi: A Case Study in Peronospora. *PLOS ONE*, 4(7), e6319. <https://doi.org/10.1371/journal.pone.0006319>
- Guglielmotti, D. M., Deveau, H., Binetti, A. G., Reinheimer, J. A., Moineau, S., & Quiberoni, A. (2009). Genome analysis of two virulent *Streptococcus thermophilus* phages isolated in Argentina. *International Journal of Food Microbiology*, 136(1), 101–109. <https://doi.org/10.1016/j.ijfoodmicro.2009.09.005>
- Hanemaaijer, L., Kelleher, P., Neve, H., Franz, C. M. A. P., de Waal, P. P., van Peij, N. N. M. E., van Sinderen, D., & Mahony, J. (2021). Biodiversity of Phages Infecting the Dairy Bacterium *Streptococcus thermophilus*. *Microorganisms*, 9(9), 1822. <https://doi.org/10.3390/microorganisms9091822>
- Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E., & Hatfull, G. F. (1999). Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5), 2192–2197. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC26759/>
- Hockenberry, A. J., & Wilke, C. O. (2021). BACPHLIP: Predicting bacteriophage lifestyle from conserved protein domains. *PeerJ*, 9, e11396. <https://doi.org/10.7717/peerj.11396>
- Hynes, A. P., Rousseau, G. M., Agudelo, D., Goulet, A., Amigues, B., Loehr, J., Romero, D. A., Fremaux, C., Horvath, P., Doyon, Y., Cambillau, C., & Moineau, S. (2018). Widespread anti-CRISPR proteins in virulent bacteriophages inhibit a range of Cas9 proteins. *Nature Communications*, 9(1), 2919. <https://doi.org/10.1038/s41467-018-05092-w>
- ICTV. (2020). The new scope of virus taxonomy: Partitioning the virosphere into 15 hierarchical ranks. *Nature Microbiology*, 5(5), 668–674. <https://doi.org/10.1038/s41564-020-0709-x>
- Koonin, E. V., Dolja, V. V., Krupovic, M., Varsani, A., Wolf, Y. I., Yutin, N., Zerbini, F. M., & Kuhn, J. H. (2020). Global Organization and Proposed Megataxonomy of the Virus World. *Microbiology and Molecular Biology Reviews*, 84(2), e00061-19, /mmb/84/2/MMBR.00061-19.atom. <https://doi.org/10.1128/MMBR.00061-19>
- Lavelle, K., Martinez, I., Neve, H., Lugli, G., Franz, C., Ventura, M., Bello, F., Sinderen, D., & Mahony, J. (2018). Biodiversity of *Streptococcus thermophilus* Phages in Global Dairy Fermentations. *Viruses*, 10(10), 577. <https://doi.org/10.3390/v10100577>
- Lavelle, K., Murphy, J., Fitzgerald, B., Lugli, G. A., Zomer, A., Neve, H., Ventura, M., Franz, C. M., Cambillau, C., van Sinderen, D., & Mahony, J. (2018). A Decade of

Streptococcus thermophilus Phage Evolution in an Irish Dairy Plant. *Applied and Environmental Microbiology*, 84(10), e02855-17, /aem/84/10/e02855-17.atom.
<https://doi.org/10.1128/AEM.02855-17>

- Le Marrec, C., van Sinderen, D., Walsh, L., Stanley, E., Vlegels, E., Moineau, S., Heinze, P., Fitzgerald, G., & Fayard, B. (1997). Two groups of bacteriophages infecting *Streptococcus thermophilus* can be distinguished on the basis of mode of packaging and genetic determinants for major structural proteins. *Applied and Environmental Microbiology*, 63(8), 3246–3253. <https://doi.org/10.1128/AEM.63.8.3246-3253.1997>
- Lefort, V., Desper, R., & Gascuel, O. (2015). FastME 2.0: A Comprehensive, Accurate, and Fast Distance-Based Phylogeny Inference Program. *Molecular Biology and Evolution*, 32(10), 2798–2800. <https://doi.org/10.1093/molbev/msv150>
- Leroy, F., & De Vuyst, L. (2004). Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology*, 15(2), 67–78. <https://doi.org/10.1016/j.tifs.2003.09.004>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>
- Levesque, C., Duplessis, M., Labonte, J., Labrie, S., Fremaux, C., Tremblay, D., & Moineau, S. (2005). Genomic Organization and Molecular Analysis of Virulent Bacteriophage 2972 Infecting an Exopolysaccharide-Producing *Streptococcus thermophilus* Strain. *APPL. ENVIRON. MICROBIOL.*, 71, 12.
- Li, Z., Jaroszewski, L., Iyer, M., Sedova, M., & Godzik, A. (2020). FATCAT 2.0: Towards a better understanding of the structural diversity of proteins. *Nucleic Acids Research*, 48(W1), W60–W64. <https://doi.org/10.1093/nar/gkaa443>
- Lucchini, S., Desiere, F., & Brüssow, H. (1998). The Structural Gene Module in *Streptococcus thermophilus* Bacteriophage ϕ Sfi11 Shows a Hierarchy of Relatedness to *Siphoviridae* from a Wide Range of Bacterial Hosts. *Virology*, 246(1), 63–73. <https://doi.org/10.1006/viro.1998.9190>
- Lucchini, S., Desiere, F., & Brüssow, H. (1999a). The Genetic Relationship between Virulent and Temperate *Streptococcus thermophilus* Bacteriophages: Whole Genome Comparison of cos-Site Phages Sfi19 and Sfi21. *Virology*, 260(2), 232–243. <https://doi.org/10.1006/viro.1999.9814>
- Lucchini, S., Desiere, F., & Brüssow, H. (1999b). Comparative Genomics of *Streptococcus thermophilus* Phage Species Supports a Modular Evolution Theory. *Journal of Virology*, 73(10), 8647–8656. <https://doi.org/10.1128/JVI.73.10.8647-8656.1999>
- Mahony, J., Casey, E., & van Sinderen, D. (2020). The Impact and Applications of Phages in the Food Industry and Agriculture. *Viruses*, 12(2), 210. <https://doi.org/10.3390/v12020210>
- Mahony, J., & van Sinderen, D. (2014). Current taxonomy of phages infecting lactic acid bacteria. *Frontiers in Microbiology*, 5. <https://doi.org/10.3389/fmicb.2014.00007>

- McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.-P., Lugli, G. A., Ventura, M., Kouwen, T. R., & van Sinderen, D. (2017). Global Survey and Genome Exploration of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Frontiers in Microbiology*, *8*, 1754. <https://doi.org/10.3389/fmicb.2017.01754>
- McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.-P., Kouwen, T., & van Sinderen, D. (2016). Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, *82*(17), 5153–5165. <https://doi.org/10.1128/AEM.00835-16>
- Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., & Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, *14*(1), 60. <https://doi.org/10.1186/1471-2105-14-60>
- Meier-Kolthoff, J. P., & Göker, M. (2017). VICTOR: Genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics*, *33*(21), 3396–3404. <https://doi.org/10.1093/bioinformatics/btx440>
- Mills, S., Griffin, C., O’Sullivan, O., Coffey, A., McAuliffe, O. E., Meijer, W. C., Serrano, L. M., & Ross, R. P. (2011). A new phage on the ‘Mozzarella’ block: Bacteriophage 5093 shares a low level of homology with other *Streptococcus thermophilus* phages. *International Dairy Journal*, *21*(12), 963–969. <https://doi.org/10.1016/j.idairyj.2011.06.003>
- Moraru, C. (2021). VirClust – a tool for hierarchical clustering, core gene detection and annotation of (prokaryotic) viruses. *BioRxiv*, 2021.06.14.448304. <https://doi.org/10.1101/2021.06.14.448304>
- Moraru, C., Varsani, A., & Kropinski, A. M. (2020). VIRIDIC—A Novel Tool to Calculate the Intergenomic Similarities of Prokaryote-Infecting Viruses. *Viruses*, *12*(11), 1268. <https://doi.org/10.3390/v12111268>
- Neve, H., Zenz, K. I., Desiere, F., Koch, A., Heller, K. J., & Brüssow, H. (1998). Comparison of the Lysogeny Modules from the Temperate *Streptococcus thermophilus* Bacteriophages TP-J34 and Sfi21: Implications for the Modular Theory of Phage Evolution. *Virology*, *241*(1), 61–72. <https://doi.org/10.1006/viro.1997.8960>
- Nishimura, Y., Yoshida, T., Kuronishi, M., Uehara, H., Ogata, H., & Goto, S. (2017). ViPTree: The viral proteomic tree server. *Bioinformatics*, *33*(15), 2379–2380. <https://doi.org/10.1093/bioinformatics/btx157>
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. <https://doi.org/10.1093/bioinformatics/btv421>
- Philippe, C., Levesque, S., Dion, M. B., Tremblay, D. M., Horvath, P., Lüth, N., Cambillau, C., Franz, C., Neve, H., Fremaux, C., Heller, K. J., & Moineau, S. (2020). Novel Genus of Phages Infecting *Streptococcus thermophilus*: Genomic and Morphological

- Characterization. *Applied and Environmental Microbiology*, 86(13), e00227-20, /aem/86/13/AEM.00227-20.atom. <https://doi.org/10.1128/AEM.00227-20>
- Pujato, S. A., Quiberoni, A., & Mercanti, D. J. (2019). Bacteriophages on dairy foods. *Journal of Applied Microbiology*, 126(1), 14–30. <https://doi.org/10.1111/jam.14062>
- Quiberoni, A., Moineau, S., Rousseau, G. M., Reinheimer, J., & Ackermann, H.-W. (2010). *Streptococcus thermophilus* bacteriophages. *International Dairy Journal*, 20(10), 657–664. <https://doi.org/10.1016/j.idairyj.2010.03.012>
- Rambaut, A. (2012). *FigTree*. <http://tree.bio.ed.ac.uk/software/figtree/>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). Mol* Viewer: Modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, 49(W1), W431–W437. <https://doi.org/10.1093/nar/gkab314>
- Simmonds, P., Adams, M. J., Benkő, M., Breitbart, M., Brister, J. R., Carstens, E. B., Davison, A. J., Delwart, E., Gorbalenya, A. E., Harrach, B., Hull, R., King, A. M. Q., Koonin, E. V., Krupovic, M., Kuhn, J. H., Lefkowitz, E. J., Nibert, M. L., Orton, R., Roossinck, M. J., ... Zerbini, F. M. (2017). Virus taxonomy in the age of metagenomics. *Nature Reviews Microbiology*, 15(3), 161–168. <https://doi.org/10.1038/nrmicro.2016.177>
- Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmeler, S., Frey, J. E., & Ahrens, C. H. (2019). Long-read based de novo assembly of low-complexity metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BMC Microbiology*, 19(1), 143. <https://doi.org/10.1186/s12866-019-1500-0>
- Stanley, E., Fitzgerald, G., Le Marrec, C., Fayard, B., & Van Sinderen, D. (1997). Sequence analysis and characterization of ØO1205, a temperate bacteriophage infecting *Streptococcus thermophilus* CNRZ1205. *Microbiology (Reading, England)*, 143 (Pt 11), 3417–3429. <https://doi.org/10.1099/00221287-143-11-3417>
- Szymczak, P., Janzen, T., Neves, A. R., Kot, W., Hansen, L. H., Lametsch, R., Neve, H., Franz, C. M. A. P., & Vogensen, F. K. (2017). Novel Variants of *Streptococcus thermophilus* Bacteriophages Are Indicative of Genetic Recombination among Phages from Different Bacterial Species. *Applied and Environmental Microbiology*, 83(5), e02748-16, e02748-16. <https://doi.org/10.1128/AEM.02748-16>
- Szymczak, P., Rau, M. H., Monteiro, J. M., Pinho, M. G., Filipe, S. R., Vogensen, F. K., Zeidan, A. A., & Janzen, T. (2019). A comparative genomics approach for identifying host-range determinants in *Streptococcus thermophilus* bacteriophages. *Scientific Reports*, 9(1), 7991. <https://doi.org/10.1038/s41598-019-44481-z>
- Szymczak, P., Vogensen, F. K., & Janzen, T. (2019). Novel isolates of *Streptococcus thermophilus* bacteriophages from group 5093 identified with an improved multiplex

- PCR typing method. *International Dairy Journal*, 91, 18–24.
<https://doi.org/10.1016/j.idairyj.2018.12.001>
- Tremblay, D. M., & Moineau, S. (1999). Complete Genomic Sequence of the Lytic Bacteriophage DT1 of *Streptococcus thermophilus*. *Virology*, 255(1), 63–76.
<https://doi.org/10.1006/viro.1998.9525>
- Turner, D., Kropinski, A. M., & Adriaenssens, E. M. (2021). A Roadmap for Genome-Based Phage Taxonomy. *Viruses*, 13(3), 506. <https://doi.org/10.3390/v13030506>
- Walker, P. J., Siddell, S. G., Lefkowitz, E. J., Mushegian, A. R., Adriaenssens, E. M., Alfenas-Zerbini, P., Davison, A. J., Dempsey, D. M., Dutilh, B. E., García, M. L., Harrach, B., Harrison, R. L., Hendrickson, R. C., Junglen, S., Knowles, N. J., Krupovic, M., Kuhn, J. H., Lambert, A. J., Łobocka, M., ... Zerbini, F. M. (2021). Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Archives of Virology*. <https://doi.org/10.1007/s00705-021-05156-1>
- Zafar, N., Mazumder, R., & Seto, D. (2002). CoreGenes: A computational tool for identifying and cataloging “core” genes in a set of small genomes. *BMC Bioinformatics*, 3(1), 12.
<https://doi.org/10.1186/1471-2105-3-12>
- Zinno, P., Janzen, T., Bennedsen, M., Ercolini, D., & Mauriello, G. (2010). Characterization of *Streptococcus thermophilus* lytic bacteriophages from mozzarella cheese plants. *International Journal of Food Microbiology*, 138(1–2), 137–144.
<https://doi.org/10.1016/j.ijfoodmicro.2009.12.008>

SUPPLEMENTARY INFORMATION (SI)**SI file 1**

<http://tiny.cc/ykwtuz>

SI file 2

<http://tiny.cc/mkjtuz>

SI file 3

<http://tiny.cc/nkjtuz>

SI file 4

<http://tiny.cc/okjtuz>

SI file 5

● P738 group ● *Moineauvirus* ● 5093 group ● 987 group ● *Brussowvirus*

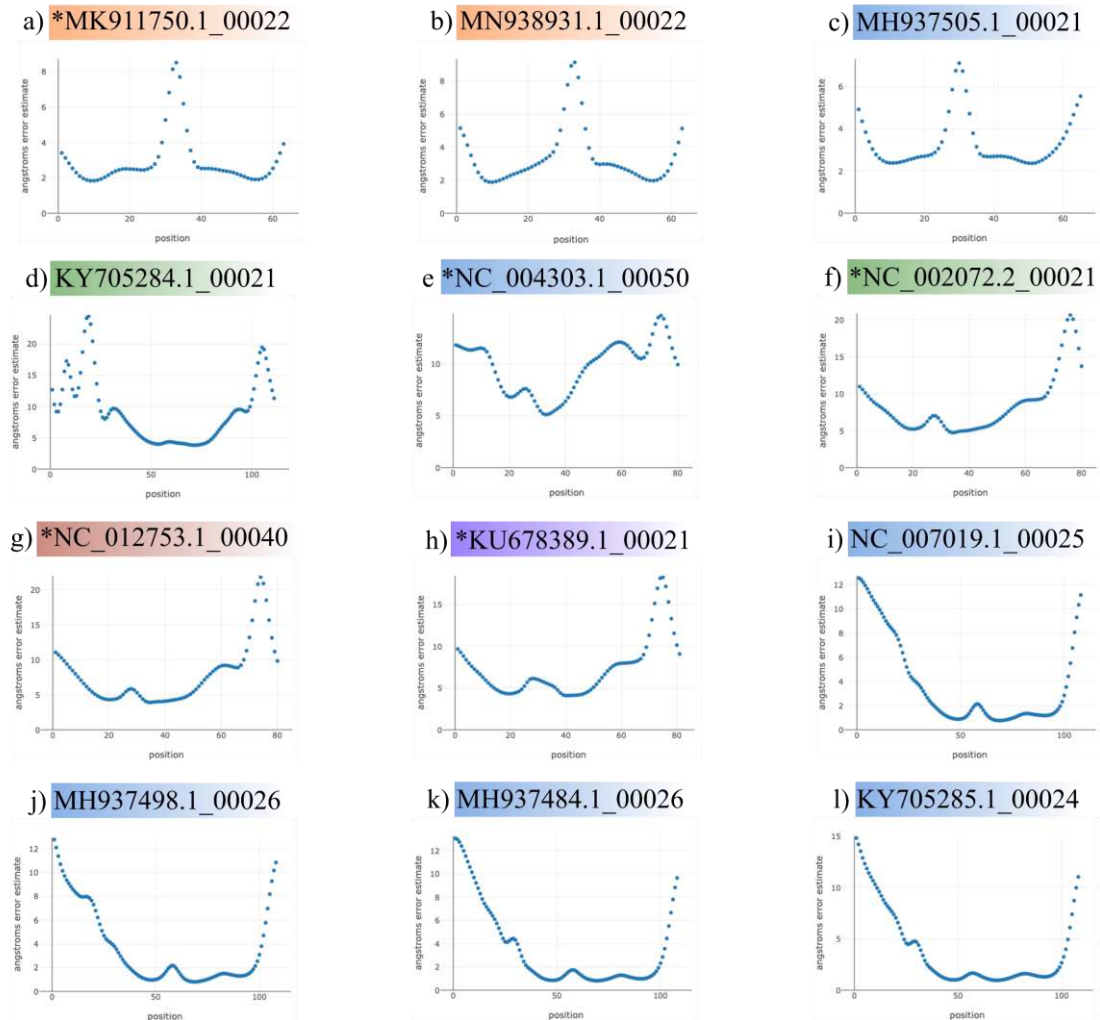


Figure S1 - Angstroms error estimate based on the amino acid position of representative holins encoded by *Streptococcus thermophilus* phages. Phage r1t holin family (PF16945) from P738 group (a, b), and *Brussowvirus* genus (c). Holin superfamily I (PF04531) from *Moineauvirus* genus (d, f), *Brussowvirus* genus (e), 5093 group (g), and 987 group (h). Superfamily VI holins (PF09682) from *Brussowvirus* genus (i, j, k, l).

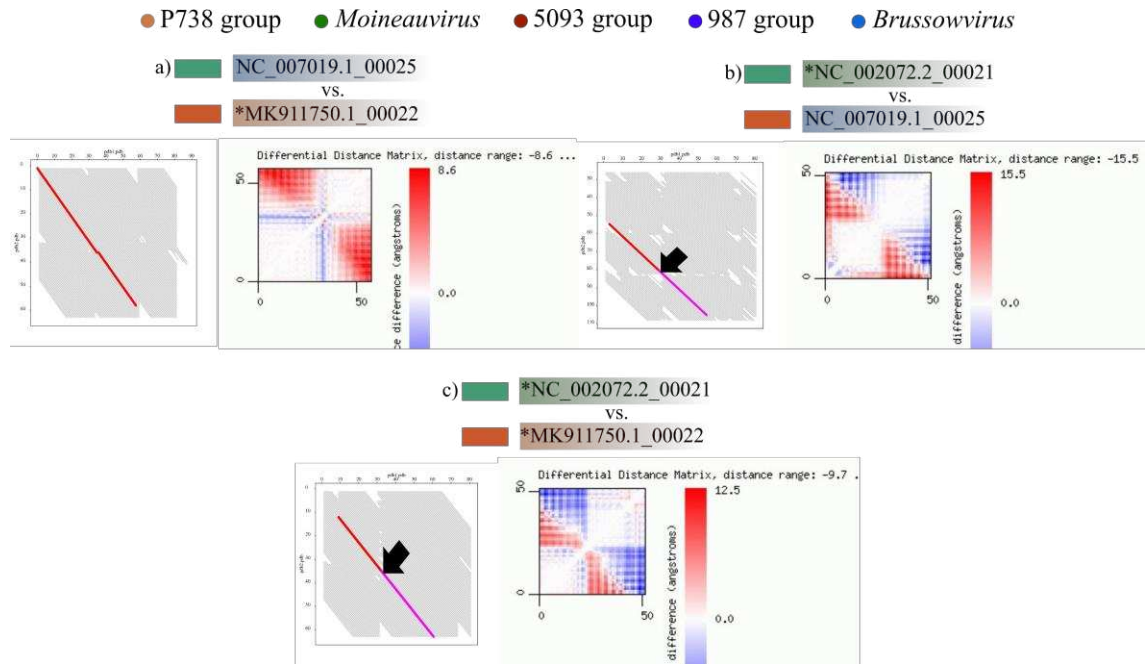


Figure S2 - Graphic representation and differential distance matrix decomposition of the pairwise alignment of representative holins encoded by *Streptococcus thermophilus* phages. Holin superfamily VI vs. Phage r1t holin family (a). Holin superfamily I vs. Holin superfamily VI (b). Holin superfamily I vs. Phage r1t holin family (c).

Table S1 - Summary of information retrieved from GenBank, VIRIDIC, and CoreGenes of the *Streptococcus thermophilus* phages of the P738 group.

Phage name	INSDC	Size (Kb)	GC%	Protein	Overall DNA sequence identity (*)	Overall % homologous proteins (**)
Streptococcus phage P738	MK911750.1	34.04	37.1	48	100	100
Streptococcus phage D4446	MN938931.1	33.66	37.0	46	90.715	93.75

(*) Determined using VIRIDIC (Moraru et al., 2020), compared to *Streptococcus* phage P738.
 (**) Determined using CoreGenes 5.0 (Zafar et al., 2002), compared to *Streptococcus* phage P738.

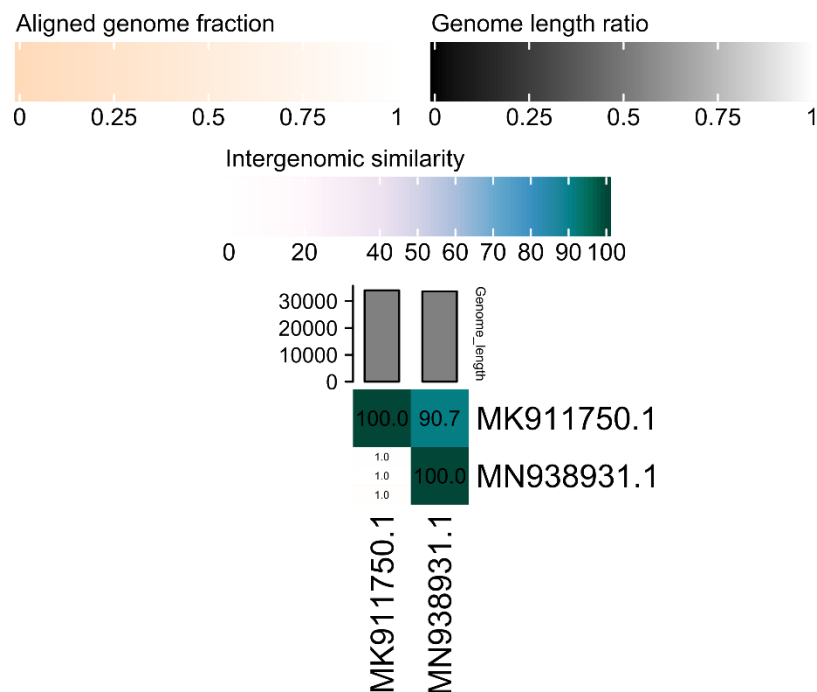


Figure S3 - VIRIDIC heat map of the distribution of the two genomes the *Streptococcus thermophilus* phages of the P738 group into two species based on intergenomic similarity values (right half) and alignment indicators (left half and top bars). In the right half of the heatmap, the color-coding allows for quick visualization of the grouping of phage genomes based on intergenomic similarity, since the higher the intergenomic similarity value between the phages, the darker the color. The numbers represent the similarity values for each genome pair. In the left half, three indicator values are represented for each genome pair, in order from top to bottom: genome aligned fraction 1 (for the genome located in this row), genome length ratio (for the two genomes in this pair) and genome aligned fraction 2 (for the genome located in this column). Darker colors emphasize low values, indicating pairs of genomes where only a small fraction of the genome was aligned (orange to white color gradient), or that there is a large difference in the length of the two genomes (black to white color gradient).

Table S2 - Summary of information retrieved from GenBank, VIRIDIC, and CoreGenes of the *Streptococcus thermophilus* phages species belonging to the *Moineauvirus* genus.

Phage name	RefSeq No.	INSDC	Size (Kb)	GC %	Protein	Overall DNA sequence identity (*)	Overall % homologous proteins (**)
Streptococcus phage SW6		MH8923 51.1	34.28	38.6	42	70.653	73.33
Streptococcus phage SW30		MH8923 75.1	34.16	38.5	41	71.166	73.33
Streptococcus phage CHPC1156		MH9375 08.1	34.91	39.0	44	72.626	84.44
Streptococcus phage D1024		MH0006 03.1	37.76	38.1	49	71.291	91.11
Streptococcus phage D1811		MH0006 04.1	34.08	39.0	40	68.231	75.56
Streptococcus phage P5641		KY7052 59.1	37.24	38.8	52	61.536	68.89
Streptococcus phage CHPC1062		MH9374 99.1	40.04	38.5	54	59.137	66.67
Streptococcus phage D4276		MF1613 28.1	39.71	38.0	54	58.616	68.89
Streptococcus phage CHPC877		MH9374 67.1	39.97	38.4	52	58.97	66.67
Streptococcus phage CHPC927		MH9374 71.1	37.3	39.0	48	62.816	71.11
Streptococcus phage CHPC930		MH9374 74.1	36.35	38.6	46	59.577	66.67
Streptococcus phage P7574		KY7052 71.1	36.3	38.7	49	56.866	66.67
Streptococcus phage CHPC663		MH9374 62.1	38.53	38.9	52	57.894	71.11
Streptococcus phage M19		MF5807 70.1	34.33	38.9	44	63.328	66.67
Streptococcus phage SW3		MH8923 54.1	37.18	38.6	45	59.11	66.67
Streptococcus phage 7A5		MF5807 59.1	35.94	38.3	44	60.925	64.44
Streptococcus phage B0		MF5807 66.1	36.77	38.5	47	62.26	62.22
Streptococcus phage B5		MF5807 67.1	35.64	38.6	42	63.003	62.22
Streptococcus phage L5A1		MF5807 69.1	36.88	38.9	44	65.337	62.22
Streptococcus phage A0		MF5807 65.1	35.48	38.8	44	63.1	62.22
Streptococcus phage C0		MF5807 68.1	34.72	39.0	42	64.604	62.22

Streptococcus phage CHPC950		MH9374 77.1	35.3	39. 0	48	65.21 8	75.56
Streptococcus phage CHPC1014		MH9374 85.1	35.26	39. 0	47	65.25 3	75.56
Streptococcus phage CHPC1045		MH9374 95.1	34.1	38. 7	43	66.75 3	64.44
Streptococcus phage CHPC1046		MH9374 96.1	34.79	38. 7	46	64.59 9	71.11
Streptococcus phage CHPC1048		MH9374 97.1	34.81	38. 7	46	64.65 3	71.11
Streptococcus phage CHPC572		MH9374 58.1	37.01	38. 5	49	66.68 8	73.33
Streptococcus phage STP1		MF5807 73.1	35.28	38. 5	41	65.73 9	62.22
Streptococcus phage STP2		MF5807 74.1	35.29	38. 5	44	65.51 1	64.44
Streptococcus phage 7201	NC_002 185.1	AF1450 54.1	35.47	38. 7	46	54.24 2	66.67
Streptococcus phage CHPC879		MH9374 68.1	36.01	38. 2	45	57.85 6	62.22
Streptococcus phage CHPC925		MH9374 70.1	34.76	38. 5	45	63.44 3	71.11
Streptococcus phage P0091		KY7052 51.1	35.65	39. 2	44	56.72 8	62.22
Streptococcus phage vB SthS VA214		MG7082 74.1	38.25	38. 2	53	58.73	77.78
Streptococcus phage CHPC642		MH9374 61.1	35.71	38. 8	49	56.64 7	71.11
Streptococcus phage CHPC1036		MH9374 90.1	36.33	38. 7	48	54.80 7	66.67
Streptococcus phage SW2		MH8923 53.1	37.22	38. 8	47	57.74 5	71.11
Streptococcus phage SW5		MH9736 61.1	36.84	38. 8	47	57.89 7	71.11
Streptococcus phage Abc2	NC_013 645.1	FJ23631 0.1	34.88	39. 0	48	66.17 2	68.89
Streptococcus phage P9902		KY7052 89.1	35.53	38. 9	50	71.83 7	71.11
Streptococcus phage P9903		KY7052 90.1	35.26	38. 7	50	72.57 9	73.33
Streptococcus phage 73		KT7170 83.1	36.73	38. 8	46	65.68 7	71.11
Streptococcus phage P9901		KY7052 88.1	35.21	38. 9	47	64.89	71.11
Streptococcus phage CHPC875		MH9374 66.1	36.58	39. 0	49	69.63 9	73.33
Streptococcus phage CHPC979		MH9374 81.1	34.28	38. 7	47	65.91 5	73.33

Streptococcus phage CHPC1148		MH9375 06.1	39.07	38. 0	50	56.48 7	75.56
Streptococcus phage Sfi21	NC_000 872.1	AF1151 03.1	40.74	37. 6	50	64.97 6	75.56
Streptococcus phage CHPC919		MH9374 69.1	37.28	38. 3	46	72.21 8	77.78
Streptococcus phage Sfi19	NC_000 871.1	AF1151 02.1	37.37	38. 3	45	71.33 3	75.56
Streptococcus phage SW11		MH8923 60.1	36.25	38. 2	42	74.9	82.22
Streptococcus phage CHPC1091		MH9375 04.1	37.03	38. 4	46	74.10 9	86.67
Streptococcus phage CHPC873		MH9374 65.1	38.26	38. 1	51	72.54	82.22
Streptococcus phage P7631		KY7052 74.1	36.16	38. 9	48	77.62 9	84.44
Streptococcus phage P7633		KY7052 76.1	35.92	39. 0	47	77.20 3	82.22
Streptococcus phage P7632		KY7052 75.1	35.99	39. 1	47	77.53 2	82.22
Streptococcus phage SW8		MH8923 57.1	33.93	39. 1	41	79.61 7	80
Streptococcus phage 128		KT7170 85.1	34.59	39. 0	40	78.65 1	80
Streptococcus phage D5842		MH0006 02.1	36.5	39. 1	43	78.20 3	82.22
Streptococcus phage SW7		MH9736 62.1	35.92	39. 1	42	76.95 7	82.22
Streptococcus phage SWK1		MH8923 77.1	36.09	39. 2	43	77.29 2	82.22
Streptococcus phage SWK2		MH8923 78.1	36.59	39. 3	42	77.34 7	82.22
Streptococcus phage 9A		MF5807 61.1	35.6	39. 0	41	77.93 2	84.44
Streptococcus phage SW21		MH8923 56.1	35.35	38. 8	39	76.60 7	80
Streptococcus phage SWK5		MH8923 81.1	35.85	38. 9	39	75.83 1	80
Streptococcus phage SWK6		MH8923 82.1	35	39. 0	38	74.86 2	80
Streptococcus phage SWK3		MH8923 79.1	35.44	38. 7	41	74.87 8	80
Streptococcus phage SWK4		MH8923 80.1	35.34	38. 7	41	74.95 3	80
Streptococcus phage P7152		KY7052 66.1	35.66	38. 9	46	80.01 9	84.44
Streptococcus phage P8921		KY7052 82.1	34.52	39. 0	46	80.66 1	86.67

Streptococcus phage P7151		KY7052 65.1	35.03	38. 8	47	80.65 4	88.89
Streptococcus phage P7154		KY7052 67.1	34.51	38. 8	42	77.04	82.22
Streptococcus phage P8922		KY7052 83.1	34.51	38. 8	43	76.97 3	82.22
Streptococcus phage P5651		KY7052 60.1	34.29	39. 1	43	83.35 3	88.89
Streptococcus phage P5652		KY7052 61.1	33.45	39. 3	43	81.72 4	86.67
Streptococcus phage SW1		MH8923 52.1	34.82	38. 9	38	75.14 4	71.11
Streptococcus phage SW9		MH8923 58.1	34	38. 7	38	77.06 7	73.33
Streptococcus phage 9B4		MF5807 62.1	34.64	38. 7	41	80.52 9	77.78
Streptococcus phage 16B8		MF5807 63.1	34.38	38. 6	38	79.88 3	77.78
Streptococcus phage 7T		MF5807 60.1	35.14	38. 7	41	80.18 8	80
Streptococcus phage 31B4		MF5807 64.1	36.44	38. 9	40	82.04	77.78
Streptococcus phage SW10		MH8923 59.1	34.34	38. 8	39	80.02 2	77.78
Streptococcus phage R1		MF5807 72.1	34.65	38. 7	39	80.91 2	77.78
Streptococcus phage V2		MF5807 75.1	34.7	38. 7	40	80.73 2	77.78
Streptococcus phage 53		KT7170 84.1	34.24	38. 9	41	80.38 5	80
Streptococcus phage CHPC595		MH9374 59.1	35.7	39. 1	46	87.98 6	95.56
Streptococcus phage DT1	NC_002 072.1 NC_002 072.2	AF0852 22.2	34.82	39. 1	45	100	100
Streptococcus phage CHPC928		MH9374 72.1	34.02	38. 8	39	77.76 3	80
Streptococcus phage CHPC1029		MH9374 87.1	35.92	39. 1	45	78.41 8	86.67
Streptococcus phage CHPC1040		MH9374 92.1	35.85	38. 9	45	76.77	84.44
Streptococcus phage P3681		KY7052 56.1	36.17	38. 4	45	74.84	86.67
Streptococcus phage P3684		KY7052 57.1	35.44	38. 9	44	77.29 4	88.89
Streptococcus phage P7601		KY7052 72.1	35.4	38. 5	51	80.15 7	82.22

Streptococcus phage P7602		KY7052 73.1	37.2	38. 7	51	75.83 6	82.22
Streptococcus phage CHPC954		MH9374 80.1	37.46	38. 7	45	75.55 6	80
Streptococcus phage P7133		KY7052 63.1	33.28	38. 8	42	85.96 3	82.22
Streptococcus phage MM25		MF5807 71.1	34.06	39. 0	41	84.01 1	82.22
Streptococcus phage CHPC1005		MH9374 83.1	37.6	38. 5	47	81.98	88.89
Streptococcus phage CHPC1027		MH9374 86.1	35.93	38. 9	46	81.79 2	86.67
Streptococcus phage CHPC1033		MH9374 88.1	36.83	38. 5	46	79.11 7	84.44
Streptococcus phage P7132		KY7052 62.1	35.84	39. 0	44	84.06	82.22
Streptococcus phage CHPC1067		MH9375 00.1	34.35	39. 0	43	83.57 7	82.22
Streptococcus phage P7134		KY7052 64.1	34.02	38. 9	45	84.39 9	84.44
Streptococcus phage CHPC1073		MH9375 01.1	34.02	38. 9	45	84.35 8	84.44
Streptococcus phage vB SthS VA460		MG7082 75.1	41.17	38. 8	56	68.32 9	86.67
Streptococcus phage CHPC1041		MH9374 93.1	38.84	38. 1	51	68.90 2	82.22
Streptococcus phage P9851		KY7052 84.1	35.84	39. 1	48	65.30 6	80
Streptococcus phage P9854		KY7052 87.1	36.98	38. 9	48	62.40 8	75.56
Streptococcus phage P7573		KY7052 70.1	37.71	38. 7	48	67.07 3	82.22
Streptococcus phage SW12		MH8923 61.1	35.38	38. 8	43	79.19 4	86.67

(*) Determined using VIRIDIC (Moraru et al., 2020), compared to *Streptococcus* phage DT1.

(**) Determined using CoreGenes 5.0 (Zafar et al., 2002), compared to *Streptococcus* phage DT1.

Table S3 - Strains of *Streptococcus thermophilus* phages in the *Moineauvirus* genus.

Name	Accession No.	Strain of:
Streptococcus phage SW30	MH892375.1	Streptococcus phage SW6
Streptococcus phage B5	MF580767.1	Streptococcus phage B0
Streptococcus phage A0	MF580765.1	Streptococcus phage L5A1
Streptococcus phage C0	MF580768.1	Streptococcus phage L5A1
Streptococcus phage CHPC1014	MH937485.1	Streptococcus phage CHPC950
Streptococcus phage CHPC1048	MH937497.1	Streptococcus phage CHPC1046
Streptococcus phage STP2	MF580774.1	Streptococcus phage STP1

Streptococcus phage D5842	MH000602.1	Streptococcus phage 128
Streptococcus phage SWK1	MH892377.1	Streptococcus phage SW7
Streptococcus phage SWK2	MH892378.1	Streptococcus phage SW7
Streptococcus phage SW21	MH892356.1	Streptococcus phage 9A
Streptococcus phage SWK5	MH892381.1	Streptococcus phage 9A
Streptococcus phage SWK4	MH892380.1	Streptococcus phage SWK3
Streptococcus phage P8922	KY705283.1	Streptococcus phage P7154
Streptococcus phage P5652	KY705261.1	Streptococcus phage P5651
Streptococcus phage 16B8	MF580763.1	Streptococcus phage 9B4
Streptococcus phage 31B4	MF580764.1	Streptococcus phage 7T
Streptococcus phage SW10	MH892359.1	Streptococcus phage 7T
Streptococcus phage R1	MF580772.1	Streptococcus phage 7T
Streptococcus phage V2	MF580775.1	Streptococcus phage 7T
Streptococcus phage CHPC1040	MH937492.1	Streptococcus phage CHPC1029
Streptococcus phage CHPC954	MH937480.1	Streptococcus phage P7602
Streptococcus phage CHPC1073	MH937501.1	Streptococcus phage P7134

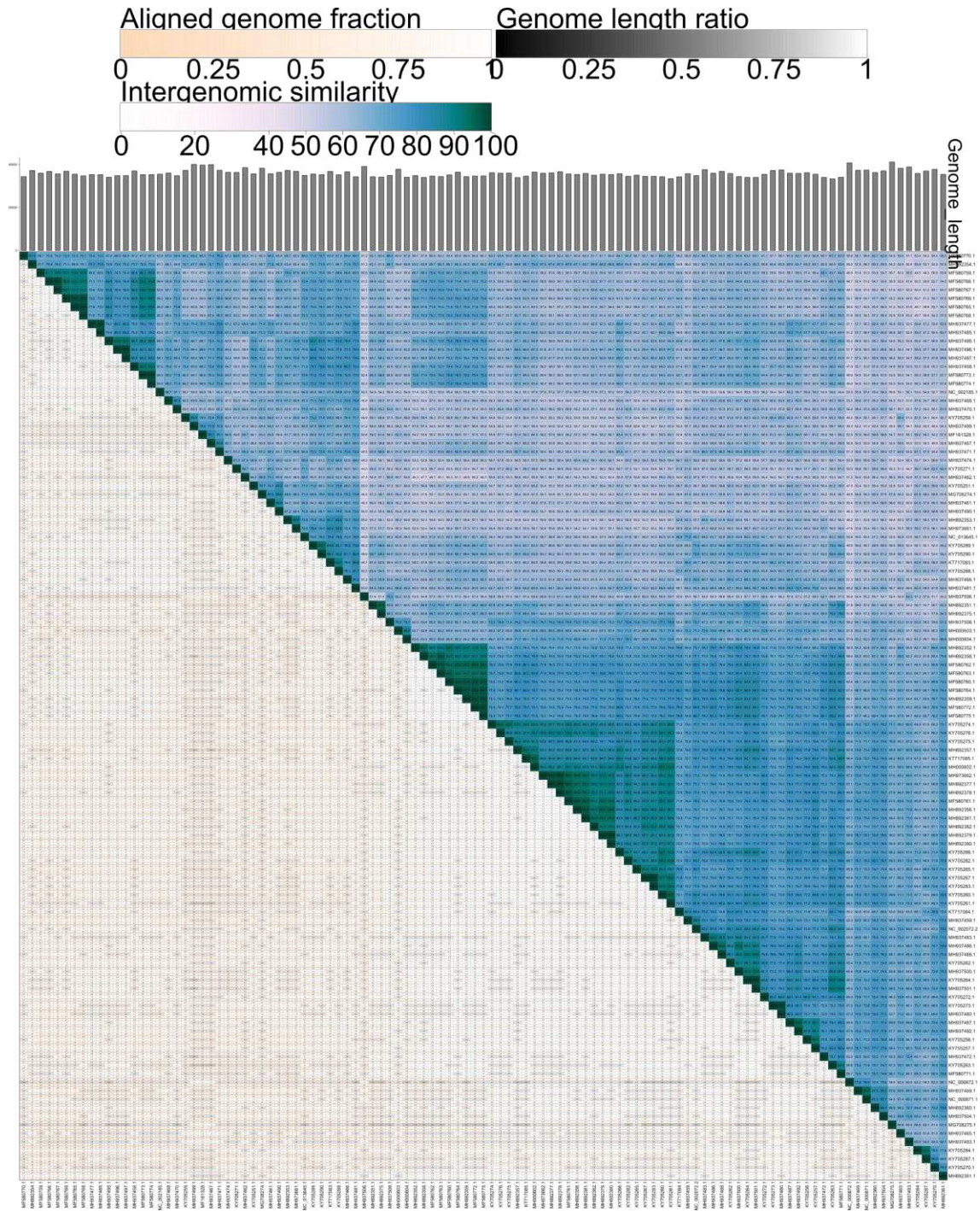


Figure S4 - VIRIDIC heat map of the distribution of the 109 genomes the *Streptococcus thermophilus* phages of the *Moineauvirus* genus into 86 species based on intergenomic similarity values (right half) and alignment indicators (left half and top bars). In the right half of the heatmap, the color-coding allows for quick visualization of the grouping of phage genomes based on intergenomic similarity, since the higher the intergenomic similarity value between the phages, the darker the color. The numbers represent the similarity values for each genome pair. In the left half, three indicator values are represented for each genome pair, in order from top to bottom: genome aligned fraction 1 (for the genome located in this row), genome length ratio (for the two genomes in this pair) and genome aligned fraction 2 (for the genome located in this column). Darker colors emphasize low values, indicating pairs of genomes where only a small fraction of the genome was aligned (orange to white color

gradient), or that there is a large difference in the length of the two genomes (black to white color gradient). See SI file 6 for high resolution version.

Table S4 - Summary of information retrieved from GenBank, VIRIDIC, and CoreGenes of the *Streptococcus thermophilus* phages species belonging to the *Brussowvirus* genus.

Phage name	RefSeq No.	INSDC	Size (Kb)	GC %	Protein	Overall DNA sequence identity (*)	Overall % homologous proteins (**)
Streptococcus phage TP-J34	NC_020197.1	HE861935.1	45.61	38.8	60	63.712	73.68
Streptococcus phage CHPC929		MH937473.1	40.87	38.5	58	68.539	77.19
Streptococcus phage TP-778L	NC_022776.1	HG380752.1	41.76	39.0	52	63.689	77.19
Streptococcus phage Sfi11	NC_002214.1	AF158600.2	39.81	38.6	51	81.115	82.46
Streptococcus phage O1205	NC_004303.1	U88974.1	43.08	38.3	57	100	100
Streptococcus phage CHPC640		MH937460.1	40.4	38.8	55	67.772	71.93
Streptococcus phage CHPC676		MH937463.1	40.4	38.8	53	67.718	71.93
Streptococcus phage P7954		KY705280.1	35.69	38.9	46	66.609	63.16
Streptococcus phage P7952		KY705278.1	37.59	39.0	48	63.285	63.16
Streptococcus phage P7953		KY705279.1	35.79	39.1	43	65.749	59.65
Streptococcus phage SW13		MH892362.1	36.59	39.2	41	64.352	59.65
Streptococcus phage SW33		MH892385.1	36.59	39.2	41	64.319	59.65
Streptococcus phage SW31		MH892383.1	36.59	39.2	41	64.319	59.65
Streptococcus phage SW32		MH892384.1	36.59	39.2	41	64.319	59.65
Streptococcus phage 858	NC_010353.1	EF529515.1	35.54	39.8	46	53.091	64.91
Streptococcus phage P9853		KY705286.1	34.47	39.7	45	58.222	63.16
Streptococcus phage P9852		KY705285.1	34.64	40.0	44	59.99	61.4
Streptococcus phage 2972	NC_007019.1	AY699705.1	34.7	40.2	44	59.169	63.16

Streptococcus phage CHPC1008		MH937484 .1	34.8 4	40.1	45	59.145	63.16
Streptococcus phage CHPC1057		MH937498 .1	34.8 5	40.1	45	59.149	63.16
Streptococcus phage ALQ13.2	NC_013598 .1	FJ226752. 1	35.5 3	39.4	44	57.827	59.65
Streptococcus phage SW14		MH892363 .1	35.9 1	39.9	46	58.814	68.42
Streptococcus phage SW1151		MH892376 .1	35.0 6	39.6	45	56.902	68.42
Streptococcus phage CHPC1152		MH937507 .1	35.3 5	39.6	44	58.621	64.91
Streptococcus phage CHPC951		MH937478 .1	36.4 7	39.3	47	60.349	64.91
Streptococcus phage CHPC1083		MH937502 .1	36.4 7	39.3	45	60.364	63.16
Streptococcus phage SW15		MH892364 .1	34.2 9	39.4	44	56.935	64.91
Streptococcus phage SW18		MH892366 .1	34.8 3	39.4	44	59.879	63.16
Streptococcus phage P7571		KY705268 .1	34.9 1	39.4	49	61.074	66.67
Streptococcus phage P7572		KY705269 .1	33.4 1	39.6	43	59.54	59.65
Streptococcus phage P4761		KY705258 .1	38.1 4	39.0	55	55.414	57.89
Streptococcus phage P7951		KY705277 .1	34.1 2	39.8	47	47.036	54.39
Streptococcus phage P7955		KY705281 .1	36.9 3	38.9	50	55.006	56.14
Streptococcus phage 20617	NC_023503 .1	HG424323 .1	48.8	40.0	68	37.007	66.67
Streptococcus phage VS-2018a		CP029253. 1	39.8 8	39.5	54	50.889	63.16
Streptococcus phage CHPC952		MH937479 .1	36.7 8	39.4	46	45.287	50.88
Streptococcus phage CHPC1084		MH937503 .1	36.7 8	39.4	47	45.287	52.63

Streptococcus phage CHPC1109		MH937505.1	33.79	39.3	42	38.142	40.35
Streptococcus phage CHPC869		MH937464.1	37.08	39.1	50	43.79	52.63
Streptococcus phage CHPC931		MH937475.1	36.49	39.2	49	43.63	52.63
Streptococcus phage CHPC1230		MH937509.1	39.38	39.5	55	43.776	59.65
Streptococcus phage CHPC1246		MH937510.1	36.88	39.2	53	45.098	57.89
Streptococcus phage CHPC1042		MH937494.1	42.02	39.3	61	45.179	61.4
Streptococcus phage CHPC1248		MH937457.1	38.38	39.5	53	41.251	54.39
Streptococcus phage CHPC1247		MH937511.1	35.54	39.8	48	42.597	52.63

(*) Determined using VIRIDIC (Moraru et al., 2020), compared to *Streptococcus* phage O1205.

(**) Determined using CoreGenes 5.0 (Zafar et al., 2002), compared to *Streptococcus* phage O1205.

Table S5 - Strains of *Streptococcus thermophilus* phages in the *Brussowvirus* genus.

Name	Accession No.	Strain of:
Streptococcus phage CHPC676	MH937463.1	Streptococcus phage CHPC640
Streptococcus phage SW33	MH892385.1	Streptococcus phage SW13
Streptococcus phage SW31	MH892383.1	Streptococcus phage SW13
Streptococcus phage SW32	MH892384.1	Streptococcus phage SW13
Streptococcus phage P9852	KY705285.1	Streptococcus phage 2972
Streptococcus phage CHPC1008	MH937484.1	Streptococcus phage 2972
Streptococcus phage CHPC1057	MH937498.1	Streptococcus phage 2972
Streptococcus phage CHPC1083	MH937502.1	Streptococcus phage CHPC951
Streptococcus phage P7572	KY705269.1	Streptococcus phage P7571
Streptococcus phage CHPC1084	MH937503.1	Streptococcus phage CHPC952

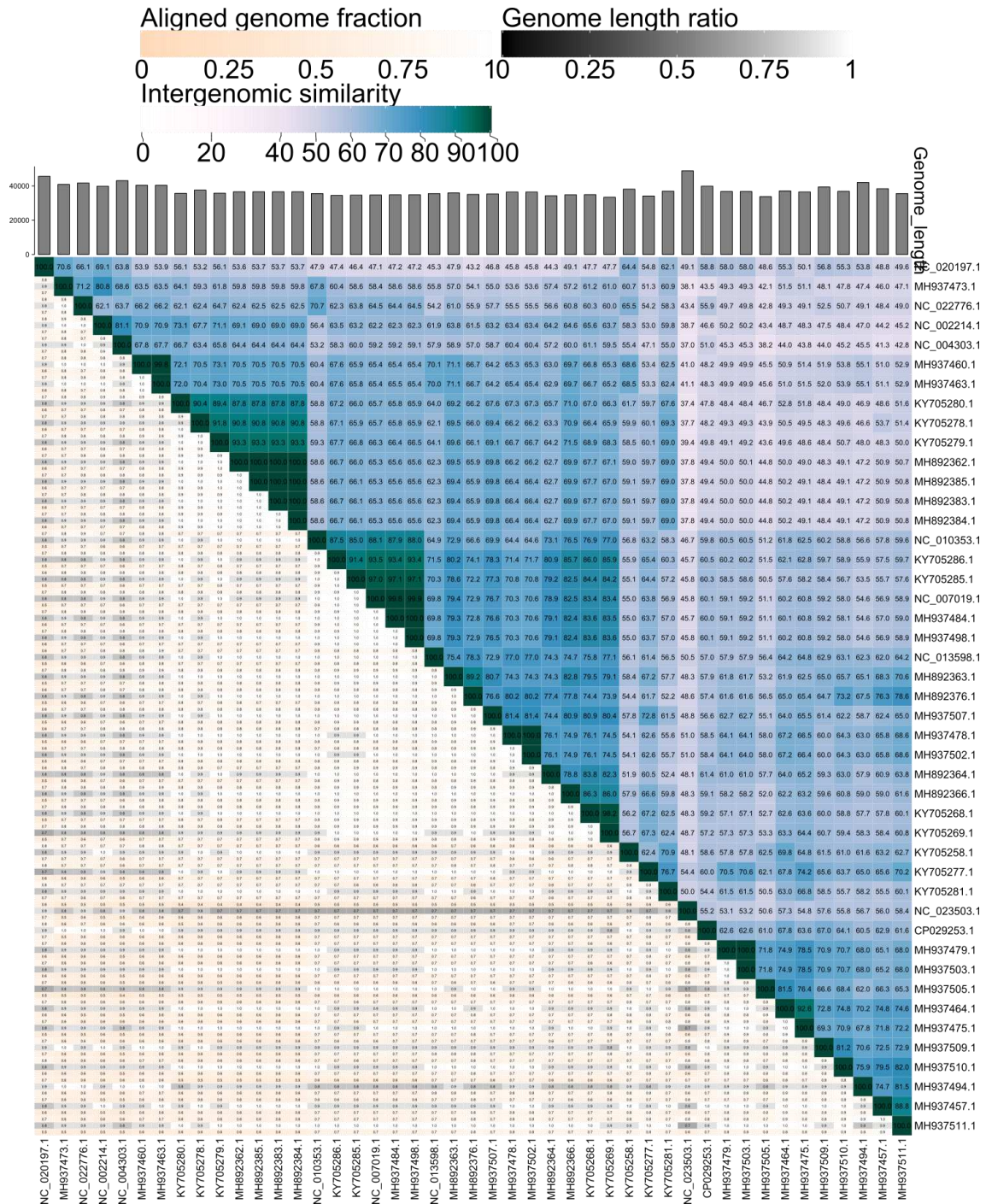


Figure S5 - VIRIDIC heat map of the distribution of the 45 genomes the *Streptococcus thermophilus* phages of the *Brussowvirus* genus into 35 species based on intergenomic similarity values (right half) and alignment indicators (left half and top bars). In the right half of the heatmap, the color-coding allows for quick visualization of the grouping of phage genomes based on intergenomic similarity, since the higher the intergenomic similarity value between the phages, the darker the color. The numbers represent the similarity values for each genome pair. In the left half, three indicator values are represented for each genome pair, in order from top to bottom: genome aligned fraction 1 (for the genome located in this row), genome length ratio (for the two genomes in this pair) and genome aligned fraction 2 (for the genome located in this column). Darker colors emphasize low values, indicating pairs of genomes where only a small fraction of the genome was aligned (orange to white color

gradient), or that there is a large difference in the length of the two genomes (black to white color gradient). See SI file 6 for high resolution version.

Table S6 - Summary of information retrieved from GenBank, VIRIDIC, and CoreGenes of the *Streptococcus thermophilus* phages of the 987 group.

Phage name	INSDC	Size (Kb)	GC%	Protein	Overall DNA sequence identity (*)	Overall % homologous proteins (**)
Streptococcus phage CHPC577	KX879641.1	35.13	36.8	50	70.098	69.39
Streptococcus virus 9871	KU678389.1	32.73	37.1	49	100	100
Streptococcus virus 9872	KU678390.1	33.11	36.8	49	91.425	95.92
Streptococcus virus 9873	KU678391.1	32.81	36.9	48	91.532	95.92
Streptococcus phage SW22	MH892369.1	31.74	37.3	45	83.706	85.71
Streptococcus phage SW25	MH892371.1	32.01	37.2	46	83.132	87.76
Streptococcus phage SW16	MH892350.1	32.11	37.0	45	83.93	83.67
Streptococcus phage SW28	MH892374.1	32.09	36.9	45	85.311	87.76
Streptococcus phage SW17	MH892365.1	32	36.8	46	85.374	87.76
Streptococcus phage SW20	MH892368.1	32.57	36.8	46	85.636	85.71
Streptococcus phage SW23	MH892370.1	31.42	37.0	45	86.736	87.76
Streptococcus phage SW26	MH892372.1	31.28	37.0	44	85.676	83.67
Streptococcus virus 9874	KU678392.1	32.65	36.6	48	74.364	77.55
Streptococcus phage CHPC926	KX879642.1	30.08	37.0	44	75.588	79.59

(*) Determined using VIRIDIC (Moraru et al., 2020), compared to *Streptococcus* virus 9871.

(**) Determined using CoreGenes 5.0 (Zafar et al., 2002), compared to *Streptococcus* virus 9871.

Table S7 - Strains of *Streptococcus thermophilus* phages in the 987 group.

Name	Accession No.	Strain of:
Streptococcus virus 9873	KU678391.1	Streptococcus virus 9872
Streptococcus phage SW25	MH892371.1	Streptococcus phage SW22
Streptococcus phage SW28	MH892374.1	Streptococcus phage SW16
Streptococcus phage SW17	MH892365.1	Streptococcus phage SW16
Streptococcus phage SW20	MH892368.1	Streptococcus phage SW16

Streptococcus phage SW23	MH892370.1	Streptococcus phage SW16
Streptococcus phage SW26	MH892372.1	Streptococcus phage SW16

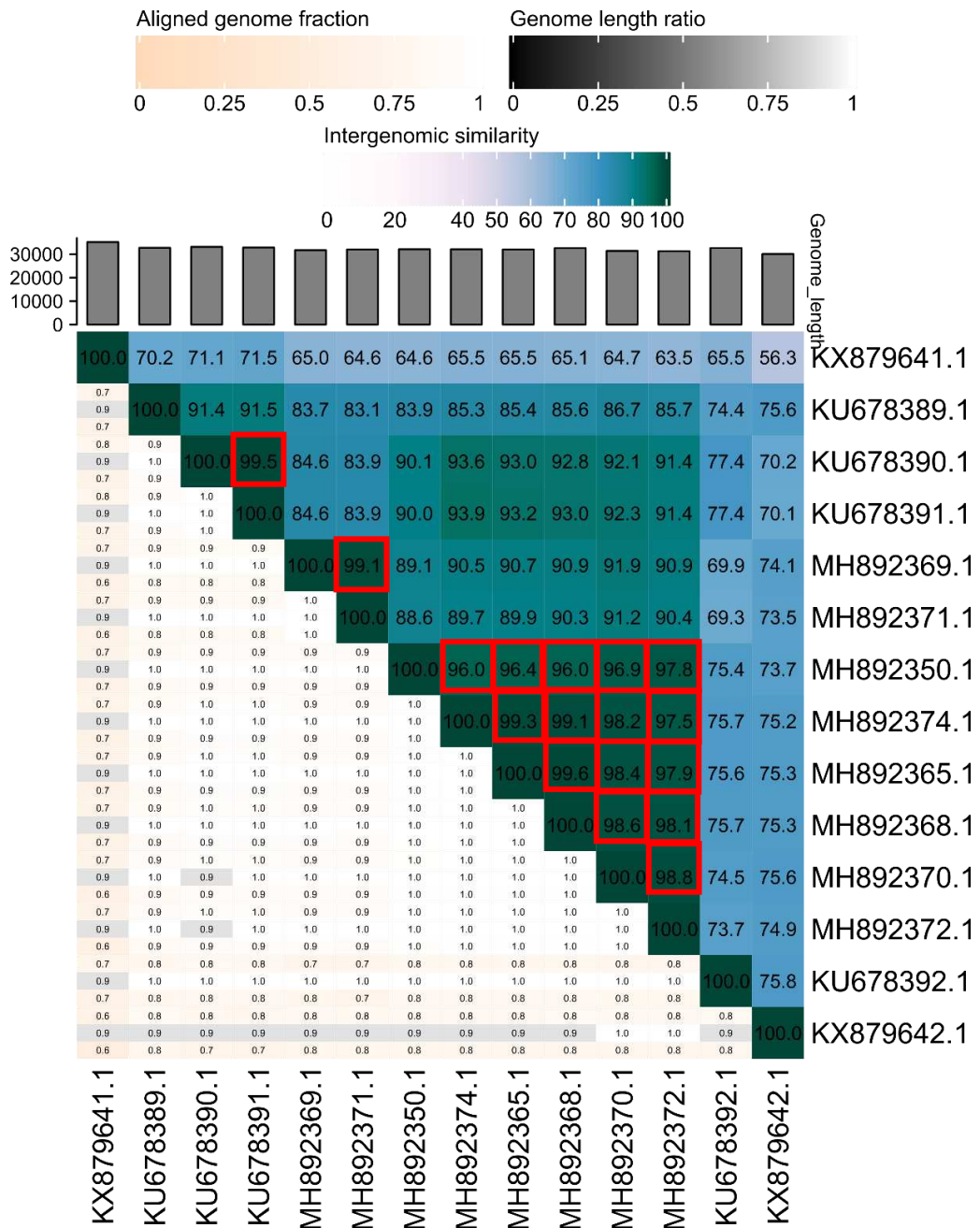


Figure S6 - VIRIDIC heat map of the distribution of the 14 genomes the *Streptococcus thermophilus* phages of the 987 group into seven species based on intergenomic similarity values (right half) and alignment indicators (left half and top bars). In the right half of the heatmap, the color-coding allows for quick visualization of the grouping of phage genomes based on intergenomic similarity, since the higher the intergenomic similarity value between the phages, the darker the color. The numbers represent the similarity values for each genome pair. In the left half, three indicator values are represented for each genome pair, in order from top to bottom: genome aligned fraction 1 (for the genome located in this row), genome length ratio (for the two genomes in this pair) and genome aligned fraction 2 (for the genome located in this column). Darker colors emphasize low values, indicating pairs of genomes where only a small fraction of the genome was aligned (orange to white color gradient), or that there is a

large difference in the length of the two genomes (black to white color gradient). The red box delineates strains.

Table S8 - Summary of information retrieved from GenBank, VIRIDIC, and CoreGenes of the *Streptococcus thermophilus* phages of the 5093 group.

Phage name	RefSeq No.	INSDC	Size (Kb)	GC %	Protein	Overall DNA sequence identity (*)	Overall % homologous proteins (**)
Streptococcus phage 5093	NC_012753.1	FJ965538.1	37.18	38.0	48	100	100
Streptococcus phage P0095		KY705255.1	35.97	38.1	54	75.591	72.92
Streptococcus phage CHPC1232		MK202160.1	31.97	38.3	45	77.348	70.83
Streptococcus phage SW27		MH892373.1	33.95	38.2	46	79.076	70.83
Streptococcus phage SW24		MH973663.1	32.09	38.3	42	82.063	72.92
Streptococcus phage CHPC1151		KX879643.1	33.51	38.5	46	78.685	72.92
Streptococcus phage P0092		KY705252.1	34.58	38.4	50	79.363	72.92
Streptococcus phage P0093		KY705253.1	34.93	38.6	51	77.801	70.83
Streptococcus phage P0094		KY705254.1	33.33	38.5	47	79.757	68.75
Streptococcus phage CHPC1282		MK202161.1	34.79	38.1	45	64.44	52.08
Streptococcus phage SW19		MH892367.1	35.15	38.1	46	66.624	56.25
Streptococcus phage SW4		MH892355.1	34.56	38.2	47	71.759	68.75
Streptococcus phage CHPC1198		MK202159.1	33.24	38.1	47	74.486	70.83

(*) Determined using VIRIDIC (Moraru et al., 2020), compared to *Streptococcus* phage 5093.

(**) Determined using CoreGenes 5.0 (Zafar et al., 2002), compared to *Streptococcus* phage 5093.

Table S9 - Strains of *Streptococcus thermophilus* phages in the 5093 group.

Name	Accession No.	Strain of:
Streptococcus phage P0094	KY705254.1	Streptococcus phage P0093

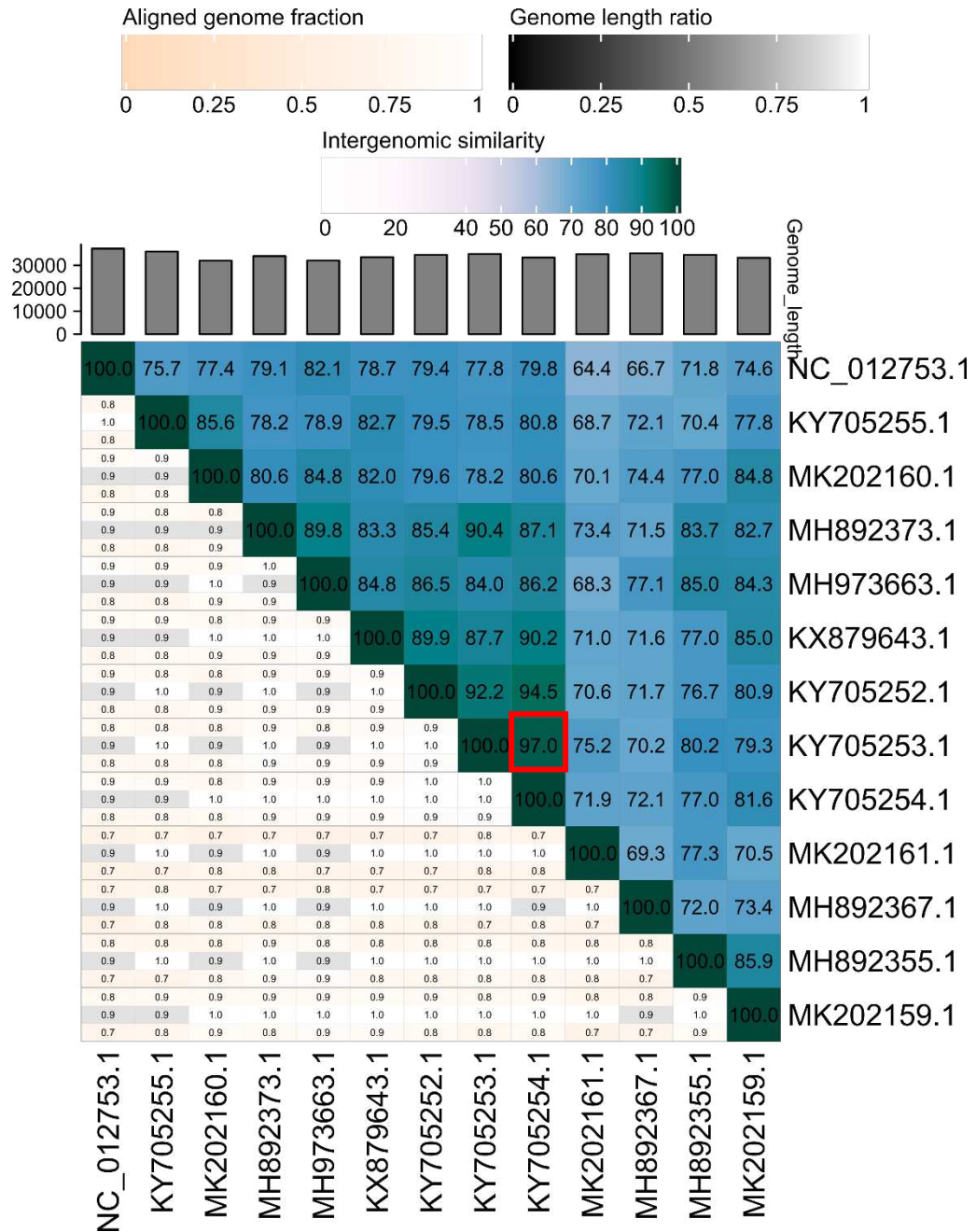
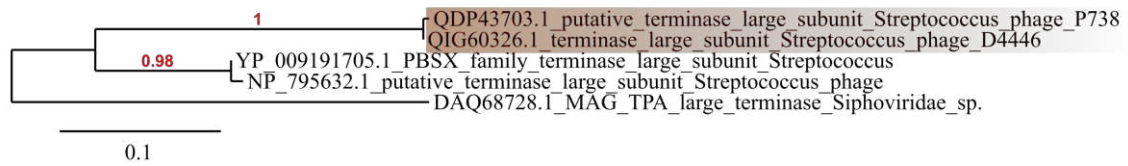


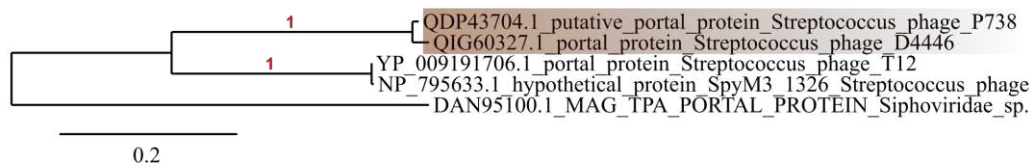
Figure S7 - VIRIDIC heat map of the distribution of the 13 genomes the *Streptococcus thermophilus* phages of the 5093 group into 12 species based on intergenomic similarity values (right half) and alignment indicators (left half and top bars). In the right half of the heatmap, the color-coding allows for quick visualization of the grouping of phage genomes based on intergenomic similarity, since the higher the intergenomic similarity value between the phages, the darker the color. The numbers represent the similarity values for each genome pair. In the left half, three indicator values are represented for each genome pair, in order from top to bottom: genome aligned fraction 1 (for the genome located in this row), genome length ratio (for the two genomes in this pair) and genome aligned fraction 2 (for the genome located in this column). Darker colors emphasize low values, indicating pairs of genomes where only a small fraction of the genome was aligned (orange to white color gradient), or that there is a large difference in the length of the two genomes (black to white color gradient). The red box delineates strains.

● P738 group

Terminase Large Subunit (TerL)



Portal Protein



Holin

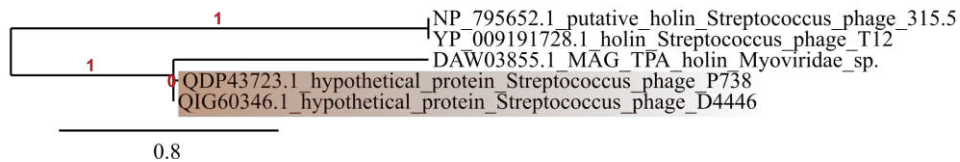


Figure S8 - Phylogenetic trees constructed for P738 group phages using the amino acid sequences of the terminase large subunit (TerL), portal, and holin proteins.

Moineauvirus

Terminase Small Subunit (TerS)

Major Tail Protein (MTP)

Portal Protein

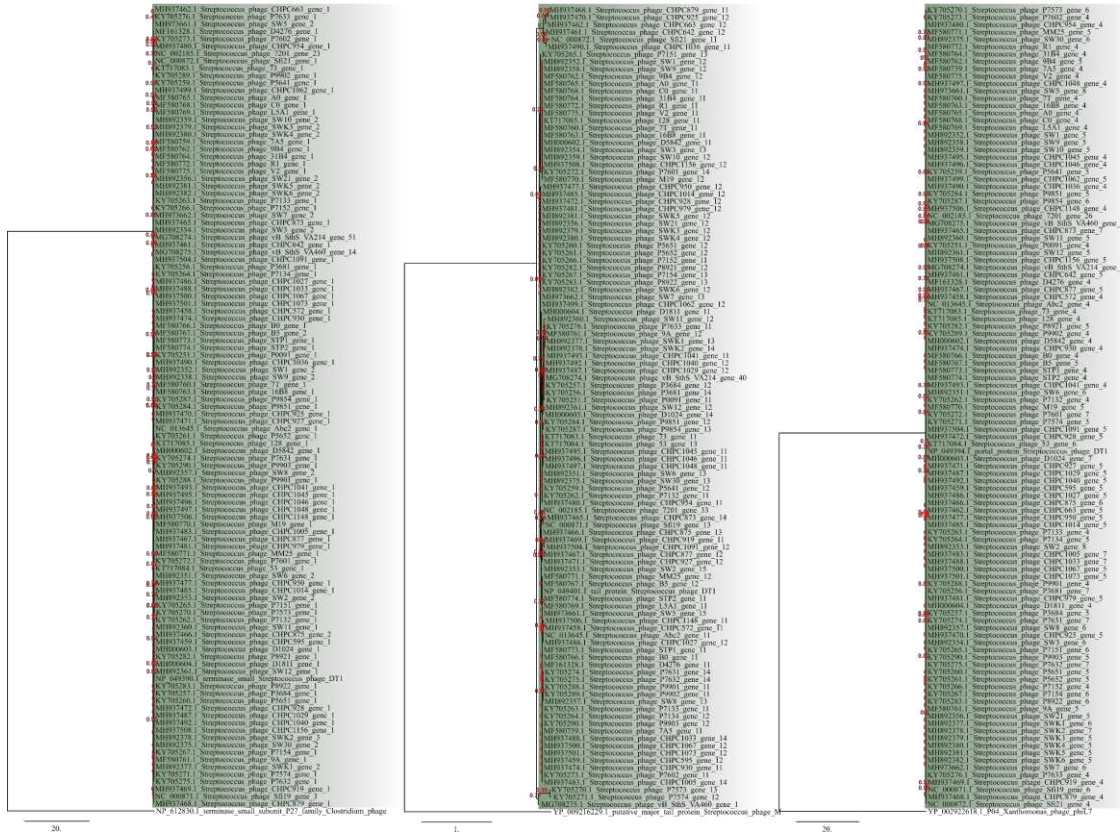
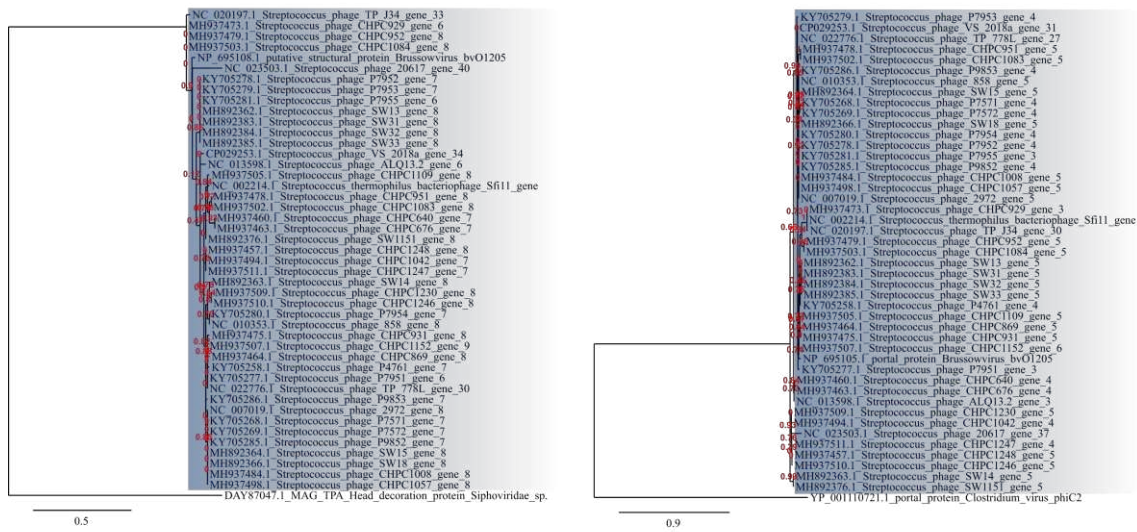


Figure S9 - Phylogenetic trees constructed for *Moineauvirus* genus phages using the amino acid sequences of the terminase small subunit (TerS), major tail protein (MTP), and portal proteins.

• *Brussowvirus*

Major Capsid Protein (MCP)

Portal Protein



Major Tail Protein (MTP)

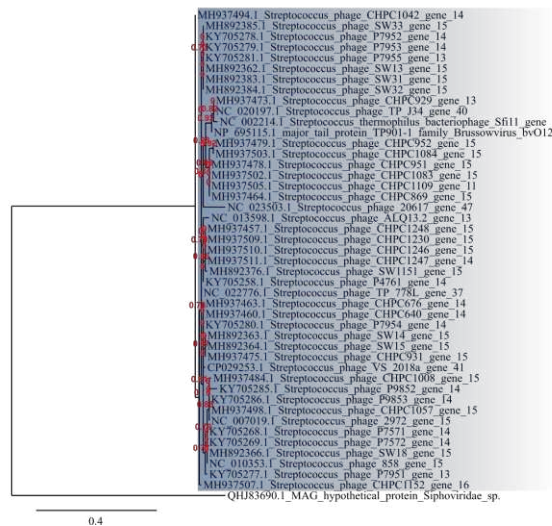
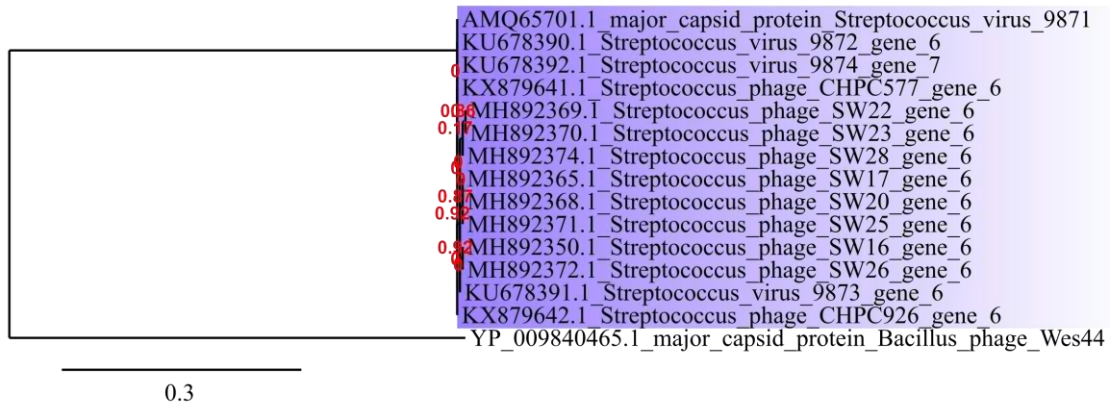


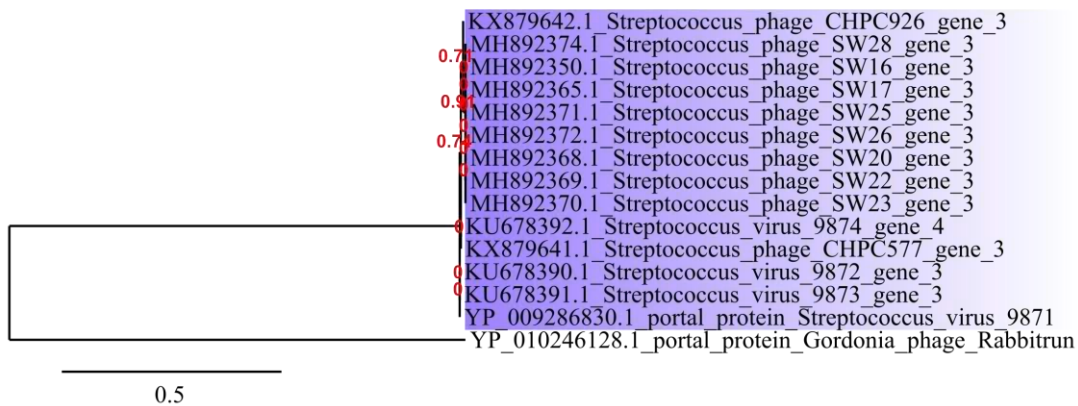
Figure S10 - Phylogenetic trees constructed for *Brussowvirus* genus phages using the amino acid sequences of the major capsid protein (MCP), portal protein, and major tail protein (MTP).

● 987 group

Major Capsid Protein (MCP)



Portal Protein



Major Tail Protein (MTP)

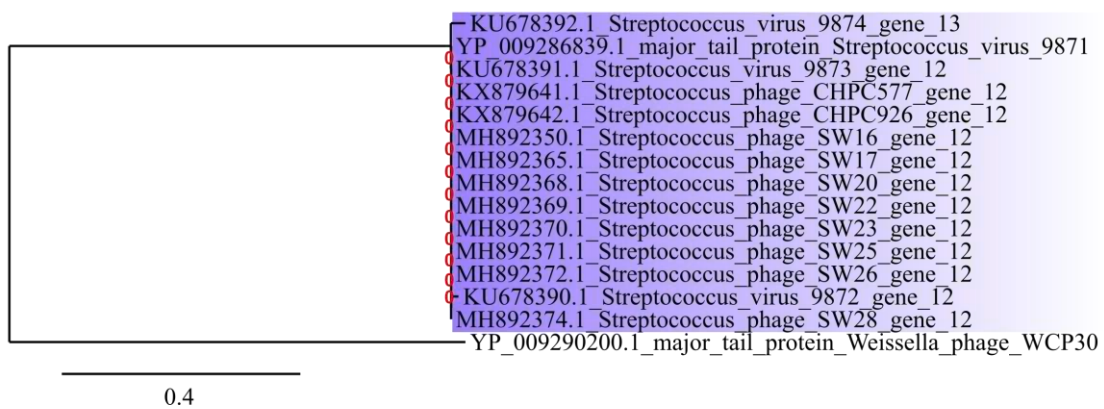
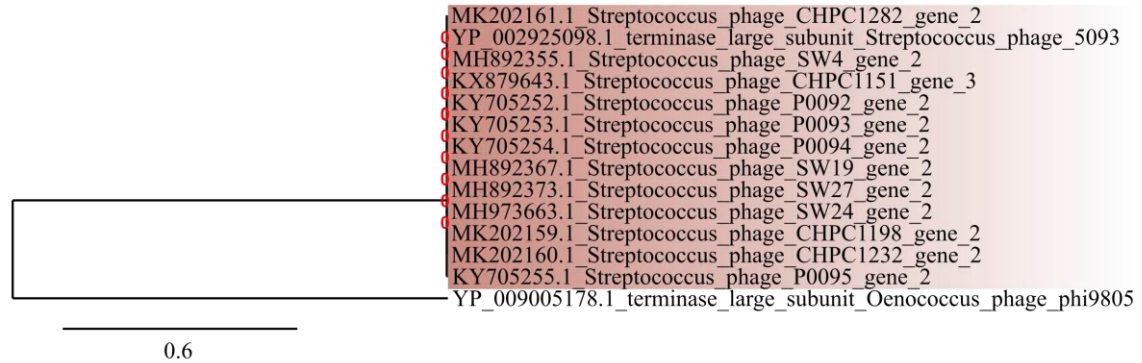


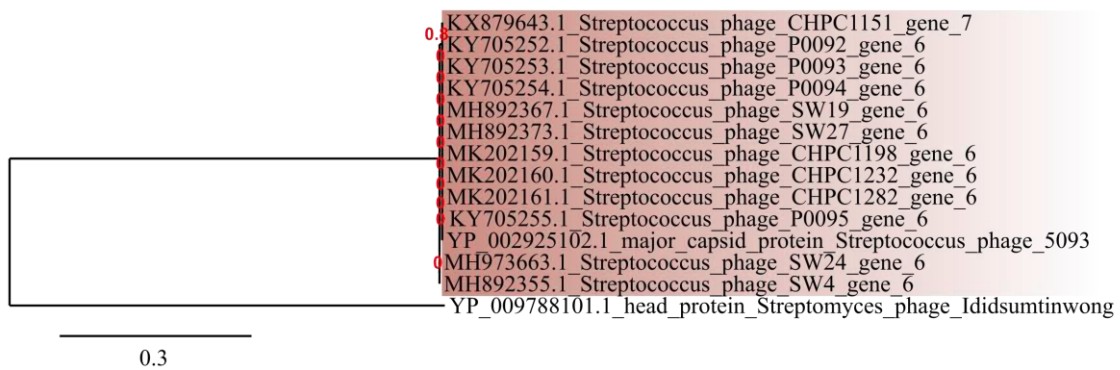
Figure S11 - Phylogenetic trees constructed for 987 group phages using the amino acid sequences of the major capsid protein (MCP), portal protein, and major tail protein (MTP).

● 5093 group

Terminase Large Subunit (TerL)



Major Capsid Protein (MCP)



Major Tail Protein (MTP)

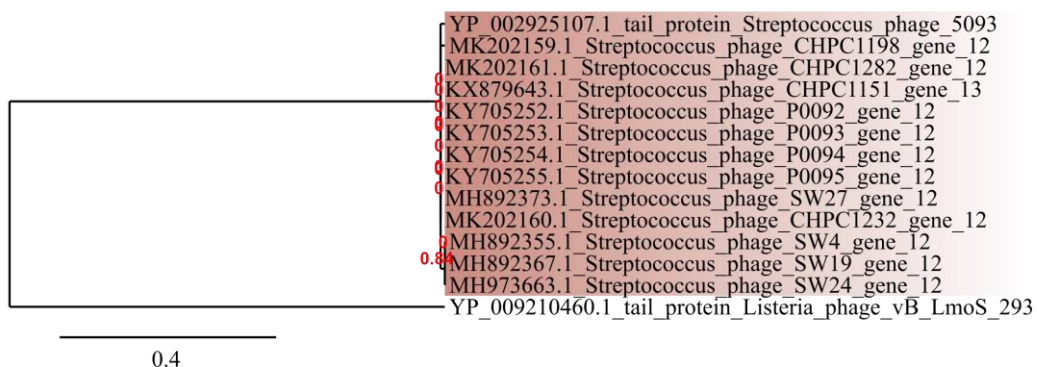


Figure S12 - Phylogenetic trees constructed for 5093 group phages using the amino acid sequences of the terminase large subunit (TerL), major capsid protein (MCP), and major tail protein (MTP).

● P738 group ● *Moineauvirus* ● 5093 group ● 987 group ● *Brussowvirus*

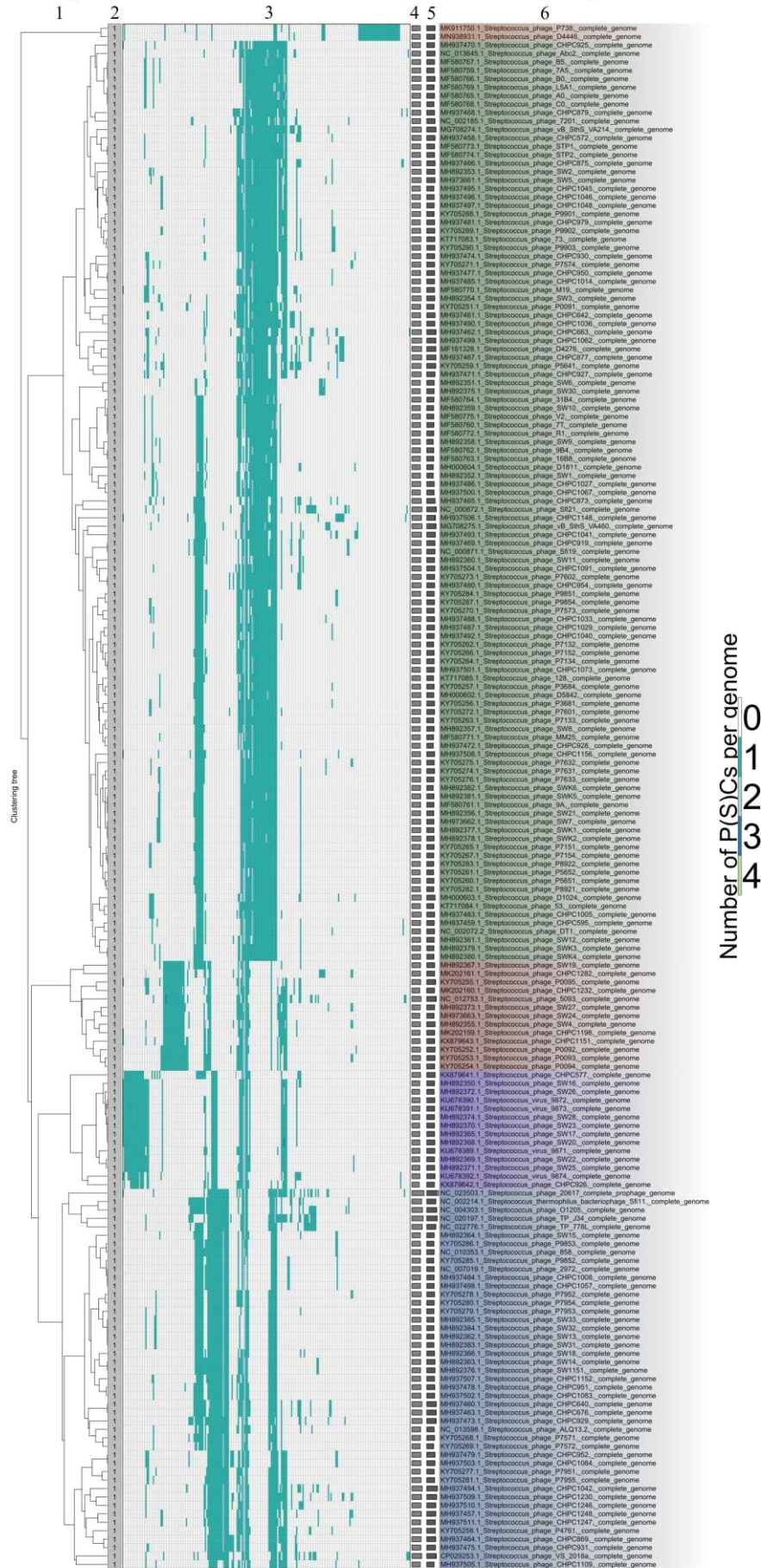


Figure S13 - VirClust hierarchical clustering of 183 phages that infect *Streptococcus thermophilus*, based on intergenomic distances calculated using the protein cluster content. 1) Hierarchical clustering tree. The tree was cut into smaller viral genome clusters (VGCs) using a 0.98 distance threshold. The VGC was proposed here as a new family. 2) VGC ID. 3) Heatmap representation of the PC distribution in the viral genomes. Rows are represented by individual viral genomes. Columns are represented by individual PCs. The ID of each PC can be read at the bottom of the heatmap. 4) Genome length (bps). 5) Fraction of proteins shared with other phages (dark gray), based on protein assignment to PCs. 6) Phage names. The colored rectangles overlapping the streptococcal phage names correspond to the genus-level grouping proposed here.

● P738 group ● *Moineauvirus* ● 5093 group ● 987 group ● *Brussowvirus*

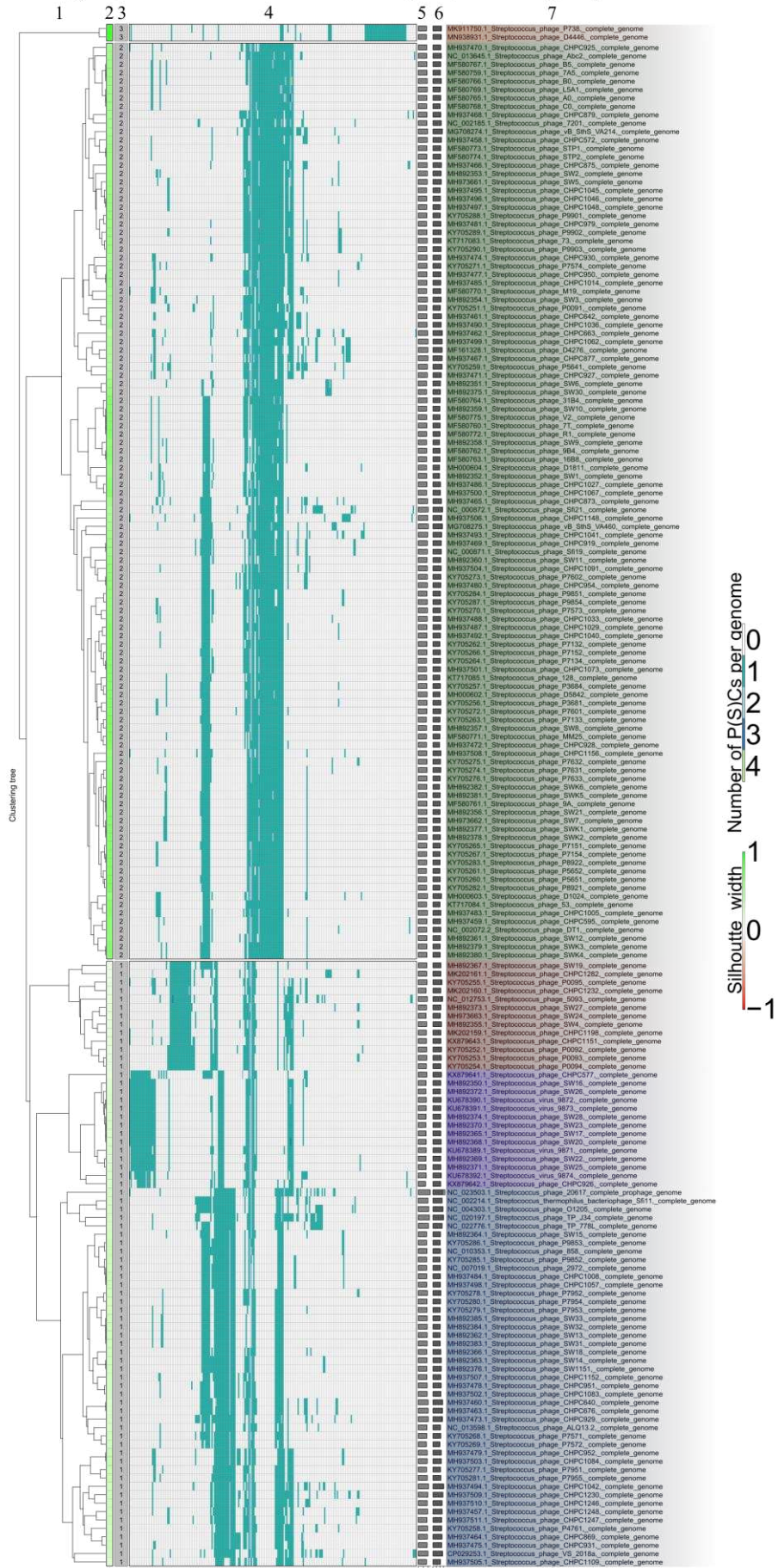


Figure S14 - VirClust hierarchical clustering of 183 phages that infect *Streptococcus thermophilus*, based on intergenomic distances calculated using the protein cluster content. 1) Hierarchical clustering tree. The tree was cut into smaller viral genome clusters (VGCs) using a 0.90 distance threshold. Each VGC was proposed here as a new family. Each VGC is framed in a rectangle in 2 and 3. 2) Silhouette width, measures how related is a virus with other viruses in the same VGCs. Similarity to other VGCs is indicated by values closer to -1 (red). Similarity to viruses in the same VGC is indicated by values closer to 1 (green). 3) VGC ID. 4) Heatmap representation of the PC distribution in the viral genomes. Rows are represented by individual viral genomes. Columns are represented by individual PCs. The ID of each PC can be read at the bottom of the heatmap and in SI file 1 Tables S17-S19. Colors encode the number of each PC per genome, with white signifying the PC absence, and the other colors signifying the various degree of replication (from 1 to 4, see legend). 5) Genome length (bps). 6) Fraction of proteins shared with other phages (dark gray), based on protein assignment to PCs. 7) Phage names. The colored rectangles overlapping the streptococcal phage names correspond to the genus-level grouping proposed here.

● P738 group ● *Moineauvirus* ● 5093 group ● 987 group ● *Brussowvirus*

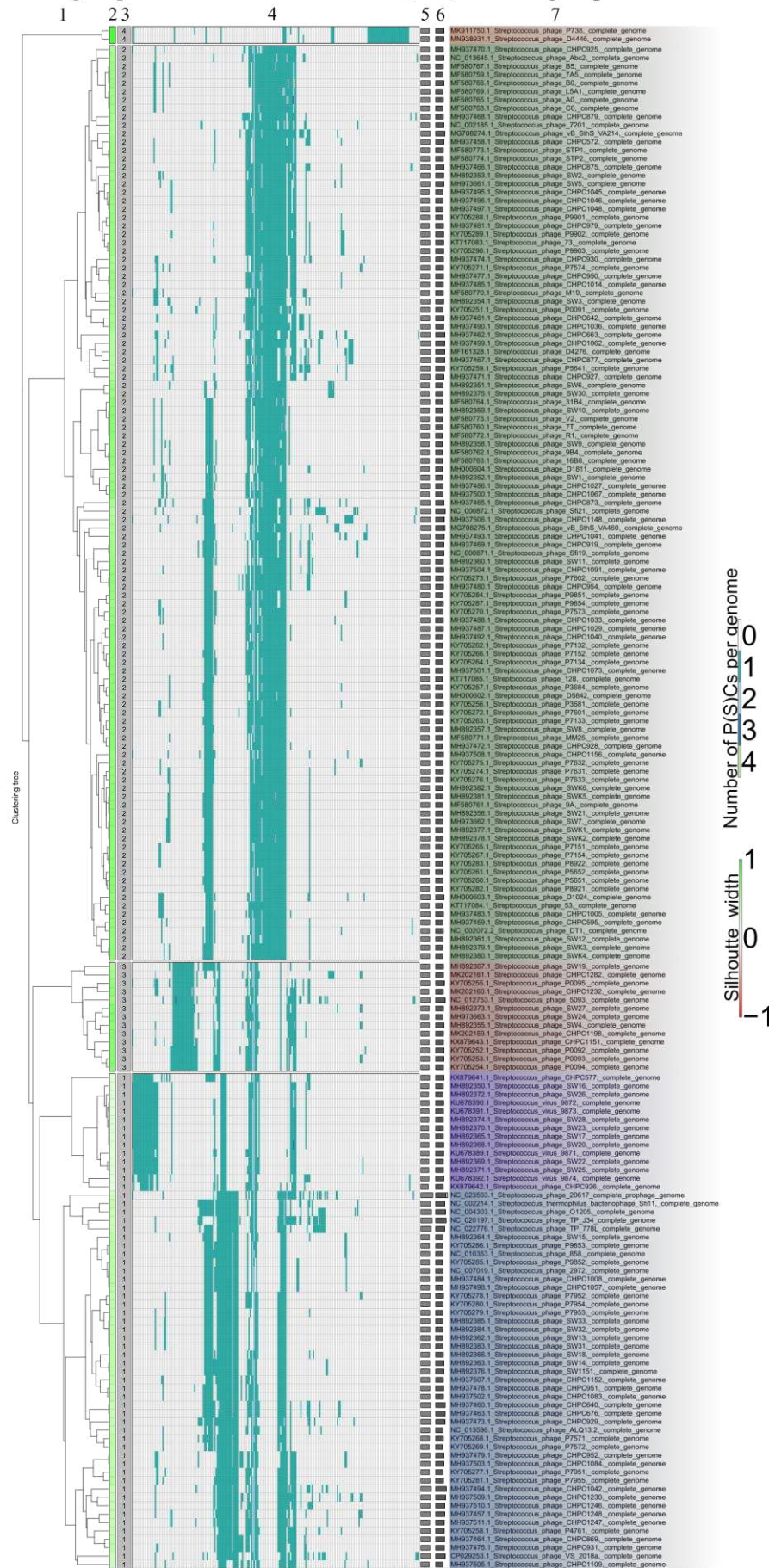


Figure S15 - VirClust hierarchical clustering of 183 phages that infect *Streptococcus thermophilus*, based on intergenomic distances calculated using the protein cluster content. 1) Hierarchical clustering tree. The tree was cut into smaller viral genome clusters (VGCs) using a 0.85 distance threshold. Each VGC was proposed here as a new family. Each VGC is framed in a rectangle in 2 and 3. 2) Silhouette width, measures how related is a virus with other viruses in the same VGCs. Similarity to other VGCs is indicated by values closer to -1 (red). Similarity to viruses in the same VGC is indicated by values closer to 1 (green). 3) VGC ID. 4) Heatmap representation of the PC distribution in the viral genomes. Rows are represented by individual viral genomes. Columns are represented by individual PCs. The ID of each PC can be read at the bottom of the heatmap and in SI file 1 Tables S22-S25. Colors encode the number of each PC per genome, with white signifying the PC absence, and the other colors signifying the various degree of replication (from 1 to 4, see legend). 5) Genome length (bps). 6) Fraction of proteins shared with other phages (dark gray), based on protein assignment to PCs. 7) Phage names. The colored rectangles overlapping the streptococcal phage names correspond to the genus-level grouping proposed here.

SI file 6

<http://tiny.cc/pkjtuz>

SI file 7

<http://tiny.cc/qkjtuz>

SI file 8

<http://tiny.cc/skjtuz>

CAPÍTULO II

Desenvolvimento de PCR multiplex para detecção de fagos que infectam *Streptococcus thermophilus*

RESUMO

As culturas *starter* de *Streptococcus thermophilus* infectadas por fagos são as principais causas de falhas na fermentação durante a produção de queijos e iogurtes nas indústrias de laticínios. A detecção precoce desses agentes representa uma oportunidade para que sejam tomadas medidas de controle antes que ocorram prejuízos à qualidade dos produtos fermentados, perdas econômicas e, eventualmente, à saúde do consumidor. Neste estudo, um método de PCR multiplex foi desenvolvido para detectar, em uma única reação, fagos que infectam *S. thermophilus*. A estratégia para a seleção dos alvos da PCR considerou o alinhamento das sequências genômicas dos genes que codificam proteínas estruturais dos fagos, resultando na proposição de três novos conjuntos de *primers* baseados na proteína do capsídeo. A amplificação de fragmentos inespecíficos na reação de PCR multiplex levou ao uso de apenas dois entre os três pares de *primers* propostos. Mesmo assim, os *primers* foram testados com sucesso em 18 genomas virais, incluindo sete fagos cujas sequências genômicas são desconhecidas. Com base nos resultados da reação de PCR multiplex e na conservação dos genomas, os pares de *primers* desenhados neste estudo foram capazes de detectar 53 (85,48%) dos 62 fagos disponíveis até 2018, destacando-se como uma ferramenta útil para a detecção rápida da maioria dos fagos que infectam *S. thermophilus*.

1. INTRODUÇÃO

Os bacteriófagos (fagos) são um problema persistente nas indústrias de laticínios pois ameaçam o desempenho das culturas *starter* de *Streptococcus thermophilus*, uma bactéria termofílica produtora de ácido láctico (BAL) e fundamental na fabricação de produtos lácteos, como queijos e iogurtes (Calasso et al., 2016; Marco et al., 2017). As consequências da infecção dessa bactéria por fagos geram problemas tecnológicos e de segurança, que podem variar entre fermentação lenta, redução da qualidade do produto final ou falha total do processo (Briggiler Marcó & Mercanti, 2021; Mahony & van Sinderen, 2015). A detecção de fagos que infectam *S. thermophilus* é, portanto, essencial para prevenir infecções durante as fermentações lácteas em indústrias de laticínios.

Sabe-se que a principal e permanente fonte de novos fagos dentro das indústrias de laticínios é por meio de leite cru, que pode conter até 10^4 Unidades Formadoras de Placa (UFP) por mL (Briggiler Marcó & Mercanti, 2021; Pujato et al., 2019), os quais podem permanecer viáveis mesmo após tratamentos térmicos (Atamer et al., 2012, 2013; Capra et al., 2013; Marcó et al., 2019) e se propagar rapidamente durante a fermentação, atingindo concentrações de até 10^9 UFP/mL em soro de leite (Atamer et al., 2011; Lavelle et al., 2018), até 10^7 UFP/g em soro em pó (Wagner et al., 2017) e até 10^5 UFP/g em concentrado proteico de soro (Briggiler Marcó et al., 2022). Respingos de líquidos e deslocamento de ar ao redor de tanques de fabricação e outras superfícies contaminadas geram aerossóis, que são reconhecidos como uma das principais vias de dispersão de fagos em laticínios (Briggiler Marcó et al., 2018; Verreault et al., 2008, 2011), atingindo concentrações de até 10^8 UFP/m³ no ar (Atamer et al., 2013; Verreault et al., 2011). Uma vez que existe uma diversidade de fagos naturalmente presentes no leite cru, capazes de permanecer viáveis após os tratamentos térmicos mais comuns e se disseminarem para diversos ambientes via aerossol, a ausência de fagos nas indústrias de laticínios é inatingível (Marcó et al., 2012; Pujato et al., 2019).

Devido às dificuldades encontradas no controle de fagos, a indústria de laticínios adotou diferentes estratégias visando reduzir os níveis desses agentes e interromper sua disseminação, incluindo tratamentos físicos (aquecimento, filtração, alta pressão, radiação ultravioleta - UV), químicos (sanitizantes), biológicos (rotação de culturas, uso de cepas com resistência a fagos - BIMs) (Pujato et al., 2019), além de controle do fluxo de ar na fábrica, programas de higienização e ausência de cepas lisogênicas (Briggiler Marcó & Mercanti, 2021). No entanto, apesar das diferentes abordagens práticas para controle, a infecção das culturas *starter* por fagos continua sendo a causa mais comum de fermentação lenta ou incompleta e produtos de baixa qualidade na indústria de laticínios (de Melo et al., 2018). A eficácia dessas estratégias é limitada principalmente devido à coevolução dinâmica entre fagos e BAL, aos curtos períodos de latência (20 a 40 min), ao *burst size* grande (até 600 UFP por célula infectada) e à termoestabilidade desses agentes aos tratamentos de pasteurização comumente aplicados na indústria de laticínios (63 °C por 30 min e 72 °C por 15 s) (Atamer et al., 2011; Brüßow et al., 1998; Daly et al., 1996; Marcó et al., 2019).

Os métodos atualmente disponíveis para detecção de fagos podem ser categorizados com base no elemento alvo, podendo ser: o fago infeccioso (contagem de placas, teste de atividade, teste de turbidez, mudanças de impedância elétrica ou de condutividade) (Adams, 1959; Anderson et al., 2011; Clokie et al., 2018; García-Aljaro et al., 2009; Gratia, 1936; Houlberg et al., 2005; Hyman & Abedon, 2009; Svensson & Christiansson, 1991), a partícula

inteira do fago (microscopia eletrônica de transmissão - MET, citometria de fluxo, microscopia óptica iluminada a laser - NanoSight, microscopia de epifluorescência e força atômica, espectroscopia Raman (Acar-Soykut et al., 2018; Ackermann, 2012; Anderson et al., 2011; Brussaard et al., 2000; Michelsen et al., 2007; Tayyarcıan et al., 2018; Zago et al., 2012), os ácidos nucleicos (reação em cadeia da polimerase - PCR simplex/multiplex, PCR quantitativa em tempo real - qPCR, transcrição reversa seguida de PCR - RT-qPCR, droplet digital PCR, sequenciamento de próxima geração - Illumina, PacBio e Nanopore) (Anderson et al., 2011; Binetti et al., 2008; Ji et al., 2020; Klumpp et al., 2012; Morella et al., 2018; Muhammed et al., 2017) e as proteínas (espectrometria de massa, imunoenaios) (Khan et al., 2015; Larsson et al., 2013; Serafim et al., 2017; Štveráková et al., 2018; Wang et al., 2019). A detecção precoce de fagos no leite, ingredientes e/ou nos coprodutos durante qualquer etapa do processo produtivo se torna extremamente útil, pois representa uma oportunidade para que sejam tomadas medidas de controle antes que ocorram prejuízos à qualidade dos produtos fermentados, perdas econômicas e, eventualmente, à saúde do consumidor. No entanto, a dinâmica das indústrias de laticínios requer ferramentas rápidas para detectar com precisão os fagos nos diferentes estágios da fabricação de produtos lácteos, por isso a amplificação de uma sequência genômica direcionada por PCR consiste em uma abordagem eficiente, já que os métodos de detecção molecular vêm se tornando cada vez mais acessíveis e econômicos (Binetti et al., 2005; Szymczak, Vogensen, et al., 2019).

A expansão contínua dos bancos de dados genômicos tem contribuído para a compreensão da biodiversidade dos fagos lácteos e fornecido novas informações para a caracterização e agrupamento desses agentes, essenciais para o planejamento de novas estratégias antifágicas (Mahony et al., 2014; Mahony & van Sinderen, 2015). Várias abordagens já foram utilizadas na tentativa de detectar e classificar os fagos que infectam *S. thermophilus* por métodos moleculares (Binetti et al., 2005; Brussow et al., 1994; McDonnell et al., 2017; Quiberoni et al., 2006; Szymczak, Vogensen, et al., 2019). No entanto, devido à alta diversidade genética entre os fagos e atualização constante do banco de dados (Lavelle et al., 2018; McDonnell et al., 2017; Szymczak, Rau, et al., 2019) se torna necessária a aplicação de estratégias de detecção que utilizem de estudos de conservação das proteínas virais para o maior alcance dessas metodologias. Neste estudo, um método de PCR multiplex foi desenvolvido para detectar, em uma única reação, fagos que infectam *S. thermophilus*.

2. MATERIAL E MÉTODOS

2.1. Agrupamento das sequências das proteínas

Em fevereiro de 2018, 62 genomas completos de fagos virulentos e temperados que infectam *S. thermophilus* estavam disponíveis no banco de dados GenBank/NCBI (<https://www.ncbi.nlm.nih.gov/genbank>) (Benson et al., 2012). As sequências das proteínas anotadas nos genomas dos fagos foram recuperadas e agrupadas usando o CD-HIT Suite (<http://weizhong-lab.ucsd.edu/cd-hit/>) (Huang et al., 2010). Os *clusters* foram determinados para reunir as sequências das proteínas estruturais que compartilhavam pelo menos 90% de identidade, sendo selecionadas a partir deles as sequências da *capsid protein* (proteína do capsídeo) por apresentarem o maior grau de conservação (menor número de *clusters* formados e maior quantidade de sequências agrupadas). Três *clusters* reuniram as sequências da *capsid protein* de 60 genomas (96,77% das sequências disponíveis): *Cluster_186* (contendo 6 sequências), *Cluster_187* (contendo 37 sequências) e *Cluster_198* (contendo 17 sequências). As sequências proteicas de referência para cada *cluster* foram usadas para o desenho dos *primers*.

2.2. Desenho dos *primers*

Os conjuntos de *primers* foram desenhados usando o programa Primer3 (<https://primer3.org/>) (Untergasser et al., 2012) considerando a amplificação de fragmentos de tamanhos distintos para a identificação dos fagos nos diferentes *clusters*. Para isso, as sequências de nucleotídeos foram obtidas a partir da tradução reversa das respectivas sequências das proteínas representativas dos três *clusters* usando a ferramenta Backtranseq do EMBOSS (McWilliam et al., 2013). A especificidade dos *primers* foi avaliada com a ferramenta Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) (Ye et al., 2012), incluindo sequências de fagos e estirpes de *Lactococcus lactis* como *off-targets*. A formação de produtos secundários (*hairpin*, *self-dimer* e *hetero-dimer*) foi verificada utilizando a ferramenta OligoAnalyzer (<https://www.idtdna.com/pages/tools/oligoanalyzer>) (Owczarzy et al., 2008). Os *primers* foram sintetizados pela empresa Invitrogen (Thermo Fisher Scientific, Itália).

2.3. Padronização das condições da PCR multiplex

Bactérias, fagos e condições de cultivo

Dez fagos e seus respectivos hospedeiros da coleção do Istituto per la Qualità e le Tecnologie Agroalimentari (Thiene, Itália) e outros 12 fagos gentilmente cedidos pela University College Cork (Cork, Irlanda) (Tabela 1) foram utilizados para a validação da PCR multiplex proposta. Dentre os 22 fagos, sete deles com sequência genômica desconhecida (162,

163, 353, 453, 459, 720 e 729) foram testados na reação para sua identificação. O DNA das bactérias hospedeiras foi utilizado como controle negativo nas análises.

Tabela 1 - Lista de fagos e estirpes de *Streptococcus thermophilus* utilizados para a validação da PCR multiplex.

Fago	Nº de acesso no GenBank	Hospedeiro	Referência
162	-	162	
163	-	163	
vB_SthS_VA214	MG708274	214	
353	-	353	
453	-	453	
459	-	459	*
vB_SthS_VA460	MG708275	460	
vB_SthS_VA698	MG708273	698	
720	-	720	
729	-	729	
SW1151	MH892376	-	
SW13	MH892362	-	
SW14	MH892363	-	
SW15	MH892364	-	
SW18	MH892366	-	
SW19	MH892367	-	
SW24	MH973663	-	(Lavelle et al., 2018)
SW27	MH892373	-	
SW31	MH892383	-	
SW32	MH892384	-	
SW33	MH892385	-	
SW4	MH892355	-	

* Micro-organismos da coleção do Istituto per la Qualità e le Tecnologie Agroalimentari (Thiene, Itália).

Fonte: Autora.

A estirpe bacteriana foi cultivada *overnight* a 44 °C em caldo M-17 (Sigma-Aldrich, EUA) a partir de colônias únicas ou do estoque e os fagos foram propagados no hospedeiro correspondente. Para isso, 10 mL de caldo M-17 contendo 1% (m/v) de CaCl₂ 1 M foi adicionado de 0,1 mL da suspensão bacteriana cultivada *overnight* a 44 °C, seguido por

incubação da mistura a 44 °C por 1 h. Após a incubação, foi adicionado 10 µL do estoque de fago e a mistura foi novamente incubada *overnight* a 44 °C. A suspensão contendo os fagos foi filtrada usando filtro de 0,20 µm (Corning, Alemanha) e armazenada a 4 °C. O título viral foi determinado pelo método de plaqueamento em dupla camada de ágar (Adams, 1959; Gratia, 1936) e expresso em termos de unidades formadoras de placas por mililitro (UFP·mL⁻¹).

Extração de DNA viral e bacteriano

A extração do DNA viral foi realizada seguindo o protocolo descrito por Binetti et al., (2005) com algumas modificações. A suspensão contendo os fagos (10⁶-10⁷ UFP·mL⁻¹) (30 mL) foi acrescentada de NaCl 1 M por 1 h a 4 °C e centrifugada a 5.400 ×g por 20 min a 10 °C. Em seguida, 10% (m/v) de polietilenoglicol 8.000 g/mol foi adicionado ao sobrenadante, seguido de incubação *overnight* a 4 °C e centrifugação a 16.000 rpm (Centrikon T-124, Kontron, Italy) por 2 h a 4 °C. O sedimento foi ressuscitado em 400 µL de tampão SM (gelatina 0,01% [m/v], Tris-HCl 50 mM [pH 7,5], NaCl 100 mM, MgSO₄·7H₂O 8 mM) e os fagos foram lisados com 100 µL de solução de lise (EDTA 0,25 M [pH 8], Tris-HCl 0,5 M [pH 7,5], SDS 2,5% [m/v]) a 65 °C por 30 min. Acetato de potássio 8 M (100 µL) foi adicionado e a mistura foi incubada em gelo por 15 min antes da centrifugação a 14.000 rpm (Centrikon T-124, Kontron, Itália) por 10 min a 4 °C. O DNA do fago foi recuperado a partir do sobrenadante adicionado de igual volume de isopropanol, mantido à temperatura ambiente por 5 min e novamente centrifugado a 14.000 rpm (Centrikon T-124, Kontron, Itália) por 10 min a 4 °C. O DNA foi ressuscitado em 630 µL de tampão TE (Tris-HCl 10 mM [pH 7,5], EDTA 1 mM [pH 8]) na presença de acetato de sódio 0,3 M, precipitado novamente com igual volume de isopropanol durante 5 min e centrifugado a 14.000 rpm (Centrikon T-124, Kontron, Itália) por 10 min a 4 °C. O DNA precipitado foi lavado duas vezes com etanol a 70% (v/v), seco e ressuscitado em tampão TE. Para a extração do DNA bacteriano, *S. thermophilus* foi cultivado *overnight* a 44 °C em caldo M-17. As células coletadas por centrifugação foram submetidas à extração de DNA total usando o protocolo de lise celular alcalino (Dellaglio et al., 1973). O rendimento e a pureza do DNA (bacteriano e viral) foram avaliados por espectrofotometria (NanoDrop 2000c, Thermo Scientific, EUA).

Reações de PCR

Para detectar os fagos que infectam *S. thermophilus* dentro de um dos três *clusters* (*Cluster_186*, *Cluster_187* e *Cluster_198*) foi desenvolvida uma PCR multiplex usando três pares de *primers* (um por *cluster*), desenhados a partir das regiões conservadas dos genes que

codificam as principais proteínas do capsídeo dos fagos, para os quais as sequências dos genomas estavam disponíveis. As condições padronizadas para as ampliações por PCR foram: 94 °C × 3 min (desnaturação inicial), seguido de 35 ciclos de 94 °C × 45 s (desnaturação), 47 °C × 1 min (anelamento), 72 °C × 1 min (extensão), com uma extensão final de 72 °C × 5 min. As reações de PCR foram realizadas em termociclador (Bio-Rad, Hercules, USA), utilizando GoTaq G2 DNA Polymerase (Promega, USA) adicionado de 10 ng de DNA (bacteriano ou viral) como molde a ser amplificado. Os produtos da PCR multiplex foram visualizados em gel de agarose 2% (m/v) em tampão TAE (Tris-Acetato-EDTA). Os fragmentos de 205, 366 e 640 pb discriminam os fagos do *Cluster_198*, *Cluster_186* e *Cluster_187*, respectivamente.

2.4. Avaliação da capacidade de detecção dos *primers*

A capacidade dos *primers* de amplificarem as sequências de DNA alvo foi avaliada *in silico* usando a ferramenta *Finding Binding Sites and Create Fragments* do programa CLC Genomics Workbench 8.5.4 (Invitrogen, EUA). Nessa análise, os genomas do banco de dados foram utilizados para a simulação de ampliações usando os três conjuntos de *primers* propostos, além de outros quatro conjuntos de *primers* desenvolvidos em estudos anteriores (Tabela 2) (Quiberoni et al., 2006; Szymczak, Vogensen, et al., 2019). As ampliações foram simuladas usando as configurações padrão para as condições de amplificação da DNA polimerase e permitindo ou não incompatibilidades (*mismatches*) no anelamento entre as sequências dos *primers* e as regiões alvo.

Tabela 2 - Lista de *primers* descritos na literatura para detecção dos fagos que infectam *Streptococcus thermophilus*.

Primer	Sequência, 5' - 3'	Amplicon (pb)	Referência
<i>cos</i>	FOR GGTTCACGTGTTTATGAAAAATGG	170	(Quiberoni et al., 2006)
	REV AGCAGAATCAGCAAGCAAGCTGTT		
<i>pac</i>	FOR GAAGCTATGCGTATGCAAGT	427	(Quiberoni et al., 2006)
	REV TTAGGGATAAGAGTCAAGTG		
987	FOR TCTGTACCTGTCTTGTTG	641	(Szymczak, Vogensen, et al., 2019)
	REV GTTTGGGGAGATGGTATT		
5093	FOR CTGATGCTGGACAAAAC	321	(Szymczak, Vogensen, et al., 2019)
	REV CACCTCCGAATTTAACAAC		

Fonte: Autora.

3. RESULTADOS

3.1. Detecção de fagos que infectam *S. thermophilus* por PCR multiplex

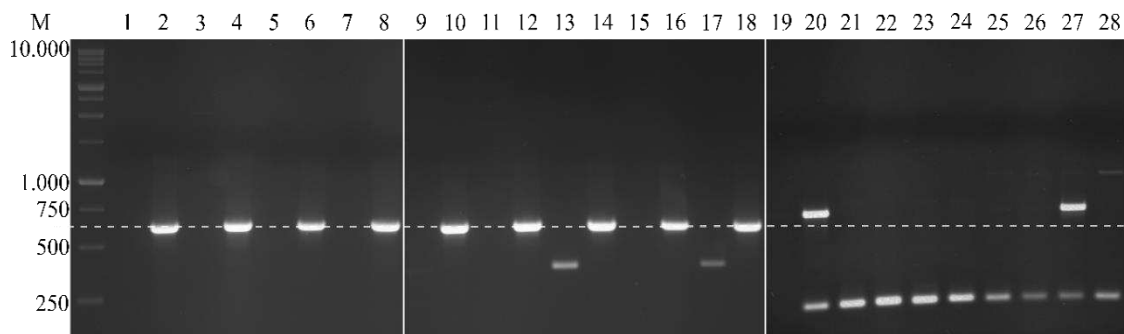
A reação de PCR multiplex contendo os três pares de *primers* (Tabela 3) gerou fragmentos inespecíficos de aproximadamente 840, 730 e 440 pb. Verificou-se que os fragmentos eram gerados a partir da combinação do conjunto de *primers Cluster_186* na presença dos *primers Cluster_187*. Devido a isto, a reação de PCR multiplex foi realizada com apenas os conjuntos de *primers Cluster_187* e *Cluster_198* (Figura 1).

Tabela 3 - Lista dos *primers* sintetizados para a diferenciação dos fagos que infectam *Streptococcus thermophilus*.

Primer	Sequência, 5' - 3'	Amplicon (pb)
<i>Cluster_198</i>	FOR CGTGATCGTGTTTCTGCTGA	205
	VER TACGCATAGCCTCAAGACGA	
<i>Cluster_186</i>	FOR TGAGGAGCAGCTTCGTTATG	497
	VER GCTGAATAGGAGCGCACATC	
<i>Cluster_187</i>	FOR TGATCTTTGGGTTGCTCAGG	640
	VER AGCAATCCAACCAGAAAGCC	

Fonte: Autora.

Figura 1 - Gel de agarose com os fragmentos amplificados a partir de DNA viral e bacteriano pelo método de PCR multiplex com os conjuntos de *primers Cluster_187* e *Cluster_198*. Linhas (1) bactéria 459; (2) fago 459; (3) bactéria 460; (4) fago 460; (5) bactéria 698; (6) fago 698; (7) bactéria 720; (8) fago 720; (9) bactéria 162; (10) fago 162; (11) bactéria 214; (12) fago 214; (13) bactéria 353; (14) fago 353; (15) bactéria 453; (16) fago 453; (17) bactéria 163; (18) fago 163; (19) bactéria 729; (20) fago 729; (21) fago SW1151; (22) fago SW18; (23) fago SW14; (24) fago SW15; (25) fago SW31; (26) fago SW13; (27) fago SW32; (28) fago SW33. Linhas 2, 4, 6, 8, 10, 12, 14, 16 e 18, amplicon de 640 pb. Linhas 20, 21, 22, 23, 24, 25, 26, 27 e 28, amplicon de 205 pb. M, marcador molecular SHARPMASS 1 kb DNA Ladder (EuroClone, Itália).



Fonte: Autora.

Os dois pares de *primers* detectaram com sucesso 18 fagos que infectam *S. thermophilus* disponíveis neste estudo. Os outros quatro fagos disponíveis (SW4, SW19, SW24 e SW27) pertenciam ao *Cluster_186* e, portanto, não puderam ser detectados por esta reação de PCR. A amplificação de fragmentos com tamanhos distintos, 640 pb (linhas 2, 4, 6, 8, 10, 12, 14, 16 e 18) e 205 pb (linhas 20, 21, 22, 23, 24, 25, 26, 27 e 28), permitiu a classificação dos fagos de acordo com o *cluster* ao qual pertenciam: *Cluster_187* (fagos 459, 460, 698, 720, 162, 214, 353, 453 e 163) e *Cluster_198* (fagos 729, SW1151, SW18, SW14, SW15, SW31, SW13, SW32 e SW33), respectivamente. A reação de PCR multiplex desenvolvida neste estudo também detectou sete fagos desconhecidos. Seis deles (459, 720, 162, 353, 453, 163) foram atribuídos ao *Cluster_187* e um fago (729) ao *Cluster_198*.

Entretanto, a reação gerou também quatro fragmentos inespecíficos: dois deles foram observados nas canaletas relativas aos genomas amplificados das bactérias 353 e 163 (394 e 400 pb, respectivamente) e dois nos genomas dos fagos 729 e SW32 (675 e 673 pb, respectivamente). Apesar disso, os fragmentos amplificados não afetaram a detecção dos fagos, uma vez que possuíam tamanho distinto dos fragmentos esperados para a amplificação específica. Com base nos resultados da reação de PCR multiplex e na conservação dos genomas reunidos nos *clusters*, os pares de *primers* desenhados neste estudo foram capazes de detectar

53 (85,48%) dos 62 fagos disponíveis até 2018, destacando-se como uma ferramenta útil para a detecção rápida da maioria dos fagos que infectam *S. thermophilus*.

3.2. Avaliação da capacidade de detecção dos *primers*

A avaliação da especificidade dos *primers* e a conservação das regiões alvo a serem amplificadas foram realizadas por PCRs simuladas computacionalmente. Na amplificação *in silico*, o anelamento dos conjuntos de *primers* nas sequências dos genomas evidenciou a presença de polimorfismos entre as sequências dos *primers* e as dos DNA alvos dos fagos de cada *cluster*. Tais polimorfismos reduzem a especificidade dos *primers* e, portanto, dificultaram o anelamento durante a PCR multiplex. A presença dos polimorfismos decorre do fato de o código genético ser degenerado e de os *primers* originais terem sido gerados a partir da sequência consenso proteica traduzida. A ferramenta utilizada para a tradução reversa utilizou uma tabela de códons que não correspondia aos códons presentes no genoma dos fagos. Logo, para a correção dos *primers*, os polimorfismos nas sequências foram identificados e corrigidos dando origem aos novos *primers* *Cluster_198.2*, *Cluster_186.2* e *Cluster_187.2* (Tabela 4). Para isso, as sequências de nucleotídeos foram obtidas a partir do alinhamento das sequências proteicas de referência contra o banco de dados de nucleotídeos traduzidos usando a ferramenta tBLASTn (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1990).

Tabela 4 - Lista dos *primers* otimizados para a diferenciação dos fagos que infectam *Streptococcus thermophilus*.

Primer	Sequência, 5' – 3'	Amplicon (pb)
<i>Cluster_198.2</i>	FOR CGTGACCGTGTTAGTGCTGA	205
	VER TACGCATAGCTTCAAGACGA	
<i>Cluster_186.2</i>	FOR GAGTGAGAGAACGTGGGACA	366
	VER TAGCCACAGTCTCCTCAACG	
<i>Cluster_187.2</i>	FOR TGACCTTTGGGTTGCTCAAG	640
	VER AGCAATCCAACCAGACAACC	

Fonte: Autora.

Os três pares de *primers* otimizados desenhados neste estudo se ligaram especificamente à região alvo nos genomas de 34 fagos (54,83%) com 100% de identidade de nucleotídeos, enquanto os quatro conjuntos de *primers* de referência (Tabela 2) se ligaram à região alvo de apenas 26 fagos (41,93%). A possibilidade de incompatibilidades (*mismatches*) entre as sequências dos *primers* e as regiões alvo na PCR computacional permitiu a amplificação de 59

fragmentos específicos em 59 genomas (95,16%) dos 62 testados, considerando o uso dos três pares de *primers* desenvolvidos neste estudo. Enquanto isso, o anelamento com *mismatches* utilizando os quatro conjuntos de *primers* de referência amplificou 513 fragmentos nos 62 genomas, evidenciando a baixa especificidade daquele grupo de *primers*.

Analisando os três genomas que não foram amplificados na PCR computacional quando se permitiu incompatibilidades entre as sequências dos *primers* e as regiões alvo, verificou-se que dois fagos pertenciam ao grupo 987 (*Streptococcus* phage CHPC926 - KX879642 e *Streptococcus* phage CHPC577 - KX879641) e um pertencia ao grupo *pac* (*Streptococcus* phage CHPC1109 - MH937505). Comparando o agrupamento em *clusters* pelo CD-HIT Suite e os grupos de fagos estreptocócicos descritos até 2018 (McDonnell et al., 2016, 2017; Szymczak et al., 2017) constatou-se que os fagos foram agrupados da seguinte forma: os fagos do grupo 5093 no *Cluster_186*, os fagos do grupo *cos*-containing no *Cluster_187* e os fagos do grupo *pac*-containing no *Cluster_198*. Já os fagos do grupo 987 não foram agrupados em nenhum dos *clusters*, pois no início da organização do banco de dados em 2018 existiam apenas dois genomas sequenciados de fagos desse grupo e, como seus genomas não se alinhavam com os demais, estes genomas não foram incluídos no agrupamento e posterior desenho dos *primers*. Assim, nenhum dos três pares de *primers* desenvolvido neste estudo é capaz de detectar fagos do grupo 987. Já o fago CHPC1109, possui regiões em seu genoma ainda mais distintas dos demais do grupo *pac*, inclusive nas regiões alvo utilizadas para o desenho dos oligonucleotídeos, e não seria detectado por qualquer conjunto de *primers* (deste estudo e de referência) já desenvolvido para esse fim.

As regiões alvo usadas neste estudo (genes que codificam a *capsid protein*) para desenhos dos *primers* são altamente conservadas e contêm apenas algumas mutações pontuais, em contraste com as sequências que codificam a *major capsid protein* (flanqueada pelos conjuntos de *primers cos, pac* e 5093) e a *receptor binding protein* (flanqueada pelo conjunto de *primers 987*), cuja proposição de *primers* levou à amplificação de regiões inespecíficas. A análise *in silico* mostrou que a região utilizada neste estudo possui maior grau de conservação que as utilizadas para a classificação desses agentes até então (Quiberoni et al., 2006; Szymczak, Vogensen, et al., 2019), sendo uma melhor alternativa para proposição do desenvolvimento de moléculas detectoras, como os *primers* aqui descritos.

4. DISCUSSÃO

Neste estudo, desenvolvemos um método molecular para detecção de fagos que infectam *S. thermophilus*, com *primers* específicos para as regiões alvo que apresentam maior

conservação genômica entre os membros de cada grupo e menor probabilidade de amplificação de fragmentos inespecíficos durante a reação de PCR multiplex.

Várias estratégias já foram utilizadas na tentativa de detectar e classificar os fagos que infectam *S. thermophilus* por métodos moleculares. O primeiro ensaio de PCR relatado para detecção desses fagos em amostras de laticínios foi baseado na amplificação de um fragmento de DNA conservado no genoma da maioria dos fagos da coleção Nestlé. Este fragmento cobria parte das ORFs 271 e 504 do fago Sfi11 (Brussow et al., 1994). Mais tarde, Binetti et al. (2005) desenvolveram um método para amplificar a sequência VR2, região variável do gene que codifica a proteína do antirreceptor, permitindo a rápida detecção e classificação de alguns fagos estreptocócicos dos grupos *cos*-containing e *pac*-containing. Um método de PCR multiplex foi desenvolvido para classificar os fagos que infectam *S. thermophilus* dentro dos grupos *cos*-containing ou *pac*-containing, com base na amplificação de regiões conservadas nos genes que codificam a *major capsid protein* nos genomas de quatro fagos do grupo *cos*-containing e três fagos do grupo *pac*-containing (Quiberoni et al., 2006). Fagos que infectam *S. thermophilus* foram os primeiros a serem detectados, quantificados e classificados em *cos*-containing ou *pac*-containing por PCR em tempo real (qPCR) entre as BAL em amostras de leite (del Rio et al., 2008).

A descoberta de fagos com novas propriedades, 5093 (Mills et al., 2011) e 987 (McDonnell et al., 2016), juntamente com a maior disponibilidade de sequências genômicas, resultou na necessidade de atualizar o sistema de detecção e classificação por PCR multiplex. Em 2017, uma melhoria no ensaio de PCR multiplex (Quiberoni et al., 2006) foi proposta incluindo dois conjuntos de *primers* adicionais para a detecção de fagos dos grupos 5093 e 987 (McDonnell et al., 2017). No entanto, o método poderia fornecer um resultado falso-positivo caso fagos que infectam outras espécies lácteas estivessem presentes, levando à classificação de fagos em grupos incorretos. A adaptação mais recente do método de PCR multiplex para detecção e classificação de fagos que infectam *S. thermophilus* foi estabelecida por Szymczak, Vogensen, et al. (2019), os quais foram usados como padrão de comparação em nosso estudo. Com base nas análises *in silico*, demonstramos que os conjuntos de *primers* desenvolvidos neste estudo são mais específicos, mesmo considerando a ocorrência de *mismatches*, apresentando menor probabilidade de amplificação de fragmentos inespecíficos durante a PCR multiplex, já que foram desenvolvidos a partir de estudos de conservação genômica.

Os conjuntos de *primers* propostos neste trabalho consistem em uma melhoria no processo de identificação de fagos que infectam *S. thermophilus* por PCR multiplex. Portanto, para a detecção de todos os fagos conhecidos independentemente do grupo ao qual pertence,

sugerimos a utilização dos conjuntos de *primers* desenvolvidos neste estudo adicionados do par de *primers* 987 (Szymczak, Vogensen, et al., 2019).

Como discutido no Capítulo I, atualmente, os fagos que infectam *S. thermophilus* estão organizados em 142 espécies e cinco gêneros (três deles novos) e que devido à sua diversidade genética, a classificação em nível de família varia de acordo com os critérios de classificação utilizados. Como perspectivas futuras, novos *primers* podem ser desenhados para detecção e classificação taxonômica desses fagos estreptocócicos, provenientes de diferentes amostras e ambientes lácteos.

5. CONCLUSÃO

Neste estudo, desenvolvemos uma PCR multiplex para a detecção rápida de fagos que infectam *S. thermophilus*, com *primers* apresentando menor probabilidade de amplificação de fragmentos inespecíficos durante a reação, por terem sido desenhados a partir de estudos de conservação genômica. Nesse sentido, o método de PCR multiplex fornece uma ferramenta adequada para detectar fagos estreptocócicos, com potencial para a identificação de fagos ainda não descritos, uma vez que se baseia em estudos de conservação genômica.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Acar-Soykut, E., Tayyarcı, E. K., & Boyacı, I. H. (2018). A simple and fast method for discrimination of phage and antibiotic contaminants in raw milk by using Raman spectroscopy. *Journal of Food Science and Technology*, 55(1), 82–89. <https://doi.org/10.1007/s13197-017-2798-y>
- Ackermann, H.-W. (2012). Bacteriophage electron microscopy. *Advances in Virus Research*, 82, 1–32. <https://doi.org/10.1016/B978-0-12-394621-8.00017-0>
- Adams, M. H. (1959). *Bacteriophages*. New York, Interscience Publishers. <http://archive.org/details/bacteriophages00adam>
- Anderson, B., Rashid, M. H., Carter, C., Pasternack, G., Rajanna, C., Revazishvili, T., Dean, T., Senecal, A., & Sulakvelidze, A. (2011). Enumeration of bacteriophage particles: Comparative analysis of the traditional plaque assay and real-time QPCR- and nanosight-based assays. *Bacteriophage*, 1(2), 86–93. <https://doi.org/10.4161/bact.1.2.15456>
- Atamer, Z., Ali, Y., Neve, H., Heller, K. J., & Hinrichs, J. (2011). Thermal resistance of bacteriophages attacking flavour-producing dairy *Leuconostoc* starter cultures. *International Dairy Journal*, 21(5), 327–334. <https://doi.org/10.1016/j.idairyj.2010.11.005>
- Atamer, Z., Neve, H., Heller, K., & Hinrichs, J. (2012). Thermal resistance of bacteriophages in the dairy industry. Em *Bacteriophages in Dairy Processing* (p. 195–214).

- Atamer, Z., Samtlebe, M., Neve, H., J. Heller, K., & Hinrichs, J. (2013). Review: Elimination of bacteriophages in whey and whey products. *Frontiers in Microbiology*, *4*, 191. <https://doi.org/10.3389/fmicb.2013.00191>
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. <https://doi.org/10.1093/nar/gks1195>
- Binetti, A. G., Capra, M. L., Álvarez, M. A., & Reinheimer, J. A. (2008). PCR method for detection and identification of *Lactobacillus casei/paracasei* bacteriophages in dairy products. *International Journal of Food Microbiology*, *124*(2), 147–153. <https://doi.org/10.1016/j.ijfoodmicro.2008.03.006>
- Binetti, A. G., Del Río, B., Martín, M. C., & Álvarez, M. A. (2005). Detection and Characterization of *Streptococcus thermophilus* Bacteriophages by Use of the Antireceptor Gene Sequence. *Applied and Environmental Microbiology*, *71*(10), 6096–6103. <https://doi.org/10.1128/AEM.71.10.6096-6103.2005>
- Briggiler Marcó, M., Machado, N., Quiberoni, A., & Suárez, V. (2022). *Streptococcus thermophilus* Phages in Whey Derivatives: From Problem to Application in the Dairy Industry. *Viruses*, *14*(4), 810. <https://doi.org/10.3390/v14040810>
- Briggiler Marcó, M., & Mercanti, D. J. (2021). Bacteriophages in dairy plants. In *Advances in Food and Nutrition Research* (Vol. 97, p. 1–54). Elsevier. <https://doi.org/10.1016/bs.afnr.2021.02.015>
- Briggiler Marcó, M., Negro, A. C., Alfano, O. M., & Quiberoni, A. D. L. (2018). New semi-pilot-scale reactor to study the photocatalytic inactivation of phages contained in aerosol. *Environmental Science and Pollution Research International*, *25*(22), 21385–21392. <https://doi.org/10.1007/s11356-017-8994-5>
- Brussaard, C. P., Marie, D., & Bratbak, G. (2000). Flow cytometric detection of viruses. *Journal of Virological Methods*, *85*(1–2), 175–182. [https://doi.org/10.1016/s0166-0934\(99\)00167-6](https://doi.org/10.1016/s0166-0934(99)00167-6)
- Brüssow, H., Bruttin, A., Desiere, F., Lucchini, S., & Foley, S. (1998). Molecular Ecology and Evolution of *Streptococcus thermophilus* Bacteriophages—A Review. *Virus Genes*, *16*(1), 95–109. <https://doi.org/10.1023/A:1007957911848>
- Brussow, H., Fremont, M., Bruttin, A., Sidoti, J., Constable, A., & Fryder, V. (1994). Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. *Applied and Environmental Microbiology*, *60*(12), 4537–4543. <https://doi.org/10.1128/AEM.60.12.4537-4543.1994>
- Calasso, M., Ercolini, D., Mancini, L., Stellato, G., Minervini, F., Di Cagno, R., De Angelis, M., & Gobbetti, M. (2016). Relationships among house, rind and core microbiotas during manufacture of traditional Italian cheeses at the same dairy plant. *Food Microbiology*, *54*, 115–126. <https://doi.org/10.1016/j.fm.2015.10.008>
- Capra, M. L., Neve, H., Sorati, P. C., Atamer, Z., Hinrichs, J., Heller, K. J., & Quiberoni, A. (2013). Extreme thermal resistance of phages isolated from dairy samples: Updating

- traditional phage detection methodologies. *International Dairy Journal*, 30(2), 59–63. <https://doi.org/10.1016/j.idairyj.2012.11.009>
- Clokier, M. R. J., Kropinski, A. M., & Lavigne, R. (Eds.). (2018). *Bacteriophages: Methods and Protocols, Volume 3* (Vol. 1681). Springer New York. <https://doi.org/10.1007/978-1-4939-7343-9>
- Daly, C., Fitzgerald, G. F., & Davis, R. (1996). Biotechnology of lactic acid bacteria with special reference to bacteriophage resistance. *Antonie Van Leeuwenhoek*, 70(2–4), 99–110. <https://doi.org/10.1007/BF00395928>
- de Melo, A. G., Levesque, S., & Moineau, S. (2018). Phages as friends and enemies in food processing. *Current Opinion in Biotechnology*, 49, 185–190. <https://doi.org/10.1016/j.copbio.2017.09.004>
- del Rio, B., Martín, M. C., Martínez, N., Magadán, A. H., & Alvarez, M. A. (2008). Multiplex Fast Real-Time PCR for Quantitative Detection and Identification of cos- and pac-Type *Streptococcus thermophilus* Bacteriophages. *Applied and Environmental Microbiology*, 74(15), 4779–4781. <https://doi.org/10.1128/AEM.00295-08>
- Dellaglio, F., Bottazzi, V., & Trovatielli, L. D. (1973). Deoxyribonucleic Acid Homology and Base Composition in Some Thermophilic Lactobacilli. *Journal of General Microbiology*, 74(2), 289–297. <https://doi.org/10.1099/00221287-74-2-289>
- García-Aljaro, C., Muñoz-Berbel, X., & Muñoz, F. J. (2009). On-chip impedimetric detection of bacteriophages in dairy samples. *Biosensors & Bioelectronics*, 24(6), 1712–1716. <https://doi.org/10.1016/j.bios.2008.08.047>
- Gratia, A. (1936). Des relations numériques entre bactéries lysogènes et particules de bactériophages. *Annales de l'Institut Pasteur*, 57. <https://orbi.uliege.be/handle/2268/267857>
- Houlberg, U., Herbsleb, P., & Sturino, J. (2005). *Method and System for Colorimetric Determination of a Chemical or Physical Property of a Turbid Medium*. <https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2005068982>
- Huang, Y., Niu, B., Gao, Y., Fu, L., & Li, W. (2010). CD-HIT Suite: A web server for clustering and comparing biological sequences. *Bioinformatics*, 26(5), 680–682. <https://doi.org/10.1093/bioinformatics/btq003>
- Hyman, P., & Abedon, S. T. (2009). Practical methods for determining phage growth parameters. *Methods in Molecular Biology (Clifton, N.J.)*, 501, 175–202. https://doi.org/10.1007/978-1-60327-164-6_18
- Ji, P., Aw, T. G., Van Bonn, W., & Rose, J. B. (2020). Evaluation of a portable nanopore-based sequencer for detection of viruses in water. *Journal of Virological Methods*, 278, 113805. <https://doi.org/10.1016/j.jviromet.2019.113805>
- Khan, M. S., Pande, T., & van de Ven, T. G. M. (2015). Qualitative and quantitative detection of T7 bacteriophages using paper based sandwich ELISA. *Colloids and Surfaces. B, Biointerfaces*, 132, 264–270. <https://doi.org/10.1016/j.colsurfb.2015.05.028>

- Klumpp, J., Fouts, D. E., & Sozhamannan, S. (2012). Next generation sequencing technologies and the changing landscape of phage genomics. *Bacteriophage*, 2(3), 190–199. <https://doi.org/10.4161/bact.22111>
- Larsson, P. A., G Puttaswamaiah, S., Ly, C., Vanerek, A., Hall, J. C., & Drolet, F. (2013). Filtration, adsorption and immunodetection of virus using polyelectrolyte multilayer-modified paper. *Colloids and Surfaces. B, Biointerfaces*, 101, 205–209. <https://doi.org/10.1016/j.colsurfb.2012.06.023>
- Lavelle, K., Martinez, I., Neve, H., Lugli, G., Franz, C., Ventura, M., Bello, F., Sinderen, D., & Mahony, J. (2018). Biodiversity of *Streptococcus thermophilus* Phages in Global Dairy Fermentations. *Viruses*, 10(10), 577. <https://doi.org/10.3390/v10100577>
- Mahony, J., Bottacini, F., van Sinderen, D., & Fitzgerald, G. F. (2014). Progress in lactic acid bacterial phage research. *Microbial Cell Factories*, 13(Suppl 1), S1. <https://doi.org/10.1186/1475-2859-13-S1-S1>
- Mahony, J., & van Sinderen, D. (2015). Novel strategies to prevent or exploit phages in fermentations, insights from phage–host interactions. *Current Opinion in Biotechnology*, 32, 8–13. <https://doi.org/10.1016/j.copbio.2014.09.006>
- Marcó, M. B., Moineau, S., & Quiberoni, A. (2012). Bacteriophages and dairy fermentations. *Bacteriophage*, 2(3), 149–158. <https://doi.org/10.4161/bact.21868>
- Marcó, M. B., Suárez, V. B., Quiberoni, A., & Pujato, S. A. (2019). Inactivation of Dairy Bacteriophages by Thermal and Chemical Treatments. *Viruses*, 11(5), 480. <https://doi.org/10.3390/v11050480>
- Marco, M. L., Heeney, D., Binda, S., Cifelli, C. J., Cotter, P. D., Foligné, B., Gänzle, M., Kort, R., Pasin, G., Pihlanto, A., Smid, E. J., & Hutkins, R. (2017). Health benefits of fermented foods: Microbiota and beyond. *Current Opinion in Biotechnology*, 44, 94–102. <https://doi.org/10.1016/j.copbio.2016.11.010>
- McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.-P., Lugli, G. A., Ventura, M., Kouwen, T. R., & van Sinderen, D. (2017). Global Survey and Genome Exploration of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Frontiers in Microbiology*, 8, 1754. <https://doi.org/10.3389/fmicb.2017.01754>
- McDonnell, B., Mahony, J., Neve, H., Hanemaaijer, L., Noben, J.-P., Kouwen, T., & van Sinderen, D. (2016). Identification and Analysis of a Novel Group of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Applied and Environmental Microbiology*, 82(17), 5153–5165. <https://doi.org/10.1128/AEM.00835-16>
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., & Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(W1), W597–W600. <https://doi.org/10.1093/nar/gkt376>
- Michelsen, O., Cuesta-Dominguez, Á., Albrechtsen, B., & Jensen, P. R. (2007). Detection of Bacteriophage-Infected Cells of *Lactococcus lactis* by Using Flow Cytometry. *Applied*

- and *Environmental Microbiology*, 73(23), 7575–7581.
<https://doi.org/10.1128/AEM.01219-07>
- Mills, S., Griffin, C., O’Sullivan, O., Coffey, A., McAuliffe, O. E., Meijer, W. C., Serrano, L. M., & Ross, R. P. (2011). A new phage on the ‘Mozzarella’ block: Bacteriophage 5093 shares a low level of homology with other *Streptococcus thermophilus* phages. *International Dairy Journal*, 21(12), 963–969.
<https://doi.org/10.1016/j.idairyj.2011.06.003>
- Morella, N. M., Yang, S. C., Hernandez, C. A., & Koskella, B. (2018). Rapid quantification of bacteriophages and their bacterial hosts in vitro and in vivo using droplet digital PCR. *Journal of Virological Methods*, 259, 18–24.
<https://doi.org/10.1016/j.jviromet.2018.05.007>
- Muhammed, M. K., Krych, L., Nielsen, D. S., & Vogensen, F. K. (2017). A high-throughput qPCR system for simultaneous quantitative detection of dairy *Lactococcus lactis* and *Leuconostoc* bacteriophages. *PloS One*, 12(3), e0174223.
<https://doi.org/10.1371/journal.pone.0174223>
- Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J. A., McQuisten, K. A., Almabrazi, H. G., Pedersen, K. F., Lin, Y., Garretson, J., McEntaggart, N. O., Sailor, C. A., Dawson, R. B., & Peek, A. S. (2008). IDT SciTools: A suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Research*, 36(Web Server issue), W163–W169.
<https://doi.org/10.1093/nar/gkn198>
- Pujato, S. A., Quiberoni, A., & Mercanti, D. J. (2019). Bacteriophages on dairy foods. *Journal of Applied Microbiology*, 126(1), 14–30. <https://doi.org/10.1111/jam.14062>
- Quiberoni, A., Tremblay, D., Ackermann, H.-W., Moineau, S., & Reinheimer, J. A. (2006). Diversity of *Streptococcus thermophilus* Phages in a Large-Production Cheese Factory in Argentina. *Journal of Dairy Science*, 89(10), 3791–3799.
[https://doi.org/10.3168/jds.S0022-0302\(06\)72420-1](https://doi.org/10.3168/jds.S0022-0302(06)72420-1)
- Serafim, V., Ring, C., Pantoja Munoz, L., Shah, H., & Shah, A. (2017). Rapid Identification of *E. coli* Bacteriophages using Mass Spectrometry. *Journal of Proteomics & Enzymology*, 06. <https://doi.org/10.4172/2470-1289.1000130>
- Štveráková, D., Šedo, O., Benešík, M., Zdráhal, Z., Doškař, J., & Pantůček, R. (2018). Rapid Identification of Intact Staphylococcal Bacteriophages Using Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometry. *Viruses*, 10(4), 176.
<https://doi.org/10.3390/v10040176>
- Svensson, U., & Christiansson, A. (1991). Methods for phage monitoring [easy to use in dairy laboratories]. *Federation Internationale de Laiterie; International Dairy Federation*. 263, 29–39.
- Szymczak, P., Janzen, T., Neves, A. R., Kot, W., Hansen, L. H., Lametsch, R., Neve, H., Franz, C. M. A. P., & Vogensen, F. K. (2017). Novel Variants of *Streptococcus thermophilus* Bacteriophages Are Indicative of Genetic Recombination among Phages from Different Bacterial Species. *Applied and Environmental Microbiology*, 83(5), e02748-16, e02748-16. <https://doi.org/10.1128/AEM.02748-16>

- Szymczak, P., Rau, M. H., Monteiro, J. M., Pinho, M. G., Filipe, S. R., Vogensen, F. K., Zeidan, A. A., & Janzen, T. (2019). A comparative genomics approach for identifying host-range determinants in *Streptococcus thermophilus* bacteriophages. *Scientific Reports*, 9(1), 7991. <https://doi.org/10.1038/s41598-019-44481-z>
- Szymczak, P., Vogensen, F. K., & Janzen, T. (2019). Novel isolates of *Streptococcus thermophilus* bacteriophages from group 5093 identified with an improved multiplex PCR typing method. *International Dairy Journal*, 91, 18–24. <https://doi.org/10.1016/j.idairyj.2018.12.001>
- Tayyarcan, E. K., Acar Soykut, E., & Boyaci, I. H. (2018). A Raman-spectroscopy-based approach for detection and discrimination of *Streptococcus thermophilus* and *Lactobacillus bulgaricus* phages at low titer in raw milk. *Folia Microbiologica*, 63(5), 627–636. <https://doi.org/10.1007/s12223-018-0604-5>
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—New capabilities and interfaces. *Nucleic Acids Research*, 40(15), e115. <https://doi.org/10.1093/nar/gks596>
- Verreault, D., Gendron, L., Rousseau, G. M., Veillette, M., Massé, D., Lindsley, W. G., Moineau, S., & Duchaine, C. (2011). Detection of Airborne Lactococcal Bacteriophages in Cheese Manufacturing Plants. *Applied and Environmental Microbiology*, 77(2), 491–497. <https://doi.org/10.1128/AEM.01391-10>
- Verreault, D., Moineau, S., & Duchaine, C. (2008). Methods for sampling of airborne viruses. *Microbiology and Molecular Biology Reviews: MMBR*, 72(3), 413–444. <https://doi.org/10.1128/MMBR.00002-08>
- Wagner, N., Brinks, E., Samtlebe, M., Hinrichs, J., Atamer, Z., Kot, W., Franz, C. M. A. P., Neve, H., & Heller, K. J. (2017). Whey powders are a rich source and excellent storage matrix for dairy bacteriophages. *International Journal of Food Microbiology*, 241, 308–317. <https://doi.org/10.1016/j.ijfoodmicro.2016.10.037>
- Wang, T., Nguyen, A., Zhang, L., & Turko, I. V. (2019). Mass spectrometry enumeration of filamentous M13 bacteriophage. *Analytical Biochemistry*, 582, 113354. <https://doi.org/10.1016/j.ab.2019.113354>
- Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., & Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13, 134. <https://doi.org/10.1186/1471-2105-13-134>
- Zago, M., Scaltriti, E., Fornasari, M. E., Rivetti, C., Grolli, S., Giraffa, G., Ramoni, R., & Carminati, D. (2012). Epifluorescence and atomic force microscopy: Two innovative applications for studying phage-host interactions in *Lactobacillus helveticus*. *Journal of Microbiological Methods*, 88(1), 41–46. <https://doi.org/10.1016/j.mimet.2011.10.006>

CAPÍTULO III

Desenvolvimento de imunoenaios para detecção de fagos que infectam *Streptococcus thermophilus*

RESUMO

O desenvolvimento de estratégias eficazes de prevenção e combate à lise por fagos que infectam *Streptococcus thermophilus* é fundamental nas indústrias de laticínios. A detecção e o monitoramento de fagos no leite, coprodutos lácteos e no ambiente industrial representam uma oportunidade para que sejam tomadas medidas de controle antes que ocorram prejuízos à qualidade dos produtos fermentados, financeiros e, eventualmente, à saúde do consumidor. Neste estudo, iniciamos o desenvolvimento de imunoenaios para detecção de fagos que infectam *S. thermophilus*, sendo um deles de fluxo lateral e o outro baseado em vesículas de polidiacetileno (PDA). Para isso, camundongos foram imunizados com suspensões de fagos com o objetivo de produzir anticorpos anti-fagos. Os soros hiperimunes coletados ao fim do protocolo de imunização apresentaram anticorpos anti-fagos e anti-bactérias, sendo necessária uma etapa de purificação para redução da resposta cruzada. Os anticorpos purificados foram então utilizados como moléculas de captura para o desenvolvimento de um imunoensaio de fluxo lateral e um imunoensaio baseado em vesículas de PDA. O imunoensaio de fluxo lateral não foi efetivamente testado em função da baixa fluidez do conjugado através da membrana de nitrocelulose. Os testes realizados com o conjugado e o anticorpo anti-IgG de camundongo e os resultados do Dot-Blot sugerem que, com os ajustes adequados, a detecção de fagos que infectam *S. thermophilus* pode ser alcançada. O imunoensaio baseado em vesículas de PDA foi efetivo para a detecção de fagos que infectam *S. thermophilus*, no entanto, o aumento da sensibilidade e a quantificação da alteração colorimétrica ainda precisam ser trabalhados. Se otimizados, os protótipos de imunoenaios aqui desenvolvidos podem se tornar uma alternativa para a detecção de fagos que infectam *S. thermophilus* na indústria de laticínios.

1. INTRODUÇÃO

O desenvolvimento de estratégias eficazes de prevenção e combate à lise por fagos que infectam *S. thermophilus* é fundamental nas indústrias de laticínios. Neste sentido, a detecção e o monitoramento de fagos no leite, coprodutos lácteos e no ambiente industrial representam uma oportunidade para que sejam tomadas medidas de controle antes que ocorram prejuízos à qualidade dos produtos fermentados, financeiros e, eventualmente, à saúde do consumidor

(Lavelle et al., 2018; McDonnell et al., 2017; Philippe & Moineau, 2021; Szymczak et al., 2017). Para esse fim, é necessária a caracterização de fagos que infectam bactérias lácticas, compreensão da dinâmica de interação com os hospedeiros, bem como o conhecimento da evolução de fagos e melhores métodos de detecção (Mahony & van Sinderen, 2015; Szymczak et al., 2017).

Dentre os inúmeros métodos de detecção, os biossensores colorimétricos têm sido amplamente utilizados devido à sua sensibilidade, versatilidade e simplicidade (Kim et al., 2021; Z. Zhang et al., 2019). O polidiacetileno (PDA) constitui uma classe de polímero conjugado (ligações duplas e triplas alternadas) que tem sido estudado em vários sistemas de detecção devido às suas propriedades ópticas únicas (Hao & Zhu, 2020; Qian & Städler, 2019; Wegner, 1969; Yadav & Tiwari, 2021). Os PDAs são formados através da 1,4-polimerização de monômeros de diacetileno automontados sob irradiação UV (254 nm) ou raios γ (Carpick et al., 2004; Wegner, 1969). A estrutura conjugada na cadeia principal do polímero permite que o PDA possua a capacidade de transição colorimétrica azul (absorção máxima em ~ 640 nm) - vermelha (absorção máxima em ~ 540 nm) quando as interações π - π na cadeia principal do PDA são perturbadas pelo ambiente externo (Kim et al., 2022; Qian & Städler, 2019). A estrutura do PDA pode ser montada na forma de filmes mono ou multicamada (Carpick et al., 2004; Geiger et al., 2002), vesículas (lipossomas) (de Oliveira et al., 2015) e outras estruturas coloidais como cristais (Park et al., 2014), tubos (Wang et al., 2018) ou hélices (Meng et al., 2017), para detecção de diferentes moléculas como bacteriocinas (Yadav & Tiwari, 2021), vírus (Roh et al., 2017; Zhao et al., 2020), bactérias (de Oliveira et al., 2015; Zhou et al., 2022), ácidos nucleicos (Jung & Park, 2015), antibióticos (Kang et al., 2012), surfactantes (Y. Zhang et al., 2017), proteínas (Jung et al., 2010) e íons metálicos (Gwon et al., 2019).

O ensaio de fluxo lateral (Lateral Flow Assay - LFA) caracteriza-se por combinar a técnica de cromatografia com uma reação imunoquímica, cujo princípio básico é o fluxo da amostra líquida contendo o analito ao longo do comprimento da fita por força de capilaridade, até que esta alcance a área de teste onde ocorre a interação antígeno-anticorpo (Koczula & Gallotta, 2016). O LFA pode ser classificado como imunoensaio de fluxo lateral (LFIA) quando a detecção é baseada na interação entre anticorpo e antígeno, e ensaio de fluxo lateral de ácido nucleico (NALFA) quando a detecção se dá pela hibridização de DNA com a sonda de DNA-alvo (Bahadır & Sezginürk, 2016). Esse ensaio é uma plataforma baseada na sobreposição de membranas de papel, amplamente utilizado para detecção quantitativa e qualitativa de diversas classes de analitos como antígenos específicos (Boisen et al., 2015), anticorpos (Nielsen et al., 2007), patógenos (Ilhan et al., 2021), ácidos nucleicos (Zheng et al., 2021), contaminantes

químicos (Ngom et al., 2010), micotoxinas (Foubert et al., 2017) e vírus (Thangavelu et al., 2022; Xiao et al., 2019).

Algumas tecnologias estão disponíveis comercialmente para a detecção de fagos (Charm Sciences®, BioVir Laboratories®, PROGEN®, Fitzgerald Industries International®, MyBioSource®, United States Biological® e Creative Diagnostics®), mas nenhuma abrange fagos que infectam *S. thermophilus*. Também não existem estudos sobre o desenvolvimento de testes imunológicos para detecção de fagos estreptocócicos. Dessa forma, o desenvolvimento de testes rápidos, práticos, sensíveis, específicos e de baixo custo para o monitoramento desses fagos, contribui para a redução de problemas tecnológicos, econômicos e de segurança nas indústrias de laticínios, além de expandir o conhecimento no campo de plataformas de sensoriamento.

2. MATERIAL E MÉTODOS

2.1. Bactérias, fagos e condições de cultivo

Fagos específicos que infectam a bactéria *S. thermophilus*, *Streptococcus* phage vB_SthS_VA214 (número de acesso no GenBank: MG708274), foram gentilmente cedidos pelo Istituto per la Qualità e le Tecnologie Agroalimentari, Veneto Agricoltura, Thiene, Itália. Como hospedeiro para a propagação do fago foi utilizada a estirpe bacteriana vB_SthS_VA214, isolada de *Fontina*, um queijo italiano de Denominação de Origem Protegida (DOP).

A estirpe bacteriana foi cultivada *overnight* a 44 °C em caldo M-17 (Sigma-Aldrich, EUA) a partir de colônias únicas ou do estoque e os fagos foram propagados no hospedeiro correspondente. Para isso, 10 mL de caldo M-17 contendo 1% (m/v) de CaCl₂ 1 M foi adicionado de 0,1 mL da suspensão bacteriana cultivada *overnight* a 44 °C, seguido por incubação da mistura a 44 °C por 1 h. Após a incubação, foi adicionado 10 µL do estoque de fago e a mistura foi novamente incubada *overnight* a 44 °C. A suspensão contendo os fagos foi filtrada usando filtro de 0,20 µm (Corning, Alemanha) e armazenada a 4 °C. O título viral foi determinado pelo método de plaqueamento em dupla camada de ágar (Adams, 1959; Gratia, 1936) e expresso em termos de unidades formadoras de placas por mililitro (UFP·mL⁻¹).

2.2. Concentração e purificação dos fagos

A suspensão contendo os fagos foi pré-concentrada por meio de uma ultrafiltração à vácuo utilizando o dispositivo Amicon Pro Purification System (Millipore, EUA) com membrana Amicon Ultra-0.5 de 100 kDa (Millipore, EUA). Em seguida, a suspensão fágica foi novamente concentrada e purificada por ultracentrifugação com gradiente de densidade de

cloreto de cério (CsCl) (Szymczak et al., 2017). Para isso, uma solução de CsCl $0,75 \text{ g}\cdot\text{mL}^{-1}$ foi adicionada à 13 mL da suspensão fágica e a mistura foi centrifugada a 40.000 rpm (rotor Beckman Ti50) durante 24 h a $4 \text{ }^\circ\text{C}$. Ao fim do processo, frações de $500 \text{ }\mu\text{L}$ do sobrenadante foram recuperadas e quantificadas pelo método do ácido bicinonínico (BCA Protein Assay kit, Boster Biological Technology). As frações com as maiores concentrações de proteínas foram avaliadas quanto à viabilidade e concentração de fagos por meio do método de plaqueamento em dupla camada de ágar (Adams, 1959; Gratia, 1936) e expresso em termos de unidades formadoras de placas por mililitro ($\text{UFP}\cdot\text{mL}^{-1}$).

2.3. Produção de anticorpos policlonais anti-fagos

Com a finalidade de obter anticorpos policlonais contra os fagos específicos para *S. thermophilus*, camundongos foram imunizados com as frações virais que apresentaram maior teor proteico. Seis camundongos adultos da linha BALB/c de 3 a 4 semanas, pesando cerca de 20 g cada, foram imunizados por inoculação subcutânea de $100 \text{ }\mu\text{g}$ com a fração fágica emulsificada com adjuvante completo de Freund (1:1 v/v) (Sigma-Aldrich, EUA). A aplicação foi repetida 7 e 14 dias após a primeira imunização. Para obtenção do soro hiperimune, os camundongos foram anestesiados em câmara de éter 25 dias após o início do protocolo de imunização e o sangue foi retirado através do plexo ocular (Dunbar & Schwoebel, 1990). O conteúdo extraído foi incubado a $37 \text{ }^\circ\text{C}$ por 4 h para coagulação completa visando a obtenção do soro. Por fim, os animais foram eutanasiados. Todo o procedimento foi aprovado pelo Comitê de Ética no Uso de Animais (CEUA) da UFV sob Protocolo número 43/2018.

2.4. Avaliação da qualidade do soro hiperimune

A presença de anticorpos anti-fagos e anti-bactéria no soro foi avaliada com o uso da técnica de Dot-Blot (Stott, 1989). Para isso, $5 \text{ }\mu\text{L}$ das suspensões purificadas de fagos ou da cultura bacteriana (controle) foram gotejadas separadamente na superfície de uma membrana de nitrocelulose com poro de $0,2 \text{ }\mu\text{m}$ (BIO-RAD, EUA), que foi mantida em temperatura ambiente até a completa secagem das gotas. Em seguida, a membrana foi incubada por 15 min em solução de bloqueio (tampão PBS [$\text{NaCl } 137 \text{ mM}$, $\text{KCl } 2,7 \text{ mM}$, $\text{Na}_2\text{HPO}_4 10 \text{ mM}$, $\text{KH}_2\text{PO}_4 1,8 \text{ mM}$], gelatina 3,0% [m/v]) e lavada duas vezes com PBS pH 7,2 acrescido de Tween 20 0,05% (m/v). Após a lavagem, a membrana foi incubada por 2 h com o soro hiperimune diluído (1:1.000) em tampão PBS pH 7,2 acrescido de gelatina 1% (m/v). Uma nova lavagem foi realizada com PBS/Tween 20. Em seguida, a membrana foi incubada por 2 h com o anticorpo secundário (anti-IgG de camundongo) (Sigma-Aldrich, EUA) conjugado com peroxidase

(HRP) diluído (1:30.000) em tampão PBS pH 7,2 acrescido de gelatina 1% (m/v). Após a última lavagem com PBS/Tween 20, a membrana foi incubada com o substrato (250 μ L 3,3'-Diaminobenzidina (DAB) 1%, 250 μ L H₂O₂, 5 mL PBS pH 7,4) e a reação foi revelada pela adição de peróxido de hidrogênio.

2.5. Purificação do soro hiperimune

Visando a redução da concentração de anticorpos anti-bactéria do soro e, conseqüentemente, a redução de reações cruzadas, foi preparado um *pool* com suspensões contendo a bactéria *S. thermophilus* vB_SthS_VA214 em diferentes fases do crescimento bacteriano (DO₆₀₀ 0,1; 0,7 e 1,20). O *pool* foi centrifugado a 5.000 \times g por 10 min a 4 °C e as células bacterianas foram ressuspensas com o soro hiperimune. Essa mistura foi deixada em repouso por 10 min para que os anticorpos anti-bactéria fossem adsorvidos nas células bacterianas. Por fim, o soro purificado contendo os anticorpos anti-fagos foi recuperado por meio de centrifugação a 5.000 \times g por 10 min a 4 °C. Esse procedimento foi realizado três vezes e a suspensão resultante foi avaliada conforme descrito no item 2.4.

2.6. Imunoensaio de fluxo lateral

2.6.1. Conjugação do anticorpo anti-fago com ouro coloidal

Os anticorpos policlonais presentes no soro hiperimune purificado foram conjugados por adsorção física às nanopartículas de ouro coloidal (AuNPs) (20 nm de diâmetro) (Sigma-Aldrich, EUA) estabilizadas em suspensão de PBS 0,1 mM (Mikawa et al., 2009). A concentração ótima de anticorpo a ser acrescido ao processo de conjugação foi determinada como descrito a seguir. Inicialmente, a concentração da solução contendo os anticorpos policlonais foi ajustada para 1 mg·mL⁻¹ utilizando tampão PBS 20 mM pH 7,4. Em seguida, 10 μ L de suspensão contendo os anticorpos em concentrações que variaram de 15,6 a 1.000 μ g·mL⁻¹ foram adicionados em 100 μ L da suspensão de ouro coloidal. A mistura foi incubada a temperatura ambiente por 15 min, seguido da adição de 10 μ L de NaCl 10 % (m/v) e incubação por 15 min. A variação de coloração da solução final foi monitorada a 540 nm em leitor de microplaca (BioTek ELx808). Foi escolhida a concentração de anticorpo na qual a solução apresentou maior absorvância.

Após a definição da concentração ótima de anticorpo, o processo de conjugação foi iniciado com a adição de 62 μ L do soro contendo os anticorpos policlonais em 1 mL de solução de ouro disperso. A mistura foi incubada à temperatura ambiente por 30 min. Em seguida, a solução do conjugado foi estabilizada com Bovine Serum Albumin - BSA (concentração final

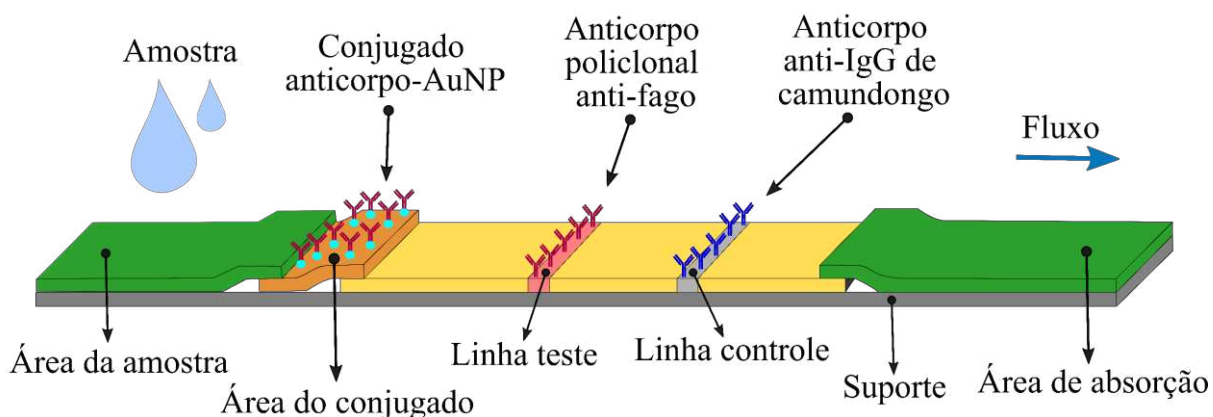
de 1% [m/v]) durante 15 min e centrifugado a $700 \times g$ por 10 min a $20 \text{ }^\circ\text{C}$. O precipitado foi descartado e o sobrenadante foi novamente centrifugado a $8.100 \times g$ por 20 min a $4 \text{ }^\circ\text{C}$ para remover os anticorpos não conjugados. Por fim, o sobrenadante foi descartado e as partículas de ouro conjugadas aos anticorpos foram ressuspensas em $100 \text{ }\mu\text{L}$ de tampão PBS $10 \text{ }\mu\text{M}$ pH 7,4 contendo sacarose 5% (m/v), BSA 0,5% (m/v), Tween 20 0,5% (v/v) e polietilenoglicol 8.000 g/mol 0,2% (m/v). A solução do conjugado ouro coloidal com o anticorpo anti-fago foi armazenada a $4 \text{ }^\circ\text{C}$.

2.6.2. Montagem da fita

Para a construção da fita do imunoenensaio de fluxo lateral foram utilizadas membranas organizadas em um sistema de sobreposição (Figura 1). Na membrana de fibra de vidro (G041, Millipore, EUA) foi fixada a solução do conjugado ouro coloidal com o anticorpo anti-fago, que tem como função a captura do antígeno (fago) caso esteja presente na amostra aplicada. Uma membrana de celulose (Sigma-Aldrich, EUA) foi utilizada na área de aplicação da amostra e na área de absorção, sendo esta responsável pelo controle do fluxo da amostra ao longo da fita. Por fim, à membrana de nitrocelulose (Hi-Flow™ Plus 135, Millipore, EUA) foi fixado, a 25 mm de sua extremidade inferior, o anticorpo de captura anti-fago. A 35 mm desta mesma fita foi fixado o anticorpo de detecção anti-IgG de camundongo, responsável pelo controle do teste.

Figura 1 - Representação esquemática da montagem da fita do imunoenensaio de fluxo lateral pela sobreposição de membranas e seus principais componentes.

● Membrana de celulose ● Membrana de fibra de vidro ● Membrana de nitrocelulose



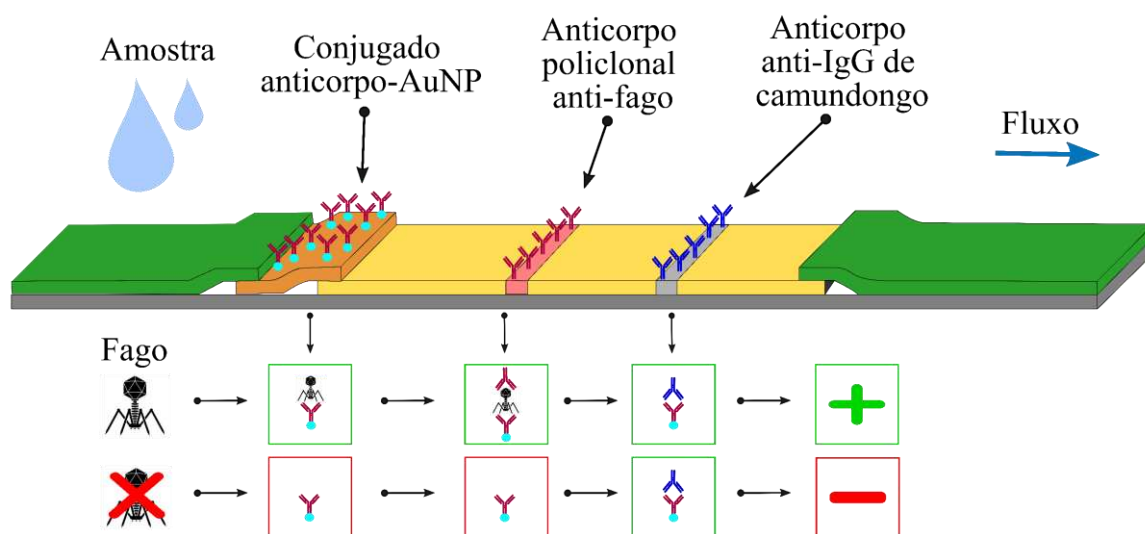
Fonte: Autora.

2.6.3. Avaliação do imunoenensaio de fluxo lateral

O sistema de membranas do imunoenensaio de fluxo lateral foi organizado sobre um suporte de plástico de 60 mm (Figura 2). Cinco microlitros da solução do conjugado ouro coloidal com o anticorpo anti-fago foram adicionados à membrana de fibra de vidro, a qual foi deixada secar à temperatura ambiente por 4 h. As membranas foram sobrepostas a partir da membrana de nitrocelulose (NC), que possuía uma região central de captura de 25 mm e duas regiões adesivas em suas extremidades. Na região adesiva da membrana de NC, definiu-se como o final da fita, fixou-se a membrana de celulose usada como almofada absorvente, garantindo que pelo menos 2 mm dessa membrana se sobrepusse na membrana de NC. Já na extremidade adesiva oposta da membrana de NC, foram adicionadas a membrana de fibra de vidro de 6 mm contendo o conjugado ouro coloidal com o anticorpo anti-fago (sobreposta na membrana de NC) e a membrana de celulose, utilizada como área de aplicação da amostra (sobreposta na membrana de fibra de vidro). Todas as membranas foram padronizadas com 5 mm de largura.

Figura 2 - Representação esquemática do imunoenensaio de fluxo lateral para detecção de fagos que infectam *Streptococcus thermophilus*.

● Membrana de celulose ● Membrana de fibra de vidro ● Membrana de nitrocelulose



Fonte: Autora.

Após a montagem da fita, foi avaliada a interação entre a solução do conjugado ouro coloidal com o anticorpo anti-fago e o anticorpo anti-IgG de camundongo como controle do imunoenensaio de fluxo lateral. Para isso, 2 μL da suspensão contendo anticorpo anti-IgG de camundongo foram aplicados sobre a membrana de NC (linha controle). Após a secagem, 200 μL de tampão PBS foi adicionado na área de aplicação da amostra para promover o fluxo do conjugado ouro coloidal com o anticorpo anti-fago até a linha controle, a fim de permitir sua

ligação aos anticorpos anti-IgG de camundongo. Para a avaliação do imunoenensaio de fluxo lateral, 2 μL da suspensão contendo anticorpo anti-IgG de camundongo foram aplicados sobre a membrana de NC (linha controle) e 2 μL da suspensão contendo anticorpo anti-fago foram aplicados sobre a membrana de NC (linha teste). Após a secagem, 200 μL da suspensão contendo fagos com título de $1,0 \times 10^6 \text{ UFP} \cdot \text{mL}^{-1}$ foram adicionados na área de aplicação da amostra.

2.7. Imunoenensaio baseado em vesículas de PDA

2.7.1. Preparação das vesículas

Os reagentes PCDA (ácido 10,12-pentacosadiinoico, Sigma-Aldrich, EUA) e DMPC (1,2-dimiristoil-sn-glicero-3-fosfocolina, Sigma-Aldrich, EUA) foram dissolvidos separadamente em clorofórmio (Merck, Brasil) até uma concentração de 1 mM. Ambas as soluções foram agitadas por 30 min a 300 *rpm* (Sartorius, Alemanha) para a completa homogeneização. As soluções de PCDA e DMPC foram reunidas em uma proporção de 3:2 (v/v) e, em seguida, o solvente foi removido por fluxo de nitrogênio. O *pellet* foi ressuspensionado em 5 mL de água ultrapura e a solução foi sonicada em desruptor de células (Unique Vira Cell, Modelo VCX 130) a uma potência de 400 W por 10 min e armazenada a 4 °C por 24 h.

2.7.2. Conjugação das vesículas com os anticorpos

As vesículas produzidas na etapa anterior foram convertidas em ésteres ativos pela adição de N-(3-dimetilaminopropil)-N'-etilcarbodiimida (EDC-HCl, Sigma-Aldrich, EUA) 2 $\text{mg} \cdot \text{mL}^{-1}$ e de N-hidroxissuccinimida (NHS, Sigma-Aldrich, EUA) 2 $\text{mg} \cdot \text{mL}^{-1}$. A suspensão foi submetida a agitação magnética a 4 °C por 3 h e, em seguida, foi adicionada a solução contendo os anticorpos anti-fagos diluída em PBS (concentração final de 1 $\mu\text{g} \cdot \text{mL}^{-1}$). Após 12 h de agitação a 4 °C, os sítios restantes de NHS foram inativados pela adição de etanolamina 2 mM sob agitação por 2 h a 4 °C. A polimerização das vesículas de PDA com os anticorpos foi desencadeada pela exposição da solução à luz ultravioleta (UV) (Spectroline) a 254 nm por 30 min. A suspensão das vesículas polimerizadas conjugadas com os anticorpos foi armazenada a 4 °C.

2.7.3. Avaliação do imunoenensaio baseado em vesículas de PDA

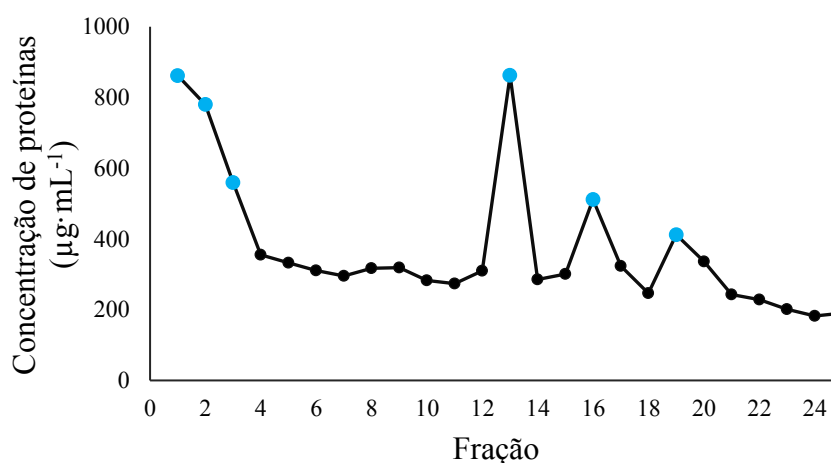
A suspensão das vesículas de PDA conjugadas com os anticorpos (900 μL) foi adicionada de 100 μL da suspensão contendo fagos com título de $1,0 \times 10^6 \text{ UFP} \cdot \text{mL}^{-1}$. A mistura foi mantida a 4 °C sob agitação por 10 min.

3. RESULTADOS

3.1. Concentração e purificação dos fagos

A pré-concentração da suspensão contendo os fagos utilizando o dispositivo Amicon resultou em um título viral de $5,0 \times 10^9$ UFP·mL⁻¹. No procedimento de ultracentrifugação, não ocorreu a formação de uma fase translúcida característica das partículas virais e, por isso, o conteúdo do tubo foi fracionado em alíquotas de 500 µL. As frações coletadas foram analisadas quanto à concentração de proteínas (Figura 3) e presença de partículas fágicas viáveis. As frações 1, 2 e 13 apresentaram as maiores concentrações de proteínas, correspondendo a 861,75 µg·mL⁻¹, 780 µg·mL⁻¹ e 862,25 µg·mL⁻¹, respectivamente. Seguidas das frações 3, 16 e 19, que exibiram concentrações de proteínas de 559 µg·mL⁻¹, 511,25 µg·mL⁻¹ e 411,75 µg·mL⁻¹, respectivamente.

Figura 3 - Concentração proteica das frações após a ultracentrifugação com CsCl. Os pontos destacados em azul representam as frações selecionadas com maior teor proteico.



Fonte: Autora.

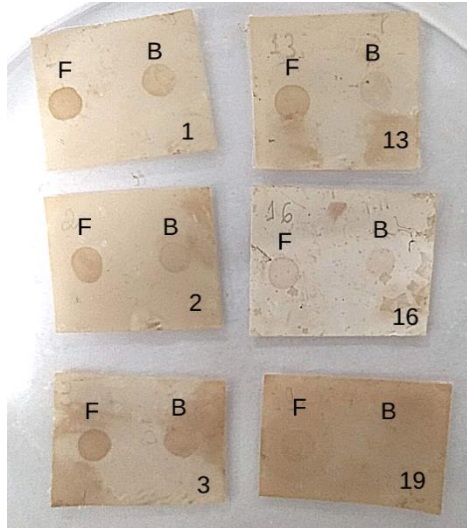
As seis frações com as maiores concentrações de proteínas apresentaram título viral entre $1,0 \times 10^2$ e $1,0 \times 10^3$ UFP·mL⁻¹, sendo utilizadas para a imunização dos camundongos.

3.2. Obtenção e avaliação do soro hiperimune

A confirmação da produção de anticorpos policlonais anti-fagos e a verificação de possível reação cruzada devido à presença de anticorpos anti-bactéria no soro hiperimune foram avaliadas pela técnica de Dot-blot. Cada um dos seis camundongos foi imunizado com uma das frações da ultracentrifugação, sendo assim, os soros foram avaliados de forma individual pela técnica de Dot-Blot (Figura 4). A presença de um halo na região onde se aplicou o antígeno

(fagos) confirmou a presença de anticorpos anti-fagos (F) nas seis amostras (1, 2, 3, 13, 16 e 19) de soro. Também foi possível observar a ocorrência de reação cruzada onde se aplicou o antígeno (bactérias), devido à presença de anticorpos anti-bactéria (B) nas seis amostras de soro.

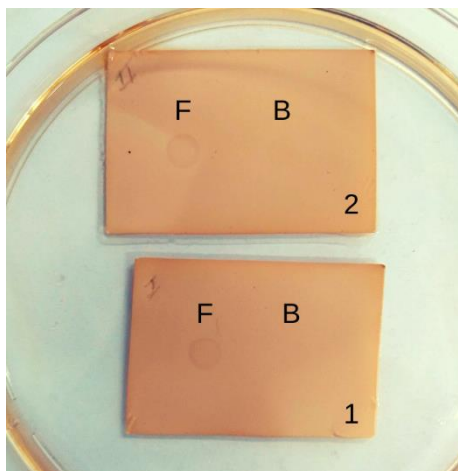
Figura 4 - Ensaio Dot-Blot das amostras de soros hiperimune confirmando a presença de anticorpos anti-fagos e anti-bactéria. Reação antígeno-anticorpo para os fagos (F); reação antígeno-anticorpo para a bactéria (B).



Fonte: Autora.

Para reduzir a concentração de anticorpos anti-bactéria, os *pellets* formados pela sedimentação de bactérias *S. thermophilus* vB_SthS_VA214 em três fases de crescimento foram ressuspensos nas amostras de soro. As amostras de soro que apresentaram maior reatividade ao fago (amostras 1, 2 e 3) foram reunidas e formaram a amostra 1. Enquanto as amostras de soro que apresentaram resultados menos reativos (amostras 13, 16 e 19) deram origem a amostra 2. Em seguida, o ensaio de Dot-Blot foi novamente realizado (Figura 5). A presença de um halo foi observada na região onde foi aplicada a suspensão contendo fagos, confirmando a reação antígeno-anticorpo para os fagos. Nenhuma marcação foi verificada na região onde se aplicou a suspensão bacteriana, confirmando a eficiência da purificação das amostras de soro.

Figura 5 - Ensaio Dot-Blot com os soros hiperimune purificados. Amostras 1, 2 e 3 (1); amostras 13, 16 e 19 (2). Reação antígeno-anticorpo para os fagos (F); reação antígeno-anticorpo para a bactéria (B).



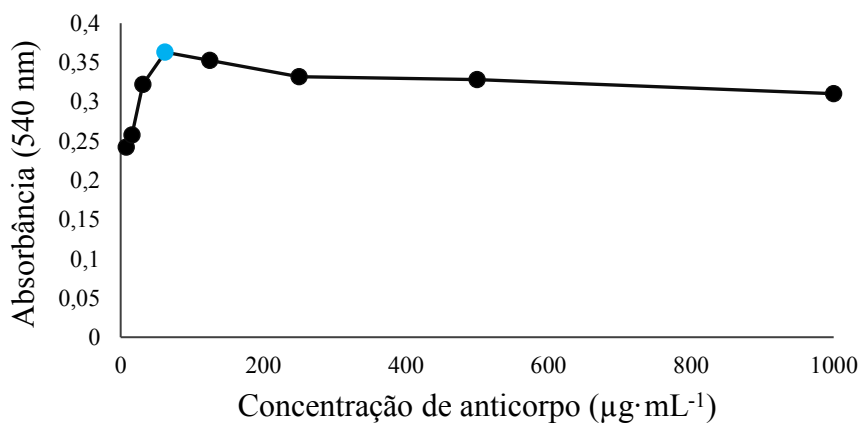
Fonte: Autora.

3.3. Imunoensaio de fluxo lateral

3.3.1. Conjugação do anticorpo anti-fago com ouro coloidal

A concentração ótima de anticorpo necessária para a conjugação com as nanopartículas de ouro coloidal (AuNPs) foi aquela em que a solução do conjugado apresentou maior absorvância, definida como $62,5 \mu\text{g}\cdot\text{mL}^{-1}$ (Figura 6). A concentração mínima de anticorpos necessária para inibir a agregação de nanopartículas de ouro na presença de excesso de sal (NaCl 10% [m/v]), é geralmente escolhida como a quantidade ideal para a conjugação com AuNPs (Safenkova et al., 2012).

Figura 6 - Variação da absorvância em função da variação da concentração de anticorpo policlonal conjugado com as AuNPs. O ponto azul corresponde à concentração ótima de anticorpo ($62,5 \mu\text{g}\cdot\text{mL}^{-1}$) para a conjugação com as AuNPs.



Fonte: Autora.

3.3.2. Avaliação do imunoenensaio de fluxo lateral

A eficácia do protocolo de conjugação foi confirmada pelo surgimento de uma marcação vermelha na região da membrana de NC onde foi aplicada a suspensão contendo anticorpo anti-IgG de camundongo (Figura 7). A mudança de cor indicou a interação entre o conjugado ouro coloidal com o anticorpo anti-fago e o anticorpo anti-IgG de camundongo.

Figura 7 - Eficácia da conjugação do anticorpo policlonal anti-fago com o ouro coloidal. A mudança de cor na membrana indica a interação entre o conjugado e o anticorpo anti-IgG de camundongo.



Fonte: Autora.

A fita do imunoenensaio de fluxo lateral foi construída pela sobreposição das membranas de NC, fibra de vidro e de celulose, e pela disposição dos elementos essenciais para o teste (Figura 8).

Figura 8 - Fita do imunoenensaio de fluxo lateral construída pela sobreposição das membranas sobre uma placa de suporte.



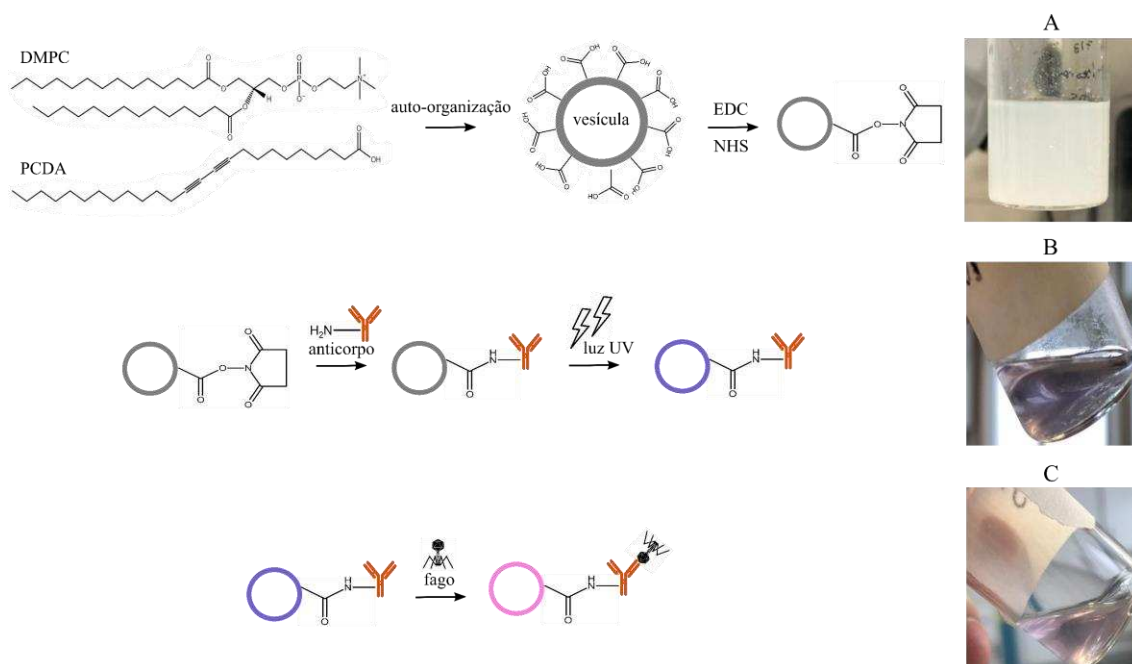
Fonte: Autora.

Após a aplicação da suspensão contendo fagos, o líquido correu por toda a fita até a área de absorção, mas o conjugado ficou retido no início da membrana de NC Hi-Flow™ Plus 135. Devido à baixa fluidez da amostra, a membrana de NC Unisart® CN 140 (Sartorius, Alemanha) foi testada em alternativa à membrana de NC Hi-Flow™ Plus 135. No entanto, o uso de tal membrana também apresentou uma baixa fluidez do conjugado ouro coloidal com o anticorpo anti-fago.

3.4. Imunoenensaio baseado em vesículas de PDA

Um imunoenensaio para detecção de bacteriófagos específicos que infectam *S. thermophilus* foi desenvolvido baseado nas propriedades de transição colorimétrica do PDA (Figura 9A). Após a exposição à luz UV, as vesículas de PDA conjugadas com os anticorpos anti-fagos sofreram polimerização formando cadeias de polímero alternadas eno-ino, resultando em uma solução de cor lilás (Figura 9B). Ao adicionar a suspensão de fagos (antígeno) à solução de PDA, foi observada uma alteração de cor do lilás para rosa claro (Figura 9C), indicando que a imunorreacção específica antígeno-anticorpo ocasionou alteração na cadeia polimérica do PDA com consequente mudança de cor da solução.

Figura 9 - Etapas de produção do imunoenensaio baseado em vesículas de PDA para detecção de fagos que infectam *Streptococcus thermophilus*. Solução de vesículas PCDA-DMPC (A); vesículas de PDA conjugadas com anticorpos anti-fagos após a polimerização com luz UV (B); detecção de fagos que infectam *Streptococcus thermophilus* com transição de cor da solução (C).



Fonte: Autora.

4. DISCUSSÃO

Neste estudo, um imunoenensaio de fluxo lateral foi parcialmente desenvolvido para detecção de fagos que infectam *S. thermophilus*. Quando a solução contendo os fagos (antígeno) foi aplicada na área de aplicação da amostra, o líquido correu por toda a fita até a área de absorção, mas o conjugado ficou retido no início da membrana de nitrocelulose. Alguns fatores podem afetar a migração do conjugado pelas membranas como a incompatibilidade entre o tamanho do conjugado e a porosidade da membrana de nitrocelulose Hi-Flow™ Plus 135 (fluxo

de 135 segundos por 4 cm de membrana). No entanto, o tamanho do conjugado foi medido (67,14 nm) por meio da técnica de dispersão dinâmica de luz (DLS) (dados não mostrados), o qual estava dentro do limite de 300 nm recomendado para moléculas a serem usadas na membrana Hi-Flow™ Plus 135. Além disso, de acordo com o manual do fabricante, a conjugação com o ouro coloidal resulta em pouca ou nenhuma restrição ao fluxo de partículas. As alterações nas características da membrana de nitrocelulose podem ser uma alternativa para melhorar o fluxo do conjugado, no entanto, o mesmo resultado foi obtido utilizando a membrana de nitrocelulose Unisart® CN 140.

Uma abordagem alternativa para o desenvolvimento do imunoenensaio de fluxo lateral, nesse caso, seria a alteração do tamanho do conjugado. As nanopartículas de ouro coloidal (AuNPs) utilizadas neste estudo foram adquiridas comercialmente com tamanho médio de 20 nm, mas essas partículas podem ser sintetizadas com diferentes diâmetros para formar o conjugado (Bastús et al., 2011). O protocolo de conjugação utilizado neste estudo se baseia em ligações hidrofóbicas, hidrofílicas e eletrostáticas entre a superfície dos anticorpos e as AuNPs para formar o conjugado. Ligações inespecíficas podem comprometer a disponibilidade dos anticorpos de detecção, considerando que o fragmento de ligação ao antígeno do anticorpo (Fab) pode se ligar às AuNPs, impossibilitando que este reconheça o antígeno. Uma alternativa para otimizar o processo de conjugação poderia ter sido a utilização da proteína A para mediar a ligação entre o anticorpo e as AuNPs. A proteína A se liga especificamente ao fragmento cristalizável do anticorpo e induz a imobilização adequadamente orientada do anticorpo, na qual os Fabs ficam acessíveis para interagir com o antígeno (Tajima et al., 2011; Tripathi & Driskell, 2018). Alternativas adicionais incluem a utilização de outros tipos de membranas e o uso de tampões com ação solubilizante, aumentando a fluidez do conjugado pela membrana e permitindo a detecção de fagos que infectam *S. thermophilus* por meio de imunoenensaio de fluxo lateral.

Para o desenvolvimento do imunoenensaio baseado em vesículas de PDA, conjugamos o éster de NHS ao PCDA por meio de uma esterificação para sintetizar o PCDA-NHS e permitir a ligação dos anticorpos anti-fagos ao monômero diacetileno (ácido 10,12-pentacoadiinóico, PCDA). Após a polimerização sob irradiação UV a 254 nm, as vesículas de PDA/anticorpos apresentaram coloração lilás, ao invés de coloração azul como amplamente relatado na literatura (Hao & Zhu, 2020; Kim et al., 2021; Qian & Städler, 2019; Z. Zhang et al., 2019). O aparecimento de cor em materiais PDA polimerizados é resultado da absorção de comprimentos de onda específicos via transição eletrônica $\pi-\pi^*$ das ligações conjugadas na estrutura do

polímero (Carpick et al., 2004). As vesículas não polimerizadas não apresentam absorção na região do espectro visível.

Após a polimerização, os PDAs aparecem na cor azul, com comprimento de onda de absorção máxima em ~640 nm, mas em alguns casos, as cores roxa, rosa, vermelha, amarela e preta podem aparecer (Okada et al., 1998). Uma série de fatores podem afetar o modo de polimerização dos monômeros de diacetileno e o aparecimento de cor azul. Se a distância entre os monômeros for muito grande ou muito pequena, a eficiência da polimerização diminui drasticamente ou a perturbação da ligação conjugada formada pela polimerização instável resulta em transições de cores incomuns, como roxo, rosa e laranja após a polimerização (Okada et al., 1998). Além da distância da repetição, a distância intermolecular de diacetileno e o ângulo de inclinação do grupo diacetileno também afetam a propagação efetiva da polimerização (Kim et al., 2021). Outro fator que poderia estar relacionado ao não aparecimento da cor azul, foi a sequência das etapas realizadas para produção das vesículas de PDA/anticorpos. Na metodologia utilizada neste trabalho, a polimerização das vesículas sob irradiação UV a 254 nm foi realizada após adição dos anticorpos anti-fagos de acordo como descrito por de Oliveira et al. (2015). Outros estudos relataram que para evitar danos aos anticorpos causados pela luz UV, as vesículas de PDA devem ser irradiadas antes da ligação dos anticorpos (Jiang et al., 2015; Patel et al., 1980).

O mecanismo de transição dos PDAs entre a fase azul e a fase vermelha ainda não foi totalmente elucidado (Chae et al., 2016; Hao & Zhu, 2020; Kim et al., 2021; Qian & Städler, 2019). Estudos indicam que uma rotação de apenas alguns graus em torno das ligações C–C altera significativamente a sobreposição do orbital π , que afeta seu estado eletrônico, causando alterações na absorção da cadeia principal do polímero (Reppy & Pindzola, 2007). Dessa forma, os PDAs sofrem uma mudança de cor de azul para vermelho (comprimento de onda de absorção máxima em ~540 nm) quando submetidos a estímulos externos (Carpick et al., 2004). Quando a solução contendo os fagos (antígeno) foi adicionada à solução com as vesículas de PDA/anticorpos, ocorreu uma imunorreação específica na superfície da vesícula, alterando a cadeia do polímero com consequente transição de cor da solução.

A mudança de cor resultante da interação entre os fagos que infectam *S. thermophilus* e as vesículas de PDA/anticorpos, confirma que o imunoenensaio desenvolvido neste estudo pode ser usado para o monitoramento e detecção de fagos em indústrias de laticínios. O desenvolvimento de um biossensor baseado em PDA para a detecção de vírus foi relatado pela primeira vez em 1993 (Charych et al., 1993) e, desde então, outros estudos demonstraram o potencial dos PDAs para aplicação de sensoriamento como biossensor (Jiang et al., 2015;

Kottappara et al., 2022; Son et al., 2019; Song et al., 2016; Zhao et al., 2020). No entanto, componentes presentes nas matrizes alimentares podem interferir em ensaios baseados em reconhecimento específico (Brussow et al., 1994; Szymczak et al., 2019). Dessa forma, etapas de preparação de amostra como purificação, centrifugação e filtração podem ser necessárias para aplicação do imunoenensaio baseado em vesículas de PDA para detecção de fagos que infectam *S. thermophilus*. Além disso, a sensibilidade do imunoenensaio baseado em vesículas de PDA desenvolvido neste estudo poderia ter sido aprimorada por meio de alterações na composição dos lipídeos, no tamanho das vesículas de PDA, na concentração de anticorpos, na proporção PDA/lipídeos das vesículas e nos tempos de incubação. Já a mudança na estrutura do PDA poderia ter sido quantificada por meio de análise óptica, como observação colorimétrica ou de fluorescência.

5. CONCLUSÃO

Neste estudo, iniciamos o desenvolvimento de imunoenensaio para detecção de fagos que infectam *S. thermophilus*. O método de obtenção de anticorpos anti-fagos foi efetivo, ainda que tenha ocorrido também a produção de anticorpos anti-bactérias. O uso de fagos purificados e metodologias de purificação de anticorpos mais sofisticadas podem ser utilizadas para eliminar a presença de anticorpos anti-bactérias.

O imunoenensaio de fluxo lateral não foi efetivamente testado em função da baixa fluidez do conjugado através da membrana de nitrocelulose. Os testes realizados com o conjugado e o anticorpo anti-IgG de camundongo e os resultados do Dot-Blot sugerem que, com os ajustes adequados, a detecção de fagos que infectam *S. thermophilus* pode ser alcançada.

O imunoenensaio baseado em vesículas de PDA foi efetivo para a detecção de fagos que infectam *S. thermophilus*, no entanto, o aumento da sensibilidade e a quantificação da alteração colorimétrica ainda precisam ser trabalhados.

Os protótipos de imunoenensaio aqui desenvolvidos, quando otimizados, podem apresentar uma alternativa promissora para a detecção de fagos que infectam *S. thermophilus* na indústria de laticínios.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Adams, M. H. (1959). *Bacteriophages*. New York, Interscience Publishers.
<http://archive.org/details/bacteriophages00adam>

- Bahadır, E. B., & Sezgintürk, M. K. (2016). Lateral flow assays: Principles, designs and labels. *TrAC Trends in Analytical Chemistry*, 82, 286–306. <https://doi.org/10.1016/j.trac.2016.06.006>
- Bastús, N. G., Comenge, J., & Puentes, V. (2011). Kinetically Controlled Seeded Growth Synthesis of Citrate-Stabilized Gold Nanoparticles of up to 200 nm: Size Focusing versus Ostwald Ripening. *Langmuir*, 27(17), 11098–11105. <https://doi.org/10.1021/la201938u>
- Boisen, M. L., Oottamasathien, D., Jones, A. B., Millett, M. M., Nelson, D. S., Bornholdt, Z. A., Fusco, M. L., Abelson, D. M., Oda, S.-I., Hartnett, J. N., Rowland, M. M., Heinrich, M. L., Akdag, M., Goba, A., Momoh, M., Fullah, M., Baimba, F., Gbokie, M., Safa, S., ... Viral Hemorrhagic Fever Consortium. (2015). Development of Prototype Filovirus Recombinant Antigen Immunoassays. *The Journal of Infectious Diseases*, 212 Suppl 2, S359-367. <https://doi.org/10.1093/infdis/jiv353>
- Brussow, H., Fremont, M., Bruttin, A., Sidoti, J., Constable, A., & Fryder, V. (1994). Detection and classification of *Streptococcus thermophilus* bacteriophages isolated from industrial milk fermentation. *Applied and Environmental Microbiology*, 60(12), 4537–4543. <https://doi.org/10.1128/AEM.60.12.4537-4543.1994>
- Carpick, R. W., Sasaki, D. Y., Marcus, M. S., Eriksson, M. A., & Burns, A. R. (2004). Polydiacetylene films: A review of recent investigations into chromogenic transitions and nanomechanical properties. *Journal of Physics: Condensed Matter*, 16(23), R679–R697. <https://doi.org/10.1088/0953-8984/16/23/R01>
- Chae, S., Lee, J. P., & Kim, J.-M. (2016). Mechanically Drawable Thermochromic and Mechanothermochromic Polydiacetylene Sensors. *Advanced Functional Materials*, 26(11), 1769–1776. <https://doi.org/10.1002/adfm.201504845>
- Charych, D. H., Nagy, J. O., Spevak, W., & Bednarski, M. D. (1993). Direct Colorimetric Detection of a Receptor-Ligand Interaction by a Polymerized Bilayer Assembly. *Science*, 261(5121), 585–588. <https://doi.org/10.1126/science.8342021>
- de Oliveira, T. V., Soares, N. de F. F., Coimbra, J. S. dos R., de Andrade, N. J., Moura, L. G., Medeiros, E. A. A., & de Medeiros, H. S. (2015). Stability and sensitivity of polydiacetylene vesicles to detect *Salmonella*. *Sensors and Actuators B: Chemical*, 221, 653–658. <https://doi.org/10.1016/j.snb.2015.06.130>
- Dunbar, B. S., & Schwoebel, E. D. (1990). Preparation of polyclonal antibodies. *Methods in Enzymology*, 182, 663–670. [https://doi.org/10.1016/0076-6879\(90\)82051-3](https://doi.org/10.1016/0076-6879(90)82051-3)
- Foubert, A., Beloglazova, N. V., Gordienko, A., Tessier, M. D., Drijvers, E., Hens, Z., & De Saeger, S. (2017). Development of a Rainbow Lateral Flow Immunoassay for the Simultaneous Detection of Four Mycotoxins. *Journal of Agricultural and Food Chemistry*, 65(33), 7121–7130. <https://doi.org/10.1021/acs.jafc.6b04157>
- Geiger, E., Hug, P., & Keller, B. A. (2002). Chromatic Transitions in Polydiacetylene Langmuir-Blodgett Films due to Molecular Recognition at the Film Surface Studied by Spectroscopic Methods and Surface Analysis. *Macromolecular Chemistry and Physics*, 203(17), 2422–2431. <https://doi.org/10.1002/macp.200290025>

- Gratia, A. (1936). Des relations numériques entre bactéries lysogènes et particules de bactériophages. *Annales de l'Institut Pasteur*, 57. <https://orbi.uliege.be/handle/2268/267857>
- Gwon, Y. J., Kim, C., & Lee, T. S. (2019). Chromatic detection of Cs ions using polydiacetylene-based vesicles containing crown-ether-like ethylene glycol units. *Sensors and Actuators B: Chemical*, 281, 343–349. <https://doi.org/10.1016/j.snb.2018.10.125>
- Hao, Y.-J., & Zhu, G.-M. (2020). Advances in Fabrication of Polydiacetylene Vesicles and Their Applications in Medical Detection. *Chinese Journal of Analytical Chemistry*, 48(2), 164–173. [https://doi.org/10.1016/S1872-2040\(19\)61213-2](https://doi.org/10.1016/S1872-2040(19)61213-2)
- Ilhan, H., Tayyarcı, E. K., Caglayan, M. G., Boyacı, İ. H., Sağlam, N., & Tamer, U. (2021). Replacement of antibodies with bacteriophages in lateral flow assay of *Salmonella* Enteritidis. *Biosensors and Bioelectronics*, 189, 113383. <https://doi.org/10.1016/j.bios.2021.113383>
- Jiang, L., Luo, J., Dong, W., Wang, C., Jin, W., Xia, Y., Wang, H., Ding, H., Jiang, L., & He, H. (2015). Development and evaluation of a polydiacetylene based biosensor for the detection of H5 influenza virus. *Journal of Virological Methods*, 219, 38–45. <https://doi.org/10.1016/j.jviromet.2015.03.013>
- Jung, Y. K., Kim, T. W., Park, H. G., & Soh, H. T. (2010). Specific Colorimetric Detection of Proteins Using Bidentate Aptamer-Conjugated Polydiacetylene (PDA) Liposomes. *Advanced Functional Materials*, 20(18), 3092–3097. <https://doi.org/10.1002/adfm.201001008>
- Jung, Y. K., & Park, H. G. (2015). Colorimetric detection of clinical DNA samples using an intercalator-conjugated polydiacetylene sensor. *Biosensors & Bioelectronics*, 72, 127–132. <https://doi.org/10.1016/j.bios.2015.04.093>
- Kang, D. H., Jung, H.-S., Ahn, N., Lee, J., Seo, S., Suh, K.-Y., Kim, J., & Kim, K. (2012). Biomimetic detection of aminoglycosidic antibiotics using polydiacetylene–phospholipids supramolecules. *Chemical Communications*, 48(43), 5313–5315. <https://doi.org/10.1039/C2CC31466E>
- Kim, C., Han, B. H., Kim, D., Lee, G., Hong, C., Kang, J. Y., & Lee, K. (2022). Multi-target polydiacetylene liposome-based biosensor for improved exosome detection. *Sensors and Actuators B: Chemical*, 355, 131286. <https://doi.org/10.1016/j.snb.2021.131286>
- Kim, C., Hong, C., & Lee, K. (2021). Structures and strategies for enhanced sensitivity of polydiacetylene (PDA) based biosensor platforms. *Biosensors and Bioelectronics*, 181, 113120. <https://doi.org/10.1016/j.bios.2021.113120>
- Koczula, K. M., & Gallotta, A. (2016). Lateral flow assays. *Essays in Biochemistry*, 60(1), 111–120. <https://doi.org/10.1042/EBC20150012>
- Kottappara, R., Paravannoor, A., & Vijayan, B. K. (2022). 25—Nanosensors for virus detection. Em A. Denizli, T. A. Nguyen, S. Rajendran, G. Yasin, & A. K. Nadda (Orgs.), *Nanosensors for Smart Agriculture* (p. 531–546). Elsevier. <https://doi.org/10.1016/B978-0-12-824554-5.00010-0>

- Lavelle, K., Martinez, I., Neve, H., Lugli, G., Franz, C., Ventura, M., Bello, F., Sinderen, D., & Mahony, J. (2018). Biodiversity of *Streptococcus thermophilus* Phages in Global Dairy Fermentations. *Viruses*, *10*(10), 577. <https://doi.org/10.3390/v10100577>
- Mahony, J., & van Sinderen, D. (2015). Novel strategies to prevent or exploit phages in fermentations, insights from phage–host interactions. *Current Opinion in Biotechnology*, *32*, 8–13. <https://doi.org/10.1016/j.copbio.2014.09.006>
- McDonnell, B., Mahony, J., Hanemaaijer, L., Neve, H., Noben, J.-P., Lugli, G. A., Ventura, M., Kouwen, T. R., & van Sinderen, D. (2017). Global Survey and Genome Exploration of Bacteriophages Infecting the Lactic Acid Bacterium *Streptococcus thermophilus*. *Frontiers in Microbiology*, *8*, 1754. <https://doi.org/10.3389/fmicb.2017.01754>
- Meng, Y., Jiang, J., & Liu, M. (2017). Self-assembled nanohelix from a bolaamphiphilic diacetylene via hydrogelation and selective responsiveness towards amino acids and nucleobases. *Nanoscale*, *9*(21), 7199–7206. <https://doi.org/10.1039/C7NR02126G>
- Mikawa, A. Y., Santos, S. A. T., Kenfe, F. R., da Silva, F. H., & da Costa, P. I. (2009). Development of a rapid one-step immunochromatographic assay for HCV core antigen detection. *Journal of Virological Methods*, *158*(1), 160–164. <https://doi.org/10.1016/j.jviromet.2009.02.013>
- Ngom, B., Guo, Y., Wang, X., & Bi, D. (2010). Development and application of lateral flow test strip technology for detection of infectious agents and chemical contaminants: A review. *Analytical and Bioanalytical Chemistry*, *397*(3), 1113–1135. <https://doi.org/10.1007/s00216-010-3661-4>
- Nielsen, K., Yu, W. L., Kelly, L., Bermudez, R., Renteria, T., Dajer, A., Gutierrez, E., Williams, J., Algire, J., & Torioni de Eschaide, S. (2007). Development of a Lateral Flow Assay for Rapid Detection of Bovine Antibody to *Anaplasma marginale*. *Journal of Immunoassay and Immunochemistry*, *29*(1), 10–18. <https://doi.org/10.1080/15321810701734693>
- Okada, S., Peng, S., Spevak, W., & Charych, D. (1998). Color and Chromism of Polydiacetylene Vesicles. *Accounts of Chemical Research*, *31*(5), 229–239. <https://doi.org/10.1021/ar970063v>
- Park, D.-H., Hong, J., Park, I. S., Lee, C. W., & Kim, J.-M. (2014). A Colorimetric Hydrocarbon Sensor Employing a Swelling-Induced Mechanochromic Polydiacetylene. *Advanced Functional Materials*, *24*(33), 5186–5193. <https://doi.org/10.1002/adfm.201400779>
- Patel, G. N., Witt, J. D., & Khanna, Y. P. (1980). Thermo-chromism in polydiacetylene solutions. *Journal of Polymer Science: Polymer Physics Edition*, *18*(6), 1383–1391. <https://doi.org/10.1002/pol.1980.180180616>
- Philippe, C., & Moineau, S. (2021). The endless battle between phages and CRISPR-Cas systems in *Streptococcus thermophilus*. *Biochemistry and Cell Biology*, bcb-2020-0593. <https://doi.org/10.1139/bcb-2020-0593>

- Qian, X., & Städler, B. (2019). Recent Developments in Polydiacetylene-Based Sensors. *Chemistry of Materials*, *31*(4), 1196–1222. <https://doi.org/10.1021/acs.chemmater.8b05185>
- Reppy, M. A., & Pindzola, B. A. (2007). Biosensing with polydiacetylene materials: Structures, optical properties and applications. *Chemical Communications*, *42*, 4317–4338. <https://doi.org/10.1039/B703691D>
- Roh, J., Lee, S. Y., Park, S., & Ahn, D. J. (2017). Polydiacetylene/Anti-HBs Complexes for Visible and Fluorescent Detection of Hepatitis B Surface Antigen on a Nitrocellulose Membrane. *Chemistry, an Asian Journal*, *12*(16), 2033–2037. <https://doi.org/10.1002/asia.201700769>
- Safenkova, I., Zherdev, A., & Dzantiev, B. (2012). Factors influencing the detection limit of the lateral-flow sandwich immunoassay: A case study with potato virus X. *Analytical and Bioanalytical Chemistry*, *403*(6), 1595–1605. <https://doi.org/10.1007/s00216-012-5985-8>
- Son, S. U., Seo, S. B., Jang, S., Choi, J., Lim, J., Lee, D. K., Kim, H., Seo, S., Kang, T., Jung, J., & Lim, E.-K. (2019). Naked-eye detection of pandemic influenza a (pH1N1) virus by polydiacetylene (PDA)-based paper sensor as a point-of-care diagnostic platform. *Sensors and Actuators B: Chemical*, *291*, 257–265. <https://doi.org/10.1016/j.snb.2019.04.081>
- Song, S., Ha, K., Guk, K., Hwang, S.-G., Choi, J. M., Kang, T., Bae, P., Jung, J., & Lim, E.-K. (2016). Colorimetric detection of influenza A (H1N1) virus by a peptide-functionalized polydiacetylene (PEP-PDA) nanosensor. *RSC Advances*, *6*(54), 48566–48570. <https://doi.org/10.1039/C6RA06689E>
- Stott, D. I. (1989). Immunoblotting and dot blotting. *Journal of Immunological Methods*, *119*(2), 153–187. [https://doi.org/10.1016/0022-1759\(89\)90394-3](https://doi.org/10.1016/0022-1759(89)90394-3)
- Szymczak, P., Janzen, T., Neves, A. R., Kot, W., Hansen, L. H., Lametsch, R., Neve, H., Franz, C. M. A. P., & Vogensen, F. K. (2017). Novel Variants of *Streptococcus thermophilus* Bacteriophages Are Indicative of Genetic Recombination among Phages from Different Bacterial Species. *Applied and Environmental Microbiology*, *83*(5), e02748-16, e02748-16. <https://doi.org/10.1128/AEM.02748-16>
- Szymczak, P., Vogensen, F. K., & Janzen, T. (2019). Novel isolates of *Streptococcus thermophilus* bacteriophages from group 5093 identified with an improved multiplex PCR typing method. *International Dairy Journal*, *91*, 18–24. <https://doi.org/10.1016/j.idairyj.2018.12.001>
- Tajima, N., Takai, M., & Ishihara, K. (2011). Significance of antibody orientation unraveled: Well-oriented antibodies recorded high binding affinity. *Analytical Chemistry*, *83*(6), 1969–1976. <https://doi.org/10.1021/ac1026786>
- Thangavelu, R. M., Kadirvel, N., Balasubramaniam, P., & Viswanathan, R. (2022). Ultrasensitive nano-gold labelled, duplex lateral flow immunochromatographic assay for early detection of sugarcane mosaic viruses. *Scientific Reports*, *12*, 4144. <https://doi.org/10.1038/s41598-022-07950-6>

- Tripathi, K., & Driskell, J. D. (2018). Quantifying Bound and Active Antibodies Conjugated to Gold Nanoparticles: A Comprehensive and Robust Approach To Evaluate Immobilization Chemistry. *ACS Omega*, 3(7), 8253–8259. <https://doi.org/10.1021/acsomega.8b00591>
- Wang, M., Yu, Y., Liu, F., Ren, L., Zhang, Q., & Zou, G. (2018). Single polydiacetylene microtube waveguide platform for discriminating microRNA-215 expression levels in clinical gastric cancerous, paracancerous and normal tissues. *Talanta*, 188, 27–34. <https://doi.org/10.1016/j.talanta.2018.05.049>
- Wegner, G. (1969). Topochemische Reaktionen von Monomeren mit konjugierten Dreifachbindungen / Tochemical Reactions of Monomers with conjugated triple Bonds: I. Mitt.: Polymerisation von Derivaten des 2.4-Hexadiin-1.6-diols im kristallinen Zustand. *Zeitschrift Für Naturforschung B*, 24(7), 824–832. <https://doi.org/10.1515/znb-1969-0708>
- Xiao, M., Xie, K., Dong, X., Wang, L., Huang, C., Xu, F., Xiao, W., Jin, M., Huang, B., & Tang, Y. (2019). Ultrasensitive detection of avian influenza A (H7N9) virus using surface-enhanced Raman scattering-based lateral flow immunoassay strips. *Analytica Chimica Acta*, 1053, 139–147. <https://doi.org/10.1016/j.aca.2018.11.056>
- Yadav, M. K., & Tiwari, S. K. (2021). Polydiacetylene vesicles acting as colorimetric sensor for the detection of plantaricin LD1. *Analytical Biochemistry*, 631, 114368. <https://doi.org/10.1016/j.ab.2021.114368>
- Zhang, Y., Northcutt, J., Hanks, T., Miller, I., Pennington, B., Jelinek, R., Han, I., & Dawson, P. (2017). Polydiacetylene sensor interaction with food sanitizers and surfactants. *Food Chemistry*, 221, 515–520. <https://doi.org/10.1016/j.foodchem.2016.09.168>
- Zhang, Z., Wang, F., & Chen, X. (2019). Recent advances in the development of polydiacetylene-based biosensors. *Chinese Chemical Letters*, 30(10), 1745–1757. <https://doi.org/10.1016/j.ccllet.2019.08.035>
- Zhao, V. X. T., Wong, T. I., Zheng, X. T., Tan, Y. N., & Zhou, X. (2020). Colorimetric biosensors for point-of-care virus detections. *Materials Science for Energy Technologies*, 3, 237–249. <https://doi.org/10.1016/j.mset.2019.10.002>
- Zheng, C., Wang, K., Zheng, W., Cheng, Y., Li, T., Cao, B., Jin, Q., & Cui, D. (2021). Rapid developments in lateral flow immunoassay for nucleic acid detection. *The Analyst*, 146(5), 1514–1528. <https://doi.org/10.1039/d0an02150d>
- Zhou, J., Duan, M., Huang, D., Shao, H., Zhou, Y., & Fan, Y. (2022). Label-free visible colorimetric biosensor for detection of multiple pathogenic bacteria based on engineered polydiacetylene liposomes. *Journal of Colloid and Interface Science*, 606, 1684–1694. <https://doi.org/10.1016/j.jcis.2021.07.155>

CONCLUSÕES GERAIS

Os diferentes capítulos nos permitiram concluir que:

- Os fagos que infectam *S. thermophilus* estão organizados em 142 espécies e cinco gêneros (três deles novos), e que devido à sua diversidade genética, a classificação em nível de família varia de acordo com os critérios de classificação utilizados.
- Não foram identificados genes significativamente conservados entre os 183 genomas avaliados, entretanto, os genes que codificam a proteína holina foram conservados em mais de 95% dos genomas.
- A análise de holinas sugere que pelo menos duas α -hélices são necessárias para a função da proteína dentro dos fagos que infectam *S. thermophilus*.
- Com base na reação de PCR multiplex e na conservação dos genomas reunidos nos *clusters*, os pares de *primers* desenhados neste estudo foram capazes de detectar 53 (85,48%) dos 62 fagos disponíveis até 2018.
- Os conjuntos de *primers* otimizados foram analisados *in silico* e apresentaram potencial para detecção de 59 dos 62 fagos disponíveis, o que consiste em uma melhoria significativa em comparação à baixa especificidade do conjunto de *primers* de referência.
- O método de obtenção de anticorpos anti-fagos foi efetivo, ainda que tenha ocorrido também a produção de anticorpos anti-bactérias, os quais foram separados do soro com sucesso por meio de uma etapa de purificação.
- O imunoensaio de fluxo lateral não foi efetivamente testado em função da baixa fluidez do conjugado através da membrana de nitrocelulose. Os testes realizados com o conjugado e o anticorpo anti-IgG de camundongo e os resultados do Dot-Blot sugerem que, com os ajustes adequados, a detecção de fagos que infectam *S. thermophilus* pode ser alcançada.
- O imunoensaio baseado em vesículas de PDA foi efetivo para a detecção de fagos que infectam *S. thermophilus*, no entanto, o aumento da sensibilidade e a quantificação da alteração colorimétrica ainda precisam ser trabalhados.
- Se otimizados, os protótipos de imunoensaios desenvolvidos neste estudo podem apresentar uma alternativa promissora para a detecção de fagos que infectam *S. thermophilus* na indústria de laticínios.